

Le projet IPFC-espagnol: développement et perspectives

Isabelle Racine

Ecole de langue et de civilisation françaises
Université de Genève

TUFS - IPFC2011-Tokyo
15 mars 2011



UNIVERSITÉ DE GENÈVE



Plan

- Introduction: le projet IPFC-espagnol
- Aspects méthodologiques
 - La question de la transcription des données du corpus
- Premières études:
 - La question des occlusives sonores
 - *L'accentuation* ⇒ *présentation S. Schwab*
 - *Les voyelles arrondies* ⇒ *présentation I. Racine (17 mars)*
- Conclusion
- Perspectives



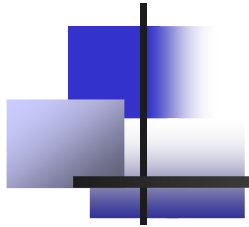
Le projet IPFC-espagnol

- 1^{er} objectif: constituer un large corpus de données orales d'apprenants hispanophones de français en utilisant un protocole et des outils identiques (IPFC).
- Ce protocole a été adapté à partir de celui du projet PFC (« Phonologie du français contemporain », Durand, Laks & Lyche 2002, 2005, 2009, <http://www.projet-pfc.net>) et inclut 6 tâches différentes :
 - Répétition d'une liste spécifique de mots lue par un natif
 - Lecture de la liste PFC
 - Lecture de la liste spécifique
 - Lecture du texte PFC
 - Entretien guidé
 - Interaction entre deux apprenants
- 2^{ème} objectif: examiner des questions méthodologiques au niveau de l'articulation entre linguistique de corpus et méthodologie utilisée par la psycholinguistique.



Le projet IPFC-espagnol

- Corpus actuel:
 - 14 apprenants avancés (B2-C1 du CECRL) de français à Genève (hispanophones d'Espagne)
 - 5 apprenants avancés de français à Madrid (hispanophones d'Espagne)
 - Centres d'intérêts:
 - Voyelles nasales (Racine et al. 2010)
 - Voyelles arrondies (Racine, 2011)
 - Consonnes occlusives sonores + /v/ (Racine et al. sous presse)
 - Accentuation (Schwab & Llisterri 2010, Schwab 2011)
 - Schwa et liaison
- + aspects méthodologiques: transcription et codage (Racine et al. sous presse)



Aspects méthodologiques: La transcription des données du corpus



Aspects méthodologiques – la transcription des données

- Selon Delais-Roussarie (2009):
«Transcrire des données sonores consiste à fournir une représentation symbolique du signal. Cette représentation n'est pas équivalente au signal, dans la mesure où elle est le résultat d'une analyse, ou plutôt d'une abstraction, des données réelles».
 - Elle distingue 4 grandes difficultés liées à cette analyse du signal:
 - 1) Les problèmes liés à la difficulté d'écoute
 - 2) Les difficultés résultant de la reconstruction perceptive effectuée obligatoirement par le transcripateur
 - 3) Le poids des préjugés linguistiques (ex. «ne» de négation)
 - 4) Les ambiguïtés liées au code oral (ex. «il mange» / «ils mangent»)
- ⇒ Nécessite l'établissement de conventions de transcriptions



Aspects méthodologiques – la transcription des données

- La difficulté liée à la **reconstruction perceptive** effectuée par le transcripateur devient prépondérante lorsqu'il s'agit de transcrire des données d'apprenants.
- Le taux de désaccord entre transcripateurs augmente considérablement lorsqu'il s'agit de données non natives (10 à 34%) vs données natives (5%) (Zechner, 2009).
 - ⇒ Impossible d'éviter un certain degré d'interprétation!
- 2 impératifs:
 - Nécessité de lisibilité des données ⇒ besoin d'un niveau orthographique simple
 - Nécessité de fidélité ⇒ dans le domaine de la phonologie, besoin d'un système permettant de rendre compte des réalisations des apprenants

⇒ Quelques exemples tirés du corpus IPFC-espagnol...



