UNIVERSITE TOULOUSE II – JEAN JAURES
DEPARTEMENT DES SCIENCES DU LANGAGE
MASTER ERGONOMIE COGNITIVE ET INGENIERIE LINGUISTIQUE

MEMOIRE DE MASTER 1

POUR UN REPERAGE AUTOMATIQUE DES STRUCTURES ENUMERATIVES A L'AIDE DES NOMS SOUS-SPECIFIES

MARINE RAOULT

SOUS LA DIRECTION DE MME LYDIA-MAI HO-DAC ET DE MME JOSETTE REBEYROLLE

REMERCIEMENTS

Je tiens à adresser mes remerciements à toutes les personnes qui ont pu contribuer de près ou de loin à la rédaction de ce mémoire de M1.

Dans un premier temps, je remercie Mme Ho-Dac et Mme Rebeyrolle, mes directrices de mémoire, pour le temps qu'elles m'ont accordé et les nombreuses solutions qu'elles ont pu apporter à mes interrogations, quelles qu'elles soient.

Merci également aux membres de l'équipe enseignante de cette année de M1, pour leur disponibilité.

Merci à la promo ECIL pour cette année riche en émotions.

Merci tout particulièrement à Sophie pour son aide, ses (nombreuses) relectures, et pour toujours m'apporter un grand bol d'air quand je sature.

Merci à mes proches, famille, amis, pour leurs conseils, leurs relectures, leurs avis.

Charte de non-plagiat

Afin de valoriser le travail personnel et pour sensibiliser les étudiants au problème du plagiat, l'Université de Toulouse II-Jean Jaurès met en œuvre un dispositif qui promeut les bonnes pratiques de la citation d'auteurs et la correcte utilisation d'idées tierces dans les devoirs, mémoires et thèses.

Définition du plagiat :

« Plagier c'est : s'approprier le travail créatif de quelqu'un d'autre et de le présenter comme sien; s'accaparer des extraits de texte, des images, des données, etc. provenant de sources externes et les intégrer à son propre travail sans en mentionner la provenance; résumer l'idée originale d'un auteur en l'exprimant dans ses propres mots, mais en omettant d'en mentionner la source. Plagier est non seulement un acte malhonnête, mais aussi une infraction qui peut entraîner des sanctions. » http://www.bibl.ulaval.ca/infosphere/sciences/evaciter1.html , consulté le 25/10/2011

Pour bien travailler :

Les enseignants et les services de documentation veilleront à transmettre les connaissances nécessaires au respect des règles de la propriété intellectuelle (citations, synthèses d'idées d'auteurs, bibliographies, notes de bas de page, etc.) Un site informatif et didactique de référence est accessible à cette adresse : http://zero-plagiat.univ-tlse2.fr

Informations concernant le contrôle du plagiat :

Les enseignants auront la possibilité de recourir à un logiciel de recherche de similitudes. Ce dernier compare le texte des travaux rendus avec une vaste base de référence. Les rapports rendus par le logiciel mettent en avant les similitudes repérées sans pouvoir les qualifier de plagiat. Ce sont les enseignants qui contrôlent le bon ou le mauvais usage des emprunts et déterminent s'il y a eu plagiat. Le recours au logiciel se fera dans le respect des clauses de confidentialité si elles existent.

Engagement :

Les étudiants joindront à tous les travaux remis à des enseignants en vue d'une évaluation une déclaration sur l'honneur remplie et signée (en annexe de la présente charte).

Déclaration sur l'honneur de non-plagiat

Je soussignée,

Nom, Prénom : Raoult Marine

Régulièrement inscrit à l'Université de Toulouse II-Jean Jaurès

N° étudiant : 21401830

Année universitaire : 2014/2015

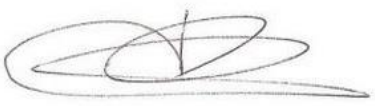
Certifie que le document joint à la présente déclaration est un travail original, que je n'ai ni recopié ni utilisé des idées ou des formulations tirées d'un ouvrage, article ou mémoire, en version imprimée ou électronique, sans mentionner précisément leur origine et que les citations intégrales sont signalées entre guillemets.

Conformément à la charte des examens de l'Université de Toulouse II-Jean Jaurès, le non-respect de ces dispositions me rend passible de poursuites devant la commission disciplinaire.

Fait à : Toulouse

Le : 15 juin 2015

Signature :



SOMMAIRE

TABLE DES FIGURES

TABLE DES TABLEAUX

INTRODUCTION

L'objet de ce mémoire de Master 1 est de proposer une analyse de noms particuliers, appelés noms sous-spécifiés, en décrivant leur fonctionnement dans des corpus multi-genres et au sein de structures dites énumératives.
Les noms sous-spécifiés sont des noms du type *problème, fait, objectif, idée, solution*… et qui sont souvent qualifiés d'abstraits puisqu'ils sont à la recherche de contenu (Legallois, 2006), fourni par un antécédent situé dans le cotexte. Ainsi dans un exemple du type :

L'objectif est de [gagner la course de dimanche].

Le nom *objectif* a donc besoin d'un contenu propositionnel (Legallois, 2008) qui lui est donné par son antécédent *gagner la course de dimanche*.
De même avec :

Paul a terminé [premier dans toutes les matières]. Que prouvent ces résultats ?

Le syntagme nominal *ces résultats*, dont la tête est un nom sous-spécifié, permet de résumer la partie gauche, on parle donc d'anaphore résumante, ou encore d'anaphore nominale résomptive. Ainsi les noms sous-spécifiés font partie d'une catégorie nominale à part, qui permet de résumer d'une phrase à un paragraphe précédent. Nous nous proposerons donc de décrire le fonctionnement de ces noms particuliers et de nous intéresser à eux plus particulièrement au sein des structures énumératives, de façon à voir s'ils peuvent permettre de les extraire de manière automatique.
Les structures énumératives sont des structures qui permettent de structurer le discours, selon différentes constructions, et qui sont composées d'un élément obligatoire : l'énumération, et de deux éléments facultatifs : l'amorce, un élément essentiel, et la clôture.
Les items de l'énumération possèdent tous un lien, qui est inféré à l'aide d'un critère de co-énumérabilité, qui peut être explicite ou implicite.

- *Item 1*
- *Item2*
- *…*
- *Item n*

*Ces **problèmes** montrent bien […]*

Dans le cas où ce critère est explicite, on le retrouve dans l'amorce et/ou dans la clôture, et nous pouvons poser pour hypothèse que ce critère prend souvent la forme d'un nom sous-spécifié, d'où l'intérêt d'étudier ces noms particuliers. Ces noms sous-spécifiés apparaissent au sein de constructions syntaxiques particulières : certaines ont déjà été décrites, d'autres ne l'ont pas encore été pour le français. Il conviendra donc de décrire le fonctionnement de ces noms et de leurs patrons syntaxiques, que ce soit de façon générale ou de façon plus spécifique, au sein des structures énumératives. Le but sera de vérifier si ces noms peuvent apparaître dans des structures propres aux structures énumératives afin de pouvoir les repérer automatiquement.
Dans ce mémoire, un premier chapitre permettra de proposer une description des anaphores, et plus précisément des anaphores résomptives, et de l'enjeu que la résolution de ces anaphores peut avoir en traitement automatique des langues. Les structures énumératives seront également décrites, en tant qu'éléments assurant la cohésion discursive. La description des structures énumératives permettra de mettre en évidence que très souvent, lorsque l'énuméraThème de la structure est explicite, il correspond à un nom sous-spécifié, raison pour laquelle nous nous intéresserons à ces derniers.
Un deuxième chapitre correspondra à l'analyse de ces noms sous-spécifiés, en règle générale et plus particulièrement dans les structures énumératives, à l'aide d'un corpus constitué de deux sous-corpus : le corpus ANNODIS et le corpus Littéracie Avancée. De plus, les noms sous-spécifiés seront analysés à travers les trois sous-corpus d'ANNODIS, afin de voir si le genre textuel entraîne un emploi particulier de noms sous-spécifiés.

CHAPITRE I : ETAT DE L'ART

I. Cohérence, cohésion et progression discursive : l'anaphorisation

Afin de mieux cerner le fonctionnement des noms sous-spécifiés (désormais NSS), il paraît important de discuter du rôle des anaphores. En effet, un NSS est « un classificateur de la valeur qui le spécifie – ce qui, au niveau textuel, a son importance (Legallois, 2006) puisque la valeur peut, dans la suite du texte, être à nouveau évoquée par la seule convocation du nom » (Legallois, 2008). Un classement des anaphores sera donc proposé, accompagné du fonctionnement de chacune d'entre-elles.

1. Les différents types d'anaphores

Un texte, en tant que signifiant linguistique, organise son contenu de manière successive. Cette organisation peut comporter des phénomènes de reprises, de répétitions, qui permettent de faire progresser le texte et d'en assurer la cohérence.

Il s'agit donc, au sein d'un texte, de répéter certaines informations afin de permettre la cohérence textuelle. Seules certaines informations seront reprises et constitueront le thème, c'est-à-dire l'information déjà connue du lecteur, par opposition au rhème, qui correspond à une information nouvelle. Ajoutons que le rhème d'une phrase peut parfaitement devenir le thème de la phrase suivante, le thème étant l'information connue. Cette opposition thème/rhème, introduite par Combettes (1983) permet de distinguer plusieurs niveaux informationnels dans une phrase, qui jouent un rôle important dans la progression discursive. Bien que critiquée pour son aspect quelque peu simpliste, cette opposition a au moins le mérite d'introduire ces différents niveaux d'organisation de l'information.

Plusieurs stratégies peuvent être mises en place afin d'assurer la reprise de certaines informations, par exemple :

(1) Marie est grande. Marie a les cheveux longs. Marie est étudiante.

On a ici une progression à thème constant (Combettes, 1983), c'est-à-dire que le thème, ici *Marie* est repris à chaque nouvelle phrase et introduit trois rhèmes :

Rhème 1 : Grande

Rhème 2 : Cheveux longs

Rhème 3 : Etudiante

Certaines stratégies peuvent permettre d'alléger cet exemple (1), comme les anaphores, qui ont pour rôle la reprise textuelle de certaines informations, de façon à assurer la cohérence textuelle. Reprenons l'exemple précédent en opérant une pronominalisation :

(2) Marie est grande. Elle a les cheveux longs et elle est étudiante.

Dans ce deuxième exemple, le nom *Marie* est repris par un pronom dit *anaphorique*, à savoir *Elle*, qui réfère à *Marie*. *Marie* est l'antécédent de *elle*. Selon la définition de Ducrot et Schaeffer (1995, p. 548), « un segment de discours est anaphorique lorsqu'il fait allusion à un autre segment, bien déterminé, du même discours, sans lequel on ne saurait lui donner une interprétation (même simplement littérale). »

Une référence est anaphorique lorsqu'elle reprend un élément antérieur mais situé dans le cotexte, c'est-à-dire dans l'environnement interne du texte. Une référence peut être anaphorique lorsqu'elle a pour antécédent un élément situé dans le cotexte gauche, et elle peut être cataphorique lorsqu'elle a pour antécédent un élément situé dans le cotexte droit, par un phénomène d'anticipation :

(3) Je veux *cette voiture*. Je trouve qu'**elle** est très belle.
(4) Quand elle part en vacances, Marie est toujours contente.

Ces deux exemples permettent d'illustrer la définition précédente, ainsi dans l'exemple (3), *elle* est une référence anaphorique dont l'antécédent *cette voiture* est situé dans le cotexte gauche, et dans l'exemple (4), *elle* est une référence cataphorique dont l'antécédent *Marie* est situé dans le cotexte droit.

Les exemples précédents permettent de représenter le cas de l'anaphore pronominale, cependant il faut noter que plusieurs types d'anaphores existent.

Plusieurs classements ont été proposés au sein de la littérature linguistique. Sans nous pencher sur tous ces classements, on retiendra ici les stratégies de reprise cotextuelles suivantes :

- Les anaphores fidèles et infidèles :
  > « On parle d'anaphore fidèle lorsqu'un référent préalablement introduit dans le texte est rappelé au moyen d'un SN défini ou démonstratif dont le nom tête est celui-là même au moyen duquel il a été introduit (« une maison… la/cette maison »). L'anaphore fidèle est donc une des figures possibles de la coréférence. On parle en revanche d'anaphore infidèle lorsque le nom de la forme de rappel est différent de celui de la forme introductrice (il s'agit alors le plus souvent d'un synonyme ou d'un hyperonyme), ou qu'il est adjoint d'une détermination quelconque (« une maison… l'habitation », « une maison…cette coquette bâtisse ») » (Apothéloz, 1995, p. 36-37).
- Les anaphores associatives :
  « On appelle anaphore associative la relation synecdochique (de la partie au tout) qui unit le représentant à son antécédent : *Ma voiture est trop vieille ; le moteur est fragile* » (Jeandillou, 2006). Selon Kleiber (1999, p. 1), l'anaphore associative consiste à voir « un phénomène de référence textuelle indirecte, c'est-à-dire l'introduction par l'expression anaphorique d'un nouveau référent via le référent de l'expression antécédent », avec l'exemple suivant qu'il donne pour appuyer cette définition : « *Il s'abrita sous un vieux tilleul. Le tronc était tout craquelé* »
- Les anaphores conceptuelles ou résumantes, dont la construction et le rôle seront détaillés dans la partie suivante

Pour ce mémoire, seules nous intéresseront et seront détaillées les anaphores résumantes, dites également conceptuelles.

2. Le cas des anaphores résumantes

Les anaphores résumantes, que l'on peut trouver décrites sous le nom d'anaphores conceptuelles ou encore d'anaphores résomptives, permettent de résumer et/ou remplacer une phrase, une partie d'un texte voire la totalité d'un texte à l'aide d'un unique syntagme nominal. Le fait de remplacer l'antécédent par un syntagme nominal permet au scripteur de créer un nouveau référent et d'y ajouter de l'information, d'où l'intérêt de ces anaphores résumantes. Apothéloz et Chanet (1997, p. 2) parlent de *nominalisation,* en la définissant comme étant « l'opération discursive consistant à référer, au moyen d'un syntagme nominal, à un procès ou un état qui a préalablement été exprimé par une proposition ».

(5) Ainsi, les escadrons d'attaque de la Force aérienne sont passés dès 1978 sous la coupe de Saddam Hussein. Plus tard, la Sûreté et les Istikhbarat, soustraits aux ministères de l'Intérieur et de la Défense, respectivement, ont de même été soumis à la tutelle d'une présidence concentrant toujours plus d'autorité.

Tout ce **processus** sera renforcé par le développement de l'image de l'ennemi intérieur, relais des " impérialistes " et autres " sionistes ", avant que l'identification des minorités irakiennes " complices " soit bientôt doublée de celle d'un ennemi extérieur autrement important : l'Iran. *(Corpus ANNODIS_me)*

L'exemple (5) illustre bien le fonctionnement d'une anaphore résumante. En effet, le simple syntagme nominal *ce processus* permet ici de résumer la totalité d'un paragraphe. De plus, notons que ce résumé

à l'aide d'une anaphore n'est pas sans but, puisque tout l'intérêt d'une reprise anaphorique telle que celle de l'exemple (5) est d'utiliser un nouveau référent pour ajouter de l'information et permettre au scripteur d'apporter un jugement sur ce qu'il a écrit plus tôt. Le scripteur souhaitait ici utiliser l'information qu'il avait donnée précédemment pour mettre en évidence quelque chose à travers une nouvelle information, or réécrire le texte n'aurait pas eu d'intérêt et aurait été bien trop lourd pour le lecteur ; malgré tout, le scripteur avait besoin de ce texte pour exprimer son opinion, c'est pourquoi il l'a résumé au moyen d'un syntagme nominal anaphorique (Schmid, 2000).

L'anaphore résumante a pour particularité de résumer une information textuelle relativement complexe par un simple syntagme nominal, sémantiquement pauvre, et qui est le plus souvent abstrait : *fait, exemple, situation*… C'est d'ailleurs en cela que ces anaphores nous intéressent, puisque rappelons que ce mémoire s'intéresse particulièrement aux NSS, qui sont justement ces noms sémantiquement pauvres, et qui permettent de résumer tout ou partie d'un texte. Ces anaphores sont relativement simples à reconnaître car elles sont le plus souvent introduites par un déterminant démonstratif, qui va fonctionner avec un nom abstrait. Ce sont les antécédents qui peuvent en revanche poser problème, puisqu'il n'est pas toujours évident de les repérer ou d'en dégager les limites. La difficulté à repérer ou limiter les antécédents constitue d'ailleurs un des problèmes pour la résolution d'anaphores en Traitement Automatique des Langues (désormais TAL), ce que nous allons pouvoir observer dans la partie suivante.

3. Résolution d'anaphores : quels apports du TAL ?

L'intérêt de la reconnaissance des éléments anaphoriques et de leurs antécédents est de permettre la compréhension de textes en langage naturel par une machine, avec des applications en extraction d'informations, traduction, résumé automatique ou encore analyse d'opinions. De plus, les reprises textuelles jouant un rôle important dans la cohérence textuelle, il paraît très important d'être capable de les reconnaître via des outils automatiques. Cependant, la résolution d'anaphores est complexe, pour trois grandes raisons :

- Les anaphores sont nombreuses et de formes variées, nous l'avons vu dans la partie précédente
- Les antécédents sont parfois difficiles à extraire ou à délimiter
- Ces antécédents sont également de longueur variable, puisqu'ils peuvent représenter un syntagme, une phrase, un paragraphe ou même un texte entier

i. Résolution des anaphores pronominales

Nous l'avons vu au début de cette première grande partie, l'anaphore pronominale est une anaphore dans laquelle un pronom réfère à un antécédent situé dans le cotexte.

(6) Paul pense qu'**il** est sage

Dans ce cas, l'antécédent est placé dans le cotexte gauche, mais prenons cet exemple :

(7) **Elle** est grande cette maison

Ici, l'antécédent est placé dans le cotexte droit, dans une relation cataphorique, ainsi les algorithmes pour la résolution d'anaphores doivent pouvoir tenir compte de ces différentes possibilités. Il faut donc qu'ils puissent localiser l'antécédent et le délimiter, que ce dernier soit dans le cotexte immédiat ou qu'il soit situé plus loin. Les anaphores sont des phénomènes complexes et un algorithme ne peut pas se contenter d'associer un pronom anaphorique au syntagme nominal le plus proche, car cela ne fonctionnerait pas dans bien des cas, par exemple :

(8) *Pierre* a dit à Paul qu'**il** allait acheter une voiture

Pour résoudre ces anaphores pronominales, la plupart des outils considèrent qu'un pronom renvoie au dernier syntagme nominal de la phrase (cotexte gauche), ou au premier si certaines contraintes ne sont pas satisfaites avec le dernier, par exemple la contrainte du genre (exemples 9 et 10). Dans l'exemple (6), le pronom *il* ne renvoie pas au premier nom propre de son cotexte gauche mais au deuxième, à savoir *Pierre*, ce qui peut poser des problèmes pour la reconnaissance automatique des anaphores.

(9) Paul a été témoin d'une collision. Il a appelé la police.
(10) Paul a été témoin d'une collision. Elle a fait plusieurs victimes.

Dans l'exemple (9), si l'on cherche l'antécédent de *il*, le premier syntagme nominal que l'on trouve dans le cotexte gauche est *une collision*, or *il* ne peut pas référer à un syntagme nominal féminin, ainsi *il* réfère forcément à *Paul*. C'est exactement l'inverse que l'on observe dans l'exemple (10), où cette fois *Elle* est un pronom féminin, qui ne peut donc référer qu'à un syntagme nominal féminin, ici le seul disponible est *une collision*. Bien que ces contraintes associées dans des algorithmes résolvent la plupart des anaphores pronominales, certains cas sont plus complexes et demanderaient une analyse sémantique en plus d'une simple analyse syntaxique. En effet, sans une analyse sémantique, il paraît difficile de savoir, dans l'exemple suivant, que *le parti* réfère à *PS* , et il est aisé de concevoir, avec un exemple de la sorte, qu'il est difficile, malgré toutes les avancées en matière de résolution automatique d'anaphores, de détecter tous les éléments de coréférence.

(11) Solferino veut travailler sur sa communication pour que la voix du *PS* soit davantage identifiable pour les militants. **Le parti** souhaite améliorer la formation des militants, qui seront désormais initiés aux dix points de la « Charte des socialistes pour le progrès humain », adoptée par les militants début décembre.

ii. Algorithmes pour la résolution des anaphores

La résolution d'anaphores repose sur des algorithmes plus ou moins complexes et un algorithme naïf consisterait à :

- Rechercher les anaphores (par exemple les pronoms anaphoriques dans le cas de la résolution des anaphores pronominales)
- Dresser la liste des candidats possibles pour être des antécédents
- Sélectionner le bon antécédent parmi les candidats, à l'aide de contraintes (genre et nombre par exemple) et de préférences (SN le plus proche du pronom favorisé, par exemple).

Le système MARS (Mitkov, 2002) peut être représentatif des systèmes de RA reposant sur des indices de surface. L'architecture de ce système, issue de Weissenbacher (2008) est la suivante

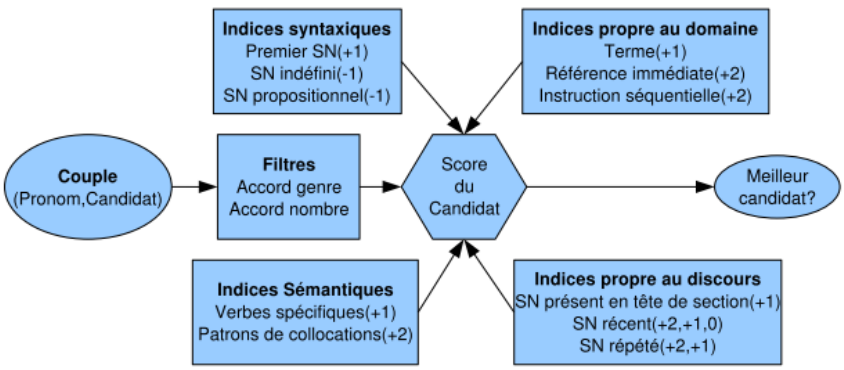


Figure 1 Architecture du système MARS

Le système commence tout d'abord par segmenter les textes, puis il applique un Pos tagging et dresse la liste des antécédents possibles. L'antécédent choisi est celui qui répond à des contraintes et à des préférences avec le pronom, dans le cas de la résolution d'anaphore pronominale.

Boudreau et Kittredge (2006) proposent également un algorithme simple pour la résolution de chaines de coréférence reposant sur 4 étapes :

- Etape lexicale : étape qui consiste à identifier les mots importants (noms propres, mots grammaticaux et mots lexicaux)
- Etape syntaxique I : cette étape construit des syntagmes à partir de l'étape lexicale
- Etape syntaxique II : attribution de fonctions syntaxiques
- Etape coréférentielle : Choix d'un antécédent

La résolution d'anaphores consiste donc en un processus automatique, selon des algorithmes, qui relie une anaphore (par exemple un pronom anaphorique) à son antécédent.

Nous l'avons vu précédemment, les algorithmes qui serviront pour la résolution automatique doivent intégrer à la fois des contraintes de genre et de nombre, mais également des contraintes sémantiques. De plus, ces algorithmes de résolution d'anaphores (désormais RA) sont souvent projetés sur des corpus annotés. Or les annotations disponibles ne sont que rarement parfaites. Les systèmes doivent donc prendre en compte ces annotations incertaines pour la résolution de leurs anaphores (Weissenbacher, 2008). La résolution des anaphores pronominales doit donc reposer, toujours selon Weissenbacher (p. 34), sur trois étapes : « la reconnaissance des pronoms impersonnels et anaphoriques, la création de la liste des candidats à l'antécédence et le choix de l'antécédent parmi les candidats ». Les systèmes de RA peuvent reposer sur des connaissances linguistiques complexes, telles que la syntaxe, la sémantique ou le discours. D'autres peuvent reposer sur des indices de surface, que sont les informations distributionnelles, syntaxiques, sémantiques et pragmatiques.

II. Les noms sous-spécifiés : acteurs de la cohésion discursive

1. Définition et rôles des NSS

Les NSS, que Shmid (2000) qualifie de *shell nouns* sont des noms dont le comportement diffère de celui des autres noms, puisque leur sens est abstrait, c'est-à-dire qu'il s'agit de noms qui vont chercher leur contenu propositionnel (Legallois, 2008) dans un autre segment du discours, d'où le terme de *sous-spécifié*. En qualifiant ces NSS de *shell nouns*, Schmid propose que ces noms fonctionnent comme des *coquilles vides*, qui se remplissent à l'aide d'un antécédent situé dans le contexte, il explique ainsi que ces noms sont incomplets sémantiquement. Prenons les exemples suivants pour mieux comprendre le fonctionnement de ces NSS, et cette idée de *coquille vide* :

(12) En intégrant les TP dans ces créneaux horaires, dont le **but** est de [faire progresser les élèves avec un programme personnalisé], nous constatons que l'enseignant maximise au mieux les conditions de travail des élèves afin de les voir réussir.

(13) Le **fait** est que [certains « mauvais compreneurs » (G. Chauveau, 2011) peuvent tout à fait arriver à comprendre le sens premier du texte lu (explicite) mais sont incapables d'inférer (implicite).] *(exemples issus du corpus « littéracie avancée »)*

Ces exemples illustrent bien la notion d'incomplétude sémantique dont parle Schmid, c'est-à-dire qu'isolés du reste de la phrase, on ne peut pas leur attribuer de sens, ainsi dans l'exemple (26), le contenu propositionnel de *but* est donné grâce à l'antécédent *faire progresser les élèves avec un programme personnalisé*. C'est également le cas dans l'exemple (27), où le nom *fait* réfère à *certains « mauvais compreneurs » (G. Chauveau, 2011) peuvent tout à fait arriver à comprendre le sens premier du texte lu (explicite) mais sont incapables d'inférer (implicite)*. Ce sont ces antécédents qui permettent de fournir le contenu propositionnel au NSS.

Les NSS permettent « d'encapsuler » le discours qui précède ou celui qui suit, et ils peuvent référer à des phrases plus ou moins complexes et plus ou moins longues, voire même à un paragraphe ou plus.

