



HAL
open science

Élaboration d'une base de données d'exemples de structures comparatives : de la grille d'annotation au système d'interrogation

Frédérique Mélanie-Becquet, Catherine Fuchs

► To cite this version:

Frédérique Mélanie-Becquet, Catherine Fuchs. Élaboration d'une base de données d'exemples de structures comparatives : de la grille d'annotation au système d'interrogation. *Corpus, Bases, Corpus, Langage - UMR 7320*, 2011, pp.273-295. halshs-00661738

HAL Id: halshs-00661738

<https://halshs.archives-ouvertes.fr/halshs-00661738>

Submitted on 20 Jan 2012

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.

**Elaboration d'une base de données d'exemples de structures comparatives :
de la grille d'annotation au système d'interrogation**

Frédérique Mélanie-Becquet & Catherine Fuchs

LaTTiCe, ENS, Paris 3

DRAFT

Le présent article retrace les étapes qui ont présidé à l'élaboration d'une base de données d'exemples de structures à subordonnées¹ comparatives du français. Ce travail a été réalisé dans le cadre du projet « Structures à Subordonnées Comparatives du français » (SCF)². Piloté par le laboratoire LATTICE sous la direction de C. Fuchs, ce projet a réuni des membres de quatre laboratoires français : B. Combettes et A. Kuyumcuyan (ATILF, Nancy), C. Guimier (CRISCO, Caen), N. Fournier et M. Morinière (ICARE, Lyon), C. Fuchs, P. Le Goffic, F. Mélanie-Becquet et M. Piot (LATTICE, Paris). Pour un aperçu de ce projet, voir Fuchs (2008 : 1-5). Pour une présentation des travaux des membres du groupe autour de la base, nous renvoyons à l'ensemble du numéro XXXI :1 de la revue *Linguisticae Investigationes* (2008).

Le but de ce projet était de rassembler des exemples de structures à subordonnées comparatives quantitatives (*plus, moins, autant ... que, etc.*) et qualitatives (*comme, tel que, ainsi que, de même que, etc.*) figurant dans les principaux dictionnaires et grammaires du français (moderne et ancienne langue) ; puis de coder et d'annoter ces exemples, afin de constituer une base de données (sous Access) ; et enfin de mettre cette base à disposition de la communauté scientifique, par l'intermédiaire d'un système d'interrogation convivial.

Dans la première partie³, on décrira les étapes de constitution de la base. La seconde partie⁴ sera consacrée à l'élaboration du système d'interrogation.

1. La constitution de la base d'exemples

¹ Sont donc exclues les structures corrélatives (*tel père, tel fils ; plus on est de fous, plus on rit*) et autres comparatives sans subordonnée (*Eve est mignonne, mais Agathe est beaucoup plus jolie*).

² Ce projet a été soutenu par l'Institut de Linguistique Française du CNRS de 2005 à 2008.

³ La première partie a été rédigée par Catherine Fuchs.

⁴ La seconde partie a été rédigée par Frédérique Mélanie-Becquet.

L'architecture générale de la base a été conçue par P. Le Goffic ; sa mise en œuvre et sa gestion ont été assurées par F. Mélanie-Becquet. Le recueil et l'annotation des exemples ont été effectués par l'ensemble des membres du groupe. La constitution de la base a été réalisée en trois grandes étapes : tout d'abord, le recueil et l'enregistrement des exemples avec codage des sources ; puis l'analyse et le codage morpho-syntaxique des exemples ; enfin, l'analyse et le codage sémantique des exemples.

1.1. Recueil des exemples avec enregistrement et codage des sources

Ont été recueillis manuellement les exemples de structures à subordonnées comparatives figurant :

- à la rubrique « comparaison » dans 18 grammaires (10 portant sur le français moderne et 8 sur l'histoire du français ou sur des états de langue anciens)
- sous les entrées correspondant aux divers termes comparatifs dans 7 dictionnaires (3 de français moderne et 4 portant sur des états de langue anciens ⁵), dont on trouvera la liste en Annexe I.

Les exemples ainsi recueillis ont été enregistrés sous Access au sein d'une première table, dite « Table des exemples recueillis » comportant 3393 exemples. Au terme d'un travail de nettoyage visant à éliminer les doublons (exemples présents dans plusieurs sources) et les scories, ainsi que d'un toilettage typographique, a été obtenue la « Table des exemples », constituée des seuls exemples pertinents, uniques et référencés. Cette table contient 2476 exemples : 1574 issus de grammaires et 902 issus de dictionnaires (dont 2127 de français moderne et 349 d'états de langue anciens).

Le codage des sources a consisté à noter dans autant de champs distincts les types d'information suivants :

- n° de l'exemple
- origine de la (ou des) source(s)
- localisation de l'exemple dans la source

⁵ Les problèmes particuliers liés aux 8 éditions du *Dictionnaire de l'Académie Française* qui couvrent 4 siècles ont conduit à constituer une Table spécifique pour les exemples de ce dictionnaire.

- type d'exemple (forgé, littéraire, oral, locution, ...)
- en cas d'exemple littéraire : auteur, œuvre, localisation dans l'œuvre
- commentaires éventuels donnés par la source.

1.2. Analyse et codage morpho-syntaxique

L'objectif étant de caractériser chaque schéma à subordonnée comparative, il a d'abord fallu résoudre le problème des exemples comportant plusieurs subordonnées comparatives. Exemple : *Ahmès, l'esclave de la maison, Nubien plus noir **que** la marmite qu'il écumait gravement, était bon **comme** une nuit de sommeil.* (exemple tiré de *Thaïs* d'Anatole France et cité par Damourette et Pichon au § 722) ⁶.