Niveau orthographique

- La transcription orthographique des conversations doit-elle refléter la morphologie du transcripteur ou celle du locuteur?
- « [ilævenu] » : - généralisation de l'auxiliaire « avoir »?
- « être » avec réalisation déviante?
- « les gens (...) ils habitent (...)
ils ne connaient pas \ il ne connaît pas




Niveau orthographique

- L'exigence de lisibilité ainsi que la compréhension globale du message passent obligatoirement par une forme d'interprétation des données orales.
- Dans la plupart des cas, une transcription orthographique peut être fournie, ce qui permet d'éviter de:
 - multiplier les segments inaudibles
 - mêler API et orthographe standard

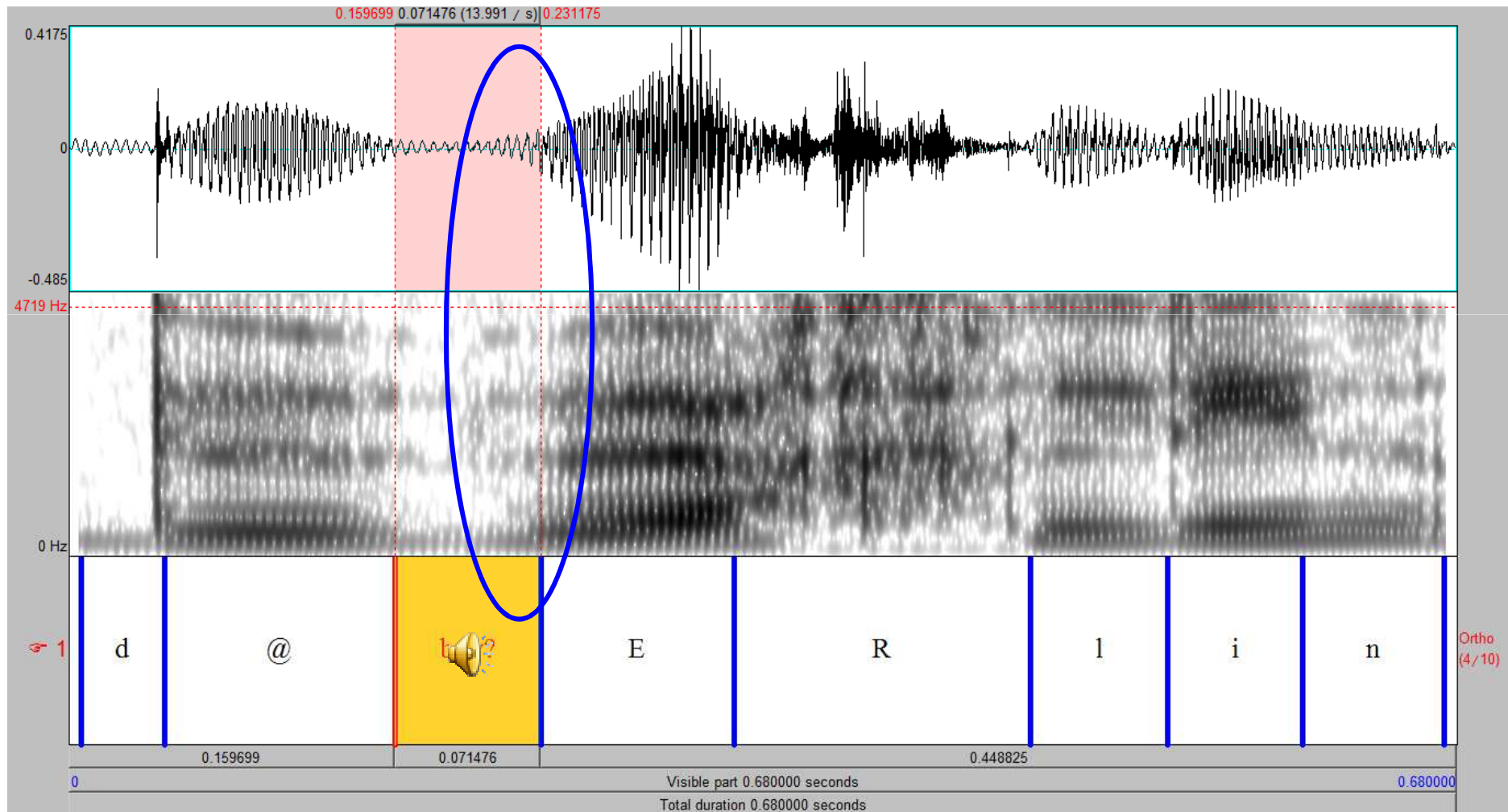
mais ce qui oblige à:

- indiquer les interprétations multiples ou peu fiables
- décider si c'est le plan morphologique ou phonologique qui est en jeu

