

# INTEX dans l'enseignement universitaire

Rolf KAILUWEIT

**Abstract:** This paper is about the experiences I made using the INTEX program at Heidelberg University in a class on Lexical grammar. In the first part of the paper I talk about the participants and the organization of the class. In the second part I look at the first six lessons more in detail in order to illustrate the knowledge I expected the students to have when I gave them a questionnaire about their experiences with INTEX. In the third part I analyze the answers of the questionnaire hoping that this will be useful for other people wanting to use INTEX for teaching. In addition, I believe that pointing out some elementary problems that experienced users might not notice any more can help to amend the program in the future.

**Keywords:** teaching, lexicon-grammar, dictionary, disambiguation, Glossanet, tagger.

**Mots clés :** enseignement, lexique-grammaire, dictionnaire, levée d'ambiguïtés, Glossanet, transducteur d'étiqutage.

## 1. Introduction

L'objectif de cet article est de présenter les expériences que j'ai faites en utilisant INTEX dans le cours de linguistique que j'ai donné au Département de Philologie Romane de l'Université de Heidelberg pendant le semestre d'été 2000. J'espère que cette façon de mettre en évidence les premières expériences que les nouveaux utilisateurs ont faites avec INTEX pourra servir à donner une orientation à ceux qui veulent utiliser le programme dans l'enseignement et, en plus, à déterminer quelques insuffisances de la version actuelle dont l'utilisateur expérimenté ne se rend peut-être plus compte.

---

✉ Dr. Rolf KAILUWEIT, Universität de Heidelberg, Romanisches Seminar  
Universität Heidelberg, Seminarstraße, 3, D-69117 Heidelberg  
Fon: mobil: +49 (0) 172 808 79 50. Büro: +49 (0) 6221 / 54 35 96  
Fax: +49 (0) 6221 / 54 35 94 e-mail : rolf.kailuweit@urz.uni-heidelberg.de

---

La première partie de cet article présentera le cours et les participants. Dans la deuxième partie j'expliquerai plus en détail le déroulement des six premières unités d'enseignement. Enfin, la troisième partie résumera les résultats d'un questionnaire que les participants ont rempli depuis les premières six unités.

## 2. Le cours et les participants

Le cours a combiné une introduction théorique au lexique-grammaire avec des exercices pratiques d'INTEX. Il était destiné aux étudiants de philologie romane et de linguistique informatique de la deuxième et troisième année. Il s'agissait d'un cours facultatif, c'est-à-dire que les étudiants l'ont choisi parmi un ensemble de cours de sujets divers qui sont équivalents au niveau du plan d'études. Le fait que 28 étudiants s'y soient inscrits est remarquable, étant donné que la moyenne des inscriptions à un cours de ce type est entre 10 et 15. Il y avait pourtant une certaine fluctuation qui est, d'ailleurs, tout à fait normale : sept personnes ont abandonné le cours au fur et à mesure que le semestre a avancé, mais 21 étudiants ont participé de façon active : ils se sont chargés d'un sujet qu'ils ont présenté en classe.

Parmi les inscrits, il y avait trois étudiants de linguistique informatique et trois étudiantes Erasmus catalanes. Les autres étaient des étudiants de philologie romane qui ont l'espagnol (5), l'italien (2) ou le français (le reste) comme première matière.

Études		Inscrits	Ayant abandonné le cours
Philologie romane	Français	15	5
	Espagnol	5	1
	Italien	2	1
Linguistique informatique		3	
Traduction et interprétation (catalan – allemand)		3	

Tous les étudiants possédaient des connaissances au moins passives d'anglais et de français. Leurs connaissances d'informatique variaient beaucoup, mais tous étaient capables de réaliser une recherche sur Internet et d'utiliser un éditeur de texte comme MS Word et un programme de messages électroniques.