Pour ce faire, à la « Table des exemples » a été associée une nouvelle table, dite « Table des exemples élémentaires » où chaque exemple n'est plus concerné que par une seule subordonnée comparative. Ainsi l'exemple ci-dessus, portant le n° 0064 dans la « Table des exemples », a-t-il été scindé en deux dans la « Table des exemples élémentaires » :

- exemple élémentaire n° 0064.01 : *Ahmès, [l'esclave de la maison,] Nubien plus noir que la marmite qu'il écumait gravement, [était bon comme une nuit de sommeil]*
- exemple élémentaire n° 0064.02 : *Ahmès, [l'esclave de la maison,] [Nubien plus noir que la marmite qu'il écumait gravement,] était bon comme une nuit de sommeil.*

Les séquences entre crochets droits indiquent les passages de l'exemple initial qui ne sont pas pertinents pour l'analyse du schéma élémentaire considéré (il peut s'agir d'une autre subordonnée comparative, ou bien d'un segment adventice comme c'est le cas de l'apposition *l'esclave de la maison*).

Au total, la « Table des exemples élémentaires » contient 2678 exemples, dont 2314 de français moderne. Chaque exemple élémentaire a alors été caractérisé, puis analysé du point de vue morpho-syntaxique.

L'étape de caractérisation concerne l'assignation, à chaque exemple élémentaire, d'un schéma syntaxique global constitué de deux segments : la matrice A (éventuellement constituée de deux

⁶ Sur les différences entre les subordonnées comparatives en *que* et celles en *comme*, voir Fournier et Fuchs (2007).

parties A' et A'' entre lesquelles vient s'insérer la subordonnée) et la subordonnée B. D'où l'étiquetage d'une structure « A B » ou « A' B A'' ». Le schéma global de l'énoncé se retrouve par concaténation des étiquetages de A', B, A'' et de la ponctuation ; ex : « N est <> adj. que N est » (le symbole <> désigne le marqueur d'égalité ou d'inégalité).

Voici, à titre d'illustration, la caractérisation du schéma global de l'exemple n° 0103 qui, en tant qu'exemple élémentaire, porte le n° 0103.00 (puisque l'exemple initial ne comporte qu'une seule subordonnée comparative) :

L'étape d'analyse morpho-syntaxique concerne l'ensemble des constituants de la matrice et de la subordonnée. Plus de 50 champs ont ainsi été constitués et renseignés. Nous en donnons ci-dessous un aperçu (liste non exhaustive).

- constituants de la matrice :
 - o déclencheur (*plus, ...*)
 - o modifieur du déclencheur (*beaucoup, un peu, ...*)
 - o support du déclencheur (*travailler, beau, vite, ...*)
 - o catégorie de ce support (verbe, adjectif, adverbe, ...)
 - o fonction de ce support

- temps du verbe
 - lexies (*d'autant plus que, préférer... plutôt que, ...*)
 - résomptif (*de même, ainsi,...*)
 - etc.
- constituants de la subordonnée :
- cheville subordonnante (*que, comme*)
 - si la subordonnée est verbale :
 - type de verbe
 - temps du verbe
 - mode du verbe
 - place du sujet par rapport au verbe
 - présence de discordanciers (*ne, non pas, ...*)
 - présence d'anaphoriques (*le, en, ...*)
 - etc.
 - si la subordonnée est averbale :
 - catégories des constituants
 - fonctions des constituants
 - déterminants
 - etc.

Voici, à titre d'illustration, le codage morpho-syntaxique de la matrice de l'exemple n° 0103.00 :

Microsoft Access - [F_Analyse_Comme]

Exemple: Menu ... **Analyse syntaxique et sémantique d'exemples de Comparaison**

Aristote n'est pas moins essentiel à toute métaphysique que ne sont les éléments d'Euclide à la géométrie.

N°Exple 0103.00 ExCible: **Aristote n'est pas moins essentiel à toute métaphysique ► que ne sont les éléments d'Euclide à la géométrie** Métacatégorie Adjectif

Frcs Mod/Anc FrMod Type: littéraire Complet/Inc complet

Schéma Syntaxique Sémantique Détail A / B Matrice - Déclencheur Kw - Subordonnée Origine / Statut Champs travail

NbDecl: 1 Décl: moins Semantq: Inégalité (infériorité) LexieGch: -
MultiDecliste: LexieDr: -
Support droit du déclencheur: essentiel MultiSupports: LexieComplexe: -
Catég du support du décl: [GAdj] Métacatégorie du supp Adjectif
Fct du support du décl: [Attribut du Sujet] Tiroir Vb matrice: [Indicatif présent]
Modif décl: - Position du modif déclencheur: (non pertinent)
Modif comme: (non pertinent)

Enr : 17 sur 1974
Mode Formulaire

et voici le codage morpho-syntaxique de la subordonnée de ce même exemple n° 0103.00 :

Microsoft Access - [F_Analyse_Comme]

Exemple: Menu ... **Analyse syntaxique et sémantique d'exemples de Comparaison**

Aristote n'est pas moins essentiel à toute métaphysique que ne sont les éléments d'Euclide à la géométrie.

N°Exple 0103.00 ExCible: **Aristote n'est pas moins essentiel à toute métaphysique ► que ne sont les éléments d'Euclide à la géométrie** Métacatégorie Adjectif

Frcs Mod/Anc FrMod Type: littéraire Complet/Inc complet

Schéma Syntaxique Sémantique Détail A / B Matrice - Déclencheur Kw - Subordonnée Origine / Statut Champs travail

Kw: que Subord: **que ne sont les éléments d'Euclide à la géométrie** LexieGch: -
LexieDr: -
LexieComplexe: -

Nature subord: verbale PositionSubord: 65 Accessoire post Ess
Catég subord: N est
Déterminant du GN: (non pertinent)
Fct du (ou des) constituants de la sub: Résomptif: -
Place subord+/lexie: postposé
Vb subord: être Place sj nmal subord: postposé
Identiq Vb matrice: Oui Nég subord: ne
Tiroir Vb matrice: indicatif présent Le / en / y: (rien)
Tiroir Vb subordonnée: indicatif présent

Pour une typologie de l'ensemble des structures syntaxiques rencontrées, nous renvoyons à Fuchs & al. (2008 : 12-25).

