

# L'apport du TAL dans l'enseignement et l'apprentissage des langues

Victoria Bobicev, Liviu Carcea  
Université Technique de Moldova

Technology on its own is not the panacea; it's how you use technology that counts.

La technologie par elle-même n'est pas la panacée;  
c'est la façon dont vous utilisez la technologie qui compte.

*Graham Davies*

Résumé : Cet article présente un concordancier online<sup>1</sup> qui a été créé dans le cadre du projet : Pour le renforcement de la recherche en vue du développement des formations francophones niveau master - La Plateforme du français technique et des technologies de la langue<sup>2</sup>. Aussi il contient des exemples d'activités possibles pour les enseignants et les étudiants qui peuvent être utilisés dans le processus d'apprentissage de la langue.

A part des outils spécialement élaborés pour l'apprentissage et basés sur les théories classiques de l'enseignement des langues, les technologies de l'information offrent d'autres ressources et outils fondamentalement nouveaux qui peuvent être utilisés dans le processus didactique.

Les corpus de textes sont les ressources les plus importantes qui ne sont pas assez valorisées et ne sont pas encore pleinement utilisées non plus.

Un problème d'utilisation efficace des corpus est leur volume. Les philologues ne sont pas habitués à ces sources d'informations. Pour une utilisation efficace des corpus dans la linguistique, la philologie, la lexicographie, l'apprentissage et l'enseignement des langues, des outils et des méthodes spécifiques sont nécessaires. L'apport du TAL est notamment l'élaboration des ces outils et méthodes.

Le concordancier représente un de ces outils. Il permet de supprimer efficacement des grandes quantités de texte et de visualiser les éléments nécessaires: mots-clés, phrases ou constructions grammaticales.

## Introduction

L'émergence et la propagation des technologies informatiques ont ouvert de nouvelles possibilités qui ne sont pas encore pleinement perçues dans le domaine de l'éducation. Les termes anglais e-learning ou computer-aided learning, traduits comme apprentissage en ligne ou enseignement assisté par ordinateur (EAO) reflètent cet aspect. EAO représente l'utilisation des médias électroniques et technologies de l'information et de la communication (TIC) dans l'éducation. En général, il comprend:

- le développement et la promotion de technologies novatrices d'enseignement, et leur transfert vers les enseignants;
- la possibilité de développer les ressources Web éducatifs;
- la possibilité d'interaction distance, aide, conseils, évaluations à distance;

---

<sup>1</sup> <http://lilu.fcim.utm.md/concordancier/menu.html>

<sup>2</sup> <http://www.beco.pub.ro/>

- la possibilité, à tout moment et en tout lieu d'obtenir des connaissances modernes du monde entier;
- travailler de façon autonome avec des matériaux en format électronique, en utilisant un ordinateur personnel ou un dispositif électronique mobile;
- création d'une communauté d'utilisateurs distribués (réseaux sociaux), ce qui conduit les activités de formation virtuelles globales.

CALL (Computer Assisted Language Learning) est un des domaines d'EAO appliqué à l'enseignement des langues qui est en pleine évolution et qui comprend une variété de méthodes, techniques, outils et ressources.

Beatty (2003) a proposé la caractérisation suivante: «une définition de l'EAO qui accueille son nature changeante est 'un processus dans lequel l'apprenant utilise un ordinateur et, par conséquent, améliore sa langue'».

L'organisation principale qui fournit des informations et des conseils sur tous les aspects de l'utilisation de la technologie pour l'apprentissage des langues est EUROCALL. EUROCALL a été mis en place comme une association professionnelle officielle en 1993, mais ses origines remontent bien plus loin - à un cours intitulé «Ordinateurs pour l'éducation et de la recherche de la langue anglaise", qui a été organisé à l'Université de Lancaster en 1984. Trois axes de développement ont été identifiés lors de la réunion lorsque le nom EUROCALL a été inventé (Davies 2000):

- ✓ L'utilisation des ordinateurs dans la salle de classe.
- ✓ Formation des enseignants.
- ✓ L'évaluation et le développement de logiciels.

#### Le traitement automatique de langage naturel (TALN)

Le traitement automatique du langage naturel ou de la langue naturelle (abr. TALN) est une discipline à la frontière de la linguistique, de l'informatique et de l'intelligence artificielle, qui concerne l'application de programmes et techniques informatiques à tous les aspects du langage humain (Charniak 1984). TALN est souvent confondu avec la linguistique informatique mais il ya une distinction claire. La linguistique informatique est un champ interdisciplinaire basé sur une modélisation symbolique (à base de règles) ou mathématique (statistique) du langage naturel établie dans une perspective informatique. L'approche de la linguistique informatique est d'abord théorique, descriptive. Son objectif principal est la création de modèles de langue. Elle propose des théories de la production ou de l'interprétation des langues comme par exemple, Saussure (système de signes, linguistique structurale), Chomsky (système de règles, linguistique générative) ou Langaker (1987) (grammaire cognitive). L'approche du traitement automatique du langage naturel est absolument pratique. Il traite de la langue ayant des tâches pratiques bien définies.

