Opération LR-1

Axe 6 : ressources linguistiques

Vendredi 1er juin 2012

Sommaire

1. Rappel des objectifs de l'opération LR-1 (C. Plancq)

1. Présentation du travail d'inventaire (S. El Ayari)
   - Grille de description des ressources
   - Présentation de la plateforme

   → Discussion sur les travaux à venir

1. Appels à intervention de l'ingénieur (S. El Ayari et C. Plancq)
   - Critères de sélection et processus de décision
   - Exemples de demandes d'intervention

Opération LR-1

LR-1 : A joint approach to language ressource development

- Opération transversale : concerne toutes les équipes, tous les axes
- Objectifs principaux : visibilité et accessibilité des ressources, interopérabilité, utilisation des standards
- Moyens : ingénieure de recherche à plein temps (CDD 3 ans)
- Actions : développement logiciel, support/conseil, formation

Inventaire des ressources

- Visites des laboratoires partenaires du projet
- Elaboration d'une grille de description compatible avec les standards de métadonnées
- Développement d'une plateforme web pour l'enregistrement et l'accès aux ressources

http://ressources.labex-efl.org

Problématique générale

Types de ressources :
- corpus
- dictionnaires
- lexiques
- outils

Modalités :
- écrit
- oral

Trouver un équilibre entre :
- la génération de métadonnées complètes
- des formulaires pas trop longs à saisir (2min/ressource)
- une utilité réelle pour les participants

Métadonnées

Standards

- Dublin Core (DC) : 15 descripteurs
- TEI Header
- OLAC : extension du Dublin Core pour les ressources linguistiques
- CLARIN (CMDI) : modules de métadonnées compatibles avec DC, TEI Header, OLAC

Catalogues

- France : Isidore (TGE Adonis)
- Europe : LRE Map, CLARIN
- International : OLAC

Initiatives (France)

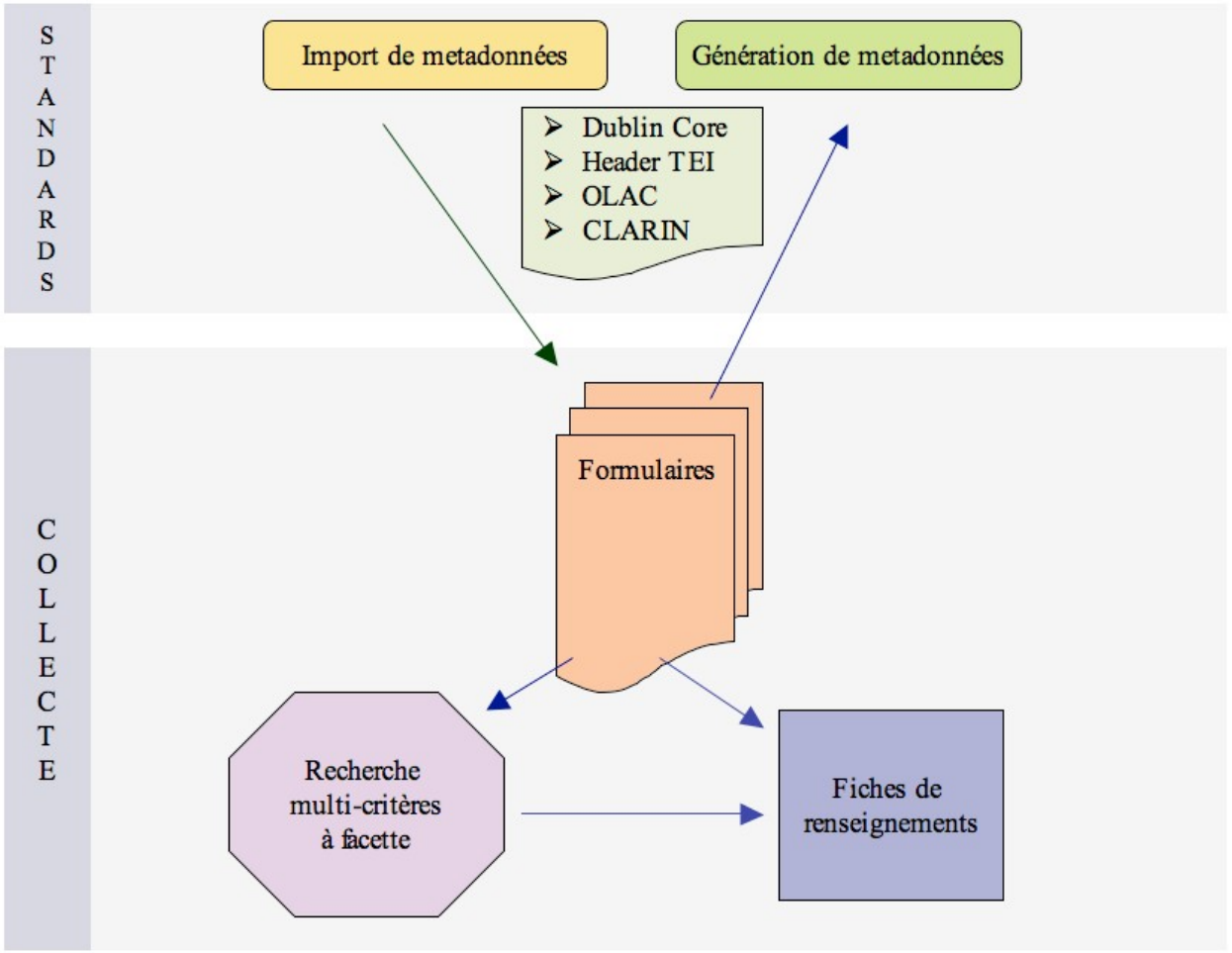
- TGIR-Corpus (IRCOM, corpus écrits)
- ORTOLANG (ATILF, LPL, MoDyCo, LLL, LORIA, INIST)
- Centre de ressources numériques : CRDO, SLDR, CNRTL

