

Au Professeur Robert Martin,
en respectueux hommage.

Grammaire du dictionnaire / dictionnaire de la grammaire : pour une nouvelle complémentarité conceptuelle

Gérard REB
Université Marc Bloch
EA 1339 LILPA-Scolia, Linguistique, Langues et Parole
Strasbourg
Neoidea@noos.fr

Idée directrice : la formalisation de la grammaire et la formalisation du dictionnaire permettent la formalisation du sens.

Les langues sont des « systèmes de formes notant des sens » (I. Tamba, 1994 : 4). En tant qu'objet linguistique, le sens est descriptible à travers les rapports que les signifiants et les signifiés verbaux entretiennent. Le traitement automatique des langues (TAL) implique alors de préciser le statut du dictionnaire et celui de la grammaire dans la définition de ces rapports. Il importe donc de montrer la complémentarité de la formalisation de la grammaire et celle du dictionnaire dans le traitement automatique du sens. Et pour ce faire, montrer comment grammaire et dictionnaire coopèrent dans la mise en forme métalinguistique des données signifiantes à travers la définition des unités de sens. Cette dernière passe nécessairement par une interrogation sur les rapports entre signifiants et signifiés verbaux, puisque la représentation informatique d'un système linguistique - sa traduction dans un système artificiel - « repose sur un jeu d'échange ou de *troc sémantique* permis par la distinction des formes signifiantes » (I. Tamba, 1994 : 47). Or, ce principe de « la distinction des formes signifiantes » associées au signifié verbal fonde, dans une théorie matérialiste du sens, l'équivalence du concept de traduction comme opération de langues naturelles à langues naturelles (traduction interlangue, traduction intralangue) et celui de traduction comme opération de langue naturelle à langage artificiel. Dans cette double acception de la traduction, se réalise en effet la même opération consistant à délier le lien du signifiant et du signifié verbal, dans une procédure d'analyse, et à reconstruire dans une procédure de synthèse le lien du signifiant et du signifié dans l'écriture du système de représentation du langage cible, langue naturelle ou langage artificiel. Conceptuellement, l'opération d'analyse et de synthèse réalisée dans la traduction de systèmes de représentations que sont les systèmes de langues valide épistémologiquement l'hypothèse de la

traduction d'une langue vers un langage artificiel de haut niveau, à savoir la compilation¹. Mais l'opération de traduction du signifié d'une langue dans le signifié d'un langage artificiel de haut niveau réalisée dans la compilation s'accompagne nécessairement de la modification, par changement de substance, du codage du signifiant analogique en signifiant numérique, du continu au discret, permise par le caractère arbitraire de la relation du signifiant au signifié dans la théorie saussurienne du signe linguistique.

De même qu'une théorie de la traduction de systèmes de langues implique une analyse sur la relation entre catégories grammaticales et catégories lexicales de la langue source et de la langue cible, l'objectif de ce travail est de définir avec précision « la part automatisable » (R. Martin, 2001 : 13) de traduction dans le traitement automatique du sens, par « la simulation du fonctionnement linguistique » mis en œuvre par la grammaire et le dictionnaire dans la représentation du sens linguistique. Le traitement automatique présuppose alors un modèle construit de la relation d'un type de grammaire et d'un type de dictionnaire reposant sur la définition des unités de sens objectivement contrôlables et reproductibles.

Quelle est la part de la grammaire, quelle est celle du dictionnaire, quelle est la part de leur association dans la représentation de ces unités de sens linguistique ?

I. Présentation du problème :

1. Définition des unités de sens

Le traitement automatique des unités de sens impose la condition irréductible de l'objectivité, ce que R. Martin (2001 : 19) appelle le « principe d'objectivité sémantique ». Or, comme « le sens ne se définit pas en soi » (ibid.), mais est accessible dans « la relation sémantique que les phrases entretiennent », « l'unité à partir de laquelle la sémantique s'organise est la *proposition*. Seule la *proposition* et la *phrase* qui la comporte ont une valeur de vérité ; le jugement sur la relation de sens porte sur des propositions et sur des phrases. La proposition et la phrase se placent ainsi au centre de la construction théorique : ce sont les unités fondatrices de tout calcul du sens. Le sens d'une phrase est l'ensemble des conséquences qu'elle entraîne, c'est-à-dire l'ensemble des phrases vraies quand elle-même est vraie. La fonction centrale est celle de l'inférence. Un analyseur sémantique débouche forcément sur un *moteur d'inférence*, capable de tirer des énoncés les conséquences que ces énoncés comportent. »

Mais en faisant reposer la notion de 'relation de sens' sur le « jugement que le locuteur compétent d'une langue est en mesure de porter sur la relation de vérité que les phrases entretiennent et sur les conséquences sémantiques que les phrases déterminent » (p. 18), R. Martin réintroduit l'intuition dans la définition de l'unité de sens qu'est la phrase. Dès lors, le principe d'objectivité sémantique qui repose sur la définition de l'unité *phrase* comme entité sémantique doit être complété par une procédure de décision qui définit la phrase comme l'unité syntaxique d'une expression bien formée récursivement d'unités plus petites, selon l'opération de compositionnalité. Comme cette dernière se fonde sur des procédures régulières descriptibles par la grammaire d'une langue, on est en droit de formuler ce que nous appellerons « le principe d'objectivité syntaxique ». L'analyseur syntaxique débouche alors sur un *moteur d'inférence*

¹ On retrouve le principe selon lequel : « l'aspect cognitif du langage, non seulement admet mais requiert, l'interprétation au moyen d'autres codes, c'est-à-dire la traduction » (Jakobson, 1963 : 84).

syntaxique, capable de générer les expressions bien formées de la phrase dans un système formel décrit par des règles de production. L'association de l'analyseur syntaxique et de l'analyseur sémantique constitue alors le modèle opératoire du TAL. L'automatisation s'organise ainsi à l'intérieur de ce système qui associe morphologie et syntaxe d'expressions bien formées et leur sémantique dans la complémentarité de la grammaire et du dictionnaire. Le modèle informatique de la compilation qui associe analyseur syntaxique et graphe de dépendances est ainsi à même de modéliser la représentation d'une part, à définir, du sens linguistique.