Les premiers travaux de recherche sur les langues utilisant les moyens informatiques ont eu lieu dès 1950, en vue de la traduction automatique, financés par l'armée américaine. Dans les années 1960, les chercheurs en intelligence artificielle pensaient que la maîtrise des langues par les machines était un objectif à court terme. Faute de résultats suffisants, en 1960, les crédits ont été coupés. Il a fallu repartir sur de nouvelles bases dans les années 1970, en formalisant la grammaire. Les travaux de Noam Chomsky (Chomsky 1969) ont fait l'hypothèse d'une similitude entre les langues naturelles et les langages informatiques. Mais les grammaires formelles ont buté sur de nombreuses difficultés pour le langage humain. Les modèles basés sur la connaissance ont pris le relais dans les années 1980 (Sabah 1989). Ils supposent que la clé de la compréhension des langues est dans les connaissances partagées. Là aussi, les résultats furent beaucoup moins rapides qu'on ne le croyait. Après des espoirs déçus de maîtrise des langues

par les ordinateurs, les informaticiens ont pris la mesure de la difficulté des problèmes et ils sont devenus modestes.

Toutefois, la croissance continue de la circulation des documents, exigeant l'étiquetage, le traitement, l'analyse, la traduction a exigé solution immédiate du problème de l'automatisation de ces processus. La croissance rapide de l'Internet, qui comprend d'innombrables textes a également provoqué un besoin urgent d'outils qui pourraient aider les gens à trouver et à traiter l'information dont ils ont besoin.

Dans ces conditions, le TALN entre en scène et essaie de résoudre ses tâches par tous les moyens, sans entrer dans les complexités du fonctionnement de la langue. En général, ce sont des méthodes statistiques. Ces méthodes permettent d'obtenir des résultats satisfaisants. En outre, ils peuvent être facilement mises en œuvre par le logiciel. La principale condition de leur succès est la grande quantité d'information nécessaire pour la création d'un modèle statistique adéquate. C'est exactement ce qui est disponible dans le monde moderne.

### Les corpus

Des grandes quantités de texte sont rassemblées dans le corpus et sont utilisées à diverses fins. Les corpus sont à la fois la matière première de l'étude des langues et des textes et le moyen de tester les modèles qui sont proposés. Un corpus est une grande quantité de matériau linguistique attesté: livres, journaux, devoirs d'étudiants, enregistrements de radio, de télévision, de discussions, etc. Les corpus numérisés sont les plus utiles pour les linguistes, les littéraires, comme pour les chercheurs en TAL. Les sources de corpus écrits sont nombreuses, tant par le Web que du fait que l'édition est numérisée depuis les années 1980.

Il y avait plusieurs ouvrages sur la linguistique de corpus avant l'ère de l'ordinateur, mais il était difficile de recueillir des preuves à partir du corpus lorsque le corpus a été imprimé sur le papier.

Chomsky cependant a critiqué la façon du corpus linguistique de l'étudier la langue. Chomsky a suggéré que le corpus ne pourra jamais être un outil utile pour le linguiste, comme le linguiste doit chercher à modéliser la compétence linguistique plutôt que la performance.

Pourtant, ce n'était pas le seul reproche que Chomsky avait pour le début de l'approche linguistique de corpus. Le caractère non-fini de la langue était l'argument principal. Tout le travail de début de la linguistique de corpus a été soutenue par deux hypothèses fondamentales, encore imparfaites: les phrases d'une langue naturelle sont limitées et peuvent être collectées et énumérées.

Mais le nombre de phrases dans un langage naturel n'est pas seulement arbitrairement grand - il est potentiellement infini. Chomsky a estimé que la seule façon de rendre compte d'une grammaire d'une langue est par la description de ses règles - pas par l'énumération de ses phrases.

