Notion de position du modèle GENELEX et structuration d'une base de données syntaxiques issue des Tables du LADL

Position in the GENELEX model and structuring of a syntactical Data Base originating from the LADL tables

Philippe Alcouffe et Bruno Revellin-Falcoz

Édition électronique

URL : <http://journals.openedition.org/praxematique/2261>

DOI : 10.4000/praxematique.2261

ISSN : 2111-5044

Éditeur

Presses universitaires de la Méditerranée

Édition imprimée

Date de publication : 1 janvier 1994

Pagination : 81-104

ISSN : 0765-4944

Référence électronique

Philippe Alcouffe et Bruno Revellin-Falcoz, « Notion de position du modèle GENELEX et structuration d'une base de données syntaxiques issue des Tables du LADL », *Cahiers de praxématique* [En ligne], 22 | 1994, document 4, mis en ligne le 01 janvier 2013, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/praxematique/2261> ; DOI : <https://doi.org/10.4000/praxematique.2261>

Ce document a été généré automatiquement le 8 septembre 2020.

Tous droits réservés

Notion de position du modèle GENELEX et structuration d'une base de données syntaxiques issue des Tables du LADL

*Position in the GENELEX model and structuring of a syntactical Data Base
originating from the LADL tables*

Philippe Alcouffe et Bruno Revellin-Falcoz

1. Introduction

- 1 Les travaux présentés ici ont été réalisés avec le Laboratoire d'Automatique Documentaire et Linguistique, Université de Paris VII (LADL) au sein du projet EUREKA GENELEX. Ce projet européen se propose d'établir un format standard de représentation des données lexicales couvrant l'ensemble des informations utiles sur la langue : le modèle conceptuel GENELEX.
- 2 Nous présentons certains aspects de ce modèle concernant la représentation des données syntaxiques (section 2). Celle-ci repose essentiellement sur la notion de Position. La souplesse d'expression du modèle qui se veut « théorie accueillant » implique une forte paramétrisation de cette notion (section 3).
- 3 Nous montrons donc comment une vue relationnelle de la Position impose des principes structurants forts au niveau du lexique décrit. Ces principes ont été mis en œuvre pour la constitution d'une base de données syntaxiques tirée des tables du LADL et ont permis de valider en grandeur réelle la modélisation choisie (section 4).
- 4 Le travail de mise au format GENELEX de tables du LADL se fait à travers le logiciel AZOTE. Ce logiciel compile des règles d'expertise exprimant de façon formelle les propriétés syntaxiques GENELEX déduites des informations contenues dans les tables,

et traduit ainsi automatiquement les données en objets GENELEX maintenus sous une base de données objet (section 5).

- 5 Nous prendrons pour exemple la règle d'expertise de la table 11 et montrerons sur la base résultante la souplesse de partitionnement obtenue grâce à la modélisation paramétrée choisie (sections 6 et 7).
- 6 Ce travail sur les tables offre pour la première fois à la communauté linguistique une lisibilité et un accès enrichi aux données et aux propriétés enregistrées par le LADL.

2. Modèle syntaxique GENELEX

- 7 Nous présentons tout d'abord une vue résumée des principaux éléments entrant dans la formalisation de la couche syntaxique du modèle GENELEX (GENELEX, 1993).
- 8 Dans le modèle syntaxique, l'entrée est l'Unité Syntaxique (*Usyn*) : associée à une Unité Morphologique (*Um*), elle en note un comportement syntaxique particulier ; une *Um* donnée peut donc avoir plusieurs *Usyn*.
- 9 Un comportement syntaxique est décrit par une *Description Syntaxique* : une *Usyn* a une *Description Syntaxique* de Base et peut avoir des *Descriptions Syntaxiques* Transformées.
- 10 La *Description Syntaxique* est l'association de deux entités : *Self* et *Construction Syntaxique*. Ce partage permet deux visions du modèle :
 - une vision atomiste où la *Description* note une structure argumentale caractérisant les liens qu'entretient l'unité décrite (*Self*) avec une liste « d'arguments » (la *Construction*).
 - une vision syntaxiste où la *Description* décrit un contexte d'apparition sous forme d'arborescence syntaxique (la *Construction*), et l'insertion de l'unité décrite (*Self*) dans ce contexte.
- 11 Dans les deux cas, *Self* porte des informations considérées « internes » à l'unité (emploi réflexif, à un temps particulier, etc), la *Construction* notant des informations « externes » (complémentation, valence, etc). *Self* est indépendant de la graphie : c'est l'*Usyn* qui porte le lien vers l'*Um*. Ainsi des *Usyn* liées à des *Um* différentes peuvent partager des *Descriptions Syntaxiques* ; de même, des *Descriptions Syntaxiques* associées à des *Self* différents peuvent partager des *Constructions Syntaxiques*.
- 12 Une *Construction Syntaxique* est décrite par une liste de *Positions* qui portent une liste d'occupants : les *Syntagmes*, par lesquels elles peuvent être instanciées dans une actualisation donnée. Chaque *Position* décrit ainsi un paradigme de réalisation et peut être typée par une fonction syntaxique et une liste de rôles thématiques. L'optionnalité des *Positions* et les conditions sur leur instanciation sont éventuellement précisées. Un *Syntagme* peut être récrit sous forme de liste de *Positions*.
- 13 Le modèle permet enfin de noter des transformations entre *Usyn*, *Descriptions* ou *Syntagmes*.

