● 論　文 ●

Oral corpora, French language education and Francophonie

—how to turn linguistic data into pedagogical resources—

Sylvain DETEY

Abstract

In this article, we examine the role that oral corpora – i.e. well-structured systematic recordings of spoken language – can play in French language education within the *francophonie* (the French-speaking world). We first provide a brief overview of the relatively thin relationship between French language education and oral corpora since the end of the 19th century, followed by an introduction to one of the most recent and renowned corpora in the field of French language studies, the PFC corpus (*Phonologie du Français Contemporain* – Phonology of Contemporary French), which was started ten years ago by a large international research group (Durand, Laks & Lyche 2002). The pedagogical exploitation of the PFC database - both online and offline - is the general objective of the ongoing PFC-EF programme (*Enseignement du français* – teaching of French), which was launched three years ago (Detey & Nouveau 2006) and which we describe in the second part of the article. Two aspects of the programme are underlined: its focus on orality and its capacity to handle linguistic variation, particularly – but not solely - from a geolinguistic viewpoint. Finally, we place the emphasis on methodological perspectives and put forward a few recommandations for turning linguistic data into pedagogical resources (Detey, Durand, Laks & Lyche to appear).

I. Introduction[1]

Before the 1950's, formal teaching of French as a foreign language, like other foreign languages, relied mainly on two types of pedagogical instruments: on the one hand the teacher's *voice*; on the other hand *texts* (textbooks, phrasebooks, grammar books and other academically recognized written documents) or other visually perceptible tools (iconographic or not)[2]. The teacher's voice was *de facto* the oral model to be imitated[3], while the grammar and texts used in class embodied the written standard to be reproduced.

Oral resources in non-French speaking areas were scarce: learners did not have many opportunities to be exposed to other voices than that of their teachers. It is thanks to technological developments and scientific progress made by speech specialists that the situation changed gradually (Martin 2008). The first wave started in 1878, when Thomas Edison received a patent for his "Phonograph". In the 1930's, tape machines started to emerge from Germany (Fritz Pfleumer's "Magnetophon K1")[4], and language laboratories were set up throughout the United States between the 1950's and the 1960's[5] (Léon 1962/1966). The second wave is more recent: the developments of computer science and digital technologies gave birth to the Internet and digital communication, which affected not only the *mass* media (in the case of French, *TV5MONDE* is *the* French-speaking international television channel with an average audience of 55 million viewers per week[6], while *Radio France International* (RFI) counted 46 million listeners in 2007[7]), but also *individual* communication media, with the rapid popularization of freely downloadable videoconference systems[8].

For a long time, following the French normative tradition established in the 17th century, the study of French relied exclusively on the grammatical and discursive models provided by the French literary canons respectful of the so-called socio-grammatical "*bon usage*" (Laks 2002)[9]. Yet, parallel to the technological evolutions we mentioned, the second half of the 20th century witnessed an unprecedented evolution within the field of linguistics. The relationship between grammar, modern linguistics – or language sciences as they are sometimes called now – and foreign language education changed (Besse & Porquier 1984): the descriptive concepts, and sometimes approaches, of modern linguistics gained stronger recognition among language educators, rocking the boat of traditional prescriptive grammars. Nowadays, the vast majority of French teachers are, if not properly trained, at least strongly aware of linguistic issues, especially the distinction between descriptivism and prescriptivism (Detey, Durand & Lyche to appear). Thus, the development of corpus linguistics (Williams 2005; Kawaguchi, Minegishi & Durand 2009) is starting to find echoes in pedagogical practices such as "data-driven learning" (Johns 1991; O'Keefe, McCarthy et Carter 2007; Boulton 2008).

Such an evolution in the approach of linguistic data (Durand 2009) goes hand in hand with an increased attention towards linguistic variation – in a broad sociolinguistic sense. In the case of French, the importance of language variation must be understood within the *francophonie* (with a lowercase *f* - the worldwide community of entirely or partially French-speaking people) and also the *Francophonie* (with an uppercase *F* – the institutional embodiment of this community of which the *Organisation Internationale de la Francophonie* (OIF) is the main operator)[10]. In this respect, the situation of French is quite similar to that of English, with its different varieties throughout the world[11]. The "renewed"[12] variational trend in French language education (Valdman 2000) is partly linked to the sociolinguistic dynamics of French language evolution and spread[13]. As is indicated in the 2006 report on the state of French in the world of the DGLFLF (p. 2): "If the French language borrows accents and takes on different status from Quebec to Paris through Tunis, Dakar, Djibouti or Port-au-Prince, it belongs to all its speakers, whether they inherited it or decided to learn it"[14]. If French is to maintain its status as an international language, French language education must make better use of its potential and take into account all of its varieties within the *francophone* world.

The evolutions mentioned above can be globally schematized as such:

Evolution	
Technology	Printing → Sound recording → Computer → Internet
Methodology	Written norm (prescriptive grammar) → Observation of oral usages (corpus linguistics)
Sociolinguistics	Imposed metropolitan French norm → French as a shared language
Pedagogy	Standard linguistic competence → Panfrancophone linguistic repertoire

Tableau 1: evolution in the field of *francophone* linguistic studies

However, not all of these shifts have reached the same stage yet: although many language centres (but not all of them!) have internet-operating facilities, and French is *de facto* a shared and evolving language, sometimes used in a functional or symbolic partnership with other local languages[15] (of which the endogenous norms are now linguistically recognized), the methodological and pedagogical evolutions are not unanimously acclaimed yet. Apart from profound conceptual and heuristic reasons raised and debated over in ongoing linguistic discussions, it is easy to see that the lack of concern or motivation on the part of language educators partly stems from a lack of available resources. Besides all the socio-educational and psycho-pedagogical factors that must be taken into account to ensure the success of a *panfrancophone* oral French curriculum, there is also a need for: 1) teaching material reflecting the contemporary French linguistic diversity within the *francophonie* and 2) oral material reflecting the authentic use of French in actual speech situations. This observation entails an examination of the existing oral resources, i.e. oral corpora.

II. Oral corpora and French language education

The first systematic recordings of spoken French date back to 1911, when Ferdinand Brunot, an illustrious grammarian and historian of the French language, set up the *Archives de la parole*[16] (Speech archives) in the Sorbonne University in Paris. He recorded the voices of celebrities such as the poet Guillaume Apollinaire, the officer Alfred Dreyfus and the sociologist Emile Durkheim. One of his main objectives was to create a phonographic linguistic atlas of France, with an inventory of all its dialects and patois (300 hours of recordings were collected between 1912 and 1913); but he also wanted to preserve “the right timbre, the perfect rhythm, the pure accent of the speech”[17] of these famous voices which could be used as linguistic models for future generations[18].

