



DGLFLF

INVENTAIRE DES CORPUS ORAUX DU FRANCAIS

Version 1.1.

Table des matières

(Cliquer pour accéder à la rubrique)

[1. CORPUS EXISTANTS](#)

[2. PROJETS DE CORPUS](#)

[1.1. France](#)

[2.1. France](#)

[1.2. Belgique](#)

[2.2. Belgique](#)

[1.3. Suisse](#)

[2.3. Suisse](#)

[1.4. Suède](#)

[2.4. Suède](#)

[1.5. Danemark](#)

[1.6. Finlande](#)

[1.7. Pays-Bas](#)

[1.8. Royaume-Uni](#)

[1.9. Allemagne](#)

[1.10. Autriche](#)

[1.11. Etats-Unis](#)

Nom du corpus	Nom du Laboratoire ou de l'équipe	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	L'accès aux données	Publications d'études sur ce corpus	Contacts
1. CORPUS EXISTANTS								
1.1. France								
Nom du corpus	LABO ou équipe	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	L'accès aux données	Publications	Contacts
AIR FRANCE	ASILA	MOREL, M-A. et DELOMIER, D., Paris 3	73 dialogues organisés en quatre parties (54000 mots)	Non média: dialogues (réservations de billets Air France et renseignements)		Accès à 9 transcriptions http://www.loria.fr/projets/asila/corpus.html		Suzanne Salmon-Alt susanne.alt@loria.fr
ALLAIRE		ALLAIRE, S., 1973. Corpus exploité pour une étude sur la subordination.	20h d'enregistrement	Média: trois séries de débats radiodiffusés (élections législatives de 1967)	Tous les enregistrements ne sont pas transcrits. 360 p. dactylographiées		ALLAIRE, S., 1973, <i>La subordination dans le français parlé devant les micros de la radiodiffusion</i> . Paris: Klincksieck	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

ALMURA	Centre de Dialectologie de Grenoble	MEDELICE, J.-E., Corpus constitué pour étudier les parlers de la région Rhône-Alpes.	Environ 50 enregistrements sur minidisques (depuis 2000)	Non média.	Tous les enregistrements ne sont pas transcrits.	Données accessibles à tous sur le net, gratuites. Publications dans le commerce.		MEDELICE Jeanine Elisa, Centre de dialectologie. Université Stendhal Grenoble 3. Jeanine.medelice@u-grenoble3.fr
ANGERS		HOUEBINE A.M. Interviews thématiques, pour l'institut de géographie d'Angers.			Transcriptions phonétiques.			
BDEDED268		Corpora constitués en vue d'études phonologiques, syntaxiques, sociolinguistiques et d'analyses conversationnelles	13 corpus numérisés (de taille variable)	Non média	Une majorité d'enregistrements ont été transcrits par des étudiants. Pas de conventions de transcription. Conventions de documentation.	http://pi-ed268.univ-paris3.fr	Publication d'études. Rapport de projet : http://pi-ed268.univ-paris3.fr/publis/rapport-final_pi-ed268-bdd.pdf	cgendrot@univ-paris3.fr ; serge.fleury@univ-paris3.fr
BDPHO	ICP		20 h de parole transcrite					
BELC		MALANDAIN J.L. " Documents sonores authentiques à des fins didactiques", à partir du corpus d'Orléans; 16 documents pédagogiques (cassettes et brochures), "langue et civilisation, 12 dossiers pour la classe", avril 1971 (et en cours d'achèvement)	Environ 6h30 d'enregistrement.					
BNF Archives de la Parole		Les Archives de la parole sont créées en 1911 à la Sorbonne. Elles préfigurent un Institut de phonétique en projet depuis plusieurs années à l'Université de Paris, et se veulent un conservatoire, un musée de la parole. A partir de novembre 1911, Ferdinand Brunot réalise				gallica.bnf.fr/ArchivesParole/		Pascal Cordereix : PASCAL.CORDEREIX@bnf.fr et site www.up.univ-mrs.fr/veronis/Atala/jecorpus/Cordereix.html

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

BNF Archives de la Parole - Section D		Section D des Dialectes. enregistrement de dialectes du monde entier, mais plus particulièrement de France, avec notamment la constitution d'un atlas linguistique phonographique des patois et dialectes de France, inspiré de l'Atlas linguistique de la France	Ferdinand Brunot ramènera près de 300 enregistrements (sur bandes magnétiques) de ces enquêtes de terrain.	Non média		gallica.bnf.fr/ArchivesParole/		Pascal Cordereix : PASCAL.CORDEREIX@bnf.fr
BNF Archives de la Parole - Section I et section O		Section I des interprètes : enregistrements de diction, de prononciation correcte du français. Section O des orateurs : professeurs, avocats... : Enregistrements inspirés des Stimmporträts, les “ portraits sonores ” réalisés par les Phonogrammarchiv	Les sections I et O correspondent à plus de 300 enregistrements sur bandes magnétiques.			gallica.bnf.fr/ArchivesParole/		Pascal Cordereix : PASCAL.CORDEREIX@bnf.fr
BNF Archives de la Parole - Section L		Section L des langues : avec l'enregistrement des langues étrangères et la mise au point de méthodes de langue grâce au phonographe (Pathégraphie).				gallica.bnf.fr/ArchivesParole/		Pascal Cordereix : PASCAL.CORDEREIX@bnf.fr
BNF Archives de la Parole - Section M		Section M des maladies de la parole, de l'expression avec l'application du phonographe au traitement des pathologies de la parole.						Pascal Cordereix : PASCAL.CORDEREIX@bnf.fr
BRANCA-PARIS 3	LPP (?)	BRANCA, Sonia. Corpus constitués pour des études sociolinguistiques	Environ 100 enregistrements soit à peu près 11 heures entre 2002 et 2005 sur divers supports	Média (une vingtaine de corpus) et non média	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants. Utilisation de conventions de transcription.	Données consultables sur place.		BRANCA, Sonia 22 avenue du Bel Air 75012 Paris branca.sonia@wanadoo.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

CAFE	ASILA (CNRS)	LUZZATI, D. Corpus constitué en vue d'études discursives et morphosyntaxiques.	10 enregistrements courts sur cassettes correspondant à 7000 mots (1975-1980).		Tous les enregistrements sont transcrits. Les transcriptions ont été faites par D. LUZZATI. Conventions de transcription.	Données (textes) - consultables sur place - accessibles à tous sur le net, gratuites - possibilité de les télécharger et de les copier. http://www.loria.fr/projets/asila/corpus.html	Publication d'études.	LUZZATI, Daniel, Univ. Du Maine Faculté des Lettres, Langues et Sciences Humaines av. Olivier Messiaen, 72085 Le Mans cedex 9 Daniel.Luzzati@iium.univ-lemans.fr
CAVILAM		"Corpus, gestes et paroles". Cf. <i>Le non-linguiste dans la communication</i> (1984), Actes du Colloque organisé à Vichy les 14 et 15 janvier 1983.						
CHLOE		BOURHIS, V., Corpus constitué en vue d'études psycholinguistiques (paramètre intonatif et acquisition)	Une dizaine d'enregistrements sur cassettes et compacts discs correspondant à 3h. (1991-1992)	Non média	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par V. Bourhis. Conventions de transcription de l'ED 268 Paris 3.	A discuter.	Publication de thèse en cours.	BOUHRIS Véronique, 1 rue Jean-Louis Sinet 92330 Sceaux bourhisveronique@yahoo.fr
CLAPI http://corpus.univ-lyon2.fr/	ICAR	Corpus collectés dans le cadre de Recherches sur l'interaction. Une partie est intégrée dans la plateforme CLAPI (Corpus de Langues Parlées en Interaction) qui offre des outils de requêtes automatisés.	70 corpus inventoriés correspondant à 600 h. d'enregistrements audio ou vidéo. 35 corpus (correspondant à 130 h.) sont numérisés. 28 corpus intégrés dans CLAPI, soit 200 transcriptions (104 h). 57 transcriptions balisées en xml (30 h).	Non média	Tous les enregistrements sont transcrits selon diverses conventions de transcription, dont la convention ICOR	Possibilité de télécharger 3 h. d'échantillons (dont 1h avec le signal). Interrogation libre par les outils de requêtes : 8 h. Sous convention de recherche : 22 h	Nombreuses publications d'études, dont 85 référencées dans CLAPI.	info-clapi@univ-lyon2.fr
CLER		BLASCO-DULBECCO, M. Equipe GARS/ DELIC. Corpus réalisé pour des études syntaxiques	104 enregistrements courts (transcrits) sur cassettes et minidisques correspondant à 26 h. (1996 à 2004)	Non média.	Les enregistrements sont transcrits depuis 1999. Les transcriptions ont été faites par des étudiants. Conventions de transcription GARS / DELIC	Données (son et textes) accessibles sur place. Une partie du corpus a été reversé à l'équipe DELIC (Aix)		BLASCO-DULBECCO, M. UFR LSH 29 bd Gergovia 63037 Clermont-Ferrand Cedex 1 Mylene.BLASCO-DULBECCO@lettres.univ-bpclermont.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

C-ORAL-ROM	DELIC	VERONIS, J. (responsable), L'objectif donc de ce projet européen est de rassembler des échantillons de discours oral de chaque langue représentée (300.000 mots/langues : français, italien, espagnol et portugais), et d'établir un outil multimédia permettant (a) d'avoir un alignement sonore des quatre langues ; (b) d'accéder à un étiquetage prosodique et pragmatique dans les quatre langues	206 enregistrements, sur minidisques correspondant à 24h. (2001-2003)	Média et non-média.	Transcrits selon les conventions DELIC plus ajout d'informations prosodiques : pause et unités mineures et majeures	Distribué par ELRA	Cresti & Moneglia C-ORAL-ROM John Benjamins (livre + DVD)	VERONIS, Jean, Université de Provence 29 avenue Robert Schuman - 13100 Aix-en-Provence jean.veronis@up.univ-mrs.fr
CORPAIX-2	DELIC	Équipe GARS. Responsable actuel Jean Véronis. Corpus réalisé notamment pour des études syntaxiques	1 702 000 mots. Inclut la partie informatisée du corpus du GARS / CORPAIX et l'a enrichi.	Média et non média.	Transcrits avec les conventions du GARS	Non distribué. Corpus de travail de l'équipe		VERONIS, Jean, Université de Provence 29 avenue Robert Schuman - 13100 Aix-en-Provence jean.veronis@up.univ-mrs.fr
CREDIF		Centre de recherche et d'étude pour la diffusion du français. "Enquête sur le langage de l'enfant français"	1965 et 1968	Non média. Conversations d'enfants de 9 ans (200p.) et 10 ans (300p.)				
CREDIF		MARTIN-BALTAR, Michel et al., Centre de recherche et d'étude pour la diffusion du français, 1989, Entretiens.				Publications dans le commerce: MARTIN-BALTAR, Michel et al., 1989, <i>Entretiens. Transcription d'un corpus oral</i> . In Cahiers du Français des années Quatre-vingts. Hors série 1, CREDIF / Didier (ENS de Fontenay-Saint-Cloud, actuellement à l'ENS de Lyon). ISBN 2-905769-15-7		
CRELEF		Corpus constitué pour l'enseignement du français.		Média: émissions radiophoniques				Université de Besançon, 30, rue Mégévand, 25030 Besançon.
CREPCO		G. AMY, F. LUBER, A. PIOLAT, S. SAUVAIRE et M. VION. Centre de recherche en psychologie cognitive. Corpus d'enfants.						

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

CRFP-1	DELIC	VERONIS, J. (responsable), "Corpus de référence du français parlé 1". Corpus constitué en vue d'études linguistiques (syntaxe notamment).	134 enregistrements courts sur minidisques correspondant à 36 h. (1999-2000). Attention, certains sont déjà signalés dans CLER, GRE, NANCY-DEBAISIEUX, PERPIGNAN et POI)	Média et non-média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des transcrip-teurs multiples. Conventions de transcription orthographiques.	http://www.up.univ-mrs.fr/delic/crpf/	Présentation dans Recherches Sur le Français Parlé N° 18.	VERONIS, Jean, Université de Provence 29 avenue Robert Schuman - 13100 Aix-en-Provence jean.veronis@up.univ-mrs.fr
DELIC	DELIC	Equipe DELIC. Corpus réalisé notamment pour des études syntaxiques	1 460 000 mots. ce corpus contient notamment ds enregistrements signalés dans CLER, POI, NANCY-DEBAISIEUX	Média et non média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des transcrip-teurs multiples. Conventions DELIC	Non distribué. Corpus de travail de l'équipe		Estelle Campione, Université de Provence 29 avenue Robert Schuman - 13100 Aix-en-Provence estelle.campione@up.univ-aix.fr
ECOLE MASSY	VALORIA	ANTOINE, Jean-Yves, GOULIAN, Jérôme, LETELLIER, Sabine (laboratoire VALORIA, Université de Bretagne Sud). "Office du Tourisme de Grenoble". Ce corpus constitué en vue d'études linguistiques porte sur un public de jeunes enfants.	31 dialogues (total 45 minutes)	Non média	Transcription orthographique. 5300 mots	http://www-valoria.univ-ubs.fr/antoine/parole_publique/OTG/index.html		Jean-Yves Antoine : Jean-Yves.Antoine@univ-ubs.fr Jerome Goulian : Jerome.Goulian@univ-ubs.fr
ESLO	CORAL	BERGOUNIOUX G. et équipe du CORAL à Orléans. Enquête Socio-Linguistique à Orléans en vue de la réalisation de manuels de FLE. (cf. Royaume-Uni)	350 h d'enregistrements sur bandes (1968 à 1969)	Non média.	Tous les enregistrements ne sont pas transcrits. Les transcriptions ont été faites par l'équipe anglaise, belge et néerlandaise. Pas de conventions de transcription.	Données accessibles à tous sur le net, gratuites ou payantes. Voir le site http://bach.arts.kuleuven.be/elicop/		BERGOUNIOUX G. UFR LSH BP 46527 Orléans cedex 2 tél. 02 38 54 23 20 (privé) gabriel.bergounioux@wanadoo.fr
FOSSAT		FOSSAT J.L. et 10 collaborateurs "Français parlé et savoirs techniques", sous-partie d'une enquête dialectologique de 6000h.			transcription phonétique, phonologique et orthographique, saisie sur ordinateur.		Fossat, Jean-Louis et Valière, Michel, 1977, Histoire de la vie rurale en Poitou. Récit d'un étalonnier. Toulouse. Université de Toulouse le Mirail	
FRANCAIS FONDAMENTAL 1		GOUGENHEIM, G., MICHEA, P., RIVENC, P. et SAUVAGEOT, A. Corpus constitué pour l'enseignement, le vocabulaire et la syntaxe. 275 témoins, de 17 régions francophones.	163 enregistrements de 10 à 30 mn correspondant à environ 75 h. (1951-1953)	Non média et quelques enregistrements médias.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par l'équipe du Français Fondamental (ENS de Saint-Cloud) sous la direction de G. GOUGENHEIM. Conventions de transcription.	Données consultables sur place: les archives sont sur microfilms.	GOUGENHEIM, G., MICHEA, R., RIVENC, P., SAUVAGEOT, A., 1964, <i>L'élaboration du français fondamental (1er degré)</i> Paris: Didier.	RIVENC Paul Résidence Bon voisin B1 - 136 ch. Salade Ponsan - 31400 Toulouse Tél. 05 61 52 31 71 rivenc@univ-tlse2.fr / rivenc@numericable.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

GARS / CORPAIX		BLANCHE-BENVENISTE Cl. et collaborateurs, Groupe Aixois de Recherche en Syntaxe. Corpus constitué principalement pour l'étude de la morphosyntaxe.	La partie "moderne" du corpus (numérisée) est appelée CorpAix. Enregistrements courts et longs sur cassettes et minidisques qui correspondent à plus d'un million de mots (des années 1990 à 2000).	quelques conversations; beaucoup de récits de vie; des récits d'accidents; de sujets d'actualité; des relations entre institutions et public enregistrées à la poste, la mairie, l'ANPE, etc.; des prises de parole professionnelles publiques; des savoirs-faire professionnels; des témoignages de personnes parlant au nom d'une association; des situations de parodie; des enregistrements de radio et TV	Les transcriptions ont été faites par l'équipe du GARS (chercheurs, enseignants, étudiants) et des étudiants de licence. Conventions de transcription orthographiques.	Données (son et textes) consultables sur place. Base de données à accès modulés. Publications d'extraits du corpus: BLANCHE-BENVENISTE, C., ROUGET, C., SABIO, F. (éds), 2002, <i>Choix de textes de Français parlé. 36 extraits</i> . Paris: Champion "Collection "Les français parlés").	Nombreuse publications des membres du GARS depuis 1977 - ouvrages et articles - Revue <i>Recherches sur le français parlé</i> (RSFP) qui comprend actuellement 18 numéros. - Publications de thèses (sous la direction de Cl. Blanche-Benveniste).	BLANCHE-BENVENISTE Cl. Claire.benveniste@freesbee.fr
GRE		SAVELLI, M. Corpus constitué en vue d'études morphosyntaxiques.	Enregistrements courts (20 mn) sur cassettes, minidisques et compacts disques correspondant à 500 000 mots	Non média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants en licence de sciences du langage. Conventions de transcription GARS / DELIC.	Données (son et textes) consultables sur place. L'intégration d'une partie de ces données dans la plateforme CLAPI est en cours : une quinzaine de corpus fin 2006.	Publications d'études.	SAVELLI Marie, Domaine Saint-Georges 38410 Uriage. marie.savelli@u-grenoble3.fr
GRENOBLE / MAILLARD		MAILLARD, M. "variétés régionales du français".	15 d'enregistrement	Non média. Interviews et conversations.	Transcription orthographique.			
GRENOBLE-1		BOUCHARD, "Conversation en classe de français langue étrangère".	1h d'enregistrement	Non média	Transcription orthographique.			
GRENOBLE-2		BOUCHARD / DABENE "Débat en classe de C.P."		Non média				
GRENOUILLES	DDL	KERN, S. Corpus constitué en vue d'études syntaxiques et psycholinguistiques	140 enregistrements courts sur cassettes (effectués en 1993) rassemblés pour des thèses et des mémoires.	Non média	Tous les enregistrements ont été transcrits par des étudiants. Utilisation de conventions.	Données (son et textes) consultables sur place. Accès sur CLAPI et fin 2006 interrogation en ligne d'une partie de ces données.	Publication d'études.	Florence.chenu@univ-lyon2.fr Harriet.jisa@univ-lyon2.fr
HP		CAPPEAU, P. Interviews d'hommes politiques dans les médias. Corpus utilisés pour des études syntaxiques	75 corpus - Environ 35 heures sur cassettes	Média	Transcription orthographique, Conventions de transcription GARS / DELIC	Données non accessibles pour l'instant	Divers articles	Paul.Cappeau@univ-poitiers.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

IMPLANTS COCHLEAIRES	DDL	HILAIRE-DEBOVE, G. Corpus constitué en vue d'études phonologiques, syntaxiques et psycholinguistiques	Enregistrements longs sur cassettes et vidéos (de 1998 à 2001).	Non média	Tous les enregistrements ne sont pas encore transcrits. Les transcriptions ont été faites par des étudiants. Utilisation de conventions existantes.	Données (textes) consultables sur place.	Publication d'études.	HILAIRE-DEBOVE, Géraldine, DDL/ISH 14 avenue Berthelot, 69363 LYON Cedex 07 Geraldine.Hilaire@ish-lyon.cors.fr
Institut Pierre GARDETTE	Université Catholique de Lyon / ICAR	Bert. M. Corpus de dialectologie collectés depuis les années 1970	300 h. d'enregistrements en cours de numérisation: francoprovençal, nord-occitan et français. Contenu : enquêtes linguistiques ou sociolinguistiques, littérature orale, ethnotextes,	Non média	Une petite partie des matériaux est transcrite	Consultation sur place. Certains corpus sont décrits dans CLAPI (http://corpus.univ-lyon2.fr/) et accompagnés d'échantillons ou	Nombreuses publications, monographies, atlas linguistiques, thèses...	michel.bert@univ-lyon2.fr
KERN	DDL	KERN, S. Corpus constitué en vue d'études syntaxiques et psycholinguistiques	Etude longitudinale de 4 enfants âgés de 8 à 25 mois	Non média	Tous les enregistrements ont été transcrits par des étudiants. Transcription en API sous LIPP.	Données (son et textes) consultables sur place. Accès sur CLAPI	Publication d'études.	Sophie.Kern@ish-lyon.cnrs.fr
L'ARCHIVE DU LACITO	LACITO	BOYD Michailovsky et JACOBSON Michel	Documentation sur les langues étudiées par les membres du Lacito (souvent des langues "minoritaires" et sans tradition d'écriture)	Non média		accès public libre : 177 enregistrements audio dont 169 avec transcription en accès protégé : 490 enregistrements audio dont 10 avec transcription dans les tiroirs : quelques centaines d'heures d'enregistrements		boydm@vjf.cnrs.fr jacobson@idf.ext.jussieu.fr
LL et KP		MULLER, C. avec LABRUNE, L., PLOOG K. Deux corpus : LL (phonologie du français d'Aquitaine) et KP, constitués en vue d'études du langage parlé et de la phonologie régionale.	Deux corpus subdivisés en séquences courtes, enregistrées sur cassettes, vidéos et minidisques. L'ensemble correspond à 14h d'enregistrement (1999 à 2001).	Non média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants et des chercheurs. Pas de conventions de transcription.	Le corpus KP est à la disposition de la DGLFLF. Le corpus LL est consultable mais il est encore utilisé dans une recherche en cours. La mise en ligne devra respecter l'anonymat des locuteurs.	Quelques articles.	MULLER, Claude , Université Michel de Montaigne. Bordeaux 3, UFR Lettres, 33607 Pessac. muller@montaigne.u-bordeaux.fr LABRUNE, Laurence laurence.labrune@wanadoo.fr PLOOG, Katja katploog@yahoo.fr
LYON II		MARCONOT, J.-M.					MARCONOT, Jean-Marie, 1985, <i>L'analyse de la conversation. Le livre de Vauvert</i> . Mussidan: FEDEROP. ISBN 3-87808-691-1.	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

LYON II- RITTAUD		RITTAUD-HUTINET, Ch.			20h de conversations en transcription phonétique et 20h en transcription orthographique			
LYON-III		J. GOUDET, Centre de linguistique appliquée et de Démodialectologie. Cf. Maunier / Crespo, 1981.	20h d'enregistrement		Transcription phonologique.			
MALECOT		"Prononciation des milieux dirigeants parisiens". Cf. Malecot, 1972, W. Ashby, 1976.	50 enregistrements.	50 entrevues faites à Paris				
MICROFUSEES		UMR GRIC (CNRS Lyon II), équipe COAST	4 enregistrements correspondant à 30 mn	Non média: dialogues qui ont pour contexte une journée d'initiation à la fabrication de microfusées.		Données accessibles sur le net (site ASILA)		Suzanne Salmon-Alt suzanne.alt@loria.fr
MONTPELLIER		Corpus de BARBERIS, BOYER, université Paul-Valéry de Montpellier, constitué pour l'analyse du discours.					BARBERIS, Jeanne-Marie (éd.), <i>Le français parlé, variété et discours</i> . Montpellier: Praxiling.	
NANCY		CARTON, F. "Le français régional de la région de Lille". Cf. F. Carton, 1971.	29 textes de variétés du français, avec cassette de 1h.		Transcription avec intonation et accentuation.	Publication dans le commerce: CARTON, F., ROSSI, M., AUTESSERRE, D. et LEON, P., 1983, <i>Les accents français</i> (texte et cassette). Paris: hachette.		Université de Nancy II, 25, rue Baron-Louis, B.P. 454, 54001 Nancy
NANCY / CANUT	ATILF (?)	CANUT, E. Corpus constitué pour l'étude de dialogues adulte-enfant (moins de six ans), axe syntaxique.	Environ 500 enregistrements courts et longs sur cassettes (1993 à 2001). Qualité médiocre (beaucoup de bruits de fond).	Non média	Les enregistrements sont presque tous transcrits. Les transcriptions ont été faites par E. CANUT et des étudiants. Conventions de transcription du centre de recherche sur l'acquisition du langage oral et écrit de Paris 3.	Données (son et textes) accessibles sur place. A discuter.		CANUT Emmanuelle 5 rue lecreux 54000 Nancy Emmanuelle.Canut@univ-nancy2.fe
NANCY / DEBAISIEUX	ATILF (?)	DEBAISIEUX, J.-M., équipe GARS / DELIC Corpus réalisé en vue d'études syntaxiques	Environ 200 enregistrements courts et longs sur cassettes et compacts disques correspondant à environ 50h. (2000 à 2004)	Média et non média.	Tous les enregistrements ne sont pas transcrits. Les transcriptions ont été faites par J.-M. DEBAISIEUX et des étudiants. Conventions de transcription de l'équipe DELIC.	Données (son et textes) accessibles sur place. Une partie du corpus a été reversée à l'équipe DELIC (Aix)		DEBAISIEUX, J.-M., Université Nancy 2 - CRAPEL - 3, place Godefroi de Bouillon BP 3397 54015 Nancy Jeanne-Marie.Debaisieux@univ-nancy2.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

NICE / CHAUVIN		(C.E.P.), M. CHAUVIN, "narrations", IDERIC	3h30 d'enregistrement.		Transcription orthographique et phonétique.			
NIMH	DDL	HILAIRE-DEBOVE, G. Corpus constitué en vue d'études phonologiques, syntaxiques et psycholinguistiques	456 enregistrements pour un total d'environ 450 h (de 2002 à 2004). Enfants de 12 à 48 mois.	Non média	Tous les enregistrements ne sont pas encore transcrits. Les transcriptions ont été faites par des étudiants (formés et rémunérés). Conventions CHILDES. Fin prévue en décembre 2006.	Données accessibles à tous sur le net, gratuites ou payantes	Publication d'études.	HILAIRE-DEBOVE, Géraldine, DDL/ISH 14 avenue Berthelot, 69363 LYON Cedex 07 Geraldine.Hilaire@ish-lyon.cnrs.fr
OTG	VALORIA	ANTOINE, Jean-Yves, GOULIAN, Jérôme, LETELLIER, Sabine (laboratoire VALORIA, Université de Bretagne Sud). "Office du Tourisme de Grenoble". Corpus de dialogue oral pour la communication Homme-Machine, constitué en vue d'études linguistiques et de traitement automatique des langues naturelles.	315 dialogues (total 2h) dans le cadre du renseignement touristique.	Non média	Transcription orthographique. 26000 mots	http://www-valoria.univ-ubs.fr/antoine/parole_public/OTG/index.html		Jean-Yves Antoine : Jean-Yves.Antoine@univ-ubs.fr Jerome Goulian : Jerome.Goulian@univ-ubs.fr
OZKAN		N. OZKAN et J. CAELEN, Corpus constitué pour étudier notamment des informations concernant le dialogue (locuteurs actes	33 dialogues correspondant à environ 11 500 mots	Non média: dialogues lors d'une tâche de positionnements d'objets.		http://www.loria.fr/projets/asila/corpus.html		Suzanne Salmon-Alt suzanne.alt@loria.fr
PARIS 3 EA1483	Langage et Langues ED268	M-A. MOREL, Centre de Recherche en morpho-syntaxe du français, corpus constitué en vue d'études syntaxiques, d'analyses intonatives et de structure du dialogue.	200 enregistrements longs sur cassettes, vidéos et compacts disques (depuis 1990)	Média et non média "A l'école": débats interviews et commentaires, 178 p. - "Conversations spontanées entre enfant et adulte", 130 p. "Débats radiophoniques", 100 p. - "Simulation d'un sondage téléphonique", 27 p. - "Conversation téléphonique", 100 p. Esquenet- Bernaudin, "Transcription d'une émission de radio de J. Chancel". Cf. Esquenet-Bernaudin, 1982, Luzzati, 1983.	Tous les enregistrements ne sont pas transcrits. Les transcriptions ont été faites par des étudiants. Conventions de transcription.	Données (son et textes) consultables sur place. Publications dans le commerce.	Exemple de publication: Corpus oraux transcrits avec courbes intonatives réalisés par les étudiants de maîtrise et de DEA, année 1992-1993, université de Paris III - Sorbonne nouvelle (env. 312p)	MOREL, Mary-Annick, 16, rue Marx Dormoy 92260 Fontenay aux Roses. Marym@ext.jussieu.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

PARIS V		Auprès de SALAZAR - ORVIG, de L. DANON-BOILEAU						
PARIS X		Corpus F. GADET, sociolinguistique, français "populaire"				Données (son et textes) consultables sur place.		
PARIS-V		D. FRANCOIS, 3 entretiens avec un informateur d'Argenteuil, 74 p. de transcription phonologique et phonétique. Cf D. François, 1974, vol. 2. Hocquet-Descamps, 1981, vol. 2, 40 p. de conversation, divers "registres", en transcription phonologique. "Recherche pédagogique sur le français oral" et "Recherche sur l'alphabétisation": Rapports de recherche des années 1973-1981; divers corpus en transcription phonétique.			Transcription phonétique et phonologique.		Publication dans le commerce: FRANCOIS, Denise, 1974, Français parlé. Analyse des unités phoniques et significatives d'un corpus recueilli dans la région parisienne. Paris: SELAF. ISBN corpus 761-838.	UFR de Linguistique Générale et Appliquée, Université de Paris-V, 12, rue de Cujas, 75005 Paris
PARIS-VII		VIOLLET, C., 3 corpus de débats constitué pour l'étude de pratiques argumentatives et de discours oral.					VIOLLET, C., 1983, <i>Pratiques argumentatives et discours oral</i> . Thèse de IIIe cycle, Paris VII, sous la direction de CULIOLI.	

PARIS-VIII		Sous la direction de P. ENCREVE, une série de corpus ayant servi de base à des thèses: Encrevé, 1967 (60h) P. CADIOT; C. CLEMENT-BOAS (40h dans la ville de St-Claude); M. de FOURNEL (ZUP de Lormont, près de Bordeaux); B. LAKS, 1980 (adolescents de la région parisienne); P. ENCREVE, L. QUERE et CONIN, conversations téléphoniques enregistrées pendant un mois (Centre d'études des Mouvements Sociaux, CNRS). Mazière, F. 1976, Conversations et monologues, enregistrés dans le limousin. R. Plas, 1979, Compréhension et mémoire d'un film: étude expérimentale						
PARIS-XIII		Corpus et transcriptions par le groupe de Ch. BACHANN et J. SIMONIN. Cf. Simonin, 1976, Bachmann, 1980.						
PASSY		Paul PASSY, 1987, <i>Les sons du français</i> , avec enregistrements sur disque. Cf. Passy / Rambau, 1987 et une étude par P. Delattre, 1966.						
PERPIGNAN		BILGER, M., équipe GARS / DELIC Corpus constitués en vue d'études syntaxiques, de régionalisme et d'enseignement.	Environ 150 enregistrements courts sur cassettes correspondant à 30 h. (1993 à 2004)	Non média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants. Conventions de transcription GARS / DELIC	Données (textes) accessibles sur place. Une partie du corpus a été reversé à l'équipe DELIC (Aix)	Publication d'études.	BILGER, Mireille, 1, rue des quinze degrés. 66000 Perpignan bilger@univ-perp.fr
PFC	MoDyCo	« La phonologie du français contemporain (PFC) : usages, variétés et structure » est un projet international sous la direction de Jacques Durand, Bernard Laks et Chantal Lyche.	60 points d'enquête (dans l'ensemble de la francophonie) avec 10 locuteurs. 900 h d'enregistrements comportant 1 lecture de mots, 1 lecture d'un texte, 1 conversation soutenue, 1 dialogue informel	Non média.	Les enregistrements sont transcrits en partie. Les transcriptions sont faites par les participants du projet: chercheurs, enseignants, étudiants. Conventions de transcription.	http://www.projet-pfc.net/?accueil.intr	Plusieurs publications dont : Delais-Roussarie & Durand Corpus et variation en phonologie du français - Méthodes et analyses, P.U. du Mirail 2 colloques par an.	Jacques Durand : jdurand@univ-tlse2.fr Bernard Laks : Laks@u-paris10.fr Chantal Lyche : Chantal.lyche@kri.uio.no

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

PIC		A. NICOLE (laboratoire GREYC, université de Caen), LPI-GRC (Université Nancy II), LPCP (Université de Caen)	1 enregistrement	Non média: dialogue de réservation		http://www.loria.fr/projets/asila/corpus.html		Suzanne Salmon-Alt suzanne.alt@loria.fr
POI		CAPPEAU Paul équipe GARS / DELIC, Corpus réalisé pour des études syntaxiques	150 enregistrements (de 15 minutes chacun) sur cassettes et minidisques depuis 1993	11 corpus viennent des médias	Tous les enregistrements sont transcrits et ont été vérifiés. Les transcriptions ont été faites par des étudiants. Conventions de transcription GARS / DELIC	La quasi totalité de ce corpus a été reversée à l'équipe DELIC (Aix)	Publication de divers articles	Paul.Cappeau@univ-poitiers.fr
RADIO-France		Cassettes et transcriptions; sélection d'enregistrements établis à l'intention des établissements scolaires par une équipe d'enseignants et responsable du C.D.I 75786 Paris cedex 16.						
REFLEX	DDL	Florence Chenu	Enregistrements numériques de -jeunes de 10 à 15 ans (120 personnes)	Non média	Transcription CLAN en cours			Florence.chenu@univ-lyon2.fr Harriet.jisa@univ-lyon2.fr
RENAULT		D. MARTINI, A. GRYL et X. BRIFFAULTLIMS	27 dialogues			http://www.loria.fr/projets/asila/corpus.html		Suzanne Salmon-Alt suzanne.alt@loria.fr
RITEL	LIMS		Recherche d'information par Téléphone. Corpus de dialogue homme / machine en vue de l'étude de recherche collaborative, de co-construction dynamique du sens et du domaine de l'interaction. Collecte septembre 2004 à janvier 2005. Questions posées par 13 personnes à un serveur : 5362 énoncés issus de 582 dialogues	Non média		Mise à disposition du corpus via "une convention de mise a disposition d'un corpus". Distribué par ELRA.	Sophie Rosset and Sandra Petel. The Ritel Corpus - An annotated Human-Machine open-domain question answering spoken dialog corpus. In International Conference on Language Resources and Evaluation, Genoa, May 2006.	rosset@limsi.fr
SPENCER		JISA-HOMBERT, H. Corpus constitué en vue d'études phonologiques, syntaxiques et psycholinguistiques	320 corpus (courts) enregistrés sur cassette	Non média	Tous les enregistrements ont été transcrits par des étudiants. Utilisation de conventions.		Publication d'études.	JISA-HOMBERT, Harriet, DDL/ISH 14 avenue Berthelot, 69363 LYON Cedex 07 Harriet.Jisa@univ-lyon2.fr

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

STRASBOURG / TABOURET- KELLER		TABOURET- KELLER, A., Corpus constitué en vue de l'enseignement, d'études sociolinguistiques et de régionalismes.	3 enregistrements sur rubans magnétiques datant de 1959 (Région de Bienne, suisse) - 2 enregistrements sur rubans magnétiques datant de 1961 (Aveyron, Hautes-Pyrénées) - 2 enregistrements sur rubans magnétiques datant de 1961 (échanges mère- enfant)	Non média.	Tous les enregistrements ne sont pas transcrits.		Pas de publication d'études.	TABOURET-KELLER, Andrée, 8, rue des Arquebusiers 6700 Strasbourg tél. 03 88 35 61 32 atekatab@club-internet.fr
THESOC	BCL (Nice)	DALBERA J.Ph. et OLIVIERI M.	Base de données multimédia. Corpus de langue occitane moderne. Données recueillies par enquêtes de terrain. Dominante lexicale. Eléments de morphologie et de syntaxe associés. Regroupe notamment l'ensemble des données de source <i>Atlas de la France par régions</i> .	Non média	3 niveaux de transcription : API, graphie phonologique, lemme	http://thesaurus.unice.fr	Diverses publications sur les données et sur leur exploitation scientifique.	olivieri@unice.fr
TOULOUSE / DUVIGNAU		DUVIGNAU Karine - étude sémantique et psycholinguistique - Corpus constitué pour une thèse	50 enregistrements qui représentent environ 5 heures	Non média - Enregistrements d'enfants en situation scolaire	Tanscription orthographique.	Données stockées sur disque dur	études publiées	duvignau@univ-tlse2.fr
TOULOUSE / GRANDATY	ISCT	Michel Grandaty	8 enregistrements en classe primaire (entre 1996 et 2005) sur cassette et en vidéo	Non média	Les transcriptions ont été faites par l'équipe de recherche	Consultables (son, texte) sur place	Plusieurs publications dans la revue Repères	michel.grandaty@toulouse.iu.fm.fr
TOULOUSE / VERGELY - PREVOT		VERGELY Pascale (Centre d'Etudes de la Navigation Aérienne et ERSS,	Le corpus "dysfonctionnement" correspond à 106726 mots,	Non média: dialogues au téléphone sur les thèmes de la résolution d'un				
TOULOUSE-1		DESPRATS, Béatrice, 1976, Le vocabulaire technique de la mégisserie à Graulhet. Textes des travailleurs des métiers de						Université de Toulouse le Mirail.
TOULOUSE-2		ICART-SEGUY, Hélène, 1976, Dialogues de femmes.						Université de Toulouse le Mirail.
TOULOUSE-3		LAFAGE, E., 1976, Les français des enfants de l'Isle en-Dodon.						Université de Toulouse le Mirail.
TOULOUSE-4		RIVENC-CHICLET, M., 1976, Eleveurs et négociants de la Haute-Saône. Dialogues de transactions.						Université de Toulouse le Mirail.

WEIL		D. WEIL, 1972, Enregistrements de 103 enfants, en milieu scolaire et de 17 enfants en milieu familial, en vue d'une étude psycholinguistique.		Non média	transcription orthographique.			
	LACITO		168 enregistrements en 32 langues	Non média				
	LIMSI			Média				
	LPL		8 h de dialogue en chambre sourde en face à face	Non média	TOE (transcription orthographique enrichie)			
	LPP							
	LIA			Média				
	LORIA			Non média				
	VALIRUN	Gudrun LEDEGEN Etude de la langue ordinaire réunionnaise (créole, français et contacts de langues	Environ 100 h (créole et français)		Transcription orthographique. Transcription alignée.	accès sur internet prochainement	Ouvrage + CD Rom sont prévus	Gudrun.Ledegen@wanadoo.fr r Gudrun.Ledegen@univ-reunion.fr
1.2. Belgique								
Nom du corpus	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	Accès aux données	Publications	Contacts	
ELICOP	Etude Linguistique de la Communication Parlée				http://bach.arts.kuleuven.be/elicop/			
ELILAP	Etude linguistique de la langue parlée. de KOCK, J., DEBROCK, M., DEBECQUE, N. et BAS, E. Le corpus est constitué de: 1. "Etude sociolinguistique sur Orléans" 2. "Le livre parlé de Tours" 3. "Voix d'Auvergne" cf. Le livre parlé de Tours dans la rubrique Royaume Uni	500 h d'enregistrements depuis la fin des années 60.		Transcriptions orthographiques (environ 100 h.) et phonétiques (12 h.) automatisées.	http://bach.arts.kuleuven.be/elicop/			
GAND	WILLEMS, D. "Les francophones de Gand" cf. Van Den Branden, 1983.	38h d'interviews		transcription orthographique		WILLEMS, Dominique, <i>Corpus des francophones de Gand</i> . Gand		

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

VALIBEL	FRANCARD, Michel et centre de recherche VALIBEL. Corpus constitué en vue d'études phonologiques, syntaxiques, sociolinguistiques, d'analyses conversationnelles et de régionalismes.	501 entrevues de 30 à 60 mn rassemblées en 22 corpus. Durée totale du corpus: 373 h. Les enregistrements ont été faits sur cassettes puis minidisques. Tous sont numérisés.	Média et non média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par le centre de recherche VALIBEL. Convention de transcriptions.	Données (son, vidéo, textes) consultables sur place. L'équipe envisage une diffusion des corpus à très court terme, pour un prix raisonnable. Le moyen de diffusion n'a pas encore été décidé (vraisemblablement, support CD-Rom).	FRANCARD, Michel, 1990, "Le français parlé des corpus oraux". Travaux de linguistique 21: 53-63. D'autres publications de corpus envisagées pour bientôt.	Francard Michel Collège Erasme, Place Blaise Pascal, B-1348 Louvain-La-Neuve, Belgique. Pour l'enquête, contacter Anne Dister, responsable du corpus: dister@tedm.ucl.ac.be Tél: 0032-10473773 (bureau) ou 0032-43804834 (domicile) Pour plus d'informations, consulter le site: http://valibel.fltr.ucl.ac.be	
1.3 Suisse								
LÜDI	Voir Fragments publiés par les cahiers de Linguistique française de Genève. (G. Lüdi, Groupe de Recherche FNRS, Bâle / Neuchâtel). Séminaires des langues et littératures romanes, Stapfelberg, 7, Ch-4051 Bâle, Suisse.	25h d'enregistrement.	Non média: entretiens avec des familles migrantes bilingues	Transcription orthographique.				
1.4. Suède								
Nom du corpus	Auteur(s), objectifs	Enregistrements	Média / Non média	Transcriptions	Accès aux données	Publications	Contacts	
FPM	FORSGREN M., Français Parlé des Médias. Corpus établi en vue d'études structurales (phonologie, syntaxe, sémantique), d'études sociolinguistiques (variations de registres) et d'analyse conversationnelles.	Environ 40h. d'enregistrements sur vidéos VHS. Numérisés sur fichiers sons wave et sur fichiers vidéo. (2000-2004).	Média: journaux FR3, Bouillon de Culture, Arrêt sur image, Rideau rouge	Tous les enregistrements ne sont pas transcrits. Les transcriptions ont été faites par 3 francophones et 3 suédois. Conventions de transcription.	Données (son, vidéo, textes) consultables sur place.			
UPPSALA / BROTH	BROTH, Mathias, "Le discours préfabriqué".	10 extraits d'enregistrements vidéos correspondant à environ 1h30.	Ce corpus comprend des segments d'enregistrement de films et de pièces de théâtre. Tous sont extraits des archives du Théâtre national de la Colline (Paris), spécialisé dans le théâtre contemporain.	Tous les extraits sont transcrits. Le système de transcription utilisé se base sur les conventions de Du Bois et al. (1993 : 45-89).			mathias.broth@romanska.u.se	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

UPPSALA / LINDQVIST	LINDQVIST, C., Corpus constitué en vue d'études phonologiques, syntaxiques, sociolinguistiques et en vue de l'enseignement	19 enregistrements longs sur vidéo (février 1993)	Média: journaux télévisés	Tous les enregistrements ne sont pas transcrits. Les transcriptions ont été faites par C. LINDQVIST. Conventions de transcription.		LINDQVIST, Christina, 2001, <i>Corpus transcrits de quelques journaux télévisés français</i> . Uppsala: Uppsala Universitet.	
UPPSALA / NOREN	NOREN, Coco, "Dialogues entre jeunes français".	9 enregistrements sur cassettes correspondant à 6 h.	Non média: dialogues entre jeunes français. 5 enregistrements ont été faits à Paris, les locuteurs sont de Rambouillet. 4 autres enregistrements ont été réalisés à Uppsala: les locuteurs sont des étudiants français venant de différentes régions de France.	Tous les enregistrements sont transcrits. Conventions de transcription orthographiques avec quelques notations prosodiques.			coco.noren@romanska.uu.se
1.5. Danemark							
CAEN	U. SAND, Jorgen et KOLSTRUP Soren, Corpus constitué pour la didactique et la recherche linguistique.	41 enregistrements sur bandes correspondant à 29 h.	Non média: Entretiens avec des écoliers (10 - 12 ans) et des apprentis, des lycéens (17-18 ans).	390 000 mots. Conventions de transcription orthographiques.			
HANSEN	HANSEN, Anita Berit, 1990, Corpus de français parisien utilisé pour l'étude sociolinguistique de la stabilisation du "e" caduc interconsonantique et l'apparition d'un "e" final.					HANSEN, Anita Berit, 1990, <i>Analyse sociolinguistique de deux évolutions linguistiques dans le français parlé à Paris. La stabilisation du "e" caduc interconsonantique et l'apparition d'un "e" final</i> . Copenhague: Université de Copenhague. (Corpus de français parisien)	
1.6. Finlande							

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

JYVÄSKYLÄ	SIHVONEN, Päivi, "Corpus de français parlé de l'Institut des langues romanes et classiques de l'université de Jyväskylä (Finlande)"	Enregistrements de 40 locuteurs sur cassettes audios et vidéos. La durée des enregistrements varie de quelques minutes à 1h 30 (1984-1993)	Média et non-média: Entretien informels ou conversations: deux locuteurs dont un (l'enquêtrice) en général non francophone; Récits, descriptions, explications: un locuteur parle d'un sujet de son choix (voyage, accident, travail, études...); Emissions de télévision ou de radio: interviews et débats.	Environ 18h d'enregistrements ont été transcrits (180 000 mots). Les transcriptions ont été faites par des étudiants et des chercheurs de l'Institut puis elles ont toutes été vérifiées par des chercheurs francophones. Conventions de transcription orthographiques adaptées de celles du GARS.			SIHVONEN, Päivi, Département des langues romanes, BP4, 00014 Université de Helsinki, Finlande. Sihvonen@helsinki.fi	
1.7. Pays-Bas								
AMSTERDAM	(Pr. B.P.F. AL), cf. Corpus d'Orléans. Depuis 1977, transcription et informatisation de 65 entrevues du corpus d'Orléans. Programmes disponibles: fabrication d'index; concordance; liste de fréquence, programmes statistiques. Université libre d'Amsterdam						Université libre d'Amsterdam	
1.8. Royaume-Uni								
Nom du corpus	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	Accès aux données	Publications	Contacts	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

BRISTOL	BEECHING, K., corpus constitué pour des études sociolinguistiques et pour l'enseignement.	95 enregistrements courts et longs sur cassettes correspondant à 17h.30 (1988-1991)	Non média. Entretiens.	Tous les enregistrements ne sont pas transcrits. Les transcriptions (155 000 mots) ont été faites par BEECHING K. Conventions de transcription	http://www.uwe.ac.uk/facults/les/staff/kb/main.html		BEECHING, K. HSL, uwe Bristol Frenchay Campus, Coldharbour lane Bristol BS 16 1QY United Kingdom Kate.Beeching@uwe.ac.uk Tél. 0044117-344-2385 (bureau) Tél. 00441225-334471 (privé)	
COVENEY	COVENEY, Aidan, corpus constitué pour l'étude de la sociolinguistique variationniste.	30 enregistrements d'environ 37 mn chacun, sur cassettes. La totalité correspond à 17h30 (1982)	Non-média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par Adian Coveney. Conventions de transcription (mais de façon limitée selon l'auteur)	Les données ne sont pas accessibles.	Plusieurs publications de Coveney sont basées sur ce corpus. Voir le site http://www.kb.nl/kb/online/	COVENEY, Aidan. Dept of French, University of Exeter, Exeter, EX4QH, Angleterre a.b.coveney@ex.ac.uk	
ESLO / ORLEANS	"Etudes sociolinguistiques sur Orléans" Cf. AMSTERDAM, BELC, ELILAP	497 enregistrements correspondant à 315 h. (1968 à 1971)	Non média. Entretiens face à face; enregistrements dans des situations sociales ou professionnelles informelles; communications téléphoniques; interviews avec des personnalités; conférences / débats (paroles publiques); enregistrements avec des témoins anonymes (parole publique différente); entretiens entre des parents d'élèves et une assistante sociale.	4 500 000 mots transcrits. Conventions de transcription.		Loneragan / Kay / J. Ross, 1974 Catalogue des enregistrements. Orleans Archive, Language Center, University of Essex, Colchester C04 3SQ, Grande-Bretagne BIGGS et DALWOOD, 1976, <i>Les Orléanais ont la parole</i> . Teaching guide and tapescript. London: Longman.		
LABEAU Cendrillon	LABEAU, Emmanuelle, Corpus réalisé pour étudier l'acquisition des temps du passé par des apprenants universitaires anglophones + groupe de contrôle francophone.	environ 100 enregistrements courts sur cassettes (automne 2003)	Non média	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants puis ont été revues par E. LABEAU	La consultation des données est à négocier.	Publication d'une étude	LABEAU, Emmanuelle School of Languages and European Studies, Aston University, Aston Triangle, Birmingham B4 7ET e.labeau@aston.ac.uk	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

LABEAU Corpus télévisé	LABEAU, Emmanuelle, Corpus réalisé pour étudier l'utilisation du passé simple en oral préparé	5 enregistrements longs sur vidéo (février 2003)	Média	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des étudiants.	La consultation des données est à négocier.	Publication d'une étude	LABEAU, Emmanuelle School of Languages and European Studies, Aston University, Aston Triangle, Birmingham B4 7ET e.labeau@aston.ac.uk	
LABEAU Modern Times	LABEAU, Emmanuelle, Corpus réalisé pour étudier l'acquisition des temps du passé par des apprenants universitaires anglophones + groupe de contrôle francophone.	67 enregistrements courts sur cassettes (automne 2002)	Non média	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par E. LABEAU	La consultation des données est à négocier.	Publications d'études	LABEAU, Emmanuelle School of Languages and European Studies, Aston University, Aston Triangle, Birmingham B4 7ET e.labeau@aston.ac.uk	
Le Livre Parlé de Tours	Corpus réalisé par des enseignants de français britanniques du College of Education de Londres. Enregistrements de 193 locuteurs de Tours (hommes / femmes), d'âge et de catégorie professionnelle différents. Voir ELILAP (section 1.2.)	193 enregistrements correspondant à 120 h. (1974)	Non média. Entretiens face à face.	Tous les enregistrements ne sont pas transcrits. Les transcriptions orthographiques correspondent à 4 h. (36 000 mots).	103 h. d'enregistrement sont disponibles au département de linguistique de K.U. Leuven.			
FLLOC	MYLES, Florence, MARSDEN Emma, RULE Sarah, MITCHELL Rosamond (Université de Southampton) Etude du développement linguistique chez des apprenants de français langue étrangère. Projet de recherche basé sur l'analyse de quatre tests oraux individuels administrés à 60 élèves de différents niveaux (en 3ème, 4ème et 5ème année d'étude du français dans le contexte du système scolaire britannique), et destinés à obtenir des structures morpho-syntaxiques variées (structure de la phrase, négation, morphologie verbale, genre, structures interrogatives, subordination, référence pronominale, langage formulaïque etc...).	635 enregistrements - 60 h. d'enregistrements numérisés au moyen de 'Sony Memory Stick IC Recorders' qui s'ajouteront à environ 200 heures provenant d'un corpus oral d'interlangue française recueilli durant l'étude longitudinale du projet 'Progression in Foreign Language Learning' (Mitchell and Dickson 1997)	Non-média.	Tous les enregistrements sont transcrits. Les transcriptions ont été faites par des chercheurs. Mise au point de conventions.	Données accessibles sur le net. Possibilité de les télécharger sur le site internet suivant : www.flloc.soton.ac.uk		fjm@soton.ac.uk	

VOIX D'AUVERGNE	Corpus réalisé par des enseignants de français britanniques. Conversations avec plusieurs locuteurs d'Auvergne.	52h d'enregistrement (1976)	Non média. Tables rondes et entretiens face à face.	17 h. d'enregistrement sont transcrits en orthographe (177 000 mots). Un seul extrait est transcrit en phonétique. Conventions de transcription.				
1.9. Allemagne								
Nom du corpus	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	Accès aux données	Publications	Contacts	
BIELEFELD	Elisabeth SCHMALE-BUTON & Günter SCHMALE, Bielefelder Text-corpora Romanischer Sprachen	Environ 1h 40 d'enregistrement sur bandes audio (1984).	Conversations téléphoniques. Thèmes abordés : communications privées, communications professionnelles (chez le docteur, chez le marchand de luminaires, chez le kinésithérapeute, chez le boucher, chez le boulanger, taxi, SNCF, etc.), conversations institutionnelles (dans un établissement scolaire – téléphone intérieur/extérieur)	Tous les enregistrements sont transcrits. Transcriptions avec intonation et prosodie.	Les enregistrements ne sont pas en libre accès afin de préserver le caractère privé des entretiens. Dans les transcriptions les noms de personnes, les lieux... ont été changés.	SCHMALE-BUTON, E., 1984, <i>Französisch I. Conversations téléphoniques</i> (Bielefeld Text-Corpora romanischer Sprachen). Bielefeld: Universistät Bielefeld. LUDWIG, Ralph, 1988, Korpus: Texte des gesprochenen Französisch. Tübingen: Narr. Wüest, Jakob Th. & Kristol, Andres M. (éds.) (1993) : <i>Aqeras montanhas. Etudes de linguistique occitane: Le Couserans</i> (Gascogne pyrénéenne). Tübingen / Basel: Francke (corpus de l'occitan gascon, mais contient quelques textes en français régional).		
GOEBL	GOEBL, HANS (Salzburg) Atlas linguistique.				CDRom dans le commerce.			
GÜLICH	GULICH, Elisabeth, et ses collaborateurs. Le corpus comprend une partie du corpus d'orléans et d'autres enregistrements divers.		Média et non média.	transcription orthographique		GULICH, Elisabeth, 1970, <i>Makrosyntax der Gliederungssignale im gesprochenen Französisch</i> . München: Fink (corpus 20-24, 35-37, 57-62, 68-70, A1-A77)	Lili Fakultät, Universität Bielefeld, 4800 Bielefeld 1, Allemagne	

HOLTUS	HOLTUS, G., "Le français chez les français". Université de Saar.	140 enregistrements	interviews	transcription orthographique				
KOTSCHI	Th. KOTSCHI et ses	10 à 12h d'enregistrement.	conversations, interviews,	transcription orthographique		ESCHMANN, Jurgen,	nstitut für romanische	
1.10. Autriche								
RINDLER-SCHJERVE	RINDLER-SCHJERVE, R."Le français en Alsace".	50 h d'enregistrement	Conversations	transcription orthographique; saisie sur ordinateur			Institut für Romanistik Universität Wien, Universitätsstrasse, 7, 1010 Vienne, Autriche.	
1.11. Etats-Unis								
VIOLIN-WIGENT / MICHIGAN	VIOLIN-WIGENT Anne (Michigan State University, Etats-Unis). Corpus constitué dans le cadre d'une thèse et en vue d'une étude phonétique sur les voyelles nasales.	3h. d'enregistrement sur cassettes.	Non média: conversations entre 8 locuteurs pendant des repas de famille.	La moitié des enregistrements sont transcrits. Les transcriptions ont été faites par Anne Violin-Wigent. Pas de conventions de transcription.			Anne Violin-Wigent : violinwi@pilot.msu.edu	
2. PROJETS DE CORPUS								
2.1. France								
Nom du corpus	Responsable, objectifs	Enregistrements	Média / non média	Transcriptions	Accès aux données	Publications	Contacts	
ACI	DELAIS-ROUSSARIE, E., équipe de recherche ERSS, Université de Toulouse le Mirail. Constitution de corpus de français parlé dans Toulouse et sa région.							
CRFP-2	VALLI, A. (responsable), "Corpus de référence du français parlé 2". Corpus constitué en vue d'études linguistiques	50 enregistrements de 1 h réalisés sur minidisques	Média (France Culture)	25 enregistrements sont transcrits et segmentés (sous Transcriber). Les transcriptions ont été faites par des transcribers multiples. Ces transcriptions sont à réviser.	L'accès à une partie est prévu sur le net		VERONIS, Jean, Université de Provence 29 avenue Robert Schuman - 13100 Aix-en-Provence jean.veronis@up.univ- mrs.fr	

2.2. Belgique								
BRUXELLES / DELVENNE - PIRLOT-SADOWSKI	PIRLOT, Barbara SADOWSKI, Gilles, DELVENNE, Sylvie (Université Libre de Bruxelles). Analyse des témoignages audiovisuels de rescapés de camps nazis; représentation écrite et exploitation linguistique des données. Ce projet porte sur l'analyse linguistique de témoignages enregistrés, face à une caméra, par des rescapés des camps nazis. Les entretiens sont menés en français avec des personnes d'origines diverses mais habitant en Belgique.	Enregistrements audiovisuels réalisés de 1992 à 1998.	Non média. Entretiens qui se constituent comme des récits de vie partant de l'enfance du narrateur jusqu'à la période d'après-guerre.	Conventions de transcription.	La majorité du matériel à traiter a été produite par la Fondation Auschwitz, en partenariat avec l'ULB et l'Université de Yale (Projet des Archives Fortunoff). Les entretiens sont déposés dans les archives de l'université et le fond géré par l'Unité de Recherche Sources Audiovisuelle en Histoire Contemporaine (URSA)		Sylvie Delvenne : sdelvenn@ulb.ac.be	
2.3. Suisse								
BALE et NEUCHATEL	MONDADA, Lorenza "La construction interactive du discours scientifique en situation plurilingue". L'objectif de ce projet est le recueil et l'analyse de conversation multi-média.	100 h d'enregistrement audio de trois groupes de chercheurs en sciences humaines. 40 h d'enregistrement vidéo pour un groupe de chercheurs en chirurgie.	Interactions professionnelles: réunions de travail de quatre groupes de chercheurs (université de Bâle, Strasbourg, Freiburg)		Corpus contenant des données confidentielles.		site : http://www.romsem.unibas.ch/linguistique/lorenza/index.html ou contacter : lorenza.mondada@unibas.ch	
NEUCHATEL / PY	PY, Bertrand "Les représentations sociales du bilinguisme et de l'apprentissage des langues". . Corpus constitué d'entretiens entre un chercheur et des groupes de 3 ou 4 personnes . Le but de ces entretiens a été de réunir des données discursives pour une étude sur les représentations sociales du bilinguisme. La plus grande partie du corpus est en français, mais il y a aussi un peu de suisse allemand.	Enregistrements réalisés entre 1998 et 1999	Entretiens entre un chercheur et des groupes de 3 ou 4 personnes (acteurs du système pédagogique concernés par les langues étrangères, y compris parents et élèves).	Transcriptions des entretiens effectués par l'équipe de Neuchâtel.	S'adresser à Berytrand PY pour obtenir une version papier contre remboursement.		site www.unine.ch . bernard.py@lettres.unine.ch	

Inventaire des corpus oraux - Projet DGLFLF - Paul CAPPEAU (et Magali SEIJIDO)

<p>NEUCHATEL / MAITRE - WILD</p>	<p>Raphaël MAITRE et Véronique WILD (Université de Neuchâtel) "Un corpus bilingue français - dialecte franco-provençal en situation de contact à Evolène en Valais." Corpus réalisé en vue d'études sur le bilinguisme. Domaines examinés: linguistique, sociolinguistique et ethnolinguistique</p>	<p>Enregistrements réalisés en octobre 2001</p>	<p>Non-média: entretiens exolingues-bilingues entre deux chercheurs et un ou deux habitants.</p>	<p>transcriptions en cours. Elles devraient aboutir à 300 000 mots. Conventions de transcription.</p>			<p>Raphaël MAITRE: rmaitre@datacomm.ch</p>	
<p>2.4. Suède</p>								
<p>INTERFRA</p>	<p>Dirigé par Inge BARTNING " Interlangue française - développement, variation et interaction". L'objectif a été d'établir un corpus de l'interlangue française d'un certain nombre d'apprenants suédois au niveau universitaire, en vue de saisir des traits typiques de leur syntaxe et de leur morphologie. intérêts aussi pour l'organisation du discours et pour les stratégies acquisitionnelles.</p>	<p>160 enregistrements d'étudiants (320 000 mots) et enregistrements de Lycéens (80 000 mots)</p>	<p>Non média: interviews, récits de bandes dessinées et de films vidéos.</p>	<p>Transcriptions orthographiques. Conventions de transcription.</p>				
<p>FPM</p>	<p>BROTH, Mathias, FORSGREN, Mats et NOREN, Coco. Collaboration entre l'université de Stockholm et l'université d'Uppsala. "Le français médiatique". Le but global est de mener une analyse à trois niveaux: structurel, argumentatif-rhétorique, interactionnel, en tenant compte notamment de l'aspect formel / informel du discours médiatique.</p>	<p>à terme 50 à 100 h. d'enregistrements.</p>	<p>Constitution d'un corpus d'émissions non-fictionnelles, débats, discussions, documentaires, qui seront transcrits dans leur totalité (à moins que le discours soit une lecture d'un manuscrit préfabriqué).</p>				<p>mathias.broth@romanska.uu.se mats.forsgren@fraitu.su.se coco.noren@romanska.uu.se</p>	