

La phrase verbale noyau en français

1. INTRODUCTION

La phrase verbale noyau en français écrit est ici définie, en première approximation, comme la chaîne des mots qui peut commencer, à gauche, par la particule de négation *ne*, par un pronom clitique non nominatif, par un auxiliaire ou par un verbe et qui finit à droite par un verbe, qui est le noyau de la phrase verbale, ou par une parmi les formes qui peut se trouver aussi à gauche du noyau. Notés en italiques, des exemples de phrases verbales noyau en français suivent :

La fille *parle*.
La fille *parle-t-elle*.
La fille *ne parle-t-elle pas*.
Ne la regarde pas.
Regarde-la.
La fille *avait regardé* le match.
La formule *ne lui avait pas été donnée*.
La formule *ne lui avait-elle pas été donnée*.
Jacques *est* astucieux et Ferdinand *l'est* aussi.
La fille *ne lui avait-elle pas donné* le livre?
Les fidèles *se sont agenouillés*.
La fille *s'est donné* une bonne excuse.
Les excuses que la fille *s'est données*.
Les excuses que la fille *s'est souvent données*.

L'expression "phrase verbale noyau en français écrit" désigne ainsi un ensemble de suites de mots (ensemble fini de par les limitations de ce travail, cf. § 2) qui peuvent se trouver dans des textes écrits produits ou susceptibles d'être produits en français. Ces suites sont observationnellement caractérisées par leurs limites gauche et droite, et par les catégories des

mots qui peuvent s'y trouver. La limite fixée à gauche coïncide largement, mais non entièrement, avec la frontière reconnue dans [Blanche-Benveniste et al., 1984, p. 74] :

[...] chaque fois qu'il y a une rencontre entre deux formes homophones [p. ex. la rencontre des deux *nous* dans *nous nous regardons*] une frontière passe entre les deux formes : celle qui sépare le sujet du bloc syntaxique formé par le verbe et ses clitiques compléments.

La non-coïncidence totale vient de ceci : tout ce qui est à gauche de la limite gauche reconnue dans ce travail — en fait, très largement, les suites qui expriment le sujet — est aussi à gauche de la frontière reconnue par Blanche-Benveniste et al., tout ce qui est accepté à droite de la frontière gauche par Blanche-Benveniste et al. est aussi à droite de la limite gauche ici reconnue — p. ex. les clitiques — mais dans ce travail on admet, à l'intérieur des limites gauche et droite, des catégories de mots — comme la négation, l'adverbe *toujours* et la forme *tous* — qui ne sont strictement pas dans “le bloc syntaxique formé par le verbe et ses clitiques compléments”.

L'expression “phrase verbale noyau en français écrit” ne désigne donc pas un “constituant” quelconque d'un quelconque “modèle” de grammaire dans une quelconque “théorie” — ou plus modestement, système descriptif —, et elle ne prétend à aucune corrélation avec un “constituant” de ce type.

Les suites de la phrase verbale noyau en français sont ici étudiées dans un cadre que, avec Philippe Blache et Caroline Hagège, nous appelons *5P*, chacun des *P*, dûment numéroté, correspondant à *Protocoles (P1)*, *Propriétés (P2)*, *Projections (P3)*, *Principes (P4)* et *Processus (P5)*, rubriques qui seront présentées ci-après § 1.2.¹. Dans la section immédiatement sui-

1. Le cadre *5P* est le résultat de la complémentation de deux points de vue, celui de [Blache 98a, et 98b], et celui de [Bès 98a], ce dernier travail cristallisant des contributions précédentes du GRIL, diffusées de manière plutôt confidentielle et restées sous veilleuse, où l'on s'était efforcés à distinguer les aspects purement déclaratifs concernant les caractéristiques des langues naturelles, des aspects algorithmiques du traitement automatique. Le travail ici présenté a beaucoup profité des discussions approfondies entre l'été 97 et l'automne 98 avec Claire Blanche-Benveniste, Caroline Hagège et Philippe Blache. En particulier, ce sont les discussions avec Blanche-Benveniste qui ont permis de détecter des faiblesses — importantes — dans les premières versions de la description proposée, et qui ont permis de faire surgir les notions de texte virtuel et de texte réel; sans les travaux de Blanche-Benveniste et les travaux qu'elle oriente au GARS, cette possibilité de voir les choses ne me serait pas apparue. Le travail en cours avec Caroline Hagège et avec Philippe Blache a permis de mûrir et améliorer sur bien des points le formalisme proposé. Merci encore à Glyn Morrill, Emmanuelle Rodier, Mohammed Abaidi, François Trouil-

vante — § 1.1. — on situe, sur le plan méthodologique, les propositions ici faites afin de donner un aperçu général des ambitions et limites du travail présenté.

1.1. LE CADRE MÉTHODOLOGIQUE

Le cadre *P1(P2, P3, P4) P5* (ou en abrégé *5P*) essaye d'intégrer des points de vue différents qui ont cours de nos jours en linguistique avec l'ambition explicite de dépasser les difficultés d'interaction entre linguistes, informaticiens et logiciens intéressés par les langues naturelles, entre linguistique et traitement automatique des langues, entre linguistique tout court et linguistique formelle (ou dite formelle). L'étude des langues humaines devrait être un travail où les efforts de linguistes, d'informaticiens et de logiciens pourraient et devraient converger. Dans le cadre ici proposé on s'efforce de distinguer des points de vue sans pourtant les isoler; la description des langues, exprimée par les Propriétés (P2), les Projections (P3) et les Principes (P4) n'est ni négligée ni subordonnée au traitement automatique. Mais cette description est exprimée dans un formalisme calculable, de telle manière qu'un lien formel puisse être établi entre ces descriptions et les Processus (P5) de traitement automatique. Ces Processus, à leur tour, doivent être effectifs : une machine existante doit pouvoir les exécuter dans des conditions d'une pratique réaliste. Enfin, le formalisme d'expression des descriptions ne se vante d'aucune naturalité ou élégance : le souci est avant tout de décrire (Propriétés (P2)) et de comprendre ou expliquer (Projections (P3) et Principes (P4)), et ceci de manière calculable et toujours et constamment testable par rapport aux Protocoles (P1). La section 8 présente une mise en perspective du paradigme général lui-même et des descriptions proposées dans les Propriétés (P2) en les comparant de manière synthétique à d'autres approches.

Il est probable que si l'on traduisait les Propriétés (P2) ici proposées dans un langage de spécification partiellement intuitif — c'est-à-dire non entièrement spécifié lui-même et laissant par là des marges d'interprétation à ceux qui l'utilisent — on pourrait faire le constat qu'aucune des Propriétés proposées, ainsi traduite, n'apporte rien de nouveau à ce que l'on savait déjà sur la phrase verbale noyau du français écrit, dans ce sens

ieux, José Rodrigo Mateos, Salah Aït-Mokhtar, Victor Castel, Daniel Rossi, Leonardo Ferres et Gérard Guieze, avec qui, individuellement ou dans des réunions de travail, j'ai pu discuter des aspects du travail ici présenté. La responsabilité de tout ce qui est ici proposé me revient entièrement; ceci est particulièrement vrai pour la mise en perspective dans § 8.

que toute Propriété ici proposée fait état d'une observation qui avait été déjà faite quelque part dans la littérature sur le sujet. Même s'il a été impossible d'examiner toutes les descriptions existantes du français, il est bien probable que cette absence de nouveauté, ainsi comprise, est réelle.

Selon l'idée que chacun se fait de ce qu'est *la science* ou *la connaissance intéressante, significative* (ou comme on voudra l'appeler pour la caractériser par une évaluation positive), cette circonstance peut ou non être considérée comme étant définitivement rédhitoire.

Le passage suivant, traduit de [Bochenski 58]², exprime un des aspects de la notion *connaissance intéressante* qui est sous-jacente à ce travail :

[La science] étant un ensemble de propositions, ses parties doivent être reliées par des relations logiques. Il est vrai que la science dans ses apports premiers se réduit à un cumul de propositions qui ne sont pas toujours reliées entre elles. Mais cette situation initiale est considérée par l'homme de science comme quelque chose de peu satisfaisant et la direction principale de toute recherche ne s'ordonne pas seulement vers la découverte de contenus nouveaux, mais aussi (peut-être avant tout) vers l'organisation logique des propositions déjà formulées. [Bochenski 58, p. 36]

S'il est probablement vrai que chacune des Propriétés (P2), propositions descriptives de la phrase verbale noyau du français écrit, peut être trouvée dans un travail déjà publié sur la question, il est non moins probablement vrai qu'il n'existe aucun travail déjà publié ou toutes les Propriétés (P2) ici proposées ont déjà été présentées exprimées dans un même système construit dans un formalisme calculable. Alors, si l'on accepte la notion de *science* ou *connaissance intéressante* qui se dégage de la citation de Bochenski — et qui peut se retrouver formulée différemment par d'autres méthodologues et philosophes de la science — la description ici proposée a la prétention de présenter quelque chose qui est différent de ce qui existe déjà et qui, comme dans beaucoup de processus dans l'histoire des sciences, tout en profitant d'au moins une partie de l'existant, l'améliore.

Pour évaluer cette prétendue différence, il faut se mettre d'accord sur ce qu'on comprend par *formalisme calculable*. La caractérisation de cette notion sera empruntée, elle aussi, à différents passages de [Bochenski 58] (toutes les italiques proviennent du texte original).

2. C'est nous qui avons traduit en français les textes cités de l'édition espagnole de l'ouvrage.

[Le formalisme] consiste à faire abstraction du *signifié* des signes qu'on emploie et à considérer ceux-ci exclusivement sous leur *forme graphique*. [p. 78].

[Le formalisme] n'est plus que l'extension d'une méthode connue depuis des siècles, la méthode du calcul. [p. 80]

[...] un signe peut avoir un sens double, celui qu'on appelle *eidétique* dans un système si nous connaissons son corrélatif *sémantique*, c'est-à-dire si nous savons ce qu'il désigne, ce qu'il signifie. En revanche, un signe a un sens purement opérationnel si nous savons seulement comment il doit être employé, si nous ne connaissons que les règles syntaxiques qui sont valables pour lui. [p. 84]

Par conséquent le formalisme est une méthode qui consiste à faire entièrement abstraction du sens eidétique des signes et à opérer avec eux en utilisant des règles qui portent exclusivement sur leur forme graphique. [p. 87]

La finalité du calcul et du formalisme est toujours un savoir. Un système formel remplit sa mission seulement lorsque ses résultats finaux sont interprétables eidétiquement. La science n'est pas un jeu. [p. 88]

De fait, dans la construction des systèmes formalisés, on procède toujours de la manière suivante : on établit des signes avec sens et on fait par la suite une abstraction de ce sens, pour construire ainsi le système formel et donner en définitive une nouvelle interprétation au système achevé. [p. 89]

Succinctement dit : les Propriétés (P2) qui sont présentées dans § 4 sont exprimées dans un formalisme calculable ; la syntaxe des expressions de ce formalisme est présentée dans § 3 conjointement avec des éléments permettant de comprendre leur interprétation sémantique formelle³. Lorsque des Propriétés (P2) sont proposées pour n'importe quelle langue et pour n'importe quelle construction dans cette langue, il existe des règles qui vont les combiner entre elles pour obtenir tous et seulement les objets qui satisfont ces Propriétés (P2). Les Propriétés (P2) opèrent sur des symboles — les catégories (cf. § 3.1. et les catégories proposées pour le français dans § 4.1) — auxquels on assigne ce que Bochenski appelle

3. Une interprétation plus rigoureuse de la sémantique est donnée dans Bès [98a] et encore améliorée dans [Bès, Blache et Hagège en préparation]. Les propriétés de calcul du formalisme utilisé sont simples et bien connues : il s'agit d'utiliser d'une part des structures de traits qui n'introduisent pas la récursion, et des listes de symboles qui se distribuent elles-mêmes dans des ensembles. Le calcul se fait sur des inclusions et des compléments entre ensembles de listes. Ceci veut dire que même si dans ce travail une interprétation complètement exhaustive du formalisme de calcul n'est pas donnée, la conjecture portant sur sa calculabilité semble bien fondée.

un sens eidétique : en fait, on donne des exemples des signifiants des mots qui doivent être associées dans le lexique à ces catégories, ou on les énumère. Mais le “sens eidétique” de ces symboles n'intervient en rien ni dans la formulation de la syntaxe qui va permettre d'obtenir les expressions des Propriétés (P2) ni dans le calcul sur la combinatoire entre les entités dénotées par ces expressions. Tout ceci se fait, pour ainsi dire, de manière aveugle par rapport à ce que chaque symbole veut dire. Mais comme tout ceci n'est pas un jeu, les objets ainsi définis — que nous appelons *les modèles*, c'est-à-dire les objets qui satisfont les Propriétés (P2) — ont une “interprétation eidétique” — toujours dans la terminologie de Bochenski — dans les phrases du français observées de certaine manière, dûment stipulée (cf. § 2). Et c'est cette interprétation eidétique dans un réel observable qui va donner à la linguistique la possibilité de proposer des hypothèses descriptives et explicatives d'exactlyement le même type que celles des sciences du réel, dites parfois naturelles ou empiriques.

Pourquoi le formalisme est utile, nécessaire, voire, nous le croyons, indispensable ? Pour les raisons signalées par beaucoup et que nous citons, empruntées toujours à Bochenski [58].

Dans une situation complexe notre vision eidétique de l'objet échoue très rapidement. [p. 89]

Étant donné que dans un système formel toutes les règles font exclusivement référence à la forme graphique, il est impossible de faire en lui une démonstration avec des règles et des axiomes non formulés. Et l'on sait que les suppositions non formulées sont dangereuses, car elles peuvent être facilement fausses et elles échappent à une vérification rationnelle du fait de ne pas avoir été clairement formulées. Le formalisme contribue fondamentalement à l'élimination de toute spécification tacite. [p. 90]

[...] dans un système axiomatique formellement développé, toutes les conséquences se déduisent des axiomes choisis avec une certaine facilité et d'une manière bien définie. [p. 90-91]

La question précédente sur la raison d'être d'un formalisme peut se compléter par la suivante : pourquoi la langue naturelle ne suffirait-elle pas pour parler d'elle-même ? Encore une fois, Bochenski est notre porte-parole pour répondre à cette question [Bochenski 58, p. 93-94] :

Dans les sciences en général [...] on emploie des concepts tellement abstraits qu'on ne trouve pas de mots adéquats dans le langage courant.

La syntaxe du langage ordinaire est peu précise, ses règles ont beaucoup d'exceptions, ce qui est un inconvénient pour bien opérer dans une science rigoureuse.

[...] la plupart des mots du langage courant sont ambigus.

Il est maintenant possible de résumer, sur le plan méthodologique, les prétentions de ce travail. On propose ici une description formelle et calculable des suites de la phrase verbale noyau du français écrit. Cette description est proposée au moyen des Propriétés (P2). Ces Propriétés (P2) sont calculables car on peut en déduire, par des opérations opérant sur les expressions qui expriment les Propriétés (P2), toutes leurs conséquences et rien que leurs conséquences. Le résultat de ce calcul peut être testé par rapport à un univers d'observation. Ces Propriétés (P2) sont donc censées être confirmées ou infirmées⁴ par rapport à cet univers d'observation. Par ailleurs, ces Propriétés (P2), parce qu'exprimées dans un formalisme calculable, peuvent être associées à des Processus effectifs qui, parmi d'autres fonctionnalités, devraient pouvoir, dans les textes effectifs, reconnaître les expressions qui satisfont ces Propriétés (P2) : le traitement automatique est ainsi distingué de la description, mais il n'est pas scotomisé de la description.

Il est absolument vrai que les Propriétés (P2) ici présentés ne sont pas tombées du ciel toutes faites ; en fait elles bénéficient des années et des années de travail de recherche accompli par beaucoup de monde. De ce fait la probabilité de les trouver exprimées non formellement ailleurs est très grande ; il s'agit plutôt d'une quasi-certitude. L'apport nouveau — si apport nouveau il y a — c'est leur présentation intégrale dans un formalisme calculable et interprétable par rapport à un réel qui permet de les tester.

Il reste à expliciter encore une limitation de ce travail. Les Propriétés (P2) ici présentées sont une description et non une explication et moins encore une théorie. Succinctement parlant, les Propriétés (P2) disent : les choses peuvent être décrites comme cela. Avec les Propriétés (P2) on ne prétend pas à l'élégance ou à la naturalité de la description, on n'essaye de convaincre personne que telle construction doit être obtenue de telle autre construction, que telle Propriété est première, ou deuxième par rapport à telle autre ou dernière par rapport à toutes les autres ; on n'affirme rien sur la redondance ou non redondance des Propriétés proposées. Ce que l'on propose est une description possible du domaine visé du français écrit. D'autres descriptions sont possibles dans le même formalisme et

4. On dirait *contrastées* ou *falsifiées* dans une terminologie plus popérienne, que nous n'adoptons pas ; on utilisera *confirmer* et *infirmer* pour exprimer les résultats respectivement positif et négatif de la confrontation avec l'observé, en suivant [Auroux 81] ; on prend donc Popper avec le (gros) grain de sel de ceux qui pratiquent effectivement les sciences du réel (cf. [Sokal et Bricmont 97], chap. 3).

dans d'autres formalismes ; il n'y a aucune raison pour qu'elles ne puissent pas être extensionnellement équivalentes à la description ici proposée, c'est-à-dire qu'elles peuvent spécifier les mêmes objets en signalant les mêmes caractéristiques de ces objets. Par ailleurs, l'observation proposée n'a aucune prétention à avoir épuisé le "vrai" réel : chaque point de vue d'observation introduit sa propre perspective et de ce fait mutile inéluctablement l'objet observé ; dans § 8 on fera état de quelques limitations criantes des observations proposées. Humblement mais clairement on se limite ainsi à dire ceci :

- si telles sont les observations sur cette partie du français écrit
- alors avec les Propriétés (P2) ici proposées on doit pouvoir rendre compte de ces observations.

En fonction de ses intérêts personnels, sa curiosité intellectuelle, ses goûts, sa position dans la hiérarchie de l'Académie, ses ambitions de carrière, chacun devrait pouvoir évaluer si cela vaut ou non la peine de continuer la lecture de ce travail étant donné ses prétentions, tellement modestes et encore non prouvées. Comme il a été déjà annoncé, une analyse plus détaillée par rapport à l'existant sera présentée dans § 8.

1.2. LES COMPOSANTS DE $5P$

À partir de la numérotation des P de $5P$ en $P1(P2, P3, P4)P5$ on décline les composants du cadre ici proposé de la manière suivante.

- $P1$ ou P de Protocoles, un Protocole étant la représentation d'une donnée à laquelle aboutit un Observateur dûment modélisé, cette donnée ayant ou non été observée dans un corpus effectif.
- $P2$ ou P de Propriétés, une Propriété pouvant s'identifier formellement à un axiome. Un ensemble fini de Propriétés spécifie en intention un ensemble (fini ou infini) de suites d'une langue donnée. On appellera modèles ces suites. Un *modèle* est ainsi une suite qui satisfait un ensemble de Propriétés. Les modèles vont être associés aux énoncés de la langue décrite.
- $P3$ ou P de *Projections*, une Projection étant une abstraction d'un ensemble de Propriétés d'une langue spécifiant les caractéristiques communes à toutes ces Propriétés.
- $P4$ ou P de Principes, un Principe -tout comme une Propriété- pouvant formellement s'identifier à un axiome, les principes étant plus abstraits et généraux que les Propriétés. Les Principes vont introduire des

contraintes générales, valables pour toutes les langues ou pour des groupes de langues, sur le type des Propriétés — et donc des modèles — qui décrivent les langues particulières.

- *P5* ou *P* de Processus, un Processus étant une procédure effective implantable ou implantée en machine, permettant d'analyser et/ou de produire un énoncé d'une langue particulière, cet énoncé étant explicitement associé à un modèle qui satisfait les Propriétés (*P2*) décrivant cette langue particulière.

Les Propriétés (*P2*) sont testées en les confrontant aux Protocoles (*P1*). Le résultat de la confrontation exprime l'adéquation externe de ce système d'hypothèses que sont les *P2*. Par ailleurs, le triplet (*P2*, *P3*, *P4*) permet d'expliquer. Ici *expliquer* est assimilé à *prévoir* : si l'on connaît les *P2* nécessaires à la description d'un corpus *C* fini — et le plus réduit possible — ces *P2* doivent, conjointement avec les *P3* et les *P4*, permettre de calculer — i.e. de prévoir — d'autres *P2* — les *P2'* — qui soit sont également nécessaires à la description de *C*, soit sont nécessaires à la description d'autres corpus de la même langue.