3. Notion de Position GENELEX

3.1 Définition

- 14 La Position est une abstraction sur plusieurs notions traditionnelles de la littérature sur la complémentation verbale (Milner, 1989) (Tesnière, 1959) (Fillmore 1968). GENELEX, plutôt que d'ajouter une nième notion, préfère offrir à chacun, à travers un seul concept, un moyen de représenter la même perception des comportements syntaxiques que celle offerte par sa notion habituelle.
- 15 La Position est donc une notion essentiellement paramétrable. Une fois ses paramètres définitoires fixés, elle impose une structuration forte du dictionnaire.
- 16 Une Position GENELEX est définie par trois éléments :
 - une fonction syntaxique
 - une liste de rôles thématiques
 - une distribution syntagmatique i.e. une liste de syntagmes dits « occupants » de la Position.
- 17 Lorsque fonction et rôles thématiques sont renseignés, il doit y avoir invariance de ces éléments pour tous les occupants de la distribution.
- 18 Ces trois éléments sont paramétrables :
 - l'ensemble des fonctions et des rôles thématiques utilisable est défini par l'utilisateur
 - la distribution est définitoire mais son mode d'obtention est libre. Elle permet, notamment dans le cas où fonction et rôles thématiques ne sont pas renseignés, d'opérer une définition « par extension » de la Position.
- 19 La notion de Position permet de partitionner l'ensemble des syntagmes apparaissant en cooccurrence avec l'unité décrite. Les listes de Positions ainsi dégagées permettent de constituer les Constructions Syntaxiques en ajoutant des informations d'ordre linéaire et d'optionnalité. Ainsi, la notion de Position est à la base de la structuration des comportements syntaxiques enregistrés dans un format GENELEX.

3.1.1 Fonction

- 20 La liste des fonctions possibles étant modifiable par l'utilisateur, GENELEX n'impose pas de définition de la notion de fonction. Une liste de valeurs est proposée : tête, sujet, objet direct, objet indirect, attribut sujet, attribut objet, épithète gauche, épithète droit, spécifieur, génitif.
- 21 On retrouve là les éléments traditionnels de la fonction syntaxique vue comme une relation entre deux syntagmes et caractérisée par des altérations morphologiques classiques : accords, marques de cas, pronominalisations en tel ou tel pronom.

3.1.2 Rôles Thématiques

- 22 GENELEX propose la liste de valeurs : agent, patient, destinataire, source, but, cause, manière, locatif, temps, instrument, thème.
- 23 L'utilisation de rôles thématiques sans définir de fonctions permet une représentation syntaxique s'apparentant aux grammaires sémantiques, les Positions devenant alors des arguments typés par leur rôle.

- 24 La possibilité d'utiliser des rôles thématiques en syntaxe n'empêche nullement GENELEX de permettre une modélisation autonome de la sémantique avec ses propres structures argumentales ainsi qu'une représentation des correspondances syntaxiques associées.

3.1.3 Distribution

- 25 La distribution est le seul élément définitoire de la Position lorsque fonction et rôles thématiques ne sont pas renseignés. La Position est alors définie en « extension » et les critères ayant présidé à cette énumération ne sont pas explicités dans la modélisation choisie. En particulier, si un quelconque « critère d'invariance » a été utilisé, il n'apparaît pas explicitement.

3.2 Construction

- 26 Les Constructions Syntaxiques sont des abstractions sur les listes de Positions définies : la Construction s'abstrait de l'ordre linéaire et des invariances induites de l'optionnalité des Positions et de la liberté d'ordre d'apparition des occupants.

3.2.1 Ordre linéaire

- 27 La seule information d'ordre portée par la Construction est l'ordre « canonique » de sa liste de Positions. Dans le cas où une entrée lexicale interdit les bouleversements linéaires habituels, cette marque ne pourra être portée en tant que telle¹. Toutefois, rien n'interdit alors d'avoir une définition de la Position l'associant explicitement à une place fixée par rapport à l'unité décrite. On peut prendre deux exemples : la pronominalisation et l'inversion sujet objet.
- 28 GENELEX conseille la représentation de la pronominalisation d'un occupant de Position par une transformation interne à la Position, le placement en surface du pronom obéissant alors aux règles syntaxiques habituelles. On peut toutefois utiliser une Position spécifique définie soit par une distribution de pronoms, soit par une fonction « clitique gauche », les conditions entre occupants de Positions distinctes excluant la réalisation conjointe en surface du syntagme et de son pronom dans le cas de non reprise.
- 29 De même, dans le cas d'entrée lexicale autorisant l'inversion sujet objet, on pourra utiliser des fonctions « sujet inversé » et « objet inversé ».
- 30 A travers ces exemples, on voit le rôle primordial qu'auront les choix définitoires de la Position dans la structuration du lexique.

3.2.2 Optionnalité

- 31 Le deuxième facteur regroupant les listes de Positions en une Construction est l'optionnalité : deux listes identiques à la présence d'une Position près sont regroupées en une seule Construction ; la Position « différentielle » est alors déclarée optionnelle quant à sa réalisation en surface.
- 32 Par exemple, les deux listes de Positions suivantes :
- qqun boire qqchse : P0 P1

- qqun boire : P0

- 33 seront factorisées en la Construction P0 (P1) : la non réalisation de P1 n'affecte pas P0 et il y a alors coïncidence des deux listes.
- 34 Des optionalités conditionnées peuvent aussi être notées : « P0 (P1) (P2) si P1 alors P2 » indiquera que l'on peut avoir P2 sans P1 mais pas le contraire.

3.2.3 Conditions entre occupants de Positions distinctes

- 35 GENELEX offre la possibilité d'exprimer des conditions de cooccurrence ou d'exclusion mutuelle entre occupants de Position, ce qui évite l'éclatement en deux constructions complémentaires.
- 36 Par exemple, si l'on choisit d'utiliser deux Positions Complément distinctes pour décrire l'unité « déboucher » (« le train débouche du tunnel dans la vallée »), chaque Position représentera par une alternative d'occupant la pronominalisation de ses occupants, une condition d'exclusion mutuelle exprimant l'impossibilité de la double pronominalisation (« y en »).