1.3. Analyse et codage sémantique

Afin de pouvoir coder sémantiquement les exemples, il est apparu nécessaire de procéder préalablement à une caractérisation du degré de parallélisme syntaxique entre la relation prédicative de la matrice A et celle de la subordonnée B. Pour ce faire, une série de champs a été élaborée et renseignée, permettant de contraster, en termes de fonctions syntaxiques, les constituants des deux propositions.

A titre d'illustration, voici les informations sur le parallélisme A/B correspondant à l'exemple n° 0103.00 :

The screenshot shows a Microsoft Access window titled 'Microsoft Access - [F_AnalyseComplète]'. The main form is 'Analyse syntaxique et sémantique d'exemples de Comparaison'. The example number is 0103, and the target text is 'Aristote n'est pas moins essentiel à toute métaphysique que ne sont les éléments d'Euclide à la géométrie.' The form is divided into sections for syntactic and semantic analysis. A red box highlights the 'Détail A / B' section, which contains the following data:

A (Supra)		B (ModP)	
Sujet	Aristote	B Sujet	les éléments d'Euclide
Nég	ne pas	B Nég	ne
Mod	est	B Mod	ne
Vb	est	B Vb	ne
Modif	moins essentiel	B Modif	ne
Ct Essentiel: Attr/Cod/Coil/...	à toute métaphysique	B Ct Essentiel: Attr/Cod/Coil/...	à la géométrie
Autre Ct		B AutreCt	
Circ		B Circ	
si A = GN : GN Déterm, GN ModifGch, GN Tête, GN Modif dr, GN Cplt		si B = GN : B GN Déterm, B GN ModifGc, B GN Tête, B GN Modif, B GN Cplt	

Below the highlighted section, there are fields for '+/-/' (Modif décl, Décl, Kw) and 'Commentaire :'. The 'Kw' field contains the word 'que'.

Ces informations concernant la structure des relations prédicatives permettent de calculer le paramètre de la comparaison (ici *essentiel*, adjectif attributif désignant une qualité gradable), le ou les constituant(s) différentiel(s) de B (ici *les éléments d'Euclide ... la géométrie*), le ou les constituant(s) parallèle(s) de A contrastant avec le(s) différentiel(s) de B (ici *Aristote ... toute métaphysique*) et, partant, d'identifier les termes donnant lieu à la comparaison (ici le terme

comparé est *Aristote par rapport à la métaphysique* et le terme comparant *les éléments d'Euclide par rapport à la géométrie*).

Sur ces bases, il était ensuite possible de procéder au codage sémantique des exemples. Deux champs ont ainsi été élaborés et renseignés :

- le champ « schéma comparatif », qui prend une série de valeurs, telles que :
 - o un procès / 2 ou plusieurs entités
 - o un procès / 2 repères temporels ou spatiaux
 - o un procès / 2 situations
 - o un procès / 2 modalités de prédicat
 - o un procès / 2 modalités de phrase
 - o un procès / un autre procès
 - o une qualité / une entité
 - o une énonciation / une autre énonciation
 - o ...

- le champ « type de comparaison », qui prend une série de valeurs, telles que :
 - o égalité de degré
 - o égalité de degré par négation d'une inégalité
 - o égalité valuative
 - o inégalité de degré (supériorité)
 - o inégalité de degré (infériorité)
 - o inégalité de degré (infériorité) par négation d'une égalité
 - o inégalité valuative
 - o identité de manière
 - o identité de manière par négation d'une différence
 - o différence de manière

- différence de manière par négation d'une identité
- ...

Pour une présentation détaillée des différents cas de figure sémantiques rencontrés, ainsi que des divers effets pragmatiques, nous renvoyons à Fuchs & al. (2008 : 25-58).

2. L'élaboration du système d'interrogation

Une fois les données annotées, nettoyées, structurées, il devenait alors possible d'élaborer le système d'interrogation (SI).

L'architecture générale du système a été conçue par C. Fuchs et F. Mélanie-Becquet – en collaboration avec l'ensemble des membres du groupe ; sa mise en œuvre a été assurée par F. Mélanie-Becquet. Il est disponible à l'adresse suivante : <http://129.199.228.10/bdd/comp/>. Pour accéder à la base, il faut préalablement faire la demande d'un login et d'un mot de passe à l'adresse suivante : frederique[dot]melanie[at]ens[dot]fr ⁷.

L'élaboration du système a nécessité, tout d'abord, une réflexion sur les enjeux du système, puis la création de formulaires dynamiques permettant d'interroger et de visualiser la base de données des exemples à subordonnées comparatives ⁸.

2.1. Les enjeux du Système d'Interrogation

Le présent article n'a pas pour vocation d'être un guide utilisateur. Il a pour objectif de décrire la structuration des pages, la navigation de page en page, l'utilisation d'une base de données relationnelle – autant de points nécessaires à la construction d'un SI ayant pour enjeu principal d'être visible sur la toile, d'être accessible à toute la communauté et simple d'utilisation.

⁷ Merci d'envoyer toutes remarques et messages d'erreur à cette même adresse. Toutes les réactions seront utiles à l'évolution du site. Nous tenons à nous excuser par avance des dysfonctionnements et erreurs possibles.