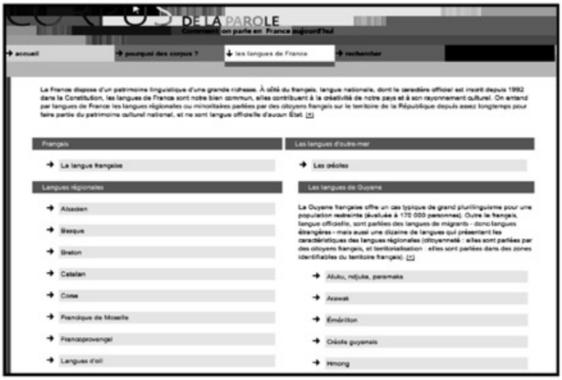
However it was only in 1953 that the results of the first systematic pedagogically-motivated linguistic analysis on spoken French (75 hours of recordings) were published: initially called the *français élémentaire* (Gougenheim, Michéa, Rivenc & Sauvageot 1956), it was renamed - and became renowned as - the *Français Fondamental* (1st degree) in 1959. The *Français Fondamental* became the linguistic basis of the language teaching methods designed within the SGAV *(Structuro-Globale Audio-Visuelle*) methodology (Rivenc 2003): *Voix et Images de France* (VIF), *De Vive Voix* (DVV), etc. The second most famous pedagogically-orientated oral corpus, the *Enquête Socio-Linguistique à Orléans* (ESLO) (350 hours of recordings) was collected in 1968 by British researchers[19]. One of its pedagogical outcomes was a textbook published in 1976 and entitled *Les Orléanais ont la parole* (Biggs & Dalwood 1976)[20]. As for the ESLO corpus, it is currently being digitalized and a second comparable corpus, ESLO 2, is under construction (Abouda & Baude 2006)[21].

Since the constitution of these two corpora, and in spite of the remarkable corpus work carried out by the GARS (*Groupe Aixois de Recherche en Syntaxe*) in the French Provence University (Blanche-Benveniste 1997) from the 70's onward, French[22] has been lagging behind in the field of oral corpora, in comparison with languages such as Portuguese (Corpus de Português Fundamental)[23] or English (British National Corpus) (Laks 2003; Blanche-Benveniste 2006). In the late 1990's, however, partly at the instigation of the DGLFLF, France launched a vast operation of listing and networking of French oral corpora (Cappeau & Seijido 2005): we find, among the most famous, the *Corpus de Référence du Français Parlé* (CRFP, Université de Provence) and the *Corpus de Langues Parlées en Interaction* (CLAPI, Université de Lyon II)[24]. In 2007 this operation culminated in the creation of an official website of the French *Ministère de la culture et de la communication* (Ministry of culture and communication) devoted to oral corpora: the website entitled *Corpus de la parole* (Speech corpora)[25]. Here is a quote from its introductory page :

> *"France can boast a remarkable linguistic wealth based on diversity. Alongside French, the national language of France that is*

> *spoken on all five continents, the languages of France constitute a unique cultural heritage. This cultural heritage is, however, not well known and although audio records exist for almost all of these languages, these data are neither easily accessible for researchers nor for the general public. More worryingly, many unique audio documents, stored on worn-out recording equipment, are in danger of disappearing for ever in the very near future. Digitalization offers the possibility not only of saving them, but also of adding extra-value by turning them into real digital linguistic resources (catalogued, transcribed and annotated recordings), thus ensuring the vitality of this diversity. These collections of recordings, called "oral corpora" by specialists, take on a scientific as much as a patrimonial value".*[26]

The website offers a display of the different *langues de France* (languages of France): French, regional languages (Alsacian, Basque, Breton, etc.), non-territorial languages (Maghrebi Arab, oriental Armenian, Berber, etc.) and overseas languages (Creoles, Guyana languages, French Polynesian languages, etc.).



La France dispose d'un patrimoine linguistique d'une grande richesse. À côté du français, langue nationale, dont le caractère officiel est inscrit depuis 1992 dans la Constitution, les langues de France sont notre bien commun, elles contribuent à la créativité de notre pays et à son rayonnement culturel. On entend par langues de France les langues régionales ou minoritaires parlées par des citoyens français sur le territoire de la République depuis assez longtemps pour faire partie du patrimoine culturel national, et ne sont langue officielle d'aucun État. [+]

Français

→ La langue française

Langues régionales

→ Alsacien
→ Basque
→ Breton
→ Catalan
→ Corse
→ Francique de Moselle
→ Francoprovençal
→ Langues d'oïl

Les langues d'outre-mer

→ Les créoles

Les langues de Guyane

La Guyane française offre un cas typique de grand plurilinguisme pour une population restreinte (évaluée à 170 000 personnes). Outre le français, langue officielle, sont parlées des langues de migrants - donc langues étrangères - mais aussi une dizaine de langues qui présentent les caractéristiques des langues régionales (citoyenneté : elles sont parlées par des citoyens français, et territorialisation : elles sont parlées dans des zones identifiables du territoire français). [+]

→ Aluku, ndjuka, paramaka
→ Arawak
→ Émérillon
→ Créole guyanais
→ Hmong

Fig. 1: the website *Corpus de la parole*

This window display, like most of the corpora it includes, has a cultural (more specifically "patrimonial") and scientific rather than pedagogical function. Nevertheless, the issue of French language teaching is dealt with in the section "*What are oral corpora for?* " of the website:

> *"We resort to oral corpora to build language learning methods, for instance for teaching English in France, or for teaching French as a foreign language (FLE). Oral corpora enable us to index the different uses of a grammatical form and provide us with all the corresponding examples. For instance, it is important to mention the fact that the word "écoute" in sentence-initial position in spoken French does not really mean "listen": it is used only to place the emphasis on the following part of the discourse".*[27]

The exploitation of these corpora, especially with a pedagogical purpose, is part of the second stage of the corpus operation initiated by the DGLFLF and ILF, in what they call *valorisation des corpus* (promoting corpora). We cannot, however, presume that all the corpora listed will be pedagogically useful or exploitable, especially when we take into account the above-mentioned needs in terms of *francophone* linguistic diversity. Among the possible candidates, which include the CRFP, which is mainly oriented towards morphosyntax, and the CLAPI, which is mostly devoted to the study of interaction, there is one large corpus conducive to a variationist approach of oral production within the *francophonie*: PFC (*Phonologie du français contemporain : usages, variétés et structure*) (Phonology of Contemporary French: usage, varieties and structure), initially focused on phonological phenomena but offering much more as is argued below.

III. The PFC corpus: a source of linguistic data

The initial objective of the international PFC project, coordinated by Prof. Jacques Durand (University of Toulouse II), Prof. Bernard Laks (University of Paris X) and Prof. Chantal Lyche (University of Oslo and Tromsø), was to build up a large corpus of spoken French within the *francophonie* using a single methodological protocol, common tools and methods of analysis, in order to collect strictly comparable data that would provide a description of the French pronunciation in its geographic, stylistic and social diversity (technically termed diatopic, diaphasic and diastratic variation). More specifically, the main objectives were to (Durand, Laks & Lyche 2002; Durand 2006):

> "*(a) give a better picture of spoken French in its unity and diversity (geographical, social and stylistic);*
> *(b) test phonological and phonetic models from a synchronic, diachronic and variationist point of view;*
> *(c) favour communication between phonological studies and speech specialists;*
> *(d) provide new material for the teaching of French language and linguistics;*
> *(e) allow for the preservation of spoken varieties of French 'conservation du patrimoine' (cultural heritage).*"

This sociophonological project was launched ten years ago and more than sixty researchers and postgraduate students from all around the world worked together in a collaborative manner, partly thanks to the website of the project (Tchobanov 2008) on which the data, tools and publications are freely available: www.projet-pfc.net.

One of the assets of PFC is the unique, yet simple, protocol it uses: each recorded subject performs the same core tasks (which can be added to according to the theoretical orientations of the investigator or the specificities of the surveyed area) to secure inter-data comparability (Durand & Lyche 2003). In each surveyed area, around 10 people, selected as representative of the local linguistic community, spanning two or three generations in a close social network (e.g. within the same family) are recorded in a natural setting in the following tasks: *Word list reading; Text reading; Led (formal) conversation; Free (informal) conversation*. An agreement form is signed by the subjects, ensuring the confidentiality of personal data and authorizing the use of the recordings for scientific or pedagogical purposes[28]. The sound files are then orthographically transcribed and coded (with an *ad hoc* PFC coding for segmental (schwa and liaison) and suprasegmental phonological phenomena) with the free software *Praat*[29]. So far, 600 subjects have been recorded and almost 400 hours of recording, partially transcribed, are already available on the PFC website.

