

## Partage d'expériences

### Le Centre national de ressources textuelles et lexicales

Dans le cadre de la recherche en SHS, la valorisation et l'exploitation scientifiques de ressources sont particulièrement importantes et stratégiques et un des aspects essentiels aujourd'hui est leur informatisation et leur disponibilité sur la toile sous une forme facilement accessible et exploitable par l'ensemble de la communauté scientifique.

C'est sur cette base qu'en 2006 le CNRS a impulsé la création de centres de ressources, dont le CNRTL pour les ressources textuelles, lexicales et dictionnaires. Adossé à l'ATILF (Analyse et Traitement Informatique de la Langue Française (CNRS/Nancy Université), son objectif initial était de réunir, au sein d'un portail unique, le maximum de ressources informatisées et d'outils de traitement pour l'étude, la connaissance et la diffusion de la langue française.

#### Les ressources et outils proposés par le CNRTL

Les ressources proposées par le CNRTL se structurent aujourd'hui autour de cinq pôles de compétence :

1. Un **portail lexical**, base de connaissances lexicales du français qui a pour vocation de valoriser et de partager, en priorité avec la communauté scientifique, un ensemble de données issues des travaux de recherche sur le lexique français menés à l'ATILF ou au sein de partenaires du CNRTL (Académie française, ARTFL Chicago, CLEE et IRIT Toulouse, CRISCO Caen, Laboratoire informatique de Tours, etc.). Projet évolutif, cette base de connaissances permet d'obtenir à partir d'une forme lexicale, par exemple *sussiez* du verbe *savoir* (cf. Figure 1), des informations

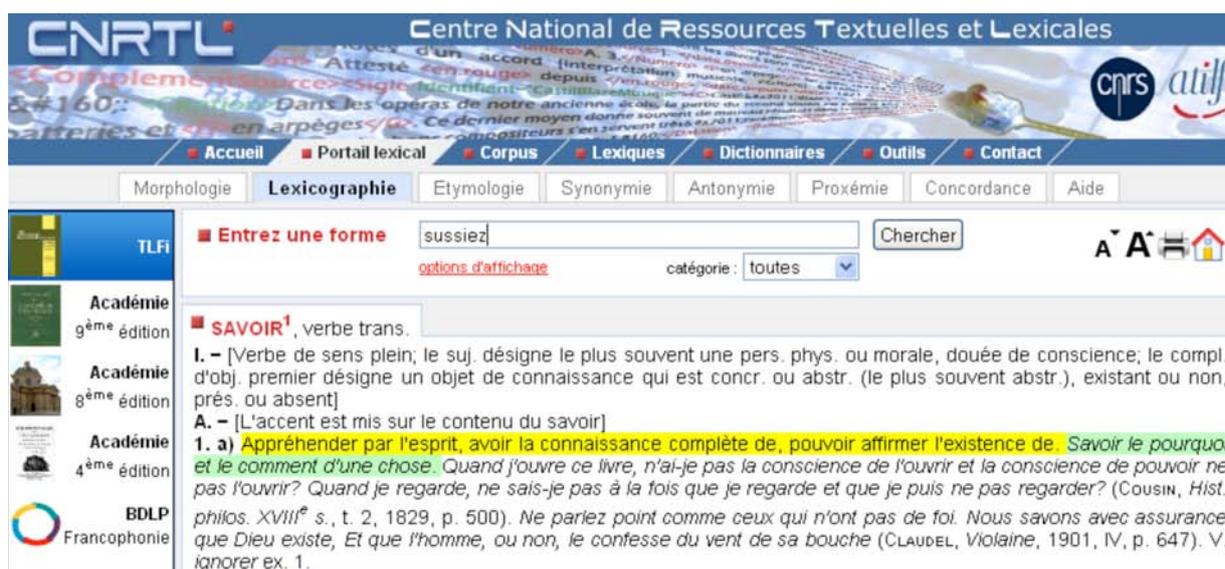


Figure 1 – Obtention d'informations lexicographiques à partir de la forme *sussiez*

Pour faciliter la mutualisation de telles ressources, nous avons choisi de doter ce centre d'une visibilité spécifique, au travers de l'acquisition des droits du domaine [cnrtl.fr](http://cnrtl.fr), offrant ainsi :

- ▶ une boîte contact : [contact@cnrtl.fr](mailto:contact@cnrtl.fr), qui aujourd'hui recueille en moyenne une quinzaine de messages par semaine ;
- ▶ un site web : [www.cnrtl.fr](http://www.cnrtl.fr), vecteur principal pour présenter le CNRTL, diffuser les ressources et permettre aux utilisateurs de proposer ou déposer des ressources auprès du CNRTL.