2. Les patrons de construction des NSS

Selon Legallois (2008), les NSS s'intègrent au sein de structures dites spécificationnelles. L'auteur nous donne un exemple de ce type de structures avec : *ce que je voudrais, c'est des vacances*, où le segment droit, à savoir *c'est des vacances* est de nature spécificationnelle, c'est-à-dire qu'il vient spécifier le segment gauche, qui lui est dit « superscripturale ». Si l'on reprend les exemples (26) et (27) cités plus haut, nous pouvons remarquer qu'ils apparaissent dans une construction quasi identique, malgré le fait que (26) introduit une proposition infinitive alors que (27) introduit une complétive. Cependant, on peut dégager un patron de construction qui serait le suivant :

Det N [Ø | ce] être [que-clause | de-inf] (Legallois 2008)

Legallois est donc parti de l'étude de quelques exemples dans lesquels apparaissaient NSS pour dégager ce patron de construction, qu'il a ensuite projeté sur un corpus d'une année du quotidien *Libération* (1995) pour tenter de dresser une liste des NSS du français.
La liste de ces noms a été divisée en trois, selon qu'ils référaient à une infinitive, à une complétive ou qu'ils étaient commun aux deux constructions. Cette liste peut être retrouvée en Annexe 1. Bien que non exhaustive, elle propose déjà beaucoup de noms potentiellement sous-spécifiés selon les structures dans lesquelles ils apparaissent, et servira d'ailleurs dans la partie exploitation des données de ce mémoire, ainsi que nous le reverrons. Une seconde liste, celle de Roze et al. (2014) est fournie en Annexe 2. Cette deuxième liste fournit les 20 NSS les plus fréquents dans le corpus de leur étude, après la projection des patrons proposés ci-après.

Malgré cela, notons que Legallois n'a projeté qu'un seul patron de construction pour obtenir cette liste, or Roze et al. (2014) ont pu faire émerger d'autres patrons qui pourraient, s'ils étaient à leur tour projetés, nous apporter une liste peut-être encore plus complète de noms sous-spécifiés. Ainsi, après avoir projeté le même patron que Legallois sur leur corpus, et après avoir relevé les NSS les plus fréquents, Roze et al. ont utilisé une méthode de fouille de données, afin d'extraire les motifs émergents pour pouvoir dégager de nouveaux patrons de construction des NSS.
Dans cette étude, les motifs émergents sont définis de la façon suivante : « [ce] sont des motifs dont le support augmente de manière significative d'un ensemble de données à un autre. Les motifs émergents sont ainsi des motifs dont le taux de croissance (« growth rate »), c'est-à-dire le rapport des supports dans deux ensemble de données, est supérieur à un seuil fixé *p* ». De nouveaux patrons ont pu être extraits à partir de ces motifs émergents, à savoir :

[V avoir] [P pour] [NSS] [P de]
« Ces deux procédures ont pour **résultat** de déplacer insensiblement le centre d'intérêt, tenu dans la première version par l'éblouissante héroïne, vers le personnage effacé et hypocondriaque de son mari]… »

[P pour] [NSS] [P de]
« Le gouvernement s'est fixé pour **objectif** de [parvenir à 1.6% du PIB en 2008 et à 2% en 2010] »

[NSS] [Ponct :]
« Leur **objectif** : [être identifiés comme des « adultes disponibles » avec lesquels on peut parler de tout et de rien] »

« **Conclusion** : [à ce jour, la monnaie unique n'a guère enrayé le malaise économie européen et l'on ne peut manquer de s'interroger sur son éventuelle responsabilité dans les difficultés économiques actuelles de la zone euro]. »

[NSS] [CS que]
« L'**idée** [que les inspecteurs puissent renifler les armes et les documents qui s'y rapportent sans l'aide des autorités irakiennes] est absurde. »

[NSS] [P de] [V inf]
« C'est un site commercial qui a été lancé en juin 2002, à New York, avec l'**objectif** de [mettre les gens en relation les uns avec les autres autour d'un sujet d'intérêt commun] »

La liste des NSS proposée par Legallois (2008) ainsi que les constructions syntaxiques qu'il a pu décrire nous serviront de base pour l'étude de ce mémoire et seront étudiés dans la partie exploitation des données. Nous ajouterons à cela les patrons de construction définis par Roze et al. (2014).
Au total, si l'on reprend la construction proposée par Legallois (2008) et les patrons mis en évidence par Roze et al. (2014), nous avons donc 6 patrons qui permettent de repérer des noms qui seraient sous-spécifiés.

3. Les NSS au sein des structures énumératives

Peu d'études ont été réalisées actuellement pour étudier les NSS, et celles qui l'ont été ne les étudient pas au sein de ces structures particulières que sont les SE, précédemment décrites. Nous l'avons vu, les SE sont des structures qui peuvent faciliter la cohésion du discours, et qui sont constituées d'une liste d'items, liés par un critère de co-énumérabilité plus ou moins explicite, et d'éléments facultatifs que sont les amorces et les clôtures. Le critère de co-énumérabilité, lorsqu'il est exprimé, prend le plus souvent la forme d'un nom, et ce nom est bien souvent sous-spécifié, puisque nous avons pu noter que ce critère n'est que rarement l'hyperonyme des items listés.

Une première intuition, qui sera à vérifier dans la suite de ce mémoire, serait qu'il y aurait plus de NSS dans les clôtures, bien que ces dernières soient plus rares que les amorces, du fait qu'il ne s'agisse pas d'un élément essentiel. Nous verrons par la suite quelles méthodes nous permettront de confirmer ou d'infirmer cette hypothèse, en exploitant des données qui seront présentées dans le chapitre 2.

(14) **L'objectif** de cette étude est double : [*Il s'agit* de mesurer le rôle du Congrès des Etats-Unis en temps de guerre, que l'actualité récente de la question irakienne éclaire d'un jour nouveau, tout en dressant un tableau plus large qui s'appuie sur les crises majeures auxquelles Washington a fait face. Un tel travail devrait également permettre de rappeler la répartition des pouvoirs instituée par la Constitution, que les parlementaires ne manquent d'ailleurs jamais de rappeler. *Il s'agit d'autre part,* à la lumière des enseignements tirés du passé et du contexte actuel, d'évaluer les possibilités offertes à l'Administration Bush par les débats qui ont précédé la récente guerre en Irak, dont les parlementaires ont accepté le principe en septembre 2002, bien avant que les opérations ne commencent, en mars 2003 : Cette étude permettra de mieux comprendre le rôle du Congrès en matière de politique étrangère, en particulier en ce qui concerne l'envoi de forces armées sur des théâtres extérieurs, qui, vu de l'étranger, en représente incontestablement l'aspect le plus significatif.]

Cet exemple (28) permet d'illustrer ces emplois de NSS au sein de SE, ainsi *l'objectif* réfère à 2 items, qui apparaissent dans la structure qui suit. A travers ces SE, les NSS apparaissent le plus souvent dans un patron qui n'a pas été décrit par Legallois (2008) et Roze et al. (2014) mais qui l'a en revanche été par Schimd (2000) et qui est le suivant : *That + N*.

En français, ce patron serait le suivant : [*Det] [NSS] [V] [Ø | CS que]*, ce qui peut se traduire par l'exemple (29) :

(15) *L'Indochine montre qu'une guerre asymétrique peut simplement être perdue par la puissance dominante. Et le Viêtnam, qu'un conflit ne se gagne pas forcément sur le champ de bataille principal*. **Ces affrontements** inégaux répètent que [la manière traditionnelle dont nos militaires conçoivent l'occupation et la manœuvre du champ de bataille n'est pas universelle].

Selon Schmid (2000), le fait d'utiliser un nouveau syntagme nominal anaphorique, à savoir *ces affrontements*, permet une réactivation, c'est-à-dire qu'une information déjà connue est reprise par ce SN, ce qui permet l'ajout d'une nouvelle information. Pour imager cet exemple et la notion de coquille vide, on peut considérer que c'est l'information en italique dans l'exemple (29) qui vient remplir cette coquille. Toutefois, ce patron reste très vague, et il conviendra de l'expliciter dans le second chapitre avec l'exploitation des données utilisées.

4. Mesures du potentiel de sous-spécification d'un nom

Les NSS sont des noms complexes, qui peuvent apparaître au sein de constructions particulières (notamment les constructions qui ont pu être décrites dans la partie précédente). Selon Schmid (2000), certains noms sont plus susceptibles d'apparaître dans ces constructions particulières que d'autres. Par exemple, un nom tel que *fait*, de plus relativement fréquent au sein du corpus utilisé, serait plus susceptible d'apparaître dans un emploi sous-spécifié qu'un nom moins fréquent tel que *malentendu*.

Selon l'auteur, la probabilité qu'un nom apparaisse dans un emploi sous-spécifié ou non dépend de sa distribution au sein du corpus. Afin de décrire cela, il propose deux mesures statistiques qui sont les suivantes :

$$attraction = \frac{fréquence\ d'un\ nom\ dans\ un\ patron}{fréquence\ totale\ du\ patron}$$

$$reliance = \frac{}{fréquence\ totale\ de\ ce\ nom\ dans\ le\ corpus}$$

L'*attraction* correspond au rapport entre la fréquence d'un nom dans un patron en fonction de la fréquence totale du patron en question. Cette mesure permet de calculer le degré d'attraction d'un patron pour un nom donné. La formule *reliance* correspond au rapport entre la fréquence d'un nom dans un patron et la fréquence totale de ce nom dans un patron. Le résultat de ce calcul permet de calculer le degré de dépendance d'un nom à un patron.

Pour cette étude, la formule *reliance* permettra de mettre en évidence les noms qui apparaissent le plus souvent dans un emploi sous-spécifié. Ainsi les noms qui apparaîtront le plus souvent au sein de patrons syntaxiques particuliers, tels que ceux décrits précédemment, seront considérés comme plus souvent sous-spécifiés que les autres.

Concernant cette formule de *reliance*, Schmid (2000) considère que des scores de plus de 15% sont intéressants, des scores de 20% peuvent être considérés comme des indicateurs qu'un nom donné est utilisé comme un NSS et que des scores de plus de 30%, 40% ou 50% et plus sont clairement spectaculaires. Le calcul de la reliance pour chaque NSS potentiel sera donc fait dans le chapitre 2, afin de mesurer le potentiel de sous-spécification d'un nom. De même, la formule attraction sera utilisée afin de voir si certains patrons sont plus susceptibles d'attirer un nom particulier ou non.

La description de ces NSS et des patrons au sein desquels ils apparaissent se fera à travers le chapitre 2 et l'analyse de données. Ces données seront issues de deux corpus, à savoir le corpus ANNODIS_me et le corpus de Littéracie avancée. Il s'agira d'observer le comportement de NSS au sein de ces deux corpus et plus particulièrement dans les amorces et les clôtures des structures énumératives, dont nous détaillons l'organisation dans la troisième partie de cet état de l'art.

III. Les structures énumératives et l'organisation discursive

Tous les exemples de cette section, hormis ceux pour lesquels une précision est apportée à la suite de l'exemple, sont issus de Wikipédia, accessible grâce au lien suivant : http ://fr.wikipedia.org/wiki/Wikip%C3%A9dia :Accueil_principal

1. Rôles et formes des structures énumératives pour l'organisation du discours

Les structures énumératives (désormais SE) (Ho-Dac et al. (2012), Bras et al. (2008)) représentent un moyen de structurer le discours. Aussi appelées séries avec des Marqueurs d'Intégration linéaires (MIL), elles permettent de réunir un ensemble d'items possédant un lien les uns avec les autres au sein d'une même structure. Une SE peut prendre différentes formes mais possède obligatoirement une énumération, c'est-à-dire une liste d'items, qui peut être entourée d'une amorce et/ou d'une clôture. Les différents items qui constituent l'énumération ne sont pas sans lien les uns avec les autres, ils sont presque toujours dans une relation d'égalité, et cette relation est marquée par un critère que l'on nomme critère de co-énumérabilité, qui peut être explicite ou non, c'est-à-dire qu'il n'a pas nécessairement besoin d'apparaître explicitement pour exister. Cette structuration peut être mise en évidence au moyen de *Marqueurs d'Intégration Linéaire* (numéral, lieu, temps) (Bras, Prévot et Vergez-Couret, 2008). Toutefois, ces MIL ne suffisent pas à définir ces structures et il convient de ne pas négliger l'organisation typo-dispositionnelle, qui permet de mettre en évidence ces structures, avec une ponctuation spécifique, des retours à la ligne etc., d'où l'apparition de la terminologie de SE.

En effet, ces recours permettent au lecteur de dégager très rapidement l'intention du scripteur (Rebeyrolle & Péry-Woodley, 2014), ainsi il sait de suite qu'il est confronté à une énumération et peut rapidement dégager le critère de co-énumérabilité : il s'agit donc d'une stratégie discursive, qui repose sur un contrat entre le lecteur et le scripteur. L'étude de Ho-Dac, Péry-Woodley et Tanguy (2010) insiste sur le fait que la composition des SE peut varier selon différents critères, par exemple la taille, puisque cette dernière peut varier de quelques mots à plusieurs pages, le nombre d'items, qui peut aller de 2 à beaucoup plus, la présence ou non d'éléments facultatifs (amorces et clôtures) et le niveau de grain des SE, c'est-à-dire leur interaction avec la structure du document, en fonction du nombre de paragraphes constituant la SE, l'alignement des items avec des titres de section et l'alignement des items avec ceux d'une liste formatée.

Les auteurs de cette même étude proposent, suite à ces constats, une typologie des SE en fonction du niveau de grain :

Type 1 : dont les items correspondent à des sections titrées

Type 2 : dont les items correspondent à des listes formatées

Type 3 : couvrant plus d'un paragraphe sans marques visuelles spécifiques

Type 4 : intra-paragraphiques

Les exemples (12) et (13) permettent de mettre en évidence cette grande variation dans la composition des SE :

(16) **Parmi les divers crimes de guerre, on peut citer :**

- les expériences pseudo-médicales de nombreux médecins nazis dans les camps de concentration, notamment du docteur Mengele :
  - **en France** (massacres commis par les nazis, ayant touché plus de 30 000 personnes) :

- exécutions d'otages par les Allemands à Châteaubriant, à Paris, à Lyon, à Limoges, etc. à partir de l'automne 1941. L'historien Serge Klarsfeld a établi la liste de 1007 otages et résistants fusillés au Mont-Valérien près de Paris, dont 117 étaient juifs,
- massacre d'Oradour-sur-Glane, le plus important avec 654 victimes dont des femmes et des enfants, exécuté par la division SS Das Reich. Il y eut 4 rescapés,
- massacres à Ascq, à Tulle, à Maillé, à Buchères, à Etobon, à Saint-Pierre-de-Clairac etc. perpétrés par les SS,
- torture et massacre des civils et des combattants du maquis du Vercors, par des unités de la Wehmacht et les milicies de Joseph Darnand (juillet 1944),
- persécutions sur les membres de familles de Français partis combattre ou restés au pays, comme Geneviève de Gaulle, nièce du général, envoyée en camp de concentration.

- **en Belgique :**
  - rationnement dramatique de la nourriture,
  - installation d'un camp de concentration à Breendonk entre Bruxelles et Anvers,
  - déportation de juifs et de résistants dans les camps de concentration allemands, révocation et arrestation de hauts fonctionnaires et d'autorités communales,
  - persécutions meurtrières exercées sur la population par les collaborateurs de l'Allemagne, entre autres le massacre de familles à Courcelles, dans la province de Hainaut, la persécution sur des parents des ministres Hubert Pierlot et Spaak du gouvernement belge libre de Londres dont certains membres seront fusillés, exécutions de notables comme le gouverneur de la Société Générale de Belgique et le batonnier Braffort,
  - déportation en Allemagne du roi, de son épouse et des enfants royaux.

- **en Italie**, occupée par l'Allemagne en 1943 :
  - massacre de 355 otages aux fosses Ardéatines près de Rome en mars 1944,
  - massacre du village de Marzabotto près de Bologne en octobre 1944, qui fit plus de 900 morts.

  - dans le « Protectorat » de Bohème-Moravie :
  - déportation de centaines d'étudiants ayant manifesté contre l'occupation (novembre 1939)
  - massacre des habitants de Lidice, en représailles à l'attentat qui abattit le chef SS et « boucher de Prague » Heydrich.
- **en Pologne :**
  - affamement et déportation du ghetto de Varsovie
  - « nettoyage » du ghetto de Varsovie par les SS après l'insurrection des derniers survivants
  - extermination de 50 000 membres des élites polonaises par les SS et la Gestapo (prêtres, aristocrates, professeurs, officiers). L'enseignement secondaire, les séminaires et les universités furent fermées, tout comme les théâtres par exemple, et ce n'est qu'à un système remarquable de cours clandestins — les *komplety* — que les Polonais parvinrent à instruire et à sauver cinq classes d'âge de bacheliers
  - massacre de 5 000 officiers polonais à Kayri, par l'armée soviétique (l'URSS a reconnu sa responsabilité après plusieurs décennies, ayant longtemps accusé les nazis d'être responsables de ce massacre)

- massacre de 10 000 autres officiers polonais en d'autres lieux, soit 15 000 personnes tuées froidement d'une balle dans la nuque par le NKVD, ancêtre du KGB.
- destruction à 90 % de Varsoviepar l'armée allemande après le soulèvement de l'Armia Krajowa du 1er août au 2 octobre 1944. La répression de l'insurrection par Himmler fit de 150 000 à 200 000 morts. Manquant de moyens pour franchir la Vistule et immobilisée par ordre de Staline pour des raisons politiques, l'Armée rouge laissa les Allemands écraser la rébellion polonaise et ne lui apporta ni armes ni aide.

- **en Union soviétique**
    - affamement et mise à mort prémédités de prisonniers de guerre russes (3 millions de morts)
    - affamement délibéré des civils de la cité de Leningrad assiégée (700 000 victimes)
    - 20 millions de citoyens de l'Union soviétique sont tués, dont un très grand nombre de prisonniers de guerre exécutés par les Allemands, et aussi des civils dont les villages et villes sont anéantis.

- **en Yougoslavie**
    - déportation de dizaines de milliers de Serbes, Juifs et Roms dans les camps de la mort (notamment dans le camp de concentration de Jasenovac) par les Oustachi croates. Ceux-ci sont responsables du massacre global de 300 000 à 400 000 personnes, ainsi que de multiples conversions forcées au catholicisme.

Cet exemple (12) est relativement long : il est constitué d'une amorce, qui vient introduire la succession d'items et qui nomme le critère de co-énumérabilité, puis on trouve la liste des différents items, mis en évidence à l'aide de moyens typo-dispositionnels (puces, ponctuation, retours à la ligne…). Chaque item permet donc de lister les différents crimes de guerre, ce qui est annoncé dans l'amorce. Cet exemple est même plus complexe, puisque l'on peut noter une SE à l'intérieur d'une autre SE : on a tout d'abord la liste des différents crimes de guerre, qui suit une puce, mais on a également une deuxième structure qui s'intègre à la première, et dont chaque item débute selon le même schéma, c'est-à-dire avec un syntagme prépositionnel dont le spécifieur est un nom propre correspondant à un pays.

(17) Les hommes ont une vie plus agréable que les femmes. Premièrement, ils se marient plus tard et, deuxièmement, ils meurent plus tôt.

Cet exemple (13) est au contraire très court, il est constitué d'une amorce, puis de deux items, mis en évidence à l'aide de Marqueurs d'Intégration Linéaire (*Premièrement* et *Deuxièmement*). Il n'y a ici pas de marques visuelles spécifiques, et il s'agit pourtant bien d'une SE, puisque l'on peut observer une amorce et que deux items apparaissent avec un parallélisme syntaxique.

Il convient également d'appuyer que les SE ne consistent pas seulement en l'énumération d'une liste d'items liés les uns aux autres : les éléments facultatifs qui les entourent font partie intégrante de ces structures et méritent d'être détaillés, de même que le critère de co-énumérabilité.

2. Anatomie des structures énumératives : amorce, clôture et critère de co-énumérabilité

i. Critère de co-énumérabilité

Les items qui constituent les SE ne sont pas seulement des éléments formant une liste sans liens les uns avec les autres. En effet, pour chaque liste de ces structures, un *critère de co-énumérabilité* assure le lien sémantique entre les différents items. Ce critère, ou cet énuméraThème (Ho-Dac, Péry-Woodley et Tanguy, 2010), peut être explicite ou implicite. Dans le cas où il est explicite, ce critère est donné soit dans l'amorce, soit dans la clôture, dont les rôles et les structures seront décrits dans les parties suivantes. Parfois, aucun alignement visuel de la structure ne permet de distinguer explicitement le critère de co-énumérabilité, même si certains critères, par exemple le parallélisme syntaxique, permettent de mettre en évidence une SE.

(18) Si la sociologie française voit en Durkheim son « père fondateur » c'est en partie parce qu'il est le premier à aborder la sociologie comme une discipline scientifique. Cela nécessite :

- d'une part un travail de clarification de son objet afin de le distinguer des discours concurrents sur la société :
    - d'un côté, le différencier de la philosophie, attachée à une démarche de pur raisonnement, de jugement normatif alors que lui, veut imposer une démarche empirique, guidée par la volonté d'établir des faits appuyés sur des données concrètes (statistiques, enquêtes monographiques).
    - De l'autre côté, rompre avec la psychologie, qui ne propose d'explications qu'au niveau individuel alors que l'étude de sa discipline se fait sur le plan collectif.
- d'autre part, il a dû aussi faire reconnaître cette discipline en constituant une équipe de chercheurs, en créant des revues et finalement, en la faisant instituer comme discipline universitaire (il a occupé le premier poste de professeur de sociologie en France).

L'exemple (14) nous montre le cas d'une structure où le critère de co-énumérabilité n'est pas explicitement mentionné au sein de l'amorce, c'est-à-dire dans la partie qui précède l'énumération des items. Concrètement, cela signifie que c'est au lecteur de déduire ce critère de co-énumérabilité et de trouver le lien qui unit les différents items. En revanche, ce critère apparaît clairement dans l'exemple (15) :

(19) Parmi celles-ci, trois *idées clefs* :

- L'esprit est cognitif, c'est-à-dire qu'il contient des croyances, des doutes, etc. La conception passée ne prenait pas en compte ce côté cognitif, ne reconnaissant que des relations logiques du style « si tu me demandes si je veux X, je te répondrai oui ». Au contraire, Chomsky explique que la façon commune de comprendre l'esprit comme ayant des croyances ou encore des états mentaux non conscients, est l'approche à privilégier ;
- Une grande partie de ce que l'esprit d'un adulte peut faire est innée. Même si aucun enfant ne naît avec la capacité de parler directement, tous naissent avec la capacité d'acquisition du langage qui leur permet d'apprendre le langage rapidement dans leurs premières années. Nombre de psychologues ont étendu cette thèse à d'autres domaines que le langage, en contradiction avec la vision du nouveau-né en tabula rasa ;
- L'architecture de l'esprit est modulaire. L'esprit est composé d'un ensemble d'interactions, de sous-systèmes spécialisés (modules), avec un flot limité d'intercommunication. Cette théorie contraste fortement avec l'ancienne conception selon laquelle chaque part d'information peut être accessible par tous les autres processus cognitifs (par exemple, on ne peut pas annuler l'effet d'une illusion optique même si on sait consciemment que c'est une illusion d'optique).

Avec une amorce telle que celle de cet exemple, le lecteur sait que les items qui vont suivre correspondent à des *idées*, il n'a pas à le déduire, avant même que la liste n'apparaisse, il sait ce qui va être abordé.

ii. Amorces

Bien qu'il s'agisse d'éléments facultatifs, les amorces sont relativement présentes au sein des SE, ce qui peut être appuyé par le corpus ANNODIS où les SE ont été annotées, et dans lesquelles les amorces sont présentes dans 95 % des structures (Rebeyrolle & Péry-Woodley, 2014), ce qui nous permet de relever le caractère essentiel de ces dernières.

Les amorces sont très fréquentes au sein de ces structures et en font partie intégrante lorsqu'elles apparaissent. L'amorce consiste en une séquence dont le rôle est d'introduire l'énumération qui va suivre. Cette amorce peut être de type complète ou de type incomplète, selon Bras et al. (2008), dont nous reprenons ici les exemples :

Amorce complète : *Notre plan se déroulera en trois parties : 1. (…) 2. (…) 3. (…).*

Amorce incomplète : *Les trois étapes de notre plan seront : 1. (…) 2. (…) 3. (…).*

Ainsi une amorce complète est une séquence qui est grammaticale sans la liste des items de l'énumération, on peut la comprendre même sans s'intéresser à la suite, même si l'on manquera d'informations pour l'interpréter, tandis qu'une amorce incomplète est une séquence dont il manque une partie pour qu'elle soit grammaticale, c'est-à-dire qu'il y a un trou syntaxique qui se doit d'être rempli par l'énumération qui va suivre. Le plus souvent, selon Bras et al. (2008), une amorce complète est caractérisée par un syntagme nominal pluriel (du type numéral ou article indéfini pluriel), qui permet d'introduire les différents items de l'énumération. Un lexical cataphorique (par exemple *ci-dessous*) peut également jouer ce rôle. Pourtant, il faut noter que les amorces peuvent être très différentes les unes des autres et que leurs structures sont très diversifiées. On peut également parler d'amorces prédictives et d'amorces non- prédictives (Rebeyrolle & Péry-Woodley, 2014*)*, qui correspondent respectivement à l'amorce incomplète et à l'amorce complète, l'amorce prédictive permettant d'inférer le critère de co-énumérabilité, tandis que l'amorce non-prédictive déclenche l'énumération et nomme le critère de co-énumérabilité. La distinction entre les deux nominations tient de leurs descriptions, qui se font au moyen de traits pour les amorces prédictives et non- prédictives (contre une simple distinction syntaxique pour les amorces complète/incomplète), tels que :

- « Un SN avec {numéral, suivants} [N]
- Une place syntaxique à remplir [S]
- De la ponctuation (deux points) [P]
- Des alinéas (typo disposition) [T] » (Rebeyrolle et Péry-Woodley, 2014)

Les auteurs considèrent également que ces traits peuvent s'organiser selon différentes configurations, dans le cas des amorces prédictives, par exemple les trois suivantes :

- Combinaison des quatre traits : on trouve alors des amorces contenant un SN (numéral ou de type *suivants*), un trou syntaxique, deux points et un retour à la ligne, ce qui appelle forcément à une énumération.

(20) Les dépenses réelles de fonctionnement se décomposent de la manière suivante :
- 67.3 M€ pour le personnel et frais assimilés
- 29.5 M€ de subventions et contingents
- 23.9 M€ en Achats de biens et services
- 9.7M€ de charges financières
- 0.2 M€ pour le reste

- Combinaison des traits : SN (numéral ou de type *suivants*), deux points et retour à la ligne.

(21) Le travail descriptif de la linguistique peut se faire selon trois axes principaux :
- Etudes en synchronie et diachronie […]
- Etudes théoriques et appliquées […]
- Etudes contextuelles et indépendantes […]

(22) Une grammaire formelle est constituée de quatre objets :
- Un ensemble fini de symboles, appelés symboles terminaux […]
- Un ensemble fini de symboles, appelés non-terminaux […]

    - Un élément de l'ensemble des non-terminaux, appelé axiome […]
    - Un ensemble de règles de production […]
- Combinaison des traits : trou syntaxique, deux points et retour à la ligne.

(23) D'un point de vue purement linguistique, la syntaxe étudie :

- L'ordre des mots
- Les catégories grammaticales ou parties du discours
- Les phénomènes de rection
- Les fonctions grammaticales

Dans le cas de cet exemple (19), le complément est à aller chercher dans la liste d'items qui va suivre.