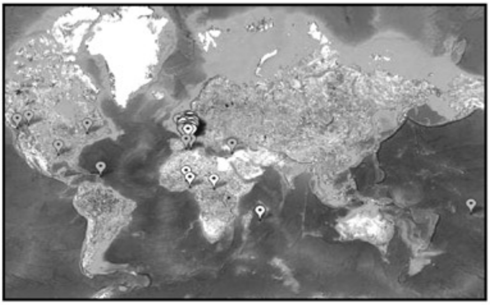


Fig. 2: the area surveyed in the PFC project in the world[30]

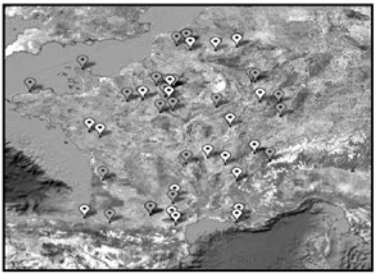
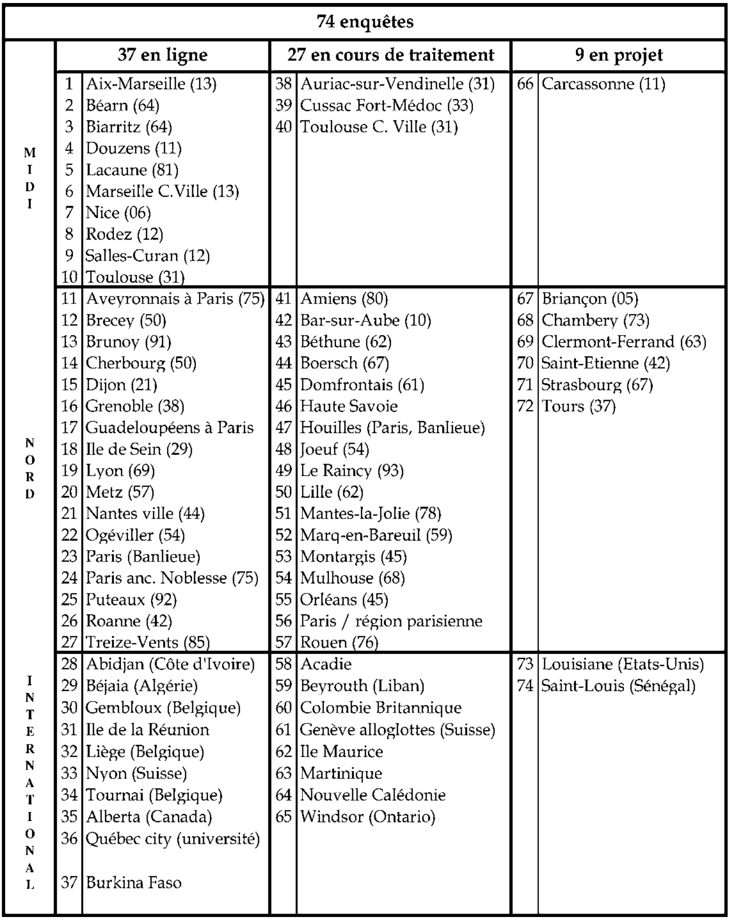


Fig. 3: the areas surveyed in the PFC project in metropolitan France



74 enquêtes						
	37 en ligne		27 en cours de traitement		9 en projet	
MIDI	1	Aix-Marseille (13)	38	Auriac-sur-Vendinelle (31)	66	Carcassonne (11)
	2	Béarn (64)	39	Cussac Fort-Médoc (33)		
	3	Biarritz (64)	40	Toulouse C. Ville (31)		
	4	Douzens (11)				
	5	Lacaune (81)				
	6	Marseille C.Ville (13)				
	7	Nice (06)				
	8	Rodez (12)				
	9	Salles-Curan (12)				
	10	Toulouse (31)				
NORD	11	Aveyronnais à Paris (75)	41	Amiens (80)	67	Briançon (05)
	12	Brecey (50)	42	Bar-sur-Aube (10)	68	Chambery (73)
	13	Brunoy (91)	43	Béthune (62)	69	Clermont-Ferrand (63)
	14	Cherbourg (50)	44	Boersch (67)	70	Saint-Etienne (42)
	15	Dijon (21)	45	Domfrontais (61)	71	Strasbourg (67)
	16	Grenoble (38)	46	Haute Savoie	72	Tours (37)
	17	Guadeloupéens à Paris	47	Houilles (Paris, Banlieue)		
	18	Ile de Sein (29)	48	Joeuf (54)		
	19	Lyon (69)	49	Le Raincy (93)		
	20	Metz (57)	50	Lille (62)		
	21	Nantes ville (44)	51	Mantes-la-Jolie (78)		
	22	Ogéviller (54)	52	Marq-en-Bareuil (59)		
	23	Paris (Banlieue)	53	Montargis (45)		
	24	Paris anc. Noblesse (75)	54	Mulhouse (68)		
	25	Puteaux (92)	55	Orléans (45)		
	26	Roanne (42)	56	Paris / région parisienne		
	27	Treize-Vents (85)	57	Rouen (76)		
INTERNATIONAL	28	Abidjan (Côte d'Ivoire)	58	Acadie	73	Louisiane (Etats-Unis)
	29	Béjaia (Algérie)	59	Beyrouth (Liban)	74	Saint-Louis (Sénégal)
	30	Gembloux (Belgique)	60	Colombie Britannique		
	31	Ile de la Réunion	61	Genève alloglottes (Suisse)		
	32	Liège (Belgique)	62	Ile Maurice		
	33	Nyon (Suisse)	63	Martinique		
	34	Tournai (Belgique)	64	Nouvelle Calédonie		
	35	Alberta (Canada)	65	Windsor (Ontario)		
	36	Québec city (université)				
	37	Burkina Faso				

Tableau 2. overview of the PFC surveys in 2007
(from Mallet & Turcsan 2007)[31]

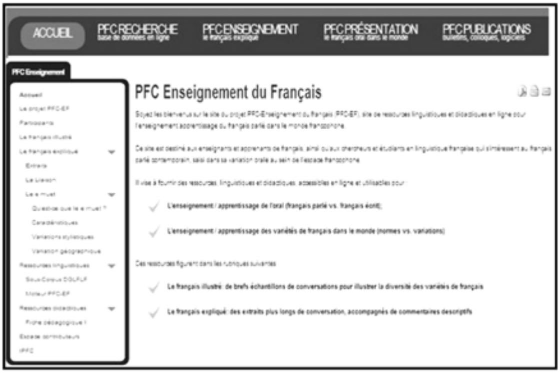
If the PFC corpus stands as a reference corpus in the area of spoken French today, it is not only from a quantitative or qualitative viewpoint: its rapid development attracted the attention of non-phonologist researchers (phoneticians, syntacticians, speech engineers, sociolinguists, psycholinguists) and the pedagogical potential of the database resulted in the setting up of a sub-project: PFC-EF (*PFC-Enseignement du français*) (PFC-Teaching of French) (Detey & Nouveau 2006).

