



HAL
open science

**NooJ, un outil TAL pour l'enseignement des langues.
Application pour l'étude de la morphologie lexicale en
FLE**

Max Silberztein, Agnès Tutin

► **To cite this version:**

Max Silberztein, Agnès Tutin. NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE. ALSIC - Apprentissage des Langues et Systèmes d'Information et de Communication, 2005, 08 (1), pp.123-134. 10.4000/alsic.336 . edutice-00109619

HAL Id: edutice-00109619

<https://edutice.hal.science/edutice-00109619>

Submitted on 25 Oct 2006

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.

NooJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE

Max SILBERZTEIN¹, Agnès TUTIN²

¹Université de Franche-Comté, France ; ²université Grenoble 3, France

Résumé : *NooJ est un système de traitement de corpus – reprenant et améliorant les fonctionnalités d'INTEX – conçu pour l'enseignement des langues et de la linguistique. NooJ intègre des outils de traitement automatique du langage qui offrent à l'enseignant des possibilités de traiter un corpus, et des procédures de recherche, de test, et d'entraînement pour l'étudiant. Nous présentons ici un exemple d'application de NooJ à l'enseignement du français langue étrangère, qui reprend quelques activités sur l'étude de la morphologie lexicale.*

- 1. Introduction
- 2. Nooj, un outil TAL pour les applications pédagogiques
- 3. Une application : l'étude des dérivés verbaux dans une perspective FLE
- 4. Conclusion et perspectives
- Références
- Annexes

1. Introduction

L'intérêt des corpus électroniques pour l'enseignement des langues et de la linguistique n'est plus à démontrer et fait désormais l'objet de conférences et d'études approfondies, surtout dans la communauté universitaire anglo-saxonne. L'exploitation des corpus peut être plus ciblée et plus fine si l'on recourt à des outils de traitement automatique du langage (TAL).

(1) Tout d'abord, par rapport à des outils informatiques ordinaires utilisant des concordances "brutes", le TAL permet des explorations complexes sur des informations linguistiques fines, comme les types grammaticaux ou les types sémantiques ([Kraif03], [Brun02]). Avec un système comme *INTEX* ([Silberztein93], [Silberztein99]), on peut, par exemple, effectuer des recherches sur des catégories grammaticales d'un certain type, certaines fonctions syntaxiques, appliquer des grammaires locales pour cerner des phénomènes précis [Garrigues99]. Ces explorations peuvent être menées de façon interactive par les apprenants ou être effectuées en amont pour un prétraitement du corpus.

(2) Par ailleurs, le TAL permet de travailler sur des volumes textuels considérables et variés, les outils TAL étant *a priori* adaptables à tout type de texte, voire même à plusieurs langues. On peut ainsi imaginer proposer la génération automatique d'activités dont le contenu serait renouvelé selon les ressources TAL et les corpus employés [Chanier00], un peu à la façon des exercices lexicaux générés à partir de corpus étiquetés et de ressources lexicales, comme dans le projet *ALFALEX* [Selva02].

(3) Enfin, de façon plus originale, les outils ordinaires du TAL peuvent être détournés et manipulés pour proposer une approche réflexive dans l'apprentissage des langues et de la linguistique. Par exemple, pour l'étude des mécanismes de morphologie lexicale, un repérage des mots construits dérivés de verbe sera proposé, ainsi qu'un classement des modes de construction.

Cependant, malgré son intérêt évident, le TAL ne s'adapte pas de façon immédiate à l'apprentissage / enseignement des langues et de la linguistique. La première difficulté est tout d'abord liée aux performances imparfaites des outils de TAL qui doivent de ce fait être utilisés avec circonspection dans les applications pédagogiques. En effet, si quelques applications simples (comme l'étiquetage morpho-syntaxique) donnent de bons résultats, elles ne sont jamais exemptes d'erreurs, ce qui est gênant dans un contexte d'enseignement. Deux options sont alors possibles : a) proposer un post-traitement des outils TAL, b) proposer une supervision par l'enseignant des outils de TAL.

Deuxièmement, les ressources linguistiques utilisées par les outils de TAL sont longues et difficiles à élaborer. Par exemple, les outils de génération lexicale développés par Thierry Selva dans le cadre d'*ALFALEX* [ALFALEXnd] reposent sur un dictionnaire informatisé complexe, le *DAFLES*, intégrant des relations paradigmatiques (synonymes, etc.) et syntagmatiques (collocations, informations sur le régime des mots). L'ajout ou la modification de ces informations lexicales nécessite une réelle expertise linguistique.

Enfin, la manipulation des outils TAL eux-mêmes est relativement complexe, et peu d'outils conviviaux ont été à ce jour développés et peuvent être mis entre les mains de non-spécialistes. Malgré les difficultés mentionnées, nous pensons néanmoins que certains outils de TAL sont extrêmement prometteurs pour les applications pédagogiques, en particulier si les applications sont elles-mêmes développées par les enseignants.

