

MÁRTON NÁRAY-SZABÓ

Anne Abeillé : *Une grammaire électronique du français*¹

La linguiste Anne Abeillé, collaboratrice du laboratoire LADL (Laboratoire d'Automatique et de Documentation Linguistique), s'inspirant des nouvelles tendances des grammaires dites d'unification, se base sur une branche relativement récente de ce courant : la grammaire « TAG » (Tree Adjoining Grammar, Grammaire à Arbres Adjoints). La théorie élaborée par l'américain Aravind JOSHI s'applique – grâce à cette thèse de doctorat de l'auteur, publiée sous forme d'un manuel abrégé de ladite grammaire – au français, en ajoutant quelques éléments d'intérêt généralisables à toutes les langues. Suivant la logique du livre insistant toujours sur les contrastes entre grammaires traditionnelles (la théorie distributionnaliste de Z. F. HARRIS et de M. GROSS et le générativisme sous une de ses formes les mieux élaborées : la théorie X-barre) et grammaires d'unification, on observe que la différence essentielle se situe dans la conception des phénomènes syntaxiques. Ici on n'a pas de structure profonde et, tandis que les transformationnalistes supposent l'existence de procédés dynamiques, les unificationnistes restent essentiellement statiques dans la mesure où ils ne parlent pas de changement de structure (« transformation »), mais de conditions de bonne formation des phrases (« règles lexicales ») sans qu'il y ait une construction privilégiée (la phrase active personnelle, sans ellipse ni ajouts chez les générativistes) engendrant toutes les autres structures (passifs, impersonnels, causatifs, etc.). Le modèle exposé ici se distingue également des grammaires traditionnelles en ce qu'il est entièrement lexicalisé : il n'utilise pas de règles généralisables contenant exclusivement des unités de sous-catégorisation (catégories grammaticales) comme entrées du côté gauche et droite, mais des arbres à ancre(s) lexicales (lexicaux). En même temps, ces arbres incorporent aussi des catégories et des fonctions grammaticales (cette dernière dans un sens absolument traditionnel), ce qui rend cette grammaire légèrement dépendante du contexte ('mildly context sensitive') : elle est ainsi plus puissante que les grammaires totalement hors contexte. Les propriétés syntaxiques et sémantiques ne sont donc pas rapprochées directement, comme chez CHOMSKY et GROSS.

¹ CNRS Éditions, Paris, 2002.

Par rapport à la théorie du Gouvernement et du Liage, on doit pourtant noter quelques désavantages. En premier lieu, sa lourdeur et le manque d'homogénéité dans le traitement et la structuration des syntagmes : on doit tout coder dans la grammaire (la composition d'un syntagme), qui se constitue essentiellement d'une série d'arbres syntactico-lexicaux avec une immense quantité de traits. Bien qu'il existe un grand nombre de généralisations, les arbres doivent être définis d'une manière quasiment unique pour chaque item lexical. C'est en ce sens qu'on peut observer un parallélisme avec la perspective grossienne, le lexique-grammaire, qui vise à décrire lexique et grammaire ensemble (sous forme de tableaux). Le TAG présente la nouveauté de réaliser cela dans une structure arborescente.

Par rapport à la théorie du Gouvernement et du Liage, on doit pourtant noter quelques désavantages. En premier lieu, sa lourdeur et le manque d'homogénéité dans le traitement et la structuration des syntagmes : on doit tout coder dans la grammaire (la composition d'un syntagme), qui se constitue essentiellement d'une série d'arbres syntactico-lexicaux avec une immense quantité de traits. Bien qu'il existe un grand nombre de généralisations, les arbres doivent être définis d'une manière quasiment unique pour chaque item lexical. C'est en ce sens qu'on peut observer un parallélisme avec la perspective grossienne, le lexique-grammaire, qui vise à décrire lexique et grammaire ensemble (sous forme de tableaux). Le TAG présente la nouveauté de réaliser cela dans une structure arborescente.