3.2.4 Réécriture de Syntagmes

- 37 Notons enfin que les occupants d'une Position peuvent eux-mêmes être décrits par une liste de Positions. On obtient une arborescence syntaxique, l'unité décrite se trouvant à n'importe quel niveau de cette arborescence. Ainsi la Description d'une unité ne se limite-t-elle pas forcément à sa projection maximale, mais peut « déborder » en un contexte syntaxique d'apparition : un nom pourra être décrit avec son verbe support, un adjectif au sein d'un groupe nominal. C'est ce que l'on dénomme la vision syntaxiste du modèle, vision que nous avons choisie ici.
- 38 Dans ces cas d'arborescence syntaxique, fonctions et rôles thématiques sont définis par rapport aux occupants d'une Position de même niveau portant la fonction tête.

4. Vers une Vue Relationnelle de la Position

4.1 Tables du LADL et Paradigme de Substitution

- 39 Pour la mise au format GENELEX des informations syntaxiques des tables du LADL, il était primordial de fixer dès le début les paramètres définitoires de la Position. En effet, la notion de Position définie en extension par une distribution prédéfinie d'occupants ne nous a pas semblé opératoire : dans le cas des tables, les seules distributions définies en tant que telles sont celles décrites par chacun des cartouches². Elles ne concernent généralement qu'un occupant³, et auraient imposé comme seule structuration celle imposée par les cartouches, c'est à dire une structuration ne reposant pas sur des principes formellement explicités dans le format d'expression et d'une granularité trop proche de la réalisation lexicale de surface pour obtenir une base permettant facilement des classifications par requêtes.
- 40 Il s'agit donc avant tout de définir précisément ce que l'on entend par « paradigme de substitution » et ce, sans utiliser de définition en extension, afin de revenir au mode opératoire permettant d'obtenir la distribution d'une Position. On définira donc un

paradigme de substitution comme classe d'équivalence d'une relation S et on fixera les paramètres de cette relation.

- 41 Pour une unité lexicale fixée u , considérons un syntagme cooccurrent avec u ; ce syntagme entretient avec u une « relation grammaticale » notée R . On dira que deux syntagmes sont en relation de substitution S si et seulement s'ils entrent dans la même relation R avec l'unité lexicale. L'ensemble des paradigmes de substitution est clairement isomorphe à l'ensemble des relations grammaticales qu'une unité lexicale précise peut sous tendre. Ayant choisi de définir une relation grammaticale par un couple fonction syntaxique, liste de rôles thématiques, il en découle que chaque distribution est définie par un tel couple.
- 42 Le passage des distributions définies pour chaque unité lexicale à la Position est immédiat : à chaque triplet correspond par définition une Position. La Position apparaît alors comme une classe d'équivalence sur l'ensemble des distributions à travers toutes les unités lexicales : deux distributions sont équivalentes si et seulement si leur extension, leur fonction et leurs rôles thématiques sont identiques. La Position est alors une distribution représentant la classe d'équivalence (e. g. la première distribution construite). Il nous semble ainsi particulièrement pertinent que la Position porte les informations qui définissent sa distribution dans le contexte de l'unité décrite (i. e. fonction et rôles thématiques).
- 43 Ce résultat immédiat est essentiel : nous n'aurions pu passer directement à la notion de Position qui nécessitait une distribution définitoire d'emblée, c'est à dire prise dans l'ensemble de toutes les distributions à travers toutes les unités, ensemble dont nous ne disposons pas, vu la nature des cartouches et le travail incrémental de la traduction des tables.
- 44 En conclusion, la vue relationnelle de la notion de Position permet la traduction au format GENELEX par examen successif des cartouches. L'expertise de traduction d'une table du LADL dans le format GENELEX consiste plus particulièrement à traduire les propriétés décrites par les cartouches en terme d'invariance par rapport à une relation grammaticale typée par une fonction et une liste de rôles thématiques.
- 45 Ce principe de structuration fort a permis de constituer incrémentalement une base de données syntaxiques dans un monde de programmation objet disposant de moyens de requêtes classifiantes.

4.2 Exemple du verbe « dériver »

- 46 Les informations de distributions sont parcellées entre cartouches. Nous montrons ici, sur l'exemple du verbe « dériver », comment le principe de structuration adopté factorise dans un moule formel les informations syntaxiques des tables. Les exemples choisis sont volontairement simples, voire simplifiés ; il ne s'agit pourtant pas d'en déduire qu'il y a forcément une correspondance un-un entre occupant de Position et intitulé de cartouche : le travail d'expertise de traduction d'une table au format GENELEX explicite souvent des dépendances ou des déductions complexes entre plusieurs cartouches (optionalité des Positions, dépendances des actualisations en surface).

- 47 Pour chaque actualisation syntaxique, nous indiquons entre crochets les compléments non concernés par le cartouche pertinent et indiquons l'intitulé du cartouche tel qu'on peut le lire dans la version papier des tables.
- 48 Le verbe « dériver de » se retrouve dans trois tables : table 8, table 16 et table 35L. Nous nous restreignons dans cet article aux tables de « Méthodes en Syntaxe » (Gross, 1975) et ne détaillerons donc pas la table 35L⁴.
- 49 Dans la suite, on référera à un cartouche (propriété) d'une table par son rang dans la liste des cartouches de cette table et on notera directement la valeur + ou - prise par l'entrée considérée pour cette propriété.
- 50 La table 8 couvre (entre autres) les comportements syntaxiques suivants. Après chaque exemple est indiquée la propriété de la table qu'il illustre.