L'application qui valide le mieux la part de signifié qu'un automate est susceptible de représenter est celle de l'interface *question-réponse*. Comme le souligne R. Martin (2001 : 138), « l'automate comprend une phrase p s'il est capable de donner de p une représentation dans un système déterminé : la compréhension consiste alors dans la traduction de p , à partir du langage où p est exprimé, dans un langage de représentation : ainsi, une question posée en langue naturelle à une Base de connaissances suppose qu'elle soit traduite - par analyse - en une représentation susceptible d'y être traitée, de manière à apporter, sous le forme de représentations nouvelles, d'éventuelles réponses, à leur tour traductibles -par génération - dans le langage initial ».

2. La notion de 'compréhension automatique' : première acception

Cette définition de la notion de « compréhension automatique » via la traduction d'une phrase exprimée en langue naturelle dans un langage de représentation artificielle implique une définition opératoire de l'unité 'phrase', de l'unité 'mot' et d'une procédure susceptible de représenter la relation entre ces deux types d'unités. La première unité étant de l'ordre de la grammaire, la seconde de l'ordre du dictionnaire, se pose la question de leur articulation dans « la traduction d'une phrase p dans un langage de représentation ». En d'autres termes, il importe donc de décider si « dans la traduction de la phrase p dans un langage de représentation artificielle, à partir du langage où p est exprimé », l'automate procède mot à mot dans une démarche ascendante, en suivant l'ordre linéaire des mots de la phrase, ou s'il part du tout unitaire de la phrase dans une démarche descendante, pour en déterminer et interpréter les parties constitutives. Les données qui permettent d'évaluer la pertinence de l'une ou de l'autre méthode résident dans l'aptitude à expliciter les critères qui guident un automate dans la mise en correspondance d'une question formulée dans une phrase interrogative avec sa ou ses réponses formulées dans des phrases déclaratives.

Prenons le cas d'école du jeu de questions mentionnées par R. Martin (2001 : 148) :

Qui était directeur général du CNRS en 1969 ?

Quand Hubert Currien est-il devenu directeur général du CNRS ?

Jusqu'à quelle date Pierre Jacquinot a-t-il été directeur général du CNRS ?

2.1 Il s'agit de montrer comment l'analyseur grammatical qui reconnaît ces phrases interrogatives comme des questions partielles guide l'automate dans la recherche de phrases déclaratives qui saturent le pronom interrogatif remplissant une certaine fonction et maintiennent comme invariants les autres constituants de la phrase interrogative et déclarative. La phrase interrogative et la phrase déclarative sont ainsi appariées par isomorphisme de leur schéma

prédicatif. En l'occurrence, sont appariés les schémas prédicatifs réalisés par les phrases déclaratives suivantes:

Hubert Currien était directeur général du CNRS en 1969.

Le 15 octobre 1969, Hubert Currien est devenu directeur général du CNRS.

Pierre Jacquinot a été directeur général du CNRS jusqu'au 15 octobre 1969.

L'analyse grammaticale des phrases interrogatives et des phrases déclaratives implique donc un modèle unitaire de représentation de la relation prédicative par le même formalisme. Ainsi, la mise en relation de l'assertion et de la question (totale ou partielle) implique :

- a) la reconnaissance de l'unité syntaxique *phrase* réalisée par les types énonciatifs de phrase (interrogatif *versus* déclaratif) et la mise en correspondance des unités grammaticales SUJET VERBE COMPLEMENT(S) instanciées par des valeurs lexicales dans les deux types de phrase,
- b) la représentation dans un même formalisme grammatical de l'encodage des fonctions primaires indépendamment des types énonciatifs de la phrase déclarative et de la phrase interrogative (c'est l'analyseur grammatical qui guide la reconnaissance de la syntaxe des fonctions propre à chaque type énonciatif de phrase),
- c) le stockage structuré dans un modèle de base qui associe données et traitements, base de type relationnelle (SGBDR), susceptible de recevoir les instances lexicales comme enregistrements dans les champs étiquetés SUJET, VERBE, COMPLEMENTS (actanciel / essentiel et circonstanciel).

Ce type de représentation

→ résout le problème de la symétrisation²,

→ autorise des calculs combinatoires et statistiques du fait de la double entrée dans la base : par mots et par enregistrements de schémas prédicatifs, ce qui permet d'articuler une analyse de corpus au module dictionnaire dans une relation d'instances au type.

2.2 La complémentarité de la grammaire et du dictionnaire passe ainsi par le couplage, d'une part d'un analyseur syntaxique qui traduit la représentation linéaire du schéma prédicatif dans la représentation hiérarchique des fonctions et, d'autre part, d'une procédure de saisie des unités lexicales qui instancient ces fonctions comme des attributs représentés par des champs dans des tables de SGBDR. Il y a ainsi un isomorphisme fondé sur le concept de relation entre les unités syntactico-lexicales de la phrase et les champs des tables de la base relationnelle.

² R. Martin (2001 : 104) constate que dans l'aménagement d'un dictionnaire électronique en dictionnaire automatique, il y a nécessité de représenter par une double entrée (« symétriser » les locutions et les syntagmes) les expressions de la collocation, car elle seule permet par appariement la résolution de la polysémie des vocables qui composent le syntagme.

3. Seconde acception de la notion de ‘compréhension automatique’ :

« L’automate comprend une phrase s’il est capable de tirer les conséquences sémantiques que cette phrase détermine ; les conséquences sémantiques d’une phrase p sont l’ensemble des phrases vraies si p est vrai, c’est-à-dire l’ensemble des inférences. » (R. Martin, 2001 : 138)

Il s’agit dans le modèle de la complémentarité de la grammaire et du dictionnaire de montrer que l’automate peut associer les inférences grammaticales inhérentes aux fonctions et les inférences lexicales inhérentes aux mots.

Cette problématique est illustrée par la définition des phrases déclaratives impliquées comme réponses appropriées par une phrase interrogative lorsque le schéma prédicatif de la phrase déclarative n’est pas isomorphe à celui de la phrase interrogative.

Soient les exemples :

Le 15 octobre 1969, Hubert Currien succède à Pierre Jacquinot à la direction générale du CNRS.

Qui était directeur général du CNRS en 1969 ?

Quand Hubert Currien est-il devenu directeur général du CNRS ?