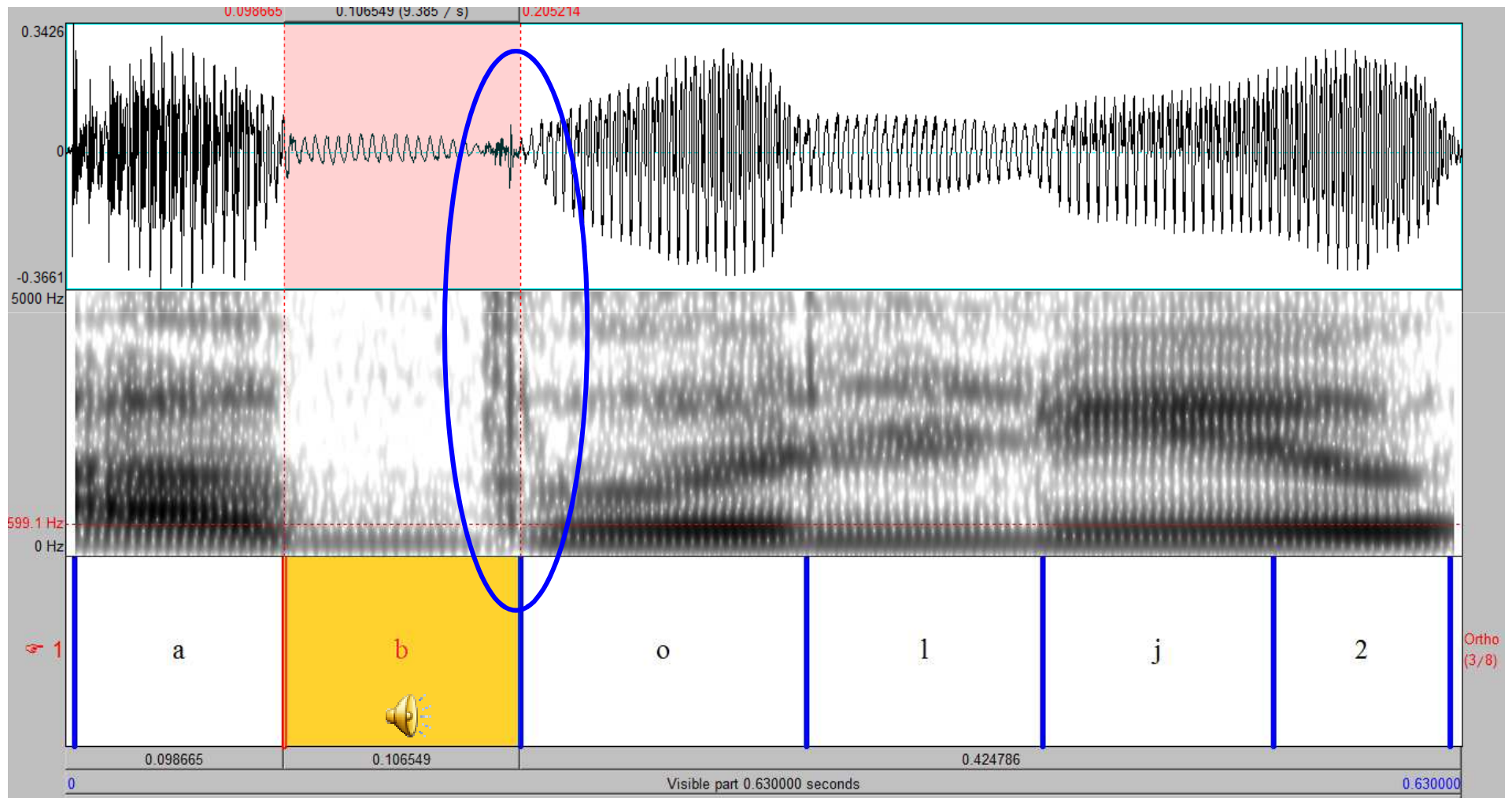
Niveau phonético-phonologique

- Transcription phonétique ⇒ question peu traitée jusqu'à récemment avec l'apparition de corpus intégrant ce type de transcriptions (cf. Neri et al., 2006; Cylwik et al., 2009; Gut, 2009; Visceglia et al., 2009; Delais-Roussarie & Yoo, 2010; Pillot-Loiseau et al. 2010).
- Unanimité: travail coûteux en temps, fastidieux et qui requiert des compétences spécifiques (cf. Gut, 2009; Makino, 2007; Wester et al., 2001).
- Problème principal: **subjectivité** due à la reconstruction perceptive (« filtre phonologique »).
- Un exemple:
« 4ème aux Jeux Olympiques de Berlin en 1936 » 
⇒ quelle transcription phonétique?

