



UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

Corpus Valjouffrey



UMR 6057 CNRS
PAROLE ET
LANGAGE
<http://lpl-aix.fr>

Médéric Gasquet-Cyrus
& Bernard Bel
Laboratoire Parole et Langage (LPL)
5 avenue Pasteur
BP 80975
13604 Aix-en-Provence

Septembre 2009

Appel à projets TUL-ILF-DGLFLF :

***Corpus en français et en langues de France (constitution,
formatage, valorisation, base de données)***

Axe B — Constitution de nouveaux corpus en français et en langues de France





UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

Présentation de l'organisme postulant

Organisme

Laboratoire Parole et Langage (LPL)
UMR 6057 CNRS - Université de Provence
5 avenue Pasteur
BP 80975
13604 Aix-en-Provence
<http://lpl-aix.fr>

Directeur : **Philippe Blache** philippe.blache@lpl-aix.fr

Responsables scientifiques du projet

Médéric Gasquet-Cyrus
mederic.gasquet-cyrus@lpl-aix.fr
Laboratoire Parole et Langage

Bernard Bel
bernard.bel@lpl-aix.fr
Laboratoire Parole et Langage

Personnes impliquées dans le projet

Conseil scientifique :

Roxane Bertrand roxane.bertrand@lpl-aix.fr

Cristel Portes cristel.portes@lpl-aix.fr
Équipe *Interactions langagières et communicatives*
Laboratoire Parole et Langage

Daniel Hirst daniel.hirst@lpl-aix.fr
Équipe *Structures linguistiques : annotations, modèles et interfaces*
Laboratoire Parole et Langage

Personnes ressources pour le terrain :

Etienne Champollion hethien@hotmail.com
Musicien et compositeur, originaire de la vallée de Valjouffrey

Gérard Jacquemin
Président de l'association « *Mémoire Battante* » consacrée à la valorisation du patrimoine culturel de Valjouffrey

Assistants :

Dessinatrice : **Aurore Champollion**
Étudiante en arts plastiques, originaire de la vallée de Valjouffrey
Université de Provence

Vidéaste : **Andréine Bel**
Secrétaire de plateau, assistante pour traduction/transcription : **Audrey Thomas**
Postproduction audio/vidéo : **Bernard Bel**





UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

Corpus Valjouffrey

Présentation du projet

Axe : B — Constitution de nouveaux corpus en français et en langues de France

Objectifs

- 1) Documentation et description d'une **langue en voie d'extinction** : le patois/parler/dialecte de **Valjouffrey** ;
- 2) Construction d'un **corpus audio/vidéo multi-locuteurs de parole spontanée** répondant aux besoins spécifiques de travaux sur la prosodie, la mimo-gestualité, les interactions langagières et communicatives, ainsi que la comparaison des langues ;
- 3) Enquête **ethnolinguistique, culturelle et historique** sur la région de **Valjouffrey**.

Terrain

La vallée de Valjouffrey dans le parc national des Ecrins, <http://www.valjouffrey.fr>

Intérêt particulier

La question des frontières linguistiques entre l'aire d'oc alpine et l'aire franco-provençale alimente de manière récurrente les travaux des dialectologues (Tuailon 1964, 1967 ; Bouvier 1979) et des linguistes (Blanchet et Pons 2003 ; Sibille 2003). Aussi précises soient-elles, les isoglosses laissent apparaître une zone floue dans laquelle l'hybridation des parlers, leurs histoires respectives et les modalités de leurs contacts sont complexes. C'est dans une telle zone que se situe Valjouffrey, terrain principal de l'enquête.

Cette vallée, située à la lisière de la zone d'oc, est particulièrement isolée (une seule voie d'accès) et se termine en cul-de-sac sur le massif des Ecrins. Les quelques témoignages issus du terrain font état d'une singularité du parler (« patois ») local par rapport à la variété proche, le valbonnetin (ce qui pourra faire l'objet de comparaisons).