Jusqu’à quelle date Pierre Jacquinot a-t-il été directeur général du CNRS ?

3.1. Il faut ainsi montrer comment la phrase déclarative mentionnée est reliée à chacune des questions, comment elle implique les relations prédicatives que chaque question réalise et comment ces relations prédicatives sont ordonnées par la phrase déclarative mentionnée.

Appliquons le raisonnement à la première question.

Pour que la correspondance puisse s’établir entre la question posée dans la phrase interrogative *Qui était directeur général du CNRS en 1969 ?* et la réponse donnée dans phrase déclarative *Le 15 octobre 1969, Hubert Currien succède à Pierre Jacquinot à la direction générale du CNRS*, il faut pouvoir montrer que le schéma prédicatif de la phrase déclarative signifie implicitement la relation prédicative réalisée par la phrase interrogative : le signifié du prédicat *être*, le constituant sujet, le signifié du constituant attribut *directeur général du CNRS* et le constituant complément circonstanciel *en 1969*.

Or, comme le schéma prédicatif de la phrase déclarative n’est pas isomorphe à celui de la phrase interrogative, il importe d’explicitier la procédure par laquelle la construction SUJET *succéder à* COMPLEMENT 1 (C 1) à COMPLEMENT 2 (C 2) *en* COMPLEMENT DE PHRASE (C P) est reliée au schéma prédicatif de la phrase interrogative. L’analyseur syntaxique ayant reconnu la construction de la phrase déclarative établit une hiérarchie entre les unités guidée par la relation prédicative modélisée par la relation associative 1 – n, où 1 vocable, le prédicat, est relié à un ensemble d’unités réalisant respectivement le SUJET, le COMPLEMENT 1 et le COMPLEMENT 2, spécifiant la valence du verbe. Il faut donc rechercher dans le dictionnaire sous l’entrée *succéder* le sens du prédicat associé à ce type de construction, en suivant le principe de R. Martin (2001 :162) selon lequel « dans la définition du verbe, la construction du verbe, ses places argumentales, sa valence, sont déterminantes ». Se pose alors le problème de la résolution de la polysémie du verbe *succéder*, résolution calculée sur les contraintes syntaxiques et sémantiques inhérentes aux relations du prédicat avec les instances

lexicales réalisant les constituants C2, C1 et S. La recherche du sens dans le dictionnaire est guidée syntaxiquement par

- 1° la construction associée à la trivalence du verbe *succéder* de la phrase occurrence,
- 2° l'instanciation des arguments aux places fonctionnelles SUJET, COMPLEMENT 1, COMPLEMENT 2.

La conjonction de ces deux propriétés permet par abduction d'apparier la construction de la phrase occurrence avec une forme propositionnelle - au sens de R. Martin (2001 : 147) « propositions dont les arguments ne sont pas instanciés et se présentent sous la forme postiche *qqn, qqc* » - liée à un emploi du verbe recensé dans le dictionnaire, caractérisé à la fois par une définition, des conditions d'emploi pour les arguments et un ou plusieurs exemples type.

Ainsi l'analyse syntagmatique de la phrase occurrence où le verbe occupe une position relative par rapport aux syntagmes fonctionnels auxquels il est relié par sa valence permet de lever la polysémie du verbe *succéder* réalisée dans la phrase occurrence et de lui associer une définition, celle du dictionnaire de référence (le TLFi mentionne *prendre la suite de qqn dans un emploi, une charge, une dignité, une fonction* ; le dictionnaire Hachette mentionne *venir après qqn et le remplacer dans une charge, dans un emploi, etc.*).

3.2. En appliquant l'analyseur syntaxique à l'énoncé de la définition, on extrait les unités de la relation prédicative qui permettent d'établir la correspondance de 1 à 1 entre les unités constitutives de la phrase occurrence analysée et celles de la phrase de définition : au prédicat *succéder* sont corrélés les prédicats *prendre la suite de*³, *venir après*, *remplacer*, et aux compléments 1 et 2 sont corrélés respectivement le nom propre *Pierre Jacquinot* et le syntagme *la direction générale du CNRS*. L'analyse syntaxique de la phrase occurrence et de la phrase type de la définition permet par appariement de leur schéma prédicatif de déduire des inférences analytiques sur les unités constitutives :

- le prédicat *succéder* est équivalent aux prédicats *prendre la suite de*, *venir après*, *remplacer*, d'où les équivalences de schémas prédicatifs : S { *prendre la suite de*, *venir après*, *remplacer* } C1 à C2,
- la relation prédicative instaurée par les prédicats { *succéder*, *prendre la suite*, *remplacer* } avec les arguments SUJET, COMPLEMENT 1 et COMPLEMENT 2 permet d'inférer les relations prédicatives :

- a) SUJET avoir C2 *fonction, charge de* substantif dérivé à valeur de nom d'activité⁴ { *la direction générale du CNRS* } ⇔ SUJET être nom d'agent lié morphologiquement { *directeur général du CNRS* }, d'où les conséquences sémantiques
 - par paraphrase : SUJET prédicat { *succéder*, *prendre la suite*, *remplacer* } C1 *comme / en tant que* nom d'agent,
 - de catégorisation du mot-tête de C2 comme hyperonyme { *fonction, charge, emploi* } et du modificateur complément du nom introduit par la préposition *de* comme hyponyme, d'où la

³ Nous nous référons à R. Martin (2001 : 157), pour qui l'hyperonyme peut prendre la forme d'un verbe simple ou celle d'un syntagme.

⁴ « activité » pris au sens de J. Lyons (1980 : 117) : « un processus qui est sous le contrôle d'un agent est une activité ».

