



UMR 6057 CNRS

Laboratoire Parole &amp; Langage

Université de Provence Aix-en-Provence, France

## SLDR - Speech and Language Data Repository - [www.sldr.org](http://www.sldr.org)

Le LPL assure la responsabilité scientifique et technique du SLDR (*Banque de données parole et langage*), un service chargé de la mutualisation et de l'archivage de données orales. Le SLDR est issu de l'initiative CRDO-Aix, en réponse à un appel à projets de la Direction de l'Information Scientifique et du Département scientifique Homme et Société du CNRS (France) en 2006. Au terme d'un projet pilote coordonné par le TGE-Adonis, le SLDR fonctionne aujourd'hui comme un dispositif interopérable et ouvert s'appuyant, pour l'archivage pérenne, sur le *Centre informatique de l'enseignement supérieur* (CINES) et, pour le partage des données, sur le *Centre de calcul de l'Institut national de physique nucléaire et de physique des particules* (CC-IN2P3).

L'objectif du SLDR est de regrouper les ressources associées à la communication orale. Une telle centralisation permet tout d'abord d'éviter que les laboratoires travaillant dans ce domaine ne soient conduits à recréer sans cesse des données. La mutualisation des ressources en sciences du langage répond par ailleurs au besoin stratégique de rapprocher les connaissances aujourd'hui dispersées dans des domaines variés comme la linguistique descriptive, formelle et computationnelle, la sociolinguistique, la littérature, la traductologie, les neurosciences, la psycholinguistique etc.

Les données partagées avec la communauté scientifique (et le public en général) comprennent des corpus de parole (spontanée ou lue), leurs enrichissements (transcriptions et annotations), des lexiques et autres bases de connaissances, ainsi que des outils adaptés au traitement des corpus. Les corpus proprement dits peuvent inclure des enregistrements audio/vidéo mais aussi des mesures de l'activité physiologique : endoscopie, potentiels évoqués, palatographie, etc.

The screenshot shows the 'Speech & Language Data Repository' interface. It includes logos for tge, DONIS, CLARIN, and Open archives. Below the header, there are navigation links for [Sign up] and [Login], and language options: / 中文 / English / español / français / . The main section is titled 'Primary data' and contains a message: 'Flash Player plugin is required to read previews. If the preview cell is empty or the player is not active, you may download Adobe Flash Player.' Below this, there is a dropdown menu for 'Display corpora in' set to 'Arabic, Standard (العربية القياسية) [arb] (1)'. A table lists one item with the following details:

Id	Name	Description	Downloading	Preview
[ARK] sldr000745 => wiki	Standard ARABIC - statements (Amina CHENTIR)	language_description/narrative Laboratoire parole et langage (LPL, Aix-en-Provence FR) Affirmative sentences in standard Arabic recorded in a sound-proof studio by 5 North-African speakers (2 men and 3 women). These sentences contain all the phonemes of standard Arabic. جمل مفيدة باللغة العربية القياسية مسجلة داخل غرفة معزولة من طرف خمسة المتكلمين من شمال إفريقيا (رجال و نساء) - تحتوي على جميع فونيمات العربية القياسية (language_documentation)	You are not logged in. Downloaded 3 time(s) (see members)	Preview icons

L'architecture du SLDR est basée sur un modèle ouvert de système d'information pour l'archivage (OAIS) et elle assure un accès pérenne aux objets par le biais d'un identifiant unique. Le dispositif mis en place au SLDR permet de reconstituer intégralement un objet archivé à partir de la version distribuée (au CC-IN2P3) et donc de libérer l'espace correspondant sur le serveur. En réponse aux besoins les plus diversifiés, l'équipe a privilégié le développement d'un système générique qui minimise les contraintes techniques sur la structure et la volumétrie des objets archivés : segmentation automatique des objets de grande taille, pas de limite sur les noms de fichiers, prise en compte de l'évolution des objets (par la diffusion de *données courantes*) et une gestion souple et précise des droits d'accès à tout document dans les phases successives d'un projet.

Le téléchargement d'un objet peut être réservé à certaines catégories d'utilisateurs. Les restrictions d'accès aux documents déposés en archive publique s'appuient sur les dérogations au principe de libre communicabilité prévues par le Code du patrimoine (art. L213-2, loi du 15 juillet 2008).

Le serveur du SLDR conserve la trace des transactions sur certaines ressources et les utilisateurs s'engagent, par le biais d'une licence non-commerciale, à les mentionner sur toute production scientifique basée sur leur partage. Ce dispositif permet de faire émerger des communautés de producteurs et d'utilisateurs favorisant l'agrégation de projets au-delà des frontières géographiques ou institutionnelles. Le site SLDR est quadrilingue : anglais, français, espagnol, chinois, et s'appuie sur l'approche Web 2.0 pour les communautés d'utilisateurs, avec des liens vers les réseaux sociaux.

Contact : <http://sldr.org/contact>

