

La syntaxe, c'est de la sémantique

Pascal Vaillant

Université Paris-13, Sorbonne Paris-Cité, LIM&Bio
(Laboratoire d'Informatique Médicale et de Bio-Informatique)

74, rue Marcel Cachin, 93017 Bobigny cedex

Courriel : vaillant@univ-paris13.fr

I Introduction

Le concept d'*afférence*, proposé par Rastier¹ pour décrire à l'échelle microscopique la construction du sens dynamique en contexte, est un mécanisme fondamental de la sémiogenèse, auquel les sciences du langage n'ont sans doute pas accordé jusqu'ici l'importance qu'il mérite. L'*afférence* est le principe par lequel une unité sémiotique peut importer, par « contagion », un sème présent dans son contexte ; ce sème vient ainsi s'agréger à son signifié dynamique.

Ce concept permet de rendre compte de la construction d'effets de sens, et il est l'une des manifestations de la dynamique de l'interprétation. Il relève donc du champ de la sémantique, et il est logique que ce soit dans ce champ qu'il ait émergé.

Nous souhaitons montrer ici qu'il permet aussi de rendre compte de phénomènes syntaxiques, comme l'attribution à une unité linguistique d'une fonction dans la phrase. Ce type de phénomènes est habituellement délégué à une tradition descriptive — celle de la syntaxe — jugée plus rigoureuse que la sémantique, et pouvant fournir un palier séparé d'analyse.

Le fait que des outils descriptifs issus de la sémantique rendent parfaitement compte de phénomènes syntaxiques n'a rien de surprenant en soi, puisque les deux niveaux d'analyse s'intéressent au fond à la même classe d'unités (des signes linguistiques) ; la seule frontière nette que l'on puisse objectivement tracer dans les classes d'unités linguistiques passant entre les signes (porteurs d'une valeur) et les segments purement distinctifs (phonèmes), et non entre « morphèmes lexicaux » et « morphèmes grammaticaux » — distinction dont la typologie et la linguistique historique démontrent qu'elle est graduelle, et variable dans le temps et l'espace. La question intéressante est donc plutôt de savoir si l'on peut se débarrasser de la séparation traditionnellement maintenue (de façon, à notre avis, trop rigide) entre les deux paliers d'analyse, tout en gardant les apports de l'un comme de l'autre.

Le présent article propose une manière d'intégrer une représentation du phénomène d'afférence dans un modèle de description formel de la syntaxe qui a fourni des preuves de son utilité et de son efficacité dans le domaine du traitement automatique de la langue (celui des grammaires d'arbres adjoints à unification de structures de traits). Nous montrerons que ce qui bloque cette intégration n'est pas le principe du formalisme lui-même. Celui-ci, au contraire, dispose de la capacité de représenter l'intégration de structures dynamiques, et fournit une modélisation naturelle de l'afférence. Ce qui bloque est la rigidité artificielle qu'introduit, dans ce modèle, l'habitude d'attribuer une catégorie syntaxique fixe (caractérisation de la *partie du discours*) à chaque mot ou syntagme.

Cette subdivision du lexique en catégories étanches, les parties du discours, n'est en réalité pas indispensable à une représentation du fonctionnement syntaxique de la langue ; en outre, elle repose

¹ RASTIER F., *Sémantique Interprétative*, Paris, PUF (coll. « Formes Sémiotiques »), 1987.

sur une tradition descriptive qui n'est pas universelle, et s'adapte assez mal à la description de certaines langues. On peut construire un modèle de la syntaxe en s'en dispensant, et ce renoncement permet, dans le cas qui nous intéresse, de transcender la séparation artificielle entre paliers de description syntaxique et sémantique.

II Situation du problème

II.1 Les catégories de parties du discours

Selon Lallot², Platon introduit la première opposition consciente entre catégories de mots dans le *Cratyle*, où il fait la distinction entre ονοματα et ῥήματα (noms et verbes). Aristote (*De l'Interprétation, Poétique*) introduit explicitement la notion de parties du discours (μερη λεξεως), et ajoute, au nom et au verbe, la conjonction (συνδεσμος) et l'article (αρθρον). La liste des parties du discours de la langue grecque se stabilise chez les grammairiens alexandrins (notamment Denys le Thrace et Apollonius Dyscole) en un inventaire de huit catégories : ονομα (nom), ῥημα (verbe), μετοχη (participe), αρθρον (article), αντωνυμια (pronom), προθεσις (préposition), επιρημα (adverbe), et συνδεσμος (conjonction).

La définition des parties du discours pour la langue française, fournie dans la *Grammaire* de Port-Royal reprend encore quasiment à l'identique cette liste, en y ayant seulement ajouté l'interjection (l'adjectif y est encore considéré comme une sous-catégorie du nom, ceux-ci étant divisés en *noms substantifs* et *noms adjectifs*).

Il est important de noter que ces catégories sont toujours définies prioritairement sur un critère essentialiste (par exemple, dans les termes d'Arnauld et Lancelot, les noms désignent « les objets de nos pensées », et les verbes « la forme et la manière de nos pensées »), mais que dès Aristote, puis de manière plus nette chez Apollonius, et très explicite enfin chez Arnauld et Lancelot, des critères morphologiques sont utilisés pour les discriminer : « Ainsi le verbe, selon ce qui lui est essentiel, est un mot qui signifie l'affirmation. Mais si l'on veut joindre, dans la définition du verbe, ses principaux accidents, on le pourra définir ainsi : *Vox significans affirmationem, cum designatione personæ, numeri et temporis* : *Un mot qui signifie l'affirmation, avec désignation de la personne, du nombre et du temps* [...]»³.

La convocation des critères morphologiques est intéressante car elle montre la difficulté à trouver des délimitations nettes en ne se fondant que sur des critères essentialistes ; et qu'elle constitue en même temps un ancrage non-avoué de ces catégories dans l'étude d'un certain type très particulier de langues (le grec, le latin, l'hébreu, le français), qui possèdent en commun une morphologie flexionnelle, et le marquage indiciel de catégories comme le genre, le nombre, la personne et le temps. Dans la variété des langues du monde, ces quatre langues sont typologiquement très voisines, et au fond assez peu représentatives de l'ensemble. Mais ni Apollonius, ni les grammairiens de Port-Royal, n'envisagent l'idée que ces critères grammaticaux puissent être autre chose que des « accidents » : la définition sémantique universelle doit primer.

² LALLOT, J., « Origines et développement de la théorie des parties du discours en Grèce », *Langages*, vol. 23, n° 92, décembre 1988, p. 11–23.

³ ARNAULD et LANCELLOT, *Grammaire Générale et Raisonnée (Grammaire de Port-Royal)*, Paris, Allia, 1997, p. 13 (1^{ère} édition 1676).

Dans cette conviction qu'il est possible de donner une définition purement essentielle du concept⁴ réside l'idée que tout mot doit pouvoir être rattaché à une catégorie — autrement dit que les catégories constituent une partition⁵ de l'ensemble des mots d'une langue :

« D'une représentation de la langue comme collection grammaticalement amorphe de noms (*onomata*), on est passé par étapes à une taxinomie rigoureuse selon laquelle *tout mot grec* relève d'une des huit classes appelées « parties du discours », et *d'une seule*.

Le critère de classement est, pour chaque mot, sa conformité à l'*idia ennoia*, la « notion propre », qui définit une classe. Cette notion propre inclut des données morphophonologiques (présence/absence de flexion, flexion casuelle/personnelle notamment), mais, pour Apollonius, ce qui la constitue essentiellement, c'est un « signifié » (*sêmainomenon*) de nature sémantico-syntaxique (le nom désigne une substance avec une qualité, l'article exprime l'anaphore du déjà connu, l'adverbe prédique le verbe, etc.). C'est donc ce « signifié » qui fournit le critère principal de classification d'un mot⁶. »

II.2 Utilisation en linguistique formelle

La grammaire formelle se développe à partir du milieu du xx^e siècle avec un double projet : rendre le fonctionnement de la langue *modélisable* pour les applications automatiques qui commencent à devenir envisageables (par exemple la traduction automatique) ; et offrir un modèle du fonctionnement de l'esprit humain lorsque celui-ci manipule le langage. Cette grammaire formelle a des ambitions explicitement universalistes (chez Chomsky).

Dans ses deux grandes variantes d'origine, la grammaire formelle hérite dès le départ de la conception alexandrine de la grammaire, avec ses parties du discours étanches, son intérêt pour la phrase comme objet central du langage, et sa conception de la phrase comme proposition logique décomposée en sujet et attribut, sans se poser (plus que les grammairiens alexandrins ou ceux de Port-Royal) la question de son adéquation à la diversité des langues.