L'absence des traits *deux points* et *trou syntaxique* fait automatiquement basculer l'amorce dans le type non-prédictive, avec des exemples du type :

(24) **Événements :**

- 23 janvier : les mutins du navire britannique le Bounty, accompagnés d'un groupe de tahitiens, atteignent l'île Pitcairn, alors inhabitée. Ils débarquent et brûlent leur embarcation. Cette communauté ne sera découverte qu'en 1808, par des baleiniers américains. Un seul des marins britanniques était encore en vie. En 1856, du fait de la surpopulation, deux cents insulaires s'installeront à l'île Norfolk, et certains reviendront plus tard à Pitcairn. L'épave du Bounty sera découverte à l'extrémité sud de l'île, en 1957.
- Février : Matsudaira Sadanobu (1758-1829), principal acteur du gouvernement du Japon, ordonne que l'on envoie massivement les mendiants (à condition qu'ils soient innocents de tout crime) en exil sur l'île deSado, dans la mer du Japon. D'autres camps de séjours (*yoseba*) seront installés dans d'autres lieux. Les itinérants sont officiellement rassemblés dans ces camps de réhabilitation par le travail pour préparer leur retour dans la société. Il apparaît rapidement que l'exil à Sado revient de fait à une condamnation au bagne, puis à la mort.
- 1er mars, Inde : début de la troisième guerre du Mysore entre Britanniques, Marathes et le nizâm de Hyderâbâd contre le sultan de Mysore Tipû Sâhib, sous le prétexte d'une attaque de ce dernier le 29 décembre 1789 contre un allié hindou de la Compagnie anglaise des Indes orientales (fin en 1792).
- 8 et 28 mars : la Constituante adopte un décret et une instruction qui écartent les colonies du droit métropolitain et crée des assemblées coloniales ouvertes aux propriétaires. Il confirme l'esclavage mais donne l'égalité de droit entre tous les citoyens libres. Le28 mai les Blancs de Saint-Domingue, qui ont élu une assemblée excluant les libres de couleur, votent une Constitution.
- 9 avril : Mulay-el-Yazid succède à son père Sidi Mohammed ibn Abd-Allah comme sultan du Maroc après une existence mouvementée. Ses abus le rendent impopulaire. Il est tué en 1792 en combattant son frère Mulay Hicham.
- 20 juin, Inde : victoire des forces marathes commandées par Benoît de Boigne sur les Rajputs de Jaipur à Patan puis sur ceux de Marwar à Merta le 20 septembre. Domination des Marathes sur les Rajputs et les Jats.
- 7 août : traité de New York entre les États-Unis et 24 chefs de la nation des Creeks.
- 19 - 22 octobre : les Iroquois conduit par le chef Michikinikwa sont victorieux à plusieurs reprises du général Harmar près de Kekionga, dans la région de l'actuel Fort Wayne, au sud du Michigan.
- 28 octobre, Saint-Domingue : soulèvement des Libres de couleur à Port-au-Prince organisé par Vincent Ogé, qui réclame l'exécution du décret du 8 mars ; battus, Ogé et ses partisans sont livrés par les Espagnols et exécutés le 26 février 1791.
- Novembre : éruption du Kīlauea à Hawaï.
- Japon : interdiction de tout autre enseignement que celui du confucianisme (doctrine de Zhu Xi, 1130-1200). La censure est renforcée et il est interdit de publier quoi que ce soit sur les défauts de l'administration.
- À la mort de Ngolo Diarra, roi de Ségou, ses fils, associés au pouvoir de son vivant, s'affrontent pour recueillir sa succession. Finalement, son second fils, Monzon Diarra, devient roi (1790). Il

multiplie les campagnes contre les Mossis, dans le royaume bambara du Kaarta (1796), le Bélédougou et le Fouladougou, et laisse le souvenir d'un grand organisateur. Après sa mort en 1808 commence le déclin de l'Empire de Ségou.
- Domoni sur l'île d'Anjouan est détruite par les pirates malgaches.

Cet exemple illustre bien cette absence des traits de ponctuation et de trou syntaxique : le titre de section (titre-amorce) constitue l'amorce de la SE qui suit, constituée de 13 items. On est ici en attente d'énumération, puisque le titre de section proposé ici suppose que l'on va lister les différents évènements en question, sans qu'aucun marqueur de ponctuation (deux points) ne soit requis, et sans la présence d'une incomplétude syntaxique. Parfois, l'énumération n'est pas forcément attendue de la part du lecteur, et l'amorce ne peut être identifiée comme telle qu'en se référant à la SE qui suit. Cette SE peut elle-même être notamment repérée grâce à des Marqueurs d'Intégration Linéaire, par exemple :

(25) Vont être opposés à ces « droits formels » des « droits substantiels » tels que le « droit au bonheur, à la santé, à la culture ». Il en résultera **tout d'abord** les despotismes éclairés, **puis** les dictatures et **enfin** la construction de l'état socialiste tel celui déterminé par les Constitutions Staliniennes qui revendiquent ces droits de l'homme

iii. Clôtures

A l'image des amorces, les clôtures sont également des éléments facultatifs aux SE, et leur définition varie selon les études. Ainsi certaines études définissent la clôture comme le dernier item d'une série, introduite par des éléments comme *…, enfin* ou encore *tout ça*, à l'oral. D'autres études en revanche tendent à considérer comme clôture l'élément qui vient suivre le dernier item d'une liste qualifiée de fermée, mais qui est cependant en rapport avec cette liste.

A la suite de Rebeyrolle et Péry-Woodley (2014), nous admettrons ici que la clôture correspond à l'élément qui suit le dernier item d'une série. En effet, si l'on parle d'amorce pour désigner le segment qui peut précéder les SE, parler de clôture pour désigner le segment qui peut suivre ces structures s'inscrit dans la continuité du raisonnement. De plus, en parlant de clôture pour désigner le dernier item d'une série sans nommer les autres, cela revient à considérer que ce dernier item serait plus intéressant ou aurait plus de valeur que les autres, alors qu'il se situe sur le même plan. Malgré tout, il est possible de se poser quelques questions sur le terme de « clôture », puisque nous verrons un peu plus loin qu'un des rôles de la clôture est de préparer la réutilisation des informations des SE dans la suite du discours, ainsi si cet élément vient effectivement bien clore une liste d'items, on peut également considérer qu'il « annonce » la suite du discours en permettant la reprise du topique; le lecteur serait donc dans l'attente si cette clôture ne servait justement qu'à clore une liste d'items, on peut donc penser que c'est l'utilisation de ce terme qui pose problème et entraîne des désaccords selon les études.

La clôture, contrairement à l'amorce, n'est pas un élément essentiel dans l'organisation des SE, preuves en sont les rares exemples que l'on peut trouver au sein de corpus où les SE ont été annotées : moins de 15% de toutes les structures annotées, selon Rebeyrolle et Péry-Woodley (2014). Pourtant, cet élément est relativement intéressant à étudier du fait qu'il permet d'observer ce qui va être fait de la SE dans la suite du discours. C'est d'ailleurs ce qui explique que cet élément ne soit pas indispensable, puisqu'on ne réutilise pas toujours les informations fournies dans les SE par la suite. Le rôle de la clôture varie selon la présence ou non d'une amorce :

- Absence d'une amorce : c'est la clôture qui nomme le critère de co-énumérabilité et permet de mettre en évidence la structure de l'énumération qui précède :

(26) **L'Allemagne** suivit dès le début des années 30 une politique différente des recettes de l'orthodoxie libérale dominante à l'époque. Sous la responsabilité financière de Herr Schacht elle se lance dans une politique d'investissement massif, au départ principalement avec des objectifs civils. Galbraith écrira dans son livre sur " la monnaie " que la politique allemande fut à cette époque une politique keynésienne complète avant l'heure. La doctrine de Keynes est en

effet qu'il faut rétablir par une politique d'investissement public l'équilibre perdu entre épargne et investissement. C'est de cette époque que date le réseau d'autoroutes allemand (dont l'équivalent en France ne sera construit que trente ans plus tard). Cette politique est menée sans aucune inflation, ce qui vaudra une réputation durable au ministre des finances, malgré son rôle dans l'appareil nazi. Le plein emploi est quasiment revenu avant même qu'Hitler oriente l'économie allemande vers la production militaire, qui d'ailleurs, est largement réalisée ... en Union Soviétique pour contourner les traités. le Pacte Germano-Soviétique a été précédé par une longue et secrète coopération militaire. **Il va de soi que l'Allemagne sortira de la guerre ruinée.**

La situation est différente en **Italie** où l'exemple allemand n'est suivi que très partiellement et où les aventures coloniales extérieures absorbent une partie importante de l'énergie nationale**. Elle sortira également de la guerre ruinée.**

La France, principalement agricole, subit la crise de plein fouet, les exportations étant pratiquement arrêtées. Elle se replie sur son Empire. Les troubles sociaux et politiques qui aboutissent au Front populaire ne permettent pas l'élaboration d'une politique constante. Alfred Sauvy dans son " histoire économique de la France entre les deux guerres " constate que les " quarante heures " bloque la reprise qui commençait à se manifester. L'effort de production militaire est tardif et n'a qu'une influence marginale sur l'activité. **La France sortira de la guerre pillée et ruinée.**

La situation est peu ou prou la même au **Royaume Uni** qui a tenté de revenir à un taux de change en or intenable pour la Livre avant même 1929 et qui a connu une stagnation plus longue que les autres. La politique d'armement ne commence vraiment que très peu de temps avant la guerre et ne peut être considérée comme la méthode qui a permis de sortir de la crise. **Elle sortira de la guerre victorieuse mais ruinée.**

Le Japon connait une période d'avant-guerre très différente des démocraties du fait de son expansionnisme militaire et de l'encadrement rigoureux de la population. Elle manque de pétrole pour ses entreprises. La guerre avec les Etats-Unis sera largement provoquée par l'embargo décidé par ce pays sur les exportations pétrolières vers le Japon**. Le pays sortira ruiné par la guerre.**

Les États-Unis connurent une période de forte activité pendant la guerre de quarante avec le retour au plein emploi, la mobilisation des hommes jeunes étant compensée par le recours massif à la main d'œuvre féminine dans les usines d'armement. D'énormes investissements furent faits dans beaucoup de domaines qui donnèrent un avantage technologique au pays après-guerre. Lorsque la guerre arriva à son terme, le retour des millions de soldats dans leurs foyers imposa une période de réajustement de l'économie. C'est cette transition qu'était censée faciliter le G.I. Bill. **Au total ce fut le seul pays important à ne pas sortir ruiné de la guerre**. La guerre avait également permis à des économistes keynésiens, sous l'influence de Hansen, de peupler l'administration qui se dote pendant la période des moyens en hommes, en idées et en droit, de son action. La paix retrouvée ils mirent en place une politique de dépense publique qui ne se relâchera plus.

Ces exemples montrent que la montée vers la guerre ne sera nulle part le secret de la fin de la crise de 1929. La guerre marquera une rupture dans les mentalités, provoquera un besoin de

reconstruction intense pendant une dizaine d'année, provoquera une concentration du pouvoir économique dans l'Etat qui est désormais partout chargé du droit au travail et à la sécurité sociale. Tous les gouvernements deviennent " keynésiens ". L'orthodoxie d'avant 1929 est morte. (*Rebeyrolle et Péry-Woodley, 2014*)

En l'absence d'une amorce dans l'exemple (22), c'est la clôture qui se charge d'expliciter le critère de co-énumérabilité, à savoir *exemples*, au moyen d'un syntagme nominal démonstratif placé au début de la clôture.

- Présence d'une amorce : la clôture mentionne une nouvelle fois le critère de co-énumérabilité, déjà cité dans l'amorce :

(27) **Ces variables** sont combinées au moyen de connecteurs logiques qui sont, par exemple :

1. le connecteur binaire disjonctif (ou), de symbole : ∨ ;
2. le connecteur binaire conjonctif (et), de symbole : ∧ ;
3. le connecteur binaire de l'implication, de symbole : → ;
4. le connecteur unaire ou monadique de la négation (non), de symbole : ¬.

Ces variables forment alors des formules complexes.

Dans le cas de cet exemple (23), le syntagme nominal *ces variables*, qui exprime le critère de co-énumérabilité est le même en amorce qu'en clôture, ainsi cette dernière permet de mentionner une nouvelle fois le critère de co-énumérabilité.

(28) **Le comportement de cette machine peut être décrit comme une boucle :**
Elle démarre son exécution dans l'état e1, remplace le premier 1 par un 0.
Puis elle utilise l'état e2 pour se déplacer vers la droite, en sautant les 1 (un seul dans cet exemple) jusqu'à rencontrer un 0 (ou un blanc), et passer dans l'état e3.
L'état e3 est alors utilisé pour sauter la séquence suivante de 1 (initialement aucun) et remplacer le premier 0 rencontré par un 1.
L'état e4 permet de revenir vers la gauche jusqu'à trouver un 0, et passer dans l'état e5.
L'état e5 permet ensuite à nouveau de se déplacer vers la gauche jusqu'à trouver un 0, écrit au départ par l'état e1.
La machine remplace alors ce 0 par un 1, se déplace d'une case vers la droite et passe à nouveau dans l'état e1 pour une nouvelle itération de la boucle.
Ce processus se répète jusqu'à ce que e1 tombe sur un 0 (c'est le 0 du milieu entre les deux séquences de 1) ; à ce moment, la machine s'arrête.

Dans le cas de cet exemple (24) en revanche, contrairement à l'exemple précédent, le syntagme qui correspond à l'énuméraThème n'est pas le même entre l'amorce et la clôture : en amorce, on trouve *le comportement*, tandis qu'en clôture, on trouve *ce processus*, ainsi le scripteur a choisi de nommer différemment un même critère, ce qui lui permet de catégoriser différemment sa liste d'items.

iv. Conclusion

Les structures énumératives permettent de structurer le discours en listant des items, à l'aide de critères typo-dispositionnels ou à l'aide de parallélismes syntaxiques. Ces SE peuvent comporter des éléments facultatifs, à savoir l'amorce et la clôture, qui permettent le plus souvent au lecteur de distinguer le critère de co-énumérabilité, qui est ainsi explicitement cité. Le plus souvent, ce critère de co-énumérabilité est exprimé à l'aide d'un syntagme nominal, parfois démonstratif, qui peut avoir un rôle cataphorique ou anaphorique, selon qu'on le trouve en amorce ou en clôture. A la tête de ce syntagme nominal, on trouve le plus souvent des noms sous-spécifiés, précédemment décrits, qui viennent prendre leur sens grâce aux items de la SE.

Une autre motivation pour l'étude de ces NSS vient du fait que l'hypothèse de certaines étude était que le critère de co-énumérabilité présent en amorce et/ou en clôture était un hyperonyme, et que les items de la liste étaient des hyponymes, or ces études ont dégagé que cette hypothèse ne se vérifiait pas et que l'énuméraThème était le plus souvent un NSS.

Observons la structure suivante :

(29)
- Poireaux
- Pommes de Terre
- Carottes
- Navet
- Courgette
- Tomate

Tous ces légumes peuvent servir à faire une soupe.

Ici, l'énuméraThème *légumes* est un hyperonyme, et chaque item est un hyponyme, puisque tous appartiennent à la classe sémantique des légumes. Contrairement à ce que pensaient certains chercheurs au début de leurs études, ces SE dont le critère de co-énumérabilité est un hyperonyme ne sont pas majoritaires et ce sont bien souvent des NSS qui permettent de résumer les informations de l'antécédent (Rebeyrolle et Péry-Woodley, 2014), d'où l'intérêt d'en proposer une étude approfondie.

CHAPITRE II : PRESENTATION ET EXPLOITATION DES DONNEES

I. Présentation des données

1. Le corpus

L'objectif est d'étudier les NSS dans les SE à travers deux corpus, à savoir le corpus ANNODIS, constitué à l'Université de Toulouse et le corpus de Littéracie Avancée (LA) constitué à l'Université de Grenoble. Le corpus ANNODIS_me a été choisi pour ses annotations de SE. Le corpus de Littéracie avancée, quant à lui, a été choisi pour étendre les résultats observés sur ANNODIS à un ensemble plus grand, en s'intéressant à des écrits d'étudiants, de la licence au master, ce qui permettra d'observer le comportement des NSS selon le genre littéraire. De plus, ce corpus pourrait permettre d'observer l'évolution de la maîtrise des NSS selon les différents niveaux.

i. Corpus issu du Projet ANNODIS

Le corpus ANNODIS est un « corpus de français écrit enrichi d'annotations discursives » (Ho-Dac et al. 2010), qui vise à faciliter l'étude de l'organisation du discours, à travers un large corpus annoté manuellement. Il est relativement diversifié, puisqu'il est composé à la fois d'articles issus de Wikipédia, du quotidien l'Est Républicain, d'Actes du Congrès Mondial de Linguistique Française 2008 (CMLF08) et de rapports de l'Institut Français de Relations Internationales.

Le corpus comprend au total 168 textes, pour 687 000 mots et il s'agit de données numérisées, accessibles via le site du CLLE-ERSS à l'adresse suivante : http ://redac.univ-tlse2.fr/corpus/annodis/ . Tous les textes ont été annotés à la main, selon un protocole d'annotation, disponible à la même adresse Web que ci-dessus. Deux types d'annotations ont été extraits:

- Une annotation en relation rhétorique
- Une annotation en structures multi-échelles : chaînes topicales et structures énumératives

Pour ce mémoire, seul le sous-corpus ANNODIS_me, qui comprend les textes annotés en structures multi-échelles a été conservé, il est composé des textes suivants :

Corpus	**Nombre d'articles**	**Nombre de mots**
WIKI (Wikipédia)	30	231 000
LING (CMLF08)	25	169 000
GEOP	32	266 000
TOTAL	87	666 000

Tableau 1 Composition du corpus ANNODIS_me

Ces annotations ont été réalisées avec l'interface GLOZZ, qui a été développée pour les besoins du projet ANNODIS. Les textes ont subi une annotation qualifiée de *naïve*, par trois étudiants en sciences du langage, puis une annotation *experte*. Cette annotation a été établie en suivant un protocole précis, décrit à la même adresse que celle citée plus haut. Un prétraitement automatique avait été mis en place, à l'aide de l'étiqueteur morphosyntaxique TreeTagger ainsi que de l'analyseur syntaxique SYNTEX (Bourigault, 2007), de façon à faciliter l'annotation manuelle par la suite.

Les annotations sont au format XML, mais nous trouvons également les textes bruts au format .ac. Notons que le site web hébergeant le projet ANNODIS possède aussi une interface pour naviguer au sein du corpus. On peut sélectionner le type de structure que l'on recherche (SE ou chaîne topicale), le sous-corpus qui nous intéresse, le type de SE, selon les typologies décrites dans la partie description des SE de ce mémoire, le nombre d'items, la présence d'amorces ou de clôtures et enfin le texte. Cet outil se présente sous la forme suivante :

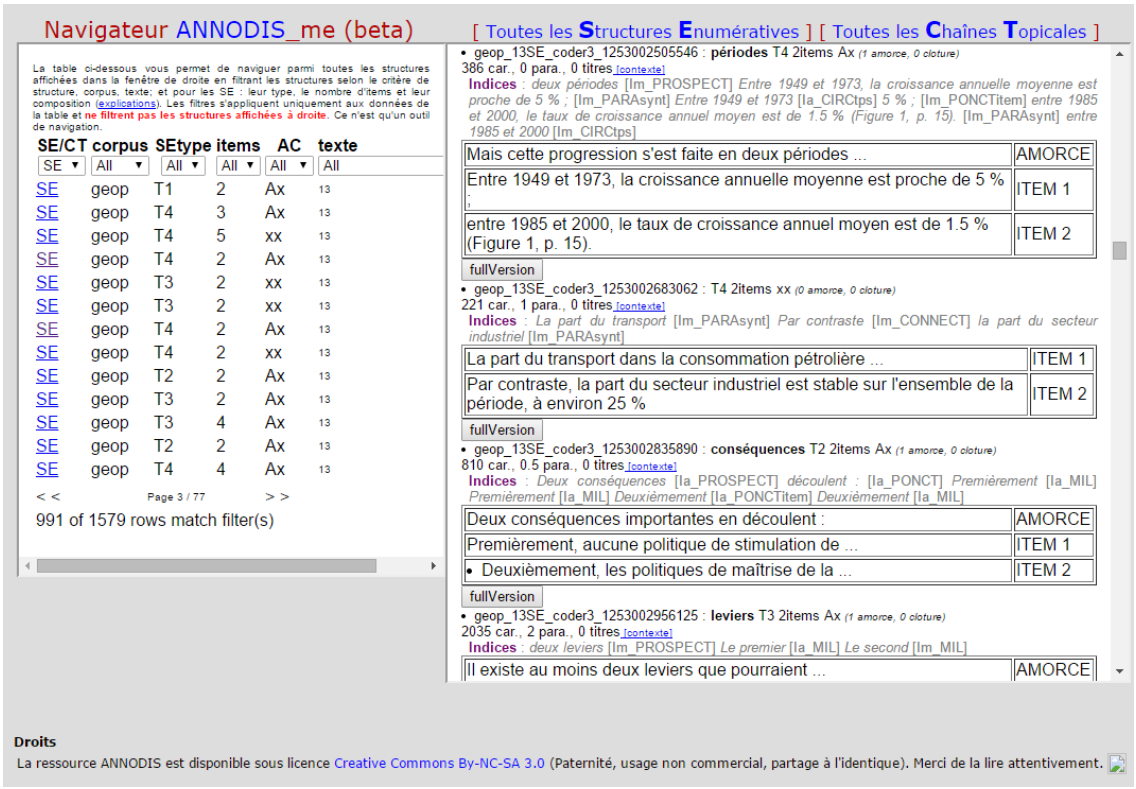


Figure 2 Navigateur ANNODIS

Une fois les informations souhaitées sélectionnées (dans la partie gauche de l'interface), nous pouvons choisir la SE qui nous intéresse, afin d'en afficher un aperçu avec les éléments qui la constituent : amorce, clôture, nombre d'items ou structure complète. Bien que limitée pour une étude plus poussée, cette interface permet de naviguer rapidement au sein des corpus. De plus, certaines données peuvent déjà être fournies par le navigateur, par exemple après avoir choisi uniquement l'affichage des SE, ce dernier nous indique que les corpus contiennent 991 SE annotées. Sur ces 991 structures, une nouvelle requête permet d'apprendre que 102 structures possèdent une amorce et une clôture, 637 possèdent seulement une amorce, 29 possèdent uniquement une clôture et 223 ne contiennent aucun de ces deux éléments facultatifs.

Ainsi la version Web du projet ANNODIS fournit une première description de l'objet de notre mémoire même si nous pousserons l'analyse de ces corpus plus loin. De plus, les annotations étant déjà présentes, cela permet un fort intérêt pour l'étude des NSS au sein des SE puisque nous n'aurons pas à annoter ces structures manuellement.

ii. Corpus de Littéracie avancée

Pour compléter les données du projet ANNODIS et offrir une possibilité d'élargissement de l'étude, nous utiliserons également le corpus de *Littéracie avancée*[1]. Le corpus *Littéracie avancée* est un corpus qui a été constitué à l'Université de Grenoble à partir d'écrits d'étudiants, de la licence 2 au master 2 dans les domaines des sciences du langage, de la didactique du français et des sciences de l'éducation. Il est composé de 11 sous-corpus décrit dans le tableau2 :

Genre	Niveau	Discipline	Nb de textes	Nb moyen de mots par texte
Dossier	L2	Sciences du langage	10	2809
Partie théorique de rapport de stage	L3	Didactique du français (licence sciences du langage)	15	4677
Analyse (sujet type CRPE – français)	M1	Enseignement 1er degré (Français)	69	1333
TER	M1	Enseignement éducation médiation 1er degré	25	10974
Compte-rendu de lecture	M1	Enseignement éducation médiation 1er degré	10	631
Compte-rendu de lecture	M2	Formateurs d'enseignants 1er et 2nd degré	10	3570
Compte-rendu professionnel	M1 et M2	Master Ecrifore (écriture formation remédiation)	22	220
Lettres de motivation	M1 et M2	Master Ecrifore	20	320
Parties théoriques de mémoires	M1	Didactique du français (Master)	10	3530
Mémoires	M1 et M2	Didactique du français (Master)	41	9645
Synthèses théoriques	M2	Didactique du français (Master)	10	1568
		TOTAL	242	39 277

Tableau 2 Composition du corpus de Littéracie avancée

Ce corpus contient au total 242 textes et 963 897 mots. Les SE n'y sont certes pas annotées manuellement, mais nous pourrons, à l'aide de ce corpus, comparer l'emploi des NSS dans un genre

[1] (Disponible sur : http ://lidilem.u-grenoble3.fr/ressources/corpus-du-labo/article/corpus-litteracie-avancee)

textuel différent de ceux d'ANNODIS. De plus, ce corpus pourrait nous permettre de faire le parallèle avec les résultats obtenus par Nur Aktas et Cortes (2008) dans leur étude sur l'utilisation des noms sous-spécifiés dans des écrits d'étudiants, et ce bien qu'il s'agisse d'étudiants ayant l'anglais comme deuxième langue. Pour leur étude, les auteurs ont pu comparer l'utilisation de NSS en fonction de plusieurs types d'écrits, et en fonction de la langue maternelle de leurs sujets (certains avaient l'anglais pour langue maternelle, d'autres parlaient d'autres langues mais écrivaient en anglais pour les besoins de leurs écrits universitaires). Ils ont par exemple pu relever que le NSS *fact*, utilisé dans un patron avec fonction cataphorique *N+clause* est très souvent utilisé par les auteurs d'articles publiés. En revanche, il ne l'est que très peu dans les écrits d'étudiants, qui semblent moins à l'aise avec cette utilisation. Il pourrait donc être intéressant de vérifier, pour le français, si l'utilisation des NSS est la même en fonction des différents genres textuels ou des différents niveaux de littéracie des scripteurs.

L'exploitation de ce corpus, combinée à celle d'ANNODIS devrait nous permettre de comparer l'emploi des NSS et de leurs antécédents en fonction du genre de textes.

2. Etiquetage morphosyntaxique

i. Choix de l'outil

Afin de pouvoir observer les distributions de NSS dans le corpus retenu, il convenait dans un premier temps de réaliser un étiquetage morphosyntaxique de ces derniers, notamment pour y projeter des patrons de construction syntaxique dans lesquels sont susceptibles d'apparaître des NSS. Pour réaliser cet étiquetage, de nombreux outils sont disponibles, tels que Tree Tagger, Talismane ou LIA tagg. Nous avons choisi d'utiliser l'outil Talismane, développé par Assaf Urieli (2013). Cet outil a été retenu pour sa robustesse et parce qu'il est entièrement configurable. Talismane permet la segmentation en phrases, le découpage en mots, l'étiquetage et le parsing. Il contient un ensemble d'options qui permettent de l'adapter à une diversité de situations. Il permet notamment de calculer l'offset d'un mot, c'est-à-dire ses coordonnées, ce dont nous avions besoin dans la démarche adoptée pour ce mémoire.