- relation d'inclusion : hyponyme 'nom d'activité' {*direction générale du Cnrs*} ETRE_UNE hyperonyme {*fonction*}
- de catégorisation du SUJET comme nom humain représenté par la forme postiche *qqn*, instanciable par un nom propre, d'où la relation d'identification *qqn* ETRE nom d'agent
- b) COMPLEMENT 1 *avoir* C2 *fonction, charge de* substantif dérivé à valeur de nom d'activité {*direction générale du CNRS*} ⇔ COMPLEMENT 1 *être* nom d'agent {*directeur général du CNRS*}, d'où les conséquences sémantiques
- de catégorisation du C1 comme nom humain représenté par la forme postiche *qqn*, d'où la relation d'identification *qqn* ETRE nom d'agent,
- c) des relations prédicatives a et b, on infère :
- l'identité de catégorisation de S et de C1 et la disjonction référentielle de S et de C1, qui déterminent l'équivalence paraphrastique de la construction réflexive du prédicat *succéder* :
S et C1 *se succéder* à C2
 - si coréférence de S et de C1, alors se réalise la construction réfléchie : S *se succéder* à lui-même ou la construction transitive S *succéder* à C1, avec l'instanciation du même signifiant⁵ en S et en C1,
 - la paraphrase⁶ S *succéder* à C1 avec ellipse syntaxique de C2 puisque S et C1 impliquent la catégorie sémantique incluse dans C2.
- d) Les relations prédicatives a et b impliquent une relation d'ordre marquée par un repérage temporel : a => [AVANT] b ⇔ APRES => AVANT .
- La relation prédicative a 'SUJET *être* nom d'agent' est vérifiée APRES la relation prédicative b 'COMPLEMENT 1 *être* nom d'agent' ⇔
- COMPLEMENT 1 *être* nom d'agent AVANT SUJET *être* nom d'agent : C1 *être prédécesseur* de S ⇔ S *être successeur* de C1, d'où les conséquences sémantiques
- d'équivalence paraphrastique C1 *précéder* S à C2 et de la relation converse S *succéder* à C1 à C2,
 - de la contrainte sur l'ordre des éléments SUJET dans la construction réflexive : S et C1 *se succéder* à C2 et non * C1 et S *se succéder* à C2,

⁵ Ce phénomène est illustré par un exemple du type *J.-P. Raffarin succède à J.-P Raffarin*, lorsque la même personne est reconduite dans ses fonctions.

⁶ Cette équivalence est le fait de nom propre en position de sujet et de complément 1 désignant des titres : *Benoît XVI succède à Jean-Paul II* ou *Louis XIII succède à Henri IV* ou *les Carolingiens ont succédé aux Mérovingiens* et est analogue à ce que J. Lyons (1980 : 120) caractérise comme « mécanisme grammatical productif permettant de diminuer ou d'augmenter ce qu'on peut appeler la valence intrinsèque d'un verbe ».

- de généralisation⁷ dans la catégorisation sémantique des entités en S et C1 reliées par les prédicats *précéder* versus *succéder* et appartenant à une même catégorie étiquetée par un hyperonyme,
 - catégorisation temporelle :
S (saison⁸, événement⁹) *précéder* C1 (saison, événement)
vs C1 *succéder* à S,
 - catégorisation spatiale où S et C1 sont les cohyponymes reliés discursivement par la relation prédicative de *succéder* à : *A la route carrossable succédait un chemin de terre*, (exemple du dictionnaire Hachette), où l'hyperonyme discursif pourrait être *voie*.
- e) Le COMPLEMENT DE PHRASE spécifie le repérage temporel par une date et catégorise la relation prédicative S *succéder* à C1 à C2 comme un événement¹⁰ :
- SUJET *être / devenir* nom d'agent à date, à partir de date
 ⇔ C 1 *être / rester* nom d'agent jusqu'à date.

De cet ensemble structuré d'inférences analytiques qui associe grammaire et dictionnaire se déduit l'appariement entre

- les questions {*Qui était directeur général du CNRS en 1969 ? Quand Hubert Currien est-il devenu directeur général du CNRS ? Jusqu'à quelle date Pierre Jacquinot a-t-il été directeur général du CNRS ?*},

- et la phrase déclarative *Le 15 octobre 1969, Hubert Currien succède à Pierre Jacquinot à la direction générale du CNRS*,

puisque les inférences a, b, c, d, e permettent d'extraire de la phrase déclarative mentionnée les relations prédicatives réalisées par les phrases interrogatives et de saturer par isomorphisme les fonctions exprimées par les pronoms interrogatifs.

La complémentarité au niveau des inférences analytiques entre les inférences grammaticales inhérentes à la syntaxe et à la morphologie dérivationnelle et les inférences analytiques inhérentes aux définitions du dictionnaire augmente la puissance descriptive du langage et donc la puissance de l'investigation des bases textuelles dans la mise en corrélation des phrases interrogatives et des phrases déclaratives. Plus généralement, la compréhension automatique des phrases redéfinit le rôle respectif du dictionnaire et de la grammaire. Le premier

⁷ Ainsi s'explique la dérivation polysémique du verbe *succéder* à, dans l'acceptation de *venir après* sans détermination du complément 2 et donc la réduction de la valence ternaire – spécifique du sens premier – à la valence binaire, propre au sens dérivé, avec catégorisation sémantique des arguments en position S et C1. Valence binaire, qui peut être augmentée par la procédure de la causativité : *Dans l'apprentissage de la lecture, il faut veiller à faire succéder des activités courtes*. C'est ainsi que le dictionnaire automatique, en explicitant les procédures de dérivation, explique la relation sémantique des signifiés apparentés d'un vocable polysémique.

⁸ Exemples : *Les saisons succèdent aux saisons ⇔ les saisons se succèdent, l'automne succède à l'été ⇔ l'automne et l'été se succèdent*.

⁹ Exemple : *La chute du Mur de Berlin a précédé la réunification allemande ⇔ La réunification allemande a succédé à la chute du Mur de Berlin*. La relation de conversion par *précéder* vs *succéder* focalise la catégorisation des entités ainsi reliées par la détermination de la différence entre temporalité et causalité.