Le modèle des PSG (*Phrase Structure Grammars*) proposé par Chomsky⁷ prend pour point de départ la structure syntagmatique des phrases. Il propose, comme mécanisme fondamental de l'articulation de la langue, des règles qui permettent de combiner des syntagmes, en fonction de leur catégorie, en syntagmes de niveau supérieur (par exemple, une phrase est la combinaison d'un groupe nominal et d'un groupe verbal). Les étiquettes de parties du discours ont dans ce modèle un rôle charnière, puisqu'ils sont les premières catégories abstraites : les catégories des syntagmes minimaux, encore appelées « pré-terminaux ».

Cette conception mécaniste de la grammaire, on le voit, ne s'intéresse pas (tout au moins dans sa première version) à la *fonction* des unités lexicales ; elle tient tout entière dans une définition rigide de leur *nature*. Dans les grammaires syntagmatiques, il n'y a pas d'agent ou de patient, il n'y a que des groupes nominaux.

⁴ Le mot *concept* est ici bien entendu un anachronisme ; le mot utilisé par Apollonius est *σημαινομενον* (*sêmainomenon*) — qu'il est tout aussi anachronique de traduire par *signifié*.

⁵ *Partition* au sens mathématique, c'est-à-dire : répartition en un ensemble de sous-ensembles disjoints dont l'union reconstitue l'ensemble total.

⁶ LALLOT, J., art. cit., p. 22.

⁷ CHOMSKY, N., *Structures Syntaxiques*, Paris, Seuil (coll. « Points / Sciences Humaines ») (éd. orig. La Haye, 1957)

L'autre grande conception fondatrice de la grammaire formelle, la théorie des grammaires catégorielles (Ajdukewicz, Bar-Hillel, Lambek), est d'inspiration logique plutôt que mécaniste. Elle donne, comme son nom l'indique, un rôle plus central encore aux catégories. La construction d'une phrase y est conçue comme une démonstration de preuve mathématique, en suivant les règles du raisonnement logique (axiome, *modus ponens* ...)

Cette autre conception est elle aussi parfaitement adaptée à une modélisation des mécanismes linguistiques destinée à des usages en traitement automatique : elle calque en effet sur le plan de la syntaxe une conception de la sémantique comme calcul référentiel, introduite par Montague⁸, et dont les catégories de base sont l'entité (e) et la proposition ou *truth value* (t). Tous les éléments de la langue y sont considérés comme des catégories dérivées permettant, par combinaison formelle, d'obtenir en dernier ressort l'une de ces deux catégories fondamentales.

L'application, dans des formalismes grammaticaux de la fin du deuxième millénaire, d'une conception de la grammaire datant du début du premier, n'a pas manqué de soulever certaines interrogations. La notion de catégorie de partie du discours a été critiquée, on le verra plus loin, par Chomsky lui-même (dès 1970). Pourtant, la linguistique formelle (et en tout cas son bras armé, l'informatique linguistique) y revient constamment, comme par habitude acquise. Ce fait tient, semble-t-il, à deux grandes raisons.

La première est que les formalismes informatiques permettant de gérer un niveau de grain un peu plus fin que ces catégories sont apparus relativement tardivement (on pense aux grammaires d'unification type HPSG ou TAG, apparus à partir du milieu des années 1970), et que cela a laissé le temps à tout un outillage conceptuel et méthodologique de se développer sur la base des CFG : algorithmes (comme l'algorithme d'analyse d'Earley) ; extensions de la théorie à d'autres domaines, comme celui du discours (RST ou DRT, étendant un formalisme assez comparable à celui des PSG à une échelle de texte plus macroscopique) ; articulations entre la syntaxe et la sémantique ; développement de bibliothèques informatiques et de banques de données ...

Les parties du discours étant le premier maillon, indispensable, de toute cette chaîne d'outils et de traitements, la tâche de l'*étiquetage* en catégories (« *Part-Of-Speech tagging* ») est devenue l'une des tâches centrales de l'informatique linguistique, et a nourri à travers le monde des centaines d'articles et de thèses.

Fortement engagée dans cette course à l'annotation en catégories, la communauté des chercheurs en linguistique formelle a eu de moins en moins le loisir intellectuel d'en critiquer les fondements. Tout au plus a-t-elle pu, avec la nécessité de décrire des constructions plus spécifiques, faire évoluer la vieille liste de huit catégories héritée de Denys le Thrace. Ainsi, le « Penn Treebank », projet international (basé à l'université de Pennsylvanie) de constitution d'un corpus « arboré », c'est-à-dire incluant une description complète de ses structures syntaxiques (dont le premier étage est bien entendu la liste des catégories de parties du discours), a-t-il institué pour ses propres besoins une liste de 36 catégories. Celles-ci restent en tout état de cause essentiellement issues d'éclatements ou de spécifications des anciennes catégories de base (on distingue ainsi le nom propre du nom commun, l'adjectif comparatif de l'adjectif au degré zéro, le pronom interrogatif, le pronom personnel, et le pronom possessif, etc.) Et elles ne remettent pas en question la définition originelle, essentialiste, de ces catégories.

⁸ MONTAGUE, R. M., « English as a Formal Language », R. THOMASON (dir.), *Selected Papers of Richard Montague*, New Haven (É.U.A.), Yale University Press, 1974, p. 188–221.

La seconde raison tient au fait que les contre-exemples ont manqué. Pendant les trente premières années de son développement, la linguistique formelle a été en très grande partie une affaire de locuteurs de langues européennes, voire même de *Standard Average European* pour importer une expression de Whorf. Le champ s'est construit sur l'analyse, en tout premier lieu, de l'anglais et du russe ; puis, dans la période 1950–1980 de l'allemand, du français, de l'italien, du tchèque... et des autres langues adossées à un état industriel. L'analyse du chinois était encore novatrice en 1980. Or on vit relativement bien avec un monde de catégories syntaxiques rigides quand on travaille sur l'anglais. L'étude des productions d'autres langues rend cette convention difficile à tenir. Nous allons l'illustrer ci-dessous à l'aide d'exemples tirés du créole martiniquais.

II.3 Le créole

Les langues créoles n'ont pas de définition génétique ou typologique : le terme est lié à un contexte historique, géographique, et social particulier, qui est celui des colonies esclavagistes européennes du XVII^e au XIX^e siècle. Les problèmes que nous allons illustrer en y puisant des exemples ne leur sont pas propres ; mais celles-ci ont l'avantage d'être assez facilement accessibles à un auditoire francophone, du fait de la proximité du stock lexical.

Les créoles sont issus d'une évolution linguistique accélérée, très spécifique, à partir du français. En un intervalle temporel d'au plus deux siècles, celle-ci conduit d'une langue possédant encore une morphologie flexionnelle fondée sur un certain nombre de catégories indicielles (genre et nombre des noms, temps et personne des verbes ...), à une langue isolante dont les unités sont difficiles à ranger dans des parties du discours étanches.

II.3.1 Descriptions traditionnelles

Je vais prendre ici mes exemples en créole martiniquais, mais une grande partie de ce qui est dit vaut aussi pour les autres langues créoles du bassin Amérique-Caraïbe (Haïti, Guadeloupe, et Guyane). On présente ici seulement quelques faits isolés ; une synthèse systématique du fonctionnement du noyau du syntagme verbal et du syntagme nominal du créole est exposée ailleurs⁹.

Les unités lexicales du créole sont invariables et autonomes. Elles ne portent pas de marques indicielles chargées de sécuriser l'interprétation en renforçant la cohésion textuelle au niveau syntaxique (comme l'accord en français), ni de marques relationnelles chargées d'assigner une fonction spécifique dans un syntagme (comme la déclinaison en latin). Les mots présentant plusieurs formes sont extrêmement rares ; les quelques exemples proviennent de formes fléchies différentes d'un mot français extrêmement fréquent, qui se sont spécialisées dans des usages distincts (ex. *prann*, *pri*, la seconde forme étant devenue un verbe à sens passif).