Cet article présente un système de traitement de corpus, *NooJ*, qui présente des fonctionnalités spécifiquement conçues pour l'enseignement des langues et de la linguistique. *NooJ* reprend certaines fonctionnalités d'*INTEX* [Silberztein99], en les améliorant et en les adaptant au développement d'activités pédagogiques. Après avoir exposé les principales fonctionnalités de *NooJ* pour le prétraitement des corpus et les emplois pédagogiques, nous présentons quelques activités pédagogiques sur l'étude de la morphologie lexicale dans un contexte d'apprentissage du français langue étrangère (FLE).

2. *NooJ*, un outil TAL pour les applications pédagogiques

NooJ, tout comme *INTEX*, est un environnement de développement linguistique permettant de construire, de tester et de gérer des descriptions formalisées à large couverture des langues naturelles, sous forme de dictionnaires et de grammaires électroniques, ainsi que de développer des

applications du TAL. Aujourd'hui, il existe des modules linguistiques *INTEX* pour une douzaine de langues, et une demi-douzaine d'applications informatiques du TAL ont été construites avec *INTEX* (cf. [INTEXnd] pour des exemples d'utilisation du logiciel).

Les fonctionnalités d'*INTEX* étaient adaptées à un public de linguistes (description de la morphologie et de la syntaxe des langues), de documentalistes (analyse de corpus) et d'informaticiens du TAL (applications d'extraction d'information). Mais le logiciel a attiré très tôt des enseignants tant en linguistique qu'en FLE [Garrigues99], ses qualités principales pour une utilisation pédagogique étant les suivantes.

- Son ouverture. Il n'y a ni traitement probabiliste, ni "boite noire" dans *INTEX*. Ses fonctionnalités sont toutes accessibles et paramétrables. Les résultats de traitement, y compris intermédiaires, sont immédiatement compréhensibles.
- Sa grande facilité d'utilisation. On peut construire sa première grammaire locale en dix minutes. Un enseignant non informaticien peut adapter rapidement les dictionnaires et les grammaires du système pour les appliquer, avec les apprenants, à des corpus de textes de son choix.
- La disponibilité et la compatibilité de données linguistiques à large couverture pour une douzaine de langues.

La nouvelle mouture du logiciel *INTEX*, appelée *NooJ* ([NooJ05], [Silberztein04]) a été réécrite en particulier pour répondre aux besoins des utilisations pédagogiques. Quatre nouvelles caractéristiques de *NooJ* devraient présenter un intérêt pour cette communauté.

- (1) **Traitement de corpus.** *NooJ* peut traiter directement des ensembles potentiellement importants de documents, qui peuvent être dans n'importe quel format : pages Internet, documents XML, *Microsoft WORD*®, etc.
- (2) **Les travaux pratiques.** *NooJ* contient des mini-applications pédagogiques qui peuvent instantanément être utilisées en séance de travaux pratiques (TP ou "Labs"). Ces mini-applications donnent aux enseignants la possibilité de monter une séance de TP autour d'un problème linguistique ou informatique spécifique. À l'heure actuelle, une demi-douzaine de TP permettent d'illustrer des problèmes comme : les différents codages informatiques des textes, l'ordre alphabétique dans chaque langue, les expressions rationnelles, la morphologie flexionnelle, etc. Par exemple, l'annexe 1 montre le TP de morphologie flexionnelle. On peut ajouter très facilement des TP dans *NooJ* : le but est d'en construire rapidement une vingtaine, mais rien n'en limite le nombre.
- (3) **Les projets.** À tout moment, on peut sauvegarder la configuration de l'environnement *NooJ* sous la forme d'un fichier "projet". Ce projet peut être rappelé ensuite à volonté. Par exemple, un enseignant peut préparer un corpus, quelques dictionnaires, quelques grammaires, utiliser certaines ressources lexicales et syntaxiques de *NooJ* (avec un jeu complexe de priorité), rédiger une documentation (ou une série d'exercices), et sauvegarder le tout sous forme d'un projet. En chargeant ce projet, les apprenants obtiennent immédiatement la configuration voulue, avec sa documentation.

- (4) **Construction, édition et gestion sophistiquées de concordances.** *NooJ* permet de construire et de gérer des concordances sophistiquées. On peut ainsi combiner plusieurs requêtes, filtrer manuellement les résultats incorrects, puis sauvegarder la concordance résultante, ce qui permet d'aborder des problèmes linguistiques incrémentalement.