Ce travail porte essentiellement sur les Propriétés (*P2*) de la phrase verbale noyau en français écrit. On se propose donc de présenter les *P2* qui vont spécifier les modèles qui doivent être associés aux suites reconues comme étant des phrases verbales noyau bien formées en français écrit.

Dans *5P* on souhaite calculer formellement et effectivement tout ce qui est formellement calculable dans une science du réel. À partir de (*P2*, *P3*, *P4*) on doit être capable de calculer d'autres Propriétés que celles dans *P2*; à partir de *P2*, on doit pouvoir calculer *M*, c'est-à-dire tous et seulement les modèles qui satisfont *P2*; enfin, à partir de *P2*, *P3* et éventuellement *P4*, on doit pouvoir calculer les sources déclaratives des Processus (*P5*) d'analyse et génération automatiques, ces processus de traitement automatique devant permettre d'interpréter ou de produire tout énoncé associé à un modèle dans *M*. Par *formellement calculable* on comprend ici un calcul exprimé par des opérations dont on connaît les capacités mathématiques formelles et effectué sur les symboles graphiques eux-mêmes (cf. l'alinéa précédent § 1.1). Par *effectivement calculable* on comprend ici un calcul qui, en plus d'être formel comme on vient de le préciser, puisse être effectué dans des conditions réalistes et effectives par une machine existante.

Mais dans *P1(P2, P3, P4)P5* on souhaite aussi tester de manière effective et systématique, et ce souhait est aussi fort que celui de calculer et de

traiter automatiquement. Tout ce qui concerne la spécification des *P1*, la confrontation des modèles spécifiés par *P2* avec les Protocoles décrivant les énoncés dans *P1*, le traitement effectif par les processus de *P5* des énoncés attestés dans des corpus effectivement utilisés, tout ceci, qui relève de l'observation et des applications effectives du traitement automatique, est aussi crucialement important dans *5P* que la formalisation permettant le calcul et le traitement automatique. Ceci a pour effet bénéfique de laisser dans un plan complètement secondaire les prétentions à la naturalité d'utilisation d'un formalisme donné ou à ses qualités d'élégance. Le point de vue adopté est qu'en science ce qui importe prioritairement n'est pas la naturalité où l'élégance du formalisme utilisé — en linguistique ce formalisme ayant été souvent importé de domaines où ils jouent un rôle complètement différent — mais de proposer des hypothèses les plus générales possibles, de tester scrupuleusement leurs conséquences, et de prévoir le plus possible à partir des observations. Et il est complètement inutile de se poser des questions de précédence : il y a interaction entre hypothèses et observations, mais pas d'exigence de précédence des unes sur les autres.

En revanche *5P* ne s'intéresse pas à des questions proches de la terminologie, et très peu à des querelles sur la question de savoir si telle observation est valable ou non au niveau des *P1*. P. ex. on ne discute pas si ce qu'on appelle ici des pronoms clitiques sont vraiment des clitiques, à partir d'une définition générale et conventionnelle de ce qu'est un clitique. Dans ces essais de définition générale on incorpore parfois des caractéristiques qui viennent de l'accentuation orale, caractéristiques que l'on ne peut pas utiliser ici, car on travaille sur des suites de caractères. De telle manière que le lecteur est invité à lire les Propriétés (*P2*), les catégories avec leurs traits (§ 3.1 et 4.1) et l'interprétation de ces catégories strictement selon les définitions données aux catégories dans § 4.1 et selon l'interprétation sémantique des Propriétés (*P2*) données — avec une formulation non complètement rigoureuse, cf. ci-dessus § 1.1 — dans § 3.2, 3.3 et 3.4. Ainsi, pour comprendre les résultats obtenus par les Propriétés (*P2*), *clitique* est un objet qui a la valeur de trait “+” de l'étiquette *cl*. La présentation de § 4.1 doit permettre l'interprétation — le sens eidétique, dirait Bochenski — de ce symbole graphique : cette présentation doit permettre d'associer des mots du français — ceux-ci compris comme des suites de caractères — à ces symboles. On observera que la forme *ne* ne reçoit pas la valeur de trait “+” pour l'étiquette *cl*. Dans ce travail, la forme *ne* n'est donc pas un clitique, mais dans ceci on ne doit

voir nulle prétention à discuter si la forme *ne* est “vraiment” ou non un clitique en français.

Par ailleurs, il n'y aura pas de discussions sur la question de savoir si les phénomènes traités relèvent de la morphologie, de la syntaxe ou de la sémantique. Chacun peut interpréter les Propriétés (P2) comme relevant d'un ou plusieurs de ces domaines. Si l'on prend les notions plus dures associées aux termes *syntaxe* et *sémantique* dans la théorie générale des langages, il est clair, comme il a été établi dans la section 1.1 précédente, que le calcul sur les Propriétés (P2) est un calcul syntaxique, que les catégories sur lesquelles sont formulées les Propriétés (P2) possèdent une sémantique, dans le sens qu'elles possèdent une dénotation dans l'ensemble des mots (signifiants) du français (et non pas dans le sens de la sémantique linguistique, car on ne s'intéresse en rien au fait de savoir si tel verbe veut dire ceci ou cela) et que le résultat du calcul — les modèles — possède une sémantique dans l'observation du réel. Mais nulle prétention dans tout ceci à édicter des normes sur le fait de savoir si “vraiment” dans les langues naturelles la “syntaxe” doit être autonome, si la “sémantique” doit être interprétative ou générative, si l'une est “prioritaire” par rapport à l'autre...

Même si les limites entre morphologie et syntaxe sont fluctuantes, il semblerait vrai qu'il y ait des zones dans un système linguistique qui résistent mieux que d'autres au changement linguistique et à la variabilité dialectale. Ces zones plus dures relèvent plutôt de ce qu'on entend couramment par “morphologie” que de ce qu'on appelle “syntaxe”. L'ambition du cadre proposé serait aussi d'aider à mieux poser ces questions. Celles-ci devraient s'inscrire dans une formulation qui explicite la capacité de l'Observateur à faire des observations dans le temps et dans l'espace portant sur des systèmes différents mais qui se “correspondent” (notion à définir elle aussi formellement). La description respective devrait intégrer la comparaison des Propriétés (P2) différentes et devrait fixer des critères pour définir la “variabilité” et la “stabilité”. Ceci étant fait, on devrait obtenir des corrélations entre d'une part types de catégories avec leur comportement différencié dans les différents types de Propriétés, et d'autre part, des observations sur la “variabilité” et la “stabilité” des catégories “correspondantes” dans un autre système. Ceci serait un embryon de description et, peut-être même, de théorie, permettant de prévoir ce type de phénomènes.

Pour ce qui est de la validité des observations, il n'y aura pas de discussion sur la question de savoir si telle construction est “vraiment possible” en français ou non, p. ex. s'il y a ou non coordination des clitiques

ou des articles en français. La problématique sous-jacente à ces questions est traitée de la manière suivante. On suppose, d'une part, la possibilité d'accès à des corpus censés être homogènes, avec des énoncés bien formés, et indéfiniment extensibles. On suppose, d'autre part, un Observateur, qui doit être dûment modélisé en lui attribuant des capacités explicites pour décrire les données avec des prédicats d'observation spécifiques.

Un corpus est par définition fermé. Si l'on suppose une source homogène, on se donne un corpus indéfiniment extensible, ce qui rend les choses plus faciles. Mais même dans ce cas on n'aura pas, dans le cumul des corpus observés à un moment donné, l'échantillonnage complet de toutes et seulement les constructions d'un certain type que l'on veut décrire. Étant donné les caractéristiques supposées des corpus enregistrés à un moment donné, on sait que toutes les constructions qui y sont attestées sont bien formées, mais une construction non attestée n'est pas forcément mal formée. C'est principalement là où l'Observateur doit forcément intervenir : il est chargé de discriminer, dans ce qui n'a pas été observé, ce qui est bien et ce qui est mal formé, sous des conditions d'observation bien définies. Avec cette vision des choses, il n'y a pas d'opposition entre les "linguistiques du corpus" et les "linguistiques de l'introspection" : elles sont complémentaires dans une hiérarchie selon laquelle ce qui est attesté dans des corpus effectifs a la priorité par rapport aux décisions de l'Observateur, dans ce sens que, en principe, si une construction est observée dans un corpus effectif, l'Observateur ne pourra pas la récuser⁵. Mais dès qu'on accepte qu'il n'existe pas de situation telle où tout ce qui est bien formé a déjà été dit dans un cumul de textes, l'Observateur et sa fonction ne peuvent pas être éliminés. Le décalage entre ce que l'Observateur admet comme bien formé tout en n'ayant pas été déjà dit et ce qui a été déjà dit est la pierre angulaire qui fonde la différence entre texte virtuel — celui qui est spécifiable selon les jugements de l'Observateur — et texte réel, celui qui a été effectivement produit dans des situations effectives de communication⁶.

5. Ce qui est dit dans le texte concerne l'utilisation des corpus du point de vue méthodologique pour tester des hypothèses linguistiques; au delà de ce point de vue, il y a toutes les exploitations possibles des observations textuelles permettant de construire une vraie théorie des corpus réels — oraux ou écrits; cf. [Blanche-Benveniste 96], [Coppieters 97] et [Habert et al. 97].

6. L'auteur du *Mio Cid*, Dante, Du Bellay, Ronsard, Mistral, parmi d'autres, ont compris et utilisé les possibilités offertes par le texte virtuel et ils les ont cristallisées en textes réels.

L'Observateur est modélisé : on lui attribue des capacités pour décrire les données en utilisant des prédicats d'observation. Ce qui est important ici est, d'une part de fixer les conditions d'observation, et d'autre part, que celles-ci soient les plus intersubjectives possibles : des observateurs individuels différents qui ont acquis la capacité d'observer et de décrire en utilisant les mêmes prédicats descriptifs, doivent aboutir, dans des marges d'erreur contrôlables, aux mêmes Protocoles, lorsqu'ils observent les mêmes données. Le problème est donc de définir les conditions d'observation et d'évaluer l'intersubjectivité de cette observation, et de se donner un système de représentation permettant de noter les Protocoles⁷.

Ceci n'empêche qu'il y ait toujours une marge de non intersubjectivité dans les observations — qu'il faudrait pouvoir déterminer au mieux. Cette non intersubjectivité semble bien relever de deux types : non intersubjectivité erratique et non intersubjectivité systématique. La première doit être la plus réduite possible. Dès qu'elle dépasse des limites intuitivement tolérables, il vaut mieux reconsidérer les prédicats d'observation. En revanche, la non intersubjectivité systématique doit être considérée comme un défi de plus auquel *5P* doit répondre.

La non intersubjectivité systématique se manifeste lorsque des observateurs différents coïncident dans leurs résultats pour un ensemble de phénomènes, mais, tout en opérant avec un même prédicat d'observation, sur un autre ensemble de phénomènes l'un donne aux constructions de cet autre ensemble, une des valeurs prévues par ce prédicat, et l'autre une valeur différente. C'est, p. ex., le cas si les observateurs s'accordent sur la qualité de bien formées d'un ensemble de suites, et que l'un donne la valeur *bien formée* aux suites d'un autre ensemble — dûment caractérisées — et l'autre la valeur *mal formée* à ces mêmes suites. Ce type de

7. Sur cette utilisation de la notion d'intersubjectivité en linguistique, cf. [Kerbrat-Orecchioni 80], p. 15; sur la nécessité de se donner un système de représentation de l'expérience, cf. [Granger 92] p. 43 : "La vérification scientifique [...] consiste donc en une mise à l'épreuve [...] d'un parti pris de *représentation* de l'expérience" (en italiques dans le texte original), cf. l'Introduction et la Première partie de l'ouvrage. On utilise ici *Protocoles* sans forcément toutes les caractéristiques données par l'empirisme à la notion d'énoncés protocolaires; nos *Protocoles* sont les propositions d'observation de [Bunge 69], dont on adopte l'exigence selon laquelle on doit, à partir des hypothèses, pouvoir déduire des formules susceptibles d'être approximativement traduites à des propositions d'observation en vue d'être comparées à celles-ci. En fait, dans ce travail, les Protocoles sont représentés très simplement par des suites de mots associés à des catégories maximales et à des jugements de l'Observateur (cf. § 2); aucun langage de représentation particulier plus sophistiqué n'est proposé, mais il reste clair que, selon le point de vue adopté, les données ne sont ni la "réalité" ni "le phénomène en soi".

variation peut, par ailleurs, être observée en dialectologie et en diachronie. Dans une situation ainsi caractérisée, les Propriétés (P2) doivent traiter de manière identique ce qui est commun aux observateurs concernés, et moyennant des modifications mineures, les P2 doivent pouvoir s'adapter aux différents constats de variation. Ce travail exclut de manière explicite de son domaine d'observation la coordination. La coordination ou non des formes clitiques ne le concerne donc pas; mais, si elle le concernait, l'idée serait non de se battre sur le fait de savoir s'il y a "vraiment" coordination des pronoms clitiques en français, question qui nous semble être, ainsi posée, complètement stérile, car il est impossible d'y donner une réponse rationnelle, mais de proposer une description qui puisse, à la demande, s'accommoder de la coordination des pronoms clitiques.

On aboutit ainsi à une dernière exigence sur 5P : on souhaite, en particulier pour (P2, P3, P4), que les hypothèses formulées soient modulaires et incrémentales. Ceci veut dire que le système doit permettre d'ajuster la finesse de l'analyse prévue aux objectifs que l'on veut atteindre, et ceci moyennant des modifications mineures. P. ex. si l'on ne souhaite pas décrire ou différencier certaines propriétés d'accord, on doit pouvoir, moyennant l'introduction ou suppression de certaines Propriétés, adapter la description à la demande, sans que le reste de la description en pâtisse. C'est ce choix méthodologique qui a dicté l'organisation des observations en niveaux (cf. § 2 et 7), et, p. ex. la présentation des clitiques avec les formes impératives en deux versions : la puriste, très peu pratiquée, et la non puriste. Le formalisme peut exprimer les deux. Que l'une soit la version couramment utilisée est encore une observation qui relève de la distance entre texte virtuel et texte réel, et qui relève d'une vraie théorie de la variabilité linguistique, de la propagation ou non dans la société d'une norme imposée culturellement, etc. mais non de la validité de la description en tant que telle.

C'est dans ce cadre général et avec ces objectifs d'ensemble que la phrase verbale noyau du français est traitée ici. Celle-ci est décrite moyennant des Propriétés (P2). En calculant sur ces P2, on obtient les modèles qui les satisfont. La vérification de l'adéquation externe des modèles se fait dans un cadre observationnel déterminé. Ces conditions de vérification sont fixées dans la section suivante (§ 2). Dans § 3 on présente les types de Propriétés susceptibles d'être utilisés dans une description particulière. Ces types de Propriétés sont utilisés dans § 4 pour décrire la phrase verbale noyau du français observée selon les conditions fixées dans § 2. Le Générateur de Modèles (désormais, en abrégé, *Gén-*

Mod) est l'outil formel qui permet de calculer les modèles à partir des *P2*. Cet outil est présenté dans § 5 conjointement à des exemples de calcul à partir des *P2* présentés dans § 4 et aux modèles effectivement générés. Ceux-ci à leur tour peuvent être utilisés comme des sources d'autres calculs; les résultats de ces calculs obtenus à partir des modèles sont présentés dans § 6. Les modèles spécifiés dans § 5 peuvent et doivent être vérifiés, ce qui permet de caractériser les limitations de l'adéquation externe des *P2* proposées. On montrera en § 7 les possibilités d'adaptation flexible des *P2* proposées et de leur extension incrémentale, ce qui permet d'améliorer l'adéquation externe du système. Dans § 8, comme il a déjà été indiqué, on met en perspective les propositions ici faites en les comparant à d'autres approches.

2. OBSERVATION ET VÉRIFICATION

L'objectif de ce travail est de décrire les phrases verbales noyau du français écrit qui comportent une forme verbale fléchie, avec ou sans participe, la plupart des pronoms clitiques, particules de négation (*ne*, *pas*), l'adverbe *toujours* et la forme *tous*.

Avec cette énumération on laisse explicitement en dehors du champ visé les expressions incises (p. ex. dans *il a, avec patience, étudié le problème*), les formes à l'infinitif (p. ex. dans *pour l'étudier*), les clitiques nominatifs *-on* et *-vous[sg]*, l'impersonnel *-il*, le pronom *-ce* et les coordinations (p. ex. dans *il a étudié et rêvé*).

Le système présenté — un lexique de catégories maximales et les Propriétés (*P2*) — doit pouvoir générer, par le biais du *GénMod*, toutes et seules les suites de catégories maximales — les modèles — qui satisfont ces Propriétés (*P2*).

Il est à noter que les Propriétés (*P2*) doivent spécifier tout ce qui est interne à la phrase verbale noyau; les relations — p. ex. d'accord —, entre les suites enchaînées à gauche et à droite avec la phrase verbale noyau sont en dehors du champ visé par les *P2* ici présentées.

Pour des raisons de clarté d'exposition autant que pour illustrer la capacité incrémentale du système, les Propriétés (*P2*) sont présentées à différents niveaux, chaque niveau intégrant tout le domaine observationnel du niveau précédent. Cette présentation en niveaux n'est nullement une exigence du formalisme : celui-ci permet, sans aucune altération de ses possibilités expressives, la formulation des *P2* pour rendre compte, p. ex.

du Niveau 3 sans passer au préalable par la formulation des *P2* correspondant aux Niveaux 1 et/ou 2.

Les niveaux distingués sont les suivants :

Niveau 1 : Suite des catégories entre la première à gauche et le noyau verbal, mais sans clitiques nominatifs, ni formes verbales à l'impératif, ni forme de l'auxiliaire surcomposé, ni négation, ni adverbe *toujours*, ni forme *tous*.

Niveau 2 : Extension du Niveau 1 par ajout des clitiques nominatifs et de l'auxiliaire surcomposé.

Niveau 3 : Extension du Niveau 2 par ajout de la négation et des formes verbales de l'impératif. Ce niveau est présenté en deux versions : version puriste et version non puriste.

Niveau 4 : Extension du Niveau 3 version non puriste par ajout de *toujours* et de la forme *tous*.

Niveau 5 : Extension du Niveau 4 par la prise en compte d'exceptions introduites par certains lemmes verbaux.

Ce qui précède définit les conditions d'observation dans les limites desquelles le système doit être testé.

Confronter les modèles générés par le *GénMod* avec les données d'observation est la tâche qui va permettre de confirmer ou infirmer le système. Comme il a été dit au § 1.2 on se donne deux types différents et complémentaires de données d'observation : celles qui proviennent de l'accès à des corpus supposés homogènes et extensibles et celles qui proviennent des capacités explicites attribuées à l'Observateur. Ainsi, on suppose que celui-ci, en plus de l'accès à ces corpus et au lexique des catégories maximales :

i est capable d'étendre à gauche et à droite par des suites, éventuellement nulles, de catégories maximales, tout modèle généré dans chaque niveau

ii ayant accès à un lexique de signifiants de mots, il est capable d'associer aux catégories maximales du lexique des catégories (cf. ces catégories maximales § 4.1) les signifiants de mots ; il peut ainsi associer le signifiant *l'* du lexique des signifiants, à deux catégories maximales différentes de pronoms clitiques, l'une qui sera associée au signifiant *l'* dans *il l'est* et l'autre qui sera associée au signifiant *l'* dans *il l'invente*, de telle manière qu'il est capable d'associer les suites de catégories maximales obtenues selon (i) à des suites de signifiants mots

iii est capable de discriminer des suites de mots qui correspondent à des énoncés bien et mal formés du français par utilisation du prédicat d'observation *être bien formé en français*.

Dans ce cadre on peut définir les conditions d'adéquation externe du système, qui vont permettre de le confirmer ou de l'infirmé :

i Toute suite de (signifiants de) mots attestée dans un texte écrit du français supposé bien formé, si elle est conforme aux conditions d'observation de la phrase verbale noyau définies ci-dessus, est un Protocole d'observation. Pour que le système soit confirmé, il doit exister un modèle spécifié à partir du lexique proposé de catégories maximales et des Propriétés (P2) tel que, si l'on associe des (signifiants de) mots aux catégories maximales du modèle, on retrouve le Protocole d'observation. Le système est infirmé dans le cas contraire.

ii Toute catégorie maximale de tout modèle généré, pour que le système soit confirmé, doit pouvoir être associée à au moins un mot du lexique des signifiants, et la liste de signifiants de mots ainsi obtenue doit pouvoir s'enchaîner à gauche et à droite avec des suites de signifiants de mots, la liste finale des signifiants de mots qui en résulte devant être jugée une suite bien formée du français par l'Observateur, cette liste ainsi obtenue étant aussi un Protocole d'observation. Le système est infirmé dans le cas contraire.