Ce parler est cependant en voie d'extinction. Déjà deux locuteurs connus sont décédés l'an dernier, et il ne reste qu'une poignée d'entre eux, (4 ? 6 au maximum ?), très âgés, qu'il conviendra d'aller rencontrer afin de collecter de précieuses données pour documenter et décrire ce parler. En effet, aucun corpus de cette variété n'a été constitué à ce jour, alors que des travaux ont été menés sur des parlers proches (Duc, par exemple). L'enquête s'appuiera sur un réseau de contacts locaux, ainsi que sur plusieurs représentants du milieu associatif qui ont déjà commencé à documenter le patrimoine culturel. Il est possible que des entretiens réalisés dans ce contexte soient intégrés au corpus.

Le projet prévoit une enquête de terrain (en plusieurs sessions), précédée de la lecture de la documentation, ancienne (Champollion 1809) ou récente (Duc 1987, 1991 ; Pons 2003) susceptible d'orienter les recherches. Les locuteurs seront enregistrés au cours de séances d'entretiens qui privilégieront les interactions spontanées, afin de permettre la construction d'un corpus de parole audio/vidéo multi-locuteurs répondant aux besoins spécifiques de travaux sur la prosodie, la mimo-gestualité, les interactions langagières et communicatives, ainsi que la comparaison des langues.

Une qualité importante du corpus proposé réside dans les techniques d'enregistrement envisagées. En plus de 2 caméscopes (pour filmer les interactions) nous prévoyons d'utiliser 4 micros serre-tête et un enregistreur numérique multicanaux (jusqu'à 8 canaux simultanés). Ceci nous permettra de mettre en scène des interactions entre plusieurs locuteurs (plus un canal pour l'intervenant) dans des situations où les fréquents recouvrements entre tours de parole pourront être analysés.

En raison de la particularité du terrain, cette enquête linguistique sera aussi d'ordre ethnographique. La richesse et la diversité des ressources patrimoniales peuvent susciter des travaux en histoire, géographie, ethnolinguistique, etc., d'où la nécessité de rassembler tous les documents numérisés



UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

(son, vidéo, images) dans un centre de ressources (CRDO) dédié à l'accès mutualisé, à l'archivage pérenne et au suivi des utilisations.

Valorisation

Ce corpus pourra faire l'objet d'enrichissements au moyen de systèmes d'annotation multidimensionnels et intégratifs (cf. projet OTIM au LPL, <http://lpl-aix.fr/projet/170>).

Compétences mises en œuvre

Médéric Gasquet-Cyrus est Maître de Conférences au département de Linguistique et Phonétique Générales de l'Université de Provence et membre de l'équipe de recherche *Dynamique des systèmes linguistiques, créoles et variétés émergentes* au LPL. Ses compétences recouvrent notamment la sociolinguistique, les accents, le français régional de Provence, les représentations et discours épilinguistiques, l'humour verbal et la lexicographie.

Bernard Bel est ingénieur de recherche au LPL, chargé du système d'information du laboratoire et responsable de la composante *Ressources et outils d'analyse* du Centre de ressources pour la description de l'oral (CRDO, <http://crdo.fr>). Ses compétences comprennent une dizaine d'années de terrain en milieu rural pour la collection de corpus de parole et de chant, l'édition de la série *Communication Processes* (3 volumes) chez Sage, et le développement d'un environnement informatique pour la mutualisation et l'archivage pérenne des ressources orales (projet pilote du TGE-Adonis, en collaboration avec le CINES et le CC-IN2P3, voir <http://crdo.fr/wiki/Developpement>). Il est aussi secrétaire du *Speech Prosody Special Interest Group* (SProSIG) de l'ISCA. Publications : <http://lpl-aix.fr/publi/2>

Formats de données

Audio : Enregistrement et archivage 44.1 Khz PCM 16 bits sur 2 à 8 canaux. Format de diffusion : WAV, 44.1 Khz PCM 16 bits 2 canaux.

Vidéo : Enregistrement sur cassette mini DV. Formats d'archivage et de diffusion : Matroska encodages H.264/Flac (ouvert au sous-titrage et pas de dégradation du signal audio) et MPEG4 encodages AVC/AAC.

Transcriptions, traductions, annotations : texte UTF8, fichiers Praat ou/et XML.