Cette absence de morphologie est, de façon tout à fait logique, corrélée à une rigidité forte de l'ordre des constituants d'un syntagme. Les prédicats qui attendent un actant unique présentent la construction intransitive classique dans les langues SVO : **Arg0 Pred** ; de même, les prédicats à deux actants présentent la construction qu'on appelle en français transitive directe : **Arg0 Pred**

⁹ VAILLANT, P., « Grammaires factorisées pour des dialectes apparentés », *TALN 2008*, Avignon, p. 159–168. Pour une description plus complète de la grammaire des créoles, on peut se reporter à VALDMAN, A., *Le créole : structure, statut et origine*, Paris, Klincksieck, 1978, pour le haïtien ; à BERNABÉ, J., *Précis de syntaxe créole*, Matoury, Ibis Rouge, 2003, pour le martiniquais et le guadeloupéen ; et à SAINT-JACQUES-FAUQUENOY, M., *Analyse structurale du créole guyanais*, Paris, Klincksieck (coll. « Études Linguistiques », vol. XIII), 1972, pour le guyanais.

Arg1¹⁰ ; les (rares) prédicats à trois actants présentent la construction inamovible : **Arg0 Pred Arg2 Arg1**, où le complément d'attribution précède toujours le complément d'objet, qu'il se présente sous la forme de pronom ou de groupe nominal plein. Le créole est sur ce point plus rigide que le français, qui permet l'alternance stylistique — lorsque les deux compléments sont des syntagmes nominaux pleins — entre « J'ai donné un couteau à mon père » et « j'ai donné à mon père un couteau ».

En créole, le temps, le mode, et l'aspect sont indiqués par des particules pré-verbales invariables. Au cœur du syntagme verbal, on a un noyau de trois particules spécialisées dans l'expression de ces catégories, et qui forment un système rigide : *té* exprime le passé ; *ké* exprime le futur, ou, lorsqu'il est en combinaison avec *té*, le conditionnel ou l'irréel¹¹ ; *ka*, enfin, exprime l'aspect imperfectif (duratif, fréquentatif ou progressif). Chacune de ces particules alterne avec l'absence de marque (∅), et les trois forment système dans un ordre fixe, si bien que les 8 combinaisons possibles sont attestées.

	∅	∅	∅	
Nou	té	ké	ka	dansé
	PASSÉ	PROSPECTIF	IMPERFECTIF	

D'autres adverbes, locutions adverbiales plus ou moins grammaticalisées, ou verbes modaux, expriment des variations autour de ce système noyau, pour exprimer des sens comme l'aspect inchoatif (*prann*) ou accompli (*fini*), le passé proche (*soti*) ou le futur proche (*kalé* > *kay*).

Cette rigidité du système laisse penser en première analyse qu'un formalisme grammatical fortement guidé par la syntaxe, comme les CFG, serait parfaitement adapté à la description du créole. Il n'y aurait même pas besoin de le doter d'un mécanisme de représentation des traits syntaxiques d'accord (comme on le fait, pour rendre compte des catégories morphologiques comme le genre et le nombre en français ou en allemand, avec les DCG¹²) — ou alors très marginalement, si l'on souhaite intégrer au niveau de la grammaire le phénomène de l'harmonie nasale.

Pourtant ces formalismes posent en réalité problème, en ce qu'ils reposent sur le postulat que l'on peut facilement attribuer à une unité une catégorie. Or ce postulat ne fonctionne pas en créole : beaucoup d'unités lexicales peuvent facilement se glisser dans le rôle syntaxique de plusieurs « parties du discours », selon la définition classique du terme : selon le contexte, elles jouent le rôle du nom, du verbe, ou de l'adjectif ; et ce phénomène est trop répandu pour qu'on puisse le résoudre, comme le faisait Apollonius pour l'article/pronom *ô* en grec homérique, en décrétant qu'on a affaire à deux unités homophones, l'une dérivant de l'autre par transposition (*μεταληπισς*)¹³.

II.3.2 Problèmes avec la description catégorielle

Définition d'une classe de verbes au sein du lexique

¹⁰ Il n'y a pas d'équivalent en créole du « complément d'objet indirect » : « il ressemble à son père » se dit « i sanm papa'y » ; « réfléchis à ce qui s'est passé » se dit « sonjé sa ki pasé » ; « sors de là » se dit « soti la ».

¹¹ Il n'est pas rare que la combinaison du futur et du passé, dans plusieurs langues du monde, serve à exprimer l'irréel ou l'incertain (un futur possible considéré depuis le passé ?). C'est l'origine, historiquement, de l'actuel conditionnel en *-erait* des langues romanes modernes.

¹² DCG : *Definite Clause Grammars* ; extension des CFG intégrant des unifications de traits, proposée par COLMERAUER, A., « Metamorphosis Grammars », *Natural Language Communication with Computers*, Berlin, Springer (coll. « Lecture Notes in Computer Science », vol. 63), p. 133–188.

¹³ LALLOT, J., art. cit.

Pour les premières études sur le créole (cf. par exemple Saint-Quentin¹⁴ sur le créole de Cayenne), la question des catégories morphologiques ne semble ne pas se poser : les unités lexicales sont attribuées à des parties du discours de la grammaire classique sur la base de l'évidence. Il y a, nous semble-t-il, au moins deux sources à ce sentiment d'évidence. D'une part, les mots créoles ressemblent la plupart du temps à des mots français, donc on leur attribue assez spontanément la catégorie de leur cognat. D'autre part, les auteurs, suivant en cela toute la tradition antérieure, croient savoir ce qu'est un nom, un verbe ou un adjectif sur la base des définitions sémantiques. Ils ne manquent pas de remarquer la labilité des catégories, mais dans ce cadre de pensée (comme pour Apollonius), cela ne remet pas en question la nature fondamentale d'un nom ou d'un verbe : cela oblige simplement à traiter le phénomène comme une transposition (une *translation*, dans le vocabulaire de Tesnière¹⁵) :

« En créole, comme dans certaines langues plus relevées, l'infinitif devient volontiers un substantif et en suit les règles¹⁶. Exemple [...] : *maré-tèt wom*, la manière de s'attacher la tête des hommes (littéralement : l'attacher tête des hommes)¹⁷. »

Le critère conceptuel pour définir ce qu'est un verbe ne résiste évidemment pas à l'examen critique, cela ne fait pas débat en linguistique. Si un verbe pouvait se définir comme un mot qui désigne une action, alors *destruction*, dans *la destruction de Troie par les Achéens* serait un verbe.

Les définitions contemporaines des parties du discours font passer au premier plan les critères morphologiques ou syntaxiques, qui, comme le suggérait Lallot, étaient en réalité présentes dès le départ (dès la *Poétique* d'Aristote), mais que les grammairiens classiques n'avaient pas envie d'ériger en critère classificatoire car ils leur préféraient une classification fondée sur l'idée.

Dans le cas du créole, aucun critère morphologique ne pouvant s'appliquer, on ne peut compter que sur d'éventuels critères syntaxiques (au sens restreint : positionnels) pour définir un ensemble adapté de parties du discours, si l'on souhaite construire un modèle opérationnel de la grammaire de cette langue. Cette contrainte engendre un programme de recherche empirique consistant à définir un ensemble de tests, à valider ou à invalider, qui permettent de définir des parties du discours.

Prenons un exemple fondamental de ce type de test : si le verbe peut (toujours) servir de prédicat, est-ce que la fonction de prédicat définit le verbe ?

La réponse à cette première question est non, car toutes sortes de mots peuvent assumer, en créole, la fonction de prédicat (exemples 1a à 1e).

- (1) a. mwen manjé
 1SG manger
 j'ai mangé
- b. mwen las
 1SG fatigué

¹⁴ SAINT-QUENTIN, A., « Étude sur la grammaire créole », in *Introduction à l'histoire de Cayenne, suivie d'un recueil de contes, fables et chansons en créole*, Antibes, J. Marchand, 1872. Réédité en Guyane en 1989, et disponible également en version numérisée sur le site BNF *Gallica* : <http://gallica.bnf.fr/ark:/12148/bpt6k82939m/>.

¹⁵ TESNIÈRE, L., *Éléments de syntaxe structurale*, Paris, Klincksieck, 1959.

¹⁶ L'auteur, on peut le noter au passage, utilise ici spontanément le mot *infinitif*, qui désigne en effet, dans les langues à flexion verbale, un mode du verbe lui permettant de fonctionner syntaxiquement comme un nom ; mais dans ces langues, il s'agit d'un mode morphologiquement marqué. En créole, il n'y a rien de tel, puisque l'unité lexicale est invariable (*maré* est tout autant un infinitif qu'une forme finie).