Le CNRTL est soutenu par le TGE Adonis et, au cours des dernières années, pour assurer sa mise en place, il a été contractualisé dans le cadre du CPER Lorrain et a bénéficié d'un soutien du FEDER lorrain (Fonds européen de développement économique des régions).

morphologiques, lexicographiques et étymologiques, des informations de synonymie, d'antonymie et de proximité sémantique (proxémie, cf. Figure 2) et une concordance utilisant le corpus de textes libres de droits de la base [FRANTEXT](http://frantext.fr). Il permet d'exporter les résultats du concordancier au format XML/TEI (Text Encoding Initiative) et, à notre connaissance, c'est le seul site permettant à un utilisateur d'importer dans un format normalisé un concordancier français d'une telle importance. Produit phare du CNRTL, ce portail sert en moyenne chaque jour plus de 350 000 requêtes venant du monde entier (cf. Figure 3). Il est intégré sous forme d'extension aux navigateurs Firefox et Chrome.

2. Un ensemble de **corpus** librement accessibles par téléchargement, avec, entre autres :





Figure 4 – Exemple d'informations fournies par Prolex

construites par dérivation mais aussi celles formées par composition savante ou néoclassique.

Ainsi à partir de *importable*, Flemm fournit l'analyse suivante :

a.1. *importable/ADJ*==> [ in [ [ *porter VERBE*] *able ADJ*] *ADJ*] (*importable/ADJ*, *portable/ADJ*, *porter/VERBE*)

a.2. "Non portable"

b.1. [ [ *importer VERBE*] *able ADJ*] (*importable/ADJ*, *importer/VERBE*)

b.2. "(lequel - Que l') on peut importer"

► **FastKwic**, un outil permettant d'indexer un texte, français ou anglais, et de produire un concordancier à partir du résultat de cette indexation.

## Une insertion nationale et internationale forte

Le CNRTL est donc engagé dans un processus de mutualisation de ressources issues des différents laboratoires, à travers la validation, l'harmonisation, la diffusion et le partage de ces ressources, qu'il s'agisse de données textuelles et lexicales informatisées ou d'outils permettant un accès intelligent à leur contenu.

Outre son implication au niveau national au sein du **TGE ADONIS** dont il est opérateur pour son champ de compétences (corpus écrits, lexiques et dictionnaires), le CNRTL, dont l'expertise scientifique est largement reconnue, est fortement impliqué au niveau européen ou international à travers :

► sa participation au projet européen **CLARIN** d'infrastructure européenne partagée pour les SHS (Common Language Resources and Technology Infrastructure) s'appuyant sur des centres régionaux « certifiés » dans leurs domaines respectifs. La mise en place d'un accord de partenariat entre l'ATILF et l'INIST (février 2009) nous a permis de positionner le CNRTL comme un des principaux centres de cette future infrastructure de recherche.

► sa fonction de centre européen support de la TEI : issue d'un partenariat entre l'ATILF, l'INIST et le LORIA, cette fonction de centre européen support de la TEI, assurée jusqu'alors par le CNRTL, est reprise pour l'avenir directement par le TGE ADONIS.

► Des collaborations directes avec des centres partenaires, en Grande-Bretagne (Université d'Oxford), en Allemagne (Centres de compétence de Trèves et de Würzburg, Université de Sarrebruck, MPI Berlin) et aux Pays Bas (Université de Nimègue).

## Et l'avenir...

Nous souhaitons à l'avenir, au sein du TGE ADONIS, accentuer notre mission d'accompagnement des laboratoires souhaitant mutual-

ser avec d'autres leurs ressources en offrant, entre autres :

► un site WEB permettant aux utilisateurs de pouvoir naviguer dans les collections disponibles sur la plate-forme CNRTL et proposant une interface de recherche conviviale dans les métadonnées au format Dublin Core ou CMDI (Clarín Metadata Infrastructure) ainsi que la possibilité de sélectionner dans l'interface graphique des ressources pour les ajouter dans un « panier » de fichiers à télécharger. Découplée des plates-formes de stockage, cette interface sera régulièrement synchronisée avec elles en moissonnant toutes les métadonnées en OAI-PMH disponibles ;

► l'export OAI-PMH des métadonnées permettant la visibilité des ressources tant sur des plateformes de recherche, par exemple **ISIDORE** qu'au sein d'agrégateurs ou d'infrastructures de recherche, par exemple **CLARIN**.

► des services WEB qui, étant donné un nom de déposant (laboratoire, UMR, équipe d'accueil), fournissent respectivement une page HTML et un fichier XML récapitulant les dépôts archivés de ce déposant ;

► la pérennisation de ces ressources via le service d'archivage à long terme offert par le TGE ADONIS avec le CINES ;

► un outil d'aide à la fabrication des métadonnées CLARIN, complétées avec la notion de déposant, et offrant une fonctionnalité permettant d'engendrer automatiquement de telles métadonnées à partir de métadonnées au format TEI ou de fabriquer des métadonnées minimales ne rendant compte que des notions de déposants. Les données ainsi produites, normalisées et stockées pourront être utilisées par des communautés élargies et répondront aux diverses exigences des standards internationaux.

**Jean-Marie Pierrel**

Responsable du CNRTL

ATILF UMR7118 Nancy-Université/CNRS

### contact&info

► Jean-Marie Pierrel

Responsable du CNRTL

Jean-Marie.Pierrel@atilf.fr

► Pour en savoir plus

[www.cnrtl.fr](http://www.cnrtl.fr)