Cependant Leech (1992) soutient que le corpus est une méthode plus puissante du point de vue de la méthode scientifique, car il est ouvert à une vérification objective des résultats. Dans le même temps, Quirk (1960) a planifié et exécuté la construction de son Enquête ambitieuse de l'utilisation d'anglais (Survey of English Usage - SEU), qui il a débuté en 1961. La même année, Francis et Kucera, a commencé à travailler sur le désormais célèbre corpus Brown, un travail qui consistait à prendre près de deux décennies. Peu à peu, le corpus est devenu lisible par l'ordinateur (machine-readable) ce qui simplifiait considérablement sa manipulation et de l'utilisation. Linguistique de corpus moderne définit un corpus comme un ensemble de textes qui ont quatre caractéristiques principales:

- il doit être en un forme lisible à la machine (format électronique);

- il doit être représentatif pour le domaine pour lequel il a été créé;
- il doit être finit;
- tous les textes doivent être codées et annotés de la même manière.

La plupart des corpus remplit ces conditions, mais certains sont en constante croissance. Par exemple, certains corpus collectés des éditions périodiques électroniques comme des journaux ou des quotidiens.

Différents types de corpus sont créées à des fins diverses. Les corpus représentatifs ont tendance à représenter l'ensemble de la langue dans toute sa diversité. Il y a des corpus nationaux pour certaines langues, comme British National Corpus<sup>3</sup> or Russian National Corpus<sup>4</sup>

Les corpus spéciaux représentent des phénomènes spécifiques de la langue. Par exemple, le corpus des créations de l'Alexandre Dumas ou un corpus de correspondance ouvrière.

Des diverses méthodes de création de corpus ont été développées, mais c'est encore un énorme travail de recueillir et organiser un bon corpus qualitatif.

Les plus précieux sont les corpus marqués or annotés mais le marquage qualitatif du corps nécessite un travail manuel de professionnels qualifiés. L'annotation de corpus représente le fait d'ajouter des informations linguistiques interprétatives à un corpus. Par exemple, une annotation de type commun est l'ajout d'étiquettes indiquant la classe morphologique à laquelle les mots d'un texte appartiennent.

Ceci est ce qu'on appelle marquage des catégories grammaticales (Part of Speech - POS tagging), et peut être utile, par exemple, pour la recherche de toutes les formes d'un mot. Les autres types d'annotations sont: l'annotation lexicale, l'annotation phonétique, l'annotation sémantique, l'annotation de sentiments, etc.

Les exemples des corpus écrits annotés et en français sont PAROLE<sup>5</sup> et MULTEXT JOC<sup>6</sup> (Véronis et Khouri, 1995), le French Treebank (P7T) (Abeillé et al., 2003), la base FREEBANK (Salmon-Alt et al., 2004) et le récent corpus Sequoia<sup>7</sup> (Candito et Seddah, 2012). Excepté la FREEBANK, ces corpus sont toujours accessibles aujourd'hui via un guichet sur le Web. Le P7T<sup>8</sup> est probablement le corpus annoté le plus utilisé et le plus référencé, et ce essentiellement pour trois raisons : il est libre d'usage pour des activités de recherche, il bénéficie d'une analyse multi-niveaux (de la structure textuelle à la structure syntaxique en passant par des annotations en morphologie) et il compte près du double de mots annotés que tous les autres corpus disponibles réunis. En pratique ce corpus se compose d'articles journalistiques issus du journal Le Monde écrits dans les années 90, soit plus de 500 000 mots annotés.

Les corpus sont utilisé dans la lexicographie par se que les exemples peuvent être facilement extraites et organisées pour l'analyse en utilisant une logicielle. Corpus sont les outils bons pour la recherche syntaxique en raison de la quantification représentant de toute une variété de langue. Le principal avantage de corpus est le fait que les données sont en grande partie naturalistes, sans surveillance et le produit de contextes sociaux réels. Ainsi, le corpus constitue l'une des sources les plus fiables de données d'origine naturelle qui peuvent être examinées.

Les corpus de textes sont les ressources les plus importantes qui ne sont pas assez valorisées et ne sont pas encore pleinement utilisées dans l'apprentissage et l'enseignement des langues.

---

<sup>3</sup> <http://www.natcorp.ox.ac.uk/>

<sup>4</sup> <http://www.ruscorpora.ru/en/>

<sup>5</sup> [http://catalog.elra.info/product\\_info.php?products\\_id=565](http://catalog.elra.info/product_info.php?products_id=565)

<sup>6</sup> [http://catalog.elra.info/product\\_info.php?products\\_id=534](http://catalog.elra.info/product_info.php?products_id=534)

<sup>7</sup> <https://www.rocq.inria.fr/alpage-wiki/tiki-index.php?page=CorpusSequoia>

<sup>8</sup> <http://www.llf.cnrs.fr/Gens/Abeille/French-Treebank-fr.php>

## Le concordancier

Un problème d'utilisation efficace des corpus est leur volume. Les philologues ne sont pas habitués à ces sources d'informations. Pour une utilisation efficace des corpus dans la linguistique, la philologie, la lexicographie, l'apprentissage et l'enseignement des langues, des outils et des méthodes spécifiques sont nécessaires. Le concordancier représente un de ces outils. Il permet de supprimer efficacement des grandes quantités de texte et de visualiser les éléments nécessaires: mots-clés, phrases ou constructions grammaticales.