¹⁷ SAINT-QUENTIN, A., *op. cit.*, p. 130.

je suis fatigué

c. mwen doktè

1SG docteur

je suis docteur

d. mwen Kaz-Pilot

1SG Case-Pilote (toponyme)

je suis à Case-Pilote

e. mwen la

1SG DEIC

je suis là

Il est donc impossible d'utiliser le critère de la fonction prédicat pour définir une classe d'unités lexicales, à moins d'englober presque toutes les unités lexicales de la langue dans cette classe — et donc, de devoir redéfinir pratiquement chaque unité comme possédant un homophone dans une certaine classe, et, en plus, un homophone dans la classe *prédicat*.

Le créole est dans une grande mesure une langue *omniprédicative*, pour reprendre le terme forgé par Launey¹⁸ à propos du nahuatl. Launey, concédant que le nahuatl est une langue assez exceptionnelle tant elle est extrême sur ce critère typologique, note au passage, plus généralement, qu'il n'est pas nécessaire que la fonction prédicative, à travers les diverses langues du monde, se spécialise en une seule classe de mots appelée *verbes* :

«La dissociation entre fonctions et classes lexicales [...] fait du nahuatl une langue typologiquement extrémiste, et relativement isolée dans ce pôle des systèmes prédicatifs.

[...] En nahuatl, on a bien les verbes comme supports de catégories aspecto-temporelles, et les noms comme support de quantification ; mais le système fonctionne selon une autre logique prédicative, qui peut s'énoncer : *tout concept* (représenté par un item lexical, pourvu le cas échéant des catégories grammaticales afférentes) *est un prédicable*.

Envisagée sous cet angle, la prédication concerne bien l'ensemble du lexique, et le pôle typologique opposé (prédicat toujours verbal et arguments toujours nominaux) apparaît comme issu, non d'une nécessité universelle, mais d'un durcissement jusqu'à la contrainte de ce qui n'est, au niveau universel, qu'une relation privilégiée¹⁹. »

Sans que le créole ait le moindre rapport ni génétique ni typologique avec le nahuatl, on peut penser qu'il illustre cette intuition de Launey.

Après l'échec de cette première hypothèse de définition d'une classe de verbes, fondée sur la possibilité de jouer un rôle de prédicat, une deuxième hypothèse soumise à validation pourrait être : est-ce que le verbe est défini par la compatibilité avec la marque de temps *té*? Là encore, cette hypothèse de critère différentiel se révèle infructueuse, car l'ensemble des énoncés 2a à 2e sont également recevables.

(2) a. mwen té manjé

¹⁸ LAUNEY, M., *Une grammaire omniprédicative : essai sur la morphosyntaxe du nahuatl classique*, Paris, Éditions du CNRS, 1994.

¹⁹ *Ibid.*, p. 59–60.

1SG PST manger
j'avais mangé

b. mwen té las
1SG PST FATIGUÉ
j'étais fatigué

c. mwen té doktè
1SG PST docteur
j'étais docteur

d. mwen té Kaz-Pilot
1SG PST Case-Pilote (toponyme)
j'étais à Case-Pilote

e. mwen té la
1SG PST DEIC
j'étais là

Il semble donc qu'une grande partie des unités lexicales du créole, non seulement puisse occuper la fonction de prédicat, mais puisse également porter des marques temporelles.

Délimitation d'une classe de verbes processifs

L'obstacle qui apparaît lorsque l'on tente de délimiter une catégorie de verbes surgit également si l'on tente de déplacer le front, en s'essayant à la délimitation nette d'une catégorie de verbes de procès, en laissant de côté les autres types de prédicats.

Toutes les grammaires du créole que j'ai pu consulter posent une distinction entre verbes statifs et verbes processifs, sous ce terme ou sous un autre. Le critère sémantique utilisé pour définir ces deux classes est reflété par les éléments de métalangue (verbe exprimant un état *vs.* verbe exprimant un processus); quant au critère syntaxique, il est invariablement celui de la combinaison possible avec la particule d'aspect *ka*. Selon les auteurs cités plus haut, les verbes statifs (comme *ni*, avoir; *kwè*, croire; *sonjé*, penser; *rété*, rester...), n'exprimant pas un procès que l'on puisse observer dans son déroulement, neutralisent la catégorie de l'aspect: ils ne sont donc pas compatibles avec *ka*. Ces descriptions posent ainsi comme principe que la phrase 3a est recevable, alors que les phrases 3b, 3c ne le sont pas.

(3) a. mwen ka manjé nyanm
1SG IPFV manger igname
je suis en train de manger de l'igname

b. *mwen ka ni lajan
1SG IPFV avoir argent
je suis en train d'avoir de l'argent

c. *mwen ka malad
1SG IPFV malade
je suis en train d'être malade

Si ce critère était vérifié, il fournirait en effet un moyen d'isoler une classe de verbes processifs, qui s'opposeraient en tant que tels à tous les autres prédicats — verbes statifs ou adjectifs. Mais il se révèle que ce critère n'est pas vérifié, le jugement de non-recevabilité porté sur 3b ou 3c étant infirmable dans certains contextes. Ainsi, dans l'énoncé 4, l'unité *ni* (avoir), réputée incompatible

avec la particule aspectuelle *ka*, entre dans une construction parfaitement acceptable avec cette particule, et prend un sens fréquentatif²⁰ :

- (4) chak simèn i ka ni lajan
 chaque semaine 3SG IPFV avoir argent
 chaque semaine, il a de l'argent

Une autre construction où *ni* est associé à *ka* est celle que l'on trouve dans l'exemple 5. Il s'agit d'un proverbe antillais dont la teneur générale correspond en gros à *la raison du plus fort est toujours la meilleure*. Dans cet énoncé, la construction avec *ka* sert à exprimer une vérité générale ou intemporelle (l'aspect général y est renforcé, en isotopie, par l'absence d'article déterminant les noms — procédé qui, comme en français classique, marque le degré de détermination générique).

- (5) ravèt pa ka ni rézon douvan poul
 ravet NEG IPFV avoir raison devant poule
 le ravet (cafard) n'a [jamais] raison devant la poule

Les exemples 4 et 5 témoignent, semble-t-il, que le jugement d'agrammaticalité traditionnellement porté dans les grammaires sur 3b résulte peut-être d'une imagination trop restreinte de la valeur que peut avoir l'imperfectif, lorsqu'une telle phrase minimaliste est considérée sans contexte. On tend, dans ce cas, à projeter sur la phrase la valeur la plus répandue de l'imperfectif, à savoir la valeur progressive, qui marque le déroulement d'un procès ; c'est oublier que l'imperfectif peut avoir d'autres valeurs, comme le fréquentatif ou le duratif, qui, elles, sont compatibles avec un verbe comme *ni* (et sont, sans surprise, les valeurs sélectionnées lorsque la particule *ka* est accolée à *ni*).

Autres tentatives de délimitation de catégories

D'autres tentatives d'utiliser des critères distributionnels pour délimiter, en créole, des parties du discours, butent sur le même obstacle : on trouve toujours des contre-exemples attestés. Nous ne pouvons développer ici le sujet faute de place, mais la question de la délimitation, dans les langues créoles, d'une classe d'adjectifs distincte de celle des verbes, a occupé sans succès des générations de linguistes. Tout au plus Damoiseau et Saint-Louis²¹ ont-ils pu affirmer qu'en créole haïtien, sur 440 unités lexicales étudiées comme candidates au statut d'adjectif, seules 8 sont spécialisées dans la fonction épithète (et ne jouent jamais le rôle du prédicat) ; parmi les autres, 232 sont même compatibles avec la marque d'aspect imperfectif, supposée réservée aux verbes de procès.

Quant à la partie du discours que l'on pourrait supposer la plus étrangère à l'expression du procès, le nom, elle-même refuse de se laisser cerner par des caractérisations morphosyntaxiques en créole : ainsi, des phrases attestent la combinaison d'un nom avec une marque d'aspect imperfectif (6).

²⁰ La plupart des exemples qui suivent sont des exemples construits collectivement « en laboratoire », au sein d'un groupe d'universitaires ou d'étudiants. Je les garde dans l'exposé pour leur vertu de concision, mais j'ai pris le soin de vérifier que des exemples comparables sont attestés dans mes corpus transcrits. Par exemple : *lè ou fè konsa ou pa ka ni piès problèm de traduction* (quand vous faites comme cela, vous n'avez pas le moindre problème de traduction), attesté dans le corpus de Lengrai, Moustin et Vaillant, à la même valeur fréquentative que l'exemple 4 (LENGRAI, Ch., MOUSTIN, J., VAILLANT, P., « Interférences syntaxiques et lexicales du français dans le créole martiniquais des émissions radiophoniques », 3. *Freiburger Arbeitstagung zur Romanistischen Korpuslinguistik : Korpora und Pragmatik*, Freiburg in Breisgau, 2006).