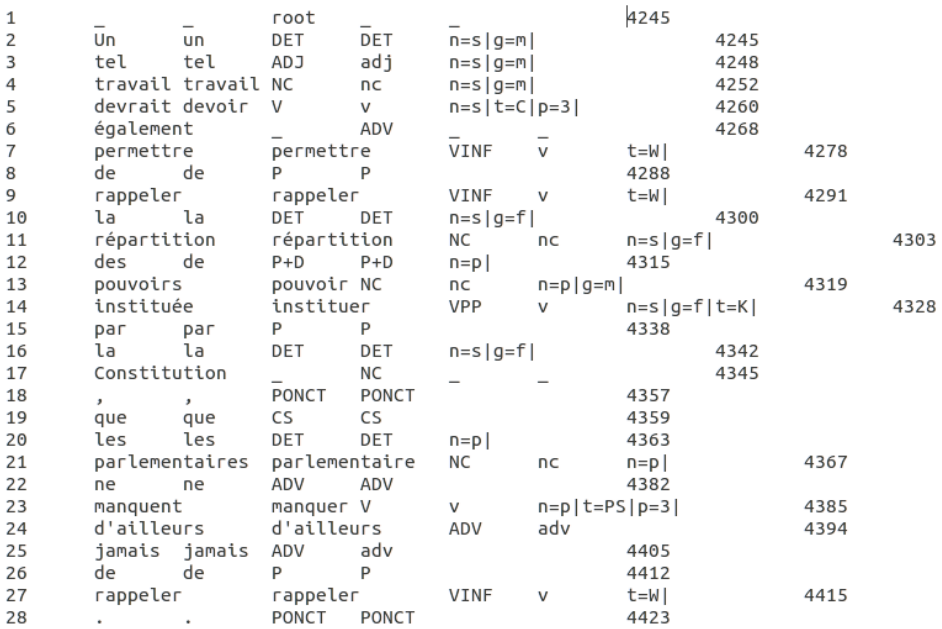
ii. Etiquetage avec Talismane

L'étiquetage morphosyntaxique a permis d'obtenir en sortie des fichiers composés de la manière suivante :

N° du mot dans la phrase	Forme de surface	Lemme	Etiquette morphosyntaxique Talismane	Catégorie morphosyntaxique du lexique LEFFF	Traits morphologiques du lexique LEFFF	Offset

Tableau 3 Format de sortie des fichiers étiquetés avec Talismane

Concrètement, l'étiquetage de la phrase *Un tel travail devrait également permettre de rappeler la répartition des pouvoirs instituée par la Constitution, que les parlementaires ne manquent d'ailleurs jamais de rappeler.* donne le résultat suivant, une fois étiqueté avec Talismane (Corpus ANNODIS_me) :



1	_	_	root	_	_	4245		
2	Un	un	DET	DET	n=s|g=m|	4245		
3	tel	tel	ADJ	adj	n=s|g=m|	4248		
4	travail	travail	NC	nc	n=s|g=m|	4252		
5	devrait	devoir	V	v	n=s|t=C|p=3|	4260		
6	également		_	ADV	_	_	4268	
7	permettre		permettre	VINF	v	t=W|	4278	
8	de	de	P	P		4288		
9	rappeler		rappeler	VINF	v	t=W|	4291	
10	la	la	DET	DET	n=s|g=f|	4300		
11	répartition		répartition	NC	nc	n=s|g=f|	4303	
12	des	de	P+D	P+D	n=p|	4315		
13	pouvoirs		pouvoir	NC	nc	n=p|g=m|	4319	
14	instituée		instituer	VPP	v	n=s|g=f|t=K|	4328	
15	par	par	P	P		4338		
16	la	la	DET	DET	n=s|g=f|	4342		
17	Constitution		_	NC	_	_	4345	
18	,	,	PONCT	PONCT		4357		
19	que	que	CS	CS		4359		
20	les	les	DET	DET	n=p|	4363		
21	parlementaires		parlementaire	NC	nc	n=p|	4367	
22	ne	ne	ADV	ADV		4382		
23	manquent		manquer	V	v	n=p|t=PS|p=3|	4385	
24	d'ailleurs		d'ailleurs	ADV	adv		4394	
25	jamais	jamais	ADV	adv		4405		
26	de	de	P	P		4412		
27	rappeler		rappeler	VINF	v	t=W|	4415	
28	.	.	PONCT	PONCT		4423		

Figure 3 Résultat de l'étiquetage morphosyntaxique

II. Objectifs de l'étude

Etape 1 : Projection de patrons syntaxiques et constitution d'une liste de NSS potentiels

Un des premiers objectifs de cette étude sera d'extraire les NSS des deux corpus, à partir des six patrons de construction de NSS qui ont été décrits plus haut et sont repris ici :

Construction décrite par Legallois (2008) :

Det N [Ø | ce] être [que-clause | de-inf]

Constructions mises en évidence par Roze et al. (2014) :

[V avoir] [P pour] [NSS] [P de]

[P pour] [NSS] [P de]

[NSS] [Ponct :]

[NSS] [CS que]

[NSS] [P de] [V inf]

Cette étape permettra de fournir une méthode d'extraction de patrons, qui pourra être réutilisée par la suite mais également d'obtenir notre propre liste de NSS potentiels (désormais NSSP), afin de pouvoir la comparer à Legallois (2008). Il a été décidé de projeter ces patrons et non pas la liste déjà proposée par Legallois puisque les NSS sont des noms particuliers, qui ne sont sous-spécifiés qu'au sein de certaines constructions, ainsi un même nom peut à la fois être non sous-spécifié dans une construction donnée (exemple 30), et à la fois sous-spécifié dans une autre construction (exemple 31). L'exemple 30 fournit une illustration du nom urgence employé dans un contexte classique, dans lequel le nom n'est

pas utilisé dans son emploi sous-spécifié, tandis que dans l'exemple 31, le contenu propositionnel d'*urgence* est donné par *récupérer nos affaires* (Legallois, 2008).

(30) Autant dire que le non-interventionnisme économique de l'Etat fédéral a été immédiatement relégué au second rang devant l'**urgence** de la situation.

(31) L'**urgence** est de récupérer nos affaires

Une fois ces patrons projetés, une liste de NSSP aura été extraite, et il conviendra de la modifier pour n'obtenir que des NSSP intéressants. Cette liste sera accompagnée de la fréquence du NSSP au sein du corpus et du degré de *reliance*, c'est-à-dire de la proportion d'occurrences de ce nom dans un patron.

Etape 2 : Distribution des NSSP

L'objectif premier de cette seconde étape sera de comparer les distributions de la liste de NSSP précédemment obtenue dans le corpus ANNODIS_me et dans les SE. Cette étape consistera donc à projeter la liste de NSSP afin de mesurer la fréquence de ces noms au sein du corpus puis dans les SE, et plus particulièrement dans les amorces et les clôtures. Il s'agira alors de vérifier si les NSS projetés sont plutôt concentrés dans les SE ou s'ils y sont significativement moins nombreux..

Etape 3 : Annotation manuelle de NSS et description de nouveaux patrons

Afin de pouvoir découvrir et décrire de nouvelles constructions qui ne l'auraient pas été par Legallois (2008) et Roze et al. (2014) et qui seraient présentes au sein des amorces et clôtures des SE, il conviendra d'annoter manuellement les NSS et leurs patrons de constructions. Pour des raisons liées au temps, seules les amorces et les clôtures des SE de type 3 seront annotées. Rappelons que les SE de type 3 correspondent à celles qui couvrent plus d'un paragraphe, sans marques visuelles spécifiques.

21 fichiers ont été analysés (7 dans chacun des trois sous-corpus qui constituent ANNODIS_me) et les NSS ainsi que leurs patrons ont été annotés dans 76 SE.

Etape 4 : Projection des nouveaux patrons obtenus

L'objectif de cette dernière étape est de vérifier s'il est possible de repérer automatiquement des SE à l'aide des NSS. Pour ce faire, les nouveaux patrons obtenus lors de l'étape précédente seront projetés, selon la méthode utilisée lors de la première étape. Dans un premier temps, ces nouveaux patrons seront projetés sur ANNODIS_me afin d'observer le nombre de fois où ces patrons permettent d'extraire des NSS au sein des SE. Dans un second temps, les patrons seront projetés sur le corpus de littéracie avancée afin de vérifier que les résultats obtenus sur ANNODIS_me peuvent être similaires sur un corpus dans lequel les SE n'ont pas été annotées : les NSS extraits permettent-ils de repérer des SE ?

Cette projection permettra donc de :

- Vérifier si ces nouveaux patrons permettent d'extraire de nouveaux NSS (en reprenant la méthode de l'étape 1) et si ces patrons sont attractifs (selon la mesure d'attraction de Schmid)
- Vérifier si ces patrons sont propres aux SE ou non, à l'aide des annotations du corpus ANNODIS_me
- Vérifier si ces patrons permettent de repérer des SE dans un corpus où elles n'auraient pas été annotées (littéracie avancée)

III. Résultats

1. Projection de patrons syntaxiques et constitution d'une liste de NSSP

i. Description des patrons

Afin de reproduire en partie la démarche de Legallois (2008) et Roze et al. (2014), il convient de projeter les patrons mis en évidence sur nos propres corpus. L'objectif est de comparer la liste de noms que nous obtiendrons à celle qu'avait pu obtenir Legallois (Roze et al. (2014) reprenant uniquement les 20 NSSP les plus fréquents chez Legallois pour découvrir de nouveaux patrons à travers la fouille de textes, nous ne nous intéresserons pas ici à cette liste). Les patrons correspondent à des constructions syntaxiques dans lesquelles des NSS sont susceptibles d'apparaître. Ainsi ces patrons permettent de supposer qu'un nom qui peut être à la fois non sous-spécifié et à la fois sous-spécifié sera plus probablement sous-spécifié dans un contexte donné.

Rappelons que les patrons qui sont projetés sont les suivants, avec chaque fois un exemple qui permet de l'illustrer (La construction de Legallois (2008) [DET] [NC] [Ø|ce][V $etre] [CS que|P de] a été divisée en deux patrons, respectivement les patrons 1 et 2 pour plus de facilités lors de l'extraction automatique, les autres constructions ont été décrites par Roze et al.) :

ID du patron	Patron	Exemple	NSSP
Patron 1	[DET]$_1$ [**NC**]$_2$ [ce]$_3$ [V être]$_4$ [CS que|P de]$_5$	[Le]$_1$ [**problème]**$_2$ [c']$_3$ [est]$_4$ [que]$_5$ la marche à suivre pour réussir ces évaluations est rarement enseignée	Problème
Patron 2	[DET]$_1$ [**NC**]$_2$ [Ø][V être]$_3$ [CS que|P de]$_4$	[Le]$_1$ [**but]**$_2$ [serait]$_3$ [de]$_4$ privilégier une écriture spontanée	But
Patron 3	[DET]$_1$ [**NC**]$_2$ [CS que]$_3$	[Le]$_1$ [**fait]**$_2$ [que]$_3$ chaque lecteur soit unique fait de la lecture un acte personnel.[est]$_4$ [que]$_5$ la marche à suivre pour réussir ces évaluations est rarement enseignée	Fait
Patron 4	[**NC**]$_1$ [P de]$_2$ [V inf]$_3$	On note ici l'[**importance]**$_1$ [de]$_2$ [transmettre]$_3$ un maximum de mots aux enfants	Importance
Patron 5	[**NC**]$_1$ [PONCT :]$_2$	nous allons aborder la [**question]**$_1$ [:]$_2$ sous quelles conditions et quels enjeux les poètes du 20ème siècle travaillent-ils leur art ?	Question
Patron 6	[P pour]$_1$ [**NC**]$_2$ [P de]$_3$ [V inf]$_4$	il y a une tendance analytique qui a [pour]$_1$ [**consigne]**$_2$ [de]$_3$ [retravailler] $_4$ les œuvres passées en utilisant d'autres concepts	Consigne

Tableau 4 Patrons sélectionnés pour l'extraction automatique de NSSP

Tous les exemples pour illustrer les patrons proviennent du corpus littéracie avancée et ont été extraits à l'aide de la projection de ces patrons, dont le programme est disponible en annexe 6.

ii. Méthode de projection des patrons

L'ensemble des 6 patrons décrits plus haut a été projeté sur les deux sous-corpus, à savoir ANNODIS_me et LA. La répartition des NSSP extraits en fonction du patron dans lequel ils apparaissent est la suivante :

	ANNODIS_me		**LA**		**TOTAL**	
	Fréquence absolue	%	Fréquence absolue	%	Fréquence absolue	%
Patron1	0	0	5	0.2	5	0.1
Patron 2	94	10	242	10	336	10
Patron 3	212	23	480	20	692	21
Patron 4	383	41	1 216	52	1599	49
Patron 5	195	21	296	13	491	15
Patron 6	45	5	97	4	142	4
Total	929	100	2336	100	3265	100

Tableau 5 Répartition des NSSP au sein des patrons

Si le patron 4 est très représenté, d'autres en revanche ne le sont que très peu, c'est le cas des patrons 1, 2 et 6. Notons également que la construction proposée par Legallois (2008) (i.e les patrons 1 et 2 ici) n'est pas la plus fréquente dans nos corpus, bien qu'elle permette de repérer un NSS, ce qui a permis à l'auteur d'établir sa liste. De prime abord, une hypothèse pour expliquer ces résultats serait qu'ils dépendent du corpus utilisé. Ici nous avons deux corpus très diversifiés, au sein desquels la distribution de la construction de Legallois est relativement la même. Ainsi, et malgré le fait que certains patrons soient très représentés et d'autres pas, on note une distribution à peu près équivalente de chacun de ces patrons dans les deux sous-corpus, ce qui peut être facilement repérable à l'aide de la figure 4 :

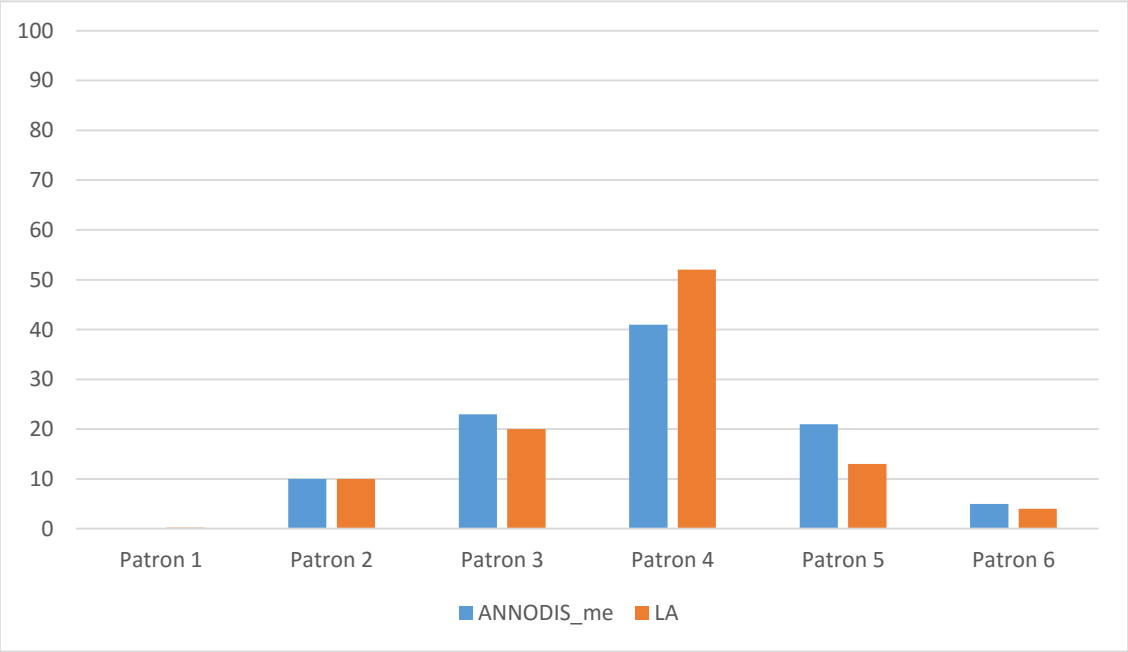


Figure 4 Répartition des patrons dans les deux sous-corpus

Le patron 2 (qui est en fait une sous-partie du patron 1) n'est presque pas présent, que ce soit dans ANNODIS_me ou dans LA. Même en additionnant les résultats du patron 1 et du patron 2 afin de n'avoir plus qu'un seul patron qui correspond à celui de base proposé par Legallois, les résultats restent très faibles par rapport à certains autres patrons. Seul le patron 5 est un peu plus distribué dans ANNODIS_me que dans LA, toutefois nous observerons qu'il s'agit du patron le moins contraint, qui ramène par conséquent le plus de résultats non pertinents.

La plupart des patrons décrits par Legallois (2008) et Roze et al. (2014) sont relativement contraints, ce qui permet de limiter le bruit dans les résultats. Seul le patron 5 ([NC] [PONCT :]) ne l'est que très peu puisque tous les noms suivis du signe de ponctuation *:* sont extraits, ce qui explique ces résultats non attendus tels que :

(32) Les enquêtes menées aux Etats-Unis et en France tendent à montrer une vision concordante sur trois types de réactions à la **mondialisation :** l'évaluation du phénomène, ses conséquences économiques, et ses conséquences sociales et culturelles.

Dans cet exemple 32, le nom *mondialisation* est considéré comme un NSSP, or il y a bien un NSS au sein de cette phrase, mais ce n'est pas celui qui est extrait par le patron 5.

Le patron 4 correspond au deuxième patron le moins contraint, et bien que les résultats de son extraction soient encourageants, nous pouvons ici aussi noter du bruit, par exemple :

(33) Plusieurs didacticiens insistent sur la nécessité pour les **élèves** de pratiquer un raisonnement grammatical complet.

En effet dans cet exemple 33, le patron 4 permet d'extraire le NSSP *élève*, alors que c'est *nécessité* qui requiert un contenu propositionnel, donné par *pratiquer un raisonnement grammatical complet*.

Pour chacun des 6 patrons projetés, 10 résultats ont pu être analysés manuellement pour chacun d'eux, (excepté pour le patron 1 qui ne contenait que 5 NSSP, qui ont donc tous été étudiés). Les NSSP alors obtenus ont pu être validés (désormais NSSV) ou invalidés manuellement et ont permis d'obtenir les résultats suivants (Annexe 7) :

- Patron 1 : 4 NSSV sur 5NSSP
- Patron 2 : 7 NSSV sur 10NSSP
- Patron 3 : 8 NSSV sur 10 NSSP
- Patron 4 : 9 NSSV sur 10 NSSP
- Patron 5 : 4 NSSV sur 10 NSSP
- Patron 6 : 10 NSSV sur 10 NSSP

Ces résultats permettent de calculer la précision du système d'extraction des patrons utilisés, qui correspond à :

$$Précision = \frac{Nombre\ de\ NSSV}{Nombre\ de\ NSSP}$$

La valeur de la précision s'élève donc à 0.76, ce qui est un score très encourageant, bien qu'il faille garder en mémoire que seulement 10 résultats pour chacun des patrons ont été validés manuellement, et que le patron 5 extrait beaucoup de bruit du fait de ses faibles contraintes.

iii. Résultats de la projection des patrons

Cette projection de patrons a permis d'extraire de nombreux NSSP qui sont à l'origine d'une liste, à l'instar de celle de Legallois (2008) qui sera utilisée par la suite. La mise en commun des noms extraits dans ANNODIS_me et dans LA ainsi que la suppression des doublons étaient deux étapes indispensables pour la création de cette liste.

	ANNODIS_me + LA
Nombre total de noms extraits	3265
Nombre total de noms après suppression des doublons	935

Tableau 6 Elimination des doublons

De nombreux doublons ont pu être supprimés, ce qui signifie que beaucoup de noms étaient présents plusieurs fois. Cette observation permet d'appuyer l'hypothèse selon laquelle certains noms seraient plus systématiquement sous-spécifiés que d'autres. En effet, certains n'avaient pas de doublons, tels que *voyelle, vecteur, usage*, en revanche d'autres étaient très largement distribués, c'est-à-dire qu'ils apparaissaient très régulièrement au sein de constructions syntaxiques propres au NSS, par exemple *fait*, qui apparaissait 369 fois, ou encore *but*, qui était présent 141 fois.

iv. Constitution de la liste définitive

La liste obtenue a ensuite été comparée à celle qu'avait pu établir Legallois (2008), à l'aide du plugin *Compare*, de l'outil *Notepad++*. Au total, 169 noms sont communs aux deux listes, ainsi la liste de Legallois comporte 158 noms qui ne sont pas présents dans notre liste, et nous avons 766 noms qui ne sont, à l'inverse, pas présents chez l'auteur.

Bien qu'il soit normal que notre liste contienne beaucoup plus de noms, puisque nous avions projeté d'autres patrons, il paraît étonnant de ne pas avoir plus de résultats en commun avec la liste de Legallois. Toutefois, Legallois a extrait les noms de sa liste d'un corpus journalistique, alors que nous avons utilisé des données très diversifiées. Il est donc possible de penser que les structures rencontrées dans des genres de textes différents ne sont pas les mêmes, et que, par conséquent, les noms rencontrés ne sont pas les mêmes, ce qu'il faudra vérifier par la suite. Parmi les noms qui sont présents dans la liste de Legallois et qui ne le sont pas dans la nôtre, nous pouvons par exemple relever *acquis, activité, affiche, auteur, catastrophe*… Ces noms communs aux deux listes et ceux qui sont différents sont tous disponibles dans l'annexe 4. De plus, Legallois ne s'est pas contenté d'extraire des noms communs, il a également extrait des noms modifiés, tels que : *fond du problème, point essentiel, élément important*… En revanche, nous n'avons pas extrait ces cas lors de la projection de nos patrons, ce qui peut expliquer que nous n'ayons pas certains noms.

Cette liste de 935 noms extraits à partir de patrons dans les deux sous-corpus a été projetée sur l'ensemble du corpus ANNODIS_me (programme fourni en annexe 8). Cette projection a permis d'extraire 70 606 occurrences.

Le rapport entre le nombre d'occurrences extraites, qui correspondent à des NSS potentiels, et le nombre total de NC dans le corpus ANNODIS_me, à savoir 147 119 est de 0.48, ce qui pourrait laisser penser

qu'environ la moitié des noms communs correspond à des NSS. Nous savons en revanche que si au moins un tiers de ces noms ont pu être extraits, ils ne sont pas tous sous-spécifiés. En effet, les NSS sont des noms au fonctionnement complexe, et les occurrences extraites correspondent toutes à des noms qui peuvent effectivement être sous-spécifiés dans certaines structures, mais qui ne le sont pas forcément systématiquement. Ainsi, il est plus juste de penser qu'environ 50% des noms du corpus peuvent être utilisés en tant que NSS, c'est-à-dire que 50% de ces noms sont des NSSP.

Considérons les exemples suivants (ANNODIS_me) :

(34) Lorsque je travaillais encore au Sénat, une fondation dont **l'objet** était de [promouvoir le tennis en tant qu'activité périscolaire] m'avait invité […]

(35) Elles se distinguent ainsi de la poste qui transmet des informations ou des **objets** sous forme physique.

Lorsque les patrons à l'origine de la liste créée ont été projetés, le nom *objet* de l'exemple 34 était correctement extrait et s'inscrivait au sein du patron 1, et le nom *objet* de l'exemple 35 n'était lui pas ramené lors de l'extraction. En revanche lors de la projection de la liste, le nom *objet* de l'exemple 35 était extrait, bien qu'il ne soit pas sous-spécifié puisque l'utilisation du lexique permet d'extraire des mots indépendamment de la construction syntaxique dans laquelle ils apparaissent. Pour autant, la projection des NSSP paraît indispensable puisque nous supposons que certains patrons, importants au sein des amorces et des clôtures dans les SE, n'ont pas été décrits par Legallois (2008) ou Roze et al. (2014) ; ces NSSP pourraient donc permettre de mettre en évidence ces nouveaux patrons.

Ainsi le nom *objet* apparaît 261 fois au total dans le sous-corpus ANNODIS, mais il s'inscrit seulement 6 fois au sein des patrons précédemment décrits. En reprenant la formule de Schmid (2000), décrite dans la partie *Objectifs de l'étude* du chapitre II, il est possible de calculer la relation de *reliance* du nom *objet* en mesurant le rapport entre son apparition dans un patron et son apparition totale. La valeur de ce calcul peut aller de 0 à 1, où 1 correspond à une dépendance parfaite, c'est-à-dire qu'un nom relevé au sein de patrons n'apparaît qu'au sein de ces patrons, donc qu'il est systématiquement sous-spécifié au sein du corpus. La valeur de *reliance* du nom objet est de 0.023, elle est relativement faible, puisque Schimd (2000) considère que seuls les noms dont le résultat de ce calcul excède 0.15 sont intéressants en termes de sous-spécification. Ainsi après avoir observé les valeurs de ce calcul, nous pouvons confirmer que le nombre de noms qui sont effectivement sous-spécifiés est largement plus petit que le nombre total de noms, tous les noms n'étant pas sous-spécifiés. Toutefois, certains noms peuvent ne pas être représentés, puisque nous avons projeté une liste qui n'est pas exhaustive, et que certains noms non présents dans la liste peuvent être sous-spécifiés s'ils apparaissent au sein de patrons qui n'ont pas été projetés, ce que nous reverrons par la suite. De plus, les occurrences étant des noms issus d'une liste, certains noms ne sont pas forcément sous-spécifiés bien qu'ils soient tout de même extraits, ce que nous avons pu illustrer avec l'exemple 35.