Toute grammaire formelle utilise comme « matière première » les éléments suivants : 1. *Axiome* ; 2. *Vocabulaire terminal* ; 3. *Vocabulaire auxiliaire* ; 4. *Ensemble de règles*. En TAG, les deux premiers existent en tant que tels, le troisième inclut les *traits* (ceux-ci sont ancrés sur chaque nœud non-terminal de l'arbre) et les catégories. Au lieu de poser l'existence de règles « omnipotentes », on a affaire ici à des arbres appelés initiaux et auxiliaires. Ces deux types d'arbres doivent décrire toutes les constructions grammaticales. Les règles dites lexicales ne font que déduire les tendances les plus importantes, qui connaissent plus d'exceptions que de cas réguliers.

L'opération d'*adjonction* (= insertion dans un non-terminal) et de *substitution* (= insertion sur un terminal) constituent les deux procédés de dérivation. Le résultat s'appelle *arbre dérivé* (AD), qui n'est pas à confondre avec l'*arbre de*

dérivation (AdD). Le premier exprime d'une manière simpliste l'ordre de constitution, le deuxième présente l'« historique » des dérivations des arbres, interprétant ainsi la sémantique de la phrase. Une phrase ne correspond pas forcément à un AD et à un AdD.

Les AdD exigent également une notation pour le lieu d'attachement et l'ordre des constituants est très strictement réglementé. Cela rend la description quelque peu obscur malheureusement. La tête de ce type d'arbre équivaut au « gouverneur » dans le sens chomskien.

Les nœuds non-terminaux et terminaux ne sont évidemment qu'une sorte de squelette pour la représentation. Les particularités morpho-syntaxiques et sémantiques (fonction) se reflètent dans les *traits* dont certains se rattachent à chaque nœud non-terminal. Ces traits ont ceci de particulier qu'ils se propagent sans limitation à l'intérieur d'une proposition. Chaque nœud comprend deux séries de traits : les premiers se répandent vers le haut (t), les deuxièmes vers le bas (b). Après adjonction, les traits de la racine du substitut sont *unifiés* avec les traits 't' de l'arbre à substituer à l'endroit de substitution ; ceux du pied avec les traits 'b' du même endroit. Après l'*unification*, les traits 't' et 'b' de chaque nœud se mettent ensemble. Les arbres « nouveaux venus » peuvent changer des deux côtés, en amont et en aval, les traits de l'arbre originel : c'est le cas des adjonctions obligatoires (de déterminants, de modifieurs). Ainsi, dans les arbres sont possibles toutes sortes de furcations (bi-, tri-, etc.), et certains posent même des adjonctions simultanées de plusieurs arbres sur un même arbre. Les traits peuvent se diffuser d'un pied à un autre (*phénomènes d'accord, etc.*), d'un pied à la racine (*interrogation, etc.*) ou d'une racine aux pieds (*mode, etc.*).

Au lieu d'imposer donc des règles dynamiques, les TAG se fonde sur les arbres dont le fonctionnement dépend uniquement des *principes de bonne formation*. Ce sont les suivants : 1. une tête lexicale ne peut être vide (sauf dans le cas de la complétive infinitive ou de l'impératif), 2. le prédicat (V, N, A) doit exprimer tous ses arguments, 3. aucun arbre élémentaire ne peut être sémantiquement vide (les mots grammaticaux ne forment pas d'unités autonomes), 4. un arbre forme une unité sémantique (les phrasèmes sont des unités à plusieurs têtes lexicales).

En effet, la notion de tête donne lieu à une confusion. Non seulement parce que le terme est utilisé dans un sens autre que celui des grammaires

structuralistes-transformationnalistes, mais parce qu'il se définit de trois manières selon qu'on parle d'un AE, d'un AD ou d'un AdD. Dans le premier cas c'est l'ancre lexicale, dans le deuxième c'est la racine, dans le troisième c'est le gouverneur. L'auteur aurait très bien pu remplacer ces termes respectivement par *ancre*, *racine* et *tête* par exemple.

Dans le chapitre consacré à la description de la phrase française, Abeillé présente toute une série d'arguments contre l'équation générative traditionnelle $P = SN + SV$. Ce dernier (le syntagme verbal) n'existe pas selon elle. À côté de quelques arguments phonétiques difficilement vérifiables (liaisons, pauses), elle énumère des facteurs distributionnels (coordination, découpage hypothétique $SV + SN$ de la P, l'existence de SV discontinus, possibilité d'intercaler le SN sujet dans SV, la composition très variée du SV, le statut ambigu des auxiliaires, des infinitifs et des participes) et des facteurs sémantiques (l'existence des expressions à sujet figé, l'identité de la représentation structurelle de la lecture littérale et idiomatique dans les expressions).