²¹ DAMOISEAU, R., SAINT-LOUIS, G., « Note sur les "adjectifs" dans quelques créoles français », *Langages*, vol. 34, n° 138, juin 2000, p. 61–69.

- (6) i ka timanmay
 3SG IPFV petit garçon
 il fait l'enfant / il se comporte comme un petit garçon

La question qui se pose à ce stade est donc la suivante : si la notion de partie du discours est si pénible à caractériser, et si toute tentative de règle définitoire engendre automatiquement une foule d'exceptions, alors cette catégorie est-elle vraiment indispensable ? Et comment peut-on expliquer, sans elle, les phénomènes syntaxiques qui, en créole comme dans toutes les langues, permettent d'identifier les fonctions des unités au sein des syntagmes ?

III Propositions

Ce que les exemples donnés plus haut illustrent à notre avis de manière frappante, c'est la pertinence de la notion d'afférence pour rendre compte de phénomènes qui sont généralement considérés comme relevant du domaine de la syntaxe.

L'afférence est un concept développé par Rastier²² pour rendre compte du sens dynamique en contexte : c'est le mécanisme élémentaire de construction de la cohérence textuelle. Une unité lexicale quelconque peut recevoir de son contexte un sème qui n'est pas nécessairement présent dans la liste de ses sèmes inhérents (signification stabilisée, « en langue ») ; réciproquement, un sème inhérent peut être partiellement ou totalement virtualisé par son contexte.

III.1 Afférence de sèmes

Reprenons pour illustrer le concept l'exemple de la phrase « le commissaire aboie », comparée à la phrase « le chien aboie ». Le premier énoncé (7a) est *isotope*, c'est-à-dire qu'il associe deux unités possédant un même sème inhérent, renforçant ainsi la cohérence de l'ensemble. Cet énoncé est compréhensible sans effort d'interprétation, et n'apporte pas énormément de nouveauté (informative ni stylistique). Le second (7b), en revanche, est *allotope* : il associe deux unités possédant des sèmes inhérents différents, voire contradictoires (/humain/ d'un côté, /chien/ de l'autre). Il est pourtant parfaitement interprétable par un locuteur du français : il porte l'idée que le commissaire ne parle pas de façon agréable.

- | | | | | | | | | | | |
|-------|----|---------|---|------------|--|----|----|-------------|---|------------------------|
| (7)a. | le | chien | | aboie | | b. | le | commissaire | | aboie |
| | | /chien/ | ↔ | /∧chien/ | | | | /chien/ | ← | /∧chien/ ²³ |
| | | | | /bruit/ | | | | /humain/ | → | /∧humain/ |
| | | | | /agressif/ | | | | | | /bruit/ |
| | | | | | | | | | | /agressif/ |

Le mécanisme de l'afférence est convoqué ici pour expliquer la cohérence de cet énoncé allotope (c'est-à-dire au fond ni plus ni moins que la possibilité pour la langue de créer du sens) : par la proximité créée par le contexte, le sème /chien/, présent dans le verbe « aboyer », s'active dans le sens dynamique du nom « commissaire » (le commissaire n'est plus si humain que ça, et devient « un peu chien ») ; tandis que symétriquement le sème /humain/ présent dans le nom « commissaire » vient atténuer la littéralité de l'aboiement.

²² RASTIER, F., *op. cit.*

²³ Nous utilisons les symboles ∧ et ↗, empruntés aux notations formelles des grammaires catégorielles de dépendance, pour indiquer des sèmes associatifs (qui se propagent du prédicat vers l'argument). La mise entre parenthèses d'un sème indique sa virtualisation ; les flèches horizontales indiquent l'afférence, et la flèche horizontale double une isotopie.

Il nous semble qu'il n'y a pas de différence fondamentale de mécanisme mis en jeu entre l'exemple « le commissaire aboie » (7b) et l'exemple « i ka timanmay » (6). Dans ce dernier, il est possible de considérer que la mise en contexte de *timanmay* dans un syntagme verbal contenant le marqueur d'aspect progressif *ka* conduit à l'activation, par afférence, du trait *progressif* dans ce lexème. Ce phénomène conduit automatiquement le lecteur à produire, sous contrainte, une interprétation du mot qui se rapproche plus de « faire l'enfant » que de « être un enfant » (10).

(8)	i		ka		→	timanmay
			/processif/			/processif/
			/progressif/			(/nominal/)
						/humain/
						/enfant/
						/mâle/

De la même manière que nous contrastions l'énoncé isotope « le chien aboie » et l'énoncé allotope « le commissaire aboie », nous pouvons contraster l'énoncé isotope « mwen ka dansé » et l'énoncé allotope « mwen ka malad » (11a ; 11b). Dans « mwen ka dansé », le sème /processif/ est inhérent à la fois dans *ka* et dans *dansé* : il y a donc une isotopie de cette catégorie syntaxique qui — comme dans l'énoncé « le chien aboie » — permet de renforcer la cohérence globale. Dans « mwen ka malad », en revanche, on peut concevoir qu'un phénomène d'afférence — analogue à celui qui se déroule dans « le commissaire aboie » — permet d'activer le trait /processif/ dans *malad*, qui ne le possède pas de manière inhérente.

(11)	a.	mwen	ka		dansé	b.	mwen	ka	malad
		/processif/	↔	/processif/			/processif/	←	/processif/
		/progressif/		/mouvement/			/progressif/	→	(/statif/)
				/corps/					/corps/
				/plaisir/					/malaise/

L'interprétation s'oriente alors vers un sens contextuel plus proche de « tomber malade » que de « être malade ». À l'opposé, le trait /statif/, *a priori* inhérent dans *malad*, se trouve dans ce contexte virtualisé.

Le même principe descriptif peut être utilisé pour rendre compte de « i ka timanmay » (10) : c'est ici pas moins que le caractère nominal du lexème qui se retrouve virtualisé, tandis que le caractère verbal (processif) est activé par afférence.

Ces réflexions posent plus généralement la question de la frontière conventionnellement établie entre syntaxe et sémantique. Dans les deux cas, me semble-t-il, les mêmes mécanismes fondamentaux sont à l'œuvre. L'isotopie crée une cohérence, elle établit les « îlots de confiance » de l'interprétation (les phénomènes d'accord, et plus généralement ce que le technolecte de la typologie linguistique appelle les *mécanismes indiciels*, ne sont ni plus ni moins que des isotopies de catégories syntaxiques) ; et l'allotopie, avec les phénomènes dynamiques possibles qu'elle peut déclencher (afférence ou virtualisation), permet de créer de nouveaux sens en contexte.

Alors, si les mêmes mécanismes sont à l'œuvre, y a-t-il une différence de nature qui justifie de tracer une frontière entre syntaxe et sémantique ? La subdivision établie classiquement se fonde souvent sur une opposition entre catégories « pleines » et catégories « vides » (ou désémantisées). Sans chercher à prétendre que cette notion est dépourvue d'intérêt, force est de constater qu'elle est éminemment graduelle. Certaines catégories présentées comme sémantiques, comme l'opposition animé/inanimé, sont fortement codifiées en langue (c'est par exemple cette opposition qui fonde la différence entre *qui* et *quoi* en français). À l'opposé, si certaines catégories grammaticales semblent

totallement dépourvues de signification (comme la première et la deuxième conjugaison des verbes français), d'autres sont indexées sur des dimensions sémantiques (comme les cas en latin, ou les classes nominales dans les langues bantoues) ; et certaines, aussi grammaticalisées qu'elles puissent paraître, peuvent être réinvesties de signification dans certaines situations.

En somme, il n'y a pas de frontière entre syntaxe et sémantique, mais un gradient de figement dans la langue, qui conduit des catégories sémiqes ayant les contraintes les plus fortes et les inventaires les plus fermés, à celles qui se combinent le plus librement et qui ont les inventaires les plus ouverts. Cette conception a entre autres pour conséquence, comme nous l'avons vu illustré avec le français et le créole, que ce qui est syntaxique dans une langue peut ne pas l'être dans une autre.