Il est donc important de noter que les noms issus de la liste qui sont les plus fréquents ne sont pas forcément ceux qui sont le plus fréquemment sous-spécifiés. En effet si l'on regarde les 10 mots les plus fréquents parmi ceux de la liste de l'annexe 3, certains n'apparaissent que très peu dans les patrons qui ont été projetés, et leur *reliance* est donc parfois très faible :

NSSP	Freq absolue	Dans Patrons	Reliance
Langue	720	1	0.0014
Sens	622	4	0.0064
Guerre	605	5	0.0083
Système	559	2	0.0036
Question	507	16	0.0316
Cas	459	4	0.0087
Terme	449	1	0.0022
Point	425	3	0.0070
Force	424	2	0.0047
Texte	394	1	0.0025

Tableau 7 Les 10 NSSP les plus fréquents

Pour constituer la liste définitive de NSSP, un seuil devait donc être fixé. Lorsque Schmid (2000) décrit le calcul de la *reliance*, il fixe ce seuil à 0.15, or il décrit bien plus de patrons syntaxiques que ceux qui ont pu être décrits ici, ainsi ce seuil était trop élevé pour cette étape. En effet, beaucoup de NSSP étaient éliminés, puisque nous passions de 935 NSSP à 43, et que beaucoup parmi ceux qui paraissent très intéressants étaient éliminés, tels que *objectif, exemple*… Afin de pouvoir fixer un seuil plus adapté, les valeurs de *reliance* ont pu être réunies selon un intervalle de 0.005, auxquelles on a associé les effectifs cumulés de NSSP :

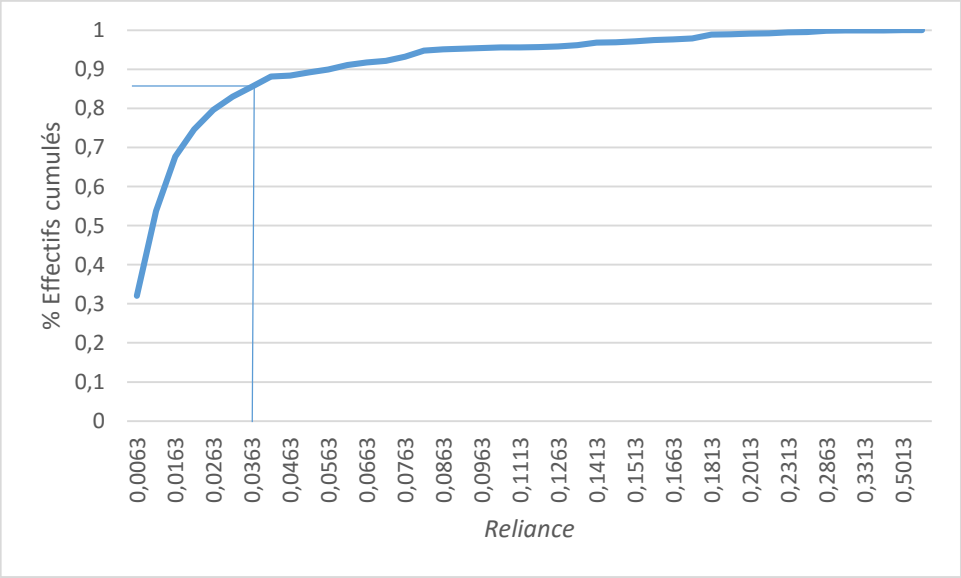


Figure 5 Evolution de la reliance des NSSP

Lorsque la *reliance* atteint 0.036, plus de 80% des NSSP sont déjà représentés, et les valeurs de *reliance* n'augmentent que peu pour les 20% restant. Il a donc été décidé de ne garder que les NSSP dont la *reliance* dépassait 0.036.

Toutefois, un deuxième critère a dû être retenu pour constituer la liste de NSSP définitive. En effet, certains noms n'apparaissent que de très rares fois au sein du corpus, et 10 d'entre eux ont une *reliance* de 1, ce qui signifie que les seuls contextes dans lesquels ils apparaissent sont sous-spécifiés. En effet, la *reliance* étant le rapport entre l'apparition d'un nom dans un patron et son apparition totale dans le corpus, si ce nom n'apparaît qu'une seule fois dans le corpus et qu'il est présent au sien d'un patron, alors sa *reliance* vaut 1. C'est le cas des noms suivants : *congénère, consternation, deuil, mathusalem, nabuchodonosor, rappelé, rétribution, salmanazar* et *sellerie*.

Cependant, ces valeurs sont à nuancer. En effet, le rapport établi permet de dire que si le nom apparaît une fois dans tout le corpus, et que cette seule apparition se fait au sein d'une construction syntaxique propre à la sous-spécification, alors ce nom est systématiquement sous-spécifié, or la conclusion ne peut être aussi nette. Tout d'abord, l'étiquetage de Talismane a laissé quelques erreurs, c'est par exemple le cas pour :

(36) Trent Lott […] et Dennis Hastert […] ont tous les deux **rappelé** […]

Dans l'exemple 36, le bruit est causé par une erreur d'étiquetage de la part de Talismane, qui a étiqueté *rappelé* comme étant un nom commun (étiquette NC).

De plus, une *reliance* dont la valeur atteint 1 ne concerne que des cas ou le nom extrait par les patrons n'apparaît qu'une seule fois, ainsi il faudrait pouvoir observer plus en détail les usages de ces noms au sein d'autres corpus afin de pouvoir généraliser ou non ce qui peut être observé ici. Pour toutes ces raisons, il a donc été décidé, en plus de ne retenir que les noms dont les résultats sont supérieurs ou égaux à 0.036, de ne pas tenir compte des NSSP dont la fréquence au sein du corpus est inférieure ou égale à 3. La liste finale obtenue est constituée des 147 noms suivants :

adepte, air, alternative, ambiance, ambition, anarchiste, appétit, apprenant, autorisation, avis, balance, besoin, bouche, but, capacité, catastrophe, cathare, certitude, cesse, chance, charia, chimie, collaboration, collectivité, conduit, constat, conviction, copie, crainte, décès, décomposition, demie, dépression, désir, devoir, diachronie, dissimulation, écueil, effet, émetteur, espoir, esprit, essentiel, exigence, façon, faculté, fait, fiction, finalité, fonctionnaire, futur, garantie, genèse, geste, globalisation, gloire, habitude, hiatus, honneur, hypothèse, idéal, idée, impératif, impossibilité, impression, incapacité, incertitude, incompatibilité, instant, intention, interdiction, interlocuteur, intuition, inventaire, jeûne, lecteur, lendemain, libéralisation, licence, littérarité, manière, médecine, mérite, métalangage, mondialisation, moyen, nécessité, nomenclature, objectif, obligation, occasion, opportunité, pacifisme, paradoxe, parasite, pari, perception, périphérie, permis, peur, plaisir, polémique, pompe, possibilité, précurseur, preuve, prière, primat, primitif, priorité, problématique, productivité, profession, profondeur, propos, prosodie, quart, question, quotidien, raison, rassemblement, ratio, rayonnement, rédaction, refus, registre, relativité, réputation, rêve, risque, sémiotique, sentiment, soin, sommation, souci, subordination, synchronie, syntaxe, tâche, tentative, texture, transitivité, veau, vecteur, vérité, vocation, volonté

Cette liste complète accompagnée des fréquences de chaque NSSP au sein du corpus, en amorce et clôture, dans des patrons et leur reliance ainsi que leur présence ou non dans la liste proposée par Legallois (2008) est donnée en annexe 5.

2. Distribution des NSSP

i. Répartition des noms dans ANNODIS_me

Le sous-corpus ANNODIS_me est constitué de 666 000 mots, dont 147 119 noms communs (étiquetés NC par Talismane), ce qui représente 24% de tous les mots du corpus. A partir de la liste définitive de NSSP établie précédemment, seuls 4% de ces noms seraient des NSSP et 1.5% de ces NSSP seraient localisés dans les amorces et/ou les clôtures de SE.

	ANNODIS_me
Nombre de noms	147 119
NSSP	6519
Amorces et clôtures	97

Tableau 8 Répartition des noms

Parmi ces NSSP, les 10 plus fréquents sont les suivants :

NSSP	**Fréquence absolue**
Question	507
Fait	372
Effet	295
Moyen	275
Façon	260
Idée	253
Manière	233
Capacité	208
Objectif	195
Raison	195

Tableau 9 Les 10 NSSP les plus fréquents

Ces NSSP les plus fréquents au sein d'ANNODIS_me sont tous présents dans la liste proposée par Legallois (2008), à l'exception de *capacité*. Bien que la méthode de l'auteur soit à relativiser du fait qu'il ne considère que deux constructions syntaxiques, les noms présents dans sa liste sont cependant tous possiblement sous-spécifiés au moins dans les constructions qu'il décrit.

Il paraît d'ailleurs intéressant, à ce stade, de se demander si ces 10 NSSP présentent une *attraction* plus forte pour certains patrons. Il s'agit alors de reprendre la mesure *attraction* de Schmid (2000), décrite dans l'état de l'art, pour chacun des 6 patrons déjà décrits. Rappelons que cette mesure est la suivante :

$$attraction = \frac{fréquence\ d'un\ nom\ dans\ un\ patron}{fréquence\ totale\ du\ patron}$$

	Attraction Patron 1	Attraction Patron 2	Attraction Patron 3	Attraction Patron 4	Attraction Patron 5	Attraction Patron 6
Question	0	0.093	0	0.032	0.043	0
Fait	0	0.070	0.39	0.036	0	0
Effet	0	0	0.007	0.004	0	0.7
Moyen	0	0.023	0	0.007	0	0
Façon	0	0	0	0.036	0	0
Idée	0	0.023	0.191	0.029	0	0
Manière	0	0	0.029	0.018	0.029	0
Capacité	0	0	0	0.029	0	0
Objectif	0	0.140	0	0.007	0	0.172
Raison	0	0	0.029	0.018	0.029	0

Tableau 10 Attraction des 10 NSSP les plus fréquents

Le tableau 10 permet de mettre en valeur le degré d'attraction de certains patrons pour un nom particulier. Par exemple le patron 2 permet d'attirer dans 14% des cas le NSS objectif, le patron 3 permet d'extraire dans 19% des cas le NSS idée, ainsi certains NSS seraient plus attirés par certains patrons, bien que ces résultats puissent éventuellement être valorisés par l'apport de nouveaux patrons.

La répartition des NSSP peut également être observée à travers différents genres textuels, afin de vérifier si les NSS sont plus ou moins utilisés en fonction du genre.

ii. Répartition des NSSP au sein des sous-corpus

Une des interrogations de cette étude était de savoir si le genre textuel pouvait avoir ou non une influence sur l'utilisation de NSS. ANNODIS étant relativement diversifié, il est possible d'observer la répartition des NSSP au sein des trois sous-corpus qui le constituent : geop, wik2 et ling.

Sous-corpus	geop	Wik2	ling	Total
Fréquence de NSSP	2 634	2 108	1 777	6 518
Fréquence relative de NSSP (pour 100 000 mots)	990	912	1051	2953

Tableau 11 Répartition des NSSP dans les trois sous-corpus d'ANNODIS_me

A première vue, il n'y a pas de différence significative dans les répartitions des NSSP au sein des différents sous-corpus, ainsi il pourrait être avancé que l'emploi de NSS ne dépendrait pas du genre textuel utilisé. Toutefois pour vérifier cela, il convient de regarder la répartition des patrons 1 à 6 au sein de ces divers sous-corpus, pour observer si cette hypothèse reste vraie avec des noms réellement utilisés en tant que NSSP.

Sous-corpus	Fréquence absolue	Fréquence relative (pour 100 000 mots)
Geop	375	140
Ling	255	150
Wik2	301	130
TOTAL	931	420

Tableau 12 Fréquences de patrons dans les trois sous-corpus

Les tableaux 12 et 13 semblent confirmer l'hypothèse selon laquelle les NSSP découverts sont utilisés de manière relativement équitable, indépendamment du genre textuel dans lequel ils apparaissent. En effet pour 100 000 mots, le nombre de patrons est relativement similaire dans les trois sous-corpus, et les différents patrons précédemment décrits sont également répartis de façon similaire. Seul le patron 5 semble un peu moins bien réparti : il apparaît en effet 15 fois pour le sous-corpus geop, 48 fois pour le sous-corpus ling et 31 fois pour le sous-corpus wik2, or il n'est pas possible d'en tirer de conclusions définitives puisque nous avions vu précédemment que le patron 5 était celui dont la mesure de précision ramenait la valeur la plus faible, du fait de son manque de contraintes.

	geop		**ling**		**Wik2**	
	Freq. absolue	**Freq. relative**	**Freq. absolue**	**Freq. relative**	**Freq. absolue**	**Freq. Relative**
Patron 1	0	0	0	0	0	0
Patron 2	35	13	24	14	35	15
Patron 3	96	36	56	33	60	25
Patron 4	173	65	88	52	122	52
Patron 5	41	15	81	48	73	31
Patron 6	30	11	4	2	11	5
TOTAL	375	140	255	150	301	130

Tableau 13 Répartition des patrons dans les trois sous-corpus

Afin d'observer ce qu'il se passe au sein des SE, la répartition des NSSP sera maintenant analysée dans les amorces et les clôtures, afin de voir s'il existe un lien entre le fait d'être un NSSP et le fait d'apparaître dans l'un de ces éléments facultatifs.

iii. Répartition des NSSP au sein des amorces et des clôtures

Le sous-corpus ANNODIS_me est constitué de 991 SE, mais toutes ne contiennent pas d'amorces et de clôtures, qui sont des éléments facultatifs. La répartition de ces éléments facultatifs au sein des SE est la suivante :

Amorces	**Clôtures**	**Amorces + clôtures**	**Total**
637	29	102	768

Tableau 14 Répartition des amorces et des clôtures

Parmi ces 768 SE qui contiennent au moins un des deux éléments optionnels, 4% de tous les noms en font partie, et 97 des 6 090 sont des NSSP.

	Fréquence NC	**Fréquence NSSP**	
	Freq. Absolue	**Freq. absolue**	**%**
Amorces	5191	74	1.4
Clôtures	899	23	2.5
Total	6090	97	1.6

Tableau 15 Répartition des noms dans les SE (amorces et clôtures)

Si l'on se fie aux résultats de la figure 15, moins de 2% des noms présents dans les amorces et les clôtures seraient effectivement des NSSP. De plus, 1.6% des 6519 NSSP seraient concentrés dans les amorces et les clôtures, ce qui porte à croire que les NSS ne seraient pas forcément plus concentrés dans ces éléments que dans le reste du corpus. Toutefois, il semblerait que les résultats dans les clôtures soient plus significatifs. En effet, les clôtures sont des éléments facultatifs des SE et ils sont beaucoup moins représentés que les amorces, cependant il semblerait qu'une clôture soit plus souvent introduite par un NSSP qu'une amorce.

Afin de vérifier si cet écart entre les amorces et les clôtures est significatif, il est possible d'utiliser un test statistique tel qu'Epsilon ou u (loi normale). La formule de ce calcul est représentée comme suit :

$$u = \frac{|P_a - P_b|}{\sqrt{\frac{P * (1-P)}{N_a} + \frac{P * (1-P)}{N_b}}}$$

Avec P_a=0.014, P_b =0.025, N_a=5 191, N_b=899 et P= ((0.014*5191) + (0.025*899)) / N_a + N_b= 0.016

Le seuil de significativité pour un risque de 5% est de 1.96.

Dans le cas présent :

U = 2.19

U > 1.96, la différence entre la répartition des NSSP dans les amorces et dans les clôtures est donc significative, ce qui peut effectivement laisser penser que les clôtures seraient plus souvent accompagnées d'un NSS, tandis que ce ne serait pas aussi évident pour les amorces.

De plus, la faiblesse du pourcentage de NSSP en clôtures peut s'expliquer de par le fait que seuls 6 patrons ont été projetés jusqu'ici. Il convient donc de vérifier si d'autres patrons peuvent être révélés, notamment le patron *that* + N de Schmid (2000) déjà décrit. En effet, une hypothèse serait que ce patron serait largement répandu dans les SE, par exemple :

(37) Ces trois premiers **principes** entraînent donc les variations héréditaires qui confèrent un avantage sélectif seront davantage transmises à la génération suivante que les variations moins avantageuses.

Dans le cas de cet exemple 37, *principes* est un NSS présent en clôture, qui va chercher son contenu propositionnel dans l'énumération qui précédait.

Certaines constructions syntaxiques n'ayant pas forcément été décrites dans d'autres études jusque-là, il conviendra de mettre en évidence ces patrons syntaxiques et de vérifier, dans le cas où ils existeraient, s'ils sont propres au SE et permettent de les repérer automatiquement, ou si au contraire le repérage de SE à l'aide de NSS n'est pas pertinent.

3. Annotation manuelle de NSS et description de nouveaux patrons

i. Annotation manuelle avec GLOZZ

Afin de pouvoir décrire et projeter tous les patrons syntaxiques dans lesquels on peut retrouver des noms sous-spécifiés, au sein des SE, nous avons choisi d'annoter manuellement ces NSS et leurs patrons. Cette annotation sera réalisée à l'aide de l'interface GLOZZ.

L'interface GLOZZ est un outil d'aide à l'annotation qui a été conçu spécialement pour l'annotation d'ANNODIS, en 2007. Cette plateforme permet d'explorer et d'annoter des corpus, elle est totalement configurable, bien qu'elle n'ait été créée à la base que pour des contextes discursifs et elle avait été pensée de manière plus large afin de permettre d'autres annotations que des annotations multi-échelles du niveau discursif.

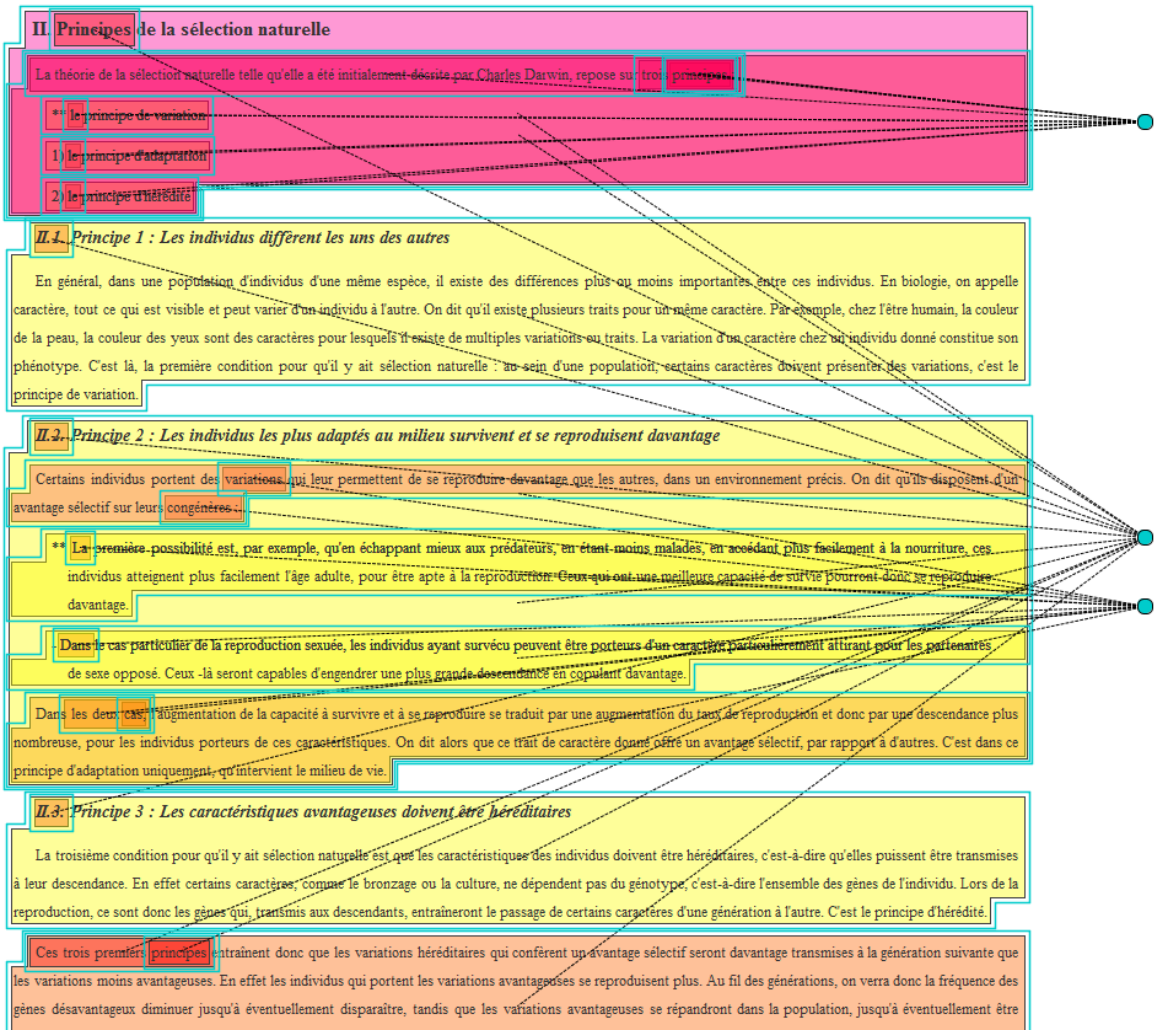


Figure 6 Interface de l'outil GLOZZ

Ce modèle repose sur l'articulation entre les notions d'unité, de relation et de schéma, avec un hyperonyme commun : l'élément. Les unités correspondent à des mots, syntagmes, propositions, paragraphes, unités thématiques ou documents. Les relations correspondent à un rapport binaire entre deux unités, peu importe l'ordre ou la distance, ce qui peut concerner tout ce qui est relations

syntaxiques, rhétoriques ou de coréférence. Les schémas correspondent à une configuration textuelle récurrente impliquant les unités et les relations, telle que la SE, composée d'une amorce, d'items et de relations hyperonymiques entre amorce et items. Cet ensemble se traduit par la représentation symbolique sous forme d'une structure de traits de l'association entre les différents éléments.

Les fichiers du corpus ANNODIS_me ont été chargés à l'aide de la plateforme d'annotation, et les NSS ainsi que leurs patrons syntaxiques ont pu être annotés dans les amorces et les clôtures.

Un modèle a été créé pour permettre ces annotations et il est le suivant :

```
<?xml version="1.0" encoding="UTF-8"?>
<annotationModel>

    <units>
        <type name="NSSV">
            <featureSet>
            </featureSet>
        </type>
        <type name="patron">
            <featureSet>
                <feature name="patron1et2">
                    <value type="boolean" default="0" />
                </feature>
                <feature name="patron3">
                    <value type="boolean" default="0" />
                </feature>
                <feature name="patron4">
                    <value type="boolean" default="0" />
                </feature>
                <feature name="patron5">
                    <value type="boolean" default="0" />
                </feature>
                <feature name="patron6">
                    <value type="boolean" default="0" />
                </feature>
                <feature name="nouveau patron">
                    <value type="free" default="" />
                </feature>
            </featureSet>
        </type>
    </units>

</annotationModel>
```

Figure 7 Modèle d'annotation des NSS et de leurs patrons

Ce modèle, au format XML, est relativement simple, puisqu'il est seulement composé de deux unités : *NSSV* et *patron*. *NSSV* correspond à un nom véritablement utilisé en tant que NSS, et l'unité *patron* permet d'annoter la construction syntaxique dans laquelle il apparaît. Ce modèle d'annotation est associé à une feuille de styles, qui permet la visualisation des annotations selon un ensemble de couleurs dans l'interface GLOZZ :

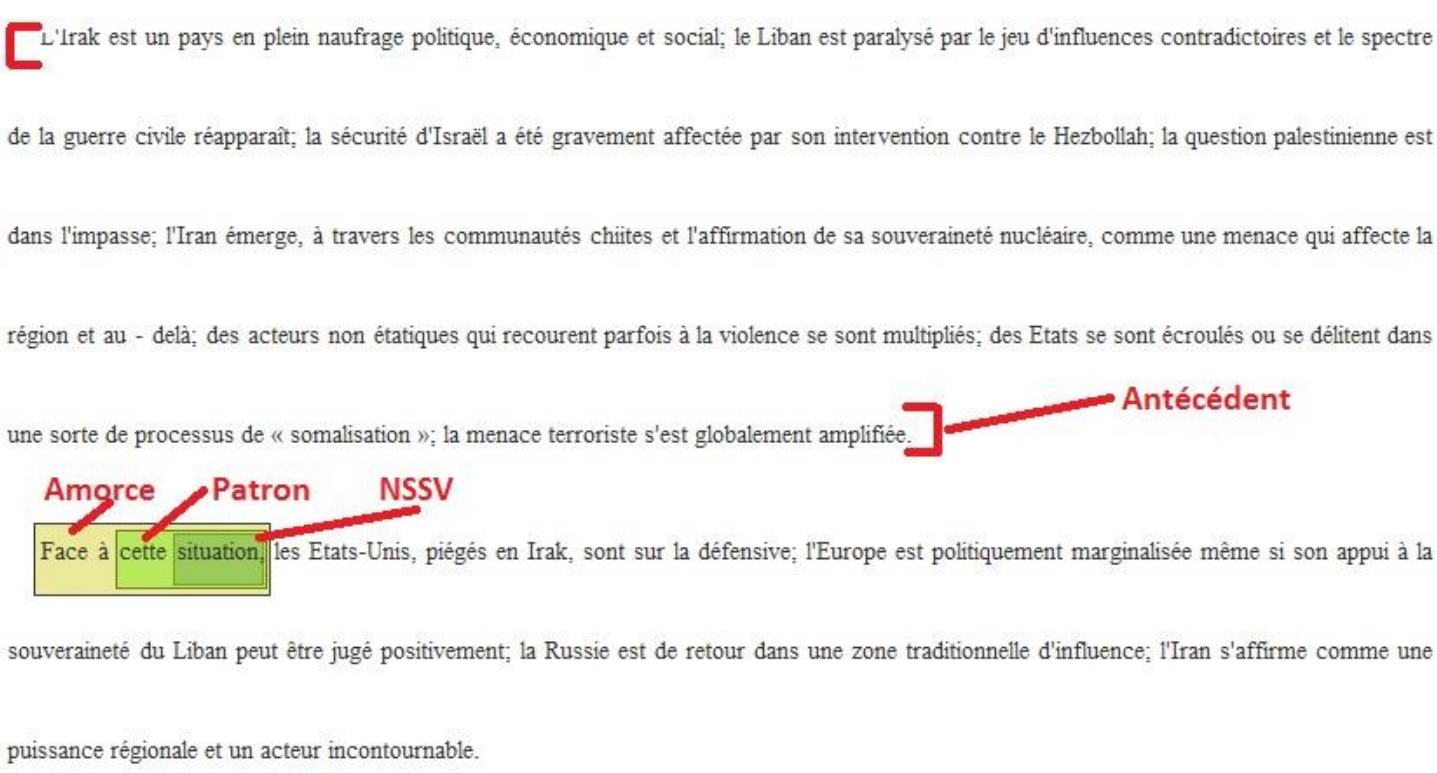


Figure 8 Annotation d'un NSSV et de son patron

Dans le cas présent, une amorce avait déjà été annotée lors de l'annotation d'ANNODIS_me, elle correspond à *Face à cette situation*, et nous avons pu annoter un NSSV ainsi que le patron de construction dans laquelle il apparaît. Dans ce cas, *situation* est ici un NSS, puisqu'il permet de résumer tout le paragraphe précédent, qui correspond à l'antécédent sur la figure 18. Le contenu propositionnel est en effet donné par le paragraphe qui précède, et le NSS apparaît dans un patron décrit par Schmid (2000) : *that + N*, que nous décrirons ici sous la forme suivante : [DEM] [NSS] ([DEM] pour *démonstratif*).