Ce sont ces deux exigences, complémentaires, qui vont définir les conditions d'adéquation externe du système, la première correspondant à l'accès à des corpus extensibles et à la notion de texte réel, la deuxième, à la capacité attribuée à l'Observateur d'utiliser le prédicat *être bien formé en français* et à la notion de texte virtuel. Les deux types de Protocoles sont susceptibles d'être emmagasinés dans une Base de données qui sera utilisée pour tester de manière systématique le ou les systèmes proposés. Les deux types d'exigence sont illustrées par la suite.

• Exemples de l'exigence (i). Dans la page 1 du *Monde* du 18/7/98 on lit la phrase

(... la commission d'instruction) *n'était pas obligée* (de suivre...)

Cette phrase, attestée dans un texte effectif, confirme les Propriétés (P2), puisque à partir de celles-ci on peut générer le modèle suivant, qui présente la suite de catégories qui, par le biais des catégories maximales, permet de retrouver la phrase attestée

ne êtreN pas 1vapN

En revanche, dans le même journal, p.3, on lit

(... la Cour Suprême) *a du même coup confirmé* (la validité...)

Cette phrase, attestée elle aussi dans un texte effectif, ne confirme ni infirme les Propriétés (P2). En effet, celles-ci, dans leur formulation actuelle, n'intègrent que *toujours* parmi les adverbes, la chaîne adverbiale *du même coup* n'est donc pas concernée.

• Exemple de l'exigence (ii). Si à partir des Propriétés (P2) le modèle suivant est spécifié (les lettres capitales indiquent l'accord en nombre — N — et en personne — P) :

ne 1vaisNP cInNP pas

ce modèle confirme les P2, car selon l'Observateur on a en français la phrase bien formée suivante (les suites enchaînées à gauche et à droite sont entre parenthèses et les suites de la phrase noyau en italiques)

(la fille) *ne regarde-t-elle pas* (le livre ?)

3. PROPRIÉTÉS

Les Propriétés sont exprimées en utilisant des catégories et des prédicats qui expriment des relations entre catégories. Les catégories sont des ensembles de traits, chaque trait dans une catégorie étant une étiquette associée à une seule valeur. On distingue trois grands types de Propriétés : les Propriétés d'existence, de linéarité et de fléchage. Dans § 3.1 la notion de catégorie et son utilisation sont présentées. Dans les trois sections suivantes, on présente successivement les Propriétés d'existence, de linéarité et de fléchage.

3.1 LES CATÉGORIES

On utilisera une notion de catégorie très proche de celle qui est utilisée dans GPSG [Gazdar et al. 85, chap. 2]. Un trait du vocabulaire des traits est une étiquette avec une ou plusieurs valeurs. Une catégorie est un ensemble non vide d'étiquettes de traits, chaque étiquette ayant une seule valeur.

Un trait booléen a comme valeurs '+' et '-'. Exemples des traits booléens qu'on va utiliser :

être ; {+, -}
 cl[itique] ; {+, -}
 v[erbe] ; {+, -}

Exemples de traits⁸ qui présentent plus de deux valeurs ou deux valeurs qui ne sont pas {+, -}

pers(onne) ; {1, 2, 3}
 sous-classe-v ; {ve, va} [ve : verbes qui exigent *être*; va verbes qui admettent *avoir* ou *être*]
 sous-classe-ve; {ve1, ve2}[exemple de ve de la sous-classe 1 : *arriver* ; exemple de ve de la sous-classe 2 : *s'agenouiller*]

Conventions de notation. Dans les catégories on note directement les valeurs des traits. Pour les traits booléens, la valeur + se note par l'étiquette. Ainsi, dans une catégorie, *être* note la valeur + de l'étiquette *être*. Le symbole ~ conjointement avec l'étiquette correspondante note la valeur - d'un trait booléen. Pour les traits non booléens, on note dans la catégorie directement la valeur du trait. Les traits d'accord sont indiqués de manière différenciée (cf. ci-dessous). Pour alléger la notation, les valeurs des traits exprimées en utilisant des entiers (p. ex. *ve1* et *ve2*) seront directement notées par les seuls nombres entiers si aucune ambiguïté ne risque de se produire.

Dans le lexique des catégories, on définit les *catégories maximales*. Étant donné un ensemble de traits, et la hiérarchie entre ces traits, les catégories maximales sont toutes les catégories que l'on peut former avec cet ensemble de traits et la relation d'héritage entre les traits, celle-ci étant définie par la hiérarchie des traits. On obtient ainsi les catégories — dites ici maximales — auxquelles on ne peut rajouter aucune valeur de trait et qui sont toutes différentes les unes des autres par le fait de posséder au moins une valeur différente de trait. P. ex. si l'on reste aux traits présentés ci-dessus et que l'on suppose la hiérarchie des traits [notée par l'indentation]

8. Les traits et les catégories sont présentées ici de manière simplifiée; la version complète est spécifiée § 4.1. Tous les exemples présentés du français dans § 3 supposent toutes les catégories et seules les catégories de § 4.1. Ainsi, p. ex., par *clitique réfléchi* on comprend toutes les formes qui reçoivent la valeur + pour l'étiquette *clr* indépendamment du fait que ces formes soient utilisées ou non dans une construction réflexive. Selon cette convention, les formes *te* dans les deux exemples qui suivent, sont des clitiques réfléchis : *tu te laves, il te regarde* ; on réserve *construction réflexive* pour le premier de ces deux exemples. De même, avec le trait *aux* on fait état de seulement les formes avec le lemme *avoir* (et non celles de *être*) dans un rôle d'auxiliaire.

v
 ve
 1
 2
 ...

les catégories maximales sont

[v, ve, 1]
 [v, ve, 2]

Les catégories non maximales sont constituées par des ensembles de valeurs de traits qui sont inclus dans les ensembles de valeurs de traits des catégories maximales.

Si l'ensemble des valeurs de traits d'une catégorie A est inclus dans l'ensemble des valeurs de traits d'une catégorie B on dit que la catégorie A subsume la catégorie B (ou que la catégorie B est subsumée par la catégorie A). P. ex. la catégorie notée *ve* subsume la catégorie notée *1ve*, car toutes les valeurs de traits dans l'ensemble de traits de la première — à savoir *ve* — sont dans l'ensemble de traits de la deuxième, qui possède les valeurs des traits *ve* et *1*⁹.

Les Propriétés (P2) s'expriment moyennant des catégories. Les modèles spécifiés par ces P2 sont ainsi des suites de catégories. Si ces catégories ne sont pas des catégories maximales — ce qui est très généralement le cas — pour obtenir les modèles complets, les catégories non maximales doivent être remplacées par toutes les catégories maximales dans lesquelles elles sont incluses. P. ex si on a défini les catégories maximales suivantes (où *f* note fléchi; *is* note indicatif-subjonctif, *im* note impératif et *va*, *1*, *2* les valeurs des traits indiqués ci-dessus) :

9. On rappelle l'utilisation de *subsumer* lorsqu'on est dans le domaine des structures de traits. Une catégorie — en tant que symbole — ["être vivant"] — subsume la catégorie ["être vivant", "humain"] (qui est composée de deux symboles) parce que l'ensemble que dénote ["être vivant"] — c'est-à-dire les objets dans le monde qui sont des être vivants — est un surensemble de l'ensemble que dénote le symbole ["être vivant", "humain"] (ou dans l'autre sens : l'ensemble que dénote le symbole ["être vivant", "humain"] est inclus dans l'ensemble que dénote ["être vivant"]): dans la dénotation tous les objets qui sont des hommes sont des objets qui sont des être vivants mais non vice-versa. Mais au niveau des symboles c'est en quelque sorte le contraire : le symbole qui dénote l'ensemble le plus large subsume l'ensemble qui dénote le moins large car il a "moins de symboles internes". L'idée est en général que, moins spécifié est un symbole, plus large est sa dénotation : c'est le fait qu'il y ait moins de spécification — dans ce cas moins de valeurs de traits — qui détermine la relation d'inclusion entre ensembles de traits et la relation de subsumption entre catégories, pensées comme des symboles constitués de symboles; dans ce cas valeurs de traits.

[va, 1, f, is]	(notée de manière compacte <i>Ivafis</i>)
[va, 1, f, im]	(notée de manière compacte <i>Ivafim</i>)
[va, 2, f, is]	(notée de manière compacte <i>Ivafis</i>)
[va, 2, f, im]	(notée de manière compacte <i>Ivafim</i>)

et qu'un modèle présente la suite

ne vaf pas

Les modèles complets, qui présentent les catégories maximales dans lesquelles sont incluses les catégories non maximales du modèle précédent, sont, en notation compacte, les suivants :

ne 1vafis pas
 ne 1vafim pas
 ne 2vafis pas
 ne 2vafim pas

Pour obtenir les suites des mots du français, chaque catégorie d'un modèle doit être remplacée par toutes les entrées dont le signifiant est associé à la catégorie du modèle, si celle-ci est maximale, ou à la catégorie maximale qui est subsumée par la catégorie du modèle. Ainsi, parmi les suites des mots du français associés aux modèles ci-dessus on a :

(le garçon) *ne regarde pas* (la fleur)
ne regarde pas (la fleur)
 (la pluie) *n'abonde pas*
n'abonde pas (en excuses)

Le compactage dans un symbole unique des traits qui notent les catégories non maximales est considéré directement interprétable : *Ivep* est le symbole qui note une catégorie avec les valeurs de traits : *I, ve, p*; *êtréf* est le symbole qui note la catégorie avec *être* et *f* comme valeurs de traits, et ainsi de suite.

3.2. PROPRIÉTÉS D'EXISTENCE

Les Propriétés d'existence expriment les catégories que l'on peut reconnaître dans un modèle et les relations d'implication d'existence entre ces catégories. Ces Propriétés n'expriment aucune relation d'ordre. Formellement elles ne spécifient que des ensembles de catégories.

Les Propriétés d'existence s'expriment moyennant les prédicats suivants : *amod*, *oblig* ou *obligdi*, *uniq*, *exig*, *exigac*, *exclu*. Le premier argument de chaque prédicat identifie le label du modèle pour lequel le prédicat est pertinent. Dans ce travail le premier argument de tous les

prédicats présentés de § 4 à § 7 est ainsi *phn* (abréviation de phrase (verbale) noyau).

amod sert à définir toutes et seules les catégories que l'on peut utiliser dans tous les modèles. P. ex. avec

amod(id, [cat1, cat2, ..., catn])

on dit :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, on ne peut utiliser que les catégories subsumées par *cat1* ou *cat2* ou par toute autre catégorie notée dans le deuxième argument de *amod*, et pour chacune des catégories notées dans le deuxième argument de *amod* il y a au moins un modèle *id* dans lequel on a une catégorie subsumée par la catégorie notée dans le deuxième argument de *amod*.

P. ex dans le Niveau 1 de la description de la phrase verbale noyau du français on a

amod(phn, [auxf, v, être, cl~n]).

Cette Propriété dit que dans une suite de la phrase verbale noyau du français, dans les limites d'observation du Niveau 1, on peut trouver une catégorie qui est subsumée par une des catégories *auxf* [forme fléchie de l'auxiliaire], *v* [forme verbale, celle-ci pouvant donc être, grâce à la relation de subsomption *ve*, ou *va*, etc.], *être* [forme *être*, celle-ci pouvant être *été* ou une forme fléchie] ou *cl~n* [forme de clitique non nominatif, le Niveau 1 n'incorporant pas les clitiques nominatifs. Dans le Niveau 2, (cf. § 7.2), qui incorpore les clitiques nominatifs, *amod* du Niveau 1 deviendra

amod(phn, [auxf, v, être, cl, ...]).

Avec cette deuxième formulation, dans laquelle la catégorie *cl* subsume aussi bien les clitiques non nominatifs que les clitiques nominatifs, on peut spécifier les modèles illustrés par les exemples (i) à (iii) qui suivent :

- i (Jacques) *les regarde* (souvent)
- ii (Jacques) *regarde-t-il* (souvent)
- iii (Jacques) *les regarde-t-il* (souvent)

uniq exprime l'exigence d'utilisation unique d'une catégorie dans un modèle. L'exemple d'utilisation de *uniq* qui suit

uniq(id, [cat1, cat2, ..., catn]).

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, s'il y a une catégorie subsumée par *cat1* ou par *cat2*, ou par toute autre catégorie notée dans le deuxième argument de *uniq*, il y a une seule catégorie subsumée respectivement par *cat1* ou par *cat2* ou par toute autre catégorie notée dans le deuxième argument de *uniq*.

P. ex dans le Niveau 1 de la description de la phrase verbale noyau du français on a

$uniq(\text{phn}, [v, \text{être}, \dots])$.

Cette Propriété dit que dans une suite de la phrase verbale noyau du français, s'il y a une forme verbale, il y a une seule forme verbale et que s'il y a une forme *être*, alors il y a une seule forme *être* (et ainsi de suite pour les symboles ici remplacés par les points de suspension).

oblig et **obligdi** sont utilisés disjonctivement (c'est-à-dire que l'on utilise l'un ou l'autre de ces deux prédicats); ils sont utilisés pour exprimer la catégorie qui sera considérée comme le noyau du modèle *id*. Le premier est employé lorsqu'il n'y a qu'une catégorie possible comme noyau (p. ex. pour les modèles du syntagme adjectival); le deuxième lorsqu'il y a plusieurs candidatures de catégories, chacune étant disjonctivement possible par rapport aux autres. C'est celui qu'on utilise pour la phrase verbale noyau du français, car dans l'analyse ici proposée plusieurs catégories peuvent être utilisées comme noyau, bien que dans chaque modèle il n'y en ait qu'une seule catégorie noyau. Pour noter le noyau, s'il est nécessaire de le distinguer en tant que tel, on utilise le symbole $^{\circ}$ antéposé à la catégorie; ainsi p. ex. $^{\circ}catX$. L'exemple d'utilisation de *obligdi* qui suit

$obligdi(id, [cat1, cat2, \dots, catn])$.

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, il y a un seul noyau, et celui-ci est une catégorie subsumée par *cat1* ou une catégorie subsumée par *cat2*, ou une catégorie subsumée par une autre catégorie notée dans le deuxième argument de *obligdi*, et pour toute catégorie notée dans *obligdi*, il y a au moins un modèle *id* dans lequel on a comme noyau une catégorie subsumée par cette catégorie dans *obligdi*.

P. ex. pour décrire le syntagme nominal noyau du français — *snnfr* — et du portugais¹⁰ — *snp* — on a besoin, respectivement, des *obligdi* suivants

obligdi(*snnfr*, [n, card, ...]).
obligdi(*snp*, [n, card, dem, ...]).

Dans les deux langues, on peut avoir un syntagme nominal noyau avec un nom ou un cardinal comme noyau. Mais une différence entre les deux langues est que le français n'admet pas un démonstratif comme noyau, alors que le démonstratif peut être noyau en portugais. Exemples :

Ces °garçons sont venus
 *Ces sont venus
 Estes °rapazes vieram
 °Estes vieram

exig exprime les exigences d'inclusion d'un sous-ensemble de catégories dans un modèle motivées par la présence dans le même modèle d'un autre sous-ensemble de catégories. L'exemple d'utilisation de *exig* qui suit

*exig*nd. ([cat1], [cat2]).

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, s'il y a un sous-ensemble avec une catégorie subsumée par *cat1*, il y a aussi un autre sous-ensemble avec une catégorie subsumée par *cat2*.

P. ex pour décrire la phrase verbale noyau du français on a besoin de

exig(*phn*, [[été], [auxf]]).

Avec cette Propriété, on exprime quelques caractéristiques des constructions passives et, plus généralement, des formes avec *été*. On y dit : si dans une phrase verbale noyau du français on a la forme *été*, dans cette phrase verbale il y a aussi une forme fléchie¹¹ de l'auxiliaire. Exemples : *il a été envoyé*, *il a été content*.

Dans l'exemple l'exigence porte sur un seul ensemble. Mais le formalisme permet d'exprimer des exigences sur l'un ou l'autre de plusieurs ensembles (disjonction non exclusive). Ainsi, p. ex. on peut écrire des *exig* comme suit :

10. Les exemples du portugais sont extraits de [Hagège et Bès 98].

11. La restriction "fléchie" est introduite par les conditions d'observation définies dans ce travail (cf. § 2); dans une description complètement exhaustive des phrases verbales noyau du français cette restriction n'a pas lieu d'être.

exig(id, [[cat1], [cat2], ..., [cat3]]).

Ici l'exigence de l'ensemble noté *[cat1]* (le premier noté sur la liste du deuxième argument) porte sur la coexistence nécessaire dans le modèle de *[cat2]* ou de *[cat3]*, ou d'une catégorie notée à droite de *[cat1]* sans exclure, pour autant, la coexistence de *[cat2]* et de *[cat3]* (si *[cat2]* et *[cat3]* ne pouvaient pas coexister dans un modèle, ceci serait indiqué dans une autre Propriété, de type *exclu*, cf. ci-dessus), avec, éventuellement, toute autre catégorie notée à droite de la première.

P. ex. dans la description du syntagme nominal noyau du portugais (snnpt) on a le *exig* suivant

exig(snnpt, [[pos], [artd], [dem]]).

car en portugais (du Portugal) on ne peut pas avoir dans le syntagme nominal noyau un possessif sans un article défini ou un démonstratif, alors qu'une Propriété analogue n'est pas adéquate pour les deux possessifs du français. Ainsi en portugais on observe :

Este meu livro
O meu livro
*Meu livro

tandis qu'en français on a

*Ces mes livres
*Les mes livres
Mes livres
Les miens livres [à peine attesté]

Dans les ensembles présentés ci-dessus, l'ensemble impliquant — celui noté en premier — et chacun des ensembles impliqués dans la disjonction non exclusive, sont des ensembles à un seul élément. Mais le formalisme permet d'exprimer un ensemble impliquant et un ou plusieurs ensembles impliqués à plus d'un élément, pourvu qu'ils soient des ensembles finis (exemples ci-dessous).

Les exigences d'accord, lorsqu'elles se combinent avec les exigences de coexistence — c'est-à-dire les exigences sur l'ajout d'au moins une catégorie à l'ensemble impliquant — se notent directement dans les *exig*. En revanche, si elles n'exigent pas nécessairement une coexistence supplémentaire par rapport à l'ensemble impliquant, elles s'expriment moyennant *exigac* (cf. ci-dessous).

Dans tous les cas, ce sont des majuscules qui notent l'accord. Les propriétés d'accord sont analysées dans ce travail, de manière assez tradition-

nelle¹² en GNP (genre, nombre, personne). L'accord est indiqué soit s'il est exigé soit s'il est interdit. Notation :

Accord exigé : variables notées par des majuscules identiques.

Accord interdit : variables notées par des majuscules identiques dont l'une est niée par le symbole ~.

Accord libre : les traits d'accord ne sont pas indiqués ou ils sont notés par des majuscules différentes.

P. ex dans la description de la phrase verbale noyau du français on a les *exig* suivants :

i *exig*(phn, [[clrP, êtreP], [1vap], [2vep]]).

ii *exig*(phn, [[le, êtreNP], [clrNP]]).

L'*exig* de (i) exprime l'exigence introduite par un ensemble avec les catégories de clitique réfléchi et une forme fléchie de *être* qui s'accordent en personne (p. ex. dans *me suis*) sur la coexistence soit avec un participe d'un verbe *va*, sous-classe 1 (p. ex. *regardé(e)*) soit avec un participe d'un verbe *ve*, sous-classe 2 (p. ex. *agenouillé(e)*).

L'*exig* de (ii) exprime l'exigence introduite par un ensemble avec les catégories de clitique objet et une forme fléchie de *être* sur la coexistence avec un clitique réfléchi s'accordant en nombre et personne avec la forme fléchie de *être* (pour, p. ex., spécifier *me le suis acheté* et éviter **le suis acheté* sans la forme du clitique réfléchi).

exigac exprime les exigences d'accord lorsqu'elles n'introduisent pas nécessairement des exigences de coexistence; formellement cette Propriété est de même type que *exig* : si telles catégories existent dans le modèle, alors il faut qu'il y ait ou qu'il n'y ait pas d'accord entre elles, ou que l'accord soit libre.