⁸ Le but de cet article n'est pas de donner des informations techniques sur la construction de formulaires dynamiques. Cependant un minimum d'information est fourni à ce sujet en Annexe II.

Notre site web se veut avant tout convivial et facile d'utilisation. Ses atouts sont les suivants :

- Le site Web est épuré. Il ne multiplie pas les polices de caractères et les couleurs, les pages ont des structures identiques et la navigation de page en page est facile.
- Le métalangage utilisé est simple. Il ne reprend pas une terminologie complexe qui ne serait intelligible que pour un nombre limité de personnes (par exemple : la terminologie mise en place et utilisée de façon interne par les participants au projet SCF). Il veut ouvrir à la réflexion sur la langue, laisser toute liberté d'observation et d'interprétation à l'utilisateur.
- Le SI procède par étapes. Les formulaires mis en place proposent des critères précis, des listes finies - que nous détaillons ci-après. Ainsi l'utilisateur a la possibilité peu à peu d'étoffer sa grille de critères. La navigation évolue au fil des choix opérés.
- L'ordre des étapes n'est pas contraint. La navigation est la plus libre possible. L'utilisateur choisit l'ordre d'apparition des formulaires. Il peut dans un premier temps définir son corpus, puis sélectionner des marqueurs de comparaison - ou inversement. Il peut à sa guise revenir sur l'un ou l'autre des formulaires, et modifier les critères préalablement choisis. C'est pourquoi, pour offrir une plus grande clarté et une continuité au système, une section *rappel*⁹ est mise en place dans les formulaires *sources* et *marqueurs de comparaison*, et lors de l'affichage des résultats. Elle a vocation à rappeler les choix validés au fil de la navigation.
- Le contenu des formulaires n'est pas contraint. Aucun remplissage minimal ou maximal des formulaires n'est requis.
- Les exemples répondant aux critères retenus sont présentés dans un tableau, tableau qui pourra ultérieurement être exporté au format Excel.

Observons en détail chacun de ces formulaires.

2.2. Formulaires : des requêtes au résultat

Notre SI est constitué de manière à guider au mieux un utilisateur - linguiste ou non - s'intéressant à la comparaison - et plus spécifiquement aux subordonnées comparatives - vers un ensemble d'exemples correspondant aux critères recherchés.

⁹ Cf. Annexe II.

Le SI élaboré permet à l'heure actuelle d'effectuer des recherches simples¹⁰. Il est constitué de deux formulaires de requêtes, correspondant à deux étapes majeures dans la recherche des exemples. Une étape est la définition du corpus : y correspond le formulaire *sources*. L'autre étape est la sélection de termes comparatifs : y correspond le formulaire *marqueurs de comparaison*. L'utilisateur a la possibilité de commencer par l'une ou l'autre des étapes, ou de ne passer par aucune de ces deux étapes et d'accéder directement à l'ensemble des exemples collectés.

2.2.1. Le formulaire *sources*

Le formulaire *sources* permet de sélectionner des critères liés aux sources (grammaires et dictionnaires pris en compte dans la base) et à leurs exemples. Il est constitué de trois champs :

(1) le type d'ouvrage : l'utilisateur peut sélectionner soit la totalité des grammaires soit la totalité des dictionnaires disponibles dans la base.

(2) l'époque : l'utilisateur peut sélectionner Ancien Français (11^e, 12^e et 13^e siècles), Moyen Français (14^e et 15^e siècles), 16^e, 17^e, 18^e, 19^e ou 20^e siècles.

(3) Le type d'exemples : l'utilisateur peut sélectionner les *exemples d'auteur*, c'est-à-dire des exemples de la littéraires, ou bien les *exemples forgés* (construits par les grammaires) *ou recueillis* (exemples attestés recueillis auprès de locuteurs)¹¹.

L'utilisateur complète le formulaire comme il lui convient, dans l'ordre qui lui convient. Il peut ne remplir aucun des trois champs – et ainsi ne porter aucune restriction sur le corpus¹². Si l'utilisateur a complété le formulaire, des informations complémentaires lui sont demandées dans deux cas précis.

(a) Premier cas, un type d'ouvrage a été sélectionné.

¹⁰ A terme, il a vocation à être enrichi et complexifié afin de permettre des recherches avancées (voir conclusion).

¹¹ Nous avons inclus dans les exemples attestés non littéraires les 11 exemples à subordonnées comparatives extraits de la grammaire de Damourette et Pichon *Des mots à la pensée. Essai de grammaire de la langue française* (voir référence complète en Annexe I) notés par les auteurs comme exemples de « locuteur ».

¹² Quand aucun champ n'est informé, en cliquant sur *valider* l'utilisateur est dirigé vers le formulaire *marqueurs de comparaison*.

En cliquant sur *valider*, l'utilisateur actualise le formulaire *sources*. Ce dernier est à nouveau généré mais enrichi d'une liste d'ouvrages - soit une liste de grammaires, soit une liste de dictionnaires (selon le choix effectué précédemment). La liste apparaît sous forme de cases à cocher. L'utilisateur coche autant d'ouvrages qu'il le désire.

Prenons un exemple. Quelqu'un souhaite isoler dans la base de données la totalité des exemples littéraires du 19^e siècle utilisés dans *Le Bon Usage*. Il remplit le formulaire *sources* à l'aide des listes de choix proposées. Il sélectionne *grammaires* comme type d'ouvrage, 19^{ème} siècle comme époque et *exemples d'auteur* comme type d'exemples. Il valide, et obtient une version rafraîchie du même formulaire. Une liste de 8 grammaires lui est proposée¹³. Il lui suffit alors de cocher *Le Bon Usage*.