Afin de découvrir de nouveaux patrons, il a été convenu de ne s'intéresser qu'aux amorces et aux clôtures de SE de type 3, qui correspondent à des SE qui couvrent plus d'un paragraphe, sans marques visuelles spécifiques, telles que :

(38) [Face à cette situation :]$_{amorce}$ [les Etats-Unis, piégés en Irak, sont sur la défensive ;]$_{item1}$ [L'Europe est politiquement marginalisée même si son appui à la souveraineté du Liban peut être jugé positivement ;]$_{item2}$ [la Russie est de retour dans une zone traditionnelle d'influence ;]$_{item3}$ [l'Iran s'affirme comme une puissance régionale et un acteur incontournable.]$_{item4}$
[Il en résulte que les perspectives d'avenir, si les tendances actuelles devaient se confirmer, sont préoccupantes.]$_{cloture}$

Les fichiers annotés, ainsi que la répartition de leurs SE, de leurs SE de type 3 et des types de SE sont décrits ci-dessous :

Corpus	Nombre de SE	SE de type 3	T3 de type Ax	T3 de type AC	T3 de type xC
Geop_19	32	8	3	1	0
Geop_10	11	5	3	2	0
Geop_avicenne	34	8	2	5	0
Geop_13	23	4	2	0	0
Geop_26	7	5	2	2	0
Geop_15	14	3	0	0	1
Geop_3	14	3	1	2	0
Ling_puech	11	9	6	1	0
Ling_kleiber	13	8	3	2	1
Ling_bouard	12	3	2	1	0
Ling_bronckart	17	4	3	1	0
Ling_prandi	14	3	1	0	0
Ling_colasBlaise	17	4	4	0	0
Ling_muller	16	7	6	1	0
Wik2_attentats11sept	19	7	4	0	1
Wik2_histoireCapitalisme	32	12	2	1	1
Wik2_titanic	15	7	2	0	0
Wik2_affaireDreyfus	16	6	1	0	0
Wik2_bicyclette	6	1	1	0	0
Wik2_liberte	15	1	1	0	0
Wik2_Islam	20	5	3	1	0
TOTAL	358	113	52	20	4
				TOTAL SE retenues	76

Tableau 16 Description des fichiers annotés

Les SE de type AX correspondent à des SE qui possèdent seulement une amorce, les SE de type AC possèdent une amorce et une clôture et les SE de type XC possèdent seulement une clôture. Les SE ne contenant aucun élément facultatif ont été supprimées puisque seuls les NSS présents en amorces ou en clôtures nous intéressaient ici.

La difficulté pour réaliser ces annotations relevait du fait que les NSS sont des noms complexes, qui ne sont pas toujours facilement repérables ou qui ne se comportent pas toujours de la même façon. Plusieurs questions devaient être posées afin de déterminer si un nom présent dans une amorce ou une clôture était bien un NSS :

- Le nom apparaît-il dans une construction déjà décrite par d'autres auteurs ? (Voir les patrons 1 à 6 définis plus haut)
- Le nom permet-il d'amorcer ou de clôturer une énumération ?
- Le nom correspond-il à un énuméraThème ?
- Le nom est-il attaché à un contenu propositionnel qui précède ou qui suit ?

Lorsque la réponse est oui au moins à une de ces questions, on peut considérer presque de manière certaine qu'il s'agit d'un NSS, et de l'annoter comme tel, avec son patron. Par exemple :

(39) De plus en plus souvent, pourtant, ils semblent faire l'objet d'ouvrages complets qui règlent l'empan de l'enquête historique (court/ moyen. Long terme) sur deux **finalités** complémentaires :
- D'une part, revenir aux nœuds qui ont scellé la crise en restituant les données qui ont scandé le développement scientifique, qui en ont constitué les acquis, qui en ont installé les apories, les obstacles surmontés ou restant à surmonter. Or, il nous semble que c'est à ce moment de clôture d'une période relativement longue pour les sciences du langage de cumulativité des résultats scientifiques, et alors même que cette cumulativité de la grammaire historique et comparée est menacée par la crise des lois phonétiques, que l'histoire émerge comme une pièce essentielle

d'un dispositif critique, moins tourné vers le passé comme source de valeurs scientifiques que comme appui pour de nouvelles fondations et refondations (la linguistique générale).
- Est -ce un hasard, d'autre part, si c'est à la charnière des deux siècles, et à l'issue de cette crise, que s'affirme avec le plus d'insistance et selon une thématique promise à un long avenir dont il faudrait faire l'histoire fine la revendication d'autonomie de la discipline linguistique, autonomie qui ne cessera d'être à la fois reprise, disputée et contestée dans la postérité saussurienne tout particulièrement, mais aussi au-delà?

Dans le cas de l'exemple 39, le nom *finalité* que l'on retrouve en amorce est bel et bien sous-spécifié. En effet, bien qu'il n'apparaisse pas dans des constructions syntaxiques déjà décrites par Legallois (2008) ou Roze et al. (2014), il est rattaché à un contenu propositionnel que l'on retrouve par la suite, dans les deux items de l'énumération. Sans ce contenu particulier, le lecteur serait en attente, puisqu'il ne saurait pas à quoi correspondraient ces « deux finalités complémentaires ». Nous pourrons également noter que ce NSS correspond également à l'énuméraThème de cette SE, et que tous les énuméraThèmes repérés correspondaient chaque fois à un NSS.

Toutefois dans certains cas, il est toujours difficile d'annoter avec certitude un NSS, par exemple :

(40) Les relations nouées depuis des siècles dans la région nous valent assurément estime et considération. Elles suscitent aussi des *attentes* et des *déceptions*.
Au Maghreb, les gouvernements attendent de nous concours et, pour chacun d'entre eux, soutien exclusif. Les populations sont plus attentives à la coopération, à la liberté de circulation et à la situation des immigrés chez nous.
Au Proche-Orient, nos prises de paroles sont scrutées et analysées dans le détail. Nous y sommes attendus, sollicités et espérés tant l'image d'une France compagnon de route des grandes causes arabes demeure encore enracinée.
L'approche est différente dans le Golf où nous sommes vus comme un partenaire privilégié pour se soustraire à un tête-à-tête trop exclusif avec les Etats-Unis.
Les *perspectives* pour la France dans tous les domaines y sont remarquables.
En témoignent tout récemment les opérations du Louvre et de la Sorbonne à Abou Dhabi.

Cette SE est constituée de trois items, d'une amorce et d'une clôture. L'amorce correspond à : *Les relations nouées depuis des siècles dans la région nous valent assurément estime et considération. Elles suscitent aussi des attentes et des déceptions* et la clôture correspond à : *Les perspectives pour la France dans tous les domaines y sont remarquables. En témoignent tout récemment les opérations du Louvre et de la Sorbonne à Abou Dhabi.*

Dans l'exemple 39, il ne fait aucun doute que finalité est réellement un NSS, or dans l'exemple 40, trois noms sont possiblement des NSS, à savoir *attentes* et *déceptions* en amorce, et *perspectives* en clôture. Dans le cas présent, les *attentes* et les *déceptions* en question sont énumérées ensuite à travers les trois items de la SE, or il est tout à fait possible d'envisager un contexte dans lequel ces items n'existeraient pas et où *attentes* et *déceptions* n'iraient pas chercher un contenu propositionnel ailleurs. En effet, le lecteur n'est pas forcément ici dans l'attente d'une énumération, bien qu'il est concevable qu'elle soit présente, alors que dans d'autres cas, tel l'exemple 39, il n'est pas envisageable qu'une énumération ne suive pas.

Le choix a été fait d'annoter ici *attentes* et *déceptions* comme étant des NSS, puisqu'ils sont effectivement utilisés comme tels dans ce contexte, mais nous noterons que certains NSS semblent appeler une énumération, tandis que d'autres non. On peut également penser que ce sont les patrons syntaxiques dans lesquels apparaissent ces NSS qui appellent ou non une énumération. Par exemple avec le NSS *attente*, une énumération n'est pas forcément nécessaire dans l'exemple 40, tandis qu'elle l'est forcément dans un contexte du type :

(41) Le lecteur a deux attentes : *item 1* et *item 2*

ii. Patrons retenus

Sur 76 SE de type 3 contenant une amorce et/ou une clôture, 73 NSS ont pu être annotés, à travers 8 patrons. Ces patrons, accompagnés d'exemples, sont les suivants :

[DET] [NSS] [V] [VINF]

(42) « L'**important** doit être [une disposition au dialogue pour autant que l'interlocuteur respecte, lui aussi, ce que nous sommes.] »

[DEM] [NSS] (Correspond au patron That N de Schmid, 2000)

(43) « L'autre forme de centralisation, plus institutionnelle, consistait à court-circuiter les hiérarchies traditionnelles dans certains secteurs sensibles. [Ainsi, les escadrons d'attaque de la Force aérienne sont passés dès 1978 sous la coupe de Saddam Hussein. Plus tard, la Sûreté et les Istikhbarat, soustraits aux ministères de l'Intérieur et de la Défense, respectivement, ont de même été soumis à la tutelle d'une présidence concentrant toujours plus d'autorité.]

Tout ce **processus** sera renforcé par le développement de l'image de l'ennemi intérieur, relais des " impérialistes " et autres " sionistes ", avant que l'identification des minorités irakiennes " complices " soit bientôt doublée de celle d'un ennemi extérieur autrement important : l'Iran. »

[P (de|par)] [ADJ] [NSS]

(44) « Il en résulte que les perspectives d'avenir, si les tendances actuelles devaient se confirmer, sont préoccupantes. Après l'échec de la Pax Americana, on peut craindre des évolutions plus radicales : [un ordre islamiste, tout au moins dans certains pays, un chaos généralisé. Au mieux se vérifierait une évolution moins extrême comme un processus de dégradation progressif et modulé selon les pays.]

De telles **évolutions** ne sont pas une fatalité. »

(45) « Le naufrage du Titanic a de nombreuses causes, tant naturelles qu'humaines. Son bilan, qui est l'un des plus lourds de l'histoire maritime, s'explique également par plusieurs **facteurs**.

[Les circonstances du naufrage sont en effet particulières. Il est en effet rare de trouver des icebergs dans cette région de l'Atlantique au mois d'avril, mais la présence de nombreuses glaces cette année-là s'explique par un hiver particulièrement doux. Ceci explique que le Titanic, qui navigue pourtant plus au sud que la route conseillée, se soit dirigé droit vers un champ de glaces. De plus, la nuit est sombre, sans Lune et sans vent, ce qui rend plus difficile le repérage des icebergs. Ceci est aggravé par l'absence de jumelles dans le nid-de-pie, suite à une négligence des officiers : selon Frederick Fleet, le veilleur qui a aperçu l'iceberg, des jumelles auraient permis de le voir à temps.

De plus, les compartiments étanches ne montent pas assez haut pour empêcher la progression de l'eau, et l'acier composant certaines parties de la coque est très cassant. La vitesse du navire au moment du choc était également trop élevée pour les circonstances (bien qu'en accord avec les règles de l'époque). Malgré une tentative de la part de la commission américaine, aucune preuve n'a pu être fournie sur le fait qu'Ismay ait poussé le commandant à aller plus vite.

Enfin, le nombre élevé de morts s'explique par le faible nombre de canots de sauvetage du navire, qui ne peuvent contenir que 1 178 personnes, mais aussi par le manque d'organisation dans leur chargement. Certains canots, comme le no 147, partent complètement vides et refusent de revenir sur les lieux du naufrage. Ceci explique que les canots soient, au final, remplis au trois quarts.]

[DET] [ADJ] [NSS]

(46) « La nouvelle donne en Afrique du Nord et au Moyen-Orient s'accompagne d'un retour spectaculaire de certains **acteurs** à l'instar de [la Russie, trop hâtivement considérée comme sortie du monde arabe et musulman, et d'une percée notable de nouvelles puissances, à savoir la Chine et l'Inde.] »

(47) « Si l'on met de côté une telle **position** extrême »

[DET numéral] [NSS]

(48) « Pour reprendre l'exemple tant débattu du Cours de linguistique générale et de son rôle dans l'histoire des idées linguistiques contemporaines, il me semble qu'on a aujourd'hui deux **manières** de considérer son statut :

[- ou bien on considère que c'est le Cours qui a effectivement joué un rôle séminal dans la genèse des différents structuralismes comme si le texte possédait en lui -même et de manière virtuelle son historicité, le principe de son devenir;

- ou bien on cherche la productivité historique de ce texte dans la manière dont on y a renvoyé, dont on s'y est référé en cherchant à caractériser le plus précisément possible les « modes de références » et les reconstructions dont il a été l'objet dans des contextes scientifiques, culturels, institutionnels les plus divers. »

[Quelques-uns | Quelques-unes] [ART des] [NSS]

(49) « Au terme de ce trop rapide parcours (pour de plus amples développements, cf. J. L. Chiss et C. Puech, 1987, 1997, 1999) dont on aura compris qu'il ne vise en rien à dévaluer le travail sur les sources, mais au contraire à l'inclure à sa place dans le continent du saussurisme en essayant de voir en quoi il est susceptible d'en faire bouger les représentations convenues et répétitives, il est peut-être temps d'énumérer quelques-unes des **difficultés** que présente à notre sens l'historiographie saussurienne. »

[NSS] [ADJ]

(50) « L'intéressant, du point de vue de la réception stylistique, c'est que le concours des niveaux de pertinence appelle, ici et là, des **régimes** différents.

[Dans Un balcon en forêt, le tiret forme système avec d'autres ponctèmes (le deux-points, le point-virgule...) pour proposer une reformulation, assumée par le locuteur, cohérente et prévisible, des tentatives d'organisation de l'espace-temps par le personnage. Récurrent sur de vastes étendues textuelles, il draine vers lui, pour s'y connecter, les isotopies figuratives et thématiques, les constructions syntaxiques auxquelles le schème structural peut être associé, et

les noue ensemble au sein d'un système semi-symbolique qui appelle une stratégie d'analyse à fort pouvoir intégratif.

Dans Le Rivage des Syrtes, le tiret vaut davantage comme une forme de l'expression répondant à la définition rhétorique du style selon G. Molinié (1994 : 206).]»

[ADV] [P de] [NSS]

(51) « [On ignore, par exemple, quelles seraient les conséquences d'une deuxième vague d'attentats. À coup sûr, la très forte tension ressentie aujourd'hui dans les villes se transformerait en réelle psychose. Mais quelles en seraient les conséquences politiques, économiques, psychologiques, tant au plan interne qu'international?

De même, on ignore si l'opinion publique restera toujours favorable à l'engagement américain en Afghanistan. Si le concept du " zéro mort " semble avoir vécu outre-Atlantique et si, d'après de récents sondages, l'opinion publique semble prête à supporter le coût d'une longue campagne, encore faudra-t-il qu'elle reste convaincue de l'adéquation entre les moyens mis en oeuvre et l'objectif poursuivi, à savoir la destruction des réseaux terroristes, entreprise longue et incertaine. Si de nouveaux attentats sont perpétrés aux États-Unis et que, par ailleurs, les forces américaines s'enlisent en Afghanistan ou subissent des pertes importantes, il faudra s'attendre à ce qu'un nombre croissant d'Américains remette en cause l'opportunité d'une guerre lointaine, à l'heure où les terroristes agissent sur le territoire national. Si, traditionnellement, la capacité des opinions publiques à supporter les coûts d'une opération militaire est liée à la conviction de mener une guerre " juste ", elle est également fonction de la lisibilité du conflit et de la conviction de mener une guerre " efficace ".

Enfin, au plan politique interne, on peut s'interroger sur la pérennité de l'union sacrée qui réunit démocrates, républicains et indépendants depuis le discours du président Bush devant le Congrès, le 20 septembre Si l'on a encore en mémoire l'image des représentants et sénateurs américains applaudissant debout leur président, tous partis confondus, à l'issue d'un discours qualifié à maintes reprises et sans surprise d'historique ", il convient aussi de signaler les tensions qui sont apparues peu de temps après entre démocrates et républicains quant au bien-fondé et au montant du plan de relance de l'économie américaine annoncé par la Maison Blanche et même quant à la fédéralisation des contrôles de sécurité dans les aéroports. Le président va-t-il toujours bénéficier des pleins pouvoirs qui lui ont été de facto accordés le 11 septembre? Ou faut-il s'attendre à ce que le jeu des partis reprenne son cours, en particulier à l'approche des élections de mi-mandat, avec - rappelons-le - l'enjeu d'un possible basculement de l'ensemble du Congrès sous majorité démocrate?] »

Autant de **questions** dont les réponses sont encore inconnues et qui dépendront dans une large mesure de facteurs exogènes. »

Ajoutons que certains NSS ont pu être annotés à travers un patron difficile à faire apparaître. En effet, certains NSS apparaissent dans des titres de section, ainsi ils ne font partie de constructions syntaxiques propres aux NSS, ils sont principalement repérés à l'aide d'éléments visuels de numérotation, par exemple :

(52) IV.3. **Opérations** militaires

[L'impact militaire le plus direct est l'invasion de l'Afghanistan, désigné comme le siège opérationnel d'Al-Quaïda, dès le mois d'octobre 2001 et le renversement du régime des Talibans quelques mois plus tard par les forces armées américaines, britanniques, canadienne, françaises, et autres.

Ce renversement et l'établissement d'un gouvernement de transition s'accompagne de l'arrestation de nombreux musulmans présumés terroristes, internés dans des camps disséminés autour de la planète, ce qui provoquera les vives réactions de nombreuses ONG, dont Amnesty International. La création de la prison de Guantanamo s'explique en partie par cet afflux important de prisonniers.
Un second impact militaire d'importance est l'invasion de l'Irak et le renversement du régime de Saddam Hussein en 2003 par les forces armées américaines et britanniques. Bien que l'Iraq de Saddam Hussein n'ait pas participé aux attentats du 11 septembre, le régime baasiste a été désigné par l'administration américaine comme un soutien actif du terrorisme international et un détenteur d'armes de destruction massive, malgré l'absence de preuves sur le terrain. Le régime de Saddam Hussein a été remplacé par un régime plus démocratique, notamment par la tenue d'élections et une représentation de la majorité chiites par rapport aux sunnites. L'invasion de l'Irak provoquera de houleux débats à l'ONU et des manifestations à travers le monde, protestant contre les véritables raisons qui seraient d'ordre économique et stratégique (indépendance énergétique vis-à-vis de l'Arabie saoudite notamment).
Il est à remarquer que le candidat George W. Bush s'était engagé pendant sa campagne sur le fait que les États-Unis ne prendraient pas l'initiative d'opérations militaires nouvelles hors de leur territoire national. Les événements du 11 septembre lui donnaient donc à nouveau les coudées franches dans ce domaine.]

Ces types de NSS ne seront pas traités dans la présente étude, nous nous limiterons à des patrons syntaxiques plus évidents à décrire, bien que nous puissions souligner que l'étude de ce genre de cas pourrait être intéressante.

Nous pourrons également noter que certains patrons se recoupent, et qu'il serait possible, à l'avenir, de privilégier un classement tel que celui proposé par Schmid (2000), en regroupant donc certains patrons proches syntaxiquement. Il y aurait alors un patron générique accompagné d'une liste de sous-patrons dérivés. Dans le cas présent par exemple, les patrons [P (de|par)] [ADJ] [NSS], [DET] [ADJ] [NSS] et [NSS] [ADJ] seraient issus de la même liste de patrons dérivés.

Parmi ces patrons, certains se trouvent beaucoup plus représentés que d'autres, la répartition des NSSV dans ces 8 constructions est la suivante :

Patron	**Nombre de NSSV**
[DET] [NSS] [V] [VINF]	3
[DEM] [NSS]	14
[P (de|par)] [ADJ] [NSS]	7
[DET] [ADJ] [NSS]	4
[DET numéral] [NSS]	29
[Quelques-uns | Quelques-unes] [ART des] [NSS]	1
[NSS] [ADJ]	4
[ADV] [P de] [NSS]	1
TOTAL	**63**

Tableau 17 Répartition des NSSV dans les nouveaux patrons

Il apparaît clairement à travers ce tableau que deux des huit patrons décrits sont bien plus représentés que les autres, à savoir le patron *[DET (ce|cette|ces)] [NSS]* et le patron *[DET numéral] [NSS]*, qui ont été annotés respectivement 14 et 29 fois. Les patrons contenant un NSS et un adjectif, peu importe que celui-ci soit placé avant ou après le NSS, sont également nombreux, puisque 10 NSS ont pu être annotés à partir de ces derniers. Certains paraissent plus anecdotiques puisque peu de NSS y ont été relevés, toutefois cela restera à vérifier après la projection de ces nouveaux patrons.

De plus, 10 NSSV ont pu être annotés dans les patrons déjà décrits dans la première partie des présents résultats. 7 apparaissent dans le patron 5 ([NC] [PONCT :]), 2 dans le patron 4 ([NC] [P de] [VINF]) et

1 dans le patron 2 ([DET] [NC] [Ø] [V être] [CS que | P de]). Cependant, il apparaît clairement que les patrons utilisés par d'autres auteurs ne sont pas les plus fréquents au sein des SE, d'où la nécessité de cette annotation manuelle.

4. Projection des nouveaux patrons obtenus et perspectives

Cette partie n'a pas pour vocation de fournir une liste exhaustive de NSSP, il s'agit d'une première approche dont le but est de permettre de vérifier s'il serait éventuellement possible de repérer des SE à l'aide de NSSP issus de patrons. Toutefois, nous verrons que ces patrons nécessitent d'être revus pour obtenir des résultats plus concluants. Cette partie représente donc plus une ouverture possible pour un travail ultérieur qui pourrait se baser sur les méthodes utilisées dans ce mémoire.

Comme dans la première étape des analyses proposées dans cette étude, les nouveaux patrons ont été extraits de manière automatique, sur le corpus ANNODIS_me et sur le corpus de LA. Cette extraction automatique a permis d'obtenir 48 479 NSSP sur ANNODIS_me et 55 454 sur LA, selon la répartition suivante :

	ANNODIS_me		**LA**	
	Freq. absolue	**Freq. Relative (100 000 mots)**	**Freq. absolue**	**Freq. Relative (100 000 mots)**
[DET] [NSS] [V] [VINF]	360	54	1 788	185
[DEM] [NSS]	5 224	784	9741	1010
[P (de|par)] [ADJ] [NSS]	996	149	1 265	131
[DET] [ADJ] [NSS]	6 613	992	9 655	1 001
[DET numéral] [NSS]	1 613	242	2 885	299
[Quelques-uns | Quelques-unes] [ART des] [NSS]	1	0	0	0
[NSS] [ADJ]	33 164	4 979	29 297	3 039
[ADV] [P de] [NSS]	508	76	823	85
TOTAL	48 479	7 276	55 454	5 750

Tableau 18 Résultats de la projection des nouveaux patrons

De plus, ces résultats sont à nuancer puisque le script perl (disponible en annexe 9) qui a permis l'extraction serait à revoir. En effet, certains patrons se recoupent, ou sont trop ou pas assez contraints, et mériteraient sûrement d'être repris. Prenons l'exemple suivant :

(53) Cet enfant adorable

Si l'on se fie au script perl utilisé, alors cet exemple 53 correspondra à la fois au patron [DEM] [NSS] et à la fois au patron [NSS] [ADJ].

Une fois les deux listes de NSSP issus des deux sous-corpus réunies, et après suppression des doublons, il reste encore 4 766 NSSP. Toutefois, comme cela a déjà pu être observé précédemment, beaucoup d'entre eux ne sont pas sous-spécifiés dans tous leurs contextes. Ces noms n'ont pas été analysés comme cela avait pu l'être fait précédemment avec la liste de l'annexe 5. Notons cependant que la même méthode que celle utilisée avec les 6 premiers patrons de Legallois (2008) et Roze et al. (2014) pourrait être reproduite à nouveau.

Certains patrons ont été trop contraints, tels que le patron *[Quelques-uns | Quelques-unes] [ART des] [NSS]*, et d'autres au contraire ne l'ont probablement pas été assez, tels que [NSS] [ADJ]. Une évaluation avec 10 NSSP pour chaque patron de chaque sous-corpus permet de mettre cela en valeur :

	ANNODIS_me			LA		
	NSSP évalués	**NSSV**	**SE**	**NSSP évalués**	**NSSV**	**SE**
[DET] [NSS] [V] [VINF]	10	6	5	10	1	0
[DEM] [NSS]	10	10	4	10	10	3
[P (de|par)] [ADJ] [NSS]	10	6	4	10	4	3
[DET] [ADJ] [NSS]	10	6	1	10	6	6
[DET numéral] [NSS]	10	4	2	10	10	10
[Quelques-uns | Quelques-unes] [ART des] [NSS]	1	1	1	0	0	0
[NSS] [ADJ]	10	2	0	10	2	1
[ADV] [P de] [NSS]	10	5	0	10	2	0
TOTAL	71	40	17	70	35	23

Tableau 19 Evaluation de la projection des nouveaux patrons

Quelques différences peuvent être observées entre les deux corpus. Dans ANNODIS_me, 40 NSS ont pu être validés sur les 71 NSSP, soit une précision de 0.56. Dans LA, 35 NSS ont pu être validés sur 70 NSSP, soit une précision de 0.50, les résultats sont donc similaires. Mais il faut tout de même noter que les résultats diffèrent entre certains patrons. Par exemple, 6 NSS ont été validés dans ANNODIS_me avec le patron *[DET] [NSS] [V] [VINF]*, contre un seul dans LA, et à l'inverse, 4 ont été validés dans ANNODIS_me avec le patron [DET numéral] [NSS] contre 10 dans LA. De plus, le rapport entre le nombre de NSSV et le nombre de SE détectées est de 42% pour ANNODIS_me et de 65% pour LA. Les patrons qui permettent de détecter le plus souvent une SE ne sont pourtant pas les mêmes entre les deux corpus. Dans ANNODIS_me, il s'agit du patron [DET] [NSS] [V] [VINF] alors que dans LA, les patrons [DET] [ADJ] [NSS] et [DET numéral] [NSS] permettent de détecter 16 SE sur les 35 attendues. Bien qu'il soit difficile de tirer des conclusions à partir de ces résultats trop peu nombreux, il est possible de poser l'hypothèse selon laquelle l'amélioration de la détection des NSS pourrait engendre l'amélioration du repérage automatique des SE. Nous avions également vu qu'il ne semblait pas y avoir de différences entre les genres textuels dans ANNODIS_me, et si les résultats obtenus sur les nouveaux patrons sont vérifiés à plus grande échelle, il faudrait vérifier si le nombre de SE utilisées dans les écrits étudiants n'est pas plus important que pour d'autres genres textuels, ce qui pourrait expliquer, si tel est le cas, que certains patrons permettent un meilleur repérage des NSS et des SE.