Le terme 'concordance' est venu d'anglais ou cela signifie une liste alphabétique des mots dans un livre ou une série de livres qui dit aussi où chaque mot peut être trouvé et souvent comment il est utilisé. Pour produire des concordances manuellement et sans l'aide d'un ordinateur était une tâche énorme et difficile. Aujourd'hui, un ordinateur peut le faire en quelques minutes.

Une première référence de l'utilisation de concordanciers électroniques dans l'enseignement de la langue peut être trouvée dans Higgins & Johns (1984), et les nombreux exemples de leur utilisation pratique dans la salle de classe sont décrits par Lamy et Klarskov Mortensen (2010).

C'était Tim Johns (1991), qui a rehaussé le profil de l'utilisation du concordancier dans la classe de langue avec son concept d'apprentissage piloté par les données (Data Driven Learning - DDL). DDL encourage les élèves à travailler sur leurs propres règles sur le sens de mots et de leur utilisation par l'utilisation d'un concordancier pour trouver des exemples dans un corpus de textes authentiques. Il est également possible pour l'enseignant d'utiliser un concordancier pour trouver des exemples d'utilisation authentique de démontrer un point de grammaire ou collocations typiques, et de générer des exercices basés sur des exemples trouvés.

Dans DDL le processus d'apprentissage ne repose plus uniquement sur l'initiative de l'enseignant, sa / son choix de sujets et de matériaux et l'enseignement explicite des règles, mais de sa propre découverte de l'apprenant des règles, des principes et des modèles d'utilisation de la langue étrangère. En d'autres termes, l'apprentissage est entraîné par les données linguistiques authentiques.

Le principal avantage du concordancier est résumée par Tim Johns, en référence à une phrase qu'il fréquemment utilisé: "la compagnie que les mots maintenir" ou "vous devez connaître le mot par la compagnie qu'il maintenir". En d'autres termes, il fait savoir non seulement sens d'un mot, mais son utilisation dans le texte avec les autres mots.

Ainsi, une concordance est un mot (appelé mot-clé), tiré d'une pièce de langue authentique (corpus), représenté avec des parties des contextes dans lesquels ils se produisent. Ceci est également connu en tant que mot clé en concordance de contexte ou un KWICconcordance. Un exemple est présenté en figure 1.

```
(A1 clocher 1:I)      penduës. Haut clocher. gros clocher. clocher pointu.
(A1 gagner 1:I)      dans les fermes. gagner gros. un bon ouvrier peut
(A1 gras 1:I)        beaucoup de graisse. Il est gros et gras. il est gras par
(A1 vin 1:I)         vin tiré en bouteilles. gros vin. petit vin. vin
(A1 vin 1:I)         en perce. vendre du vin en gros et en détail. éclaircir du
(A1 vin 18:I)        tant pour le pot de vin. un gros pot de vin. le pot de vin
(A2 clocher1 1:I)    Eglise. Haut clocher. gros clocher. clocher pointu.
```

Figure 1. Les concordances pour le mot gros. Source:

<http://homes.chass.utoronto.ca/~wulfric/academie/cachelst.htm#gros>.

Le concordancier créé dans le cadre de ce projet<sup>9</sup>

Dans ce projet le concordancier est créé comme un ensemble d'outils pour l'enseignement et l'apprentissage du français de spécialité. Il se compose de plusieurs outils qui sont présentés ci-dessous.

Liste de mots. Le premier outil crée la liste de mots à partir du texte introduit dans la fenêtre d'entrée du concordancier. Cet outil compte tous les mots dans le texte et les présente dans une liste ordonnée. Cela montre les mots qui sont les plus fréquents dans le texte et les mots qui figuraient dans le texte qu'une seule fois.

Liste de mots clés. Cet outil select les mots qui sont les plus fréquents dans le texte et les présente dans une liste ordonnée. Il utilise l'algorithme de traitement de texte le plus sophistiqué parmi les outils dans cet ensemble. Il supprime les mots les plus fréquents qui sont toujours des prépositions, des articles, pronoms, etc. et qui sont communes pour tous les textes. Il supprime aussi les mots les moins fréquentes qui ne sont pas caractéristiques du texte. Ensuite, il compare la fréquence des mots du texte avec leur fréquence plus générale sur la base du corpus de référence créé à partir des textes de différents domaines. Les mots qui sont plus fréquentes dans le texte que dans le corpus de référence sont considérés comme spécifiques pour ce texte.