Les majuscules notant, comme indiqué ci-dessus, l'accord, l'exemple d'utilisation de *exigac* qui suit

exigac(id, [[cat1GN], [cat2GN]]).

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors les catégories *cat1* et *cat2* s'accordent en genre et nombre. Le formalisme permet de noter plus de deux catégories, pourvu qu'elles ne soient pas contradictoires sur les exigences d'accord.

12. Et de manière peu satisfaisante; cf. les limitations introduites sur les formes *on* et *vous* § 4.1, et § 8.2.

De même, il est possible d'exprimer l'exigence de l'inexistence de l'accord. L'exemple d'utilisation de l'*exigac* qui suit

exigac(id, [[cat1GN], [cat2~GN]]).

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors les catégories *cat1* et *cat2* ne peuvent pas s'accorder en genre.

Pour décrire le syntagme nominal noyau du français — *snnfr* — et du portugais — *snnpt* — on a besoin d'*exigac* très simples, comme p. ex. :

exigac(*snnfr*, [[artGN], [nGN]]).

Cet *exigac* indique simplement qu'un article s'accorde en genre et nombre avec le nom. En revanche, l'analyse présentée ci-dessous § 4 de la phrase verbale noyau en français a exigé des Propriétés *exigac* moins triviales, p. ex. :

- i *exigac*(*phn*, [auxfP, clr-P]).
- ii *exigac*(*phn*, [auxfN, été, 1vapN]).

L'*exigac* de (i) interdit qu'un auxiliaire fléchi ait la même personne qu'un clitique réfléchi; on évite ainsi **vous t'avez regardé* (cf. dans § 7 la possibilité d'exprimer ce type d'exigence à différents niveaux de finesse).

L'*exigac* de (ii) dit ceci : si dans une phrase verbale noyau il y a une forme d'auxiliaire fléchie, la forme *été* et un participe d'un verbe *va*, sous-classe 1, alors l'auxiliaire et le participe doivent s'accorder en nombre (p. ex. dans *elles ont été regardées*).

exclu exprime les exigences d'exclusion d'un sous-ensemble de catégories dans un modèle, motivées par la présence dans le même modèle d'un autre sous-ensemble de catégories. Sur le plan formel, *exclu* combine l'implication et la négation. L'exemple d'utilisation de *exclu* qui suit

exclu(id, [[cat1], [cat2], ..., [catn]]).

doit s'interpréter :

dans tous les modèles qui satisfont les Propriétés d'existence *id*, s'il y a un sous-ensemble avec une catégorie subsumée par *cat1*, on ne peut pas avoir un autre sous-ensemble avec une catégorie subsumée par *cat2* ou par toute autre catégorie notée dans *exclu*.

P. ex dans la description de la phrase verbale noyau du français on a les *exclu* suivants :

- i $\text{exclu}(\text{phn}, [[\text{lui}], [\text{y}]])$.
- ii $\text{exclu}(\text{phn}, [[\text{\textasciitilde}tref], [\text{auxf}], [1\text{vaf}], [2\text{vaf}], [1\text{vef}], [2\text{vef}]])$.

L'*exclu* de (i) interdit qu'un clitique *lui* coexiste avec un clitique *y*. L'*exclu* de (ii) interdit la coexistence de deux quelconques des catégories qui y sont incorporées : on y dit que deux formes verbales quelconques fléchies ne peuvent pas coexister dans la même phrase verbale noyau.

De même que pour *exig*, le formalisme autorise à utiliser des ensembles avec plus d'un élément. Par ailleurs, on peut également noter dans l'expression plusieurs ensembles, de telle manière que l'exclusion est exigée pour une paire quelconque de ces ensembles. Ainsi, p. ex. on peut écrire des *exclu* comme celui qui suit :

$\text{exclu}(\text{id}, [[\text{cat1}, \text{catn}], [\text{cat2}, \text{catm}], [\text{cat3}]])$.

Ici l'exigence d'exclusion vaut pour toutes les paires de ces trois ensembles, c'est-à-dire les deux membres de chacune des paires $\langle [\text{cat1}, \text{catn}] \text{ et } [\text{cat2}, \text{catm}] \rangle$, $\langle [\text{cat1}, \text{catn}] \text{ et } [\text{cat3}] \rangle$, $\langle [\text{cat2}, \text{catm}] \text{ et } [\text{cat3}] \rangle$ ne peuvent pas coexister.

3.3. PROPRIÉTÉS DE LINÉARITÉ

Si les Propriétés d'existence définissent les ensembles de catégories susceptibles de former les modèles, les Propriétés de linéarité introduisent les relations d'ordre entre les éléments des ensembles ainsi définis. C'est-à-dire que, formellement, les Propriétés de linéarité transforment les ensembles en listes.

On utilise deux schémas d'expressions, illustrés par les exemples suivants, pour exprimer les Propriétés de linéarité :

- i $\langle \text{id}, \{ _ \}, \text{précède}(\text{cat1}, \{\text{cat2}\}) \rangle$.
- ii $\langle \text{id}, \{ \{ \text{catm} \} \setminus \{ \text{catn} \} \}, \text{précède}(\text{cat1}, \{\text{cat2}\}) \rangle$.

L'exemple (i) s'interprète ainsi :

dans tout modèle *id* s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat2*. Ce schéma est utilisé, p. ex., pour exprimer que la forme *été* précède une forme de participe.

L'exemple (ii) s'interprète ainsi :

dans tout modèle *id* où il y a une catégorie subsumée par *catm* et où il n'y a pas une catégorie subsumée par *catn*, s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat2*.

Dans (ii), le symbole $\{ _ \}$ de (i) a été remplacé par $\{ \{ \textit{catm} \} \setminus \{ \textit{catn} \} \}$. Cette dernière expression permet d'introduire des exigences sur les ensembles sur lesquelles vont opérer les exigences de linéarité. En général on a

{AVEC \ SANS}

Cela veut dire qu'à droite de l'antislash on note les catégories qui ne peuvent subsumer aucune catégorie dans le modèle et à gauche de l'antislash on note les catégories qui subsument obligatoirement des catégories dans le modèle.

Le deuxième argument de *précède*, aussi bien dans (i) que dans (ii) ci-dessus, est noté entre les accolades $\{ \}$; ce symbole note ainsi un ensemble avec un seul élément (en l'occurrence *cat2*). Mais dans cet ensemble on peut avoir plusieurs éléments. Dans ce cas on dira que *cat1* précède chacune des catégories qui y sont notées. De cette manière, l'exemple

iii $\langle \textit{id}, \{ _ \}, \textit{précède}(\textit{cat1}, \{ \textit{cat2}, \textit{cat3}, \dots, \textit{catn} \}) \rangle$.

s'interprète ainsi :

dans tout modèle *id* s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat2*, s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat3*, alors la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat3*. En général, s'il y a *cat1* et toute catégorie entre accolades, *cat1* précède cette catégorie.

Il est à remarquer que l'ordre dans la notation des catégories dans l'ensemble n'est pas pertinent, de telle manière que (iii) exprime exactement la même Propriété que

iv $\langle \textit{id}, \{ _ \}, \textit{précède}(\textit{cat1}, \{ \textit{cat3}, \textit{cat2} \}) \rangle$.

P. ex dans la description de la phrase verbale noyau du français (Niveau 1), on a les *précède* suivants :

i $\langle \textit{phn}, \{ _ \}, \textit{précède}(\textit{été}, \{ \textit{p} \}) \rangle$.

ii $\langle \textit{phn}, \{ _ \}, \textit{précède}(\textit{le}, \{ \textit{lui}, \textit{en}, \textit{y} \}) \rangle$.

La Propriété *précède* dans (i) dit que s'il y a une forme *été* et un participe (noté par *p*), alors la forme *été* précède le participe.

La Propriété *précède* dans (ii) dit

- s'il y a une forme du clitique *le* et une forme du clitique *lui*, alors la forme du clitique *le* précède la forme du clitique *lui*
- s'il y a une forme du clitique *le* et le clitique *en*, alors la forme du clitique *le* précède le clitique *en*
- s'il y a une forme du clitique *le* et le clitique *y*, alors la forme du clitique *le* précède le clitique *y*.

3.4. PROPRIÉTÉS DE FLÉCHAGE

Les Propriétés d'existence et les Propriétés de linéarité ont permis de spécifier des ensembles de listes. Les Propriétés de fléchage vont permettre d'exprimer des relations entre les éléments disposés dans ces listes. On appelle souvent ces relations des *dépendances*, sans savoir très bien ce qu'est formellement une dépendance : de manière très vague, on utilise en général *dépendance* pour parler d'un lien quelconque entre deux éléments d'une phrase; ici on utilise fléchage pour se démarquer de cet usage vague, bien que la notion de fléchage soit proche parente de celle de dépendance.

Les Propriétés de fléchage seront utilisées pour calculer les représentations sémantiques. Si plusieurs catégories flèchent sur une autre catégorie (celle-ci étant souvent mais non toujours le noyau), les catégories qui flèchent et la catégorie sur laquelle elles flèchent peuvent être formellement considérées comme les arguments d'une fonction qui permet de calculer la représentation sémantique qui doit être associée à ces catégories. Dans ce travail les fonctions calculées à partir du fléchage ne sont pas présentées ; on indique simplement la possibilité de leur formulation, cf. § 6.

Pour exprimer les Propriétés de fléchage on utilise les schémas d'expressions (i) à (iv), illustrés par les exemples suivants :

- i <id, {_i}, flèche(cat1_i, [cat2_j])>.
- ii <id, {{cat_m}\{cat_n}}, flèche(cat1_k, [cat2_l])>.
- iii <id, {_i}, flèche(cat1_i, [cat2_j, cat3])>.
- iv <id, {{cat_m}\{cat_n}}, flèche(cat1_k, [cat2_l, cat3])>.

Les Propriétés de fléchage portent sur des éléments déjà ordonnés dans des listes. Elles peuvent donc s'exprimer en utilisant les relations d'ordre. Celles-ci seront notées par l'ordre alphabétique des symboles utilisés dans les indices, mais l'utilisation d'indices — et donc de l'ordre entre les éléments qui flèchent, n'est pas formellement exigée.

L'exemple (i) s'interprète ainsi :

dans tout modèle *id* s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, et si la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat2*, alors la catégorie subsumée par *cat1* flèche sur la catégorie subsumée par *cat2*. Ce schéma est utilisé, p. ex., pour exprimer qu'une forme d'auxiliaire fléchie flèche sur la forme *été*.

En revanche, l'exemple (v)

v <id, {₋}, flèche(cat1, [cat2])>.

s'interprète ainsi :

dans tout modèle *id* s'il y a une catégorie subsumée par *cat1* et une catégorie subsumée par *cat2*, alors la catégorie subsumée par *cat1* flèche sur la catégorie subsumée par *cat2*.

L'exemple (ii) utilise la notation {AVEC\ SANS} déjà présentée ci-dessus. Comme pour (i) on peut utiliser des indices ou non pour exprimer les relations d'ordre.

Les exemples (iii) et (iv) notent [*cat2j*, *cat3*] comme deuxième argument de *flèche*. Ceci ne veut nullement dire qu'une catégorie flèche sur deux catégories différentes dans un même modèle, mais exprime une ambiguïté de fléchage, et donc une ambiguïté d'interprétation sémantique du modèle.

L'exemple (iii) s'interprète ainsi :

dans tout modèle *id* s'il y a une catégorie subsumée par *cat1*, une catégorie subsumée par *cat2* et une catégorie subsumée par *cat3*, alors ou bien, si la catégorie subsumée par *cat1* précède la catégorie subsumée par *cat2*, la catégorie subsumée par *cat1* flèche sur la catégorie subsumée par *cat2*, ou bien la catégorie subsumée par *cat1* flèche sur la catégorie subsumée par *cat3*.

Il est clair que (iv) est passible du même type d'interprétation.

Quelques exemples dans la description de la phrase verbale noyau du français. Dans le Niveau 1, on a, parmi d'autres, les *flèche* suivants :

i <phn, {₋}, flèche(auxf, [été])>.
ii <phn, {\ {été}}, flèche(auxf, [p])>.

La Propriété *flèche* dans (i) dit que s'il y a une forme *auxf* et la forme *été*, la forme *auxf* flèche sur la forme *été*. La Propriété *flèche* dans (ii) introduit une condition. Elle dit que s'il y a une forme *auxf* et un participe, pourvu qu'il n'y ait pas la forme *été*, la forme *auxf* flèche sur le participe.

Dans le Niveau 5, on a, parmi d'autres, le *flèche* qui suit :

<phn, {Ne[_pl,_]}, flèche(tous, [°X])>.

Cette Propriété *flèche* dit que s'il n'y a pas une forme de clitique *les*, la forme *tous* flèche sur le noyau.

Les Propriétés *flèche*, comme d'ailleurs toutes les Propriétés, se présentent écrites de manière ordonnée, mais l'ordre d'écriture n'est pas significatif : p. ex., l'ordre dans l'écriture de (i) et (ii) du Niveau 1 pourrait s'inverser et le résultat serait le même.

4. LES PROPRIÉTÉS DE LA PHRASE VERBALE NOYAU EN FRANÇAIS

En utilisant le formalisme présenté dans § 3, les Propriétés (P2) du Niveau 1 sont présentées ci-dessous dans § 4.2, 4.3 et 4.4. Les extensions exigées par les Niveaux 2 à 5 sont présentées dans § 7. Le lexique des catégories utilisées dans la formulation des (P2), dans tous les Niveaux, est présenté dans § 4.1. Dans chaque niveau des conventions sur ce qu'on accepte dans le lexique des catégories et une parmi les Propriétés — celle qui est exprimée par le prédicat *amod* — explicitent les catégories utilisées dans ce niveau. Puisque dans le Niveau 3 on propose une description pour les phrases verbales noyau ayant incorporé des formes à l'impératif, et ceci dans deux versions, on se donne ici les catégories nécessaires à la formulation de cette double description, les catégories nécessaires à la description puriste étant dans § 4.1 signalées par un astérisque.

Les traits et les catégories proposées n'ont pas la prétention d'être LES BONNS traits et LES BONNES catégories. Ils sont plutôt conçus comme une pièce d'un engrenage plus vaste et plus compliqué qui les utilisent et qui doit être évalué dans son ensemble.

4.1. LE LEXIQUE DES CATÉGORIES

Les catégories sont définies à l'aide des valeurs de traits (cf. § 3.1). Les traits utilisés sont présentés par la suite.

• *Vocabulaire des traits*

vx ; {+, -}

v ; {+, -}

aux ; {+, -}

être ; {+, -}

été ; {+, -}

cl; {+, -}
 cln; {+, -}
 clr; {+, -}
 se; {+, -}
 sousclasse-v; {ve, va}
 sousclasse-ve; {1, 2}¹³
 sousclasse-va; {1, 2}
 sousclasse-cl; {1, 2}
 forme-flexion; {f, p} [f : forme fléchie ; p : participe]
 sousclasse-forme-fléchie; {is, im} [is : indicatif-subjonctif ; im : impératif]
 valeurs-clit; {l', le, lui, y, en}
 sousclasse-valeurs-clit; {-l, 'y, 'en, fo, fa} [fo : forte ; fa : faible]

• Hiérarchie des traits

Les traits du Vocabulaire ci-dessus s'organisent et s'utilisent dans la hiérarchie qui suit, indiquée par l'indentation.

vx	[toute forme verbale]
eu	[forme <i>eu</i> pour l'auxiliaire surcomposé]
être	[<i>être</i> n'est ni auxiliaire ni verbe] ¹⁴
été	
f	[formes fléchies de <i>être</i>]
is	
im	
aux	
f	[formes fléchies de <i>avoir</i> , qui est le seul auxiliaire]
is	
v	[cf. ci-dessous le classement des lemmes verbaux]
ve	
1	
f	
is	
im	
p	
2	
f	
is	

13. Rappel : les valeurs des traits de cette étiquette notées "1" et "2" sont, dans une notation complète, respectivement *ve1* et *ve2*; *mutatis mutandis* la même observation est valable pour la notation des valeurs des traits des étiquettes *sous-classe-va* et *sous-classe-cl*; cependant, pour alléger la notation, comme ces valeurs de traits sont toujours utilisées dans des catégories où aucune ambiguïté n'est possible, on note simplement "1" et "2" (cf. les conventions de notation § 3.1).

14. Cf. dans § 7.5 la justification de l'analyse adoptée pour les formes avec *être*; sur l'interprétation du trait *aux* cf. en rappel la note 8 dans § 3.1

			im	
		p		
	va			
		1		
		f	is	
			im	
		2		
		p		
		f	is	
			im	
		p		
cl				
	cln			{-je, -tu, il(s), -elles(s), -nous, -vous}
	cl~n			[clitiques non nominatifs]
		1		[clitiques à gauche de la forme verbale]
		clr		[clitiques réfléchis]
		se		{se, s'}
		~se		{me, m',te, t', nous, vous}
		cl~r		[clitiques non réfléchis]
		l'		{l', le}[Pour <i>Jacques l'a été</i>]
		le		{le, l', la, les} [Pour <i>Jacques la regarde</i>]
		lui		{lui, leur} [Pour <i>Il lui parle, il lui court après</i>]
		y		{y} [Pour <i>Il y va, il y ajoute</i>]
		en		{en} [Pour <i>Il en parle, il en connaît cinq</i>]
		2		[clitiques à droite d'une forme verbale à l'impératif]
		clr		
		fo		[Formes fortes, pour <i>Parle-moi, parle-nous</i>]
		fa		[Formes faibles, pour <i>Va-t'en</i>]
		cl~r		
		l'		{-le}[Pour <i>Sois-le</i>]
		le		
		~l		{-le, -la, -les}[Pour <i>Regarde-le</i>]
		-l		{-l}[Pour <i>Accompagne-l'y</i>]
		y		
		~'y		{-y}[Pour <i>Vas- y</i>]
		'y		{-l}[Pour <i>Accompagne-l'y</i>]
		en		
		~'en		{-en}[Pour <i>Prends -en</i>]
		'en		{'en}[Pour <i>Va-t'en</i>]

Pour pouvoir traiter les formes *ne*, *pas*, *toujours* et *tous* on ajoute des traits booléens *ad hoc* qui seront utilisés toujours avec la valeur + et qui sont respectivement notés *ne*, *pas*, *tj* et *tous*.

La hiérarchie des traits détermine les relations d'héritage. P. ex. la catégorie qui sera associée aux formes du clitique objet *le* associé aux formes orthographiques *le*, *la*, *les*, *l'* (dans, p. ex., *le regarde*, *la regarde*, *les regarde*, *l'élabore*) héritent, étant donné la hiérarchie des traits, des valeurs de traits : *cl*, *cl-n*, *l*, *cl-r*, *le*.

Le vocabulaire des traits conjointement avec la hiérarchie des traits et la relation d'héritage permettent de définir le lexique des catégories maximales. Les Propriétés (P2) sont exprimées à l'aide des catégories qui subsument ces catégories maximales : c'est-à-dire que les ensembles des valeurs de traits utilisées dans l'expression des P2 sont inclus dans les catégories maximales présentées par la suite.

Dans ce travail, on suppose que chaque entrée lexicale — c'est-à-dire chaque mot, considéré en tant que signifiant orthographique (suite des caractères) — est associée à une seule catégorie du lexique des catégories maximales (assignation non ambiguë) ou à plus d'une de ces catégories (assignation ambiguë). Ainsi par exemple, pour le type de descriptions ici proposées, le mot *se* aura une assignation non ambiguë, tandis que *été* et *juge* auront des assignations ambiguës.

Une très grande majorité des lemmes verbaux du français peut se partitionner de la manière suivante :

- lemmes dont la catégorie maximale à laquelle ils sont associés comporte les traits *1ve*. Ce sont des lemmes qui exigent dans les formes composées une forme *être* et qui ne possèdent pas de construction réflexive. Ex. : *arriver*.
- lemmes dont la catégorie maximale à laquelle ils sont associés comporte les traits *2ve*. Ce sont des lemmes qui exigent dans les formes composées une forme *être* et qui, avec les traits *2ve* ne s'attestent que dans des constructions réflexives. Ex. : *s'agenouiller*.
- lemmes dont la catégorie maximale à laquelle ils sont associés comporte les traits *1va*. Pour qu'un lemme soit associé à la catégorie *1va* il faut et il suffit qu'il puisse être attesté dans au moins une construction avec complément d'objet direct. Ex. : *chanter*, *donner*, *promener*, *commencer*, *laisser*.
- lemmes dont la catégorie maximale à laquelle ils sont associés comporte les traits *2va*. Pour qu'un lemme soit associé à la catégorie *2va* il faut et il suffit qu'il n'admette pas de forme composée avec *être*. (À

noter que selon la définition des lemmes de la classe précédente, aucune des constructions possibles d'un lemme associé à la classe *2va* ne doit être un complément d'objet direct). Ex. : *abonder*.