Pour obtenir le certificat prévu pour ce cours, les étudiants devaient présenter un sujet en classe et réaliser un petit travail de recherche pendant les vacances d'été suivantes où ils étaient tenus d'appliquer quelques théorèmes du lexique-grammaire à un corpus établi à l'aide d'INTEX. À la bibliothèque du Département, ils avaient à leur disposition un ordinateur où les versions française, espagnole et italienne d'INTEX ont été installées. Ils y trouvaient en plus un classeur avec quelques lectures de base sur le lexique-grammaire et sur INTEX. J'ai composé un site Internet (<http://www.rzuser.uni-heidelberg.de/~at4/PSLexGram.htm>) où se trouvent le programme du séminaire avec les objectifs d'apprentissage pour chaque unité d'enseignement, des exercices, une bibliographie essentielle sur le lexique-grammaire et quelques informations sur le LADL et la communauté INTEX.

### 3. Les unités d'enseignement

Les six premières unités d'enseignement<sup>1</sup> se sont déroulées de la façon suivante :

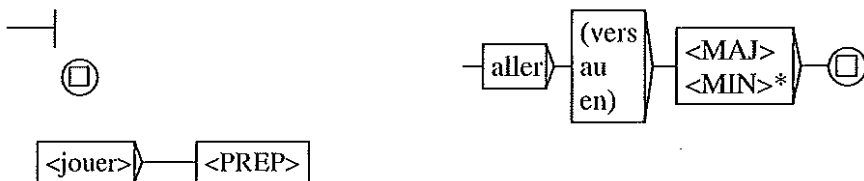
Nous avons commencé avec une introduction générale sur les problèmes de la représentation syntaxique et lexicale à l'égard des applications informatiques. Mon objectif était de donner une première impression sur les insuffisances de la grammaire scolaire et des dictionnaires d'usage et sur la différence entre l'approche du lexique-grammaire et d'autres théories syntaxiques.

---

<sup>1</sup> Je décrirai ces unités en détail pour donner une impression des connaissances des étudiants au moment où ils ont rempli de questionnaire. Les 7 unités de la deuxième moitié du semestre étaient destinées à quelques sujets du lexique-grammaire. Pendant ces unités nous n'avons pas cherché à connaître plus en détail le programme INTEX, mais à utiliser les fonctions élémentaires du programme pour illustrer les résultats de recherche en lexique-grammaire.

La deuxième unité d'enseignement était destinée complètement à l'introduction au programme INTEX. Deux étudiants de linguistique informatique se sont chargés de démontrer les opérations de base à l'aide des instructions du manuel de Max Silberstein (1999–2000). Les étudiants devaient apprendre à charger un texte, appliquer un dictionnaire, rechercher une séquence de mots dans le texte avec une expression régulière et créer un simple graphe de reconnaissance. Je leur ai donné l'exercice de créer un simple graphe, de l'appliquer à un texte et de m'envoyer le graphe et les occurrences via un message électronique.

Les résultats, il faut l'avouer, n'étaient pas tout à fait satisfaisants. Seuls 15 étudiants ont fait l'exercice et parmi des graphes acceptables, il y avait aussi des exemplaires moins réussis comme les suivants :



Il est apparu que le manuel et l'introduction globale en classe sans des exercices guidés n'étaient pas suffisants pour faire apprendre les opérations élémentaires à tous les étudiants. Il fallait faire des tutorats avec des petits groupes d'étudiants.

Pendant la troisième unité, nous avons discuté de l'article de Maurice Gross (1994), « Constructing Lexical Grammars », qu'un étudiant a présenté et que les autres avaient lu par avance. Mon objectif était de donner un aperçu des concepts de base du lexique-grammaire et de l'état de la recherche réalisée au LADL. En plus, nous nous sommes occupé du problème des ambiguïtés au niveau de la représentation FST d'INTEX. Une étudiante a expliqué ce qu'est un transducteur pour la levée d'ambiguïté par grammaires locales en se basant sur les articles de Anne Dister (1999 et 2000).

La quatrième unité était réservée à la présentation de *Méthode en syntaxe* de Maurice Gross (1975) et de *Sentential complementation in Spanish* de Carlos Subirats-Rüggeberg (1987). Nous avons récapitulé

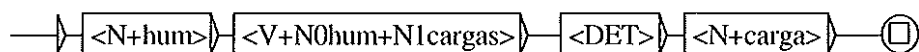
les différences principales entre l'approche du lexique-grammaire et de la grammaire chomskienne, mais mon objectif était surtout pratique : je voulais apprendre aux étudiants à lire les tables du lexique-grammaire pour les rendre capables de travailler avec l'information qu'ils y trouvent. Après cette unité-là, ils devaient être en situation de faire l'exercice suivant : choisir une propriété distributionnelle ou transformationnelle d'une table quelconque, créer un graphe de reconnaissance qui représente plus au moins cette propriété et tester ce graphe avec un certain nombre de verbes pendant quelques jours à l'aide de Glossanet. Cet exercice était encore trop exigeant pour la grande majorité des participants, étant donné qu'ils dominaient les fonctions élémentaires d'INTEX d'une façon assez précaire. Il fallait attendre la fin du semestre : une fois les problèmes techniques surmontés, les étudiants se montraient capables de faire cet exercice.