Ce théorème dérive [de cet ensemble d'axiomes]

2 + : Sujet / « Nr »

Le fait que ce théorème est faux dérive [de cet ensemble d'axiomes]

3 + : Sujet / « le fait Qu P »

Que ce théorème est faux dérive [de cet ensemble d'axiomes]

3 + : Sujet / « le fait Qu P »

[Ce théorème] dérive de ce que j'ai démontré auparavant

9 + : Compléments Indirects / « que P »

[Ce principe] dérive de Chomsky

22 + : Compléments Indirects / « N hum »

[Ce théorème] dérive de cet ensemble d'axiomes

23 + : Compléments Indirects / « N -hum »

[Ce théorème] dérive du fait que cet autre théorème est faux

24 + : Compléments Indirects / « le fait Qu P »

[Ce théorème] en dérive

25 + : Compléments Indirects / Pronom / « ppv »

- 51 Ce qui se traduira dans le format GENELEX par 2 Positions P0 et P1 :

- 52 Optionalité : P0 (P1)

- 53 P0 =

[Sujet, Agent]

{SN, Pro[qqch], SN[le fait que P], PCompl}

- 54 P1 =

[Objet indirect, Source]

{Pro[en], SP[P3 P4]}

où P3 = {Prep[de]}

et P4 = {SN, Pro[qqun], Pro[qqch], SN[le fait que P], PCompl}

- 55 La table 16 note les comportements suivants :

Pierre dérive [de cet ensemble d'axiomes que ce théorème est faux]

1 + : Sujet / « N hum »

[Pierre] dérive que ce théorème est faux ...

9 + : Compléments Directs ou Indirects / « que P »

[Pierre] dérive le fait que ce théorème est faux ...

9 + : Compléments Directs ou Indirects / « le fait Qu P »

[Pierre] dérive ce théorème ...

20 + : Compléments Directs ou Indirects / « N -hum »

[Pierre] dérive [ce théorème] de cet ensemble d'axiomes

36 + : Compléments Indirects / « N -hum »

[Pierre] dérive [ce théorème] du fait que cet autre est faux

37 + : Compléments Indirects / « le fait Qu P »

56 Ce qui se traduira en GENELEX par 3 Positions P5 P6 et P7 :

57 Optionalité : P5 (P6) (P7)

58 Condition d'Optionalité : si P7 est actualisée alors P6 l'est également

59 P5 =

[Sujet, Agent]
{SN, Pro[qqun]}

60 P6 =

[Objet direct, Thème]
{SN, Pro[qqch], SN[le fait que P], PCompl}

61 P7 =

[Objet indirect, Source]
{Pro[en], SP[P8 P9]}
où P8 = {Prep[de]}
et P9 = {SN, Pro[qqch], PCompl}

62 L'utilisation d'un format unifiant et d'une base de données permet d'opérer facilement certains regroupements : ainsi on pourra se demander, au vu de la proximité des Positions P1 et P7 qui ne diffèrent que de l'occupant « Pro[qqun] », s'il s'agit d'une incomplétude dans la description de la table 16 et, dans ce cas, s'il n'y a pas lieu de relier les deux comportements par une relation de transformation de type neutralité (P0 et P6 couvrant la même distribution). De tels recoupements n'auraient pas pu être faits automatiquement sur la seule base des informations disponibles sous support informatique, i.e. des + et des -.

5. Des Tables au Format GENELEX

5.1 Les tables

63 Les tables du LADL constituent la plus grosse base de connaissance syntaxique du français. Chaque table définit minimalement une classe de comportements syntaxiques par des critères d'entrée dans cette table et par un ensemble de propriétés distinctives pertinentes sur cette classe. La nature et le nombre des propriétés examinées dans ces tables varient donc pour chacune. Le format informatique dont nous disposons pour un travail automatisé est très simple. Il s'agit d'un fichier texte comportant autant de lignes que d'entrées décrites, chaque ligne possédant deux zones d'information :

- champs lexicaux : graphie de l'entrée, préposition régie, particules préverbaux (pronoms figés, négation), etc.
- propriétés syntaxiques : chaque cartouche a pour valeur + ou -. Leur nombre varie suivant la table, ainsi que leur libellé et leur interprétation. La sémantique des + et - est elle-même variable : + peut exprimer, selon les cas, que la propriété exprimée est obligatoire, possible ou dépendante de la valeur d'un autre cartouche.

64 Le libellé des cartouches utilisés ou implicites, leur interprétation et autres informations connexes, telles que les critères d'entrée dans la table, les exemples utilisés pour fonder certaines attestations, font partie de documents papier d'accompagnement ou de la tradition orale.

65 On conçoit aisément l'impact d'une traduction de l'ensemble de ces données sous un format standard et unifiant tel que GENELEX. La base de données ainsi constituée est

non seulement lisible humainement mais constitue aussi une source d'informations linguistiques réellement exploitable enrichie de tout l'environnement de travail d'une BD (requêtes, extractions, navigations, recoupements, mises à jour, etc).

5.2 Le logiciel AZOTE

- 66 Nous avons choisi de clairement distinguer dans AZOTE l'expression de la connaissance et les traitements informatiques appliqués, à savoir l'expertise de transfert et son exploitation automatique (Alcouffe et al., 1993).
- 67 On se donne dans ce but un cadre formel de représentation dans lequel on s'attache à traduire de façon homogène et unifiée les propriétés contenues dans les tables. La règle d'expertise ainsi écrite pour une table donnée contient la description d'une entrée potentielle maximale de la table. L'ensemble de cette expertise fournit à lui seul une source de connaissance unique sur l'interprétation des tables.
- 68 Le logiciel développé pour exploiter l'expertise formalisée permet l'analyse de la règle et le chargement immédiat des données. Leur fusion incrémentale est opérée au fur et à mesure de l'incorporation de nouvelles tables, les objets de contenus identiques étant partagés.