¹⁰ « événement » pris au sens de J. Lyons, 1980 : 116 : « une situation dynamique, (...) si elle est momentanée, c'est un événement ». On pourrait ajouter que le CP détermine le « moment » de l'événement par une date.

en devenant automatique reçoit le statut théorique de mémoire où les vocables sont des entrées structurées par les relations prédictives qui caractérisent les emplois différenciés des vocables en fonction de paramètres syntaxiques, morphologiques et sémantiques. Le sens d'un vocable est ainsi défini par un type d'emploi différencié dans un ensemble de valeurs appariées à un ensemble ouvert d'occurrences réalisées dans les instances de phrases d'un corpus. L'analyseur grammatical - qui s'applique aux données des phrases d'un corpus et des énoncés d'un dictionnaire - a, quant à lui, comme rôle de traduire les définitions des vocables en schémas prédictifs pour déterminer les inférences analytiques contenues dans les définitions du dictionnaire et les appairer avec le schéma prédictif d'une instance de phrase de texte. Le couplage de la grammaire et du dictionnaire constitue donc le moteur d'inférences syntaxiques et sémantiques qui permet de résoudre la polysémie des vocables en contexte et ensuite de contrôler la compréhension automatique de la phrase. La résolution automatique de la polysémie des vocables et la compréhension automatique de la phrase mise en œuvre dans l'appariement des phrases interrogatives et des phrases déclaratives sont ainsi unifiées dans un modèle de traitement. Ce modèle hiérarchise l'analyse lexicale de la reconnaissance des signifiants et de leur bonne formation, l'analyse syntaxique et sémantique des syntagmes fonctionnels par rapport au prédicat. Cette dernière associe comme attributs aux symboles grammaticaux (vocabulaire auxiliaire et vocabulaire terminal de la grammaire) les inférences lexicales induites par le dictionnaire automatisé à double entrée : celle des schémas prédictifs - sous forme de SGBDR- et celle des définitions de mots.

Cette complémentarité se représente par l'association de règles de production et de règles sémantiques inférentielles : ce que représente adéquatement le modèle de la compilation définie comme traduction dirigée par la syntaxe, qui associe à l'analyseur syntaxique un graphe de dépendances pour représenter la hiérarchie des inférences sémantiques liées aux syntagmes (les nœuds) et aux vocables (les feuilles de l'arbre syntaxique). Ce modèle correspond à la partie frontale d'un programme de compilation par analyse et synthèse tel que défini par Aho *et al.* (1991 : 16) : « Il y a deux parties dans la compilation : l'analyse et la synthèse. La partie *analyse* partitionne le programme source en ses constituants et en crée une représentation intermédiaire. La partie *synthèse* construit le programme cible désiré à partir de cette représentation intermédiaire. »

1^{ère} conclusion : la formalisation de la grammaire est complémentaire de la formalisation d'un dictionnaire modélisé sous forme d'un SGBDR représentant les relations prédictives et les inférences lexicales associées, comme le montre la mise en correspondance de la classe des assertions et de la classe des questions.

2^{ème} conclusion : la formalisation du dictionnaire implique une formalisation du **contenu** des objets typés du dictionnaire informatisé (notamment les définitions, les conditions d'emploi, les exemples) : d'où la complémentarité du dictionnaire électronique (type TLFi) avec un SGBDR pour augmenter la richesse des inférences.

Formalisation de la grammaire et formalisation du dictionnaire induisent un TAL où la linguistique de corpus fournit les phrases instances extraites de textes au compilateur qui, associant arbre syntaxique et graphe de dépendances, permet de modéliser la signification de la phrase instance par appariement avec les relations prédictives types et leurs inférences extraites des objets typés du dictionnaire.

II. LA GRAMMAIRE :

La présentation en sera schématique¹¹ :

- Grammaire formelle : dans une théorie formelle, selon laquelle les théorèmes sont dérivés d'un ensemble de termes et d'axiomes de base obtenus par induction et déduction dans une démarche hypothético-déductive :

1- par induction : observation diachronique et synchronique (G. Reb, 1988) -> généralisation à une cinquantaine de langues non apparentées génétiquement (E. Geniusiene, 1987) -> théorisation (dans une théorie de l'observation et de la généralisation), Benveniste, PLG, I, p. 41 « Ce n'est que par une opération d'abstraction et de généralisation que nous pouvons délimiter un (pareil) objet d'étude. Saussure y insiste, seul le point de vue crée cette substance. Tous les aspects du langage que nous tenons pour donnés sont le résultat d'opérations logiques que nous pratiquons inconsciemment ». La vertu du TAL est précisément d'exiger l'explicitation de ces « opérations logiques ».

⇔ Formes observables -> formes abstraites -> forme théorique symbolique métalinguistique.

Quelles formes observées et quelles relations entre les formes ? (car « le lien qu'on établit entre les choses préexiste aux choses elles-mêmes, et sert à les déterminer », *ibid.*, p.41, Benveniste citant Saussure).

- complémentarité entre morphème grammatical (nombre fini dans un état de langue) et morphème lexical (classe ouverte dans un état de langue) type *se répéter* vs *repeat itself* pour signifier l'équivalence en linguistique générale entre la forme complexe du prédicat réflexif et le syntagme verbal [Verbe – Pronom réflexif] et l'opération de commutation sur la classe d'équivalence (Pronom réflexif vs Pronom non réflexif / SN),

- construit théorique = axiome dans le système d'une LN et métaaxiome dans le système général des langues apparentées, représenté à un niveau abstrait par les symboles de la logique combinatoire de Curry et Feys (cf. G. Reb, 1988),

2- par déduction, à partir de l'axiome, se dérivent des théorèmes par application de règles d'inférence explicites sous forme de règles de production,

⇔ la théorie du langage s'inscrit dans la théorie des langages qui, d'une part, unifie l'informatique fondamentale et la linguistique formelle et, d'autre part, est à la base de la réalisation des compilateurs.

- Grammaire non contextuelle, fondée sur l'association de règles syntaxiques et de règles sémantiques selon le principe des « définitions dirigées par la syntaxe » (Aho *et al.*, 1991 : 311 sq.), dans une théorie de la valeur :

« chapitre IV, La valeur linguistique », *Cours de linguistique générale* (1972 : 168) : la langue est pour ainsi dire une algèbre qui n'aurait que des termes complexes. Parmi les oppositions qu'elle comprend, il y en a qui sont plus significatives que d'autres ; mais unités et faits de grammaire ne sont que des noms différents pour désigner des aspects divers d'un même fait général : le jeu des oppositions linguistiques. Cela est si vrai qu'on

¹¹ Nous renvoyons à G. Reb, « Le schéma d'axiome de la phrase à anaphore réflexive », à paraître.

pourrait très bien aborder le problème des unités *en commençant par les faits de grammaire* ».

« Les catégories grammaticales ont une teneur sémantique élevée » (Jakobson, 1963 : 84).