La cinquième unité d'enseignement était destinée aux ressources lexicales. Deux étudiantes ont présenté en détail les différents dictionnaires disponibles dans les fichiers DELAF et DELACF d'INTEX. Une autre étudiante a approfondi la question de la représentation lexicale à l'égard des applications informatiques en se basant sur quelques articles (BLANCO 1998 et 1999 ; BLANCO et BUVET 1998) que Xavier Blanco a mis à ma disposition. L'objectif était de faire connaître aux étudiants les détails de la représentation lexicale dans INTEX pour les rendre capables de faire des recherches plus raffinées et de créer non seulement des dictionnaires supplémentaires pour les formes simples inconnues, mais aussi de nouveaux dictionnaires qui contiennent des informations sémantiques.

C'est pour cela que nous avons écrit un petit dictionnaire inspiré par des réflexions de Xavier Blanco (1999) pour le verbe espagnol *llevar* qui peut entre autre sélectionner des vêtements et des charges :

corbata, corbata. N+ropa:fs  
 lleva, llevar. V+N0hum+N1cargas:P3s:Y2s  
 lleva, llevar. V+N0hum+N1ropa:P3s:Y2s  
 llevó, llevar. V+N0hum+N1cargas:P3s:Y2s  
 llevó, llevar. V+N0hum+N1ropa:P3s:Y2s  
 maleta, maleta. N+carga:fs

Le graphe suivant reconnaît les constructions *llevar* + charge :



Enfin, la sixième unité d'enseignement a concerné les transducteurs pour la levée d'ambiguïté par grammaires locales. Deux étudiantes ont présenté quelques-uns des graphes que Anne Dister a mis à la disposition de la communauté INTEX. Elles ont ensuite expliqué comment créer un transducteur d'étiquetage. En plus, une étudiante de philologie italienne a comparé les graphes d'Anne Dister à ceux que Simona Vietri a construits pour l'italien. Elle nous a donné une première impression sur les possibilités et les limites de la création des graphes analogues pour l'italien<sup>1</sup>.

#### 4. Le questionnaire

Depuis la sixième unité d'enseignement, j'ai distribué aux étudiants un questionnaire sur les premières impressions et les difficultés qu'ils ont eues en utilisant INTEX, dont je résumerai les résultats. Je ne les présenterai que de façon qualitative. Une présentation statistique n'aurait pas beaucoup de sens, étant donné le nombre total de réponses que j'ai à ma disposition.

<sup>1</sup> J'ai donné une heure de cours hebdomadaire à part pour les étudiantes catalanes Erasmus. Puisqu'il n'y a pas de version catalane d'INTEX, elles ne peuvent faire les exercices dans la langue de leurs études. Aussi voulais-je profiter de la présence de trois étudiantes de langue maternelle catalane pour commencer à créer des ressources pour un futur INTEX catalan. Nous avons affiché la liste d'environ 7000 lexèmes d'un petit texte catalan afin de commencer à créer un mini-DELAF catalan. Les entrées que nous avons faites s'apparentent aux entrées des DELAF français et espagnol. En faisant le dictionnaire, les étudiantes ont trouvé difficile de penser toujours aux ambiguïtés d'une forme qui ne correspondent pas à l'usage concret du texte. Il faut se demander toujours par exemple si un substantif du texte analysé peut être un adjectif dans un autre texte. Ces difficultés ont été résolues à l'aide du dictionnaire standard catalan de l'Institut d'Estudis Catalans. Quant aux substantifs et adjectifs, nous avons travaillé avec la fonction du menu « DELA » « Inflect » après avoir créé des automates correspondant à la morphologie catalane. Mais pour faciliter le travail consistant à remplir les paradigmes verbaux, j'ai préféré programmer des outils externes au programme INTEX. Nous avons fait l'expérience que la création d'un automate d'INTEX pour la flexion verbale était trop inconmode.