Après avoir énuméré les catégories dans leur intégralité dans le chapitre 3, l'auteur décrit en détail leurs propriétés morpho-syntaxiques et sémantiques et propose à chaque fois une méthode de représentation dans le modèle. On peut soit permettre quelque chose à l'aide d'une nouvelle règle, soit interdire la réalisation agrammaticale par des traits ($\langle \text{mode} \rangle \neq \text{ppart}$ dans les arbres de clitiques, ce qui leur interdit de se placer directement devant un participe passé), soit dégager la propriété directement de la structure arborescente (l'arbre des Aux n'indique que le V régi et ses compléments, donc l'auxiliaire ne peut pas avoir d'arguments indépendants), soit créer un conflit entre les traits en amont et en aval (donner + en amont et – en aval pour le trait $\langle \text{det} \rangle$ pour exiger l'insertion d'un Det).

Les catégories grammaticales sont au nombre de neuf : A, Adv, C (Complémenteur), Cl (Clitique), Conj, D (Déterminant), N, Prép, V. Les syntagmes n'ont pas de description particulière sauf le SP : il se compose d'une Prép et d'un autre élément (N / SP). La structure profonde n'existe donc pas ici. Les mouvements au sens chomskien (mouvement du SN (*NP-movement*), mouvement Qu (*WH-movement*)) ne sont que de simples inversions : c'est un avantage de plus du modèle. Les arguments (sujet inclus !) sont au même niveau que le prédicat et la phrase est conçue comme la projection d'une tête lexicale

prédicative. Celle-ci détermine la topologie de l'arbre. Les complémenteurs partagent également le niveau de leurs phrases subordonnées.

Les verbes se subdivisent en verbes *déficients* (auxiliaires (+ part. passé), modaux, aspectuels (+ inf.), attributifs (+ attribut) et causatifs (+ inf.)), verbes à *argument phrastique* éventuel (complétive à l'indicatif, au subjonctif ou infinitive) et verbes à *complément* exclusivement *nominal*. Des traits indiquent le caractère copulatif ou causatif du verbe, et les propriétés des compléments. À cela on aurait envie d'ajouter les traits généraux pour les prédicats de G. GROSS : événements, états, actions.

Les prépositions et les conjonctions de subordination sont unifiées sous la catégorie Prép, et sont toujours des têtes d'ajouts (*et, ou, selon, selon que, etc.*). Dans le cas de coordination de modifieurs (*une fille qui danse bien et que tous admirent*) l'analyse de l'auteur ([N0 modif1] et [ϵ^2 modif2]) pourrait être améliorée en posant N0 [modif1 et modif2]. Les adverbes sont soit des arguments, soit des modifieurs, ce qui se reflète dans les arbres des prédicats uniquement. On a des adjectifs épithètes (le trait <modif> devient +), attributs (avec copule obligatoire [<cop> = +]) ou déficients (type *facile à, susceptible de* ; ils n'ont pas d'arguments autres que ceux de l'infinitif régi [<facile> = +]). Les noms sont classés comme suit : non-prédicatifs, prédicatifs, pronoms (excepté les clitiques), quantifieurs. Les déterminants sont des ajouts dans le modèle. Leurs traits sont d'un nombre particulièrement élevé. Les clitiques sont traités séparément des autres pronoms pour empêcher de les coordonner avec des syntagmes N ou SP. Le clitique sujet est distingué des autres : il supporte l'insertion d'auxiliaires modaux, d'adverbes, etc., il est donc d'un étage plus haut dans l'arbre.