Il existe un nombre réduit de lemmes qui n'entrent pas dans cette partition, car ils relèvent de deux classes différentes, p. ex. le lemme *monter*, qui est associé aux deux classes *1ve* et *1va*.

Les valeurs des traits des catégories maximales sont présentées en colonne et selon les conventions établies dans § 3.1. Pour rendre plus facile la lecture des *P2*, en particulier des Propriétés d'existence (cf. ci-dessous § 4.2) au dessus de chaque colonne on a noté, en caractères gras, la catégorie en notation compacte qui est fréquemment utilisée dans l'expression de ces Propriétés. On remarquera qu'un même symbole en gras — qui note une catégorie non maximale — peut être placé au dessus de plus d'une catégorie maximale : ceci est dû au fait que le symbole en gras note un ensemble de traits qui est inclus dans plusieurs catégories maximales différentes. Exemple :

1vaf	1vaf
vx	vx
v	v
va	va
l	l
f	f
is	im

Les deux colonnes spécifient deux catégories maximales dont les valeurs de traits sont identiques à l'exception du dernier : celle de gauche est *is* (indicatif-subjonctif) alors que celle de droite est *im* (impératif), les deux catégories étant des formes verbales (*vx*), des verbes — et non des auxiliaires — (*v*), de la sousclasse-*v* (*va*), de la sousclasse-*va* (*l*) et des formes fléchies (*f*). Les mots *regarde, regardais, regardons, regardera...* à l'indicatif et au subjonctif, sont, parmi d'autres, associés dans le lexique des entrées à cette catégorie maximale. Les mots *regarde, regardez, ...* à l'impératif sont, parmi d'autres, associés dans le lexique des entrées à la catégorie maximale notée dans la colonne de droite.

Pour les catégories maximales autres que celles des formes verbales, à droite du symbole en gras on indique entre accolades les signifiants orthographiques des entrées auxquelles elles sont associées.

• *Catégories maximales des formes verbales*

auxf	eu	êtréf	êtréf	été
vx	vx	vx	vx	vx
aux		être	être	être
f	eu	f	f	été
is		is	im	

1vef	1vef	1vep	2vef	2vef	2vep
vx	vx	vx	vx	vx	vx
v	v	v	v	v	v
ve	ve	ve	ve	ve	ve
1	1	1	2	2	2
f	f	p	f	f	p
is	im		is	im	

1vaf	1vaf	1vap	2vap	2vap	2vap
vx	vx	vx	vx	vx	vx
v	v	v	v	v	v
va	va	va	va	va	va
1	1	1	2	2	2
f	f	p	f	f	p
is	im		is	im	

• *Catégories maximales des formes clitiques*

Les symboles en gras notent ici une seule catégorie maximale. Le système ne rend pas compte de la bonne utilisation des consonnes de liaison.

1l' {l', le}	2l {-le}*	
cl	cl	
1	2	
cl~n	cl~n	
cl~r	cl~r	
l'	l'	
1le {l', le, la, les}	2le {-le,-la, -les}	-l {-l}*
cl	cl	cl
1	2	2
cl~n	cl~n	cl~n
cl~r	cl~r	cl~r
le	le	le
	~l	-l

1lui {lui, leur}	2lui {-lui, -leur}		
cl	cl		
1	2		
cl~n	cl~n		
cl~r	cl~r		
lui	lui		
1y {y}	2y {-y, -z-y}	'y {'y}*	
cl	cl	cl	
1	2	2	
cl~n	cl~n	cl~n	
cl~r	cl~r	cl~r	
y	y	y	
	~'y	'y	
1en(en)	2en {-en, -z-en}	'en {'en}*	
cl	cl	cl	
1	2	2	
cl~n	cl~n	cl~n	
cl~r	cl~r	cl~r	
en	en	en	
	~'en	'en	
1clr~se {me, m', te, t', nous, vous}	1se {se}	2cl {-moi, -toi, -nous, -vous}	2clr {-m, -t}*
cl	cl	cl	cl
1	1	2	2
cl~n	cl~n	cl~n	cl~n
~se	se	fo	fa
clr	clr	clr	clr
cln {-je, -tu, -il, -elle, -nous, -vous, -ils, -elles}			
cl			
cln			

Catégories maximales des formes ad-hoc

ne {ne, n'}	pas {pas}	tj {toujours}
ne	pas	tj

On indique par la suite les propriétés d'accord pour les différentes catégories; le symbole '_' indique que la forme à laquelle ce symbole est associé ne possède pas de valeur d'accord pour le trait correspondant, cette forme pouvant s'accorder, en ce qui concerne ce trait, avec un trait

qui possède une valeur quelconque (ainsi une forme avec ‘_’ comme valeur de G peut s’accorder avec les valeurs *m* et *f* de G d’une autre forme).

1vapGN	auxfNP				
1vepGN	êtréfNP	im	clrNP	clnGNP	leGN
2vepGN	is				
[m, sg, _]	[_ , sg, 1]		[_ , sg, 1]	[_ , sg, 1]	[m, sg, _]
[f, sg, _]	[_ , sg, 2]		[_ , sg, 2]	[_ , sg, 2]	[f, sg, _]
[m, pl, _]	[_ , sg, 3]	{ _ , sg, 2}	[_ , _ , 3]	[f, sg, 3]	[_ , pl, _]
[f, pl, _]	[_ , pl, 1]	{ _ , pl, 1}	[_ , pl, 1]	[f, sg, 3]	
	[_ , pl, 2]	[_ , pl, 2]	[_ , pl, 2]	[_ , _ , 1]	
ou bien :	[_ , pl, 3]			[_ , _ , 2]	
[_ , _ , _] ¹⁵				[m, pl, 3]	
				[f, pl, 3]	

• *Quelques remarques*

Dans cette version nous ne considérons pas les clitiqes nominatifs *-on*, *-vous* [- , sg, 2], l’impersonnel *-il* et le pronom *-ce* (cf. § 2).

On a essayé d’utiliser au maximum des traits booléens; c’est ainsi que l’on peut noter la valeur positive du trait directement par l’étiquette elle-même.

Les symboles en gras (notation compacte) ne sont pas les seuls qui seront utilisés dans l’expression des Propriétés. Pour noter celles-ci, on peut utiliser n’importe quel ensemble non vide de valeurs de traits. En fait, pour exprimer les formes verbales, on utilise effectivement *f* — pour capter toutes les formes verbales fléchies —, *vx* — pour capter toutes les formes verbales —, *p* — pour capter toutes les formes verbales qui sont des participes, etc. Pour exprimer les formes clitiqes en notation compacte, on utilise fréquemment les symboles **1cl~n**, **1cl~r**, **le**, **lui**, **y**, **en**, ... Le symbole **1cl~n** note une catégorie qui possède les valeurs de traits *l* et *cl~n*; cette catégorie subsume donc les catégories maximales notées ci-dessous par *ll’*, *lle*, *llui*, *ly*, *len*, *1clr~se*, *lse*. De même le symbole **le** note une catégorie qui possède la valeur de trait *le*; cette catégorie subsume donc les catégories maximales notées ci-dessous par *lle*, *2le*, *-l*.

15. Les traits [_ , _ , _] sont utilisés pour les catégories avec les traits *2vap* et pour certains lemmes associés aux catégories *lvap* (cf. ci-après § 7.5).

4.2. LES PROPRIÉTÉS

Dans la section qui suit, les Propriétés qui vont permettre de définir les modèles de la phrase verbale noyau sont données en deux versions : l'une respecte la syntaxe de formulation de Propriétés présentée dans § 3. Comme il a été indiqué dans § 1.1 note 3, les formules (syntaxiques) qui expriment les Propriétés possèdent une sémantique et c'est pour cette raison que les formules sont calculables : cela veut dire qu'à partir d'un ensemble de catégories et des formules par lesquelles les Propriétés sont exprimées, on peut calculer quelles sont les suites de catégories qui satisfont les Propriétés. Ceci n'est pas possible à partir des expressions en langue naturelle, pour les raisons déjà signalées dans les citations correspondantes de Bochenski de § 1.2 : des expressions telles que *il existe, il y a, il y a aussi, si ... alors* de la langue naturelle autorisent beaucoup trop d'interprétations différentes, ce qui n'est pas le cas des opérateurs booléens (et simples) qui sont utilisées dans l'interprétation sémantique formelle. Mais ces expressions en langue naturelle donnent une bonne idée de ce que les formules "veulent dire" et montrent, par ailleurs, qu'avec un effort minimum, des linguistes qui ne sont pas habitués à des formalismes — souvent inutilement complexes ou mal présentés — peuvent exprimer beaucoup de choses parfaitement calculables.

4.2.1. Les Propriétés d'existence

Dans le Niveau 1, décrit dans cette section, on suppose un lexique de catégories maximales où toutes les catégories avec la valeur de trait *f* ont la valeur de trait *is* ; c'est-à-dire qu'on ne prend pas en compte les formes verbales à l'impératif. De même, dans ce lexique on suppose que toutes les catégories maximales des formes clitiques sont des formes à gauche de formes verbales; étant donné ceci, on n'utilise pas le trait *sousclasse-cl* ; $\{1,2\}$ dans la notation des catégories avec la valeur de trait *cl*. Dans la colonne de gauche on identifie les Propriétés, mais l'ordre introduit par la numérotation n'est pas pertinent

P1	amod(phn, [auxf, v, être, cl~n]).
P2	uniq(phn, [auxf, v, être, l', clr, le, lui, y, en]).
P3	obligdi(phn, [v, être]).
P+1	exig(phn, [[vepN], [étrefN]]).
P+2	exig(phn, [[1vap], [étref], [auxf]]).
P+3	exig(phn, [[2vap], [auxf]]).
P+4	exig(phn, [[2vepN], [étrefNP, clrNP]]).
P+5	exig(phn, [[2vefNP], [clrNP]]).
P+6	exig(phn, [[été], [auxf]]).

P+7	exig(phn, [[clrP, êtrefP], [1vap], [2vep]]).
P+8	exig(phn, [[seNP], [1vafNP], [2vefNP], [êtrefNP]]).
P+9	exig(phn, [[leGN], [1vaf], [1vapGN]]).
P+10	exig(phn, [[le, êtreNP], [clrNP]]).
Pac1	exigac(phn, [auxfP, clr~P]).
Pac2	exigac(phn, [2vafP, clr~P]).
Pac3	exigac(phn, [auxfN, été, 1vapN]).
Pac4	exigac(phn, [êtrefP, clrP], [êtrefPN, clrPN]).
Pac5	exigac(phn, [1vafP, clrP], [1vafPN, clrPN]).
Pac6	exigac(phn, [[1vapGN, êtref], [1vapGN, êtrefN], [1vapGN, êtrefNIP, clrNIP]]).
P~1	exclu(phn, [[°être], [v]]).
P~2	exclu(phn, [[l'], [v]]).
P~3	exclu(phn, [[été], [ve], [2vap]]).
P~4	exclu(phn, [[êtref], [auxf], [1vaf], [2vaf], [1vef], [2vef]]).
P~5	exclu(phn, [[clr], [lui]]).
P~6	exclu(phn, [[le], [été]]).
P~7	exclu(phn, [[lui], [y]]).
P~8	exclu(phn, [[y], [en]]).
P~9	exclu(phn, [[l'], [cl~n]]).

• Formulation intuitive des Propriétés d'existence

[Dans la suite d'une phrase verbale noyau]

P1	On peut trouver l'auxiliaire, un verbe, une forme <i>être</i> , un clitique non nominatif
P2	Il y a un seul auxiliaire, un seul verbe, une seule forme <i>être</i> , un seul <i>l'</i> , un seul clitique réfléchi, un seul <i>le</i> , un seul <i>lui</i> , un seul <i>y</i> , un seul <i>en</i> .
P3	Il y a obligatoirement un verbe ou une forme <i>être</i> comme noyau
P+1	S'il y a un participe passé d'un verbe <i>ve</i> , alors il y a un être fléchi accordé avec lui en nombre
P+2	S'il y a un participe passé d'un verbe <i>1va</i> , alors il y a un auxiliaire fléchi ou un être fléchi
P+3	S'il y a un participe passé d'un verbe <i>2va</i> , alors il y a un auxiliaire fléchi
P+4	S'il y a un participe passé d'un verbe <i>2ve</i> , alors il y a un être fléchi et un clitique réfléchi accordés en nombre et personne
P+5	S'il y a un forme fléchie d'un verbe <i>2ve</i> , alors il y a un clitique réfléchi accordé avec elle en nombre et personne
P+6	S'il y a <i>été</i> , alors il y a un auxiliaire fléchi
P+7	S'il y a un être fléchi et un clitique réfléchi accordés en personne, alors il y a un participe passé d'un verbe <i>1va</i> ou un participe passé d'un verbe <i>2ve</i>
P+8	S'il y a <i>se</i> , alors il y a un forme fléchie d'un verbe <i>1va</i> ou un forme fléchie d'un verbe <i>2ve</i> ou un être fléchi accordés avec <i>se</i> en nombre et personne.
P+9	S'il y a <i>le</i> , alors il y a une forme fléchie d'un verbe <i>1va</i> ou un participe passé d'un verbe <i>1va</i> accordée en genre et nombre avec <i>le</i>

- P+10 S'il y a *le* et un être fléchi, alors il y a un clitique réfléchi accordé en nombre et personne avec le être fléchi
- Pac1 S'il y a un auxiliaire fléchi et un clitique réfléchi, alors l'auxiliaire fléchi et le clitique réfléchi ne peuvent pas s'accorder en personne
- Pac2 S'il y a une forme fléchie d'un verbe 2va et un clitique réfléchi, alors la forme fléchie du verbe 2va et le clitique réfléchi ne peuvent pas s'accorder en personne
- Pac3 S'il y a un auxiliaire fléchi, *été* et un participe passé d'un verbe 1va, alors l'auxiliaire fléchi et le participe passé du verbe 1va s'accordent en personne
- Pac4 S'il y a un clitique réfléchi accordé en personne avec le être fléchi, alors le clitique réfléchi s'accorde en nombre avec le être fléchi
- Pac5 S'il y a un clitique réfléchi accordé en personne avec une forme fléchie d'un verbe 1va, alors le clitique réfléchi s'accorde en nombre avec la forme fléchie du verbe 1va
- Pac6 S'il y a un participe passé d'un verbe 1va et un être fléchi, alors le participe passé du verbe 1va s'accorde en nombre avec le être fléchi, ou il y a un clitique réfléchi accordé en nombre et personne avec le être fléchi.
- P~1 Une forme *être* noyau et un verbe ne peuvent pas coexister
- P~2 Un clitique *l'* et un verbe ne peuvent pas coexister
- P~3 Deux quelconques parmi les formes *été*, un verbe *ve*, un participe passé d'un 2va, ne peuvent pas coexister
- P~4 Deux quelconques parmi les formes être fléchi, auxiliaire fléchi, 1va fléchi, 2va fléchi, 1ve fléchi, 2ve fléchi, ne peuvent pas coexister
- P~5 Un clitique réfléchi et le clitique *lui* ne peuvent pas coexister
- P~6 Le clitique *le* et *été* ne peuvent pas coexister
- P~7 Le clitique *lui* et le clitique *y* ne peuvent pas coexister
- P~8 Le clitique *y* et le clitique *en* ne peuvent pas coexister
- P~9 Le clitique *l'* et un clitique non nominatif ne peuvent pas coexister

4.2.2. Les Propriétés de linéarité

Dans le Niveau 1, les Propriétés de linéarité sont les suivantes :

- P11 < phn, { _ }, précède(f, { été, p }) >.
- P12 < phn, { _ }, précède(été, { p }) >.
- P13 < phn, { _ }, précède(cl~n, { is }) >.
- P14 < phn, { _ }, précède(clr, { cl~r }) >.
- P15 < phn, { _ }, précède(le, { lui, en, y }) >.
- P16 < phn, { _ }, précède(lui, { en }) >.

• Formulation intuitive des Propriétés de linéarité

[Dans la suite d'une phrase verbale noyau]

- P11 s'il y a une forme fléchie et une forme *été*, alors la forme fléchie précède la forme *été*.

- s'il y a une forme fléchie et un participe, alors la forme fléchie précède le participe.
- P12 s'il y a une forme *été* et une forme *p*, alors la forme *été* précède la forme *p*.
- P13 s'il y a une forme de clitique non nominatif et une forme verbale fléchie, alors la forme de clitique non nominatif précède la forme forme verbale fléchie.
- P14 s'il y a une forme de clitique réfléchi et une forme de clitique non réfléchi, alors la forme de clitique réfléchi précède la forme de clitique non réfléchi.
- P15 s'il y a une forme *le* et une forme *lui*, alors la forme *le* précède la forme *lui*.
s'il y a une forme *le* et une forme *en*, alors la forme *le* précède la forme *en*.
s'il y a une forme *le* et une forme *y*, alors la forme *le* précède la forme *y*.
- P16 s'il y a une forme *lui* et une forme *en*, alors la forme *lui* précède la forme *en*.

4.2.3. Les Propriétés de fléchage

Dans le Niveau 1, les Propriétés de fléchage sont les suivantes (rappel : °X note une variante sur le noyau) :

- F11 < phn, { _ }, flèche(auxf, [été]) >.
F12 < phn, (\ {été}), flèche(auxf, [p]) >.
F13 < phn, { - }, flèche(êtréf, [p]) >.
F14 < phn, { - }, flèche(cl,[°X]) >.

• Formulation intuitive des Propriétés de fléchage

[Dans la suite d'une phrase verbale noyau]

- F11 s'il y a une forme d'auxiliaire fléchie et la forme *été*, alors la forme d'auxiliaire fléchie flèche sur la forme *été*.
- F12 s'il n'y a pas une forme *été*,
s'il y a une forme d'auxiliaire fléchie et une forme de participe, alors la forme d'auxiliaire fléchie flèche sur la forme de participe.
- F13 s'il y a une forme *être* fléchie et une forme de participe, alors la forme *être* fléchie flèche sur la forme de participe.
- F14 s'il y a un clitique et un noyau, alors le clitique flèche sur le noyau

5. LE CALCUL SUR LES PROPRIÉTÉS

Les Propriétés définissent en intension un langage : c'est l'ensemble de modèles. Le GénMod (Générateur de Modèles) est l'outil informatique qui doit permettre de calculer, étant donné un lexique de catégories avec ses relations de subsomption, et un ensemble de Propriétés, tous et seulement les modèles qui satisfont ces Propriétés. Cela veut dire que le GénMod permet de passer de l'intension à l'extension.

Le GénMod *évalue* (c'est-à-dire *vérifie* ou *contrôle*) chacune des Propriétés. Il le fait en ordre, mais l'ordre d'évaluation ne doit pas être déclaré dans l'expression des Propriétés. Le GénMod est une machine algorithmique qui ne se modifie pas pour évaluer des Propriétés différentes, pourvu que celles-ci aient été déclarées selon les exigences générales de formulation des Propriétés.

Le GénMod n'est pas un analyseur automatique tel qu'on conçoit ceux-ci habituellement, bien que formellement il puisse être utilisé comme analyseur ; il serait probablement inefficace. Son rôle est d'exhiber les modèles qui vérifient les Propriétés afin, en particulier, de pouvoir les tester.

Le GénMod, pour évaluer les Propriétés d'existence, procède en calculant à partir de ce qu'ici on appelle un *alinéa*. À partir d'un alinéa il calcule, moyennant des règles, l'alinéa immédiatement suivant. Le schéma d'un alinéa est le suivant :

LP		
x ₁	Pa ₁ ;	D ₁
...		
x _n	Pa _n ;	D _n

Dans ce schéma, *LP* note une liste de Propriétés ; ce sont les Propriétés qui ont été déjà évaluées.