1. *Est-ce utile de combiner l'enseignement d'une théorie syntaxique avec des exercices pratiques d'INTEX ?*

Les étudiants trouvaient que c'était une bonne idée en général. Mais pour certains d'entre eux, le cours avait avancé trop vite. Ils n'avaient pas encore les connaissances suffisantes en linguistique générale pour comprendre sans difficultés les théorèmes de lexique-grammaire. Ils auraient préféré un tutorat hebdomadaire pour récapituler la théorie et faire des exercices à l'ordinateur.

2. *Comment jugez-vous le manuel de Max Silberztein ?*

Les réponses étaient en général très positives. Les étudiants trouvaient la structure du manuel très claire et les explications compréhensibles. Ils aimaient bien les figures et surtout les reproductions des écrans qui aidaient beaucoup au travail pratique avec le programme, bien que les informations que donnent les figures (par exemple le nombre des lexèmes ou des étiquettes) ne correspondent pas toujours aux informations du texte.

Quelques-uns ont dit que le manuel était trop détaillé pour les débutants. Par contre, à mon avis, la quantité des informations est juste ce qu'il faut pour apprendre à dominer le programme. D'autres ont critiqué le fait que le manuel utilise une terminologie française qui leur est inconnue et qui en plus se distingue de la langue du programme qui est l'anglais. En fait, il n'est pas facile pour un utilisateur non francophone de se débrouiller avec cette terminologie. Au début du manuel on trouve parfois les concepts anglais correspondant entre parenthèses, mais il serait utile d'avoir un glossaire français-anglais à la fin du manuel. Pourtant, je crois que nous sommes ici devant un problème plus général : le mélange des langues. Nous avons un programme en anglais, avec des informations supplémentaires sur les dictionnaires, sur les .fst, etc. dans les langues diverses selon le goût du programmeur de ces ressources (par exemple : des explications pour le DELAFm ou le Roman\_Numerals.fst anglais sont en français, les explications pour les ressources italiennes sont en anglais). Parmi les symboles formels qu'il faut utiliser pour les recherches indépendamment de la langue figure le mot français <MOT>. Les symboles et les codes flexionnels des DELAF sont abstraits ou bien correspondent

par un heureux hasard aux deux langues français et anglais en même temps. Mais le fait que le DELAF italien brise ce formalisme en le remplaçant par la terminologie grammaticale italienne est à mon avis gênant pour les comparatistes.

3. *L'accès au programme : lancement d'INTEX, chargement d'un texte, application des ressources lexicales, recherche des séquences simples. Y avait-il des difficultés ou des problèmes ?*

Il y a quelques inconvénients au niveau de la programmation des fenêtres : il n'y pas de commande de menu pour quitter le programme. Il faut le quitter en cliquant sur la croix au coin supérieur droit. Par contre, il n'est pas possible de fermer de cette façon le texte, la liste des lexèmes (*token list*), la liste des mots (*word list*), le texte-fst et la concordance. Il faut recourir à la commande « Close » dans le menu texte pour fermer toutes ces fenêtres à la fois.

Quant aux ressources lexicales, le manuel n'explique pas qu'il faut presser sur « Ctrl » et « Enter » à la fois pour choisir plusieurs dictionnaires. Sans presser sur « Ctrl », on ne peut que choisir un ou bien tous les dictionnaires.

Il est déjà bien que la version 4.21 permette de choisir les dictionnaires par défaut (*set default*). Dans le menu « preprocessing » il serait aussi bien de pouvoir modifier les choix par défaut. Il est gênant qu'il faille toujours désélectionner les fonctions « appliquer un .fst dans le mode 'Replace' » ou « appliquer les dictionnaires par défaut », si on n'a pas besoin de ces fonctions.

Quant aux recherches, dans la liste des codes flexionnels manque le symbole pour le subjonctif imparfait. Il faut recourir aux dictionnaires pour vérifier que c'est le symbole <:T>.