Application de la théorie de la valeur à la relation prédicative réflexive :

La relation dynamique (en synchronie comme en diachronie) entre le morphème grammatical et le morphème lexical qui construit soit un prédicat complexe (ex : le français) soit un syntagme (ex : l'anglais) signifie, indépendamment de son mode de réalisation dans la substance, la catégorie grammaticale de l'aspect dynamique correspondant à la notion de *processus*, défini par J. Lyons, 1980 : 116, comme « situation dynamique [qui] a une étendue dans le temps ». Ce processus peut concerner un participant, codé par le morphème grammatical du pronom réfléchi avec la fonction syntaxique *objet* (cf. les langues à cas) dont l'identité référentielle est apportée par la catégorie grammaticale réalisant la fonction *sujet*.

ex : il se fait tard vs

il se répète / Jean se répète (= clôture analytique de la phrase minimale définie par la valence intrinsèque du verbe)

=> les propriétés de la grammaire (formelle et non-contextuelle) caractérisent une représentation du fonctionnement linguistique isomorphe avec un *système formel* (cf. D. Hofstadter, 1985) qui montre la correspondance entre la linéarité de la *phrase de texte* et la hiérarchie de la *phrase de système* (au sens de J.Lyons, 1980 : 247), d'où une grammaire

- définie par un quadruplet : vocabulaire terminal (dictionnaire de formes de mots), vocabulaire auxiliaire (code : 66 variables grammaticales), règles de production (grammaire d'une centaine de règles), axiome (la phrase minimale représentant le schéma prédicatif),
- générative, de type syntagmatique : elle définit un langage contrôlé par extension de règles de production (correspondant à un ensemble ordonné de phrases de système) qui caractérisent à la fois une procédure de décision et une description structurelle spécifique (les fonctions syntactico-sémantiques, illustrées notamment par les questions partielles et le problème de la récursivité aux différents niveaux d'analyse).

Le langage défini par la grammaire correspond par isomorphisme aux phrases de textes d'un corpus (documentaire en l'occurrence), ce qui permet de valider la procédure de construction des règles de grammaire à partir d'énoncés attestés.

III. FORMALISATION DU CONTENU DES OBJETS TYPES : APPLICATION DE LA GRAMMAIRE AUX OBJETS DU TLFi

3.1. Les objets de niveaux 1 et 2, les vedettes/catégories grammaticales (n, v, adj, adv) et leurs définitions

3.1.1. L'application de l'analyseur syntaxique aux définitions permet l'extraction des schémas prédicatifs et des contenus propositionnels de la définition par réduction de la phrase complexe ou étendue à la phrase simple isomorphe de formes propositionnelles instanciées.

Soit l'exemple de la définition du substantif *émail* :

- *Email, matière fondante composée de différents minéraux, rendue très dure par l'action de la chaleur, destinée à recouvrir par la fusion, le métal, la céramique, la faïence, la porcelaine pour des fins de protection ou de décoration, et prenant alors des couleurs inaltérables.*

a) *Email, matière fondante composée de différents minéraux* ⇔ l'émail être une matière fondante composée de différents minéraux => la matière fondante être composée de différents minéraux ⇔ différents minéraux composer la matière fondante ⇔ composition de l'émail

b) *Email, matière fondante composée de différents minéraux, rendue très dure par l'action de la chaleur* ⇔ la chaleur rendre très dure la matière fondante (composée de...)

c) *Email, matière fondante composée de différents minéraux, rendue très dure par l'action de la chaleur, destinée à recouvrir par la fusion, le métal, la céramique, la faïence, la porcelaine pour des fins de protection ou de décoration* ⇔ la matière fondante (...) être destinée à recouvrir par la fusion, le métal, la céramique, la faïence, la porcelaine pour des fins de protection ou de décoration

d) *Email, matière fondante composée de différents minéraux, rendue très dure par l'action de la chaleur, destinée à recouvrir par la fusion, le métal, la céramique, la faïence, la porcelaine pour des fins de protection ou de décoration, et prenant alors des couleurs inaltérables* ⇔ la matière fondante (...) prendre des couleurs inaltérables.

- *Email, substance extrêmement dure, translucide, composée à 95% de matières minérales, qui recouvre l'ivoire de la couronne dentaire.*

a) *Email, substance extrêmement dure, translucide* ⇔ l'émail être une substance extrêmement dure, translucide => la substance être extrêmement dure, translucide

b) *Email, substance extrêmement dure, translucide, composée à 95% de matières minérales* ⇔ la substance (...) être composée à 95% de matières minérales

c) *Email, substance extrêmement dure, translucide, composée à 95% de matières minérales, qui recouvre l'ivoire de la couronne dentaire* ⇔ la substance (...) recouvre l'ivoire de la couronne dentaire.

L'explicitation par la grammaire des schémas prédicatifs contenus dans les définitions respectives de *émail* permet de confronter et de différencier les propriétés définitoires des deux entités et donc de résoudre leur polysémie en langue.

Et la typologie des définitions selon la distinction *partie classifiante* vs *partie spécifiante* codée par des opérateurs lexicographiques (R. Martin, 2001 : 164) augmente la puissance du moteur d'inférences lexicales.

La formalisation de la polysémie en langue et l'explicitation des inférences analytiques inhérentes aux types de définitions favorisent ainsi la compréhension automatique des instances de phrases en faisant du dictionnaire un *outil*, qui « comporte des indices qui, discursivement orientent vers tel ou tel sens plutôt que tel autre » (R. Martin, 2001 : 67).

3.1.2. L'analyseur syntaxique, en structurant les propositions contenues dans une définition, hiérarchise les inférences lexicales.

Soit l'exemple de *rançon* (R. Martin, 2001 : 178) :

- *rançon, somme d'argent, valeur que l'on exige contre la remise en liberté d'une personne retenue captive.*

a) *rançon, somme d'argent, valeur* ⇔ la rançon être une somme d'argent, une valeur => syntagme stéréotypique *payer somme d'argent* ⇔ *payer rançon*

b) *rançon, somme d'argent, valeur que l'on exige contre la remise en liberté d'une personne retenue captive* ⇔ on exige la somme d'argent contre la remise en liberté d'une personne retenue captive => - on retient une personne captive
- on remet en liberté la personne captive => AVANT *qqn payer somme d'argent*

3.1.3. Bien que les types de définition dépendent du type grammatical du vocable défini (n, v, adj, adv), l'analyseur grammatical unifie leur traitement par l'extraction des relations prédicatives contenues dans les définitions :

Etudier, par un effort intellectuel, acquérir des connaissances dans un domaine précis.