III.2 Implication théorique

Si l'on admet la possibilité de modéliser les phénomènes de construction de syntagmes de la manière qui vient d'être exposée plus haut, alors l'implication théorique en est qu'il faut faire exploser la notion de partie du discours (conçue comme ensemble contenant des familles de mots homogènes) pour la remplacer par un faisceau de propriétés élémentaires (dont chacun réponde à une question précise sur les comportements syntaxiques prévisibles de chaque unité lexicale, et soit éventuellement actualisable ou virtualisable en contexte). Ainsi, au lieu de se demander si un lexème donné est un verbe ou un adjectif (et d'épuiser un siècle de recherche sans parvenir à répondre à cette question), les bonnes questions deviennent celles de savoir si ce lexème peut avoir une fonction de prédicat dans la phrase, s'il peut être épithète dans un groupe nominal, s'il se combine ou non avec une particule de temps, avec un déterminant, ou avec un pronom relatif... En somme, il s'agit de définir des catégories grammaticales qui sont contenues dans les unités de la langue, et non plus celles qui les contiennent.

En soi, cette idée n'est pas terriblement neuve : Chomsky l'a proposée en 1970, et Gazdar, Klein, Pullum et Sag en ont fait un modèle cohérent en 1985²⁴.

Selon Chomsky, « there is no reason to retain the notion of category at all, even for the base. We might just as well eliminate the distinction of feature and category, and regard all symbols in the grammar as sets of features²⁵ ». Plus loin, il suggère la possibilité de définir les catégories grammaticales majeures de l'anglais comme des croisements de deux traits distinctifs binaires N et V, définis comme :

[±N]	having nominal properties	(exists in space, set-denoting)
[±V]	having verbal properties	(exists in time, situation-denoting)

Dans cette combinatoire de deux axes binaires, on peut définir les quatre catégories considérées comme majeures selon le schéma suivant :

Noms :	<N> = +	<V> = -
Verbes :	<N> = -	<V> = +
Prépositions :	<N> = -	<V> = -
Adjectifs :	<N> = +	<V> = +

²⁴ GAZDAR, G., KLEIN, E., PULLUM, G., SAG, I., *Generalized Phrase Structure Grammar*, Oxford (R.U.), Blackwell, 1985.

²⁵ CHOMSKY, N., « Remarks on nominalization », R.A. JACOBS et P.S. ROSEBAUM (dir.), *Readings in English Transformational Grammar*, Waltham (É.U.A.), Ginn, 1970, p. 184–221 : p. 208.

Cette idée faisait éclater la notion classique de partie du discours pour la remplacer par une combinatoire de traits ; mais elle n'allait pas jusqu'au bout de sa logique puisque, ne contenant aucun mécanisme dynamique, elle condamnait chaque unité lexicale à être toujours la même chose (à être certes {<N>=+, <V>=+} au lieu de **Adj**, mais néanmoins toujours la même chose). La linguistique formelle ne s'est donc pas vue encouragée à abandonner les étiquettes de parties du discours, d'autant que, comme nous l'avons dit plus haut, le travail sur l'anglais ou le français s'en accomode généralement assez bien.

Dans un travail de modélisation portant sur le créole, en revanche, il s'agit de prendre au sérieux ce programme de redéfinition « par le bas » des catégories sous forme de faisceaux de traits, en faisant l'inventaire des traits qui ont une pertinence dans les contraintes de distribution des unités, et en les incorporant dans la définition des lexèmes ; par exemple :

1. le mot *moun* (personne) a une fonction nominale parce qu'il forme la plupart du temps un syntagme avec des déterminants (“an moun”, “moun-lan”, “sé moun-lan”, “sé moun-tala”, “tout moun” ...) → il doit avoir le trait inhérent /nominal/ ;
2. le mot *malad* peut avoir la fonction d'épithète parce qu'il peut s'adjoindre à un nom dans un GN (“an moun malad”, “sé moun malad-la” ...) → il doit avoir le trait inhérent /épithète/ ;
3. le mot *dansé* peut dans beaucoup de contextes se combiner avec la marque aspectuelle de l'imperfectif/progressif *ka* → il doit avoir le trait inhérent /processif/ ;
ce qui n'empêche pas que :
4. le mot *malad* n'est pas toujours épithète, il peut avoir une fonction de prédicat (*mwen malad*) → le trait inhérent /épithète/ peut être virtualisé ;
5. le mot *dansé* peut aussi dans certains contextes se combiner avec des déterminants et des épithètes (*sé an bel dansé*) → le trait /nominal/ peut être afférent.

Ce modèle reste en principe opérationnalisable et calculatoirement borné.

En réalité, la seule chose qui puisse, dans cette manière de concevoir les catégories, être gênant pour les habitudes de pensée de la syntaxe formelle, c'est le fait qu'elle se débarrasse de la notion de nature, l'un des deux pôles de la distinction nature/fonction. Dans ce modèle des faits syntaxiques, la fonction reste une chose bien définie (comme l'ensemble des traits activés en contexte), mais la nature se retrouve réduite à une notion faible (à savoir un ensemble de traits inhérents, qui ne sont même pas obligatoires).

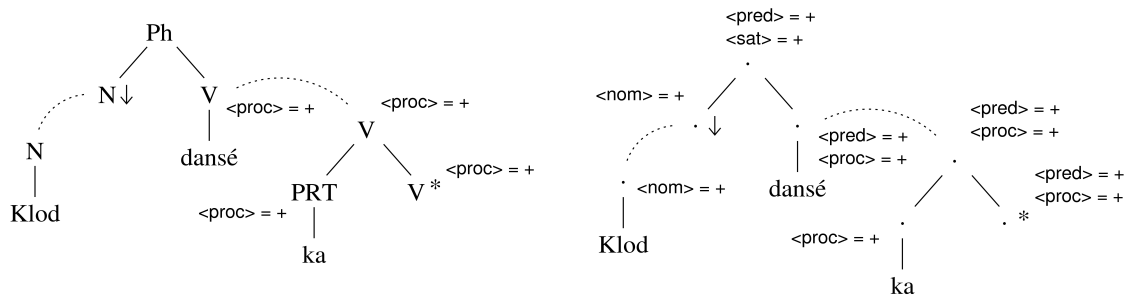
III.3 Possibilités de modélisation

La proposition faite ci-dessus de traiter les faits de syntaxe avec l'outillage de la sémantique interprétative est implémentable grâce aux formalismes de grammaire reposant sur le principe de l'unification de structures de traits (comme FS-TAG²⁶ ou HPSG²⁷). Elle impose simplement une exigence nouvelle : celle de se débarrasser complètement des étiquettes correspondant à la nature des syntagmes représentés par les nœuds de l'arbre, puisque l'on souhaite partir du principe que ces étiquettes ne sont plus définies *a priori*.

²⁶ VIJAY-SHANKER, K., JOSHI, A., « Feature-structure based Tree Adjoining Grammars », *COLING 1988*, Budapest (Hongrie), p. 714–719.

²⁷ POLLARD, C., SAG, I., *Head-driven Phrase Structure Grammar*, Chicago (É.U.A.), University of Chicago Press, 1994.

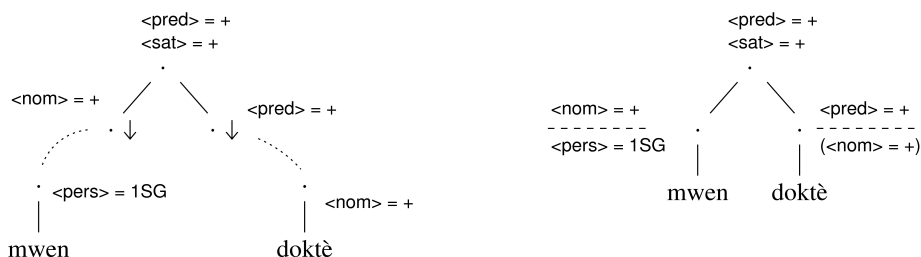
Cette exigence ne modifie en réalité pas en profondeur les mécanismes de base du modèle FS-TAG : au lieu d'unifier deux arbres — au cours d'une opération de substitution par exemple — en prenant en compte la nature du nœud *et* les structures de traits qui lui sont rattachées, on unifie deux arbres en ne prenant en compte que les structures de traits (quitte à ce que la même information soit présente, mais descendue du « piédestal » de l'étiquette du nœud). La figure présentée ci-dessous (ill. 2) illustre ce principe avec trois arbres élémentaires TAG, combinables pour former une phrase simple comme *Klod ka dansé* (Claude danse).