IV. The PFC-EF project: towards pedagogical resources

Following one of PFC's initial objectives (*provide new material for the teaching of French language and linguistics*), the main aim of the sub-project PFC-EF (Detey, Durand, Laks, Lyche & Nouveau 2007) was to increase the usability of the PFC database and turn linguistic data into pedagogical resources for the teaching of French, not only as a foreign language (Nouveau & Detey 2007), but also as a second (Boutin, Brou, Kouadio N'Guessan & Nebout-Arkhurst 2007) or even first language (Delamotte-Legrand & Penloup 2007). Two formats were adopted: one online (a website), the other one offline (a source book).

IV.1. Online resources: the PFC-EF website[32]

On the PFC website, which was initially devoted to fundamental research in phonology, a pedagogical space for PFC-EF was created with four different categories (Tchobanov, Detey & Lyche 2007; Detey, Lyche, Tchobanov, Durand & Laks 2009): *Le français illustré* (French illustrated), *Le français expliqué* (French explained), *Ressources linguistiques* (Linguistic resources), *Ressources pédagogiques* (Pedagogical resources).



PFC Enseignement du Français

Soyez les bienvenus sur le site du projet PFC-Enseignement du français (PFC-EF), site de ressources linguistiques et didactiques en ligne pour l'enseignement apprentissage du français parlé dans le monde francophone.

Ce site est destiné aux enseignants et apprenants de français, ainsi qu'aux chercheurs et étudiants en linguistique française qui s'intéressent au français parlé contemporain, saisi dans sa variation orale au sein de l'espace francophone.

Il vise à fournir des ressources, linguistiques et didactiques, accessibles en ligne et utilisables pour :

✓ **L'enseignement / apprentissage de l'oral (français parlé vs. français écrit);**

✓ **L'enseignement / apprentissage des variétés de français dans le monde (normes vs. variations)**

Ces ressources figurent dans les rubriques suivantes

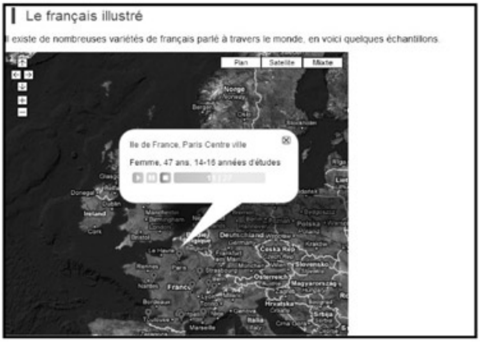
✓ **Le français illustré: de brefs échantillons de conversations pour illustrer la diversité des variétés de français**

✓ **Le français expliqué: des extraits plus longs de conversation, accompagnés de commentaires descriptifs**

Fig. 4: the PFC-EF space

1. French illustrated

In this category, the user (teacher or student alike) can listen to 30-second samples from conversations drawn from the database to illustrate the diversity of French.



Le français illustré

Il existe de nombreuses variétés de français parlé à travers le monde, en voici quelques échantillons.

Ile de France, Paris Centre ville

Femme, 47 ans, 14-16 années d'études

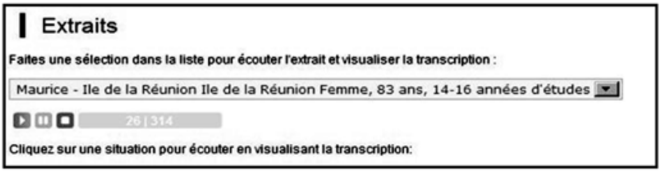
Fig. 5: French illustrated

Examples:

- Marseille (southern France): *Tout se passe très très bien, je suis un joyeux luron, euh bon je euh. Je, je dynamise les, les groupes quand euh, quand on est euh, dans les brigades tout ça euh. Je, je pense être un bon chef de cuisine parce que euh, j'ai toujours su faire la part des choses.*
- Abidjan (Ivory Coast): *Comme je l'ai dit, quand j'étais petite, j'étais au village jusqu'à l'âge de trois, quatre ans. Et là, euh, j'ai ma grand-mère, la tante à ma maman qui est venue me récupérer pour m'amener à Divo. Là, et en quatre ans, trois ans, quatre ans, je peux pas maîtriser, donc, j'ai fait un moment à Divo.*

2. French explained

This category is divided into three sections. The first one offers longer and pedagogically usable samples from conversations (approximately 5 minutes each): the user can listen to the sound file and read the transcription on a synchronized reader[33]. Several of the samples are presented with descriptive linguistic commentaries provided to facilitate their observation and analysis. The second and third sections are respectively devoted to *schwa* and *liaison*, which are two fundamental phenomena in French phonology (Durand & Eychenne 2004; Lacheret, Lyche & Morel 2005; Durand & Lyche 2008) and essential to the teaching/learning of French.



Extraits

Faites une sélection dans la liste pour écouter l'extrait et visualiser la transcription :

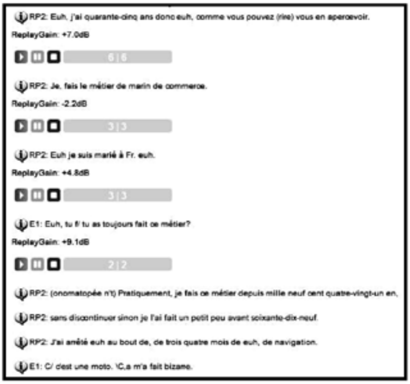
Maurice - Ile de la Réunion Ile de la Réunion Femme, 83 ans, 14-16 années d'études

Cliquez sur une situation pour écouter en visualisant la transcription:

Fig. 6: French explained

3. Linguistic resources

This category offers what can be described as "pre-pedagogical" linguistic resources. The first part consists of a sub-corpus made from the conversations of 32 speakers (around 5 hours of recording) from 9 investigation points (Paris, Vendée, Aix-Marseille and Pays Basque in France, Nyon and Genève in Switzerland, Gembloux and Liège in Belgium, Quebec city in Canada and Abidjan in Ivory Coast). The anonymization and orthographic transcription of this sub-corpus have been thoroughly double-checked and improved, while the conversations have been segmented, thematically labelled (within large categories such as "Work" or "Family stories", and sub-categories such as "My husband is an engineer" or "a Basque grandmother") and provided with a synchronized reader.



RP2: Euh, j'ai quarante-cinq ans donc euh, comme vous pouvez (rire) vous en apercevoir.

ReplayGain: +7.0dB

RP2: Je, fais le métier de marin de commerce.

ReplayGain: -2.2dB

RP2: Euh je suis marié à Fr. euh.

ReplayGain: +4.8dB

E1: Euh, tu f/ tu as toujours fait ce métier?

ReplayGain: +9.1dB

RP2: (onomatopée n't) Pratiquement, je fais ce métier depuis mille neuf cent quatre-vingt-un en,

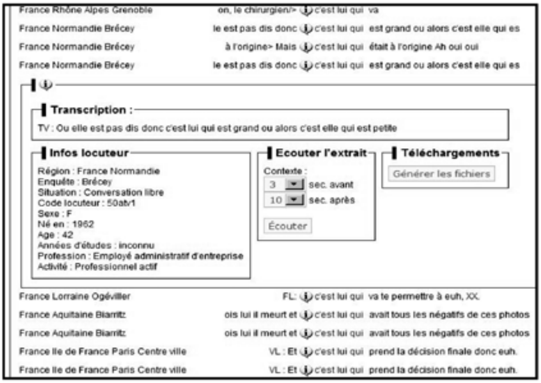
RP2: sans discontinuer sinon je l'ai fait un petit peu avant soixante-dix-neuf.