Clusters (n-grammes). Cet outil montre des clusters basés sur la condition de recherche. Il balaye l'ensemble du texte pour trouver des clusters de longueur 'N' (par exemple 1 mot, 2 mots, etc. ...). Cela souvent trouve des expressions courantes dans un texte.

Concordances avec un mot et avec un fragment. Cet outil affiche les résultats de recherche dans un format 'KWIC' (mot-clé dans le contexte). Cela vous permet de voir comment les mots et les phrases sont couramment utilisés dans le texte. L'outil permet utilisateur d'entrer plusieurs formes du mot dans le formulaire de saisie et recherche toutes ces formes. Les concordances avec un fragment peuvent être utilisées pour la recherche de plusieurs mots ayant la même racine (par exemple, explication, explicatif, etc.). Dans ce cas, l'utilisateur peut entrer seulement une partie commune de ces mots (pour l'exemple précédent, le fragment 'explicat') et l'outil va rechercher tous les mots commençant par ce fragment.

Co-occurrences avec un mot et avec un fragment. Cet outil montre les co-occurents d'un mot de recherche pour étudier les tendances non-séquentielles de la langue.

Tous ces outils sont statistiques et donnent de bons résultats en travaillant avec des grands volumes de texte. Il y a cependant un problème avec le concordancier en ligne: le volume du texte transmis. La connexion Internet lente peut créer des problèmes pour les grands textes.

Le concordancier peut être utilisé pour l'enseignement et l'apprentissage des langues de manières asses variées. Certaines d'entre elles sont citées ci-dessous:

- l'enseignant peut utiliser un concordancier pour trouver des exemples d'utilisation authentique pour démontrer les caractéristiques du vocabulaire, des collocations typiques, un point de grammaire ou même la structure d'un texte ;
- l'enseignant peut générer des exercices basés sur des exemples tirés d'une variété de corpus, par exemple des exercices et des tests à combler les lacunes ;
- les apprenants peuvent travailler sur des règles de grammaire ou d'usage et des caractéristiques lexicales pour eux-mêmes par la recherche de mots clés dans leur contexte. Selon leur niveau, ils peuvent être invités à remettre en question certaines des règles basées sur l'observation des modèles dans la langue authentique ;

---

<sup>9</sup> <http://www.beco.pub.ro/>

- les apprenants peuvent être plus actifs dans leur apprentissage du vocabulaire: en fonction de leur niveau, ils peuvent être invités à découvrir des nouvelles significations, d'observer les collocations habituelles, de relier des mots à la syntaxe, ou de critiquer les entrées du dictionnaire ;
- Les apprenants peuvent être invités à réfléchir sur l'utilisation de la langue en général, sur la base de leurs propres explorations d'un corpus des textes, se transformant ainsi en chercheurs.

Il y a des nombreuses activités qui sont possibles en classe mais doivent être préparées par l'enseignant avec le concordancier et les textes respectifs :

- Comme le suggère le site Cobuild Web, exécutez un concordancier sur un mot particulier et puis supprimez le mot clé. Les élèves doivent ensuite essayer de deviner le mot clé du contexte. Cela encouragera les élèves à déterminer le mot de contexte et non pas en utilisant le dictionnaire.
- Créer une liste KWIC pour un mot particulier et demander aux élèves de déterminer la préposition qui précède ou qui suit ce mot clé (Tribble et Jones, 1997). Une recherche de concordance pourrait également produire des listes de contextes pour différents temps de verbes qui pourraient être comparés dans leurs contextes (Ketteman, 1998).
- Créer une liste KWIC d'un mot qui a plusieurs sens et demander aux élèves de déterminer que plus d'un sens existe et que ces significations sont. Cela ne fonctionnera pas très bien avec les mots qui apparaissent dans les expressions idiomatiques et littérales.
- Créer des listes KWIC pour deux ou plusieurs mots avec un sens similaire; par exemple, créer des listes pour dire et parler et demander aux élèves de discuter en petits groupes les caractéristiques uniques de chaque terme (Tribble et Jones, 1997). En utilisant les mêmes listes, supprimer les mots-clés et demander aux élèves de déterminer à partir du contexte quels mots seraient les mieux places dans la phrase.

Suivant sont certain exemples d'activités possibles avec un concordancier en classe<sup>10</sup>.

Activité 1: Devinez le mot mystère.