Ill. 1 : Combinaison d'arbres élémentaires dans le formalisme FS-TAG pour obtenir la phrase *Klod ka dansé*. À gauche : avec étiquetage des catégories syntaxiques ; à droite : sans étiquetage.

Grâce au mécanisme de FS-TAG, un sème inhérent peut être représenté comme un trait aval (trait qui « remonte » du lexique vers les syntagmes de plus haut niveau), et un sème afférent comme un trait amont (trait présent dans le cadre du syntagme, et qui s'impose par le haut sur les nœuds qu'il domine).

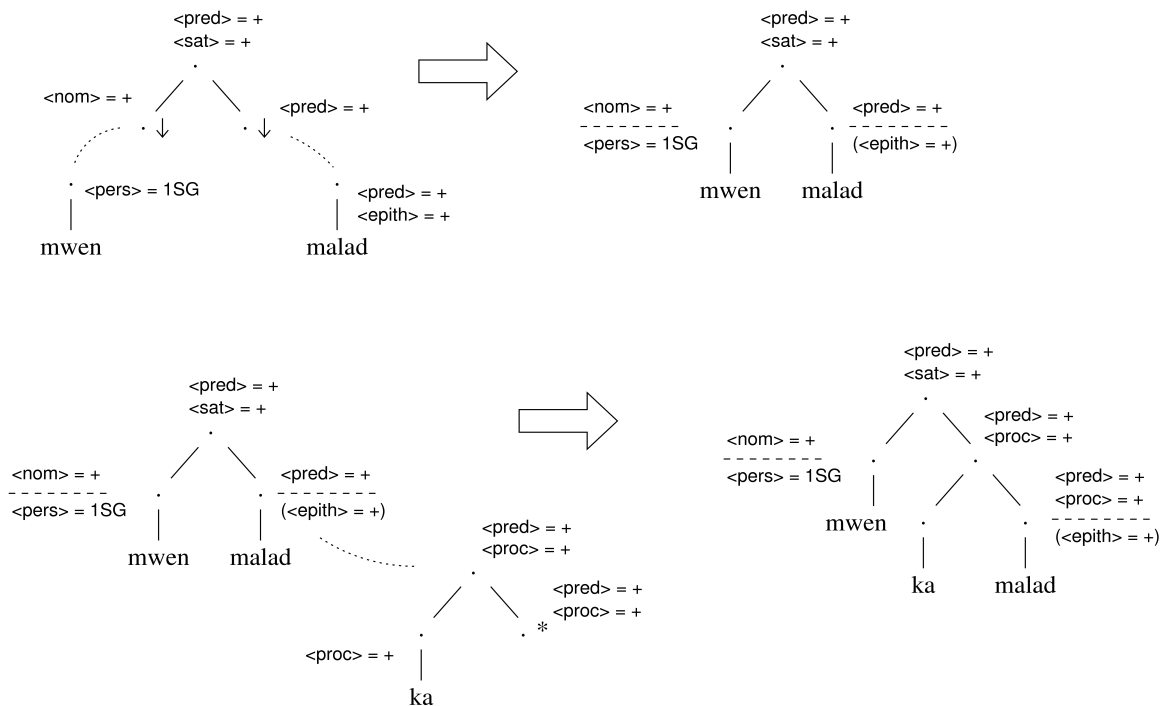
- Dans les cas d'énoncés isotopes, l'unification valide le fait que les valeurs respectives de deux traits de même nature sont compatibles, et la clôture de l'arbre par unification des traits amont et aval ne présente pas d'obstacle (c'est la situation représentée dans l'ill. 1).
- Dans le cas où un trait n'est pas présent dans l'une des unités, mais où il y est activé par afférence, on est dans le cas d'un trait « amont » qui se propage en mode descendant ; la clôture, en unifiant traits amont et aval, valide l'activation de ce trait au niveau de l'unité (ou du syntagme) situé en aval (ex. ill. 2 — noter la propagation du trait /pred/).
- Dans le cas où un trait est présent de manière inhérente dans l'une des unités, mais où il est virtualisé par le contexte, on est dans le cas d'un trait « aval » qui ne bloque pas l'unification, mais n'est pas propagé vers l'amont. Il reste donc à l'état de potentialité à l'échelle de l'unité lexicale qui le porte, mais n'est pas mobilisé pour contribuer au sens du syntagme global (ex. ill. 2 — noter la non-propagation du trait /nom/).



Ill. 2 : Virtualisation modélisée par le formalisme FS-TAG dans l'exemple de la phrase *mwen doktè* (« je suis docteur »). À gauche : arbres élémentaires ; à droite : après unification et clôture.

L'un des problèmes encore ouverts concerne le traitement des allotopies contradictoires. Comme l'illustrent les exemples des ill. 2 et 3, l'allotopie « simple » (non-contradictoire, c'est-à-dire avec des attributs qui sont différents) est résolue automatiquement, dans le mécanisme

d'unification de structures de traits, en ce que *tous* les sèmes sont conservés au niveau de l'unité concernée. Certains se propagent aux niveaux supérieurs, alors que d'autres non ; mais même ceux qui ne se propagent pas restent présents en tant que potentialité dans l'unité lexicale concernée (comme le caractère nominal de *doktè* dans *mwen doktè*, ou la possibilité de jouer un rôle d'épithète pour *malad* dans *mwen ka malad*).



Ill. 3 : Afférence et virtualisation de traits inhérents représentés en FS-TAG avec l'exemple de l'énoncé allotope *mwen ka malad* (3c).

En revanche ce phénomène n'est pas possible pour l'allotopie contradictoire (même attribut, valeurs opposées, par exemple <pred> = + vs. <pred> = -).

Le traitement par défaut de ce type d'allotopie, dans le mécanisme FS-TAG, est qu'il rend l'unification impossible : il définit donc les quelques cas d'énoncés réellement non-recevables. On peut considérer que ce traitement est légitime (il définit en tout cas l'un des rôles possibles du mécanisme syntaxique).

Deux autres traitements possibles sont imaginables pour gérer les cas d'allotopie contradictoires :

- la neutralisation : l'un des deux traits incompatibles est éliminé de la structure résultante, par un mécanisme de priorité (le trait amont écrase le trait aval) ; cette possibilité revient à traiter l'allotopie contradictoire comme l'allotopie simple, et donc à ignorer les impossibilités combinatoires inscrites dans le lexique ;
- la prise en compte d'un score de compatibilité : les deux traits contradictoires resteraient dans ce cas présents, mais une sorte de « pénalité de dissonance » serait calculée.

En faisant la somme des dissonances sur un syntagme, on obtiendrait une mesure de son « degré d'allotopie ». On peut convenir par exemple qu'un score de -1 correspond à une phrase totalement inacceptable, et qu'un score de 0 correspond à une phrase acceptable.

L'inacceptabilité, entre les deux, serait une affaire de degré. Il ne semble pas nécessaire *a priori*, si l'on adopte ce traitement, de postuler qu'il faut définir un seuil d'acceptabilité : on peut admettre qu'il y ait des énoncés « plus ou moins » surprenants.

Il faut noter également que la proposition de traitement exposée ci-dessus — et c'est d'un certain point de vue un inconvénient — ne permet pas une lexicalisation totale de la grammaire. En effet, si un verbe peut potentiellement être sujet, et qu'un nom peut potentiellement être prédicat, alors seule la place dans la structure de la phrase va définir ce qui est sujet (la première position), et ce qui est prédicat (la deuxième position). Il faut donc accepter que certains traits soient entièrement imposés par un cadre syntaxique non ancré dans le lexique, c'est-à-dire que notre grammaire contienne des arbres élémentaires non-lexicalisés²⁸.

IV Perspectives

IV.1 Intérêt de l'approche

L'approche que nous proposons est l'un des moyens de traiter le problème de la plasticité des catégories grammaticales en créole ; un traitement par des grammaires fondées sur des catégories échoue à rendre compte d'une partie des phénomènes observés en corpus, à moins de construire une longue série de règles permettant de gérer les exceptions, ou d'accepter une très grande indétermination dans les résultats obtenus (ce qui serait une façon d'accuser réception, dans le domaine de l'étude linguistique, du préjugé courant selon lequel « le créole n'a pas de grammaire »).

En outre, cette approche offre un traitement unifié de la syntaxe et de la sémantique qui rend partiellement caduques (tout au moins au niveau micro-sémantique) les questions d'« intégration syntaxe-sémantique » engendrées par les modèles fondés sur une séparation de ces deux composantes.