4. *En général, comment jugez-vous la structure des menus ?*

Il serait utile d'intégrer la fonction « nouveau » (*new*) dans les menus « texte » et « dela ». Pour les débutants qui veulent tester leurs premiers graphes, il est troublant qu'ils ne puissent accéder directement à un éditeur pour écrire un petit texte de test.

Il y a des fonctions prévues dans les menus auxquelles ne correspond aucune ressource. Ces fonctions vides perturbent les débutants.



5. *Résultats des recherches : est-ce que le programme trouve ce que vous cherchez ?*

Les codes flexionnels italiens posent des problèmes : si on veut par exemple rechercher les participes passés, l'expression <V:K> donne des résultats corrects pour le français, mais l'expression <V:PP> trouve les participes passés, les participes présents et les formes du présent de l'indicatif italiens.

L'usage des opérateurs booléens est très limité. Si on veut par exemple rechercher toutes les formes fléchies d'un verbe, il est impossible de le faire par la simple exclusion des formes non fléchies (<V:!W:!G:!K>).

6. *Créer un graphe : y avait-il des difficultés ou des problèmes ?*

Les débutants ont eu des difficultés avec les opérations de connexion et déconnexion des nœuds, mais ces difficultés semblent inévitables et disparaissent avec un peu d'expérience. Il est pourtant un peu gênant qu'au moment de cliquer sur un nœud pour le connecter, ses étiquettes soient en danger. On pourrait peut-être réserver le double-cliquer vite pour la modification des étiquettes. En plus, la ligne de commande pour faire entrer les étiquettes n'est pas très commode pour les listes plus longues. Peut-être serait-il possible de modifier les étiquettes dans la boîte du nœud ?

7. *Glossanet : est-ce que l'installation d'un compte de Glossanet ou le chargement des graphes ont posé des problèmes ? Comment jugez-vous les résultats fournis par Glossanet ?*

Au début, beaucoup d'étudiants n'ont pas reçu d'occurrences bien qu'ils aient chargé des graphes très simples et généraux. Peut-être y a-t-il de temps en temps des problèmes techniques pour la consultation des journaux. Mais il semble que les possibilités de recherche de Glossanet sont beaucoup plus limitées que celles d'INTEX. Glossanet manque d'un manuel qui explique ces limitations de façon systématique.

8. *Ressources lexicales : comprenez-vous les structures des entrées ? Si vous travaillez avec une autre langue que le français : comment jugez-vous les différences au niveau des codes flexionnels ?*

Les codes italiens sont presque tous auto-explicatifs<sup>1</sup>. C'est pour cela que beaucoup d'étudiants ont préféré l'étiquetage italien à l'étiquetage français. Cependant un système de codes homogène au moins pour toutes les langues romanes rendrait les recherches comparatives plus faciles.

9. *Écrivant des dictionnaires supplémentaires, avez-vous eu des difficultés techniques?*

Il faut se rappeler que les dictionnaires doivent avoir l'extension .dic. Beaucoup d'étudiants ont eu des difficultés à retrouver leurs dictionnaires parce qu'ils les ont enregistrés comme fichiers .txt selon le choix par défaut de l'éditeur de texte standard de Windows.

10. *Utilisations des ressources lexicales : est-ce que la recherche de codes sémantiques ou flexionnels pose des problèmes?*

Les étudiants n'étaient pas très satisfaits quant aux résultats des recherches de codes sémantiques. Les résultats d'une recherche + <+hum> ou <-hum> étaient sans valeurs à cause du grand nombre des ambiguïtés. C'est pour cela qu'il me semble utile de préciser les informations sémantiques selon les suggestions de Xavier Blanco (1999).

11. *Levée d'ambiguïtés par grammaires locales : application des ressources, construction d'un simple transducteur, compilation, étiquetage. Est-ce qu'il y a des problèmes techniques?*

Les transducteurs italiens de Simona Vietri utilisent le code « V:U » qui ne correspond pas au DELAF italien.

Une fois le texte étiqueté, il est impossible d'annuler ce processus. Il faut recharger le texte en Windows-Ansi.

