



HAL
open science

Mise en oeuvre d'un outil de lecture active de textes latins

Mathieu Mangeot

► **To cite this version:**

Mathieu Mangeot. Mise en oeuvre d'un outil de lecture active de textes latins. [Rapport de recherche] Laboratoire Litt&Arts. 2019. hal-02165651

HAL Id: hal-02165651

<https://hal.science/hal-02165651>

Submitted on 26 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mise en œuvre d'un outil de lecture active de textes latins

Mathieu Mangeot
Laboratoire LIG, équipe GETALP
Bâtiment IMAG CS 40700
38058 GRENOBLE CEDEX 9, FRANCE
mathieu.mangeot@imag.fr

Introduction

Le principe de la lecture active est d'enrichir un texte avec des informations permettant de mieux en comprendre le sens. Le but est d'augmenter les compétences du lecteur dans une langue donnée contrairement à la traduction automatique qui vise à augmenter ses performances. Certaines informations rajoutées sont liées à la nature de chaque mot : prononciation, catégorie grammaticale, etc. D'autres peuvent être une traduction dans une langue mieux connue du lecteur. Ces dernières ne sont affichées qu'au survol de la souris pour ne pas gêner la lecture du texte et éviter que le lecteur ne se focalise uniquement sur les traductions.

Dans le contexte du projet Lateo, il nous a semblé pertinent de proposer un outil de lecture active de textes en latin à destination des apprenants francophones qui puisse afficher des informations spécifiques aux mots (par exemples les déclinaisons) ainsi que des traductions françaises de certains mots.

1. Description de l'outil

L'outil de lecture active (Abdellaoui et al., 2018) procède de la manière suivante :

1. Le texte reçu en entrée est envoyé à un analyseur morphologique afin d'obtenir le lemme et les informations grammaticales de chaque mot.
2. À l'aide des lemmes reçus de l'analyseur, l'outil consulte ensuite un dictionnaire monolingue ou bilingue afin d'obtenir des informations lexicales sur chaque lemme. Il peut s'agir d'une définition ou d'une traduction.
3. L'outil affiche le texte reçu en entrée en y ajoutant les informations supplémentaires soit au dessus ou en dessous du texte, soit en affichant les mots avec des couleurs différentes, soit en ajoutant un code qui affichera une fenêtre surgissante lors du survol du mot par la souris.

Pour le latin, nous avons utilisé les ressources suivantes :

L'analyseur morphologique du latin est le Tree Tagger (Schmid, 1995), disponible en ligne à cette adresse :

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

La ressource lexicale utilisée est une version électronique du dictionnaire Gaffiot qui a été convertie au format XML puis importée sur la plateforme Jibiki de gestion en ligne de ressources lexicales.

Les deux figures suivantes montrent un texte latin entré dans l'outil de lecture active et son résultat analysé avec le survol de la souris sur le mot « praesertim »

Cornelius Tacitus (nescimus utrum ei Gaius an Publius fuerit praenomen) fuit senator et rerum scriptor Romanus. Plenam scientiam de rebus Romanis primo saeculo gestis a Tacito narratam habemus. Sed praesertim innotuit nonnullis libris quibus historiam Britanniae et Germaniae temporum antiquorum narravit.

Figure 1 : Texte latin en entrée de l'outil de lecture active

Note : passez la souris sur un mot pour afficher ses traductions.
Code couleurs : nominatif, vocatif, accusatif, génitif, datif, ablatif

Cornelius Tacitus (nescimus utrum ei Gaius an Publius fuerit praenomen) fuit senator et rerum scriptor Romanus. Plenam scientiam de rebus Romanis primo saeculo gestis a Tacito narratam habemus. Sed praesertim innotuit nonnullis libris quibus historiam Britanniae et Germaniae temporum anti

adv. surtout (entre autres choses), notamment : [sert à détacher une condition, une circonstance, un petit stature ; surtout quand nous avons dans la nature même un maître et un guide || ind. surtout au moment où ; subj. surtout étant donné que, vu que surtout ; subj. quoique surtout, quand même notamment, cf. || subj. etc., surtout étant donné que ; subj. quand bien même || étant questeur, tu n'as pas obéi à un tribun de la plèbe auquel pourtant, fait notable, ton collègue obéissait || surtout puisque || surtout si.

Figure 2 : Affichage de lecture active sur le texte entré précédemment

L'outil est codé en PHP et accessible en ligne à l'adresse suivante :

<https://jibiki.fr/lecture/tacitus.php>

2. Préparation de la ressource lexicale

La ressource lexicale que nous avons trouvé sur le Web est au format LaTeX. Il s'agit d'une version du 2 mai 2016 publiée sous licence Creative Commons « Attribution-NonCommercial-NoDerivatives 4.0 International » . Les auteurs sont Gérard Gréco, Mark De Wilde, Bernard Maréchal et Katsuhiko Ôkubo. Le fichier contient 72 166 articles.

