

Présentation d'Unitex

Pascal Amsili

ILPGA, Sorbonne Nouvelle

Mars 2021

Unitex/Gramlab : ensemble de logiciels

- permettant de traiter des textes en langues naturelles
- en utilisant des ressources linguistiques

Ressources :

- dictionnaires électroniques
- grammaires
- tables de lexique-grammaire

Origine : travaux de Maurice Gross au LADL (Paris Diderot, Marne-La-Vallée)

Extension à d'autres langues : réseau de laboratoires RELEX

dictionnaires électroniques

... décrivent les mots simples et composés d'une langue en leur associant un lemme ainsi qu'une série de codes grammaticaux, sémantiques et flexionnels.

- différence majeure avec outils usuels de recherche (concordanciers) : informations linguistiques disponibles
- formalisme DELA
- français, anglais, grec, italien, espagnol, allemand, thaï, coréen, polonais, norvégien, portugais, etc...

« grammaires »

... représentations de phénomènes linguistiques par réseaux de transitions récursifs (RTN), un formalisme proche de celui des automates à états finis.

- pertinence du formalisme aux niveaux morphologique, syntaxique, phonologique, etc.
- le formalisme utilisé est (un peu) plus puissant que les automates
- grammaires représentées sous forme de graphes que l'utilisateur peut aisément créer et mettre à jour

tables de lexiques-grammaires

...matrices décrivant les propriétés de certains mots.

élaborées pour tous les verbes simples du français dont elles décrivent les propriétés syntaxiques.

L'expérience ayant montré que chaque mot a un comportement quasi unique, ces tables permettent de donner la grammaire de chaque élément de lexique, d'où le nom de lexique-grammaire.

Unitex permet de construire des grammaires à partir de telles tables.

Environnement Unitex

Unitex est un moteur permettant d'exploiter ces ressources linguistiques.

→ essentiellement pour faire des recherches/extractions

Caractéristiques techniques : portabilité, modularité, possibilité de gérer des langues possédant des systèmes d'écritures particuliers, ouverture (logiciel libre)

Caractéristiques linguistiques : précision, exhaustivité et prise en compte des phénomènes de figement, notamment en ce qui concerne le recensement des mots composés.

Installation

The screenshot shows the website unitexgramlab.org/fr with a navigation menu including ACCUEIL, A PROPOS, SAISIES D'ÉCRAN, TÉLÉCHARGEMENTS (highlighted), and PLUS. The main content area is titled "Télécharger Unitex/GramLab" and features two version options: 3.2 and 3.3alpha. Below this, there are four download categories: Windows 32 bits, Linux 32 bits, OS X 10.7+, and Autres 32 bits. Each category includes a brief description of compatibility and a "Télécharger" button. At the bottom, there is a footer with links for version v3.2.130, a "1 année plus tôt" link, license information (LGPL+LGPLLR), and links for Manuel, Aide, Bugs, and Autres versions.

unitexgramlab.org/fr

Rechercher avec Goo...

ACCUEIL A PROPOS SAISIES D'ÉCRAN **TÉLÉCHARGEMENTS** PLUS

France

Télécharger Unitex/GramLab

3.2 3.3alpha

Windows 32 bits

Unitex/GramLab 3.2 pour Windows. Compatible Windows XP, Vista, 7, 8 and 10. Les installateurs sont disponibles pour les systèmes 64 bits et 32 bits

Linux 32 bits

Unitex/GramLab 3.2 pour Linux. Les installateurs sont disponibles pour les systèmes 64 bits et 32 bits

OS X 10.7+

Unitex/GramLab 3.2 pour Mac OS X. Compatible OS X 10.7 (Lion), 10.8 (Mountain Lion), 10.9 (Mavericks), 10.10 (Yosemite), 10.11 (El Capitan) et 10.12 (Sierra)

Autres 32 bits

Unitex/GramLab 3.2 pour autres systèmes comprend ressources nécessaires pour compiler cette version

Télécharger Télécharger Télécharger Télécharger

v3.2.130 | 1 année plus tôt | LGPL+LGPLLR | Manuel | Aide | Bugs | Autres versions