12. *Comment jugez-vous vos connaissances après les six premières unités d'enseignement? Avez-vous des suggestions didactiques?*

Les étudiants moins expérimentés en informatique ont désiré encore des tutorats. Sans l'aide d'un utilisateur avancé, ils se trouvaient par-

---

<sup>1</sup> Cela semble aussi nécessaire, puisque les explications des catégories que donnent les auteurs du DELAF italien sont incomplètes et incorrectes à la fois. Ils y parlent par exemple d'une catégorie DET qui dans le DELAF italien se trouve remplacée par ART.

fois incapables d'exécuter devant l'ordinateur les opérations expliquées en classe. Ceux qui ont déjà présenté un sujet en classe et avec lesquels j'ai fait des tutorats à l'ordinateur n'ont plus eu peur d'éprouver les fonctions du programme.

13. *Voulez-vous vous servir d'INTEX dans le futur pour rechercher des informations ou analyser des textes littéraires ?*

Beaucoup d'étudiants ne se sentent pas encore en situation de profiter du programme. Quelques-uns ont l'intention de faire des recherches d'information complexes pour des mémoires de littérature ou de sciences humaines qu'ils préparent.

Les étudiants de linguistique informatique sont tous désireux de continuer à travailler avec INTEX. Nous avons déjà décidé de programmer ensemble quelques outils pour la génération automatique des ressources lexicales catalanes.

## 5. Conclusion

En résumé, le plus positif de cette expérience didactique est le grand intérêt du côté des étudiants. Découragés par les cours de linguistique abstraits qui ne visent à aucune application, ils se sentent motivés par la possibilité de prouver leurs connaissances en faisant des exercices qui peuvent même avoir des résultats utiles comme des dictionnaires supplémentaires. Mais, après six unités d'enseignement, beaucoup d'étudiants n'étaient pas encore en situation de dominer le programme et encore moins de faire un travail linguistique sérieux à l'aide d'INTEX. À la fin du cours, les résultats étaient déjà plus positifs. Mais je doute qu'un semestre de 13 ou 14 unités d'enseignement soit suffisant pour apprendre à dominer le programme. Il faudrait continuer le travail dans un cours suivant, ce qui n'est pas prévu par notre plan d'études. Il ne reste qu'à espérer que l'un ou l'autre perfectionnera ses connaissances de façon autodidacte et se servira d'INTEX pour faire un mémoire de maîtrise. Si cela se réalisait, mon initiative ne serait pas en vain et toute la communauté INTEX pourrait en profiter.

## Bibliographie

- BLANCO (Xavier) : 1998, « Dictionarios bilingües en soporte electrónico. Problemas de desambiguación », in *Actes del III Congrès Internacional sobre Traducció*, mars 1996 (Barcelone : Universitat Autònoma).
- BLANCO (Xavier), BUVET (Pierre-André) : 1998, « El tratamiento de los determinantes en procesamiento automático del lenguaje natural », *Terminologie et Traduction*, 3 (Luxembourg : Commission Européenne).
- BLANCO (Xavier) : 1999, « Cómo leer un diccionario. Algunos aportes de la lingüística informática », in *Actas del Coloquio Artibytes* (Santarem).
- DISTER (Anne) : 1999a, « La levée d'ambiguïtés morphologiques par grammaires locales. Le système Intex », in *Actes des XIIIes Journées de Linguistique, (25 et 26 mars 1999)*, (Québec : Faculté des Lettres, Université Laval), p 57–66
- DISTER (Anne) : 2000, « Construire des grammaires de levée d'ambiguïtés pour Intex », in *FAIRON (Cédric) : Analyse syntaxique et lexicale. Le système INTEX, Linguisticae Investigationes* (Amsterdam/Philadelphie : John Benjamins), p. 231–245.
- GROSS (Maurice) : 1975, *Méthode en syntaxe* (Paris : Hermann).
- GROSS (Maurice) : 1994 « Constructing Lexicon-Grammars » in Atkins (B.T.S.) and Zampolli (A.) eds. : *Computational Approaches to the Lexicon*, (Oxford : University Press), p. 213–263.
- SILBERZTEIN (Max) : 1999-2000, *Intex*. <http://www.ladl.jussieu.fr/INTEX/intex.html>.
- SUBIRATS-RÜGGERBERG (Carlos) : 1987, *Sentential complementation in Spanish. A lexico-grammatical study of three classes of verb* (Amsterdam/Philadelphie : John Benjamins).