Conclusion

L'analyse des NSS, en tant qu'acteurs de la cohésion discursive, a permis de décrire leur fonctionnement ainsi que les constructions syntaxiques dans lesquelles ils apparaissent préférentiellement. Cette étude s'est basée dans un premier temps sur les travaux déjà réalisés dans la littérature linguistique française. Deux études ont permis d'amorcer l'analyse proposée : celle de Legallois (2008) et celle Roze et al. (2014). Il s'est donc agit de réutiliser la construction spécificationnelle proposée par Legallois pour décrire les NSS, et les patrons décrits par Roze et al. Ces constructions ont alors pu être projetées sur les deux corpus utilisés (ANNODIS_me et LA) afin de vérifier d'abord si les NSS pouvaient être repérés de manière automatique, ensuite si les noms extraits étaient plus ou moins souvent sous-spécifiés, et enfin si ces derniers pouvaient permettre de découvrir de nouveaux patrons tels que ceux déjà décrits par Schmid (2000) pour l'anglais. Les résultats ont alors confirmé, notamment à l'aide du calcul de la *reliance*, que certains noms semblaient plus systématiquement sous-spécifiés que d'autres, c'est notamment le cas de *fait* ou *objectif*.

La première étape de cette analyse a également mis en évidence le fait que certains patrons sont plus fréquents que d'autres, par exemple le patron [NC] [P de] [VINF], qui permet d'extraire près de 50% de NSSP.

Nous nous sommes ensuite interrogés quant à la répartition de ces NSSP en fonction du genre textuel, avec les trois sous-corpus d'ANNODIS_me : geop, ling et wik2. En effet il paraissait légitime de se demander si un genre textuel particulier pouvait influencer l'utilisation de NSS. Les résultats ont semblé montrer que non, puisque la répartition de NSSP entre les différents sous-corpus était relativement équivalente.

Les NSS ont également été observés au sein des SE, et plus particulièrement dans les amorces et les clôtures de ces dernières. Des évaluations statistiques ont permis de montrer qu'il y aurait un écart significatif entre le nombre de NSSP dans les amorces et dans les clôtures. Les clôtures auraient tendance à recourir plus fréquemment à l'utilisation d'un NSS. De plus, les annotations manuelles des NSS de 76 SE ont montré que chaque énuméraThème explicite correspond à un NSS, ce qui confirme l'hypothèse de départ que nous posions dans notre mémoire.

Malgré cela, les NSS ne permettent pas forcément de repérer automatiquement les SE, même lorsque des patrons issus de SE ont été relevés, bien que le nouveau patron décrit [DET numéral] [NSS] semble intéressant dans ce but pour le corpus de LA. De façon générale, les NSS semblent globalement bien répartis dans les textes, peu importe le genre et peu importe les patrons au sein desquels ils apparaissent.

Pour confirmer ou infirmer tous ces résultats, il pourrait être intéressant de proposer une méthode d'extraction automatique de NSS qui combine l'utilisation de patrons et l'utilisation d'une liste de NSS dont la valeur de *reliance* est supérieure à un seuil qu'il faudra définir (à moins de reprendre celui de 15% proposé par Schmid, 2000). En effet, l'amélioration de la précision pour l'extraction automatique de NSS pourrait permettre de proposer une méthode plus fiable pour l'extraction automatique des SE. Une autre perspective serait le repérage et la catégorisation des antécédents des NSS, puisqu'il est possible de penser que ces derniers pourraient être des indices pour le repérage de structures telles que les SE. Enfin, un point intéressant n'a pas pu être étudié dans cette étude, il s'agit de l'analyse des NSS en fonction du degré de littéracie des scripteurs. Une des poursuites de ce travail pourrait donc consister à faire cette analyse et à mettre en parallèle les résultats qu'avaient pu obtenir Nur Aktas et Cortes (2008).

BIBLIOGRAPHIE

APOTHELOZ D., (1995) *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*, Genève/Paris, Librairie Droz

APOTHELOZ D., CHANET C., (1997) Défini et démonstratif dans les nominalisations, in W. De Mulder, L. Tasmowski-De-Ryck, C. Vetters (éds) : *Relations anaphoriques et (in)cohérence*, Amsterdam : Rodopi, 159-186.

BOUDREAU S., KITTREDGE R., (2006) Résolution d'anaphores et identification des chaînes de coréférence : ne approche « minimaliste », in *8e journées internationales d'analyse statistique des données textuelles (JADT 2006),* Besançon

BOURIGAULT D., (2007) *Un analyseur syntaxique opérationnel : Syntex*. Mémoire d'HDR, Université de Toulouse

BRAS M., PREVOT L., VERGEZ-COURET M., (2008) Quelle(s) relation(s) de discours pour les structures énumératives ?, *in* Eds Durand J. Habert B., Laks B., Congrès Mondial de Linguistique Française

CORNISH F., (2011) Strict anadeixis, discourse deixis and text structuring, in *Language Sciences*, 33, 753-767

CHARAUDEAU P., MAINGUENEAU D., (2002) *Dictionnaire d'Analyse du Discours*, in Ed. du Seuil

COMBETTES B., (1983) *Pour une grammaire textuelle. La progression thématique*, Ed. De Boeck/ Duculot

DUCROT O., SCHAEFFER J-M., (1995) *Nouveau dictionnaire encyclopédique des sciences du langage*, Ed. du Seuil

GRAY B., (2010) One the use of demonstrative pronouns and determiners as cohesive devices : A focus on sentence-initial this/these in academic prose, in *Journal of English for Academic Purposes 9*, 167-183

GUILLOT C., (2006) Démonstratif et déixis discursive : analyse comparée d'un corpus écrit de français médiéval et d'un corpus oral de français contemporain, in *Langue Française*, 152, 56-69

HO-DAC L-M., PERY-WOODLEY M-P., TANGUY L., (2010) Anatomie des structures énumératives, *TALN*, Montréal

JACKIEWICZ A., (2005) Les séries linéaires dans le discours, in *Langue Française*, 148, 95-110

JEANDILLOU J-F., (2006) *L'analyse textuelle*, Ed. Armand Colin

KLEIBER G., (1999) Anaphore associative et relation partie-tout : condition d'aliénation et principe de congruence ontologique, in *Langue Française*, 122, 70-100

LEGALLOIS D., (2006) Quand le texte signale sa structure : la fonction textuelle d'une certaine catégorie nominale, in *Corela*, numéro spécial Organisation des textes et cohérence des discours

LEGALLOIS D., (2008) Sur quelques caractéristiques des noms sous-spécifiés, in *Scolia*, 23 :109-127

MITKOV R., (2002) *Anaphora Resolution*, Ed. Longman (Pearson Education)

NUR AKTAS R., CORTES V., (2008) Shell nouns as cohesive devices in published and ESL student writing, in *Journal of English for Academic Purposes* 7, 3-14

REBBEYROLLE J., PERY-WOODLEY M-P., (2014) Enumération et structuration discursive, *in* Congrès Mondial de Linguistique Française- CMLF 2014, SHS Web of Conferences

ROZE C., CHARNOIS T., LEGALLOIS D., FERRARI S., SALLES M., (2014) Identification des noms sous-spécifiés, signaux de l'organisation discursive, *21ème Traitement Automatique des Langues Naturelles*, Marseille

SCHMID H-J., (2000) *English abstract nouns as conceptual shells: from corpus to cognition*,

Topics in English Linguistics, 34
URIELI, A., (2013) *Analyse syntaxique robuste du français : concilier méthodes syntaxiques et connaissances linguistiques dans l'outil Talismane*, Thèse de doctorat, Toulouse 2
VERGEZ-COURET M., BRAS M., PREVOT L., VIEU L., ATTALAH C., (2011) Discourse contribution of Enumerative structures involving pour deux raisons, *in* Eds N. Asher, L. Danlos, Proceesings of the 4th Constraints in Discourse Workshop (CID 2011), (Agay Roches Rouges : INRIA), 1-10
WEISSENBACHER D., (2008) *Influence des annotations imparfaites sur les systèmes de Traitement Automatique des Langues, un cadre applicatif : la résolution de l'anaphore pronominale*, Thèse de doctorat : spécialité informatique. Paris-Nord- Paris 13, 170p
WIDLOCHER A., MATHET Y., (2009) La plate-forme Glozz : environnement d'annotation et d'exploration de corpus, in *Actes de TALN*, Senlis, France
WIEDERSPIEL B., (2012) Anaphores, stratégies discursives et genres textuels, in *Echo des études romanes*, VIII/Num.1, 241-254

Annexe 1

Liste des NSS proposée par Legallois (2008)

« NSS pour la CS avec inf. : *acte action alternative ambition apport argument art astuce atout attitude audace avantage avenir axe bataille boulot but caractéristique cauchemar chance charge choix chose clé cœur concept confort conseil conséquence consigne consolation contre-feu contribution controverse conviction coup-de-génie courage crainte critère culot danger débat décision défaut démarche dénominateur-commun désir dessein destin deuxième devoir difficulté dignité direction directive discours discussion don effet éloge énigme enjeu enseignement erreur espoir esthétique étape éthique exemple exigence face facilité façon faux-pas fin-du-fin finalité fonction fond-de-notre-méthode force gageure geste grâce grandpied grande-affaire habileté hantise hommage idéal idée illusion impératif inclination inconvénient initiative inquiétude intelligence intention intérêt l'essentiel l'important le-plus-économique le-mieux le-moindre-mal le-plus-désagréable leplus-difficile le-plus-dur le-plus-éclairant le-plus-formidable le-plus-grand plaisir le-plus-important le-plus-passionnant le-plus-remarquable le-plus-simple le-plussûr le-plus-urgent leitmotiv liberté ligne logique maîtrise mandat manière mérite métier mission mode moment motif motivation moyen mystère nature nec-plus-ultra nécessité normalité objectif objet obsession opinion orientation originalité paradoxe pari parti-pris particularité partie-de-l'art penchant naturel performance plaisir plan point-de-départ pratique premier principal principe priorité problème programme projet propos proposition propre prouesse provocation question question-clé raison raison-d'être réaction réflexe règle remède réponse reproche responsabilité ressort résultat réussite rêve révolution risque rôle ruse sagesse satisfaction sens solution souci souhait spécialité spécificité sport stratégie suggestion surprise suspense tâche tactique technique tendance tentation terme tort tournant tout tradition travail trouvaille truc urgence usage vocation volonté*

NSS pour la CS en *que* : *alibi ambition analyse apparence argument argumentmassue argumentation aspect atout attrait avantage avenir avis axiome beauté-dusport bénéfice bienfait bon-côté bonne-chose but calcul caractéristique certitude chance charme choix chose clé comble conclusion condition conséquence consolation constante constat contrepartie conviction côté-positif coup-de-théâtre crainte critère croustillant danger démonstration dénominateur-commun désir deuxième différence difficulté distinction donnée drame écueil effet-boomerang élément-déterminant élément-important ennui enseignement équité espoir évaluation évidence exigence explication faiblesse fait fin-du-fin fin-mot fond-de-l'affaire fonddu-problème force gag génie grandeur grief hic hypothèse idéal idée impression inconvénient information intérêt ironie jugement l'essentiel l'étonnant l'extraordinaire l'important lacune le-plus-cruel le-plus-curieux le-plus-déplorable le-plus-déprimant le-plus-désolant le-plus-terrible le-plus-fascinant le-mieux lemoins-que-l'on-puisse-dire le-pire le-pire le-plus-absurde le-plus-ahurissant le-plusamusant le-plus-beau le-plus-cocasse le-plus-déroutant le-plus-difficile le-plusdrôle le-plus-dur le-plus-étonnant le-plus-étrange le-plus-fort le-plus-fou le-plusfrappant le-plus-grave le-plus-important le-plus-impressionnant le-plus-incroyable le-plus-inquiétant le-plus-insensé le-plus-intéressant le-plus-intrigant le-plusmarquant le-plus-probable le-plus-spectaculaire le-plus-stupéfiant le-plussurprenant le-plus-triste le-plus-troublant le-plus-vicieux le-plus-vraisemblable lepoint-capital leçon ligne-rouge limite logique malheur marque-du-moment merveilleux message miracle morale mythe noeud non-dit normalité nouveauté nouvelle objectif objection opinion originalité paradoxe pari particularité peur philosophie piment-de-l'affaire point point-essentiel point-négatif point-noir pointpositif point-troublant position postulat premier principal probabilité problème procédé propos propre raison raisonnement réalité récompense regret réponse reproche résultat réussite rêve revers-de-la-médaille risque scandale sentiment signe singularité singulier souci souhait soulagement sujet surprise talent thèse tout trouvaille truc utilité valeur vérité vertu voeu volonté*

NSS communs aux deux sous-constructions : *ambition argument atout avantage avenir but caractéristique chance choix chose clé conséquence consolation conviction crainte critère danger désir deuxième difficulté enseignement espoir exigence idéal idée inconvénient l'essentiel l'important le mieux le plus difficile le plus dur le plus important logique normalité objectif opinion originalité paradoxe particularité pari premier principal problème propos propre raison réponse reproche résultat réussite rêve risque souci souhait surprise tout trouvaille truc volonté*. »

Annexe 2

Liste des 20 NSS les plus fréquents, proposée par Roze et al. (2014)

Objectif

Problème

But

Question

Idée

Rôle

Ambition

Mission

Essentiel

Enjeu

Intérêt

Priorité

Important

Risque

Difficulté

Chose

Souci

Solution

Mérite

Vérité

Annexe 3

Liste des NSS obtenue par projections des patrons

accent, acception, accident, accompagnement, accord, achèvement, acquis, acquisition, acte, action, activité, adepte, adjectif, administration, adulte, adversaire, affaire, affichage, affiche, affirmation, âge, agriculture, aide, air, aise, album, allégorie, alphabet, alternative, ambiance, ambition, an, analogie, analyse, anarchiste, ancêtre, anglais, année, apparition, appétit, apport, apprenant, apprentissage, arbre, ardoise, argument, argumentation, arme, armée, art, article, artiste, aspect, atelier, attaquant, attente, attention, attitude, auditeur, auditoire, augmentation, auteur, automatisme, automobiliste, autonomie, autorisation, avantage, avant-bras, avion, avis, axe, balance, balthazar, banane, barbu, barreau, base, beauté, bénéfice, besoin, biais, bibliothèque, bien, bienfait, binôme, bois, bouche, bouteille, bras, bureau, but, cadre, cage, cahier, camarade, campagne, capacité, capitalisme, caractère, caractéristique, carnet, cartouche, cas, case, catastrophe, catégorie, catégorisation, cathare, centre, certitude, cesse, chaîne, challenge, champagne, chance, changement, charge, charia, chariot, chef, chiffre, chimie, choix, chose, chronologie, chute, citoyenneté, civilité, classe, classeur, clavier, client, clou, coalition, codage, code, coin, collaboration, collectivité, collocation, colonie, communication, communisme, compétence, complément, complexité, comportement, composition, compréhension, compte-rendu, concept, conception, conceptualisation, conclusion, condition, conduit, conflit, confusion, congénère, conjugaison, connaissance, conscience, conseil, consensus, conséquence, consigne, consommateur, constat, consternation, construction, consultation, conte, contenu, contexte, contour, contrainte, contraire, contrat, convergence, conversation, conviction, coopération, coopérative, copie, corpus, correction, côté, couloir, coup, cour, courant, courrier, coutume, couverture, crainte, création, créativité, cri, crime, critère, critique, culture, curiosité, curriculum, cuvée, danger, débat, débutant, décennie, décès, déchiffrage, décision, déclaration, décomposition, découpage, découverte, défaveur, défense, défi, définition, demande, démarche, demie, demi-heure, démission, démocratie, dépendance, dépression, dérivation, désaccord, désavantage, désir, dessin, destinataire, destination, détail, deuil, développement, devoir, diachronie, dialogue, dictée, diction, dictionnaire, différence, différenciation, difficulté, dilution, direction, discipline, discours, discussion, dispositif, disposition, dissimulation, distance, distinction, divergence, division, document, doigt, domaine, domicile, donnée, doute, droit, durée, dynamique, eau, échange, échec, écho, école, écoute, écran, écrit, écriture, écrivain, écueil, effectif, effet, efficacité, effondrement, effort, égard, élément, élève, émetteur, émoi, émotion, emploi, encre, énergie, enfant, engagement, enjeu, ennemi, ennui, énoncé, enquête, enrichissement, enseignant, enseignement, ensemble, entourage, entraînement, entreprise, entretien, envi, envie, environnement, époque, épreuve, erreur, esclavagisme, espace, espoir, esprit, essentiel, étage, étape, état, étonnement, étranger, étudiant, évaluation, événement, éventualité, évolution, exemple, exercice, exigence, expérience, explication, exploit, expression, fable, facilité, façon, faculté, faiblesse, fait, famille, fantasme, fatalité, faute, femme, ferme, feuille, fichier, fiction, film, fin, finalité, flûte, fois, fonction, fonctionnaire, fonctionnement, fond, fondation, force, formateur, formation, forme, formule, forum, fourmilière, français, frein, frise, fruit, frustration, futur, garantie, garçon, gendre, général, genèse, génie, genou, genre, gentillesse, geste, globalisation, globalité, gloire, gomme, goût, gouvernement, grammaire, grammaticalité, graphème, graphie, graphisme, grille, groupe, guerre, guise, habitant, habitude, hauteur, héros, hétérogénéité, heure, hiatus, histoire, homologue, honneur, hypothèse, idéal, idée, identité, illusion, image, imaginaire, imagination, impact, impeachment, impératif, implication, importance, important, impossibilité, impression, incapacité, incertitude, incompatibilité, index, indice, individu, individualisme, inférence, information, initiative, inscription, inspiration, instant, instinct, instituteur, institution, instruction, intelligence, intention, interaction, interdiction, intérêt, intérieur, interlocuteur, interprétation, intertextualité, intervention, intuition, inventaire, invention, investissement, invitation, irréel, islam, jéroboam, jeton, jeu, jeûne, jour, jubilation, justice, justification, juxtaposition, kilomètre, lacune, langage, langue, leçon, lecteur, lecture, légitimation, lendemain, lettre, lexique, liaison, libéralisation, liberté, licence, lien, lieu, ligne, limite, linguiste, linguistique, liste, littérarité, littérature, livre, locuteur, logiciel, logique, loi, loisir, lutte, main, maison, maître, mal, malaise, manière, manifeste, manoeuvre, marché, marketing, marque, matériel, maternelle, mathématique, mathusalem, matière, médecine, médiation, méditation, médium, membre, mémoire, menace, mer, mère, mérite, message, messianisme, mesure, métalangage, méthode, métier, miellat, milieu, mineur, minuscule, minute, mission, mobilité, mode, modèle, modification, mois, moment, monde, mondialisation, montre, morphosyntaxe, mort, mot, motard, motif, motivation, mouvement, moyen, musique, musulman, nabuchodonosor, natation, nature, nécessité, négociation, niveau, nom, nomade, nomenclature, norme, notateur, notation, note, notion, nouveau-né, nouveauté, objectif, objet, obligation, observateur, observation, obstacle, occasion, oeil, ogre, oiseau, opération, opinion, opportunité, opposition, optique, oral, orateur, ordinateur, ordre, organisation, orientation, original, orthographe, outil, ouvrage, ouvroir, pacifisme, page, pair, paix, palais, panique, papi, papier, paradigme, paradoxe, paralysie, parasite, parcelle, parcours, parent, pari, parlementaire, parole, part, partage, partenaire, particularité, partie, passage, passeur, patience, paysage, pédagogie, peine, peinture, pensée, perception, père, période, périphérie, permis, personnage, personne, personnel, personnification, perspective, pertinence, pétrole, peuple, peur, phase, phénomène, philosophie, phonème, photo, phrase, pièce, pied, pierre, pilote, piscine, place, plaisir, plume, poème, poésie, poète, poids, point, polémique, polysémie, pompe, ponctuation, population, portée, portfolio, position, possibilité, postposition, postulat, poutre, pouvoir, pratique, précurseur, préface, prénom, préposition, présence, président, presse, preuve, prière, primaire,

primat, primitif, principe, priorité, privilège, probabilité, problématique, problème, procès, processus, producteur, production, productivité, produit, professeur, profession, profondeur, programme, progrès, progression, projection, projet, pronom, prononce, propos, proposition, propriété, prosodie, psychologue, public, publicité, qualité, quantité, quart, question, questionnaire, quotidien, raison, raisonnement, rang, rappelé, rapport, rassemblement, rasta, ratio, rature, rayonnement, réaction, réalisation, réalité, récepteur, recette, recherche, récit, reconnaissance, rédaction, réécriture, référent, réflexe, réflexion, réflexivité, réforme, refus, regard, régime, région, registre, règle, réinvestissement, relation, relativité, relecture, remarque, remédiation, renommée, répertoire, réponse, représentant, représentation, reprise, républicain, réputation, réserve, responsabilité, ressenti, ressource, restitution, résultat, résumé, retour, rétribution, rétroaction, réunion, réussite, rêve, révision, rigueur, rime, risque, rituel, rôle, roman, rond, route, rythme, sablier, salmanazar, santé, satisfaction, savoir, scénario, scientifique, scientologie, scolarité, score, scripteur, séance, sécurité, sein, sellerie, semaine, sémantique, sémiotique, sénateur, sens, sensation, sentiment, séquence, série, siècle, signal, signature, signe, signification, silence, similitude, simplicité, simplification, site, situation, slogan, société, soin, solution, sommation, son, sondage, sonorité, sorcier, sorte, sortie, souci, souhait, source, spécificité, stage, statut, stratégie, structuralisme, structure, style, stylistique, stylo, sûr, subjectivité, subordination, suggestion, suite, sujet, supercherie, support, surface, surpopulation, surprise, surréalisme, survie, syllabe, synchronie, syntaxe, synthèse, système, tableau, tache, tâche, talent, technologie, temps, tendance, tentative, terme, terrain, terre, terrorisme, test, tête, texte, texture, thématique, thème, théologie, théorie, timidité, titre, toile, ton, touche, tour, trace, tracé, traduction, trait, transformation, transitivité, transparence, travail, tribunal, trou, trouble, tutorat, type, unité, usage, utilisateur, utilisation, utilité, valeur, veau, vecteur, vélo, vendredi, verbalisation, verbe, vérité, versification, victime, vie, violation, violence, virgule, visée, vitesse, vocabulaire, vocation, voie, voiture, voix, volontariat, volonté, voyelle

Annexe 4

Noms présents dans la liste de Legallois (2008) et dans la liste établie pour notre mémoire

acte, action, affaire, alternative, ambition, analyse, apport, argument, argumentation, art , aspect, attitude, avantage, avis, axe, beauté, bénéfice, bienfait, but, caractéristique , certitude, chance, charge, choix, chose, concept, conclusion, condition, conseil, conséquence, consigne, constat, conviction, côté, coup, crainte, critère, danger, débat, décision, démarche , désir, devoir, différence, difficulté, direction, discours, discussion, distinction, donnée, écueil, effet, élément, enjeu, ennui, enseignement, erreur, espoir, essentiel, étape, évaluation, exemple, exigence, explication, facilité, façon, faiblesse, fait, fin, finalité, fonction, force, génie, geste, hypothèse, idéal, idée, illusion, impératif, important, impression, information, intelligence, intention, intérêt, lacune, leçon, liberté, ligne, limite, logique, mal, manière, marque, mérite, message, méthode, métier, mission, mode, moment, motif, motivation, moyen, nature, nécessité, nouveauté, objectif, objet, opinion, orientation, paradoxe, pari, particularité, partie, peur, philosophie, plaisir, point, position, postulat, pratique, principe, priorité, probabilité, problème, programme, projet, propos, proposition, question, raison, raisonnement, réaction, réalité, réflexe, règle, réponse, responsabilité, résultat, réussite, rêve, risque, rôle, satisfaction, sens, sentiment, signe, solution, souci, souhait, spécificité, stratégie, suggestion, sujet, sûr, surprise, tâche, talent, tendance, terme, thèse, travail, usage, utilité, valeur, vérité, vocation, volonté

Noms présents dans la liste de Legallois (2008) mais pas dans celle établie dans notre mémoire

absurde, ahurissant, alibi, amusant, apparence, astuce, atout, attrait , audace, avenir, axiome, bataille, beau, boulot, calcul, cauchemar, charme, cocasse, coeur, comble, confort, consolation, constante, contre-feu, contrepartie, contribution, controverse, courage, croustillant, cruel, culot, curieux, défaut, démonstration, dénominateur, déplorable, déprimant, déroutant, désagréable, désolant, dessein, destin, deuxième, difficile, dignité, directive, don, drame, drôle, dur, éclairant, économique, éloge, énigme, équité, esthétique, éthique, étonnant, étrange, évidence, extraordinaire , face, fascinant, faux-pas, formidable, fort, fou, frappant, gag, gageure, grâce, grandeur, grave, grief, habileté, hantise, hic, hommage, impressionnant, inclination, inconvénient, incroyable, inquiétant, inquiétude, insensé, intéressant, intrigant, ironie, jugement , leitmotiv, malheur, mandat, marquant, merveilleux, mieux, miracle, morale, mystère, mythe, naturel, nec-plus-ultra, noeud, non-dit, normalité, nouvelle, objection , obsession, originalité, passionnant, penchant, performance, pire, plan, premier, principal, probable, procédé, propre, prouesse, provocation, récompense, regret, remarquable, reproche, ressort, revers, révolution, ruse, sagesse, scandale, simple, singularité, singulier, soulagement, spécialité, spectaculaire, sport, stupéfiant, surprenant, suspense, tactique, technique, tentation, terrible, tort, tournant, tout, tradition, triste, troublant, trouvaille, truc, urgence, urgent, vertu, vicieux, voeu, vraisemblable