5.3 Règle d'Expertise

- 69 Une règle d'expertise est un objet arborescent (dit Objet Conditionné) associé à un objet du modèle GENELEX (dit Objet Associé) et qui donne pour chaque attribut la valeur qu'il reçoit lorsqu'une Condition qui lui est associée est vérifiée. La valeur de l'attribut est soit une chaîne de caractères, soit un autre Objet Conditionné, lui-même associé à un objet du modèle. Ainsi la structure arborescente d'un Objet Conditionné suit la structure arborescente de son Objet Associé qui le contraint sémantiquement.
- 70 Un Objet Conditionné représente l'ensemble des instances de la classe décrite par son Objet Associé, telles qu'elles existent dans l'ensemble combinatoire des Conditions associées. Pour chaque contexte de conditions, un objet peut être créé qui prend pour ses attributs les valeurs correspondant à l'évaluation des conditions dans ce contexte.
- 71 On crée pour chaque table un Objet Conditionné associé à l'Objet Usyn. Les Conditions sont des vérifications de + et - dans des cartouches. L'Objet Conditionné ainsi créé est l'expression en un seul objet de toutes les Usyn potentielles de la table. Pour chacune des lignes de la table, l'instanciation des conditions permet la construction d'une Usyn différente. Ainsi un seul Objet Conditionné encode toute l'expertise de traduction d'une table en objets GENELEX Usyn, attribut par attribut.

6. Exemple d'expertise de traduction sur la table 11

- 72 Nous présentons ici le travail mené pour la traduction de la table 11 :
 « Les verbes de cette table ont un complément direct substantival et un complément indirect en à (complétive ou infinitive). Les complétives se classent en deux catégories principales, selon qu'elles se réduisent à une infinitive avec sujet N0 ou avec sujet N1. » (Gross, 1975 :197)
- 73 Parmi les réalisations lexicales décrites par cette table, on trouvera :

Luc dresse le chien à sauter sur les passants
Luc consacre du temps à ce que la maison soit en ordre
Luc consacre du temps à ranger la maison
Faire des maths a conduit Luc à être enseignant
Faire des maths a conféré beaucoup d'importance à Luc
La situation limite l'activité de Paul à amuser Marie
Luc s'aide de ce que le tronc flotte à sortir de l'eau
Luc est stimulé à réussir de ce qu'il voit les autres perdre
Luc est avantagé de ce qu'il y a du vent
Les soldats sont sacrifiés par leur supérieur à défendre la colline
Il a été affecté cette somme par l'administration à acheter des stylos

74 Dans l'expertise d'une table, on peut distinguer 5 étapes successives :

- Première détermination des Positions de la description de base
- Examen des propriétés - cartouches - de la Table ; chaque cartouche fournira en général une information sur un des aspects suivants :
 - optionalité d'une Position
 - occupant possible d'une Position
 - descriptions transformées de la description de base
 - raffinement des Positions identifiées en première approximation : cela peut conduire à éclater ce qui était initialement pressenti comme une unique Position en deux Positions distinctes, voire à représenter la table par deux Unités Syntaxiques distinctes
- Détermination d'exemples illustrant les comportements identifiés dans la table
- Ecriture de la règle proprement dite : formalisation de l'expertise de transfert
- Chargement de la Table et vérification des données obtenues

75 Dans ce qui suit, après avoir rappelé la liste des cartouches de la table 11, nous examinons pour chacun d'entre eux (et pour leurs combinaisons éventuelles) la manière dont il se traduit en une propriété GENELEX. Nous donnons enfin quelques exemples des comportements syntaxiques GENELEX ainsi obtenus.

6.1 Cartouches de la Table 11

1				Nhum
2				Nnr
3	Sujet			le fait Qu P
4				V1 W
5				V2 W
6				N0 V
7				N0 V N1
8		Noms		Nhum
9				N-hum
10	Complément			N0 est V-ant pour N1

11	direct			N1 se V de ce Qu P
12				N1 est Vpp de ce Qu P
13				[passif par]
14				[passif de]
15				N0 V à N2
16				Que P
17				Que Psubj
18				[pc z.]
19				V0 W
20				V1 W
21		Complétives		V1c W
22				Tc = « passé »
23				Tc = « futur »
24				ce (ci + là)
25			Pronoms	ppv
26				là
27				Nhum
28				ppv
29		Noms		N-hum
30				le fait Qu P
31				ppv
32				[extrap] [passif]
33				[passif par]
34				[passif de]

6.2 Description de Base et Positions

- 76 Les entrées de la table 11 peuvent comporter deux compléments distincts pouvant se réaliser simultanément, on prendra donc en première évaluation pour la description de base une structure à trois Positions : une Position Sujet et deux Positions Complément.
- 77 Cartouches 1 à 5 : Position Sujet – Agent
- 78 Cartouches 7 à 12 : Position Objet Direct – Patient
- 79 Cartouches 16 à 31 : Position Objet Direct / Indirect
- 80 Cartouche 18 (pc. z.) : cette propriété indique la possibilité d'effacement de l'introducteur (« à ce ») de la Complétive Indirecte (« à ce que P »), i.e. la possibilité d'avoir une Complétive Directe (« que P ») en deuxième Position objet. Selon que l'on souhaite privilégier la distribution ou la fonction dans la définition de la Position, on pourra soit éclater la deuxième Position objet en deux Positions de fonctions respectivement objet direct et objet indirect, soit conserver une seule Position regroupant les complétives directe et indirecte sous une même fonction à déterminer. Dans le cas de la table 11, ce problème ne se pose en pratique pas dans la mesure où toutes les entrées de cette table prennent la valeur – sur ce cartouche.
- 81 On notera dans la suite les trois Positions de notre description de base PSujet, PDirect et PIndirect. La virgule indique la conjonction de cartouches.