On demandera tout d'abord à l'étudiant d'identifier certains termes ou constructions dans un corpus, tout d'abord de façon naïve (par exemple, rechercher tous les mots finissant par *-ment*, pour identifier des adverbes, ou rechercher tous les nombres pour identifier des dates). Puis, au vu de la concordance résultante, on aidera l'apprenant à affiner petit à petit la caractérisation des phénomènes, d'une part en améliorant les requêtes (par exemple en tenant compte du contexte des mots), d'autre part en classifiant les résultats obtenus. Chaque étape fera alors l'objet d'exercices, dont le résultat (une concordance) peut être édité, sauvegardé puis contrôlé par l'enseignant.

3. Une application : l'étude des dérivés verbaux dans une perspective FLE

Pour illustrer cette dernière fonctionnalité, nous présentons quelques activités développées avec *NooJ* autour de l'étude des dérivés verbaux dans une perspective FLE, destinées à des apprenants de niveau intermédiaire et avancé.

L'étude du lexique en FLE, comme d'ailleurs en langue maternelle, a longtemps été délaissée au profit d'activités plus attrayantes mettant en jeu les habiletés communicatives. L'enseignement du lexique, bien qu'incontournable pour la maîtrise d'une langue seconde, était perçu comme difficile à systématiser (le lexique étant le lieu de l'"idiosyncrasie"), donc peu gratifiant sur le plan didactique. Il ne pouvait être abordé en tant qu'activité autonome, "pour soi", mais était associé à des activités de compréhension ou de production [Cuq04].

Depuis peu, les études sur le lexique et son acquisition connaissent cependant un nouvel essor. Au-delà des classiques champs lexicaux de la sémantique structurale, les travaux des linguistes mettent au jour des régularités dans la formation des mots (par exemple, travaux du laboratoire *SILEX* autour du modèle développé par D. Corbin en morphologie lexicale [Corbin87]) et dans les associations lexicales comme les collocations.

Par ailleurs, il est démontré [Bogaards94] qu'un apprentissage systématique du lexique est possible s'il est accompagné d'une description riche des phénomènes lexicaux et d'une réflexion et d'une implication personnelle de l'apprenant. En outre, l'importance d'un apprentissage lexical en contexte est également soulignée dans [Treville96].

L'activité que nous proposons ici est l'observation sur corpus de dérivés verbaux pour en étudier de façon inductive le mode de formation. Nous partons de l'hypothèse qu'une activité métalinguistique menée de façon active par l'apprenant à partir d'exemples sur corpus renforcera ses compétences lexicales. Plus précisément, l'objectif visé est la prise de conscience de mécanismes morphologiques réguliers dans la formation des mots, les règles associées aux bases lexicales permettant d'accroître largement, au moins en compréhension, le stock lexical disponible. L'activité est ici prévue en présentiel, sous forme d'atelier, avec l'accompagnement de l'enseignant.

3.1. La consultation et le filtrage du corpus brut : la recherche de noms déverbaux

Une première application simple de *NooJ* est la consultation et l'extraction d'informations à partir du corpus "brut", *i. e.* ne comportant pas de balises spécifiques. *NooJ* contient un analyseur morphologique qui permet d'effectuer des recherches et des traitements dans les textes à partir d'expressions régulières intégrant des formes, des lemmes, des catégories syntaxiques ou toute information lexicale. Dans l'exemple 1, l'expression suivante permet de rechercher dans un corpus tous les noms qui finissent par *-ion*, *-age* ou *-ment*.

(1)

```
<N+MP="(ion|age|ment)$">
```

(Requête simple pour rechercher tous les noms (N) qui se terminent (\$) par *-ion*, *-age* ou *-ment*)

On peut demander à l'étudiant d'extraire les occurrences des noms déverbaux se terminant par une suite de morphèmes, dans un corpus de petite taille choisi par l'enseignant, par exemple un article de presse simple. Pour ce faire, l'étudiant devra formuler des requêtes adaptées, extraire les concordances, et filtrer les résultats inadéquats. Il pourra s'aider du TP sur les expressions PERL (ce qui est écrit entre guillemets dans la requête de l'exemple 1).

La copie d'écran suivante (Figure 1) montre une concordance extraite à partir de la requête précédente. L'étudiant a alors pour tâche de sélectionner parmi les entrées de la concordance obtenues celles qui correspondent vraiment à des noms déverbaux (et d'effacer les noms comme *lion*, *nation*, etc.).