Dur : qui a une solide résistance physique et morale, qui supporte fermement la maladie, la fatigue, la douleur.

Elégamment : d'une manière élégante.

3.2. Les objets de niveau 3, les crochets et les indicateurs

Dans l'article *étudier*, II A 1 a [*l'objet* désigne une discipline d'enseignement]

Dans l'article *attaquer*, I B 2 b [*le sujet* désigne un phénomène naturel ou un agent inanimé (l'accent étant mis sur le caractère inévitable de son action destructrice)].

Les noms métalinguistiques *objet* et *sujet* renvoient aux instances lexicales associées par le graphe de dépendances aux syntagmes COMPLEMENT et SUJET qui représentent des nœuds de l'arbre syntaxique.

3.3. Les objets de niveau 5, les exemples

Les exemples représentent sous une forme réduite la matière d'un corpus ; dès lors l'application de l'analyseur à des textes autres que littéraires enrichirait la description de la polysémie.

Dans l'article *email*, A- P. *métonymie. Objet d'email, en email.* Objet recouvert d'une couche d'email :

C'est seulement une fois que je vis, un matin, la cuisine froide, la casserole d'email bleu pendue au mur, que je sentis proche la fin de ma mère. (Colette, *La maison de Claudine*, 1922, p. 212).

Le plus souvent au pluriel. Ouvrage d'orfèvrerie fabriqué en email. *Les émaux de Bernard Palissy, des émaux de Limoges.*

Un large gorgerin, composé de fins émaux cloisonnés de traits d'or, cerclait la base du col et descendait en plusieurs rangs. (Gautier)

Dans l'article *Email*, B,

Il existe des animaux où l'émail enveloppe la dent de toute part, tel est le morse.

Dans l'article *étudier*, II A 1 a,

Je croyais, dit timidement M. Delteil, qu'il n'y avait pas moyen d'étudier la botanique sans connaître un peu de latin.

Dans l'article *attaquer*, I B 2 b,

La rouille attaque le fer, la maladie attaque l'organe.

On obtient ainsi une *représentation UNIFIEE* des contenus des objets, exploitable par inférence dans un système de traitement automatique associant dictionnaire et grammaire :

- 1) le traitement unifié du métatexte et du texte permet une généralisation (grâce à la relation hiérarchique hyperonyme/hyponyme) par appariement du contenu des crochets (l'hyperonyme) spécifiant les conditions d'emploi d'un constituant, Sujet ou Complément, avec le syntagme identifié (hyponyme) comme Sujet ou Complément dans l'exemple.

D'où les déductions : *la botanique* être *une discipline d'enseignement*.

la rouille être *un phénomène naturel ou un agent inanimé*.

la maladie être *un phénomène naturel ou un agent inanimé*.

- 2) il permet - par généralisation - la définition systématique de la relation de l'objet de niveau 4, les constructions, avec les formes propositionnelles (au sens de R. Martin, 2001 : 147) dont les arguments (identifiés par les formes postiches *qqn/qqc*) sont instanciés par des vocables appartenant à des « classes lexicales » (R. Martin, 2001 : 165) définies par la relation d'hyper-/hyponymie. L'objet Construction du TLFi peut alors être analysé comme protase dans le système d'inférences qui établit la structure de la base de données relationnelles (SGBDR de paraphrases et de classes lexicales). La base autorise également une *représentation unifiée* des inférences à fondement lexical construites sur la relation de dépendance du prédicat avec ses arguments et une mise en correspondance *multilingue* de leurs sens.

IV. APPLICATION DE LA GRAMMAIRE A L'INTERROGATION DU TLFi

4.1. Le remplissage du formulaire de recherche est guidé par les termes de la question modélisée par des opérateurs lexicographiques :

Soient les exemples du site du TLFi :

Quelles sont les locutions populaires *ayant trait* à la religion?

-> Quelles locutions populaires ont un rapport avec la religion?

Quels sont les verbes *empruntés* à l'italien et qui, lorsqu'ils sont *employés* avec un sens péjoratif, sont *illustrés* par une définition empruntée à l'Académie?

-> Quels sont les verbes empruntés à l'italien, employés avec un sens péjoratif, et illustrés par une définition empruntée à l'Académie?

Quelles sont les définitions *contenant* le mot mets, *utilisées* dans le domaine de l'art culinaire, et *illustrées* par des exemples de Zola?

4.2. Interrogation sur le contenu des définitions

1) par appariement d'un terme de la question avec l'objet de niveau 1, vedette/mot grammatical :

Quel *verbe* exprime le cri de l'éléphant ?

Comment *nomme-t-on* la partie de la physique qui traite des sons ?

dans *acoustique* II B 1 :

Discipline qui traite des aspects physiques de la production et de la transmission des sons et des bruits.

2) par appariement d'un terme générique de la question avec une partie identifiée du contenu de la définition, partie classifiante vs partie spécifiante :

Qu'est-ce que l'émail ?

Quelles sont les *propriétés* de l'émail ?

Un instrument pour enfoncer des pilotis dans la terre ?

Je cherche quelque chose pour enfoncer des pilotis dans la terre.

V. MOTEUR D'INFERENCE ET BASE DE CONNAISSANCES

Les objets du TLFi, notamment les définitions et les exemples, modélisés dans un même langage de représentation, permettent de constituer une base de connaissances.

Soit, l'article *attaquer* :

A ce moment, l'émail formé est assez dur pour résister à l'attaque de l'acide. = exemple 6, sens I B, Emplois métaph. 2 b,

Les exemples enchaînés : La plaque est prête pour la morsure au perchlorure de fer. Le mordant va attaquer le cuivre partout où il est mis à nu par le départ de l'albumine soluble. Il n'attaque pas les régions couvertes d'émail correspondant aux blancs du négatif, c'est-à-dire aux noirs du sujet.