Échéancier

- Octobre 2009 : repérages sur le terrain
- Novembre 2009 - mars 2010 : missions sur le terrain
- Avril - juin 2010 : transcriptions, traductions, alignement, postproduction audio et vidéo

Bibliographie

- Bec, P. (1978 [1963]), *La langue occitane*, Paris, PUF, Que sais-je ?
- Bertrand, R.; Blache, P.; Espesser, R. et al. (2008), « Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle », *Traitement Automatique des Langues*, vol. 49, no. 3., p. 1-30. <http://hal.archives-ouvertes.fr/hal-00349893>
- Blanchet, P. (1992), *Le provençal. Essai de description sociolinguistique et différentielle*, Série Pédagogique de l'Institut de Linguistique de Louvain, n° 15, Louvain-la-Neuve, Peeters.
- Blanchet, P., Pons P. (dirs) (2003), *Les langues régionales ou minoritaires de l'Arc alpin*, actes du colloque international Gap-Charance, Unioun Prouvençalo.
- Blanchet, P. (2003), « Les langues et cultures « minoritaires » de l'Arc alpin : présentation générale », in Blanchet, P., Pons P. (dirs) (2003), *Les langues régionales ou minoritaires de l'Arc alpin*, actes du colloque international Gap-Charance, Unioun Prouvençalo, p. 6-11.
- Bouvier, J.-C., Martel C. (1975-1979), *Atlas linguistique et ethnographique de la Provence*, Paris, CNRS Editions.
- Bouvier, J.-C. (1979), « L'occitan en Provence : le dialecte provençal, ses limites et ses variétés », in *Revue de linguistique romane*, t. 43, p. 46-62.
- Champollion, Jean-Jacques (1809), *Nouvelles recherches sur les patois ou idiomes vulgaires de la France et en particulier sur ceux de l'Isère*, Paris, Librairie Goujon.
- Duc, A. (1990), *Les régionalismes du canton de La Mure*, CNRS INaLF, coll. Matériaux pour l'étude des régionalismes du français, n° 5, Klincksieck.





UMR 6057 CNRS

Laboratoire Parole & Langage

Université de Provence Aix-en-Provence, France

- Duc, J. (1987), « Les patois matheysins », in B. De la Fayolle (dir.), *Le pays de la Mure, cœur du Dauphiné*, Grenoble, Didier et Richard.
- Duc, J. (1991), *Les Patois du pays de La Mure*, Grenoble, Centre Alpin et Rhodanien d'Ethnologie.
- Durauffour, A. (1969), *Glossaire des patois franco-provençaux*, Paris, CNRS.
- Gardette, P. (1967-1976), *Atlas linguistique et ethnographique du Lyonnais*, Paris, CNRS Editions.
- Germi, C. (1996), *Mots du Champsaur. Hautes-Alpes*, ELLUG, Grenoble, Université Stendhal.
- Martin, J.-B., Tuaillon, G. (1971-1978), *Atlas linguistique et ethnographique du Jura et des Alpes du Nord*, Paris, CNRS Editions.
- Pons, P. (1982), *Le provençal haut-alpin*, Société d'Etudes des Hautes-Alpes, Gap.
- Pons, P. (2003), « Le patrimoine linguistique haut-alpin », Blanchet, P., Pons P. (dirs) (2003), *Les langues régionales ou minoritaires de l'Arc alpin*, actes du colloque international Gap-Charance, Unioun Prouvençalo, p. 19-22.
- Sibille, J. (2003), « Le francoprovençal », in B. Cerquiglini (dir.), *Les langues de France*, Paris, PUF, p. 117-124
- Sibille, J. (2003), « L'occitan ou langue d'oc », in B. Cerquiglini (dir.), *Les langues de France*, Paris, PUF, p. 173-190.
- Tuaillon, G. (1964), « Limite nord du provençal à l'est du Rhône », in *Revue de linguistique romane*, XXVIII, p. 127-142.
- Tuaillon, G. (1967), « Principes pour distinguer français et francoprovençal », in *Revue de linguistique romane*, XXXI, 292-296.
- Tuaillon, G. (1968), « A propos du patois matheysin », in R. Raymond, *Pierre-Châtel, hier et aujourd'hui*, Gap, Louis-Jean.