RP2: J'ai arrêté euh au bout de, de trois quatre mois de euh, de navigation.

E1: C/ c'est une moto. \C,a m'a fait bizarre.

Fig 7: the sub-corpus PFC-EF

The second part of this category consists of the search engine. It enables the user to extract sound-aligned textual units (graphemic, lexical, grammatical) in the PFC database according to several criteria (geographical location, age, etc.) within its original context, as most concordancing softwares do, with an option for right and/or left audio context enlargement.



France Rhône Alpes Grenoble	on, le chirurgien/>	c'est lui qui	va
France Normandie Brécey	le est pas dis donc	c'est lui qui	est grand ou alors c'est elle qui es
France Normandie Brécey	à l'origine> Mais	c'est lui qui	était à l'origine Ah oui oui
France Normandie Brécey	le est pas dis donc	c'est lui qui	est grand ou alors c'est elle qui es

Transcription :

TV : Ou elle est pas dis donc c'est lui qui est grand ou alors c'est elle qui est petite

Infos locuteur

Région : France Normandie
Enquête : Brécey
Situation : Conversation libre
Code locuteur : 50atv1
Sexe : F
Né en : 1962
Age : 42
Années d'études : inconnu
Profession : Employé administratif d'entreprise
Activité : Professionnel actif

Ecouter l'extrait

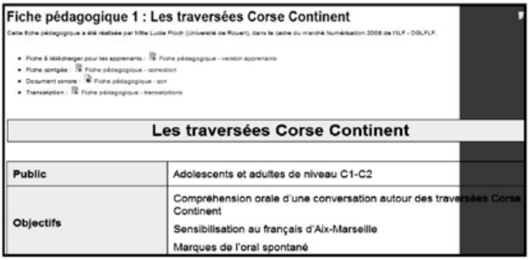
Contexte :
3 sec. avant
10 sec. après

France Lorraine Ogéviller	FL :	c'est lui qui	va te permettre à euh, XX.
France Aquitaine Biarritz	ois lui il meurt et	c'est lui qui	avait tous les négatifs de ces photos
France Aquitaine Biarritz	ois lui il meurt et	c'est lui qui	avait tous les négatifs de ces photos
France Ile de France Paris Centre ville	VL : Et	c'est lui qui	prend la décision finale donc euh.
France Ile de France Paris Centre ville	VL : Et	c'est lui qui	prend la décision finale donc euh.

Fig. 8: results of the search for “*c'est lui qui*” (it's him who)

4. Pedagogical resources

This category provides ready-to-use pedagogical frames including different language learning and/or linguistic activities, either for an online use or for an offline use with the possibility of downloading the audio and text files.



Fiche pédagogique 1 : Les traversées Corse Continent

Cette fiche pédagogique a été réalisée par Mlle Lucie Floch (Université de Rouen), dans le cadre du marché Numérisation 2008 de l'ILF - DGLFLF.

- Fiche à télécharger pour les apprenants : Fiche pédagogique - version apprenants
- Fiche corrigée : Fiche pédagogique - correction
- Document sonore : Fiche pédagogique - son
- Transcription : Fiche pédagogique - transcriptions

Les traversées Corse Continent

Public	Adolescents et adultes de niveau C1-C2
Objectifs	Compréhension orale d'une conversation autour des traversées Corse Continent Sensibilisation au français d'Aix-Marseille Marques de l'oral spontané

Fig. 9: pedagogical frame

Pedagogical activities range from classic listening exercises (characteristics of the speakers, topic of the conversation) using quizzes, questionnaires, transcriptions and gap-filling tasks, to more specific activities on the linguistic aspects of the conversation: rewriting and paraphrasing activities, observation of sociolinguistic and style markers, and so on. Some activities also use the corpus as a source of raw linguistic data in data-driven tasks aiming at shedding light on the morphophonological structure and pragmatic function of grammatical units (for instance the *–ment* ending adverbs in French such as *franchement* and *carrément*).

IV. 2. Offline resources: a source book

The second pedagogical outcome of the PFC-EF project is a source book covering the varieties of French spoken in the main *francophone* areas (Detey, Durand, Laks & Lyche to appear), to which around 30 specialists from around the globe have contributed. One of the main functions of this book is to provide material and guidelines for the study of the different varieties of spoken French: it provides not only a *well-structured* and transcribed oral corpus, but also non-technical information and hints for understanding the specificities of these samples of spoken French and their distinguishing features *vis-à-vis* the traditional descriptions of French provided by written-based prescriptive grammars. The audio files are included in a DVD and the book's structure is as follows:

Part	**Chapter**	**Content**
I	1	Methodological tools
	2	Linguistic tools
	3	Socio-phonological variation
	4	Orality, discourse and syntax
	5	Pedagogical use of the data
II	1-2	French from northern metropolitan France: a synthesis
	3-9	Descriptive analyses
III	1	French from southern metropolitan France: a synthesis
	2-6	Descriptive analyses
IV	1	French from Belgium: a synthesis
	2-6	Descriptive analyses
V	1	French from Switzerland: a synthesis
	2-5	Descriptive analyses
VI	1	French from Africa and the overseas territories: a synthesis
	2-6	Descriptive analyses
VII	1	French from North America: a synthesis
	2-5	Descriptive analyses

Tableau 3: structure of the book
"Les variétés du français parlé dans l'espace francophone"

Due to resource limitations, not all *francophone* areas and types of variation have been included in the book. The focus has been set on diatopic (i.e. geographic) variation, and one speaker from each selected investigation point has been chosen, which amounts to a total of 30 speakers and approximately two and a half hours of conversation, selected according to the audio quality and the thematic content of the recording for an optimal pedagogical use. All the descriptive analyses respect the following structure:

1. Sociolinguistic profile of the speaker
2. Cultural and lexical aspects
3. Syntactic and discursive aspects
4. Phonetic and phonological aspects

Tableau 4: structure of the descriptive chapters

This book can be used by teachers (of French or French linguistics), students and researchers alike. The resources it provides are *linguistic*[34] (recordings and transcriptions) and *metalinguistic*[35] (syntheses and analyses). The latter have been written in a plain academic textbook style (non-technical and few references within the text) to provide easy-reading texts for non-specialists.

V. Conclusion

Building up oral corpora is a very costly and time-consuming activity, and the resulting material can potentially fulfil several functions in the field of language education. Yet, turning a linguistic corpus into a pedagogical database is neither an easy nor an obvious enterprise, as the multi-usage feature of corpora is one of the still often debated methodological issues (see for instance Gadet 2006).

In the case of the PFC corpus, the PFC-EF programme has outlined the broad features of possible modalities of pedagogically-oriented corpus exploitation:

- *Secure the usability of primary data*, to provide a “clean” sub-corpus[36]: anonymous data, agreements signed by the recorded subjects for pedagogical use of the data, alignment of sound and text (if possible), verification and harmonization of the transcriptions, ethically acceptable content, acoustic quality of the recordings, standardization of the formats, length of the audio files (segmentation of the recordings), data identification (not only speech types - according to the classic sociolinguistic criteria: age, sex, geographical and social position, style - but also degree of interactivity), length and theme).

This first step is not as simple as it might seem (Detey, Le Gac, Floch, Coquillon, Navarro, Nouveau & Tchobanov 2008): over-anonymization can make the data pedagogically unexploitable (for people but also for brands, legal entities, places, etc.), unless a very time-consuming (and not always satisfactory)[37] systematic replacement procedure is performed. Other difficulties stem from the lack of visual information, hindering the interpretation and processing of the content of the recordings (onomatopoeia, situational implicitness, deixis, mimics and body information, etc.).