Il faut noter que l'ordre dans lequel l'utilisateur remplit le formulaire, a une incidence. En effet, dans l'exemple énoncé précédemment, l'utilisateur remplit la totalité des champs, puis clique sur le bouton « valider ». Il demande ainsi au système d'interrogation d'exclure du résultat souhaité, les grammaires qui :

- n'illustrent pas leur propos à l'aide d'exemples littéraires, mais construisent leur propres exemples (forgés). C'est le cas de *La Grammaire d'aujourd'hui* et de *la Systématique de la langue française*¹⁴.
- n'utilisent pas d'exemples littéraires du 19^e siècle.

C'est pourquoi seules 8 grammaires, parmi les 18 que comprend la base, sont affichées à l'étape suivante.

(b) Deuxième cas, *Dictionnaire de l'Académie* est coché.

L'utilisateur choisit de visualiser des exemples de dictionnaires et plus spécifiquement ceux du *Dictionnaire de l'Académie*. Le formulaire est alors à nouveau rafraîchi. Comme les exemples de la base ont été recueillis dans les huit éditions successives de ce dictionnaire, l'utilisateur peut choisir quelle(s) édition(s) il souhaite retenir. La liste des éditions lui est proposée, il peut cocher la ou les éditions souhaitée(s). Comme le souligne Pruvost (2006 : 38), les éditions successives de ce dictionnaire « représentent un corpus unique pour tout un ensemble d'études comparatives sur la

¹³ Cliquer sur *valider* entraîne l'exécution d'une requête SQL qui retourne la liste des grammaires utilisées. Cette liste est présentée sous forme de cases à cocher, à savoir des contrôles de type *checkbox*.

¹⁴ Voir référence complète en Annexe I.

langue française d'une période à l'autre ». Notre SI est à ce titre un moyen d'accéder aux exemples forgés par les académiciens pour illustrer la comparaison au fil des années, des siècles.

En sélectionnant les 8 éditions du *Dictionnaire de l'Académie*, c'est l'ensemble des exemples collectés qui est disponible et qui peut être observé. Cela permet d'obtenir des informations précieuses concernant l'adressage et la modification d'exemple.

- l'adressage.

Certains exemples sont utilisés deux fois dans une même édition. C'est le cas des exemples suivants :

- (1) Ce vin-là est encore pire que le premier.
- (2) Tant plein que vide.
- (3) Elle est moins jolie que sa sœur.

Les exemples (1) et (2) apparaissent une seule fois dans les 5 premières éditions du *Dictionnaire de l'Académie*, respectivement sous l'entrée « pire » et « tant ». Mais à partir de 1835, année de la sixième édition, se met en place un double adressage – présence de l'exemple sous deux entrées dictionnairiques : celle du 'déclencheur' et celle de la 'cheville subordonnante' (*que*).

L'exemple (3) est utilisé pour la première fois en 1835. Dans cette édition et les suivantes, il apparaît sous l'entrée « moins » et sous l'entrée « que ».

Ainsi le double adressage semble être une pratique caractéristique des trois dernières éditions. Cependant, le phénomène n'est pas généralisé à la totalité des exemples avec déclencheur.

- la modification d'exemple.

Plusieurs siècles séparent la première édition (1694) de la dernière (1935). Au fil des années la société évolue, certains exemples sont modifiés, refondus. C'est ce que nous illustrons ci-dessous :

- Exemples présents de la première à la huitième édition.

(4a) *ce que je vous en dis est moins pour vous faire de la peine, que pour vous **marquer**, etc.*

(4b) *Ce que je vous en dis est moins pour vous faire de la peine que pour vous **avertir**.*

(5a) *Comme le Soleil chasse les tenebres, ainsi la **grace** chasse **le péché**.*

(5b) *Comme le soleil chasse les ténèbres, ainsi la **science** chasse **l'erreur**.*

(6a) *Ce **vaisseau** contient autant que l'autre.*

(6b) *Ce **vase** contient autant que l'autre.*

L'exemple (4) est placé sous l'entrée « moins » : (4a) utilisé en 1694 laisse la place à (4b) dans les 7 éditions suivantes.

L'exemple (5) est placé sous l'entrée « ainsi » : (5a) est utilisé dans les deux premières éditions puis il laisse la place à (5b).

L'exemple (6) est placé sous l'entrée « autant » : (6a) est utilisé dans les 5 premières éditions, avant de laisser la place à (6b).

- Exemple présent dans un nombre restreint d'éditions.

(7a) *Rien n'anime le soldat, comme **l'espoir du pillage**.*

(7b) *Rien n'anime le soldat comme **l'exemple des chefs**.*

L'exemple apparaît dans trois éditions consécutives, sous l'entrée « comme » : (7a), utilisé en 1762 et 1798, est modifié en (7b) en 1835, avant d'être évincé des éditions suivantes.

Ces quelques exemples pris pour illustrer notre propos, montrent l'intérêt de la base ainsi constituée.

2.2.2. Le formulaire *marqueurs de comparaison*

Le formulaire *marqueurs de comparaison* utilise deux critères significatifs pour la sélection d'exemples à subordonnées comparatives. Il s'agit pour l'utilisateur de choisir, d'une part le terme qui introduit la subordonnée - appelé *terme subordonnant* (1) -, et d'autre part le terme 'déclencheur' marquant le degré, l'identité ou l'altérité - appelé *terme comparatif* (2) - qui sert de support au terme subordonnant. L'utilisateur peut décider de remplir les deux champs du formulaire, aucun d'entre eux ¹⁵, ou seulement l'un des deux champs.

(1) Les termes subordonnants.

Chacun des exemples de la base contient une structure à subordonnée comparative ¹⁶. Ainsi chacun des exemples est constitué d'au moins un terme subordonnant : soit *que*, soit *comme*.

(2) Les termes comparatifs.