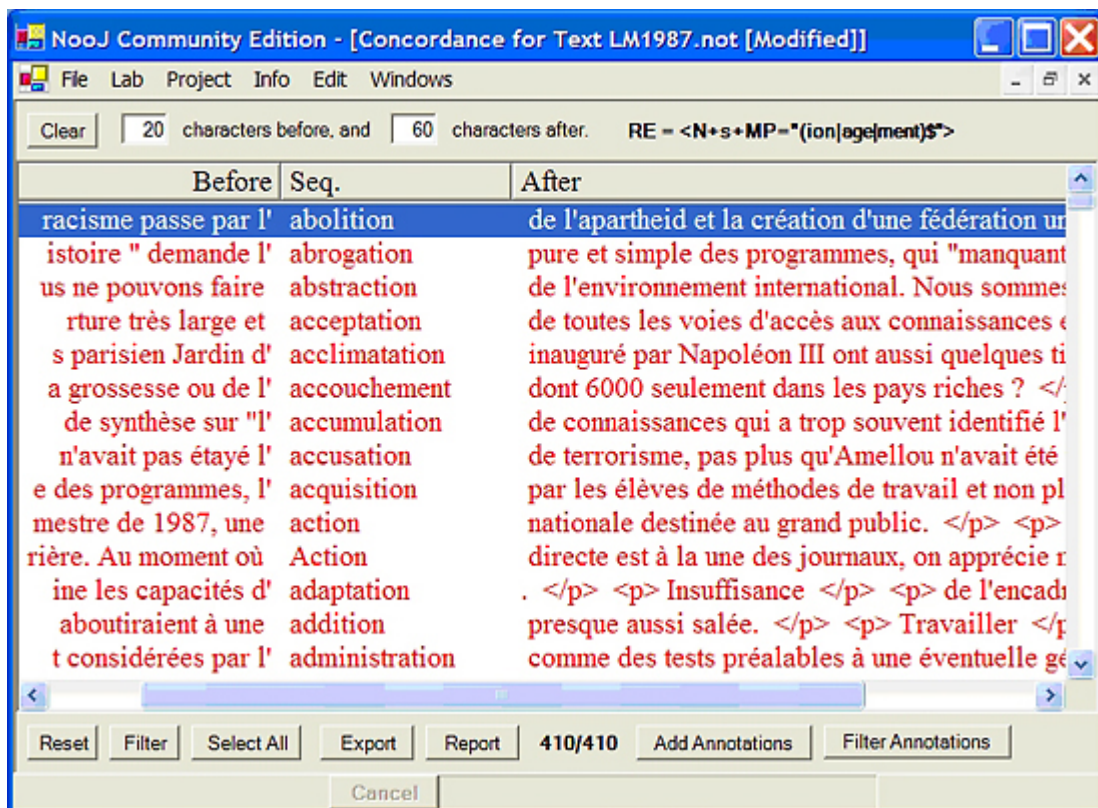


Figure 1 - Une première concordance.

Pour repérer les noms déverbaux, l'étudiant peut exploiter un corpus plus large à sa disposition (*NooJ* permet d'ouvrir simultanément plusieurs corpus) en essayant de retrouver le verbe à l'origine de la dérivation, à partir de la troncation. Par exemple, l'étudiant est confronté au mot **investissement** dans le corpus d'étude. Pour déterminer s'il s'agit d'un déverbal, l'étudiant exploite un corpus plus large, par exemple un gros corpus de presse comme *Le Monde*, et applique la requête de l'exemple 2. En examinant la concordance résultante, il détermine que le nom **investissement** est bien dérivé du verbe **investir**.

(2) <V+MP="^invest">

(Requête pour rechercher tous les verbes qui commencent par **invest**)

On peut imaginer des prolongements à ce simple relevé, comme par exemple le repérage des différents allomorphes pour une dérivation (par exemple, **-ment** vs **-issement**) et la recherche du paramètre motivant la forme du suffixe (ici, la forme du verbe de base).

3.2. Annotation interactive du corpus

NooJ permet également d'appliquer des grammaires (représentées graphiquement) à un corpus. Par exemple la grammaire en annexe 2 reconnaît des expressions de date du français. On pourra présenter cette grammaire à des étudiants de FLE, leur demander de l'appliquer à un corpus de textes journalistiques, et éventuellement de la compléter (pour détecter des dates comme **avant Noël, deux jours après**, etc.).

Lorsqu'on applique une grammaire à un corpus, le résultat est là aussi une concordance (dont la colonne du milieu contient cette fois une séquence de mots). L'utilisateur peut valider ou non chaque entrée de la concordance ; le résultat peut alors être utilisé pour annoter le texte. L'annotation d'un corpus peut être utilisée par l'enseignant pour, d'une part, adapter les résultats des traitements TAL effectués automatiquement (lorsqu'il prépare le corpus), et, d'autre part, pour vérifier et corriger les traitements effectués par l'apprenant.

Pour le sujet qui nous intéresse (les dérivés verbaux), plusieurs annotations peuvent être appliquées aux corpus. Premièrement, on observera effectivement si le nom en **-ion**, **-age** ou **-ment** est ou non dérivé d'un verbe. Bien entendu, il ne s'agit pas de transformer les étudiants en morphologues, et on ne s'intéressera ici qu'aux dérivations perceptibles en synchronie (par exemple, le terme **télévision** ne sera pas retenu). Deuxièmement, on observera si le nom dérivé a une valeur prédicative (il exprime un procès ou une **action**) ou s'il est concret. Par exemple, **construction** dans **la construction de ces établissements** sera considéré comme prédicatif, alors que dans **cette construction est impressionnante**, il est un nom concret. L'exemple de la figure 2 présente une telle annotation, sous format XML, qui sera proposée à l'apprenant.