Voici un exemple d'article au format original :

```
\sansqu{sutrinum}
\entree{sūtrīnum,} \des{ī,} \gen{n.} \etyml{(sutrīnus),} boutique
ou métier de cordonnier :
\aut{Sen.} \oeuv{Ep.} \refch{90, 23}.
```

Figure 3 : Article "sutrinum" du dictionnaire Gaffiot au format LaTeX

Nous l'avons transformée au format XML à l'aide d'une série d'expressions rationnelles en suivant une méthodologie de conversion décrite dans (Enguehard & Mangeot, 2014). Voici le même article un exemple d'article au format XML :

```
<article id="lat.sutrinum.1">
<vedette>sutrinum</vedette>
<prononciation>sūtrīnum</prononciation><des>ī</des><genre>n.</genre><etymonlatin>sutrinus</etymonlatin> boutique ou métier de cordonnier :
<auteur>Sen.</auteur><œuvre>Ep.</œuvre><refch>90, 23</refch>.
</article>
```

Figure 4 : Article "sutrinum" au format XML après transformation

Les différentes parties d'information ont été balisées au format XML. Par contre, aucune structuration n'a encore été effectuée. Il manque en particulier le balisage de la traduction.

3. Mise à disposition sur la plateforme Jibiki

Jibiki (Zhang et al. 2016) est une plateforme de gestion de ressources lexicales en ligne¹. Elle permet d'importer de consulter et d'éditer toute ressource lexicale au format XML. Nous avons importé le Gaffiot sur l'instance Papillon de la plateforme.

Le dictionnaire est consultable en ligne à cette adresse :

<http://papillon.imag.fr/papillon/ConsultPage.po?lang=fra>

Il faut sélectionner « Gaffiot_lat_fra » comme volume puis entrer le mot à consulter.

La figure montre le résultat de la consultation du mot « sutrinum » dans le Gaffiot en ligne. La colonne de gauche permet de naviguer dans le dictionnaire par ordre alphabétique.

The screenshot shows the Jibiki platform interface. At the top, there is a navigation bar with links: Accueil, Projet, Données, Contacts, Aide, Connexion, and a language dropdown menu set to 'lang/语言/法/...'. Below the navigation bar, the main header reads 'La plateforme Jibiki'. A search bar is present with the text 'Chercher dans le volume Gaffiot_lat_fra' and a search input field containing 'Tacitus'. To the right of the search bar are two buttons: 'Recherche générique' and 'Recherche avancée'. Below the search bar, there is a list of words in a table on the left, with 'sutrinum' selected. The main content area displays the definition for 'sutrinum': 'sutrinum /sūtrīnum, /ī, [n.] (sutrinus), boutique ou métier de cordonnier : Sen. Ep. 90, 23.' Above the definition, there are several action buttons: 'FINISHED by', 'edit', 'duplicate & edit', 'delete', 'view axie', and 'view history'.

Figure 5 : Affichage de l'article "sutrinum" du dictionnaire Gaffiot sur la plateforme Jibiki

¹ <http://hub.docker.com/r/mangeot/jibiki/>

Conclusion

Un premier outil fonctionnel de lecture active de textes en latin a été achevé. Plusieurs perspectives sont maintenant envisageables.

Concernant les données, le dictionnaire Gaffiot XML pourrait être d'avantage structuré avec l'identification des traductions et ainsi permettre l'affichage uniquement des traductions au lieu de l'article complet, ce qui gêne la lecture du texte. La plateforme Jibiki permet également d'éditer les ressources. Nous pourrions lancer un projet coopératif d'enrichissement de la ressource XML nouvellement construite en recueillant des contributions en ligne sur la plateforme. La ressource gagnerait à être valorisée d'avantage en permettant aux utilisateurs de la télécharger dans son intégralité.

Concernant l'intégration avec l'environnement du projet Lateo, le travail reste à faire. Nous envisageons d'analyser au préalable les textes de Tacite ainsi que les commentaires pour proposer une lecture active immédiate sur tout le texte.

Références

- Slimane Abdellaoui, Valérie Bellynck, Mathieu Mangeot & Christian Boitet (2018) Outillage de l'accès aux textes par la lecture active étymologique multilingue pour apprenants berbérophones et arabophones. Actes de l'atelier Traitement Automatique des Langues Africaines TALAf 2018, conférence LTT 2018, Grenoble, France, 26 septembre 2018, 15 p.
- Chantal Enguehard & Mathieu Mangeot (2014) *DILAF : des dictionnaires africains en ligne et une méthodologie*. Actes du colloque "Francophonie et Langues Nationales", Dakar, Sénégal, 22 novembre 2014, 9 p.
- Helmut Schmid (1995) *Improvements in Part-of-Speech Tagging with an Application to German*. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland.
- Ying Zhang, Mathieu Mangeot, Valérie Bellynck & Christian Boitet (2014) Jibiki-LINKS: a tool between traditional dictionaries and lexical networks for modelling lexical resources. Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex) 2014 (Eds. Michael Zock, Reinhard Rapp, Chu-Ren Huang), Dublin, Ireland, 23 August 2014, 12 p.