Dans les exemples à subordonnée comparative, « deux grands types sont à distinguer, selon que la matrice comporte un 'déclencheur' de la comparaison (*plus, aussi, autre, ...*) ou non » (C. Fuchs & al., 2008 : 12). La liste de ces déclencheurs est proposée à l'utilisateur.

Aux critères sélectionnés précédemment, dans le formulaire *sources* ("*Le Bon Usage*", "19^{ème} siècle et "exemples d'auteurs"), l'utilisateur peut ajouter un marqueur de comparaison. Comme le montre la copie d'écran ci-dessous, il choisit d'ajouter comme critère le *terme comparatif* « si » :

¹⁵ Quand aucun champ n'est informé, en cliquant sur *valider* l'utilisateur est dirigé vers le formulaire *sources*.

¹⁶ Nous rappelons que notre base est constituée uniquement d'exemples à subordonnées comparatives.

SÉLECTION DES MARQUEURS DE COMPARAISON

Accéder à la page

- Accueil
- Sources
- Marqueurs
- Exemples

Critères entrés dans le formulaire sources :

- un type de source: grammaires
- ... 1 source(s) choisie(s):
"Le Bon Usage"
- une période : "19e"
- un type d'auteur:exemples d'auteurs

*En cliquant sur valider vous actualisez l'étape suivante
En cliquant sur exemples vous accédez aux exemples

Illustration de la terminologie utilisée

Pierre est **plus** bête **que** méchant

↑ ↑
terme comparatif terme subordonnant

Vous pouvez maintenant sélectionner les marqueurs de comparaison suivants :

terme subordonnant :

terme(s) comparatif(s) :

"aussi"
"autre"
"davantage"
"meilleur"
"mieux"
"moins"
"plus"
"plutôt"
"préférer"
"si"

Exemples *

Il est à noter que seuls 10 déclencheurs – parmi les 29 que compte actuellement la base ¹⁷ - lui sont proposés. Ce sont la totalité des déclencheurs présents dans les *exemples littéraires* du 19^e utilisés dans *Le Bon Usage*.

L'utilisateur aurait tout aussi bien pu remplir dans un premier temps le formulaire *marqueurs de comparaison*, sélectionner le terme comparatif « si », valider le formulaire, et dans un second temps informer le formulaire *sources*, sélectionner comme type de sources *grammaires*. En procédant ainsi, l'utilisateur aurait eu à choisir *Le Bon Usage* parmi 16 grammaires ¹⁸.

Il s'agit d'un outil de recherche élémentaire :

- Seuls deux des nombreux champs créés par les membres du groupe SCF sont exploités. Ils ont été jugés suffisants et significatifs pour une première version du SI. Les nombreuses autres informations obtenues suite aux annotations des exemples seront utilisées ultérieurement pour le développement d'un outil de recherche avancée.
- L'interrogation porte sur une chaîne de caractères. Il n'y a pas de possibilité à l'heure actuelle d'interroger sur une contiguïté possible du *marqueur de comparaison* et du *terme subordonnant*.

¹⁷ Pour l'instant, la liste contient uniquement des déclencheurs de français moderne.

¹⁸ Dans la base, 2 grammaires n'ont pas d'exemples à subordonnées comparatives ayant comme déclencheur « si » : la *Grammaire de la langue française du 17^{ème} siècle* et *La grammaire d'aujourd'hui* (références complètes en Annexe I). Dans ce dernier ouvrage, nos relevés d'exemples proviennent des paragraphes 'Complément du comparatif et superlatif' (p.43-44) et 'Les circonstanciées de comparaison (comparatives)' (p.113). Une référence est faite aux subordonnées comparatives en « si » : « La comparaison peut également être marquée par *comme, ainsi, que, de même que (...), d'autant plus que, tel, autant, aussi, si... que, le même que*, ou un adjectif ou un adverbe au comparatif suivi de *que* » (p.113), mais aucun exemple est donné.

2.2.3. Le formulaire *résultats*

Il est possible qu'aucun exemple ne corresponde aux critères retenus. Dans ce cas, le tableau ne s'affiche pas, le message suivant est délivré : "*Aucun exemple ne correspond aux critères retenus*".

Si des exemples correspondent aux critères sélectionnés, alors les informations affichées sont les suivantes :

- le *nombre d'exemples* extraits de la base de données et qui répondent aux requêtes établies.

Puis, pour chaque exemple :

- l'*exemple brut* : l'exemple tel qu'il a été relevé dans la source, sans repérage aucun de la subordonnée comparative.

- le *nombre de sources* dans lesquelles se trouve l'exemple : cette information figure dans le tableau si le tri demandé n'est pas un tri par *ordre alphabétique des sources*.

- les références *auteur* et *œuvre* : quand il s'agit d'un exemple non forgé (par exemple un exemple littéraire).

- l'*ouvrage source* dans lequel se trouve l'exemple : la localisation précise est donnée (le numéro de la page, ainsi que les références de paragraphe)¹⁹.

A ce stade, l'utilisateur peut agir sur l'ordre d'affichage des exemples. Divers tris sont possibles :

- "afficher les résultats par *ordre alphabétique de l'exemple*"

- "afficher les résultats par *ordre alphabétique des sources*" : il s'agit d'un classement par ordre alphabétique des titres pour les dictionnaires, et d'un classement par ordre alphabétique des auteurs pour les grammaires.

- "afficher les résultats par *ordre alphabétique des noms d'auteur*".

¹⁹ C'est pourquoi l'utilisateur pourra trouver en ligne la liste des sources utilisées ainsi que leur édition.

En cliquant sur un de ces boutons, le même tableau est généré, mais cette fois les exemples sont classés comme souhaité.