- *Secure data accessibility: online* (downloading time, hypertextual navigation, compatibility of transcription fonts (especially for diacritics in the case of French), etc.) or *offline* (cost of the product and format compatibility according to the area, in the case of DVD for instance).

- *Provide ad hoc data working tools*: search engine with appropriate parameters, text-sound synchronized reading software, etc.

In the case of PFC, the search engine is particularly suited to phonetic/phonological work, but less efficient – at the time being – on the morphosyntactic level, due to the initial objectives of the PFC project (no morphosyntactic tagging, but coding for schwa, liaison and prosody). Apart from the grammatical tagging and general ergonomic work on the website interface to increase its user-friendliness, other tools should be developed or integrated in the “Pedagogical resources” section of the PFC-EF website, using ASR (Automatic Speech Recognition) and CAPT (Computer Assisted Pronunciation Training) technology, for instance (see Neri, Cucchiarini & Strik 2008). Collaborative workspaces for teachers and students should also be part of the agenda.

- *Provide examples of pedagogical utilization of the corpus*: from basic oral comprehension tasks to more complex data-driven metalinguistic observation tasks, for instance on the adverbial system in spoken French (Detey & Nouveau to appear).

Therefore, in the case of the PFC corpus, the future perspectives for the PFC-EF programme include the following: integration of complementary material (different speech types, authentic and pedagogical), grammatical development (morphosyntactic coding, spoken French grammar mining tools), improvement of the multimedia hyperlinking of the existing data, and construction of a well-structured language study programme based on French language variation in the *francophone* world (Valdman 2000).

Beyond their roles as primary linguistic data-providers, oral corpora might bring a new light, and new questions, into the field of language education. What should their function(s) be and which change(s) can they bring about? The first question echoes the one that motivated, in the middle of the last century, the building up of the *Français Fondamental*, i.e. “which French should be taught?” (Detey 2009b). In a world where the status and functions of French are not strictly the same as they were at the time of Rousselot and Brunot (at the turn of the 19th-20th century) or Gougenheim, Michéa, Rivenc and Sauvageot (in the 1950’s), the question takes on a new meaning, and researchers must provide a renewed and up-to-date answer. In our view, oral corpora have an important role to play as linguistic and pedagogical tools in the promotion of panfrancophone plurilingualism within the *francophonie* (Detey, Durand, Laks, Lyche & Nouveau 2007).

[1] The collective work presented in this paper pertains to the PFC-EF programme within the PFC project (*cf. infra*). I would like to thank all the collaborators, especially Prof. Jacques Durand (Toulouse), Prof. Bernard Laks (Paris) and Prof. Chantal Lyche (Oslo), the directors of PFC, and Dr. Atanas Tchobanov (CNRS), who is in charge of the website of the project. In 2008 and 2009, the PFC-EF programme has been supported by grants from the *Institut de Linguistique Française* (ILF), as one of the *Valorisation des corpus oraux* projects of the 2008 *Numérisation* plan, and from the *Délégation Générale à la Langue Française et aux Langues de France* (DGLFLF) of the French *Ministère de la culture et de la communication*.

Many thanks to Professor Jacques Durand and to an anonymous reviewer of the WGF board for their helpful comments on an earlier version of this article.

[2] As early as 1658 with the Czech educationalist J. A. Coménius (Besse 1985), and by the end of the 19th century in western Europe (Germany and France) and the United States (thanks to the emigration of M. D. Berlitz in 1878 and F. Gouin in 1881) (Léon 1962/1966), several educators tried to reinforce the role of oral education in modern language pedagogy, particularly within what came to be known as the "direct methodology" (Puren 1988). After the second world war, the priority of oral education became a basic tenet of the American audio-oral method, of the European audio-visual structuro-global methodology (*méthodologie Structuro-Globale Audio-Visuelle* (SGAV), see Rivenc 2003) and of other approaches such as the Silent way.

[3] However, "speaking machines" (phonographs or gramophones) were used at a very early stage in France (e.g. Summer course of the *Alliance française* in 1911, thanks to the French phonetician and dialectologist Abbé Rousselot) and in the Unites States (Professor C. C. Clarke of Yale University has been reported to have used it from 1906) (Léon 1962/1966). Also, the "*Pathégraphe*" was commercialised in France in 1913, and was known as an audiovisual language learning method (the "Louis Weill" method in German, then in English and Spanish): students would listen to the disk and read the paper-based integrated scrolling transcription and its translation in French at the same time. For more information, refer to the archives of the *Bibliothèque Nationale de France* (French National Library):

http://chroniques.bnf.fr/archives/decembre2006/numero_courant/coulisse/pathegraphe.htm.

For a detailed retrospective on the history of language laboratories before 1940, see (Léon 1962/1966).

[4] In 1937, there were around 200 *Magnetophon* actually in use (Daniel, Denis Mee & Clark 1998: 61).

[5] In the United States, their number amounted to a hundred in 1957, and to more than 10 000 by the end of 1966 (Léon 1962/1966).

[6] http://www.tv5.org/TV5Site/tv5monde/presentation.php.

[7] http://www.rfi.fr/pressefr/articles/072/article_30.asp.

[8] A Japanese freshman from the School of International Studies of Waseda University (Japan) achieved surprisingly good oral skills in French, both in production and com-

prehension, after only a one-week stay in a French-speaking environment (France). Highly motivated, he managed to make friends on Internet with young French learners of Japanese, and would chat and speak with them very regularly for hours using the free software *Skype*^{tm}. During their sessions, they would take some time to correct each other's mistakes in French and in Japanese.

[9] Exceptions include pioneers such as Abbé Rousselot, who, at the end of the 19th century, had already introduced experimental phonetics in the summer course of the *Alliance française* (Léon 1962/1966). Pronunciation and oral skills were among his top priorities, although his views were strongly purist and normative, which must be understood within the ideological and political European context of the time (Nishiyama 2005). As for the "direct method", which appeared in the French official *instructions* for modern language education at the beginning of the 20th century, it did not survive after the first world war (Puren 1988).

[10] According to the 2006 report of the DGLFLF: 115 millions of native speakers of French ("mother tongue"), 61 millions with a partial command of French, around 89 millions of young and adult learners, and 900 000 French teachers (among them more than 400 000 in French speaking Africa, 100 000 in Maghreb and 70 000 in North America).

[11] See for instance the academic journal entirely devoted to this issue that has been published since 1981: *World Englishes, journal of English as an international and intranational language.*

[12] We must remember that the superficial homogeneity of contemporary French, at least in metropolitan France, is the outcome of a long historical process of dialectal eradication and linguistic assimilation. This process was partly natural, partly politically motivated by a need for national unification (not only under the monarchy regime, but even more so during the 1789 French revolution): from the Ordinance of *Villers-Cotterêts* in 1539, by which French becomes an official language at a time when Latin on the one hand and regional languages on the other hand were omnipresent, to the *Jules Ferry* laws of 1881-1882, by which schooling becomes free, laic, mandatory and in French for all French children (Perret 2008).

[13] There are also possible acquisitional motives behind the introduction, under certain conditions, of a rich and diversified linguistic input. On the phonetico-phonological level see (Detey 2009a).