Le folklore hongrois avait sa place aux côtés de la littérature ou des Beaux-arts français, la musique classique allemande, et les Beatles... universels. Pour les Hongrois, il est clair, que l'entrée de la Hongrie dans l'Europe unie est une <deverbal pred="non">obligation</deverbal> historique, qui arrive même un peu tard. En réalité, avec le <deverbal pred="oui">changement</deverbal> politique, l'entrée en tant que telle ne change pas grand-chose. Aujourd'hui, les Hongrois voyagent déjà beaucoup, font du tourisme, suivent les conférences internationales, les marchés financiers.

Figure 2 - Annotation en format XML d'un corpus, extrait du site de TV5, témoignage du journaliste hongrois Andras Davis
http://www.tv5.org/TV5Site/enseignants/apprendre_francais.php

NooJ permet d'interroger ensuite le corpus en exploitant ces annotations. L'étudiant peut alors effectuer des consultations fines sur le corpus annoté, comme celle de l'exemple 3. Ici aussi, le résultat de la requête est une concordance qui peut être combinée avec d'autres, filtrée, puis sauvegardée avec son historique.

(3)

```
<DEVERBAL> (<A> + <E>) (à + de) <V+INF>  
(déverbal, suivi d'un adjectif optionnel, suivi de à ou de,  
suivi d'un verbe à l'infinitif)
```

NooJ peut consulter des corpus préalablement annotés à l'aide de balises XML. Dans ce cas, les balises XML ont le même statut que les annotations produites par *NooJ*, qu'elles soient lexicales, syntaxiques ou sémantiques.

4. Conclusion et perspectives

NooJ présente des fonctionnalités de TAL qui paraissent prometteuses pour l'enseignement des langues et de la linguistique. Son principal avantage est sa simplicité d'utilisation : il permet à la fois à l'enseignant non spécialiste de TAL de constituer des ressources linguistiques (à l'aide d'interfaces simples) et de les paramétrer afin de constituer des projets pédagogiques destinés aux apprenants. Pour développer une activité pédagogique à l'aide de *NooJ*, il n'est donc pas nécessaire de recourir à une chaîne de traitement complexe, ni de faire appel à des spécialistes d'informatique ou de TAL. Cela garantit une cohérence entre les traitements linguistiques et les exploitations pédagogiques qui en sont faites.

Il reste encore à développer plusieurs fonctionnalités pour les applications pédagogiques. Tout d'abord, à court terme, la génération automatique d'activités contextuelles, un peu à la façon du générateur d'activités de *MIRTO* [Antoniadis04], mais en exploitant les corpus annotés avec *NooJ*. L'enseignant définirait une requête destinée à l'étudiant (par exemple, "Discriminez les différents types de constructions déverbales"). Les exercices seraient alors générés automatiquement, en échantillonnant de façon aléatoire une concordance *NooJ*. La qualité des données extraites serait alors garantie, puisque le corpus a été préalablement annoté à l'aide d'outils TAL dont l'application peut être vérifiée (et au besoin corrigée au cas par cas) par l'enseignant.

Par ailleurs, à moyen terme, d'autres fonctionnalités plus avancées de *NooJ*, en particulier les transformations syntaxiques, pourraient aussi être exploitées comme démonstrateurs d'opérations

linguistiques [Buvetnd]. Par exemple, dans le cas des déverbaux, l'étudiant pourrait jouer avec une série de transformations permettant de classer la construction à laquelle il a affaire. En sélectionnant l'expression *l'investissement de la France*, il pourrait alors obtenir *la France a investi*, etc.

Les activités proposées doivent maintenant faire l'objet d'évaluations pédagogiques, qui détermineront à leur tour les développements ultérieurs de *NooJ*.

Références

Les liens externes étaient valides à la date de publication.

Bibliographie

[Antoniadis04]

Antoniadis, G., Echinard, S., Kraif, O., Lebarbé, T., Loiseau, M. & Ponton, C. (2005). "Modélisation de l'intégration de ressources TAL pour l'apprentissage des langues : la plateforme MIRTO". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*. vol. 8. http://alsic.org/v08/antoniadis/alsic_v08_04-rec4.htm

[Bogaards94]

Bogaards, P. (1994). *Le vocabulaire dans l'apprentissage des langues étrangères*. Paris : Crefid-Hatier.

[Brun02]

Brun, C., Parmentier, T., Sándor, A. & Segond, F. (2002). "Les outils de TAL au service de la e-formation en langues". In Segond, F. (dir.) *Multilinguisme et traitement de l'information*. Paris : Hermès. pp. 223-249.