Niveau phonético-phonologique



Niveau phonético-phonologique





« 4^{ème} aux Jeux Olympiques de Berlin en 1936 »



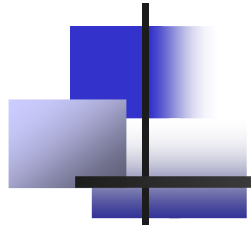
- ⇒ problème d'identification d'une différence phonétique fine liée à la reconstruction perceptive effectuée par le transcripteur mais que, une fois identifiée, on peut transcrire à priori en ayant recours à l'API de la L1 ([β]) après avoir vérifié le signal sonore (présence ou non d'une barre d'explosion).



Niveau phonético-phonologique

- La subjectivité induite par la double catégorisation effectuée par:
 - « l'oreille » du/des transcrip-teur(s)
 - la représentation catégorielle qui découle de l'usage d'un alphabet phonétique.pose problème.
- Il semble donc nécessaire:
 - de s'appuyer sur la visualisation du signal sonore
 - de recourir à plusieurs transcrip-teurs
 - que ceux-ci connaissent les deux systèmes en jeu
 - de tester la perception (= catégorisation phonémique) dans une étape ultérieure par le biais de tests perceptifs effectués par des sujets natifs non spécialistes (cf. Strange et al. 2005; Racine et al., 2010).

↪ **Tâche de transcription phonétique extrêmement lourde!**



Premières études: La question des occlusives sonores

Un exemple d'interférence

Quel est le mot?



l'arabe



dourer



adourer

⇒ **Confusion [g, γ, ʁ]**





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).

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>

Un exemple d'interférence

- Mais cela engendre des interférences entre [g, γ, ʁ]:

- [g] est réalisé [γ]

⇒ possibles difficultés de compréhension



- [ʁ] est réalisé [γ]

⇒ Quelle catégorisation?



⇒ Différence hors contexte/en contexte?

- [ʁ] est réalisé [g]



⇒ compréhension très difficile voire impossible à comprendre

- Problème lorsque [g] et [ʁ] sont proches



- Problème dans le groupe [gʁ]





Un exemple d'interférence

- 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: [ɤ] 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: [ɤ] par [ɣ].

Perspectives:

Décisions à prendre au niveau de la transcription du corpus:

- Viser un niveau phonologique?
- Viser un niveau phonétique?
- Les deux?
















Elaboration de tests perceptifs (cf. méthodologie utilisée pour l'évaluation des voyelles nasales et arrondies)



La question du /b/ et du /v/

- Systèmes de la L1 et de la L2:
 - En espagnol:
 - Un phonème avec deux réalisations allophoniques:
 - /b/ réalisé [β] sauf après pause ou consonne nasale
Ex. vaca [**b**aka] vs una vaca [unaβaka]
 - Pas de fricative labiodentale sonore (/v/)
 - A l'écrit, et <v> coexistent pour des raisons étymologiques et sans correspondance terme à terme avec les deux réalisations allophoniques
 - En français:
 - 2 phonèmes:
 - Une occlusive bilabiale sonore: /b/ (écrit , p. ex. « boire »)
 - Une fricative labiodentale sonore: /v/ (écrit <v>, p. ex. « voir »)

La question du /b/ et du /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 ou erreurs avec autocorrection 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.**



Conclusion

- Point commun de toutes les premières études (voyelles nasales et arrondies et observations sur les occlusives sonores):

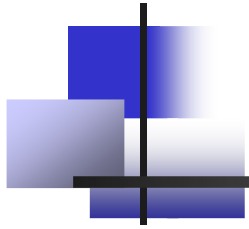
**Les systèmes interphonologiques des apprenants
diffèrent en fonction de la tâche**

- Deux conclusions:
 - Au niveau méthodologique: nécessité d'étudier la phonologie L2 avec une méthodologie incluant des tâches multiples
 - Au niveau pédagogique: nécessité de fournir un environnement d'apprentissage équilibré incluant un input à la fois oral et écrit



IPFC-espagnol - Perspectives

- Corpus:
 - Terminer la constitution des 2 corpus (20 apprenants par région)
 - Aspects méthodologiques:
 - Finaliser les conventions de transcription
 - Transcription phonétique ⇒ système de codage?
 - Etudes:
 - Poursuivre le travail sur les voyelles nasales et arrondies
 - Approfondir les observations sur les occlusives sonores
 - Développer une méthodologie pour l'étude de l'accentuation
 - Examiner les réalisations de la liaison et du schwa
- avec une méthodologie similaire pour chaque thème étudié:
- Analyse des productions des différentes tâches du protocole
 - Expériences pour évaluer la perception des productions par des natifs.
- Examiner les résultats à la lumière des modèles théoriques existants.