[14] Our translation of: "*Si la langue française emprunte des accents et revêt des statuts différents de Québec à Paris en passant par Tunis, Dakar, Djibouti ou Port-au-Prince, elle appartient à tous ceux qui la parlent, qu'ils l'aient reçue en héritage ou qu'ils aient choisi de l'apprendre*".

[15] The label "partner language" (*langue partenaire*) pertains to the *Francophonie* specific terminology.

[16] See http://www.bnf.fr/PAGES/collections/archives_sonores2.htm and http://gallica.bnf.fr/?lang=fr

[17] Our translation of: "*la parole au timbre juste, au rythme impeccable, à l'accent pur", Le Musée de la parole*, Paris–Journal, March 21st 1910.

[18] The archives were divided into five sections: *Interpreters* (diction and pronunciation conform to the orthoepic norm of the time, such as the one of the French actress Cécile Sorel recorded in a excerpt from the *Misanthrope* of Molière), *Orators* (professors, lawyers), *Foreign languages, Dialects* and *Speech pathologies*. The first two are known as the "Famous voices" of the time. Along with this prestigious catalogue, Brunot also recorded "ordinary" conversations with Parisian craftsmen, following the phonetician Paul Passy in his then-original and innovative interest in "common" spoken French (Cordereix 2002).

[19] For a detailed historical account, see (Bergounioux, Baraduc & Dumont 1992).

[20] Biggs D. & M. Dalwood (1976). *Les Orléanais ont la parole*. London: Longman.

[21] See also: http://www.univ-orleans.fr/eslo/spip.php?rubrique1.

[22] At least in France: the situation in Quebec was different, which was partly due to the sociolinguistic context and politico-linguistic struggles of the 1960's.

[23] Partly inspired from the *Français Fondamental* (Rivenc 2000).

[24] Two other new corpora, with ambitious scientific goals in a *francophone* comparatist perspective, are currently being constituted: the *Corpus International et Ecologique de la Langue Française* (CIEL-F) and the *Français Contemporain en Afrique et dans l'Océan Indien* (CFA) (Dister, Gadet, Ludwig, Lyche, Mondada, Pfänder, Simon & Skattum 2008).

[25] http://www.corpusdelaparole.culture.fr/.

[26] Our translation of: "*La France dispose d'une richesse linguistique fondée sur la diversité. A côté du français, langue nationale, présente sur les cinq continents, les langues de France constituent un patrimoine culturel unique. L'ensemble de ce patrimoine est méconnu, et si des documents sonores existent pour la quasi-totalité de ces langues, ils ne sont accessibles ni à l'ensemble de la communauté scientifique, ni au grand public. Plus grave encore, de nombreux documents sonores uniques, conservés sur des supports physiques à bout d'usage, sont voués à disparaitre à jamais dans un délai très bref. La numérisation offre non seulement la possibilité de sauver ces documents, mais aussi de les valoriser en les transformant en de véritables ressources linguistiques numériques (enregistrements catalogués, transcris et annotés), assurant ainsi la vitalité de cette diversité. Ces collections d'enregistrements appelés "corpus oraux" par les spécialistes, prennent alors une valeur autant scientifique que patrimoniale*".

[27] Our translation of: "*On a recours aux corpus oraux pour construire les manuels de langue étrangère, que ce soit par exemple l'anglais en France, ou le français comme langue étrangère (FLE). Cela permet entre autres de répertorier les différentes utilisations d'une forme grammaticale et de fournir tous les exemples correspondants. Il faut par exemple préciser que « écoute » placé en début de phrase, n'a pas vraiment de sens à l'oral : il sert seulement à insister sur ce qui va suivre.*"

[28] By doing so, the PFC protocol follows the recommendations of the *Guide des bonnes pratiques pour la constitution, l'exploitation, la conservation et la diffusion des corpus oraux* (guidebook for the creation of oral corpora) coordinated by Olivier Baude (2005).

[29] See: http://www.fon.hum.uva.nl/praat/.

[30] Dark-coloured flag = data not yet available online; light-coloured flag = data available online.

[31] 74 surveys: 37 online, 27 under process, 9 in preparation. For up-to-date figures, see: www.projet-pfc.net/ .

[32] Still under construction.

[33] Thanks to the free software *Cantare*. See: http://rea.ccdmd.qc.ca/ri/cantare/ .

[34] With several possibilities of pedagogical use according to the format: a) *audio* only (e.g. orthographic or phonetic transcription task, repetition task, note-taking task, etc.); b) *orthographic* only (e.g. grammatical observation task, reading aloud task, rephrasing task, etc.); c) *audio and orthographic* synchronously or asynchronously (e.g. morphophonological observation taks, global and detailed oral comprehension task, dictation task, etc.).

[35] For teacher training and student courses but also for pre-pedagogical linguistic descriptions used to build up language learning activities.

[36] This echoes the recommendations of the oral corpus ethics guidebook by Baude et al. (2005).

37 For instance, how should we deal with acronyms such as "SNCF" *(Société Nationale des Chemins de fer français)* (the French national railway company)?

References

Abouda L. & O. Baude (2006). Constituer et exploiter un grand corpus oral: choix et enjeux théoriques. Le cas des ESLO. In F. Rastier & M. Ballabriga (eds) *Corpus en Lettres et Sciences sociales: des documents numériques à l'interprétation*. Paris: Texto, 143–50.

Baude O., C. Blanche Benveniste, M-F. Calas, P. Cordereix, I. de Lamberterie, L. Goury, C. Marchello-Nizia & L. Mondada (2005). *Guide des bonnes pratiques pour l'exploitation, la conservation, et la diffusion des corpus oraux*. Paris: CNRS.

Bergounioux G., J. Baraduc, & C. Dumont (1992). L'Etude sociolinguistique sur Orléans (1966-1991), 25 ans d'histoire d'un corpus. *Langue française* 93, 74-93.

Besse H. (1985). *Méthodes et pratiques des manuels de langue*. Paris: Didier-Credif.

Besse H. & R. Porquier (1984). *Grammaires et didactique des langues*. Paris: Hatier.

Blanche-Benveniste C. (1997). *Approches de la langue parlée en français*. Paris: Ophrys.

Blanche-Benveniste C. (2006). Linguistic Analysis of Spoken Language: the case of French. In Y. Kawaguchi, S. Zaima & T. Takagaki (eds) *Spoken Language and Linguistic Informatics*. Amsterdam: John Benjamins, 35-66.

Boulton A. (2008). Esprit de corpus: promouvoir l'exploitation de corpus en apprentissage des langues. *Texte et Corpus* 3, 37-46.

Boutin B.A., A. C. Brou-Diallo, J. Kouadio N'Guessan & P. Nebout-Arkhurst (2007). De l'intérêt du projet PFC-EF pour la didactique du français, langue seconde en Côte d'Ivoire. In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7: *PFC: enjeux descriptifs, théoriques et didactique*. Toulouse: CLLE-ERSS-UTM, 65-86.

Cappeau P & M. Seijido (2005). *Inventaire des corpus oraux en langue française*. Paris: DGL-FLF.

Cordereix P. (2002). Des Archives de la parole au Département de l'Audiovisuel de la Bibliothèque nationale de France – 1911 - 2002 : un siècle de français parlé enregistré. Journée d'étude *Constitution et exploitation de corpus du français parlé*, Paris.

Daniel E. D., C. Denis Mee & M. H. Clark (eds) (1998). *Magnetic Recording: the first 100 years*. Piskataway: Wiley-IEEE Press.