La phrase de texte de l'exemple A ce moment, l'émail formé est assez dur pour résister à l'attaque de l'acide correspond à une proposition complexe composée de quatre propositions calculées par inférences grammaticales et lexicales :

a) l'émail être formé

b) l'émail être dur

c) l'émail résister à l'attaque de l'acide

d) l'acide attaquer ou quelque chose attaquer l'acide : proposition non décidable du fait de l'homotaxie du syntagme *l'attaque de l'acide* à complément du nom objectif ou subjectif.

Les propositions a) et b) sont reliées à la définition sous A de *émail* : *Matière fondante, composée de différents minéraux, rendue très dure par l'action de la chaleur,(...)*

- la relation est opérée par appariement entre les synonymes *formé/composé* et par équivalence de l'adjectif *dur*.

La proposition c) est définie par la construction et la sous-catégorisation de *résister à* : *qqc. résister à qqc.* sous I - [Le sujet désigne un inanimé] *Faire obstacle à une action ou à une force.* A. *Ne pas céder sous l'effet d'une force* ou, B. *Ne pas s'altérer sous l'effet d'un agent extérieur.*

La construction et la définition synonymique *résister/faire obstacle* ou *ne pas céder* ou *ne pas s'altérer* induisent la définition par hyperonyme/hyponyme : *action/attaque* et *agent/acide* et permettent de lever l'ambiguïté de la proposition d) en faisant correspondre le rôle de valence AGENT au sujet grammatical *acide* et le prédicat ACTION au verbe transitif *attaquer*.

Dès lors, le lieu lexicographique de *attaquer* est bien **I B 2 b** [Le sujet désigne un phénomène naturel, un agent inanimé (l'accent étant mis sur le caractère inévitable de son action destructrice)]. La rouille attaque le fer, la maladie attaque un organe. *Synon. Affecter, altérer, corroder.* Le synonyme *altérer* qui relie la définition de *attaquer* et celle de *résister à* permet de décider du sens **I B** pour *résister à*.

Les inférences grammaticales (par calcul sur l'arbre syntaxique) et lexicales déduites des propositions c) et d) permettent d'inférer la proposition e) :

- e) L'acide n'attaque pas l'email.

Le traitement des exemples enchaînés –correspondant à des microtextes - augmentera encore la base de connaissances en permettant les inférences interphrastiques associées au traitement de l'anaphore.

Une représentation unifiée des contenus des objets hiérarchiques du TLFi permet ainsi de faire de ce « trésor » une base de connaissances contrôlée, aisément articulable à un système de gestion documentaire et à un moteur de recherche fondé sur une sémantique linguistique. Cette articulation est en effet établie sur la hiérarchie des inférences analytiques, selon laquelle les inférences prédictives construites dans la clôture empirique du dictionnaire associé au système de la grammaire contrôlent les inférences sémantiques en langue (linguistique du dictionnaire automatique) et en discours (linguistique de corpus).

UNE PERSPECTIVE EN CONCLUSION

La théorie saussurienne du signe, en servant de cadre conceptuel à la définition des relations entre morphèmes lexicaux et morphèmes grammaticaux, fournit un modèle au traitement automatique du sens noté par les systèmes de langues. La part de la sémantique automatisable est ainsi circonscrite par la puissance de représentation du métalangage grammatical et lexical. En prenant comme entrées les phrases de texte issues d'un corpus et celles issues des objets typés du dictionnaire, le compilateur modélise la relation des instances au type, d'une manière contrôlable et toujours extensible, dans un moteur d'inférence syntaxique et sémantique, qui associe des attributs sémantiques aux symboles métalinguistiques de la

grammaire. De là procède sa limite, car seul est représentable le signifié associable à du signifiant. De ce fait substantiel, est seul représentable ce que la langue peut traduire réflexivement dans une métalangue, elle-même traduite dans un langage artificiel de haut niveau. Ce faisant, le TAL implique la nature analytique du sens, qui se fonde sur les mécanismes grammaticaux et lexicaux de la relation sémiotique constitutive du langage. La réflexivité métalinguistique porte ainsi sur le codage de ces mécanismes dont chaque langue définit les critères de validité internes à sa grammaire et à son dictionnaire. La compilation d'une langue naturelle implique donc la formalisation de ces mécanismes dans une écriture métalinguistique avant son écriture dans un langage informatique de haut niveau. La réflexion sur la part de sens automatisable d'une langue s'inscrit alors dans une théorie générale de la traduction selon le double point de vue du dictionnaire et de la grammaire, comme le souligne Jakobson (1963 : 80-81) : « Il est difficile de surestimer le besoin urgent, l'importance théorique et pratique, de dictionnaires bilingues différentiels, qui définiraient soigneusement et comparativement toutes les unités correspondantes, en extension et en compréhension. De même, des grammaires bilingues différentielles devraient définir ce qui rapproche et ce qui différencie deux langues du point de vue de la sélection et de la délimitation des concepts grammaticaux ». Dans la réflexion qu'il promeut sur le codage du sens dans la théorie de la compilation, le TAL va plus loin, en exigeant le codage de la relation analytique du sens lexical et du sens grammatical propre à chaque langue, en amont de sa numérisation.

BIBLIOGRAPHIE

- Aho A., Sethi R., Ullman J., 1991, *Compilateurs. Principes, techniques et outils*, Paris, InterEditions
 Benveniste E., 1966, *Problèmes de linguistique générale*, Paris, Gallimard
 Hagège C., 1982, *La structure des langues*, Paris, PUF
 Hofstadter D., 1985, *Gödel, Escher, Bach*, Paris, InterEditions
 Genusiene E., 1987, *The Typology of Reflexives*, Berlin, New York, Amsterdam, Mouton de Gruyter
 Jakobson R., 1963, *Essais de linguistique générale*, Paris, Les Editions de Minuit
 Lyons J., 1980, *Sémantique linguistique*, Paris, Larousse
 Martin R., 2001, *Sémantique et automate*, Paris, PUF
 Reb G., 1988, *Valeurs intrinsèques des verbes pronominaux en français moderne et en ancien français. Etude descriptive et théorique*, Thèse d'Etat, Université de Strasbourg
 Reb G. 1995, éd., *Langages et ordinateurs*, Scolia n°4, USHS
 Saussure F. de, 1972, *Cours de linguistique générale*, Paris, Payot
 Tamba I., 1994, *La sémantique*, Paris, PUF