Les lignes de l'alinéa sont constituées d'un identifiant (noté x_j dans le schéma), et de deux ensembles, notés Pa_j et D_j dans le schéma. L'ensemble Pa (ou *Paquet*) note les catégories qui doivent obligatoirement être incorporées à un modèle pour vérifier les Propriétés enregistrées dans la liste *LP*. L'ensemble D (ou *Disponibles*) note les catégories qui, restant disponibles, peuvent être incorporées à un modèle tout en vérifiant les Propriétés enregistrées dans la liste *LP*.

Les règles du GénMod permettent de passer d'un alinéa à l'alinéa suivant. On note avec ' $>$ ' la ou les lignes qui ont été modifiées dans l'alinéa précédent lors du passage à l'alinéa suivant par application d'une règle et on note avec ' \bullet ' la ou les lignes de l'alinéa suivant qui ont été modifiées par ces règles. Lorsque on passe d'un alinéa au suivant, on ajoute à la liste *LP* la Propriété qui a été évaluée par application de la règle.

Exemple (cet exemple comme tout ce qui suit dans cette section 5 illustre les Propriétés d'existence du Niveau 1 présentées dans § 4.2.1) :

Soit les deux alinéas qui suivent, le premier précédant immédiatement le deuxième dans le calcul effectué par le GénMod.

P1, P2, P3		
>1	$^{\circ}v$;auxf, être, l', clr, le, lui, y, en
2	$^{\circ}\text{être}$;auxf, v, l',clr, le, lui, y, en
P1, P2, P3, P+1		
• 1.1	$^{\circ}\text{vepN, \text{êtréfN}}$;auxf, l', clr, le, lui, y, en
• 1.2	$^{\circ}\text{vef}$;auxf, être, l', clr, le, lui, y, en
• 1.3	$^{\circ}\text{va}$;auxf, être, l', clr, le, lui, y, en
2	$^{\circ}\text{être}$;auxf, v, l',clr, le, lui, y, en

Dans le premier alinéa on dit que les Propriétés P1, P2 et P3 ont déjà été évaluées (ce sont les Propriétés introduites par les Prédicats *amod*, *uniq* et *obligdi*). De ce fait, toutes les catégories dans *D* ne peuvent subsumer qu'un seul symbole par modèle. Cet alinéa dit aussi : il existe deux manières, exprimées par deux lignes différentes, les lignes 1 et 2, de satisfaire les Propriétés P1, P2 et P3.

Une manière consiste à incorporer à un modèle le Paquet $\{^{\circ}v\}$ avec une catégorie que *v* subsume comme noyau; dans ce cas, et toujours si on se limite aux Propriétés P1, P2 et P3, on a encore comme catégories disponibles, pour être ajoutées à $\{^{\circ}v\}$, les catégories qui sont subsumées par les catégories dans l'ensemble $\{\text{auxf, être, l', clr, le, lui, y, en}\}$.

L'autre manière consiste à incorporer à un modèle le Paquet $\{^{\circ}\text{être}\}$ avec une catégorie que *être* subsume comme noyau; dans ce cas, et toujours si on se limite aux Propriétés P1, P2 et P3, on a encore comme catégories disponibles, pour être ajoutées à $\{^{\circ}\text{être}\}$, les catégories qui sont subsumées par les catégories dans l'ensemble $\{\text{auxf, v, l',clr, le, lui, y, en}\}$.

Dans le premier alinéa on a ajouté le symbole *P+1* à la liste *LP* de l'alinéa précédent. On dit de cette manière que dans le deuxième alinéa on a évalué la Propriété *P+1*, dont l'évaluation vient s'ajouter à l'évaluation déjà effectuée dans l'alinéa précédent. *P+1* est la Propriété d'existence ici rappelée

P+1 exig(phn, [[vepN], [êtréfN]]).
 [Version intuitive : s'il y a un participe passé d'un verbe *ve*, alors il y a une forme *être* fléchi accordée avec lui en nombre]

Dans la première ligne de l'alinéa précédent -celle notée avec '>' — il y a une catégorie *v*, notée $^{\circ}v$, car elle est noyau. Cette catégorie a, à sa droite, des catégories "disponibles" dans *D* (Ce sont les catégories *auxf*, *être*, *l'*, *clr*, *le*, *lui*, *y*, *en*). Selon les catégories maximales présentées dans § 4.1, la valeur de trait *v* peut subsumer plusieurs catégories maximales. La Propriété *P+1* exprime une exigence seulement lorsqu'il y a les valeurs de traits *ve* et *p*. C'est la raison pour laquelle la ligne 1 de l'alinéa

précédent “s'éclate” dans les lignes 1.1, 1.2, 1.3 de l'alinéa suivant. La ligne 1.1 est concernée par P+1; pour satisfaire cette Propriété on a incorporé dans cette ligne les valeurs de traits *êtréf* exigés par P+1, et, comme il y a unicité sur *être*, on a enlevé *être* de D. En revanche, les lignes 1.2 et 1.3 ne sont pas concernées, pour ainsi dire, par P+1. Elles représentent tout ce qui était ν dans l'alinéa précédent mais qui n'est pas *vep*. Les catégories *vef* et *va* notent ainsi les catégories qui sont subsumées par ν mais qui ne sont pas des *vep*. La Propriété P+1 n'explicite aucune contrainte sur ces catégories. Par conséquent, tout ce qui est disponible dans D pour ces catégories dans l'alinéa précédent reste disponible dans cet alinéa. C'est la raison pour laquelle on reconduit, à droite de ces catégories, toutes les catégories qui étaient disponibles à droite de ν dans l'alinéa précédent. Ce sont des formes verbales qui soit (notées *vef*) relèvent des formes fléchies des verbes *ve* qui ne sont pas formes de participe, soit relèvent des verbes *va*.

On voit donc que dans le passage d'un alinéa au suivant tout ce qui devait être vérifié l'est, et que, en même temps, “rien ne se perd” : tout ce qui était possible avant l'évaluation d'une Propriété et qui n'est pas concernée par les exigences de cette Propriété, reste possible.

Les Propriétés sont évaluées dans un ordre prédéterminé. D'abord les Propriétés d'existence, ensuite les Propriétés de linéarité et enfin les Propriétés de fléchage. Les Propriétés d'existence sont aussi évaluées selon un ordre prédéterminé dans le GénMod. Mais l'ordre d'évaluation ne modifie pas l'ensemble de modèles spécifié par les Propriétés. Celui qui exprime les Propriétés n'a pas à se soucier de l'ordre de leur évaluation.

On présente par la suite les alinéas qui permettent l'évaluation des Propriétés d'existence P1 à P3 et P+1 à P+5.

P1		;auxf, v, être, cl~n
P1, P2		;auxf, v, être, l', clr, le, lui, y, en
P1, P2, P3		
>1	ν	;auxf, être, l', clr, le, lui, y, en
2	$^{\circ}$ être	;auxf, v, l',clr, le, lui, y, en
P1, P2, P3, P+1		
• 1.1	$^{\circ}$ vepN, êtréfN	;auxf, l', clr, le, lui, y, en
• 1.2	$^{\circ}$ vef	;auxf, être, l', clr, le, lui, y, en
>• 1.3	ν va	;auxf, être, l', clr, le, lui, y, en
2	$^{\circ}$ être	;auxf, v, l',clr, le, lui, y, en



La phrase verbale noyau en français

P1, P2, P3, P+1, P+2		
1.1	°vepN, êtrefN	;auxf, l', clr, le, lui, y, en
1.2	°vef	;auxf, être, l', clr, le, lui, y, en
•1.3.1	°1vap, êtref	;auxf, l', clr, le, lui, y, en
•1.3.2	°1vap, auxf	;êtref, l', clr, le, lui, y, en
•1.3.3	°vaf	;auxf, être, l', clr, le, lui, y, en
>•1.3.4	°2vap	;auxf, être, l', clr, le, lui, y, en
2	°être	;auxf, v, l',clr, le, lui, y, en
P1, P2, P3, P+1, P+2, P+3		
>1.1	°vepN, êtrefN	;auxf, l', clr, le, lui, y, en
1.2	°vef	;auxf, être, l', clr, le, lui, y, en
1.3.1	°1vap, êtref	;auxf, l', clr, le, lui, y, en
1.3.2	°1vap, auxf	;être, l', clr, le, lui, y, en
1.3.3	°vaf	;auxf, être, l', clr, le, lui, y, en
•1.3.4	°2vap, auxf	; être, l', clr, le, lui, y, en
2	°être	;auxf, v, l',clr, le, lui, y, en
P1, P2, P3, P+1, P+2, P+3, P+4		
•1.1.1	°2vepN, êtrefNP, clrNP	;auxf, l', le, lui, y, en
•1.1.2	°1vepN, êtrefN	;auxf, l', clr, le, lui, y, en
>1.2	°vef	;auxf, être, l', clr, le, lui, y, en
1.3.1	°1vap, êtref	;auxf, l', clr, le, lui, y, en
1.3.2	°1vap, auxf	;être, l', clr, le, lui, y, en
1.3.3	°vaf	;auxf, être, l', clr, le, lui, y, en
1.3.4	°2vap, auxf	; être, l', clr, le, lui, y, en
2	°être	;auxf, v, l',clr, le, lui, y, en
P1, P2, P3, P+1, P+2, P+3, P+4, P+5		
1.1.1	°2vepN, êtrefNP, clrNP	;auxf, l', le, lui, y, en
1.1.2	°1vepN, êtrefN	;auxf, l', clr, le, lui, y, en
•1.2.1	°2vefNP, clrNP	;auxf, être, l', le, lui, y, en
•1.2.2	°1vef	;auxf, être, l', clr, le, lui, y, en
1.3.1	°1vap, êtref	;auxf, l', clr, le, lui, y, en
1.3.2	°1vap, auxf	;être, l', clr, le, lui, y, en
1.3.3	°vaf	;auxf, être, l', clr, le, lui, y, en
1.3.4	°2vap, auxf	; être, l', clr, le, lui, y, en
>2	°être	;auxf, v, l',clr, le, lui, y, en

Dans le processus d'évaluation des Propriétés d'existence, il y a un stade où tous les Pa spécifiés sont associés à des D sur lesquels il ne reste qu'appliquer des *exclu* dont aucune catégorie n'est dans un Pa. On obtient ainsi les *lignes basiques*. Celles que l'on obtient par évaluation des Propriétés d'existence présentées dans § 4.2 sont les suivantes.

Lignes basiques

I	°2vepN, êtrefNP, clrNP	;y, en
II	°1vepN, êtrefNP	;clr~P~se, lui, y, en
III	°2vefNP, clrNP	;y, en
IV	°1vefP	;clr~P~se, lui, y, en
V	°1vapGN, êtrefN1P, leGN, clrseN1P,	;y, en
VI	°1vapGN, êtrefN1P, clrseN1P,	;y, en
VII	°1vapGN, êtrefN	;lui, y, en
VIII	°1vapGN, êtrefN1P, clrN1P~se, leGN	;y, en
IX	°1vapGN, êtrefN1P, clrN1P~se	;y, en
X	°1vapGN, êtrefNP	;clr~P~se, lui, y, en
XI	°1vapGN, auxfNP, été	;clr~P~se, lui, y, en
XII	°1vapGN, auxfP	;clr~P~se, leGN, lui, y, en
XIII	°1vafNP	;clrNP, le, lui, y, en
XIV	°1vafP	;clr~P~se
XV	°2vafP	;clr~P~se, lui, y, en
XVI	°2vap, auxfP	clr~P~se, lui, y, en
XVII	°ête, auxfP	l', clr~P~se, lui, y, en
XVIII	°euefP	;l', clr~P~se, lui, y, en

Pour obtenir les modèles qui satisfont les Propriétés d'existence, il ne reste qu'à appliquer les *exclu* sur les ensembles *D*. On en présente par la suite deux exemples sur les lignes basiques II et V.

• Exemples de calculs des exclusions sur *D* [on applique P~5, P~7, P~8, P~9]

II	°1vepN, êtrefNP	;{clr~P~se, lui, y,
	en}	
i	°1vepN, êtrefNP	; {}
ii	°1vepN, êtrefNP	; {clr~P~se}
iii	°1vepN, êtrefNP	; {lui}
iv	°1vepN, êtrefNP	; {y}
v	°1vepN, êtrefNP	; {en}
vi	°1vepN, êtrefNP	; {clr~P~se, y}
vii	°1vepN, êtrefNP	; {clr~P~se, en}
viii	°1vepN, êtrefNP	; {lui, en}
V	°1vapGN, êtrefN1P, leGN, clrseN1P	; {y, en}
i	°1vapGN, êtrefN1P, leGN, clrseN1P	; {}
ii	°1vapGN, êtrefN1P, leGN, clrseN1P	; {y}
iii	°1vapGN, êtrefN1P, leGN, clrseN1P	; {en}

En supposant que les Propriétés de linéarité ont opéré aussi, on trouve les exemples suivants pour illustrer des modèles qui satisfont les lignes basiques II et V et les exclusions sur *D*.

• **Exemples d'énoncés associés aux Pa des modèles calculés ci-dessus**

II	${}^{\circ}1\text{vepN}$, étrefNP (un cadeau) est arrivé *(un cadeau) sont arrivé (un cadeau) lui est arrivé *(un cadeau) s'est arrivé (un cadeau) t'est arrivé *(un cadeau) t'es arrivé	; {clr~P~se, lui, y, en}
V	${}^{\circ}1\text{vapGN}$, étrefNIP , leGN , clrseNIP (il) se les est regardés *(il) se les est regardé	; {y, en}

6 CALCULS POSSIBLES SUR LES MODÈLES

Si les Propriétés d'existence et de linéarité organisent les modèles en assignant aux catégories qui "existent" dans ces modèles des relations d'ordre, les Propriétés de fléchage expriment les relations entre ces catégories, dans l'ordre où elles se trouvent, relations qui vont permettre de calculer les représentations sémantiques¹⁶. Or, dans une Propriété de fléchage qui s'exprime par *A flèche sur B*, on a les informations sur

16. Dans toute la discussion qui suit on suppose une représentation sémantique qui soit une formule bien formée dans une logique quelconque — donc, une représentation sémantique calculable ou qui peut devenir calculable — cette représentation étant associée à un univers de dénotation par la Fonction d'assignation. On ne propose aucune représentation particulière parmi toutes les candidatures possibles (DRT à la Kamp, logique intensionnelle à la Montague ou tout simplement logique de premier ordre ou logique propositionnelle, ou autre] l'idée étant que l'expressivité du système proposé devrait permettre de choisir celle qui convient en fonction des objectifs qu'on veut atteindre avec la représentation choisie, celle-ci pouvant être réduite, si l'on ne veut pas faire des calculs d'inférence sur les représentations sémantiques, à, simplement, des fonctions propositionnelles, où les quantificateurs restent "flottants", c'est-à-dire où l'on n'a pas indiqué leur portée. Mais le problème central — cf. [Hintikka 94], p. 35-75 — auquel SP prétend apporter un commencement de solution, est le problème de la construction de la représentation sémantique à partir de formules syntaxiques de la langue naturelle, celles-ci manifestant une espèce d'incompatibilité profonde avec l'organisation de la syntaxe des expressions de tout langage artificiel, et en particulier, avec la syntaxe des formules logiques utilisées le plus couramment.

l'existence du fléchage lui-même, sur son orientation, et sur ce qu'est A en tant que catégorie source du fléchage et sur ce qu'est B en tant que catégorie cible du fléchage.

Dans 5P on cherche à calculer une représentation sémantique de manière compositionnelle : cela veut dire que la représentation sémantique résultante doit être le résultat d'une Fonction calculée à partir de la représentation sémantique apportée par les entrées lexicales à chaque catégorie et à partir des relations de fléchage entre ces catégories. Mais le calcul compositionnel n'implique pas — contrairement aux approches inspirés par Montague- des calculs “en parallèle” : dans 5P on ne respecte pas la contrainte selon laquelle à chaque règle de la syntaxe doit correspondre une règle de la sémantique. Et cette contrainte n'est pas respectée pour deux raisons. La première est qu'il n'y a plus de règle syntaxique en 5P : ce qu'une règle dans une grammaire habituelle exprime est dissous en 5P dans une pluralité de Propriétés. La deuxième raison est que la contrainte du parallélisme paraît particulièrement pernicieuse : ce que l'on peut exprimer dans la syntaxe des langues naturelles — syntaxe ici et toujours au sens de la théorie générale des langages — est quasiment aux antipodes des contraintes de non ambiguïté dans la résolution des parenthèses qui organisent les représentations sémantiques, et de la concaténation par adjacence, contraintes qui ont inspiré la définition de la syntaxe des langages artificiels tels que ceux de la logique.

La convergence de plusieurs flèches sur une cible doit permettre de calculer la représentation sémantique sur le point occupé par la cible à partir des catégories qui fléchent sur la cible et de la cible elle-même. On donnera par la suite quelques exemples de cette possibilité sans pourtant aboutir à une représentation sémantique en bonne et due forme.

On aura observé que dans la description proposée de la phrase verbale noyau il n'y a pas une catégorie particulière pour des formes verbales passives, et ceci contrairement à ce que l'on fait dans les grammaires dites calculables ou formelles, comme p.ex. GPSG [Gazdar et al. 85], bien que les propriétés permettent de spécifier des modèles qui doivent être associés à des interprétations actives et des modèles qui doivent être associés à des interprétations passives¹⁷. P.ex. la ligne basique XII dans § 5 doit être

17. Le type d'analyse proposée par la suite dans le texte présente une convergence que nous croyons non triviale avec les solutions proposées dans [Blanche-Benveniste et al. 84]; cf. en particulier ce qui y est dit (p. 116-118) sur le passif en tant que construction qui ne peut “être décrite que dans un ensemble beaucoup plus large” et, plus généralement sur la sémantique qui y est caractérisée comme “primitive” : “[Cette sémantique] est dégagée à partir des éléments pronominaux et de leur combinatoire dans la relation

associée à un modèle qui doit recevoir une interprétation active — modèle illustré ci-après par (i) — et la ligne basique XI doit être associée à un modèle qui, si les exigences sur l'entrée verbale sont respectées, doit recevoir une interprétation passive — modèle illustré ci-après par (ii).

- (i) ligne basique XII : °1vapGN, auxfP
(les étudiants) ont manifesté
- (ii) ligne basique XI : °1vapGN, auxfP, été
(les doléances) ont été présentées

Le problème est donc de pouvoir spécifier, à partir des catégories et du fléchage, les conditions nécessaires à l'établissement d'une représentation sémantique qui soit va recevoir une interprétation active (ci-dessous notée *iactive*), soit une interprétation passive (ci-dessous notée *ipassive*) soit l'une ou l'autre.

Ces conditions sont stipulées par la suite, où on a utilisé la notation qui suit.

- iactive* : interprétation active
ipassive : interprétation passive
sans X : sans une catégorie X fléchant sur le noyau
avec X : avec une catégorie X fléchant sur le noyau
si X = Y : si la catégorie X subsume la catégorie Y

Les conditions proposées sur les modèles spécifiées dans le Niveau 1 suivent :

- | | |
|----------------------------|---------------|
| Si le noyau est 1vap | |
| <i>sans être</i> | alors iactif |
| <i>avec être</i> | |
| • <i>si être = été</i> | alors ipassif |
| • <i>si être = êtrefNP</i> | |
| • <i>sans clrNP</i> | alors iactif |
| • <i>avec clrNP</i> | |

entre le verbe et ses termes" (p. 34). Le fléchage entre catégories syntaxiques disposées dans la linéarité du modèle permet de calculer cette sémantique primitive; la représentation sémantique associée dans chaque entrée dans le lexique permet d'induire ce qui, de la sémantique individuelle de chaque entité, est apporté à la construction de la représentation sémantique complète. Encore une fois, comme dans le reste de SP on factorise des informations de type différent, le fléchage entre catégories permettant de factoriser la sémantique primitive de [Blanche-Benveniste et al. 84]. Le résultat de cette factorisation exprime ainsi des conditions nécessaires mais non suffisantes à la construction de la représentation sémantique d'ensemble.

• avec le	<i>alors</i>	iactif
• sans le		
• avec se		
• avec êtreN 1vapN	<i>alors</i>	iactif <i>ou</i> ipassif
• avec êtreN 1vap~N	<i>alors</i>	iactif
• sans se	<i>alors</i>	iactif
Si le noyau n'est pas 1vap	<i>alors</i>	iactif

D'autres conditions analogues qui portent sur la construction de la représentation sémantique peuvent être exprimées à partir du fléchage entre catégories; on en donnera ci-après quelques exemples.