Conclusion

Le système d'interrogation actuel est une première étape. Des améliorations et enrichissements de divers ordres sont envisagés :

- faciliter l'utilisation et l'export des données : mise en place d'un bouton *export*, permettant d'enregistrer les exemples trouvés dans un tableau Excel.
- augmenter le nombre d'exemples contenu dans la base de données en étendant à des exemples issus de sources non encore explorées : (i) des sources de même type : grammaires ou dictionnaires non encore explorés ; (ii) des sources de types différents : œuvres littéraires, articles scientifiques, etc. Ce travail devrait être facile à effectuer et de moindre coût, la grille de recueil et de collecte d'information étant mise en place.
- Enrichir et affiner le système d'interrogation : par intégration des divers niveaux d'analyses (analyse des structures syntaxiques et analyse sémantique). Comme nous l'avons signalé ces analyses sont effectuées mais ne sont pas encore intégrées dans le système d'interrogation. Elles le seront par constitution d'un formulaire de recherche avancée donnant, par exemple, la possibilité de choisir des exemples d'un auteur ou une œuvre littéraire particuliers dans une liste donnée ou d'interroger sur la non contiguïté entre marqueurs de comparaison et subordonnant.

Ces enrichissements sont envisagés et seront facilités par le fait que la base de données et le système d'interrogation associé sont structurés.

La consultation et l'interrogation en ligne de la base sont un atout pour son enrichissement et sa pérennisation. Un système de champ libre de type *commentaire* est envisagé. Il permettra au système d'évoluer, de collecter des informations, de prendre en considération les remarques émises par les utilisateurs.

Références bibliographiques

Fournier N. et Fuchs C., 2007, « *Que* et *comme* marqueurs de comparaison », *Lexique*, 18, p. 69-107.

Fuchs C., Fournier N. et Le Goffic P., 2008, « Structures à subordonnée comparative en français : problèmes de représentations syntaxiques et sémantiques », *Linguisticae Investigationes*, XXXI : 1, p. 11-61.

Fuchs C., 2008, « Autour de la base SCF », *Linguisticae Investigationes*, XXXI :1, p. 1-10.

Habert B., 2005, *Instruments et ressources électroniques pour le français*, Gap / Paris, Ophrys.

Habert B., 2009, *Construire des bases de données pour le français (tome I : Notions)*, Paris, Ophrys.

Heurtel O., 2010, *PHP 5.3 Développez un site Web dynamique et interactif*, St Herblain, ENI.

Pruvost J., 2006, *Les dictionnaires français, outils d'une langue et d'une culture*, Gap / Paris, Ophrys.

ANNEXE I

Liste des sources des exemples

- 10 grammaires du français moderne :

- Arrivé M., Gadet F., Galmiche M., 1986, *La Grammaire d'aujourd'hui*, Paris, Flammarion.
- Brunot F., 1965³ [1926], *La Pensée et la langue*, Paris, Masson.
- Chevalier J.C, Blanche-Benveniste C., Arrivé M., Peytard J., 1997¹¹ [1964], *Grammaire du Français contemporain*, Paris, Larousse-Bordas.
- Damourette J., Pichon E., 1911-1940, *Des mots à la pensée. Essai de grammaire de la langue française*, Paris, d'Artrey.
- Grevisse M., 1986¹²[1936], *Le Bon Usage*, refondue par A. Goosse, Paris-Gembloux, Duculot.
- Moignet G., 1981, *Systématique de la langue française*, Paris, Klincksieck.
- Riegel M. et al., 2004³ [1994], *Grammaire méthodique du français*, Paris, PUF.
- Sandfeld Kr., 1977² [1936], *Syntaxe du français contemporain. Les propositions subordonnées*, Genève, Droz.
- Wagner R.L., Pinchon J., 2004¹⁰ [1962], *Grammaire du français classique et moderne*, Paris, Hachette.
- Wilmet M., 1998² [1997], *Grammaire critique du français*, Louvain-La-Neuve, Duculot - Hachette.

- 8 grammaires portant sur l'histoire du français ou sur des états de langue anciens :

- Brunot F., 1905-1938, *Histoire de la langue française*, Paris, Colin.
- Buridant C., 2000, *Grammaire Nouvelle de l'Ancien Français*, Paris, Sedes.

- Gougenheim G., 1974² [1951], *Grammaire de la langue française du 16^{ème} siècle*, Nouvelle édition, entièrement refondue, Paris, Picard.
- Haase A., 1898, *Syntaxe française du 17^{ème} siècle*, éd. Traduite et remaniée, Obert M., Paris, Delagrave, 1975.
- Oudin A., 1632, *Grammaire française rapportée au langage du temps*, Reprint, Genève, Slatkine, 1972².
- Maupas C., 1618, *Grammaire et syntaxe française*, Reprint, Genève, Slatkine, 1973.
- Spillebout G., 1985, *Grammaire de la langue française du 17^{ème} siècle*, Paris, Picard.
- Vaugelas C.F. de, 1647, *Remarques sur la langue française*, Reprint, Paris, Droz, 1934.

- 3 dictionnaires de français moderne :

- *Le Grand Robert*, version électronique, Paris, Dictionnaires Le Robert, 1994.
- *Le Nouveau Petit Robert*, version électronique, Paris, Dictionnaires Le Robert, 2001² [1996].
- TLFi, 1971-1994, version électronique du *Trésor de la Langue Française. Dictionnaire de la langue du 19^e et 20^e siècle*, Paris, CNRS-Gallimard, 2004.

- 4 dictionnaires portant sur des états de langue anciens :

- Furetière A., 1690, *Dictionnaire universel*, Reprint, Paris, SNL – Le Robert, 1978.
- Huguet E., 1925, *Dictionnaire de la langue française du 16^{ème} siècle*, Paris, Champion.
- Richelet P., 1679-1680, *Dictionnaire français*, Reprint, Genève, Slatkine, 1994.
- 1694-1935, *Dictionnaire de l'Académie française*, Edition électronique, Redon.