Delamotte-Legrand R. & M.-C. Penloup (2007). Intérêt et usages des documents PFC en langue première. In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7*: PFC: enjeux descriptifs, théoriques et didactique*. Toulouse: CLLE-ERSS-UTM, 55-64.

Detey S. (2009a). Phonetic input, phonological categories and orthographic representations: a psycholinguistic perspective on why oral language education needs oral corpora. The case of French-Japanese interphonology development. In Y. Kawaguchi, M. Minegishi & J. Durand (eds) *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins, 179- 200.

Detey S. (2009b). Normes pédagogiques et corpus oraux en FLE: le curseur apprenabilité / acceptabilité et la variation phonético-phonologique dans l'espace francophone. In B. Olivier & I. Schaffner (eds.) *Quel français enseigner? La question de la norme dans l'enseignement/apprentissage*. Paris: Editions de l'Ecole Polytechnique, 155-168.

Detey S., J. Durand, B. Laks & C. Lyche (eds.) (to appear). *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*. Paris: Ophrys.

Detey S., J. Durand, B. Laks, C. Lyche & D. Nouveau (2007). Voix de la francophonie, éducation langagière et corpus numérisé : PFC-EF, des ressources pour la didactique du français. In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7: *PFC: enjeux descriptifs, théoriques* *et didactique*. Toulouse: CLLE-ERSS-UTM, 11-29.

Detey S., J. Durand & C. Lyche (to appear). Éléments de linguistique pour la description de l'oral. In S. Detey, J. Durand, B. Laks & C. Lyche (eds.) *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*. Paris: Ophrys

Detey S., D. Le Gac, L. Floch, A. Coquillon, S. Navarro, D. Nouveau & A. Tchobanov (2008). Valorisation des corpus oraux: développements récents de PFC-EF. Colloque *Phonologie du français contemporain: variation, interfaces, cognition*. Paris : MSH.

Detey S., C. Lyche, A. Tchobanov, J. Durand & B. Laks (2009). Ressources phonologiques au service de la didactique de l'oral: le projet PFC-EF. *Mélanges CRAPEL* 31, 223-236.

Detey S. & D. Nouveau (2006). PFC pour l'enseignement du français: lancement du projet PFC-EF. Colloque *Phonologie du français: du social au cognitif*. Paris: MSH.

Detey S. & D. Nouveau (to appear). Des données linguistiques à l'exploitation didactique. In S. Detey, J. Durand, B. Laks & C. Lyche (eds.) *Les variétés du français parlé dans l'espace francophone: ressources pour l'enseignement*. Paris: Ophrys.

Dister A., F. Gadet, R. Ludwig, C. Lyche, L. Mondada, S. Pfänder, A. C. Simon & I. Skattum (2008). Deux nouveaux corpus internationaux du français: CIEL-F et CAF. *Revue de Linguistique Romane* 285-286, 295-314.

Durand J. (2006). Mapping French Pronunciation: the PFC project. In J.-P. Montreuil & C. Nishida (eds) *New Perspectives on Romance Linguistics*. Vol. 2: *Phonetics, Phonology and Dialectology*. Amsterdam: John Benjamins, 65-82.

Durand J. (2009). On the scope of linguistics: Data, intuitions, corpora. In Y. Kawaguchi, M. Minegishi & J. Durand (eds) *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins, 25–52.

Durand J. & J. Eychenne (2004). Le schwa en français: pourquoi des corpus? *Corpus* 3, 311-356.

Durand J., B. Laks & C. Lyche (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*. Tübigen: Gunter Narr Verlag, 93-106

Durand J. & C. Lyche (2003). Le projet « Phonologie du français contemporain et sa méthodologie ». In E. Delais-Roussarie & J. Durand (eds.) *Corpus et variation en phonologie du français. Méthodes et analyse*. Toulouse: PUM, 213-278.

Durand J. & C. Lyche (2008). French liaison in the light of corpus data. *Journal of French Language Studies* 18, 33-66.

Gadet F. (2006). Que le linguiste sache ce qu'il fait: quelques réflexions sur les grands corpus. Colloque *Phonologie du français: du social au cognitif*. Paris: MSH.

Gougenheim G., R. Michéa, P. Rivenc & A. Sauvageot (1956). *L'élaboration du français élémentaire*. Paris: Didier.

Johns T. (1991). Should you be persuaded - two samples of data-driven learning materials. In T. Johns & P. King (eds) *English Language Research Journal* 4: *Classroom Concordancing*. Birmingham University, 1-13.

Kawaguchi Y., M. Minegishi & J. Durand (eds) (2009). *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins.

Lacheret A., C. Lyche & M. Morel (2005). Phonological Analysis of Schwa and Liaison within the PFC Project (Phonologie du Français Contemporain): how determinant are the prosodic factors ? *Proceedings of INTERSPEECH 2005*, 1437-1440.

Laks B. (2002). Description de l'oral et variation : la phonologie et la norme. *L'information grammaticale* 94, 5-10.

Laks B. (2003). Les grandes enquêtes phonologiques en France. *La Tribune Internationale des Langues Vivantes* 33, 10-17.

Léon P. (1962/1966). *Laboratoire de langues et correction phonétique*. Paris: Didier.

Mallet G.-M. & G. Turcsan (2007). Situation et perspectives du projet PFC. In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7: *PFC: enjeux descriptifs, théoriques et didactique*. Toulouse: CLLE-ERSS-UTM, 7-10.

Martin P. (2008). *Phonétique acoustique*. Paris: Armand Colin.

Neri A., C. Cucchiarini & H. Strik (2008). The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch. *ReCALL 20* (02), 225-243.

Nishiyama N. (2005). Cours de vacances de l'Alliance française sous la IIIe République et genèse de la formation des maîtres étrangers de français. *Enseignement du français au Japon* 33, 19-34.

Nouveau D. & S. Detey (2007). Enseignement/apprentissage du schwa et apprenants néerlandais : des données de la base PFC à l'espace-ressource en ligne du projet PFC-EF. In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7: *PFC: enjeux descriptifs, théoriques et didactique*. Toulouse: CLLE-ERSS-UTM, 87-106.

O'Keefe A., M. McCarthy & R. Carter (2007). *From Corpus to Classroom: language use and language teaching*. Cambridge: CUP

Perret M. (2008). *Introduction à l'histoire de la langue française*. Paris: Armand Colin.

Puren C. (1988). *Histoire des méthodologies de l'enseignement des langues*. Paris: Cle International.

Rivenc P. (2000). *Pour aider à apprendre à communiquer dans une langue étrangère*, Paris: Didier Erudition / Mons: CIPA.

Rivenc P. (ed.) (2003). *Apprentissage d'une langue étrangère/seconde. Tome 3: Méthodologie.* Bruxelles: De Boeck-Université.

Tchobanov A. (2008). Evolution de la base et du site PFC: PFC-EF, interface TLF et codages prosodie. Colloque *Phonologie du français contemporain: variation, interfaces, cognition.* Paris: MSH.

Tchobanov A., S. Detey & C. Lyche (2007). Vers un espace numérique tripartite pour la recherche, la diffusion et l'enseignement du français parlé, en présentiel ou à distance: évolution du site PFC (novembre 07). In S. Detey & D. Nouveau (eds) *Bulletin PFC* 7: *PFC: enjeux descriptifs, théoriques et didactique.* Toulouse: CLLE-ERSS-UTM, 31-40.

Valdman A. (2000). Comment gérer la *variation* dans l'enseignement du français langue étrangère aux Etats-Unis. *The French Review* 75 (4), 648-666.

Williams G. (ed.) (2005). *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes.