

Corpus : 52

Le tableau ci-dessous présente l'ensemble des corpus de l'équipe A – Linguistique – du laboratoire, un point central de l'identité de l'équipe : <http://forellis.labo.univ-poitiers.fr/vie-du-laboratoire/corpus/>

Corpus de français :

Type	Intitulé	Contenu	Taille	Partenaire(s)	Accessibilité / Outillage	Contact
Oral Apprenants allophones	AvEx-FLE	Productions orales argumentatives d'apprenants de niveau avancé et expérimenté.	51 productions 26 000 mots	Universités de Nis et de Novi Sad (Serbie), de Banja Luka (Bosnie Herzégovine), de Varsovie (Pologne) et Centre FLE Université de Poitiers.	.txt et .doc et .mp4 pour les enregistrements vidéo.	Antonin BRUNET antonin.brunet@univ-poitiers.fr
Oral Apprenants allophones		Productions orales argumentatives de locuteurs francophones natifs.	44 productions 39 000 mots		.txt et .doc et .mp4 pour les enregistrements vidéo.	Antonin BRUNET antonin.brunet@univ-poitiers.fr
Oral Apprenants allophones	Corpus Chy-FLE	Productions orales guidées et libres	6h33min43sec	Univ. Chypre	Transcriptions au format .txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr
Oral Apprenants allophones	Corpus Emo-FLE	Productions orales guidées et libres (dont une partie dans le cadre du projet Emo-FLE) - d'apprenants chypriotes hellénophones - d'apprenants grecs hellénophones - d'apprenants polonophones - d'apprenants sinophones - de locuteurs français	Emo-FLE CY : 22 productions soit 2h36min55sec GR : 20 productions soit 2h05min23sec minutes PL : 21 productions soit 3h01min29sec CH : 13 productions soit 1h17min12sec FR : 19 productions soit 1h11min9sec Entretien général –	Univ. Chypre, Univ. de Thessaloniki, Univ. de Varsovie	Transcriptions au format .txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr

			discours argumentatif CY : 22 productions soit 3h20min08sec GR : 20 productions soit 4h01min37sec minutes PL : 21 productions soit 2h51min59sec CH : 13 productions soit 2h33sec FR : 19 productions 46min41sec			
Oral Apprenants allophones	Corpus IPFC	- Répétition d'une liste de mots français (environ 65 mots) - Lecture à voix haute un court texte en français - Production orale guidée et libre	20 apprenants grecs 35 apprenants chypriotes		Annotation des liaisons et du schwa	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr
Oral	Corpus scolaire. Les gestes langagiers de l'enseignant	interactions langagières enseignant/élèves (9-10 ans)	environ 30h d'enregistrements filmiques (32 séances d'apprentissage)	Projet interne INSPE	Transcriptions au format .doc	Muriel CORET muriel.coret@univ-poitiers.fr Malika.KAHERAOUI malika.kaheraoui@univ-poitiers.fr Stéphanie VOLTEAU stephanie.volteau@univ-poitiers.fr Pierre SÉMIDOR
Oral	Corpus scolaire. Argumentation dans les débats philosophiques à l'école et collège.	Productions orales. interactions enseignant/élèves (5 à 11 ans)	environ 25h d'enregistrements filmiques	Projet interne INSPE, en collaboration avec X. Lerner	Transcriptions au format .doc	Malika KAHERAOUI malika.kaheraoui@univ-poitiers.fr

Oral	Corpus scolaire - Nom du projet : "interactions langagières et pratique d'expression créative à partir du médiateur argile à l'école maternelle	Interactions langagières entre enseignant / élèves et entre élèves dans des ateliers d'arts à l'école maternelle (recueil longitudinal > même groupe d'élèves de la petite section à la grande section)	19 séances (ateliers de 45 minutes à 60 minutes)	Projet interne de l'INSPE (axe : apprentissages et apprenants)	Transcriptions au format .txt	Stéphanie VOLTEAU stephanie.volteau@univ-poitiers.fr
Oral	Corpus Dekhissi sur le cinéma des banlieues	45 films du cinéma de banlieue	environ 67h d'enregistrement	A. Coveney Univ. Exeter, R.U.	Transcriptions au format .doc	Laurie DEKHISSI Laurie.Dekhissi@univ-poitiers.fr
Oral	Engagements des parents entendants lors de l'annonce de la surdité de leur enfant	Entretiens semi guidés	115 entretiens semi-guidés menés entre la France, le Canada, la Suisse et la Belgique	Université de Moncton	Guide In Vivo ((logiciel Nud*ist NVivo)(Qualitative Research Solution, Pty Ltd, Australia)	Charles GAUCHER charles.gaucher@umoncton.ca Stéphanie GOBET stephanie.gobet.jacob@univ-poitiers.fr
Écrit Apprenants allophones	Didacquis	Textes narratifs Textes argumentatifs	156 copies 43803 mots	Université de Sarajevo	Transcription au format .doc	Emir SISIC emirsisic@hotmail.com
Écrit Apprenants allophones	Corpus Chy-FLE	Textes argumentatifs d'apprenants chypriotes hellénophones de niveau intermédiaire/avancé	150.000 mots	Université de Chypre	Transcriptions au format .txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr
Écrit Apprenants allophones	Corpus CFLE	Copies du test de placement des niveaux B1-B2 d'apprenants d'origines diverses étudiant le français en France.	+ 75.000 mots		Transcriptions au format .txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr
Écrit Apprenants allophones	Corpus Emo-FLE	Productions écrites d'apprenants permettant l'étude de l'expression des émotions : hellénophones, polonophones, sinophones et francophones	30.000 mots	Univ. Chypre, Univ. de Thessaloniki, Univ. de Varsovie,	Transcriptions au format .txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr

<p>Ecrit Monolingue Diachronique</p>	<p><i>Observations de l'Académie française sur les Remarques de Vaugelas</i> (1704)</p>	<p>Les remarques élucident des problèmes d'acceptabilité linguistique</p>	<p>191 000 mots</p>	<p>B. Colombat, univ. Paris VII; W. Ayres-Bennett, univ. Cambridge, R.U. Marc Bonhomme Simone Delesalle Isabelle Ducharme A.M. Fryba-Reber André Horak Marcus Keller Douglas Kibbee Francine Mazière Magali Seijido Gilles Siouffi B. von Gemmingen Chantal Wionet</p>	<p>Version balisée, en ligne sur le portail « Grammairiens » et remarqueurs des XVI^e-XVIII^e siècles » chez Garnier numérique (accès par abonnement) ; édition «diplomatique»; possibilité d'afficher une graphie modernisée et accès à un fac-simile (en mode image) du texte.</p>	<p>Philippe CARON philippe.caron@univ-poitiers.fr</p>
<p>Ecrit Diachronique</p>	<p><i>Dictionnaire critique de la langue française</i> de J.-F. Féraud (1787-1788)</p>	<p>Cette œuvre présente une mise à jour fine et critique de l'usage autorisé des années 1760-1780</p>	<p>Environ 25000 articles à une ou plusieurs entrées</p>	<p>G.E.H.L.F -ENS de Paris Univ. Chicago (ARTFL) INaLF C.N.R.S. Equipe de Sciences du Langage de l'Université de Limoges IUFM du Limousin ATILF Nancy M.S.H.S. Poitiers</p>	<p>Version complète et balisée au format XML. Accès libre sur les sites ATILF, LDI-Cergy, MAHA-Poitiers, Univ. Montréal; texte très fidèlement resaisi en tenant compte de ses propriétés typographiques, à l'exception de la lignaison Cahier de saisie disponible à l'URL http://homes.chass.utoronto.ca/~wulfic/siehl/dfa/feraud/</p>	<p>Philippe CARON philippe.caron@univ-poitiers.fr et Louise DAGENAIS (Univ. Montréal) louise.dagenais@umontreal.ca</p>
<p>Ecrit Diachronique</p>	<p><i>Dictionnaire critique de la langue française</i> et son supplément manuscrit de J.-F. Féraud</p>	<p>Cette œuvre présente une mise à jour fine de l'usage Environ</p>	<p>25000 articles à une ou plusieurs entrées autorisés des années 1760-1780</p>	<p>G.E.H.L.F -ENS de Paris UMR METADIF (Cergy) puis LDI (Cergy) Equipe de Sciences du Langage de l'Université de Limoges</p>	<p>Même saisie que la version précédente pour la version imprimée plus le scan des trois volumes du supplément manuscrit</p>	<p>Philippe CARON philippe.caron@univ-poitiers.fr et Louise DAGENAIS (Univ. Montréal) louise.dagenais@umontreal.ca</p>

				ATILF Nancy M.S.H.S. Poitiers		
Écrit Synchronique		Corpus de discussions Whatsapp dans un contexte professionnel de télétravail	texte+emoji		txt, docx	Inaam JAFFEL inaam.jaffel@univ-poitiers.fr
Écrit Synchronique	Le français académique	Les rapports de stages des L3/M1	500.000 mots		.txt	Freiderikos VALETOPOULOS fvaletop@univ-poitiers.fr
Écrit Synchronique	LexiFLE	Scannage de manuels FLE (30 manuels FLE publiés après 2010), élaboration d'un référentiel, analyse du lexique des manuels et évaluation des productions des apprenants FLE	en cours		création d'une base de données	F. Valetopoulos freiderikos.valetopoulos@univ-poitiers.fr E. Lamprou effrosyni.lamprou@univ-poitiers.fr M. Nauge michael.nauge@univ-poitiers.fr
Écrit Français	Procédés définitoires à l'oral	Transcription d'enquêtes orales pour demander de définir des mots. Étiquetage de procédés définitoires	11524 mots		Interne. Annotation manuelle assistée	Marie-Hélène Lay Marie- helene.lay@univ- poitiers.fr
Multimodal Apprenants allophones	Corpus MVT ANL	productions orales en classe de langue interactions enseignants apprenants ou apprenants- apprenants	13 séances de classes de 40min filmées + 64 entretiens filmés de 15min chacun enseignant- apprenants	Université Normale de Chine du Sud	Transcriptions au format .doc	Julie RANCON julie.rancon@univ-poitiers.fr
Multimodal Apprenants allophones et apprenants français	Corpus TRAIMA	productions orales en classe de langue interactions enseignants- apprenants ou apprenants- apprenants	27h de cours de FLE et FLM filmés	Laboratoire TECHNE	Transcriptions au format .doc	Julie RANCON julie.rancon@univ-poitiers.fr
Multimodal	Corpus scolaire : Étude longitudinale de la	interactions langagières entre enseignant et élève(s) lors de dictées à l'adultes individuelles	292 séances enregistrées (plus de 40h) avec une		Transcriptions d'un corpus de travail de 108 séances + 108 productions	Gilles TABOURDEAU

	pratique de la dictée à l'adulte en classe multi-âge de maternelle	et collectives + productions écrites finales. Recueil longitudinal de la petite à la grande section de maternelle.	cohorte de 38 élèves. + 292 productions écrites finales		finale au format texte	gilles.tabourdeau@univ-poitiers.fr
Multimodal	Rugby Coupe du Monde 2015	15 matchs de rugby (10 en français et 5 en anglais)	1.05 Go	Région Nouvelle Aquitaine Université Bordeaux Montaigne	Dépôt des enregistrements sur la plateforme HumaNum	Responsables du projet : Sylvie Hanote (UP), Catherine Mathon (U. Bordeaux Montaigne)

Corpus d'autres langues / Corpus bilingues :

Langue(s)	Type / Intitulé	Contenu	Taille	Partenariats	Accessibilité / Outillage	Contact
Albanais	Corpus de littérature	Auteurs albanais et auteurs traduits en albanais du XXe s.	2 500 000 mots			Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr
Albanais	Corpus de presse	Presse du Kosovo et d'Albanie	1 000 000 mots		En cours d'étiquetage et de balisage	Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr
Allemand	Corpus mots morphologiquement simples écrit+oral (oral à consolider et intégrer)	Diachronie (Vha > Nha) + Synchronie (20°) Formes de Nha + formes de pré-Nha, Mha et Vha si disponibles	~ 20000 entrées		Accessible sur demande à des fins de recherche Disponible en version Access	Emilie CARATINI emilie.caratini@univ-poitiers.fr
Allemand	écrit+oral Corpus hypocoristiques	Synchronie (21°) Formes « pleines » (= non tronquées) + hypocoristique(s) et enregistrements correspondants	250 mots		Accessible sur demande à des fins de recherche Disponible en version Access	Emilie CARATINI emilie.caratini@univ-poitiers.fr
Anglais (R.U.) Diachronique	Corpus lexicophonétique comparative (DicoDiachro)	8 dictionnaires orthoépiques : Bailey (1727), Barclay (1774), Buchanan Ladies (1757a), Buchanan (1766), Stephen Jones (ca 1805), Walker (1791, [1809]),	260 963 entrées lexicales		Base de données interopérable ; outil d'interrogation : Python,	Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr

		<p>Sheridan (1780), Wells 1ère ed. (1990).</p> <p>En cours de numérisation / traitement : Buchanan Senior (1757b), Wright, 5 vol. (1852-1856).</p> <p>A venir : Afzelius (1909) et Pierce (1910), ainsi que les dictionnaires d'anglais américains Worcester (1830, 1846, 1860), Webster (1806, 1820), Morris (1908), Kenyon & Knott ainsi que la 4^{ème} édition de A. Spiers (1851)</p>			<p>Document computationnel : Jupyter Notebook</p> <p>Balisage XML-TEI pour Walker</p> <p>Démo : https://videot.heque.univ-poitiers.fr/video.php?id=ar1hj59pjzh8vw5zctc&link=5swajkajh4klb3a7sw4cl2nnszbvzbz</p>	<p>HANOTE Sylvie sylvie.hanote@univ-poitiers.fr</p> <p>Michaël NAUGE michael.nauge@univ-poitiers.fr</p> <p>Nicolas VIDEAU nicolas.videau@univ-poitiers.fr</p> <p>Associés : Nicolas Trapateau (U. Côte d'Azur) : nicolas.Trapateau@univ-cotedazur.fr</p> <p>Franck Zumstein (U. Paris) : franck.zumstein@univ-paris-diderot.fr</p>
Anglais écrit / oral	Corpus de néologismes par abrègement de l'anglais	Corpus d'abrègements collectés à partir de divers médias sociaux (2015-2020)	environ 800 entrées lexicales		Tableau Excel par entrées lexicales + Fichiers audios	Pauline LEFEBVRE pauline.lefebvre@univ-poitiers.fr
Grec	Corpus de textes journalistiques	<p>Quotidien Φιλελεύθερος</p> <p>Quotidien Σημερινή</p> <p>Quotidien Χαραυγή</p>	<p>2.400.000 mots</p> <p>1.150.000 mots</p> <p>1.450.000 mots</p>			Freiderikos VALETOPOULOS freiderikos.valetopoulos@univ-poitiers.fr
Occitan oral/écrit	OcOr (Occitan Oral) : corpus comparatif de narration orale en occitan	Une soixantaine de contes et récits	plus de 50 000 mots	<p>Accessible sur zenodo :</p> <p>https://zenodo.org/record/1451753#.X89FR25Fxoo</p> <p>Ce corpus a été réalisé dans le cadre du projet ExpressioNarration, financé par une bourse Marie Sklodovska Curie Fellowship (2016-2018, n°655034) en collaboration avec Janice Carruthers</p>	<p>Transcriptions disponibles au format XML TEI P5 (brutes ou annotées : temps verbaux, cadres de discours et connecteurs)</p>	Marianne VERGEZ-COURET marianne.vergez_couret@univ-poitiers.fr

				(Université Queen's de Belfast)		
Poitevin-Saintongeais Ecrit	Corpus TELPOS (Textes Electroniques en Poitevin-Saintongeais)	Diachronie – Synchronie (16°-20°) Une centaine d'écrits allant du 16° au 20° siècle.	400 000 mots	Accessible sur demande à des fins de recherche Ce projet a bénéficié par deux fois d'un soutien de la Région Poitou-Charentes.	1. Disponibles en version .txt. (ASCII) 2. Documents encodés TEL sous une base XTF. (UTF-8)	Marie-Hélène LAY marie-helene.lay@univ-poitiers.fr & Jean-Christophe DOURDET jean-christophe.dourdet01@univ-poitiers.fr
Poitevin-Saintongeais Ecrit	TELPOS annoté	4 textes de la base TELPOS partiellement annoté thématiquement	120000 mots		Interne, Annotation manuelle assistée	Marianne VERGEZ-COURET marianne.vergez_couret@univ-poitiers.fr
Bilingue Anglais / Français Oral	Corpus « Parole »	Données authentiques issues d'émissions radiophoniques britanniques, américaines et françaises	120 000 mots		Aligné audio/transcriptions sous Praat. Accessibilité sur le site FoReLLIS et via Humanum pour les membres du projet	Sylvie HANOTE sylvie.hanote@univ-poitiers.fr Nicolas VIDEAU nicolas.vidEAU@univ-poitiers.fr
Bilingue Français, Anglais, Oral	Corpus d'enfants bilingues	Transcriptions d'enfants bilingues anglais-français	25 enfants de 2 à 6 ans		Site du laboratoire FoReLLIS	Pierre-Don GIANCARLI pierre-don.giancarli@univ-poitiers.fr
Bilingue Anglais-Français, Ecrit	Corpus juridique	Section 6 de l'acquis communautaire de l'UE + 20 ans de décisions bilingues de la Cour Suprême du Canada	30 000 000 mots 45 000 000 mots		Bilingue aligné	Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr & Manuel TORELLAS manuel.torrellas@univ-tours.fr
Bilingue Arabe - Français, Ecrit	Corpus économique et financier	Documents publiés en français et en arabe par la Banque Mondiale, l'ONU, et le Fonds Monétaire international			Bilingue aligné	Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr

Bilingue Espagnol - Français, Ecrit	Corpus juridique	Acquis communautaire de l'Union Européenne : en espagnol et français	200 000 000 mots		Bilingue aligné	Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr & Manuel TORELLAS manuel.torrellas@univ-tours.fr
Bilingue Français/Langue des signes, Oral	Corpus Bébé signe le bébé signe et l'accompagnement à la communication chez les bébés	Apprentissage des signes par les enfants et productions lexicales vocaux et signes (combinaisons de mots et combinaisons de mots et de signes).	Minimum de 40 enfants France	CNRS Lyon	Corpus vidéo transcrit avec Elan	Stéphanie GOBET stephanie.gobet.jacob@univ-poitiers.fr
Bilingue Anglais Français, Oral	Corpus de discours vidéoludique (League of Legends)	Corpus de commentaires e-sportifs du jeu League of Legends + divers autres types de discours oraux (streams, échanges entre joueurs, vidéos Youtube)	200 000 mots		transcription orthographique + TextGrid Praat pour une grande partie du corpus	Sophie KRAEBER sophie.kraeber@univ-poitiers.fr
Bilingue Roumain Français Ecrit	Guides touristiques	Corpus comparable de guides touristiques, originaux, rédigés en roumain et en français, sur la Roumanie	870 000 mots		txt et docx	Ioana Daniela BALAUTA ioana.daniela.balauta@univ-poitiers.fr
Multilingue Anglais, Français, Espagnol, Grec, Roumain Ecrit	Corpus « GRAFE »	Corpus littéraire bidirectionnel (parallèle et comparable) : originaux des 5 langues et traductions dans les 5 langues bébé	1 million de mots dans les originaux et les traductions	Univ. Limoges (Projet PRES), univ. Chypre, univ. Tours	Accessible aux membres du laboratoire et aux extérieurs sur demande sur HumaNum Aligné à la phrase Alinéa (M. Nauge) interrogeable avec concordancier multilingue (ParaConc)	Raluca NITA raluca.nita@univ-poitiers.fr & Ramon MARTI-SOLANO ((U. Limoges)
Multilingue Anglais, Français, Espagnol	Corpus « GRAFE »	Corpus scientifique, originaux FR et EN, traduits en EN, FR, ES	869.224 mots (originaux et traductions)	Univ. Limoges (Projet PRES), univ. Chypre, univ. Tours	Accessible aux membres du laboratoire et aux extérieurs sur demande sur HumaNum Aligné à la phrase Alinéa (M. Nauge)	Raluca NITA raluca.nita@univ-poitiers.fr & Ramon MARTI-SOLANO (U. Limoges)

Ecrit					interrogeable avec concordancier multilingue (ParaConc)	
Multilingue Anglais, Français, Allemand Ecrit	Corpus « PLECI »	- Corpus de presse: originaux français et anglais, traductions en anglais, français et allemand (1999-2012)	Originaux français: 370 000 mots Originaux anglais: 330 000 mots Originaux allemand: 135 000 mots	Université Catholique de Louvain (CECL)	Une partie du corpus (environ 400 000 mots) est accessible en ligne: http://cabal.rezo.net/ -Accessible aux membres du laboratoire et aux extérieurs sur HumaNuM Aligné manuellement à la phrase, interrogeable avec concordancier multilingue (ParaConc) -Enrichi (2018 : Stages Licence)	Raluca NITA raluca.nita@univ-poitiers.fr
Bilingue Anglais, Français, Ecrit	Corpus « PLECI »	- Corpus de romans (20 ^e s.): originaux français et anglais, traductions en anglais et français	Originaux français: 350 000 mots Originaux anglais: 370 000 mots	Université Catholique de Louvain (CECL)	-Accessible aux membres du laboratoire et aux extérieurs sur HumaNuM -Aligné manuellement, interrogeable avec concordancier multilingue (ParaConc) -Enrichi (2018 : Stages Licence)	Raluca NITA raluca.nita@univ-poitiers.fr
Multilingue Ecrit Français, Anglais, Roumain	Corpus juridique	Documents contractuels notariés authentiques	300 000 mots par langue			Jean-Louis DUCHET jean-louis.duchet@univ-poitiers.fr

<p>Multilingue</p> <p>Occitan</p> <p>Poitevin-Saintongeais, Portuguais, Wallon, Hongrois</p> <p>Oral</p>	<p>Corpus VALIERE Civraisien MSHS</p>	<p>Diachronie</p> <p>Synchronie</p> <p>(19°-21°)</p> <p>contes, chansons, récit de vie, récits techniques</p>	<p>Heures d'écoute</p> <p>(200 bandes magnétiques déposées à la MSHS, copies des originaux)</p>	<p>Accessible sur demande à des fins de recherche</p>	<p>Fichiers audio non compressés de type .wav</p> <p>Numérisation par les Archives Départementales de la Vienne.</p>	<p>Estelle DUPUY</p> <p>estelle.dupuy@univ-poitiers.fr</p> <p>&</p> <p>Jean-Christophe VALIERE</p> <p>Jean-christophe.valiere@univ-poitiers.fr</p> <p>&</p> <p>Jean-Christophe DOURDET</p> <p>jean.christophe.dourdet01@univ-poitiers.fr</p>
<p>Multilingue</p> <p>Français</p> <p>Langue des signes Française</p> <p>Oral/écrit</p>	<p>Corpus MOSE (Modalité Orale Signée et Ecrite)</p>	<p>Productions orales (LSF et français) et productions écrites</p>	<p>environ 75 apprenants</p>	<p>CERCA, Université de Lille</p>	<p>Transcriptions et à venir avec Elan</p> <p>Accessibilité sur demande</p>	<p>Stéphanie GOBET</p> <p>stephanie.gobet.jacob@univ-poitiers.fr</p>