ANNEXE II

Structuration du SI : des langages efficaces et performants.

Comme il a été dit au paragraphe 1, le travail d'annotation a nécessité la constitution d'une base de données relationnelle, la création de tables et de champs permettant de décrire non seulement les sources, mais aussi la syntaxe et la sémantique des exemples à subordonnées comparatives. De la constitution de ces tables résulte une structuration possible du SI :

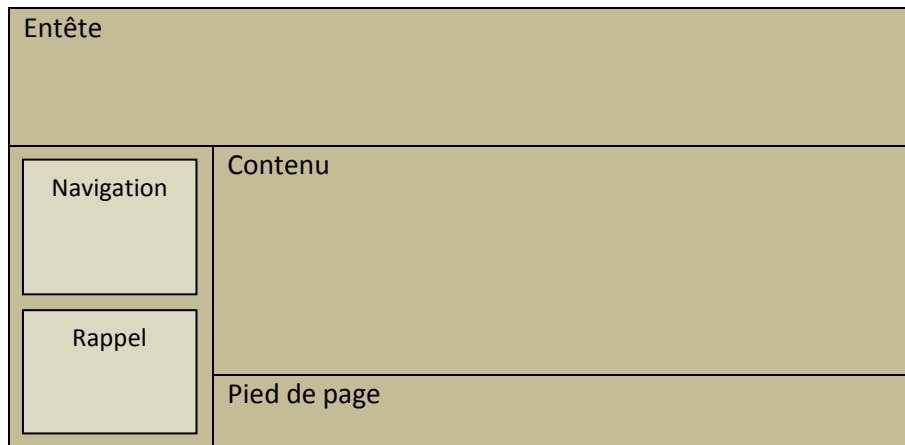
- Des interrogations sur les sources : ce niveau utilise les informations recueillies lors du codage des sources.
- Des interrogations sur les marqueurs de la comparaison : ce niveau utilise quant à lui certaines des informations recueillies lors du codage syntaxique ²⁰.
- La présentation des résultats : les exemples répondant aux critères sélectionnés sont présentés dans un tableau ; leur est associée une référence précise de source(s).

HTML et la création de pages Web

Nos pages Web sont des documents HTML. La feuille de style CSS (Cascading Style Sheets) est utilisée pour déterminer toute mise en forme : les couleurs de pages et de polices, les polices de caractères utilisées, etc. Ainsi forme et contenu sont séparés, ce qui facilite la maintenance des fichiers et assure une meilleure pérennisation des données.

Le corps de chacune des pages est structuré de manière identique :

²⁰ L'analyse sémantique n'est à l'heure actuelle pas intégrée dans l'outil d'interrogation, elle le sera ultérieurement, de même que l'analyse des structures syntaxiques.



La zone *entête* contient les logos des institutions hébergeant le projet ainsi que le nom du formulaire.

La zone *bandeau gauche*. Elle est située à gauche de l'écran et est subdivisée en deux zones :

- Une zone *navigation* qui contient les boutons de type *submit* permettant de naviguer de page en page.
- Une zone *rappel* qui permet à l'utilisateur de visualiser les choix effectués dans les pages précédemment fréquentées.

La zone *contenu* varie selon les pages. Elle est :

- description – comme le sont les pages d'accueil ;
- formulaire de requête – contenant listes de choix et checkbox ;
- tableau de résultat – contenant les exemples correspondants aux critères retenus.

La zone *pied de page* contient le bouton *Valider* qui permet – comme son nom l'indique – de valider les informations entrées dans le formulaire.

PHP : un site dynamique.

Le passage entre les pages de notre site est assuré par l'utilisation de scripts PHP. Ce langage « permet de réaliser des pages Web dynamiques dont le contenu peut-être complètement ou partiellement généré au moment de l'appel de la page, grâce à des informations récupérées dans un formulaire ou extraites d'une base de données » (Heurtel, 2010 : 14).

Une contrainte majeure est cependant à prendre en considération : « A la différence d'un langage comme JavaScript, où le code est exécuté côté client (dans le navigateur), le code PHP est exécuté côté serveur. Le résultat de cette exécution est intégré dans la page HTML qui est envoyé au navigateur » (Heurtel : *ibid.*). Ce qui signifie que l'information entrée dans un formulaire est traitée uniquement à partir du moment où le formulaire est envoyé au serveur, uniquement à partir du moment où l'utilisateur valide ce qu'il a saisi dans la page.

Les scripts PHP sont placés dans le fichier `Index.php`. Ce fichier est placé à la racine du site et gère toute ouverture de pages que ce soit l'accès au site ou l'ouverture d'une page consécutivement à la validation d'une autre page.

MySQL : une base de données.

La base de données dont il est question dans le présent article a initialement été élaborée sous Access. Il s'agit dans un premier temps de l'exporter vers MySQL ²¹. Il sera ainsi possible d'interroger la base via SQL (Structured Query Language). SQL est un langage de requêtes, permettant des requêtes fines. Il a pour avantage d'être utilisé par une large communauté ; il s'adapte à de nombreux environnements et est très documenté.

Les requêtes SQL sont utilisées dans les formulaires de requêtes pour créer les listes de choix, sous forme de listes déroulantes (*Select*) ou de cases à cocher (*Checkbox*), dans le formulaire *résultat* pour lister les exemples correspondant aux critères retenus.

L'utilisation de l'ensemble de ces outils et langages permet la création d'un site web dynamique.

²¹ MySQL est un Système de Gestion de Bases de données libre et gratuit. Nous ne précisons pas ici les aspects techniques de cette exportation, le lecteur intéressé peut se rendre sur le site du SI pour trouver ces informations.