Merci de votre attention!



Références

- Companys, E. (1966). *Phonétique française pour hispanophones*, Paris : Hachette/Larousse.
- Cylwik, N., Wagner, A. & Demenko, G. (2009). The EURONOUNCE corpus of non-native Polish for ASR-based Pronunciation Tutoring System. *Proceedings of SlaTE 2009 – 2009 ISCA Workshop on Speech and Language Technology in Education*. Birmingham, UK.
- Delais-Roussarie, E. (2009). *Conventions CHAT de Transcription des données*. Document interne, BDD Interlangue, janvier 2009.
- Delais-Roussarie, E. & Yoo, H. (2010). The COREIL corpus: a learner corpus designed for studying phrasal phonology and intonation. *Proceedings of New Sounds 2010*, 3-5 May 2010, Poznan.
- 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.
- Durand, J., Laks, B., Lyche, C. 2009. Le projet PFC: une source de données primaires structurées. In: Durand, J., Laks, B., Lyche, C. (eds), *Phonologie, variation et accents du français*. Paris: Hermès. 19-61.
- Gut, U. (2009). *Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Wien: Peter Lang.
- Makino, Takehiko, 2007, « A corpus of Japanese speakers' pronunciation of American English : preliminary research », *PTLC 2007*, UCL, London.
- Meng, , Tseng, , Kondo, , Harrison, & Viselgia, (2009). Studying L2 suprasegmental features in Asian Englishes: a position paper. *Proceedings of Interspeech 2009*, Brighton, R-U.
- Neri, A., Cucchiaroni, 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.



Références

- Pillot-Loiseau, C., Amelot, A. & Fredet, F. (2010). Contributions of experimental phonetics to the didactics of the pronunciation of the French as a Foreign language: stage 1: reflection around the establishment of a speaking materials. *Proceedings of New Sounds 2010*, 3-5 May 2010, Poznan.
- Racine, I., Detey, S., Bühler, 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 Dziubalska-Kolaczyk, K, Wrembel, M. & Kul, M. (éds.), *Proceedings of New Sounds 2010 - Sixth International Symposium on the Acquisition of Second Language Speech [CD-ROM]*, Poznan: Adam Mickiewicz University, 2010, 367-372.
- Racine, I., Detey, S., Zay, F. & Y. Kawaguchi (sous presse). Des atouts d'un corpus multitâches pour l'étude de la phonologie en L2 : l'exemple du projet « Interphonologie du français contemporain » (IPFC). Dans A. Kamber et C. Skupiens (éds). *Recherches récentes en FLE*. Berne : Peter Lang.
- Racine, I., Detey, S., Zay, F. & Kawaguchi, Y. (sous presse), « Transcription de corpus d'apprenants multilingues de FLE et analyse interphonologique : enjeux méthodologiques ». 24ème colloque du CERLICO « Transcrire, écrire, formaliser », Université de Tours, juin 2010. In *Travaux Linguistiques du CerLiCO 24*, Rennes: PUR.
- Schwab, S. & Llisterri, J. (2010). Vers une méthodologie pour l'étude de l'accentuation en langue étrangère. *Journée IPFC2010 : Interphonologie, corpus et français langue étrangère*, MSH, Paris, 8 décembre 2010.
- Schwab, S. (2011). Les apprenants hispanophones et l'accentuation en L2: quelques pistes méthodologiques. *IPFC2011 Tokyo*, 15-17 mars 2011.
- Strange, W., Bohn, O.-S., Trent, S. A. & Nishi, K. (2005). Contextual variation in the acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 118 : 1751-1762.



Références

- Visceglia, Tseng, Kondo, Meng & Sagisaka (2009). Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *Proceedings of Oriental-COCOSDA*, Urumuqi, Chine.
- Wester, Mirjam, Kessens, Judith, Cucchiarini, Catia & Strik, Helmer, 2001, Obtaining Phonetic Transcriptions : A Comparison between Expert Listeners and a Continuous Speech Recognizer, *Language and Speech*, 44 (3), 377-403.
- Zechner, K. (2009). What did they actually say? Agreement and Disagreement among Transcribers of Non-Native Spontaneous Speech Responses in an English Proficiency Test. *Proceedings of the ISCA SLaTE-2009 Workshop*, Wroxall, UK, September.