But de l'activité: Familiariser les étudiants avec l'apparence physique d'une concordance KWIC et de l'importance de contexte gauche et de contexte droit lorsque vous travaillez avec des mots-clés. C'est pour être utilisé avec les gens qui sont débutants complets à concordancier.

Feuille de travail: Lire la grille ci-dessous, où le mot non-sens "gloup" a été inscrit à la place d'un vrai mot. Votre travail consiste à décider avec votre groupe ce qui est vrai mystère mot. Lorsque vous avez fait votre choix, discuter avec votre groupe ce que les réponses aux trois questions ci-dessous la grille devraient être.

1. pport critique sur certaines utilisations abusives de la	<b>gloup</b>	est devenu un geste banal plus qu'une décision.
2. que pour beaucoup d'entre nous le fait d'allumer une	<b>gloup</b>	.
3. laquelle on est pris pour gens qui "s'abrutissent à la	<b>gloup</b>	", dans une proportion croissante depuis 1896
4. Tous les grands moments de	<b>gloup</b>	superposent un message recherché et un messa
5. sieurs postes et l'augmentation du temps de diffusion (	<b>gloup</b>	du matin et de la nuit).
6. dailleurs 21% des Français reconnaissent regarder la	<b>gloup</b>	Même si le programme les ennue. 34% seulem
7. rmettent une plus grande maîtrise individuelle de la	<b>gloup</b>	. Les comportements des téléspectateurs en ont é
8. ux Pays-Bas que l'on regarde le moins longtemps la	<b>gloup</b>	: 89 minutes par jour, contre 228 en Grande-Bret
9. publications, diffusent des émissions de radio ou de	<b>gloup</b>	. Et découvrent les vertus des communications

<sup>10</sup> [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm)

i. Lors de la lecture de la grille, à quel point (c'est à dire à qui le numéro de ligne) avez-vous commencé à soupçonner la vérité?

ii. Quel est le mot ou les mots que vous a donné l'idée la plus utile?

iii. Les lignes 1 et 2 pourraient faire penser à un mot mystère très différent. Qu'est-ce que c'est et quels indices pourraient vous faire croire cela?

Des idées pour la création de documents à distribuer: Essayez des exercices semblables avec des mots qui ont un sens double ou des significations différentes variétés de la langue. Avec ces mots, contextes contribueront fortement à la conjecture exigée. Par exemple, dépanneur français (et son sens français québécois).

Activité 2: Donc on peut dire que...

Objectifs de l'activité: Pour faire prendre conscience aux élèves des différences stylistiques entre les différentes positions de "Donc" dans les phrases. La mise en place de "Donc" en français dans les textes écrits formels par rapport aux informels.

Feuille de travail: En français, vous pouvez placer un mot comme «Donc» au début d'une phrase (par exemple, " Donc on peut dire que ..."). Mais vous le trouverez souvent entre le verbe principal et ce qui suit le verbe principal (par exemple, " On peut donc dire que ..."). Est l'un meilleur que l'autre? Où devriez-vous mettre "Donc"? Pour le savoir, regardez les deux listes ci-dessous, toutes les deux de Québec.

#### Liste A

Il devient	<b>donc</b>	difficile de proposer un plan de gestion de ces troupeaux.
Elle permet	<b>donc</b>	la circulation des avions de grandes dimensions en cas de besoin.
Ce droit exclusif n'atténue	<b>donc</b>	pas du tout les droits de ces derniers puisque ils ont accès à toutes les
Les autochtones ont	<b>donc</b>	priorité quant à la récolte.
Il serait	<b>donc</b>	intéressant de comparer des données plus récentes afin de
Ces chiffres révèlent	<b>donc</b>	une tendance à la baisse entre 1976 et 1980 mais sûrement aussi
Les données furent	<b>donc</b>	suggéré de répartir les territoires de chasse selon des zones “
Ils ont	<b>donc</b>	peur de ne pas avoir suffisamment obtenu de terrains pour permettre
Il faut	<b>donc</b>	déterminer les espèces touchées et leur importance relative
C'est	<b>donc</b>	l'épaisseur calculée en fonction du gel qui prime
Ils peuvent	<b>donc</b>	théoriquement rencontrer la demande de transport pour les années