[Buvetnd]

Buvet, P.-A. & Silberztein, M. (à paraître). "Procédures de désambiguïsation sémantique pour la détermination du domaine d'un texte". *Revue Québécoise de Linguistique*.

[Chanier00]

Chanier, T. & Selva, Th. (2000). "Génération automatique d'activités lexicales dans le système ALEXIA". *Sciences et techniques éducatives (STE)*, vol. 7, 2. pp. 385-412.

[Corbin87]

Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Lille : Presses Universitaires de Lille.

[Cuq04]

Cuq, J.-P. (2004). "Le lexique en situation d'apprentissage guidé : pour une méthodologie d'enseignement interventionniste dans l'enseignement du français langue étrangère". In Calaque, E. & David, J. (dir.) *Didactique du lexique*. Bruxelles : de Boeck.

[Garrigues99]

Garrigues, M. (1999). "Nouvelles concordances pour l'enseignement des langues", *Linguisticae Investigationes*, Tome XXII (1998-1999). pp. 59-70.

[Kraif03]

Kraif, O. (2003). "Propositions pour l'intégration d'outils TAL aux dispositifs informatisés d'apprentissage des langues". *Lidil*, n° 28. pp. 53-165.

[Selva02]

Selva, Th. (2002) "Génération automatique d'exercices contextuels de vocabulaire". *Actes de TALN*, Nancy 2002. pp. 185-194.

[Silberztein93]

Silberztein, M. (1993). *Dictionnaires électroniques et analyse de textes : le système INTEX*. Masson : Paris.

[Silberztein99]

Silberztein, M. (1999). "Indexing large corpora with INTEX". *Computer and the Humanities*, n° 33-3. pp. 265-280.

[Silberztein04]

Silberztein, M. (2004). "NooJ : an oriented object approach". In Royauté, J. & Silberztein, M. (dir.) *INTEX pour la Linguistique et le Traitement Automatique des Langues. Proceedings of the 4th and 5th INTEX workshop*, Bordeaux, mai 2001 et Marseille, mai 2002. Besançon : Presses universitaires de Franche-Comté.

[Treville96]

Tréville, M.-C. & Duquette, L. (1996). *Enseigner le vocabulaire en classe de langue*. Paris : Hachette.

Logiciels et sites

[ALFALEXnd]

ALFALEX (nd). *Site présentant le logiciel ALFALEX, génération d'exercices lexicaux*, développé par Th. Selva. <http://www.kuleuven.ac.be/alfalex>

[INTEXnd]

INTEX (nd). *Site du logiciel INTEX*, disponible et gratuitement téléchargeable. <http://intex.univ-fcomte.fr>

[NooJ05]

NooJ (2005). *Site du logiciel NooJ*, disponible et gratuitement téléchargeable. <http://www.nooj4nlp.net>