6.3 Optionalité

- 82 On déduit l'optionalité des Positions des cartouches 6, 7 et 15.

6+ ou 15+ :	PDirect optionnelle	(i)
6+ ou 7+ :	PIndirect optionnelle	(ii)
6+, 15- :	si PIndirect actualisée alors PDirect actualisé	(iii)
6+, 7- :	si PDirect actualisée alors PIndirect actualisée	(iv)
6-, 7+, 15+ :	si PIndirect effacée alors PDirect actualisée	(v)

- 83 L'entrée *aider* illustre les points (i), (ii) et (v) :

Pierre aide son fils à tondre la pelouse
Pierre aide à tondre la pelouse
Pierre aide son fils
** Pierre aide*

- 84 *Gaspiller* : (i), (ii) et (iv)

Pierre gaspille son argent à acheter des bonbons
** Pierre gaspille à acheter des bonbons*

- 85 *Inciter* : (i)

Pierre incite son fils à utiliser des préservatifs
** Pierre incite son fils*

6.4 Examen des occupants de Position

86 Dans cette section, nous indiquons en regard de chaque cartouche les occupants de Positions qui lui correspondent dans la traduction au format GENELEX ainsi que les éventuelles transformations opérant entre ces occupants.

87 On utilise les notations suivantes pour les occupants GENELEX :

- *SN* : Syntagme Nominal
- *Pro[X]* : Pronom de graphie X
- *Prep[X]* : Préposition de graphie X
- *PInf* : Infinitive
- *PCompl*, *PInd*, *PSubj* : Complétives avec ou sans indication du mode
- *SN[le fait que P]* : *le fait que PInd/PSubj*
- *SP* : syntagme introduit par une préposition
- *SP[P0 P1]* : réécriture d'un syntagme par une liste de Positions
- Les coréférences entre syntagmes sont indiquées par la présence sur les syntagmes concernés de traits de valeur *I*, *J*, *K* ou *L*. Pour les complétives et infinitives, l'information de coréférence porte sur le sujet. On aura ainsi *SN[K]*, *PInf[J]* ...
- Une transformation entre occupants est notée nom de transformation : *occupant origine => occupant résultat*

1 + :	SN
	Pro[qqun]
	Pronominalisation : SN => Pro[qqun]
2 + :	SN
	Pro[ceci]
	Pro[cela]
	Pro[qqch]
	Pronominalisation : SN => Pro[qqch]
3 + :	SN[le fait que P]
4 + :	PInf[J] – coréférent au SN[J] en Position PDirect
5 + :	PInf[K] Lire Marx a converti Marie au communisme
19 + :	SN[I] Jouer aux courses rapporte de l'argent à Luc

4 + ou 20 + :	SN[J]
8 + :	SN

	Pro[qqun]
	Pronominalisation : SN => Pro[qqun]
9 + :	SN
	Pro[qqch]
	Pronominalisation : SN => Pro[qqch]
21 + :	SN de SN[L] – syntagme nominal récrit dont le complément déterminatif porte une coréférence

6.4.3. PIndirect

- 88 Les cartouches 16 à 26 concernent les complétives et infinitives, les cartouches 27 à 31 les syntagmes nominaux.
- 89 16 + ou 17 + ou 19 + ou 20 + ou 21 + ou 24 + :

	SP[P0[Prep[à]], P10] avec pour occupants de P10 :	
16+ :	PInd	
17+ :	PSubj	
19+ :	PInf[I]	Pierre consacre ses loisirs à voyager
20+ :	PInf[J]	Ses amis le portent à faire des bêtises
21+ :	PInf[L]	Pierre étend les attributions de son employé à traiter avec la Chine
16+, 19+ :	PInd[I]	et Réduction : PInd[I] => PInf[I]
16+, 20+ :	PInd[J]	et Réduction : PInd[J] => PInf[J]
16+, 21+ :	PInd[L]	et Réduction : PInd[L] => PInf[L]
17+, ...	Traitement similaire pour les PSubj => PInf	
24+ :	Pro[ceci]/Pro[cela]	
16+, 24+ :	Pronominalisation : PInd => Pro[ceci]/Pro[cela]	
17+, 24+ :	Pronominalisation : PSubj => Pro[ceci]/Pro[cela]	
25+ :	Pro[y]	
25+ et (16 + ou 17 + ou 19 + ou 20 + ou 21 + ou 24 +) :		

	Pronominalisation : SP[P0[Prep[à]], P10] => Pro[y]
26+ :	Pro[là]
26+ et (16 + ou 17 + ou 19 + ou 20 + ou 21 + ou 24 +) :	
	Pronominalisation : SP[P0[Prep[à]], P10] => Pro[là]
27+ :	SP[P0[Prep[à]], P11] avec pour occupants de P11
	SN
	Pro[qqun]
	Pronominalisation : SN => Pro[qqun]
5+ :	SN[K]
28+ :	Pro[lui]
27+, 28+ :	Pronominalisation : SP[P0[Prep[à]], P11] => Pro[lui]
29+, 31+ :	Traitement similaire :
	SP[P0[Prep[à]], P12[SN =>Pro[qqch]]] => Pro[y]

90 Dans la description qui précède :

- on ne rend pas compte des indications de temps sémantique (22 et 23),
- des transformations pouvant s'appliquer à différents syntagmes dont certains sont plus spécifiques que d'autres ne sont enregistrées explicitement que sur les plus génériques ; ainsi, on note PInd => Pro[ceci] mais pas PInd[I] => Pro[ceci].