Grille de description

Description générale	Description particulière
• Nom • Description • URL • Accès • Metadonnées • Licence • Documentation • Publication • Projet associé • Objectifs scientifiques • Lien avec autre ressource	***Données*** • Modalité • Type de données • Source des données • Format • Taille • Langue • Annotations • Codage des caractères • Etat d'avancement
Référents •Producteur •Référent interne au Labex	***Outils*** •Type d'outil •Environnement •Langage •Interface •Format d'entrée •Format de sortie

Inventaire des ressources

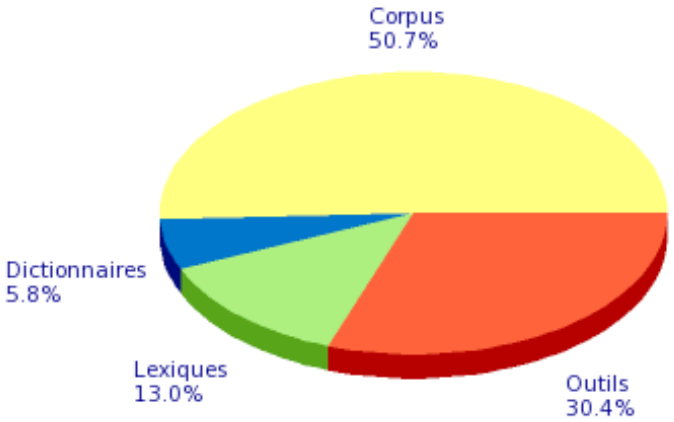
Schéma des fonctionnalités



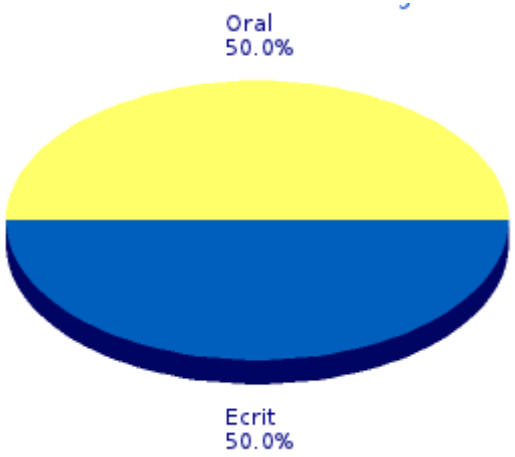
La plateforme en ligne

- Enregistrement des ressources >>
- Module de recherche à facettes >>
- Fiches descriptives avec URL stables >>
- Génération de metadonnées aux formats standards : *Dublin Core*, *TEI Header*, *OLAC*, *CLARIN* >>
- Boîte à outils >>
  conversions de metadonnées et conversions de formats de fichiers
- Liens vers des ressources externes au LabEx >>

Répartition des ressources



Par type de ressources



Par modalité

Synthèse des données

- 72 ressources, dont :
  - 21 outils
  - 51 données

- Langues représentées
  - Français, Anglais, Latin, Tamoul, Allemand, Perse, Bulgare, Croate, Tchèque, Estonien, Persan, Ouldémé, Ihanzu, Langi, Mahorais, Mankon, Mbugwe, Nagazidja, Nyilamba, Yucuna, Bahing, Hayu, Koyi, Lazé, Limbu, Mizo, Na, Naxi, Nepali, Prinmi, Tamang, Thulung, Espagnol, Galicien, Kurmanji, Polonais, Slovène, Japonais, Danois, Grec, Finnois, Macédonien, etc.

- Objectifs renseignés
  - Apprentissage d'analyseurs statistiques en dépendances, analyse syntaxique, TAL, linguistique expérimentale, observation des SP en tête de phrase, aide à la recherche en linguistique, enquêtes linguistiques, traduction automatique, machine learning.

Discussion

Qui se charge de diffuser les métadonnées ?

Appels à intervention

- Critères de sélection
  - Ressource cataloguée dans l'inventaire
  - Opération financée ou soutenue par le labex
  - Plusieurs laboratoires concernés
  - Durée de l'intervention précisée

- Processus de sélection
  - Réception des demandes au fil de l'eau
  - Sélection par un comité *ad hoc* de l'opération
  - Validation par comité scientifique restreint tous les trimestres ?