Le chapitre portant sur les familles d'arbres élémentaires énumère les types de phrases et les compare avec les classes d'équivalence de M. GROSS. La fonction grammaticale dans la phrase ne se déduit pas directement de la position du composant, comme chez CHOMSKY ou GROSS, mais est déterminée préalablement par le trait <fonc>. Deux types de règles lexicales existent : les *redistributions fonctionnelles*, qui peuvent changer la topologie de l'arbre : ajouter des nœuds, changer leur ordre et modifier les traits (passif, impersonnel,

² Tête vide.

causatif, réfléchi, neutre, moyen, symétrie), et les *réalisations d'arguments*, qui opèrent après les redistributions en mettant à jour les équations sur les traits (ajout de ϵ dans les positions vides, réalisations de clitiques, inversions).

L'auteur, non content de décrire les types d'extractions un à un, propose des équations sur les traits de façon à inclure tous ces types et leurs combinaisons en même temps à l'aide de trois traits : <inv> (inversion du sujet pronominal ou nominal), <princ> (on est dans une principale ou subordonnée) et <qu> (interrogative ou relative). Ces traits permettent la totalité des phrases grammaticales de ce genre et interdisent toutes celles qui sont agrammaticales. D'ailleurs, on n'a pas besoin du signe ϵ ni de la notion compliquée de mouvement comme chez les transformationnalistes. On pourrait dire que c'est la partie la plus élaborée et la plus géniale aussi de ce modèle.

Les quelque trente pages consacrées au traitement des expressions figées (EF) dans le cadre du TAG offrent une vue particulièrement intéressante. Abeillé s'oppose nettement à l'idée d'une réanalyse des EF comme verbes composés pour des raisons différentes. Les EF, qui constituent un continuum allant du syntagme libre à la phrase entièrement figée, sont opaques sémantiquement. Ce caractère est à vérifier par des tests. Abeillé adopte donc l'image d'un arbre élémentaire à co-têtes lexicales pour les EF, dont la profondeur est supérieure à un en général et dont les arguments libres sont contraints par la catégorie (sous forme de nœuds pieds). Les compléments nominaux ou prépositionnels peuvent avoir une fonction : cela permet ou bloque certaines constructions (transformations). Les arbres figés sont engendrés à partir d'arbres élémentaires (conception dynamique du figement), mais redeviennent élémentaires justement grâce à ce processus (les têtes sont ainsi sémantiquement équivalentes). C'est ce qui explique bien leur caractère double : flexibilité syntaxique relative d'un côté, figement sémantique de l'autre. Les avantages incontestables de cette conception sont : la possibilité de conservation des traits de fonction (cela rend possible la passivation et beaucoup d'autres redistributions) ; des mots comme *escampette* ou *haro* ne figurent pas comme entrée lexicale artificielle dans le lexique, mais entrent *directement* dans un arbre figé ; les discontinuités mentionnées ci-dessus sont désormais prévisibles / gérables ; par la sous-spécification de certains traits, on représente aisément des phénomènes d'accord comme *N0 casser Poss0 pipe*, où

Poss = mon, ton, son, etc. ; le reflet de la variabilité lexicale par les disjonctions sur les co-têtes (perdre [le Nord / la boussole / sa tête]) ; l'insertion obligatoire à l'aide de traits en amont et en aval en conflit ; l'insertion facultative permise et l'insertion interdite par des traits (par exemple : <poss> = -) ; la réglementation de l'insertion de déterminants et de modifieurs grâce à des équations sur les traits ; les AdD visualisent directement la lecture non-idiomatique et idiomatique. On est capable de rendre compte du parallélisme par l'isomorphisme des AdD entre les phrases suivantes : *Jean a pris une vraie veste. Jean a vraiment pris une veste. Ils apportent de l'eau (1) au moulin (2) indépendantiste / de l'indépendantisme.* Abeillé propose l'application de TAGs synchrones (mis ensemble avec les TAGs syntaxiques) visualisant purement les composants sémantiques de la phrase.

Tout compte fait, le modèle présente une vue radicalement nouvelle dans la mesure où il rompt avec la bipartition traditionnelle de la phrase (SN + SV) et étend – grâce aux arbres – la notion de localité dans la grammaire. Deux phénomènes se présentent dans ce paradigme de façon tout à fait particulière : les extractions et le figement.

MÁRTON NÁRAY-SZABÓ

Université Eötvös Loránd, Budapest
Courriel : mnaray@yahoo.com