6.5 Descriptions Transformées

- 91 La table 11 donne lieu à 4 descriptions transformées que nous ne détaillerons pas.
- 92 Pour les descriptions transformées, on doit lors de l'expertise étudier pour chacun des occupants possibles des Positions de la description de base s'il est conservé ou non dans la transformée. Par exemple, les infinitives en Position sujet de la description de base sont absentes de la Position complément d'agent du passif.

6.6 Exemples

- 93 Ci-dessous figure une visualisation dite « synthétique » de quelques données de la base. Cette apparence générée automatiquement par sélection d'un occupant pour chaque Position donne des informations succinctes sur l'entrée.
- 94 aguerrir

P[inf] aguerrir qqun/qqch à ce que P[subj]
 qqun/qqch être aguerrri par qqun/qqch à ce que P[subj]
 il est aguerrri qqun/qqch par qqun/qqch à ce que P[subj]

95 aider

P[inf] aider qqun/qqch à ce que P[subj]
 qqun/qqch s'aider à ce que P[subj] de ce que P[subj]
 qqun/qqch être aidé à ce que P[subj] de ce que P[subj]
 qqun/qqch être aidé par qqun/qqch à ce que P[subj]
 il être aidé qqun/qqch par qqun/qqch à ce que P[subj]

7. Volumétrie

- 96 A ce jour, 37 tables ont été traduites, produisant environ 5.000 comportements syntaxiques pour 8.000 graphies.
- 97 Nous présentons quelques chiffres correspondant au chargement dans la base des tables de « Méthodes en Syntaxe » (17 tables). Nous avons dénombré les objets construits, mais également les partitions qu'induisaient les objets structurants Position et Construction.

7.1 Objets construits dans la base

Unités Morphologiques	2348
Unités Syntaxiques	4422
Descriptions	3163
Constructions	2975
Selfs	42
Positions	1797
Occupants	557

7.2 Structuration par type de Position

- 98 On appelle type de Position le couple fonction, liste de rôles thématiques. Ce type définit, dans le contexte d'une unité, la distribution de la Position.
- 99 22 types de Positions sont recensés dans la base. Nous indiquons le nombre de Positions de chaque type, et pour chacun, nous dénombrons :
- le nombre de Constructions Syntaxiques utilisant ce type de Position
 - le nombre d'Unités Syntaxiques utilisant ce type de Position dans sa Construction de Base
 - le nombre d'Unités Syntaxiques utilisant ce type de Position dans une de ses Constructions Transformées

Type de Position	Positions	Constructions	Usyns / B	Usyns / T
Sujet /	140	2093	3087	2686
Sujet / Agent	29	380	1335	315
Sujet / Thème	87	259	0	1400
Sujet / Patient	13	267	0	1354
Objet Direct (OD) /	166	392	955	459
OD / Locatif	2	3	17	0
OD/Thème	307	708	1398	742
OD/Patient	14	446	982	361
Objet Indirect (OI) /	476	1484	1209	1397
OI / Agent	29	1216	0	4702
OI / Locatif	23	68	204	77
OI / Thème	1	183	499	758
OI / Patient	11	134	105	68
OI/Destinataire	16	532	1184	2044
OI / Cause	1	54	53	53
OI / But	44	28	16	12
Attribut Sujet/	12	67	0	555
Attribut Objet/	12	81	575	328
/ Locatif	4	22	39	270
/ But	2	84	228	322
/ Source	1	2	32	5
/ Theme	59	119	78	57

7.3 Structuration par type de construction

- 100 L'ensemble des Constructions Syntaxiques peut se partitionner en classes d'équivalences, par exemple :
- les Constructions Syntaxiques de même Optionalité
 - les Constructions Syntaxiques de même Optionalité et ayant une même liste de types de Position
- 101 Pour chaque classe d'équivalence définie, on peut définir l'ensemble des Unités Syntaxiques utilisant les Constructions de cette classe (en Base ou en Transformée).
- 102 A travers ces deux partitions, nous avons un exemple de granularité croissante : 32 classes dans le premier cas, 142 dans le second.

Optionalité	Constructions	Usyns / B	Usyns / T
P0	6	7	0
P0 P1	696	575	79
P0 (P1)	520	1267	1511
P0 P1 P2	318	651	243
P0 (P1) P2	211	129	321
P0 P1 (P2)	386	360	602
P0 (P1) (P2)	384	388	904
P0 (P1) (P2) si P1 alors P2	22	21	7
P0 (P1) (P2) si P2 alors P1	78	243	0
P0 (P1) (P2) si P1 alors P2 si P2 alors P1	6	10	0
P0 (P1) (P2) si !P1 alors P2	49	69	0
P0 P1 P2 P3	18	42	4
P0 (P1) P2 P3	2	3	0
P0 P1 (P2) P3	58	1	102
P0 P1 P2 (P3)	100	582	53
P0 P1 (P2) (P3)	253	51	1088
P0 (P1) P2 (P3)	12	1	15
P0 (P1) (P2) P3	33	0	35

P0 (P1) (P2) (P3)	112	4	229
6 Optionalités à 4 Positions avec Conditions	19	18	4
7 Optionalités à 5 Positions	79	0	615

- 103 Pour la deuxième partition en 142 classes, nous indiquerons simplement le nombre de classes en fonction de leur nombre d'éléments.