#### Liste B

Aussi le Père Joseph. Guay aussi ce sont tous des invalides.	<b>Donc</b>	je crois qu'il serait à propos de leur donner quelque chose
on sait que vous en envoyez et on n'en a pas alors tâchez	<b>donc</b>	s'il vous plaît d'être assez bon de nous envoyer à leur nom
le grand besoin avec une famille de 9 enfants ça fait bien dur.	<b>Donc</b>	Je compte sur votre secours afin de pouvoir passer l'hiver
mes enfants à la maison et les priver de l'instruction. Je compte	<b>donc</b>	sur votre grande générosité pour nous tirer d'embaras.
seront payés et ceux qui ne travailleront pas n'auront droit à rien	<b>donc</b>	il faudra qu'ils travaillent pour avoir de l'aide si non rien.
transport qui m'empêche de finir cette transaction. Je vous serais	<b>donc</b>	bien obligé de me dire quels moyens je pourrais prendre pour obtenir un prix réduit
Notre Dame Du Lac Co Témiscouata. Cher Mr Je viens	<b>donc</b>	vous écrire pour vous demander si je peux avoir de l'octroi pour
le bâtir tout de suite pourvu que je sois certain d'avoir ma prime	<b>Donc</b>	je me fie entièrement à vous pour régler cette affaire-là et je
va être fini mais cette année la Compagnie Fraser l'achète.	<b>Donc</b>	espérant recevoir une bonne réponse de vous le plus tôt
mais je n'ai pas d'argent et ils demandent déjà un bon prix	<b>donc</b>	s'il vous plaît enseignez-moi les moyens à prendre pour le
Bien cordialement. Soyez	<b>donc</b>	assez bon de me dire s'il y a encore de bons lots à prendre
Je connais la terre étant fils de cultivateur.	<b>Donc</b>	Monsieur le Curé je sais que si vous le voulez je pourrais aller
bien si vous pouviez venir inspecter ce chemin	<b>donc</b>	je veux pas vous ennuyer avec cela



### Activité 3: Le roman s'agit d'un amour malheureux

But de l'activité: à amener les élèves de dériver (par induction), la règle selon laquelle «s'agit» ne prend jamais un sujet autre que le pronom «il» impersonnel.

Feuille de travail: "Le roman (ou "le poème" ou "la pièce") s'agit d'un amour Malheureux ... " Quand les professeurs de français lisent ce genre de phrase dans les essais des étudiants, ils sont susceptibles de fouetter leurs stylos rouges et marquer les deux premiers mots. Pourquoi? Pour répondre à cela, lisez ce qui suit. Puis, avec votre groupe, d'expliquez la réaction des enseignants, et décidez quant aux précautions que vous devez toujours prendre lors de l'utilisation du verbe «s'agit».

1. e de s'adapter au monde contemporain. Il	<b>s'agit</b>	de savoir si l'on table, oui ou non
2. es, de Beurs ni de Blacks (hélas). Il ne	<b>s'agit</b>	pas d'une bande dessinée mais
3. is gaulois. On en use à présent quand il	<b>s'agit</b>	d'évoquer les solutions apportées
4. oins comme ami, simplement parce qu'il	<b>s'agit</b>	de quelqu'un de différent
5. une femme. Et réciproquement. Quand il	<b>s'agit</b>	de cette différence-là il y a
6. olution partiellement dans le prototype. Il	<b>s'agit</b>	de définir les autorisations
7. uation de l'élève faite par le système. Il	<b>s'agit</b>	donc de recueillir ces informa
8. le "welfare State", l'état-providence, il	<b>s'agit</b>	d'une approche globale et
9. péen Hans van den Broeck, selon lequel il	<b>s'agit</b>	d'"aider" la Grèce à répondre
10. fres et François Maspéro, montrent qu'il	<b>s'agit</b>	bien d'univers distincts, qui
11. meront peut-être certains, pensant qu'il	<b>s'agit</b>	une fois de plus d'un ouvrage
12. virtuel des différends (chapitre VI). Il	<b>s'agit</b>	là d'un ensemble de mesures lo
13. Aucun établissement ne part de rien. Il	<b>s'agit</b>	donc de faire l'inventaire de
14. cours multimédia d'anglais (HELLO). Il	<b>s'agit</b>	d'une réalisation de la BBC en
15. ont représentés. Statutairement, il	<b>s'agit</b>	d'une simple association (type
16. Cela prend énormément de temps. Il ne	<b>s'agit</b>	pas seulement de les lire. Il

### Conclusion

L'histoire de l'EAO suggère que l'ordinateur peut servir une variété d'utilisations pour l'enseignement des langues. Il peut être un tuteur qui propose des exercices de langue ou de la pratique des compétences; un stimulant pour la discussion et l'interaction; ou un outil pour l'écriture et la recherche. Avec l'avènement de l'Internet, il peut aussi être un moyen de communication mondial et une source de matériaux authentiques illimités.