Annexes

Annexe 1. Travaux Pratiques autour de la morphologie flexionnelle

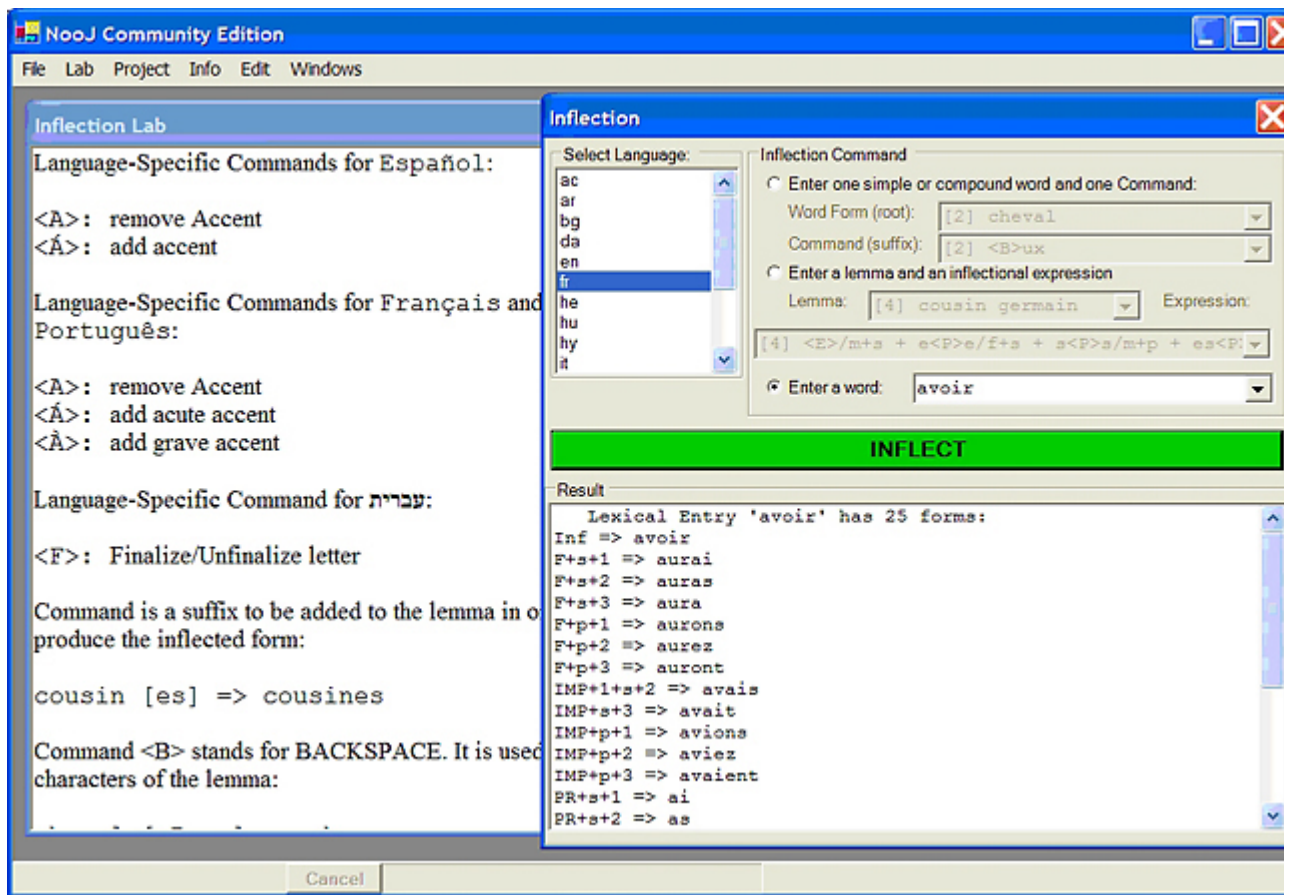


Figure 3 - Exemple de TP.

La séance de TP Inflectional Lab permet d'étudier la morphologie flexionnelle d'une douzaine de langues. L'étudiant sélectionne une langue (**Select Language**), puis utilise trois options.

- (1) Il peut entrer un lemme (par ex. **cheval**), puis un opérateur morphologique (ux), afin de construire une forme fléchie donnée (**chevaux**). Lorsqu'il clique sur le bouton vert INFLECT, une petite animation apparaît, qui illustre en temps réel le calcul de la forme fléchie à partir du lemme donné : l'étudiant visualise ainsi d'éventuelles erreurs. Les opérateurs morphologiques de chaque langue sont décrits dans la documentation.
- (2) Il peut entrer un lemme (par ex. **cousin germain**), puis une expression morphologique, pour générer l'ensemble des formes fléchies associées au lemme (**cousin germain**, **cousine germaine**, **cousins germains**, **cousines germaines**). Les expressions morphologiques sont des listes d'opérateurs morphologiques associés à des codes linguistiques (par ex. +f+p pour "féminin pluriel"). Là aussi, la documentation explique comment construire des expressions morphologiques.
- (3) Il peut entrer un mot (simple ou composé) ; *NooJ* consulte alors ses dictionnaires pour retrouver le modèle de flexion du mot, puis génère l'ensemble des formes fléchies du mot.