	< 10 élts	10 à 50	50 à 100	100 à 200	200 à 500	> 500 élts
Nb de Classes de Constructions	76	48	14	5	0	0
Nb d'Ensembles d'Usyn / B	23	24	9	8	5	1
Nb d'Ensembles d'Usyn / T	33	26	10	6	10	1

- 104 L'ensemble d'Usyn B le plus grand comprend 529 unités et correspond à la classe d'équivalence des Constructions (classe de seulement 5 éléments) :
- 105 P0 (P1) : P[Sujet Agent] P[Objet Direct Patient]
- 106 La classe d'équivalence de Constructions la plus grande contient 139 Constructions utilisées seulement en transformées et est du type :
- 107 P0 (P1) (P2) : P[Sujet Patient] P[OI Agent] P[OI]
- 108 L'examen des ensembles à petit nombre d'éléments permet éventuellement de remettre en cause certaines des acceptations enregistrées dans les tables ou de détecter certaines erreurs. Dans le cas de l'Optionalité P0 (P1) (P2) où P1 et P2 ne peuvent être actualisées l'une sans l'autre, la classe regroupe 10 Usyns dont l'Usyn « urger » : « Pierre urge Marie de terminer » où l'effacement total n'est en fait pas possible et correspondrait à une autre construction absente de la table.
- 109 Des granularités plus fines de classes d'équivalence peuvent être obtenues pour descendre jusqu'à la granularité de la Construction (soit 2975 classes). Au delà des classes d'équivalences, des ensembles de comportements sont obtenus par simples requêtes : ensemble des constructions ayant tel et tel occupant dans la Position de tel type et acceptant telle transformation, etc. On permet ainsi des regroupements qui auraient été impossibles à faire automatiquement dans le format de départ des tables.

8. Conclusions

- 110 Nous espérons avoir montré à travers quelques exemples immédiats la richesse potentielle de la structuration opérée dans la base de données syntaxiques. Cette structuration repose sur le typage d'une relation de substitution entre occupants syntaxiques et paramètre une notion clef du modèle GENELEX, la notion de Position. Au delà de la validation du modèle, le principe structurant adopté permet la définition de classes d'équivalences et d'ensembles de comportement de granularité variable.

- 111 Outre les données obtenues, l'approche logicielle AZOTE a permis d'obtenir également une formalisation de la connaissance linguistique exprimée dans les cartouches des tables du LADL (indépendamment des valeurs prises par les entrées).
- 112 Les données syntaxiques du LADL, particulièrement riches, tant par leur large couverture de la langue française que par leur finesse de description, étaient jusqu'à présent restées difficiles d'accès, en raison d'un formalisme de représentation opaque. Notre travail rend directement lisibles les informations syntaxiques contenues dans les tables et unifie leur représentation par l'usage d'un cadre formel structurant : le modèle GENELEX. La base de données lexicales obtenue constitue un fonds de premier intérêt pour toute la communauté linguistique, alliant, à la richesse des données sur les comportements verbaux du français, les capacités de manipulation et de consultation d'une base de données.
-

BIBLIOGRAPHIE

ALCOUFFE Ph., REVELLIN-FALCOZ B., ZAYSSER L. (1993), *Azote : Des tables du LADL au format GENELEX*, ILN 93 – Nantes.

FILLMORE C. (1968), *A Case for Case*, E. Bach et R. T. Harms, 1-88.

GENELEX Consortium (1993), *Rapport sur la Couche Syntaxique V4. 0*.

GROSS M. (1975), *Méthodes en Syntaxe – Régime des Constructions Complétives*, Hermann

MILNER J. C. (1989), *Introduction à une Science du Langage*, Seuil.

TESNIÈRE L. (1959), *Éléments de syntaxe structurale*, Klincksiek.

NOTES

1. À moins de forger un trait spécifique e.g. « inversion objet1-objet2 interdite » porté par la Construction.
2. Un « cartouche » note une propriété syntaxique d'une table, cf section 5.1.
3. On notera toutefois le cas du cartouche « Nr » : « Le test employé consiste à placer dans la position à étudier les groupes nominaux *Paul* ('humain'), *ce gâteau* ('concret', 'énumérable' ou non), *la sincérité* ('abstrait'). *Qu P*, et *V-inf Ω* » (Gross, 1975:50)
4. Cette table couvre l'emploi locatif « le bateau dérive d'une rive à l'autre ».

RÉSUMÉS

Les données syntaxiques du LADL, riches tant par leur large couverture de la langue française que par leur finesse de description, étaient jusqu'à présent restées difficiles d'accès, en raison d'un formalisme de représentation opaque. Notre travail rend directement lisibles les informations syntaxiques contenues dans les tables de verbes et unifie leur représentation par l'usage d'un cadre formel structurant : le modèle GENELEX. La base de données obtenue constitue un fonds de premier intérêt pour toute la communauté linguistique, alliant, à la richesse des données sur les comportements verbaux du français, les capacités de manipulation et de consultation d'une base de données.

Syntactical data from LADL detail with an extremely fine granularity French verb syntactic behavior. Up until now, these data have been difficult to process due to a representation framework unsuited to computer processing. Our work renders these syntactic data readable and unifies their format by the use of a highly structuring framework : the GENELEX model. The resulting database is highly interesting for linguistics : it combines the richness of data on French verb behavior with database abilities.

AUTEURS

PHILIPPE ALCOUFFE

SEMA GROUP

alcouffe@metradt.metra.fr

BRUNO REVELLIN-FALCOZ

SEMA GROUP