Mais comme l'a souligné Garrett (1991), "l'utilisation de l'ordinateur ne constitue pas une méthode". Il s'agit plutôt d'un «milieu dans lequel une variété de méthodes, d'approches et de philosophies pédagogiques peut être mise en œuvre» (p. 75). L'efficacité de l'EAO ne peut pas résider dans le milieu lui-même, mais seulement dans la façon dont il est mis à profit.

Une des principales caractéristiques de l'apprentissage par ordinateur est l'individualisation du processus d'apprentissage. Cette fonction est utilisée dans les approches plus récentes à l'EAO qui favorisent une approche exploratoire centrée sur l'apprenant plutôt que d'une approche centrée sur l'enseignant. L'approche exploratoire est caractérisée par l'utilisation des programmes de concordance dans la classe de langues - une approche décrite comme Data Driven Learning (DDL). Cette approche est basée sur des textes réels organisés dans des corpus et étudiés au moyen d'un concordancier.

Le concordancier online<sup>11</sup> présenté dans cet article a été créé dans le cadre du projet : Pour le renforcement de la recherche en vue du développement des formations francophones niveau

<sup>11</sup> <http://lilu.fcim.utm.md/concordancier/menu.html>

master - La Plateforme du français technique et des technologies de la langue<sup>12</sup>. Aussi il contient des exemples d'activités possibles pour les enseignants et les étudiants qui peuvent être utilisés dans le processus d'apprentissage de la langue.

### Bibliographie

- ABEILLÉ, A., CLÉMENT, L. et TOUSSENEL, F., Building and using Parsed Corpora, chapitre Building a treebank for French. Language and Speech series, Kluwer, Dordrecht, 2003.
- AWT (Agence Wallonne des Télécommunications) Qu'est-ce que l'e-learning? [archive], version mise jour 2008-01-28, consultée 2012-07-07, 2008.
- BEATTY, K., Teaching and Researching Computer Assisted Language Learning, New York, Longman, 2003.
- CANDITO, E. et SEDDAH, D. Le corpus Sequoila: annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical, In 19e conférence sur le TALN, Grenoble, France, 2012.
- CHARNIAK, E., Introduction to artificial intelligence, page 2. Addison-Wesley, 1984.
- CHOMSKY, N., La Linguistique cartésienne suivi de La Nature formelle du langage, Éditions du Seuil, 1969.
- DAVIES, G., The History of EUROCALL: an article produced to celebrate the dawning of the new millennium, 2000 - regularly revised, <http://www.camsoftpartners.co.uk/EuroHist.htm>
- GARRETT, N., Technology in the service of language learning: Trends and issues. Modern Language Journal, 75(1), 74-101, 1991.
- HIGGINS, J. & JOHNS, T. Computers in language learning, London: Collins, 1984.
- KETTEMANN, B. On the use of concordancing in ELT. TELL & CALL, 1995.
- LAMY, M-N. & KLARSKOV MORTENSEN, H. J., Using concordance programs in the Modern Foreign Languages classroom. Module 2.4 in Davies G. (ed.) Information and Communications Technology for Language Teachers (ICT4LT), Slough, Thames Valley University, 2011 [Online]: [http://www.ict4lt.org/en/en\\_mod2-4.htm](http://www.ict4lt.org/en/en_mod2-4.htm)
- LANGACKER, R. W., Foundations of Cognitive Grammar, Volume 1, Theoretical Prerequisites. Stanford: Stanford University Press, 1987.
- LEECH, G., Corpora and theories of linguistic performance. In Svartvik J. (ed.) Directions in corpus linguistics, Berlin: Mouton de Gruyter, 1992.
- LEVY, M., CALL: context and conceptualisation, Oxford: Oxford University Press, 1997.
- QUIRK, R., Towards a description of English usage, Transactions of the Philological Society: 40-61, 1960.
- SABAH, G., L'intelligence artificielle et le langage, Représentations des connaissances, Processus de compréhension, Hermès, 1989.
- SALMON-ALT, S., BICK, E., ROMARY, L. et PIERREL, J.-M., La FReeBank : vers une base libre de corpus annotés. In Traitement Automatique des Langues Naturelles - TALN'04, Fès, Maroc, 2004.
- TRIBBLE, C., & JONES, G. Concordances in the classroom: A resource guide for teachers (2<sup>nd</sup> ed.). Houston TX: Athelstan. 1997.
- VÉRONIS, J. et KHOURI, L., Etiquetage grammatical multilingue : le projet MULTEXT. Traitement Automatique des Langues, 36(1/2):233-248, (1995).

---

<sup>12</sup> <http://www.beco.pub.ro/>