Un intérêt accessoire de cette approche est qu'elle permet, en n'offrant plus une conception binaire de la recevabilité des phrases, de réduire les échecs d'analyse ou le nombre de cas non-traités : à la place des phrases non-traitées car incomplètes ou non-conformes au modèle grammatical, elle produit comme résultat des phrases « plus ou moins » conformes aux phrases canoniques, assorties d'un score d'acceptabilité (qui fluctue entre 0 et 1). Ceci n'est pas un sujet sans intérêt pour le traitement automatique des langues, puisque des travaux récents au LPL (thèse de Prost²⁹) cherchent à développer un mécanisme de ce type, fondé sur des grammaires de contraintes.

IV.2 Faisabilité

Les mécanismes proposés ici reposent sur un modèle algébrique de traitement de la langue (le formalisme FS-TAG) qui est éprouvé, bien reçu par la communauté, et dont les propriétés

²⁸ Les grammaires lexicalisées sont considérées, dans le domaine du traitement automatique des langues, comme supérieures, car elles offrent un avantage du point de vue calculatoire lors des processus d'analyse automatique de texte. En effet, chacune de leurs règles étant ancrée dans le lexique, elles rendent impossible les phénomènes de boucle récursive piégeante, et bornent le temps d'analyse de chaque chaîne puisque les éléments du lexique pré-sélectionnent toujours un sous-ensemble restreint de structures grammaticales à appliquer. Elles ne sont cependant pas équipées pour représenter les phénomènes positionnels, dans lesquels le sens est imposé par un carcan syntaxique ; et c'est un point que notre problématique concernant les langues à catégories non-étanches met à jour.

²⁹ PROST, J.-P., *Modélisation de la gradience syntaxique par analyse relâchée à base de contraintes*, thèse de doctorat, Université de Provence (Laboratoire Parole et Langage) / Macquarie University, 2008.

« computationnelles » sont bien connues (complexité d'analyse polynomiale notamment). Comme bien souvent, la véritable difficulté si l'on veut mettre en œuvre cette approche réside dans la quantité de ressources nécessaire pour coder le lexique et la grammaire.

En outre, plus on veut sortir du noyau dur de la syntaxe et intégrer des couches de micro-sémantique « proches du noyau », plus il est prévisible que la complexité combinatoire du travail nécessaire pour énumérer les sèmes inhérents augmente. La mise en œuvre de mécanismes automatiques d'apprentissage des traits (qui joueraient, dans ce modèle, le rôle que jouent les étiqueteurs morpho-syntaxiques dans les modèles fondés sur des parties du discours) est envisageable, mais elle ne fait pas disparaître la contrainte des ressources d'expert nécessaires — elle les décale : en effet l'apprentissage automatique repose sur de grandes quantités de corpus transcrit et annoté.

IV.3 Extensibilité

L'idée qui est exposée ici est conçue pour résoudre un problème³⁰ posé par les langues créoles, celui de la plasticité des unités lexicales par rapport aux catégories syntaxiques. D'autres langues ont de ce point de vue les mêmes propriétés que le créole : c'est souvent le cas des langues isolantes en général, et c'est notamment un problème bien connu pour le chinois. Le phénomène peut être illustré par un exemple donné par Paris³¹ : (12), (13).

(12) 沒 有 什 麼 用
méi yǒu shénme yòng
NEG avoir quoi utilisation
cela n'a pas d'utilité

(13) 計 算 器 他 會 用
jìsuànqì tā huì yòng
calculatrice 3SG savoir utilisation
la calculatrice, il sait l'utiliser

Paris rappelle également que ce phénomène avait retenu l'attention, notamment, de Humboldt et de Maspéro.

Les langues non-isolantes sont réputées ne pas être exposées à ce type d'incertitude, puisqu'elles marquent les catégories morphologiques directement sur les unités lexicales, et que la faculté de recevoir tel ou tel type de marque définit les parties du discours de manière non-ambiguë. Cependant, l'argument de la morphologie n'est bien souvent qu'une manière élégante de cacher les difficultés de la grammaire sous le tapis de la frontière du mot. Dans des langues à morphologie riche comme certaines langues amérindiennes, on peut observer des phénomènes d'indétermination catégorielle analogues à ceux que nous avons exposés pour le créole, mais qui se déroulent au

³⁰ Ce problème n'en est un, bien entendu, que dans le cadre imposé par les grammaires fondées sur les parties du discours.

³¹ PARIS, M.-C., « Nom et verbe en chinois : ressemblances et différences », *Modèles linguistiques*, vol. 6, fasc. 1, 1984, p. 110–115, p. 104.

niveau de la compatibilité avec des marqueurs clitiques au sein du mot, au lieu de se dérouler au niveau de la compatibilité avec les particules isolées au sein du syntagme.

Ainsi, une phrase extraite d'un corpus de purépecha³² recueilli et annoté par Claudine Chamoreau dans le cadre du projet CLAPOTY³³ contient des mots-phrases comme 14, dans lesquels une première série d'affixes nominaux (diminutif, pluriel) s'appliquent à une unité racine *wichu* (chien), suivis immédiatement d'une série d'affixes verbaux (prédication, aoriste, assertif, indice sujet de nombre et de personne, évidentiel). Ce « mot » est-il un nom ou un verbe ?

- (14) *wichu -itu -echa -e -s -ti =ksĩ =na*
 chien DIM PL PRED AOR ASS3 3PL EV
On dit qu'ils (que ce) sont les petits chiens

En étudiant une autre langue du Mexique, le nahuatl classique, Launey a mis en évidence la nature fondamentalement *omniprédicative* de cette langue :

« Dans une langue où le sujet et plus généralement les arguments ne sont pas nécessairement représentés par des groupes syntagmatiques, la première propriété du prédicat est de pouvoir être le seul constituant de la phrase. [...] Une deuxième propriété du mot-prédicat, plus propre au nahuatl mais certainement vérifiable ailleurs, est le fait qu'il peut être précédé d'une ou plusieurs particules modales [...] On peut alors, à titre d'expérience, voir comment le lexique réagit à ces critères de prédicativité. La réponse est catégorique : à ce test, *la quasi-totalité du lexique réagit positivement*³⁴. »

Launey montre qu'il est possible de trouver en nahuatl des exemples symétriques de paires de phrases prédicatives également valables et structurellement isomorphes A-B / B-A, où chaque unité joue tantôt le rôle de sujet et tantôt celui de prédicat (15a, 15b ; note : le nahuatl est une langue à verbe initial).

- (15) a. *chōca in piltōntli*
 pleurer DET enfant
l'enfant pleure
- b. *piltōntli in chōca*
 enfant DET pleurer
c'est l'enfant qui pleure (c'est l'enfant, celui qui pleure)

Dans les exemples 15a et 15b, le morphème *in* joue tantôt le rôle d'un déterminant nominal (*in piltōntli*, l'enfant), tantôt celui d'un relatif permettant de subordonner un syntagme verbal (*in chōca*, celui qui pleure). Dans une perspective omniprédicative, on peut gloser le morphème *in* par « ce qui » ou « celui qui », qui s'applique à toutes les unités lexicales prédicatives pour les transformer en groupe nominal déterminé : dans cette façon de voir les choses, *chōca* correspond à « pleurer », et *piltōntli* à « être un enfant » ; ainsi, *in chōca* est « celui qui pleure », et *in piltōntli* est « celui qui est un enfant » (donc : « l'enfant »).

La question de l'existence de catégories de parties du discours plus ou moins étanches nous semble donc se poser au-delà des frontières de certaines catégories typologiques (langues SVO vs.

³² Purépecha : langue amérindienne du Nord-Ouest du Mexique, anciennement appelée « tarasque ».

³³ Projet CLAPOTY : *Contacts de Langues, Analyses Plurifactorielles assistées par Ordinateur, et conséquences Typologiques* : <http://clapoty.vjf.cnrs.fr>.

³⁴ LAUNEY, M., *op. cit.*, p. 50.

langues VSO, langue à affixes *vs.* langues isolantes, etc.) puisqu'elle concerne, on l'a vu, des langues aussi différentes que le créole, le chinois ou le nahuatl.

Par ailleurs, même dans les langues (comme le français) où ces catégories sont assez étanches, les phénomènes de labilité, bien que marginaux, existent. Ils sont traités comme des phénomènes exceptionnels (la *translation* de Tesnière étant une version moderne de la μεταληψις d'Apollonius), mais un cadre permettant de les analyser comme des phénomènes d'afférence « normaux », faisant partie de la vie ordinaire de la parole, peut offrir une perspective alternative.