Annexe 2. Une grammaire d'expressions de date en français

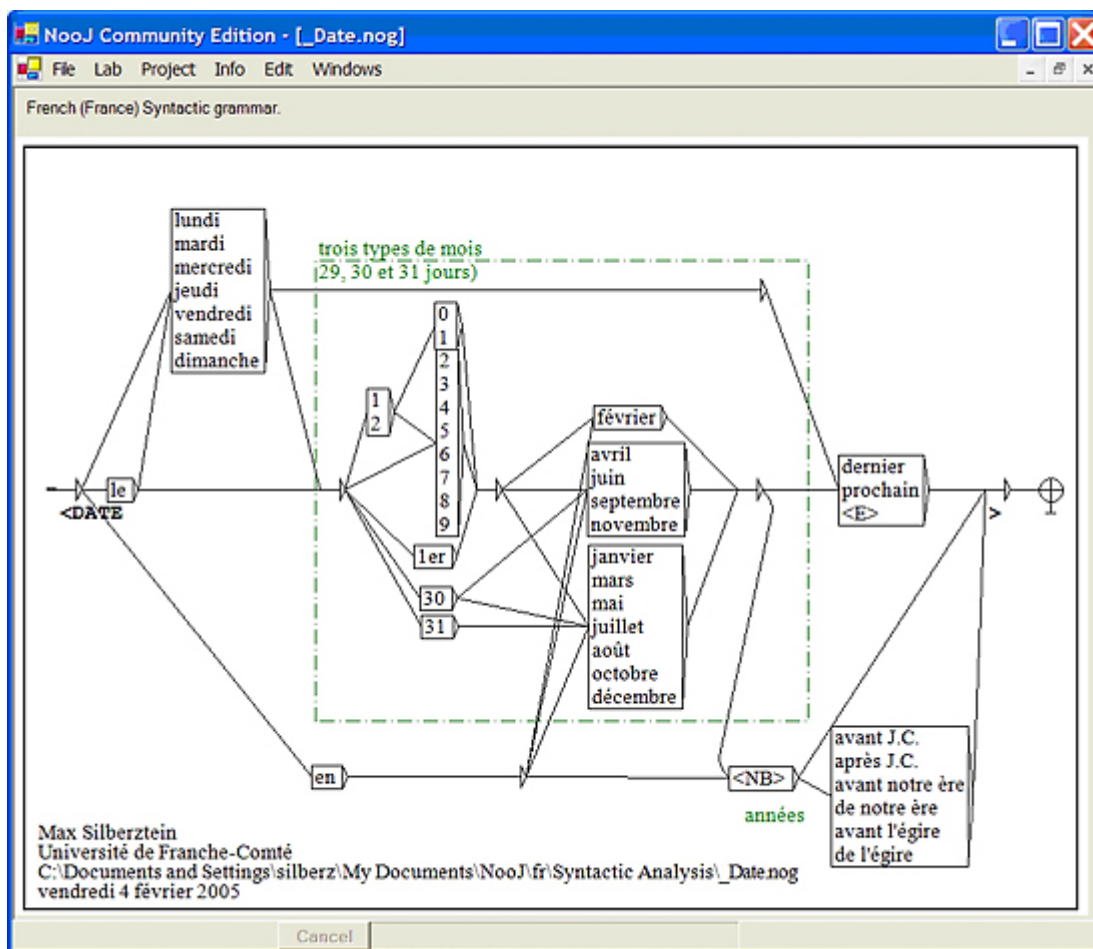


Figure 4 - Grammaire des dates.

Cette grammaire reconnaît des expressions comme celles figurant dans l'exemple 4.

(4)

le lundi 15 février prochain, mardi 1er avril 1999, en 52 avant J.C.

On présentera cette grammaire à l'étudiant, en lui demandant de l'appliquer à un corpus pour construire la concordance des séquences reconnues. L'étudiant devra ensuite distinguer les deux cas de figure de l'exemple 5.

(5)

Luc a détecté **en 2000** qu'il y avait ... (date correctement reconnue)

Luc a détecté **en 2000** endroits différents ... (date reconnue à tort)

Une fois filtrée, la concordance peut être utilisée pour introduire une annotation (ici <DATE>) dans le corpus. Les annotations sont instantanément utilisables dans d'autres requêtes NooJ. Par exemple, la requête <V+Fut> <DATE> recherche un verbe au futur suivi d'une date. Les possibilités d'exercices sont nombreuses. On peut demander à des étudiants d'étudier ou d'appliquer cette grammaire à différents types de textes (en deux heures) ou, pour des étudiants linguistes ou pour

des projets plus complets, d'en construire une version simplifiée, voire d'autres grammaires similaires (les heures, les adresses, etc.).

À propos des auteurs

Max SILBERZTEIN est professeur d'informatique à l'UFR lettres et sciences humaines de l'université de Franche-Comté depuis 1997.

Courriel : max.silberztein@univ-fcomte.fr

Toile : www.nooj4nlp.net

Adresse : Laseldi, université de Franche-Comté, 30 rue Mégevand, 25000 Besançon, France.

Agnès TUTIN est maître de conférences, chercheuse au laboratoire Lidilem en sémantique lexicale et TAL.

Courriel : agnes.tutin@u-grenoble3.fr

Toile : www.u-grenoble3.fr/tutin

Adresse : Laboratoire de Linguistique et Didactique des Langues Étrangères et Maternelles (Lidilem), université Grenoble 3, BP 25, 38040 Grenoble Cédex 09, France.

Ce texte fait partie des textes de la journée Atala 2005 qui font l'objet d'un numéro spécial d'*Alsic*.

Date de réception de l'article, 1er novembre 2004 ; date d'acceptation, 30 mai 2005.

Référence de l'article :

Silberztein, M. & Tutin, A. (2005). "NoaJ, un outil TAL pour l'enseignement des langues. Application pour l'étude de la morphologie lexicale en FLE". *Apprentissage des langues et systèmes d'information et de communication (ALSIC)*, vol. 8, n° 1. pp. 123-134. http://alsic.u-strasbg.fr/v08/silberztein/alsic_v08_20-rec11.htm, mis en ligne le 15/12/2005.



[ALSIC](#) | [Sommaire](#) | [Consignes aux auteurs](#) | [Comité de rédaction](#) | [Inscription](#)

© *Apprentissage des Langues et Systèmes d'Information et de Communication*, décembre 2005