Les catégories clitiques fléchent sur le noyau verbal. Alors, on peut formuler la condition suivante qui est, cette fois, nécessaire et suffisante pour exprimer la relation de la dénotation qui doit être associée par le biais de la représentation sémantique aux formes clitiques décrites dans le Niveau 1. Cette condition peut être formulée de la manière qui suit.

Chaque catégorie_i de pronom clitique qui flèche sur le noyau de la phrase verbale noyau doit être associée dans la représentation sémantique de la phrase verbale noyau à un symbole qui sera associé par la Fonction d'assignation à une entité dans l'univers de dénotation qui sera différente de toute autre entité qui sera associée par la Fonction d'assignation à tout symbole dans la représentation sémantique qui aura été associé à toute autre catégorie_j de pronom clitique dans la phrase verbale noyau.

Simplement dit, cette condition fixe l'exigence bien connue de non identité de référence entre formes clitiques, les constructions réflexives n'étant pas considérées. Mais il semble bien qu'on possède aussi les éléments pour formuler des conditions plus compliquées.

Chaque fois qu'il y a convergence de fléchage sur une catégorie, on peut calculer la représentation sémantique qui sera associée à cette catégorie, et si cette catégorie est noyau dans son modèle — disons modèle_i — la représentation sémantique qui sera associée à un modèle — disons modèle_j — dont le modèle_i est noyau (sur la question de l'insertion d'un modèle dans un autre, cf. § 8.2 *in fine*). Ainsi, p. ex. le modèle de la phrase noyau qui doit être associé à *la regarde* peut être considéré le noyau du modèle de phrase qui doit être associé à *La fille la regarde*. Dans une représentation simplifiée, où *m-X* note modèle de type *X* et *c-X* note catégorie de type *X*, *m-sn* modèle de syntagme nominal, *m-phv* modèle de la phrase verbale noyau, *m-phv* modèle de phrase verbale, on aura

((La °fille) m-sn °(la °regarde) m-phv)m-phv

Dans cette représentation on exprime le fait que *regarde* est le noyau de la phrase verbale noyau, et que la phrase verbale noyau est le noyau de la phrase verbale. On peut ainsi "transporter" au niveau de la phrase verbale, le fléchage interne à la phrase verbale noyau.

Ceci est censé pouvoir apporter une solution au problème que les dépendances qui "se croisent" posent à la construction de la représentation sémantique. Ainsi, p. ex., on a les énoncés qui suivent ; cf. [Blanche-Benveniste et al 84, p. 79-80].

- i La fille lui court après.
- ii La fille en connaît la raison.
- iii La fille lui parle.
- iv La fille en revient.

Dans les quatre exemples les formes clitiques fléchent sur le noyau de la phrase noyau. Mais la phrase noyau est, pour ainsi dire, noyau de manière différente dans (i) et (ii) et, respectivement, dans (iii) et (iv). Dans (i) et (ii) les noyaux *court* et *connaît* sont des noyaux d'une phrase verbale noyau qui, elle-même, reçoit en fléchage des compléments (*après* et *la raison*) avec lesquels les clitiques de la phrase noyau doivent établir des relations pour construire la représentation sémantique. On a ainsi pour (i) la représentation simplifiée du modèle correspondant qui suit :

((La °fille) m-sn °(lui °court) m-phvn) (après)c-prep)m-phv

Si l'on fait "monter" les informations de fléchage entre catégories de la phrase verbale noyau et on peut en tenir compte lorsqu'on calcule les conséquences du fléchage dans la phrase verbale, on peut calculer pour (i) et (ii) des représentations sémantiques qui seraient en tout parallèles aux représentations sémantiques qui doivent être associées à, respectivement, *La fille court après Jacques* et *La fille connaît la raison de cela*.

7. ADAPTATION ET DÉVELOPPEMENT DES PROPRIÉTÉS

La question fondamentale de la variation linguistique a été évoquée § 1.2 et, avec elle, le rôle de l'Observateur, la relation entre texte virtuel et texte réel, et la formulation modulaire et incrémentale des Propriétés. Le défi est de pouvoir produire des descriptions formellement calculables mais non rigidement stabilisées, de telle manière que, avec des modifications, elles aussi susceptibles d'être exprimées formellement, on puisse passer d'un système à un autre. Sur ce point également, l'objectif des descriptions proposées est d'essayer d'intégrer, tout en restant dans une des-

cription cohérente, des points de vue différents. Il s'agit de faire cohabiter une vision purement statique des langues naturelles avec une vision qui puisse intégrer au moins quelques aspects de leur dynamique¹⁸.

Les Propriétés, parce qu'analytiquement exprimées, peuvent se spécifier selon la finesse souhaitée. Celle-ci peut varier selon les capacités adjudgées à l'Observateur, l'intersubjectivité des observations, l'utilisation qu'on en fera, les goûts personnels du descripteur, etc. Mais dans tous les cas, il s'agit de pouvoir exprimer les relations entre systèmes formellement décrits en indiquant ce que l'on doit rajouter ou enlever à l'un pour retrouver un autre avec des possibilités expressives modifiées. Dans cette section 7 cette possibilité sera explorée dans les différents Niveaux annoncés § 2. Dans § 7.1 on propose plusieurs modifications pour traiter les observations relevant du Niveau 1, modifications qui remodelent sur certains points la description proposée § 4. Dans § 7.2 à 7.5 la variabilité se manifeste sous la forme de l'incrémentalité : il s'agit de passer d'un système qui décrit moins d'observations à un système qui en décrit plus, chacune des sections correspondant à un Niveau différent (Niveaux 2 à 5). Mais le Niveau 3, celui qui concerne l'incorporation des formes à l'impératif, est présenté dans deux versions différentes, l'une puriste et l'autre non puriste, chacune de ces versions étant incrémentale par rapport à la description du Niveau 2.

Les Niveaux 4 et 5 sont traités moins exhaustivement que les précédents : le propos est ici, au moyen de quelques exemples qui semblent significatifs, de pousser la description proposée jusqu'à ses limites afin de rendre ses carences bien perceptibles.

7.1. VARIATIONS DANS LE NIVEAU 1

Les formes clitiques présentent certaines rigidités dans leurs relations de linéarité, mais aussi des variations sur ce qui est admis comme suites bien formées, et ceci au delà de la possibilité ou non de coordination. Dans [Blanche-Benveniste et al 84] on remarque, p. ex., que les suites *refléchi+ lui* ne se trouvent que dans les constructions dites de datif

18. Cette dynamique pouvant être considérée un indice de plus de la capacité d'adaptation à leur milieu des êtres vivants, dans ce cas l'adaptation de l'être humain à son univers linguistique caractérisé par la distance entre texte réel et texte virtuel. Il existe dès maintenant des données systématiquement recueillies et généralisées ou généralisables, qui devraient pouvoir s'exprimer sous la forme de contraintes sur les variations possibles; en lignes générales (cf. [Valli 95]), il semblerait qu'une des origines de la variation entre différentes variétés du français, est l'association des mots à des catégories maximales différentes, tout en conservant les mêmes Propriétés.

éthique, que les suites *lui+y* sont difficiles et que la suite *y+en* ne se trouve guère que dans l'expression *il y en a*. De par les Propriétés d'exclusion (celles de type *P~*), aucune de ces trois suites n'est spécifiée par le système. Elles sont respectivement interdites par *P~5*, *P~7* et *P~8*. Si on souhaite les décrire, il suffit de supprimer une ou plusieurs des Propriétés d'exclusion correspondantes sans rien toucher au reste.

Encore un exemple concernant les clitiques, mais plus compliqué. Soit *Pac1*, dans § 4.2, rapellée ici.

Pac1 exigac(phn, [auxfP, clr~P]).

permet d'éviter la spécification des modèles correspondant à

- (i) * (Vous) t'avez regardé
- (ii) * (tu) vous as regardés
- (iii) * (Nous) m'avons regardé
- (iv) * (Je) nous ai regardés

Selon l'Observateur, aucune de ces phrases n'est bien formée ou bien les quatre sont bien formées ou bien les trois premières sont mal formées et la quatrième acceptée comme bien formée. Dans § 4.2 on a adopté la première possibilité¹⁹.

La deuxième possibilité peut s'exprimer comme il suit. On modifie *Pac1* de la manière suivante et rien dans le reste des Propriétés n'est modifié.

exigac(phn, [auxf, clr], [auxfN, clr~N], [auxfP, clr~P]).

[Version intuitive : s'il y a un auxiliaire fléchi et un clitique réfléchi, alors l'auxiliaire fléchi et le clitique réfléchi soit ne s'accordent pas en nombre, soit ne s'accordent pas en personne, soit les deux].

La troisième possibilité peut s'exprimer comme suit. Il faut supprimer *Pac1* et rajouter les Propriétés qui suivent, sans rien modifier au reste des Propriétés.

- (i) exigac(phn, [[auxf[_,_,2], clr], [auxf[_,_,2], clr[_,_,1]], [auxf[_,_,2], clr[_,_,3]]]).
- (ii) exigac(phn, [[auxf[_,sg,1], clr], [auxf[_, sg,1], clr[_,pl,_,_]], [auxf[_, sg,1], clr[_, sg,2]], [auxf[_, sg,1], clr[_, sg,3]]]).
- (iii) exigac(phn, [[auxf[_,pl,1], clr], [auxf[_,pl,1], clr[_,_,2]], [auxf[_,pl,1], clr[_,_,3]]]).

19. [Grevisse 93], § 747, observe que (iv) "n'est pas rare", que (i) et (ii) n'ont pas été relevés (dans des textes réels, dirions-nous) et donne quelques exemples littéraires de (iii) tout en remarquant que (iii) est plus rare que (iv); cf. un classement analogue dans [Gross 68], p. 49.

[Version intuitive de (i) : s'il y a un auxiliaire fléchi de 2^e personne et un clitique réfléchi, alors le clitique réfléchi est soit de 1^{ère} personne soit de 3^e personne.

Version intuitive de (ii) : s'il y a un auxiliaire fléchi singulier de 1^{ère} personne et un clitique réfléchi, alors le clitique réfléchi est soit pluriel, soit singulier de 2^e personne soit singulier de 3^e personne.

Version intuitive de (iii) : s'il y a un auxiliaire fléchi pluriel de 1^{ère} personne et un clitique réfléchi, alors le clitique réfléchi est soit de 2^e personne, soit de 3^e personne].

Si l'on voulait exprimer la hiérarchie de jugements qui se dégage de la note 19, (iii) doit être remplacé par (iii'), qui rajoute, par rapport à (iii), la possibilité d'avoir un clitique réfléchi singulier de première personne avec un auxiliaire pluriel de première personne.

(iii') exigac(phn, [[auxf[_pl,1], clr], [auxf[_pl,1], clr[_.,2]], [auxf[_pl,1], clr[_.,3]], [auxf[_pl,1], clr[_sg, 1]]].

Remarquons que dans tous les cas on évite ce qu'il faut absolument éviter : une forme de clitique réfléchi en accord en nombre et en personne avec une forme de l'auxiliaire *avoir*²⁰.

20. Des solutions à des observations analogues à celles traitées dans le texte sont discutées dans [Blanche-Benveniste 75] et [Blanche-Benveniste et al. 84]. Les observations dans ces ouvrages concernent plutôt la relation des formes cliticques dans la phrase noyau avec les compléments comportant des formes pronominales en dehors de la phrase noyau — c'est-à-dire qu'elles concernent la question cruciale de la proportionalité, question qui reste au delà des limites de ce travail (cf. ci-après § 7.5 *in fine* et § 8) — plutôt que les relations entre catégories dans la phrase noyau, mais la filiation entre les deux types de problèmes est assez évidente. En effet, les situations ici évoquées pourraient être traitées par des Propriétés exprimant des effets analogues à ceux de la Règle n° 5 dans [Blanche-Benveniste 75], p. 211-214, ou à la Règle n° 2 dans [Blanche-Benveniste et al. 84], p. 102-104. Dans les deux formulations de ces règles il est crucial d'accepter qu'elles portent sur des formes pronominales dont l'une "implique dans sa composition" l'autre. Dans le premier texte on observe que des restrictions analogues s'observent en anglais, allemand, espagnol et italien (p. 212) — ce qui est absolument vrai — et on rappelle la contrainte proposée par Chomsky dans "Conditions on Transformations" sur la non-intersection de référence. Le cadre proposé et utilisé ici — 5P — ne sait pas, à l'heure actuelle exprimer complètement les représentations sémantiques calculables et, de ce fait, ne sait pas associer ces représentations à univers de dénotation; il ne sait pas non plus intégrer les facteurs liés à l'énonciation — incontournables pour parler de la dénotation des pronoms personnels — aux facteurs qui règlent la linéarité des formes linguistiques. Mais ne pas savoir traiter un phénomène ne veut pas dire nier l'existence de ce phénomène (cf., p. ex., le problème du chômage ou de la faim dans le monde !). Ce qui ici, de manière laborieuse est traité moyennant des Propriétés sur les formes, doit vraisemblablement pouvoir être traité de manière plus régulière et explicative en introduisant des Propriétés sur les représentations sémantiques, et sur les relations de celles-ci avec les conditions d'énonciation bien formulées et avec l'univers de dénotation. Mais justement,

7.2. NIVEAU 2

On conserve les présuppositions sur le lexique du Niveau 1. Dans ce niveau on incorpore l'auxiliaire surcomposé et les clitiques nominatifs.

• *Propriétés d'existence*

Étant donné l'incorporation à la description de la forme *eu* et des clitiques nominatifs, P1 et P2 du Niveau 1 sont modifiées, tandis que P3 du Niveau 1 reste sans altération.

P1[Niveau 2] amod(phn, [auxf, eu, v, être, cl]).
 P2[Niveau 2] uniq(phn, [auxf, eu, v, être, cln, l', clr, le, lui, y, en]).

P+1 à P+10 du Niveau 1 restent sans altération. On doit rajouter :

P+11 exig(phn, [eu], [auxf]).
 [Version intuitive : s'il y a une forme *eu*, alors il y a un auxiliaire fléchi]

Pac1 à Pac6 du Niveau 1 restent sans altération. On doit rajouter :

Pac7 exigac(phn, [fNP, clnNP]).
 [Version intuitive : s'il y a une forme verbale fléchie et un clitique nominatif, alors la forme verbale fléchie et le clitique nominatif s'accordent en nombre et personne]

Pac8 exigac(phn, [vepGN, clnGN]).
 [Version intuitive : s'il y a un participe d'un verbe *ve* et un clitique nominatif, alors le participe du verbe *ve* et le clitique nominatif s'accordent en genre et nombre]

Pac9 exigac(phn, [[1vap, être, cln], [1vapGN, été, clnGN], [1vapGN, êtrefNP, clnGNP], [1vapGN, êtrefN1P, clr N1P, clnG1N1P]]).
 [Version intuitive : s'il y a un participe d'un verbe *1va*, une forme *être* et un clitique nominatif, alors, si la forme verbale *être* est *été*, le participe passé *1va* s'accorde en genre et nombre avec le clitique nominatif (Ex. : *ont-elles été amées*) ou si la forme verbale *être* est *êtref*, le participe passé *1va* s'accorde en genre et nombre avec le clitique nominatif, le participe passé, la forme

le projet de 5P est de traiter ces phénomènes AUSSI de manière sémantique et non À LA PLACE de la manière syntaxique. En effet, on ne souhaite pas, dans 5P, avoir recours dans P2 à un système de filtrage, dont la solution par le tout sémantique semble avoir besoin, l'idée étant que c'est l'interaction entre les deux manières de voir un problème qui va permettre de le comprendre, et que, dans certains cas — mais pas dans tous — on peut aboutir au même résultat moyennant deux voies différentes, dont l'une est beaucoup plus simple que l'autre.

*être*f et le clitique nominatif s'accordent en nombre, la forme *être*f et le clitique nominatif s'accordent en personne (ex :*sont-elles aimées ?*).

ou si la forme verbale *être* est *être*f et il y a un clitique réfléchi, la forme *être*f et le clitique nominatif s'accordent en nombre et personne, et il n'y a pas d'exigence d'accord en nombre et genre entre le participe passé et le clitique nominatif (ex : *(quels livres)se sont-elles achetées ?*)

P~1 à P~9 du Niveau 1 restent sans altération.

• Propriétés de linéarité

Pl1 du Niveau 1 est modifiée comme suit :

Pl1[Niveau 2] < phn, { _ }, précède(f, {cln, eu, été, p}) >.

Pl2 à Pl6 du Niveau 1 restent sans altération. On doit rajouter :

Pl7 < phn, { _ }, précède(cln, {été, eu, p}) >.

[Version intuitive :

s'il y a un clitique nominatif et *été*, alors le clitique nominatif précède*été*

s'il y a un clitique nominatif et *eu*, alors le clitique nominatif précède *eu*

s'il y a un clitique nominatif et un participe, alors le clitique nominatif précède le participe].

Pl8 < phn, { _ }, précède(eu, {été, p}) >.

[Version intuitive :

s'il y a une forme *eu* et *été*, alors la forme *eu* précède *été*

s'il y a une forme *eu* et participe, alors la forme *eu* précède le participe].

• Propriétés de fléchage

F11 et F12 du Niveau 1 sont modifiées de la manière suivante :

F11[Niveau 2] < phn, { _ \ {eu} }, flèche(auxf, [été]).

[Version intuitive : s'il y a un auxiliaire fléchi et *été* et qu'il n'y a pas *eu*, alors l'auxiliaire flèche sur *été*]

F12[Niveau 2] < phn, { \ {eu, été} }, flèche(auxf, [p]).

[Version intuitive : s'il y a un auxiliaire fléchi et un participe, et qu'il n'y a ni *été* ni *eu*, alors l'auxiliaire flèche sur le participe]

F13 et F14 du Niveau 1 restent sans altération. On doit rajouter :

F15 < phn, { }, flèche(auxf, [eu]).

[Version intuitive : s'il y a un auxiliaire fléchi et *eu*, alors l'auxiliaire flèche sur *eu*]

7.3.1. Niveau 3-Version puriste

Étant donné l'incorporation de toutes les formes verbales (trait *vx*), des formes *ne* et *pas*, de toutes les formes clitiques (trait *cl*), P1 et P2 du Niveau 2 doivent être modifiées, tandis que P3 reste sans altération. On donne la présentation intuitive de certaines Propriétés au fur et à mesure de leur spécification, comme dans la section précédente; d'autres, en particulier celles qui concernent les formes clitiques à droite de l'impératif, sont commentées groupées après présentation de toutes les Propriétés les concernant; dans la présentation qui suit *vpu* est abréviation de *version puriste*.

• Propriétés d'existence

P1[Niveau 3-vpu]	amod(phn, [vx, cl, ne, pas]).
P2[Niveau 3-vpu]	uniq(phn, [auxf, v, être, eu, clr, cln, l', le, lui, y, en, ne, pas]).

Les Propriétés P+1 à P+11 du Niveau 2 restent sans altération. On doit rajouter :

P+12	exig(phn, [[pas], [ne]]).
	[Version intuitive : s'il y a une forme <i>pas</i> il y a une forme <i>ne</i>]

Les autres Propriétés de type P+ à rajouter, qui concernent les clitiques, sont présentées par la suite et commentées ci-dessous :

P+13	exig(phn, [[cl2], [im]]).
P+14	exig(phn, [[imp, cl1], [ne]]).
P+15	exig(phn, [[-l], ['y], ['en]]).
P+16	exig(phn, [[clrfa], ['y], ['en]]).
P+17	exig(phn, [['y], [[-l], [clrfa]]).
P+18	exig(phn, [['en], [[-l], [clrfa]]).

Les Propriétés Pac1 à Pac9 du Niveau 2 restent sans altération.

Les Propriétés P~1 à P~9 du Niveau 2 restent sans altération. On doit rajouter les Propriétés suivantes qui, toutes, concernent les clitiques et qui seront commentées ci-dessous :

P~10	exclu(phn, [[cln], [im]]).
P~11	exclu(phn, [[imp, ne], [cl2]]).
P~12	exclu(phn, [[-l], [clr], [lui]]).
P~13	exclu(phn, [[clrfa], [~ -l]]).
P~14	exclu(phn, [[~-l[_ , sg, _], [~'y], [~'en]]).
P~15	exclu(phn, [[clrfo[_ , sg, _], [~'en]]).