Annexe 5

NSSP	Freq totale	enAmorce	enCloture	dansPatrons	Chez Legallois	Reliance
ratio	4	0	0	2	Non	0,5000
pacifisme	5	0	0	2	Non	0,4000
sommation	6	0	0	2	Non	0,3333
interdiction	25	0	0	8	Non	0,3200
mérite	14	0	0	4	Oui	0,2857
possibilité	117	7	1	33	Non	0,2821
demie	4	1	0	1	Non	0,2500
dissimulation	4	0	0	1	Non	0,2500
écueil	4	1	0	1	Oui	0,2500
hiatus	4	0	0	1	Non	0,2500
prosodie	4	0	0	1	Non	0,2500
nécessité	97	2	0	22	Oui	0,2268
souci	27	1	1	6	Oui	0,2222
rassemblement	9	0	0	2	Non	0,2222
finalité	15	1	1	3	Oui	0,2000
impossibilité	10	0	0	2	Non	0,2000
nomenclature	10	0	0	2	Non	0,2000
incompatibilité	5	0	0	1	Non	0,2000
inventaire	5	0	0	1	Non	0,2000
jeûne	5	0	0	1	Non	0,2000
métalangage	5	0	0	1	Non	0,2000
texture	5	0	0	1	Non	0,2000
impression	22	0	0	4	Oui	0,1818
fait	372	11	7	66	Oui	0,1774
crainte	17	1	0	3	Oui	0,1765
ambition	24	2	0	4	Oui	0,1667
faculté	18	0	0	3	Non	0,1667
appétit	6	0	0	1	Non	0,1667
bouche	6	0	0	1	Non	0,1667
chimie	6	0	0	1	Non	0,1667
fiction	6	1	0	1	Non	0,1667
littérarité	6	0	0	1	Non	0,1667
parasite	6	0	0	1	Non	0,1667
pompe	6	0	0	1	Non	0,1667
synchronie	6	1	0	1	Non	0,1667
veau	6	0	0	1	Non	0,1667
chance	37	0	0	6	Oui	0,1622
but	129	3	1	20	Oui	0,1550
sentiment	60	2	1	9	Oui	0,1500
obligation	41	0	0	6	Non	0,1463
médecine	14	0	0	2	Non	0,1429
ambiance	7	1	0	1	Non	0,1429
cesse	7	0	0	1	Non	0,1429
décomposition	7	0	0	1	Non	0,1429
globalisation	7	0	0	1	Non	0,1429
primat	7	0	0	1	Non	0,1429
idée	253	6	2	35	Oui	0,1383
volonté	112	2	0	15	Oui	0,1339

conviction	24	1	0	3	Oui	0,1250
charia	8	1	0	1	Non	0,1250
diachronie	8	0	0	1	Non	0,1250
primitif	8	2	0	1	Non	0,1250
subordination	8	0	0	1	Non	0,1250
habitude	25	1	0	3	Non	0,1200
plaisir	9	0	0	1	Oui	0,1111
conduit	10	0	0	1	Non	0,1000
dépression	10	1	0	1	Non	0,1000
périphérie	10	0	0	1	Non	0,1000
permis	10	0	0	1	Non	0,1000
précurseur	10	1	0	1	Non	0,1000
refus	31	0	0	3	Non	0,0968
honneur	21	0	0	2	Non	0,0952
productivité	21	1	0	2	Non	0,0952
futur	33	1	0	3	Non	0,0909
soin	22	0	0	2	Non	0,0909
gloire	11	0	0	1	Non	0,0909
intuition	11	0	0	1	Non	0,0909
rayonnement	11	1	0	1	Non	0,0909
garantie	24	0	0	2	Non	0,0833
rêve	24	0	0	2	Non	0,0833
anarchiste	12	1	1	1	Non	0,0833
décès	12	0	0	1	Non	0,0833
genèse	12	0	0	1	Non	0,0833
incapacité	12	1	0	1	Non	0,0833
propos	37	2	0	3	Oui	0,0811
intention	75	2	0	6	Oui	0,0800
autorisation	25	0	0	2	Non	0,0800
risque	163	4	1	13	Oui	0,0798
quotidien	13	0	0	1	Non	0,0769
relativité	13	0	0	1	Non	0,0769
moyen	275	8	0	21	Oui	0,0764
effet	295	10	2	22	Oui	0,0746
opportunité	28	0	0	2	Non	0,0714
balance	14	1	0	1	Non	0,0714
certitude	14	1	0	1	Oui	0,0714
idéal	14	0	0	1	Oui	0,0714
impératif	14	1	0	1	Oui	0,0714
pari	14	1	0	1	Oui	0,0714
rédaction	14	0	0	1	Non	0,0714
tâche	59	0	0	4	OUi	0,0678
espoir	30	1	0	2	Oui	0,0667
incertitude	30	0	0	2	Non	0,0667
apprenant	15	0	0	1	Non	0,0667
prière	15	0	0	1	Non	0,0667
profession	15	0	0	1	Non	0,0667
vecteur	15	0	0	1	Non	0,0667
objectif	198	12	1	13	Oui	0,0657
catastrophe	16	0	0	1	Non	0,0625

devoir	16	0	0	1	Oui	0,0625
geste	16	0	0	1	Oui	0,0625
exigence	51	1	0	3	Oui	0,0588
collectivité	17	0	0	1	Non	0,0588
vocation	17	1	0	1	Oui	0,0588
hypothèse	104	1	0	6	Oui	0,0577
tentative	53	1	0	3	Non	0,0566
raison	195	13	1	11	Oui	0,0564
copie	18	1	0	1	Non	0,0556
fonctionnaire	18	0	1	1	Non	0,0556
paradoxe	18	1	0	1	Oui	0,0556
polémique	18	3	0	1	Non	0,0556
sémiotique	18	0	0	1	Non	0,0556
occasion	73	0	0	4	Non	0,0548
air	37	2	0	2	Non	0,0541
interlocuteur	38	1	1	2	Non	0,0526
lecteur	19	0	0	1	Non	0,0526
licence	19	0	0	1	Non	0,0526
réputation	20	1	0	1	Non	0,0500
perception	42	2	1	2	Non	0,0476
cathare	21	1	0	1	Non	0,0476
manière	233	9	2	11	Oui	0,0472
émetteur	22	0	0	1	Non	0,0455
peur	22	2	0	1	Oui	0,0455
registre	22	1	0	1	Non	0,0455
constat	24	1	0	1	Oui	0,0417
problématique	49	2	0	2	Non	0,0408
adepte	25	0	0	1	Non	0,0400
quart	25	1	0	1	Non	0,0400
essentiel	51	3	1	2	Oui	0,0392
façon	260	13	4	10	Oui	0,0385
capacité	208	4	5	8	Non	0,0385
libéralisation	52	2	0	2	Non	0,0385
priorité	52	2	0	2	Oui	0,0385
lendemain	26	2	0	1	Non	0,0385
transitivité	26	1	0	1	Non	0,0385
syntaxe	79	1	0	3	Non	0,0380
esprit	133	7	0	5	Non	0,0376
collaboration	27	1	0	1	Non	0,0370
mondialisation	114	6	0	4	Non	0,0351
preuve	57	0	0	2	Non	0,0351
instant	29	0	2	1	Non	0,0345
profondeur	29	3	0	1	Non	0,0345
alternative	30	3	0	1	Oui	0,0333
avis	30	0	0	1	Oui	0,0333
désir	31	4	0	1	Oui	0,0323
vérité	31	0	1	1	Oui	0,0323
question	507	36	4	16	Oui	0,0316
besoin	127	2	1	4	Non	0,0315
TOTAL	**6519**	**233**	**43**	**559**		

Annexe 6

```
#19 mai 2015
#Lancement du programme : perl -CIOA Annexe_6.pl CHEMINVERSLECORPUS
use locale;
use utf8;


my $Corpus = $ARGV[0];

opendir('DIR',$Corpus) or die "USAGE $0: must put all files in folder
$Corpus";

my @aDir = readdir(DIR);
foreach my $file (@aDir){

    my $entree=$Corpus.$file;

    if ($file =~ /\.tal$/) {

      open (TAL, "<", $entree) or die "USAGE $0: no file $entree";
      binmode(TAL, ":encoding(utf8)");

      my (@Occ, @Lem, @Pos, @Nombre, @Genre, @Offset, @NSSoffset_debut,
@NSSoffset_fin);
      my $nummot = 0;

      while (my $ligne = <TAL>) {
          chomp $ligne;

          if ($ligne =~ /^\d/){

            my @tal = split("\t",$ligne);

            $nummot = $tal[0];
            $Occ[$nummot]=lc($tal[1]);
            $Lem[$nummot]=lc($tal[2]);
            $Pos[$nummot]=$tal[3];
            $Nombre[$nummot] = $Genre[$nummot] = "";
            if ($tal[5]=~/n=(s|p)/){
                $Nombre[$nummot]=$1;
            }
            if ($tal[5]=~/g=(f|m)/){
                $Genre[$nummot]=$1;
            }
            $Offset[$nummot]=$tal[7];
            my $corpus=$file;
            $NSSoffset_debut[$nummot]=$Offset[$nummot];

      $NSSoffset_fin[$nummot]=$NSSoffset_debut[$nummot]+length($Occ[$numm
ot]);


            }
            else {
                  for (my $i = 1; $i <= $nummot; $i++){
```

```
                #correspond au patron (DET) (NC) (Ce) (V Ãªtre) ((que-
clause|de-inf)) :
                if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^NC$/ &&
$Lem[$i+2] =~ /^ce$/ && $Lem[$i+3] =~ /^Ãªtre$/ && $Occ[$i+4] =~
/que$|de$/){
                      print $file, "\t","Patron 1\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (DET) (NC) (V Ãªtre) ((que-
clause|de-inf)) :
                if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^NC$/ &&
$Lem[$i+2] =~ /^Ãªtre$/ && $Occ[$i+3] =~ /que$|de$/){
                      print $file, "\t","Patron 2\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }

                #correspond au patron (NC) (CS que) :
                if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^NC$/ &&
$Pos[$i+2] =~ /^CS$/ && $Lem[$i+2] =~ /^que$/ && $Pos[$i+3] =~ /^DET$/){
                      print $file, "\t","Patron 3\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (NC) (P de) (V inf) :
                if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^NC$/ &&
$Occ[$i+2] =~ /^de$/ && $Pos[$i+3] =~ /^VINF$/){
                      print $file, "\t","Patron 4\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (NC) (PONCT :) :
                if ($Occ[$i] =~ /^(le|la|leur|leurs)$/ && $Pos[$i+1] =~
/^NC$/ && $Occ[$i+2] =~ /:/){
                      print $file, "\t","Patron 5\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (P pour) (NC) (P de) :
                if ($Occ[$i] =~ /^pour$/ && $Pos[$i+1] =~ /^NC$/ &&
$Occ[$i+2] =~ /^de$/ && $Pos[$i+3] =~ /^VINF/){
                      print $file, "\t","Patron 6\t",$Lem[$i+1],"\t",
$NSSoffset_debut[$i+1], "\t", $NSSoffset_fin[$i+1],"\n";
                }


           }
#Purge des tableaux
     @Occ = @Lem = @Pos = @Nombre = @Genre = @Offset = @NSSoffset_debut
= @NSSoffset_fin = ();

         }
      }
    }
}
```

Annexe 7

Patron	Phrase	NSS
1	Pour moi la **poésie**, c'est de la musique	Non
1	Le seul **moyen** de combler cette envie est d'acheter le produit	Oui
1	Le **problème** c'est que la marche à suivre pour réussir ces évaluations est rarement renseignée	Oui
1	La seule **façon** d'apprendre le langage, c'est de le pratiquer	Oui
1	La seule **manière** d'apprendre le langage, c'est de l'utiliser en communiquant	Oui
2	**L'enjeu central** de la recherche présenté dans cet article est de favoriser avec les jeunes enseignants le renouvellement de pratiqu	Oui
2	Sa **durée** est de quarante minutes.	Non
2	**L'objectif** est de repérer le temps des verbes du texte précité.	Oui
2	**L'objectif** est de rechercher le pronom personnel de chaque verbe conjugué.	Oui
2	**L'utilité** est de savoir se réguler, se faire confiance.	Oui
2	La **consigne** est de repérer les erreurs et de réécrire la dictée sans erreur sur la feuille A2.	Oui
2	La **consigne** est de regrouper les mots par famille.	Oui
2	La **compétence** est de connaître la formation des mots afin de les comprendre.	Oui
2	La procédure la plus payante pour retenir un **mot** est de passer par la catégorisation.	Non
2	Le but de cette omniprésence de l'écrit dans la **classe** est de faire comprendre aux enfants à quoi sert l'écrit.	Non
3	Elles montrent à leurs **élèves** que chaque communauté a son écrit, son alphabet.	Non
3	Le **fait** que le crayon dépasse de leur main doit les gêner.	Oui
3	Il est ressorti lors de ces **entretiens** que les trois enseignantes créent un langage à la portée des enfants.	Non
3	Outre le **fait** que les verbes structurent l'action	Oui
3	Le **fait** que la communication ne consiste pas seulement à un envoyer un message mais aussi à savoir comment il a été reçu.	Oui
3	La **notion** que les méthodes éprouvées étaient bien préférables.	Oui
3	Le **fait** que l'industriel n'était implanté que dans un seul état.	Oui
3	Des études en génétique aboutissent à **l'idée** que la population originelle pour tous les humains se situait en Afrique.	Oui
3	Le **fait** que les deux espèces de Chimpanz& soient considérées comme les espèces vivantes les plus proches de l'Homme.	Oui
3	La démarche polygénétique part de **l'idée** que la vie évolue des formes les plus simples aux plus organisées.	Oui
4	Convaincus de la **nécessité** de respecter l'être humain	Oui
4	Dans le **but** de modifier la descendance de la personne	Oui
4	Ce qui reviendrait à **l'interdiction** de porter atteinte à l'intégrité de l'espèce humaine	Oui
4	Les rastas le concoivent comme un mode de vie, une **façon** de concevoir le monde et tout ce qui le constitue depuis sa création	Oui
4	**L'idée** de voir enfin le roi des rois	Oui
4	Le **débat** de savoir si les dreads sont nécessaires à un rasta est encore important de nos jours	Oui

4	Les rastas vont ainsi inventer un grand nombre de mots qui reflètent leur **façon** de voir le monde	Non
4	Il avait pris **l'habitude** de conserver ses cheveux	Oui
4	Il me semble quon peut avoir deux **manières** de considérer son statut :	Oui
4	Ils ont désormais la **volonté** de faire une place à un plus grand nombre de pays	Oui
5	Deux autres résolutions concurrentes avaient été proposées à la **Chambre** : La résolution Durbin-Benett et la résolution Gephardt-	Non
5	Le **quart** : 18,75 ou 20cl	Non
5	La **demie** : 37,5 cl	Non
5	Le **médium** : 60 cl	Non
5	Le **mathusalem** : 6l	Non
5	Les diplomates peuvent y contribuer en recherchant de nouvelles méthodes permettant d'atteindre plus souplement le **consensus**	Oui
5	Otto Jespersen et Roman Jakobson ont ensuite déplacé la **problématique** : d'acte individuel de production, l'énonciation devient t	Oui
5	à la **question** : s'agit-il d'authentiques démocraties ?	Oui
5	L'information fournie se mesure selon la **formule** : I=logPx.	Oui
5	Ce débat est propre à l'ensemble des Etats de la **région** : la démobilisation entraînera-t-elle un recul d'influence des militaires ?	Non
6	Les organisations ont davantage pour **fonction** de procurer des revenus non imposables	Oui
6	Ce qui eut pour **effet** de modifier en profondeur la politique étrangère du pays	Oui
6	Cette loi a pour **objet** de limiter les pouvoir du président	Oui
6	La guerre froide a eu pour **effet** de neutraliser totalement les parlementaires dans le domaine des relations internationnales	Oui
6	Ce mépris de la constitution a eu pour **effet** de renforcer les rangs des opposants	Oui
6	En fait, les débats décrits ici n'ont pas eu pour **effet** de bloquer la décisions de Georges W Bush	Oui
6	Cela a pour **effet** de réduire encore davantage l'importance des querelles partisanes	Oui
6	Les orientations ne sont pas toutes connues mais ont eu pour **objectif** de renforcer la sécurité intérieure	Oui
6	Les propositions de l'administration en matière de sécurité dans les aéroports ont ainsi eu pour **effet** de diviser les membres de Co	Oui
6	Les prises de position contre une guerre en Irak ont également eu pour **effet** de diviser le camp démocrate au Sénat	Oui

Annexe 8

```
#commande : perl extractNSS_offset_XXX.pl CORPUS LEXIQUE
#CORPUS = chemin du dossier contenant les fichiers
#entrée : fichiers issus de l'annotation ANNODIS_me : .aa (xml contenant
les annotations) et .tal (fichiers .ac contenant le texte brut analysés
(POStaggés) par Talismane)
#sortie :  STDOUT, sortie tabulée où chaque ligne correspond à une
occurrence de NSS potentiel
use strict;
use locale;
use warnings;
use utf8;

my $Corpus = $ARGV[0];
my $Lexique = $ARGV[1];

open (RESULT, ">result.txt");
open (CONTEXT, ">context.csv");
#en-têtes de la sortie
print RESULT
"idPhrase\tcorpus\toccurrence\tlemme\tgenre\tnombre\tenAmorce\tenCloture\
n";
print CONTEXT "id\tcontexte\n";

opendir('DIR',$Corpus) or die "USAGE $0: must put all files in folder
$Corpus";

my @aDir = readdir(DIR);

my %lex;

open (LEX, "<", $Lexique) or die "Problème ouverture fichier $Lexique\n";
binmode(LEX, ":encoding(utf8)");

while (my $mot = <LEX>){
    chomp $mot;
    if ($mot =~ /^([^ ]+)/){#on ne prend que le premier mot
      $lex{lc($1)}=1;
    }
}
close (LEX);


foreach my $file (@aDir){
    my @item; #tableau qui contiendra les offset des items
    my @cloture; #tableau qui contiendra les offset des clôtures
    my @amorce; #tableau qui contiendra les offset des amorces

    my $entree=$Corpus.$file;

    if ($file =~ /\.aa$/) {
warn "Traitement de $file\n";
           open (AA, "<", $entree) or die "USAGE $0: no file $entree";
binmode(AA, ":encoding(utf8)");

     my $unit = ""; #variable qui indique le type d'unité dont on lit
l'annotation
```

```
     my $offset = ""; # variable qui contiendra la position de début et
de fin de l'unité

     while (my $ligne = <AA>) {
         chomp $ligne;
         if ($ligne eq "</unit>"){
           #si des infos ont été enregistrées car une unité de type
amorce, item ou cloture a été rencontrée, on enregistre les offset dans
le tableau correspondant
           if ($unit eq "amorce"){
               push (@amorce, $offset);
           }
           elsif ($unit eq "cloture"){
               push (@cloture, $offset);
           }
           elsif ($unit eq "item"){
               push (@item, $offset);
           }

           #réinitialisation des variables à chaque nouvelle unité
           $unit = "";
           $offset = "";
         }
         elsif ($ligne =~/<type>(amorce|cloture|item)<\/type>/) {
           $unit = $1;
         }
         elsif ($unit ne "" && $ligne =~/<singlePosition index=\"([0-
9]+?)\" \/>/ && $offset eq "") {#on est sur la position start
           $offset = $1;
         }
         elsif ($unit ne "" && $ligne =~/<singlePosition index=\"([0-
9]+?)\" \/>/) {#on est sur la position end
           $offset = $offset."-".$1;
         }
     }
   close (AA);
   #a partir de là les tableaux contiennent les offset des segments
annotés
   #il faut maintenant lancer la recherche des NSS sur le .ac ou le
.acTalismané (.tal)
   $entree =~ s/\_(coder.|gold)\.aa/\.tal/;
     warn "Traitement de $entree\n";

   open (TAL, "<", $entree) or die "USAGE $0: no file $entree";
binmode(TAL, ":encoding(utf8)");
   my $idPhrase=1;
   my $idNSS=0;
   my $context="";
   my $NSS = 0;

     while (my $ligne = <TAL>) {
         chomp $ligne;
           if ($ligne ne ""){
                 my @tal = split("\t",$ligne);
                 if ($context ne "" && $tal[0] eq "1"){
                       print CONTEXT $file."Phr$idPhrase\t$context\n";
                       $idPhrase++;
                       $context = "";
```

```
                }
                else {
                      my $Occ=$tal[1];
                      if ($Occ){
                            $context = $context." ".$Occ;
                      }

                      #chercher les NSS uniquement pour les lignes avec
POStag NC
                      if ($ligne =~ /\tNC\t/){

                      my $Lem=lc($tal[2]);
                      my $Pos=$tal[3];
                      my $Nombre = my $Genre = "";
                      if ($tal[5]=~/n=(s|p)/){
                            $Nombre=$1;
                      }
                      if ($tal[5]=~/g=(f|m)/){
                            $Genre=$1;
                      }
                      my $Offset=$tal[7];
                      my $corpus=$file;
                      $corpus =~ s/^.*(geop|wik2|ling).*$/$1/;
                      my $NSSoffset_debut=$Offset;#variable qui contient
l'offset de début du NSS
                      my
$NSSoffset_fin=$NSSoffset_debut+length($Occ);#variable qui contient
l'offset de fin du NSS =$NSSoffset_debut+length(NSS)

                      #Si c'est un NSS potentiel, imprimer les
informations de cette occurrence
                      if (defined($lex{$Lem}) and $Pos eq "NC"){
                            $idNSS++;
                            print RESULT
$file."Phr".$idPhrase."\tNSS".$idNSS."\t$corpus\t$Occ\t$Lem\t$Genre\t$Nom
bre";

                            #Si c'est un NSS, alors passer en revue les 3
tableaux pour voir si le nom apparaît entre les offset d'un segment de SE
                            my $nbIn = 0;#variable qui permet d'imprimer
si oui ou non cette occurrence de NSS est apparu dans l'unité amorce /
cloture / item

                            for my $unit (@amorce){
                                  if ($unit =~ /([0-9]+)\-([0-9]+)/){
                                        my $debut = $1;
                                        my $fin = $2;
                                        if ($NSSoffset_debut >= $debut &&
$NSSoffset_fin <= $fin){
                                              #alors le NSS est dans une
amorce et on modifie la valeur de la variable $print
                                              $nbIn++;
                                        }
                                  }
                            }
                            print RESULT "\t$nbIn";
                            $nbIn = 0; # on réinitialise
```

```
                              for my $unit (@cloture){
                                    if ($unit =~ /([0-9]+)\-([0-9]+)/){
                                          my $debut = $1;
                                                            my $fin = $2;
                                          if ($NSSoffset_debut >= $debut &&
$NSSoffset_fin <= $fin){
                                          #alors le NSS est dans une
cloture
                                                $nbIn++;
                                          }
                                    }
                              }
                              print RESULT "\t$nbIn\n";
                              $nbIn = 0; # on réinitialise


                        }#si lemme NC et défini dans liste NSS

                        }#si ligne fichie TAL = NC
                  }#si tab[0] ne "1"
            }#si ligne de TAL non vide
      }#while dans TAL

      }#si fichier .aa

  }#while dans le dossier corpus
```

Annexe 9

```
#Lancement du programme : perl -CIOA Annexe_9.pl CHEMINVERSLECORPUS
use locale;
use utf8;


my $Corpus = $ARGV[0];

opendir('DIR',$Corpus) or die "USAGE $0: must put all files in folder
$Corpus";

my @aDir = readdir(DIR);
foreach my $file (@aDir){

    my $entree=$Corpus.$file;

    if ($file =~ /\.tal$/) {

      open (TAL, "<", $entree) or die "USAGE $0: no file $entree";
      binmode(TAL, ":encoding(utf8)");

      my (@Occ, @Lem, @Pos, @Nombre, @Genre, @Offset, @NSSoffset_debut,
@NSSoffset_fin);
      my $nummot = 0;

      while (my $ligne = <TAL>) {
          chomp $ligne;

          if ($ligne =~ /^\d/){

            my @tal = split("\t",$ligne);

            $nummot = $tal[0];
            $Occ[$nummot]=lc($tal[1]);
            $Lem[$nummot]=lc($tal[2]);
            $Pos[$nummot]=$tal[3];
            $Nombre[$nummot] = $Genre[$nummot] = "";
            if ($tal[5]=~/n=(s|p)/){
                $Nombre[$nummot]=$1;
            }
            if ($tal[5]=~/g=(f|m)/){
                $Genre[$nummot]=$1;
            }
            $Offset[$nummot]=$tal[7];
            my $corpus=$file;
            $NSSoffset_debut[$nummot]=$Offset[$nummot];

      $NSSoffset_fin[$nummot]=$NSSoffset_debut[$nummot]+length($Occ[$numm
ot]);


            }
            else {
                  for (my $i = 1; $i <= $nummot; $i++){

                  #correspond au patron (DET) (NSS) (V) (VINF) :
                  if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^NC$/ &&
$Pos[$i+2] =~ /^V$/ && $Pos[$i+3] =~ /^VINF$/){
```

```
                    print $file, "\t","Patron DET NSS V
VINF\t",$Lem[$i+1],"\t", $NSSoffset_debut[$i+1], "\t",
$NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (DET ce cette ces) (NC) :
                if ($Lem[$i] =~ /^(ce|cette|ces$)/ && $Pos[$i+1] =~
/^NC$/){
                    print $file, "\t","Patron DETce cette ces
NSS\t",$Lem[$i+1],"\t", $NSSoffset_debut[$i+1], "\t",
$NSSoffset_fin[$i+1],"\n";
                }

                #correspond au patron (DET) (ADJ) (NC) :
                if ($Pos[$i] =~ /^DET$/ && $Pos[$i+1] =~ /^ADJ$/ &&
$Pos[$i+2] =~ /^NC$/){
                    print $file, "\t","Patron DET ADJ
NSS\t",$Lem[$i+2],"\t", $NSSoffset_debut[$i+1], "\t",
$NSSoffset_fin[$i+1],"\n";
                }
                #correspond au patron (P de|par) (ADJ) (NC) :
                if ($Occ[$i] =~ /^(de|par)$/ && $Pos[$i+1] =~ /^ADJ$/ &&
$Pos[$i+2] =~ /^NC$/){
                    print $file, "\t","Patron P de| par ADJ
NSS\t",$Lem[$i+2],"\t", $NSSoffset_debut[$i+2], "\t",
$NSSoffset_fin[$i+2],"\n";
                }
                #correspond au patron (NC) (ADJ) :
                if ($Pos[$i] =~ /^NC$/ && $Pos[$i+1] =~ /^ADJ$/){
                    print $file, "\t","Patron NSS
ADJ\t",$Lem[$i],"\t", $NSSoffset_debut[$i], "\t",
$NSSoffset_fin[$i],"\n";
                }
                #correspond au patron (ADV) (P de) (NC) :
                if ($Pos[$i] =~ /^ADV$/ && $Occ[$i+1] =~ /^de$/ &&
$Pos[$i+2] =~ /^NC$/){
                    print $file, "\t","Patron ADV Pde
NSS\t",$Lem[$i+2],"\t", $NSSoffset_debut[$i+2], "\t",
$NSSoffset_fin[$i+2],"\n";
                }
                #correspond au patron (DET numeral) (NC):
                if ($Occ[$i] =~ /^deux|trois|quatre|cinq|six|sept$/ &&
$Pos[$i+1] =~ /^NC$/){
                    print $file, "\t", "Patron DET numeral NSS\t",
$Lem[$i+1], "\t", $NSSoffset_debut[$i+1], "\t",
$NSSoffset_fin[$i+1],"\n";
                }


# correspond au patron (Quelques-uns | quelques-unes) (ART des) (NC):
                if ($Occ[$i] =~ /^quelques$/ && $Occ[$i+2] =~ /unes|uns/
&& $Occ[$i+3] =~ /^des$/ && $Pos[$i+4] =~ /^NC$/){
                    print $file, "\t", "Patron DET ART NSS\t",
$Lem[$i+4], "\t", $NSSoffset_debut[$i+4], "\t",
$NSSoffset_fin[$i+4],"\n";
                }


            }
```

```
      @Occ = @Lem = @Pos = @Nombre = @Genre = @Offset = @NSSoffset_debut
= @NSSoffset_fin = ();

           }
      }
    }
}
```