• *Propriétés de linéarité*

Les Propriétés P12 et P13, P15 et P16, restent sans modification. Les Propriétés P11, P14 et P17 sont modifiées de la manière qui suit :

P11[Niveau 3-vpu]	< phn, { _ }, précède(f, {cln, été, p, pas}) >.
P14[Niveau 3-vpu]	< phn, { _ }, précède(cl1r, {cl1~r}) >.
P17[Niveau 3-vpu]	< phn, { _ }, précède(cln, {pas, été, eu, p}) >.

La modification introduite à P11 et à P17 est minimale : elle consiste à introduire *pas* à l'intérieur des accolades pour indiquer que les formes verbales fléchies, dans P11, et les clitiques nominatifs, dans P17, précèdent la forme *pas*, incorporée à ce Niveau 3. P14 est modifiée afin que la nouvelle formulation ne concerne que les formes clitiques avec *cll*, qui seront celles qui sont placées à gauche de l'impératif (cf. ci-dessous).

On doit rajouter les Propriétés de linéarité qui suivent et qui, concernant les formes clitiques, seront commentées ci-dessous :

P18	< phn, { _ }, précède(cl1, {im}) >.
P19	< phn, { _ }, précède(im, {cl2}) >.
P110	< phn, { _ }, précède(ne, {vx, cl}) >.
P111	< phn, { _ }, précède(pas, {eu, été, p}) >.
P112	< phn, { _ }, précède(-l, {'y, 'en}) >.
P113	< phn, { _ }, précède(clrfa, {'y, 'en}) >.
P114	< phn, { _ }, précède(~-l, {clfo}) >.
P115	< phn, { _ }, précède(clfo, {~en}) >.

Les Propriétés d'existence P+13, P+14 et P~11, conjointement avec P~10 sont censées spécifier les formes de clitiques qui peuvent s'attester avec une forme verbale à l'impératif. P~10 est la Propriété la plus intuitivement directe : elle établit que les formes nominatives de clitique ne peuvent pas coexister avec une forme à l'impératif. P+13, P+14 et P~11 expriment l'exigence selon laquelle les formes qui ont reçu les traits *cl2* doivent se construire avec un impératif (P+13), mais cet impératif ne peut pas être accompagné d'un *ne* (P~11). Par ailleurs, s'il y a une forme d'impératif et une forme de clitique qui a reçu les traits *cll*, alors il faut aussi une forme *ne* (P+14).

Les propriétés de linéarité ordonnent les formes de clitique par rapport à la forme de l'impératif. Elles disent que les formes qui ont les traits *cll* vont à gauche de l'impératif (P18) et la forme de l'impératif va à gauche des formes de clitique qui ont les traits *cl2* (P19).

Dans cette version puriste, il y a deux formes différentes pour les formes clitiques qui présentent les traits qui suivent : *le*, *y*, *en* et *clr2* de 1^{re} et 2^e personne du singulier. Dans les commentaires qui suivent, on va

utiliser l'expression *forme(s) faible(s)* pour parler des formes qui sont associées aux catégories maximales présentant les traits *-l, 'y, 'en et fa*²¹.

Les Propriétés P+15 à P+18 créent les implications entre les formes faibles, alors que P~12 et P~13 s'assurent que dans les énoncés où il y a les formes faibles il n'y a que des formes faibles. P112 et P113 ordonnent les formes faibles entre elles. On a ainsi à droite d'une forme impérative non accompagnée de *ne* à sa gauche, une quelconque parmi toutes et seulement les suites suivantes :

-l'y
-l'en
-m'y
-m'en
-t'y
-t'en

Par les Propriétés commentées jusqu'ici, on sait que le système n'autorise à droite de l'impératif que les formes clitiques avec les traits *cl2*. On vient d'énumérer toutes et seules les suites que l'on peut spécifier avec une forme faible de clitique. On ne peut donc avoir à droite de l'impératif que les suites de formes de clitique associées à des catégories maximales avec les traits *cl2* et qui n'ont aucun trait propre aux formes faibles. Autrement dit, si l'on se rapporte aux catégories maximales de § 4.1, on ne peut avoir que des suites de formes clitiques qui ont un des traits qui suivent : *l', ~-l, lui, ~'y, ~'en, clrfo*. Les formes avec les traits *~-l* et *clrfo* peuvent être au singulier et au pluriel. Ces formes sont rappelées dans les exemples qui suivent :

Trait	Exemple
<i>l'</i>	sois-le
<i>~-l[_ , sg, _]</i>	regarde-le
<i>~-l[_ , pl, _]</i>	regarde-les
<i>lui</i>	parle-lui
<i>~'y</i>	vas-y
<i>~'en</i>	parlons-en
<i>clrfo[_ , sg, _]</i>	regarde-moi
<i>clrfo[_ , pl, _]</i>	regarde-nous

21. Pour saisir sans problèmes les commentaires qui suivent, le lecteur est invité à se rappeler la différence entre les suites de caractères qui notent les signifiants orthographiques, qui sont présentés entre accolades au § 4.1, et les suites des caractères qui notent les valeurs de traits et qui sont présentés sans accolades ou autre diacritiques dans les catégories maximales, présentées, elles aussi, dans § 4.1.

Toutes ces formes sont des *cl2* et rien ne les empêche d'apparaître à droite de l'impératif. Le problème est de déterminer quelles suites de plus d'une parmi ces formes sont possibles.

Les exclusions du Niveau 2 restent sans altération. Celles qui concernent les clitiques qui nous occupent sont rappelées ici par la suite.

P~5	exclu(phn, [[clr], [lui]]).
P~7	exclu(phn, [[lui], [y]]).
P~8	exclu(phn, [[y], [en]]).
P~9	exclu(phn, [[l'], [cl~n]]).

On sait que ces Propriétés sont exprimées moyennant des catégories, celles-ci étant des ensembles de traits. Si un trait dans une catégorie d'une Propriété apparaît dans une catégorie maximale, cette Propriété concerne cette catégorie. Autrement dit, une Propriété d'exclusion telle que, p. ex., P~5 ne concerne pas seulement les formes de clitique à gauche de l'impératif (celles qui sont *cl1*), mais aussi celles de droite (celles qui sont *cl2*). De telle manière que, sans rien modifier aux Propriétés d'exclusion que l'on vient de rappeler, on a les effets suivants :

- par P~9 on ne peut pas avoir dans un même énoncé une forme avec le trait *l'* conjointement avec une quelconque des autres formes clitiques
- par P~5 on ne peut pas avoir dans un même énoncé une forme avec le trait *lui* conjointement avec une forme avec le trait *clrfo*
- par P~7 on ne peut pas avoir dans un même énoncé une forme avec le trait *lui* conjointement avec une forme avec le trait *~'y*
- par P~8 on ne peut pas avoir dans un même énoncé une forme avec le trait *~'y* conjointement avec une forme avec le trait *~'en*.

Comme par ailleurs on a P2 qui exige l'unicité de formes de clitique, on sait que $\sim\text{l}[_{sg}, _]$ et $\sim\text{l}[_{pl}, _]$, et que $\text{clrfo}[_{sg}, _]$ et $\text{clrfo}[_{pl}, _]$ ne peuvent pas coexister. De telle manière que l'on a les ensembles de deux éléments qui suivent pour pouvoir construire les suites à droite de l'impératif.

i	{ $\sim\text{l}[_{sg}, _]$, lui}
ii	{ $\sim\text{l}[_{pl}, _]$, lui}
*iii	{ $\sim\text{l}[_{sg}, _]$, $\sim'y$ }
iv	{ $\sim\text{l}[_{pl}, _]$, $\sim'y$ }
*v	{ $\sim\text{l}[_{sg}, _]$, $\sim'en$ }
vi	{ $\sim\text{l}[_{pl}, _]$, $\sim'en$ }
vii	{ $\sim\text{l}[_{sg}, _]$, $\text{clrfo}[_{sg}, _]$ }
viii	{ $\sim\text{l}[_{pl}, _]$, $\text{clrfo}[_{pl}, _]$ }
ix	{ $\sim\text{l}[_{sg}, _]$, $\text{clrfo}[_{pl}, _]$ }

x	{~1[_pl,_], clrfo[_sg,_]}
xi	{lui, ~'en}
*xii	{clrfo[_sg,_], ~'en}
xiii	{clrfo[_pl,_], ~'en}

Les ensembles (iii) et (v) sont interdits par P~14 et l'ensemble (xii) l'est par P~15. Tous les autres sont possibles.

Comme les Propriétés de linéarité P15 et P16, déjà requises pour le Niveau 1 et 2 sont restées sans modification, pour les mêmes raisons que celles commentées ci-dessus, elles vont spécifier le bon ordre pour les éléments dans les ensembles (i), (ii), (iv), (vi) et (xi). Les éléments des ensembles (vii) à (x) sont ordonnés par P114 et ceux de l'ensemble (xiii) par P115. On a les exemples qui suivent des suites ordonnées bien formées.

i	donne-le-lui
ii	donne-les-lui
iv	ramène-les-y
vi	ramène-les-en
vii	donne-la-moi
viii	donne-les-nous
ix	donne-la-nous
x	donne-les-moi
xi	parle-lui-en
xiii	expliquez-vous-en

Étant donné les Propriétés proposées et le vocabulaire de catégories disponibles, à droite des suites à deux élément, on ne peut rajouter que la forme avec le trait ~'en aux suites illustrées par les exemples (i), (ii), (viii) et (ix). On peut ainsi obtenir, p. ex. à partir de (viii) *ramène-les-nous-en*.

• Propriétés de fléchage

Les propriétés de fléchage F11 à F15 du Niveau 2 restent sans altération. On doit introduire celles qui suivent, motivées par l'incorporation de *ne* et *pas*.

F16	< phn, {}, flèche(ne, [°X]).
F17	< phn, {}, flèche(pas, [°X]).

7.3.2. Niveau 3-Version non-puriste

On suppose un lexique de catégories sans les catégories maximales présentant les traits *-l*, *'y*, *'en* et *fa*. Ceci veut dire que l'on suppose un lexique de catégories maximales sans les formes qui ont été nommées formes faibles dans la section précédente. Comme pour la version puriste, on incorpore les formes verbales à l'impératif et les formes *ne* et *pas*. En lignes générales, on décrira un système où l'ordre est libre entre les clitiques à droite de la forme impérative et on suppose toujours la négation exprimée par *ne*, ce qui n'est pas toujours le cas (cf. [Blanche-Benveniste 84, p. 73]). En lignes générales, la version non puriste présentée — une donc parmi les versions non puristes — va conserver les Propriétés de la version puriste qui ne concernent pas les formes faibles. Les Propriétés du Niveau 2 — et non celles de la version puriste de la section précédente — sont le point de référence pour décider de leur conservation ou modification dans cette version du Niveau 3, mais les commentaires déjà faits dans la section précédente ne seront pas reconduits ici; dans la présentation qui suit *vnpu* est abréviation de *version non puriste*.

• Propriétés d'existence

P1[Niveau 3-vnpu]	amod(phn, [vx, cl, ne, pas]).
P2[Niveau 3-vnpu]	uniq(phn, [auxf, v, être, eu, clr, cln, l', le, lui, y, en, ne, pas]).

Les Propriétés P+1 à P+11 du Niveau 2 restent sans altération. On doit rajouter :

P+12	exig(phn, [[pas], [ne]]).
P+13	exig(phn, [[cl2], [im]]).
P+14	exig(phn, [[imp, cl1], [ne]]).

Les Propriétés Pacl à Pac9 et les Propriétés P~1 à P~9 du Niveau 2 restent sans altération. On doit rajouter les Propriétés qui suivent :

P~10	exclu(phn, [[cln], [im]]).
P~11	exclu(phn, [[imp, ne], [cl2]]).

• Propriétés de linéarité

Les Propriétés Pl2 et Pl3 du Niveau 2 restent sans modification. Les Propriétés Pl1, Pl4, Pl5, Pl6 et Pl7 sont modifiées de la manière qui suit :

Pl1[Niveau 3-vnpu]	< phn, { _ }, précède(f, { cln, été, p, pas }) >.
Pl4[Niveau 3-vnpu]	< phn, { _ }, précède(cl1r, { cl1~r }) >.

P15[Niveau 3-vnpu]	< phn, { _ }, précède(le1, {lui1, en1, y1}) >.
P16[Niveau 3-vnpu]	< phn, { _ }, précède(lui1, {en1}) >.
P17[Niveau 3-vnpu]	< phn, { _ }, précède(cln, {pas, été, eu, p}) >.

Les modifications de P14, P15 et P16 sont motivées par l'ordre libre des clitiques à droite de la forme impérative. On doit rajouter les Propriétés de linéarité qui suivent :

P18	< phn, { _ }, précède(cl1, {im}) >.
P19	< phn, { _ }, précède(im, {cl2}) >.
P110	< phn, { _ }, précède(ne, {vx, cl}) >.
P111	< phn, { _ }, précède(pas, {eu, été, p}) >.

• Propriétés de fléchage

Les propriétés de fléchage F11 à F15 du Niveau 2 restent sans altération. On doit introduire celles qui suivent, motivées par l'incorporation de *ne* et *pas*.

F16	< phn, { }, flèche(ne, [°X]).
F17	< phn, { }, flèche(pas, [°X]).

7.4. NIVEAU 4

Dans ce Niveau 4 on proposera une description embryonnaire de l'adverbe *toujours* et on discutera, à sa suite, la possibilité de généraliser la description proposée à la forme *tous*.

L'adverbe *toujours* — comme d'autres adverbes mais non comme tous les adverbes — est attesté dans la phrase verbale noyau dans plusieurs "positions", et notamment, après ou avant des formes de participe, et après seulement des formes fléchies. Les exemples qui suivent illustrent ces possibilités.

i	Ils travaillent toujours.
ii	*Ils toujours travaillent.
iii	Travaillons[im] toujours.
iv	*Toujours travaillons[im].
v	Travaillent-ils toujours ?
vi	*Toujours travaillent-ils ? ²²
vii	Ils ont toujours travaillé.

22. Avec un *toujours* en position dite topique — comme p. ex. dans *Toujours les enfants ont aimé le cinéma* — cet exemple est bien formé, mais on reste dans cette discussion dans les suites de la phrase verbale noyau. On rappelle que la description proposée ne concerne pas les formes à l'infinif.

- viii Ils ont travaillé toujours.
- ix Ont-ils toujours travaillé.
- x Ont-ils travaillé toujours ?

Dans tous les cas, il n'existe pas la possibilité d'avoir deux formes *toujours*. Un exemple suit :

- xb *Ils ont toujours travaillé toujours

Si maintenant on intègre ces observations avec celles concernant *pas*, on remarque facilement que *toujours* et *pas* peuvent coexister. Quelques exemples suivent :

- xi Ils ne travaillent pas toujours
- xii Ils n'ont pas toujours travaillé
- xiii Ils n'ont pas travaillé toujours

Alors que *toujours* et *pas* ont des Propriétés de linéarité strictes avec tous les autres éléments de la suite, entre eux l'ordre est libre. Quelques exemples suivent :

- xiv Ils ne travaillent toujours pas.
- xv Ils n'ont toujours pas travaillé

On peut rendre compte de ces observations en étendant avec très peu de modification les Propriétés du Niveau 3 version non puriste (mais exactement les mêmes modifications seraient nécessaires pour la version puriste). Les modifications consistent à introduire la catégorie avec le trait *tj* dans P1, P2, P11 et P17. Le restant de Propriétés du Niveau 3 sont sans modification. On a ainsi les Propriétés modifiées qui suivent.

• Propriétés d'existence

- P1[Niveau 4] amod(phn, [vx, cl, ne, pas, tj]).
- P2[Niveau 4] uniq(phn, [auxf, v, être, eu, clr, cln, l', le, lui, y, en, ne, pas, tj]).

• Propriétés de linéarité

- P11[Niveau 4] < phn, { _ }, précède(f, { cln, été, p, pas, tj }) >.
- P17[Niveau 4] < phn, { _ }, précède(cln, { pas, été, eu, p, tj }) >.

On remarquera que la forme *tj* est ordonnée par rapport aux formes fléchies dans P11, mais aucune relation de linéarité n'est introduite ni par rapport aux formes au participe ni par rapport à *pas*. C'est cette absence d'exigence qui permet de placer la forme *tj* à gauche et à droite des formes de participe et de *pas*.

Si on élargit les capacités de l'Observateur et on lui assigne aussi la possibilité de discriminer les suites qui doivent être associées à des représentations sémantiques différentes par rapport aux suites qui doivent être associées à une même représentation, on peut accepter qu'il considère les exemples (xi) et (xiv) comme exigeant des représentations sémantiques différentes, qu'il considère que (xii) et (xiii) admettent la même représentation, alors que (xv) en exige une différente par rapport à (xii) et à (xiii).

Les Propriétés de fléchage ont la fonction d'indiquer les liens qui doivent être respectés pour calculer la représentation sémantique. Comme exigence minimale, il faut que les fléchages soient différents pour les interprétations différentes, si l'on ne souhaite pas changer la sémantique propre de chaque entrée, c'est-à-dire si l'on ne souhaite pas régler le problème en multipliant les catégories associées à des formes homomorphes.

La solution suivante est proposée : si la forme *tj* n'est pas suivie de *pas* elle va flécher sur le noyau verbal et elle sera interprétée comme la plupart des modifieurs verbaux. Si la forme *tj* est suivie de *pas*, elle sera interprétée comme un modifieur de *pas* et flèche sur celui-ci. Tout ceci est exprimé par les Propriétés de fléchage qui suivent.

• Propriétés de fléchage

Les Propriétés F11 à F17 sont conservées. On ajoute celles qui suivent :

- F18 < phn, {\pasj}, flèche(tj, [°X]) >
 F19 < phn, {pasj\}, flèche(tj, [pasj]) >

Pour clore la description de ce niveau, on se limitera à remarquer que la description de la forme *tj* s'inscrit dans un cadre très général, dont on donne partiellement ici par la suite un autre exemple, avec la forme *tous*.

Le lecteur est invité à revenir sur tous les exemples (i) à (x) ci-dessus et à y changer *toujours* par *tous* : on retrouve exactement les mêmes effets que pour *tj*; la même observation étant valable pour (xb) et pour (xi) à (xiii). Il est facile de vérifier que *tous* peut coexister avec soit *pas*, soit *toujours*, soit *pas* et *toujours*; quelques exemples suivent :

- Ils ont toujours tous travaillé.
 Ils n'ont pas tous travaillé.
 Ils n'ont pas tous toujours travaillé.

On avait observé que l'ordre entre *toujours* et *pas* peut être échangé. L'ordre entre *tous* et *toujours* peut être échangé, mais pas celui entre *tous* et *pas*; quelques exemples suivent :

Ils ont tous toujours travaillé.
 Ils n'ont pas toujours tous travaillé.
 Ils travaillent tous toujours.
 Ils travaillent toujours tous.
 Ils ne travaillent pas tous.
 *Ils ne travaillent tous pas.
 Ils n'ont pas tous travaillé.
 *Ils n'ont tous pas travaillé.

On sait maintenant comment incorporer au moindre coût toutes ces observations. Il s'agit de reprendre P1 et P2 parmi les Propriétés d'existence, P11 et P17 parmi les Propriétés de linéarité et d'y rajouter *tous*, comme on l'a déjà fait pour *tj*. Comme, contrairement à *tj*, *pas* précède *tous*, il faut aussi modifier P11 et introduire *tous* parmi les formes précédées par *pas*.

La forme *tous* permet d'illustrer avec un exemple différent du précédent les fonctionnalités du fléchage. Étant donné que celui-ci est un élément essentiel pour construire les représentations sémantiques, il doit collaborer à l'établissement des relations dites d'"anaphore", au sens large du terme. Or la forme *tous* présente un comportement analogue à celui de *tj*, dans ce sens qu'elle doit être interprétée différemment selon la présence ou absence de certaines catégories dans les suites de la phrase noyau. Les exemples qui suivent donnent une idée du problème.

i (Ils) travaillent tous.
 ii (Ils) ont tous travaillé.
 iii Ont-ils tous travaillé.
 iv (Ils) les travaillent tous.
 v (Ils) les ont tous travaillé.
 vi Les ont-ils tous travaillé.

Il semble que, sans équivoque, dans (i) à (iii) la sémantique doit être calculée en faisant porter la sémantique propre à *tous* sur celle du sujet. Ceci peut s'exprimer en faisant flécher *tous* sur le noyau verbal sur lequel porte aussi le sujet, qu'il se trouve exprimé (p. ex. dans (iii)) ou non (p.ex. dans (i) et (ii)) dans la phrase verbale noyau. En revanche, il semblerait aussi que dans (v), *tous* porte sur *les* (cf. p.ex., *Jacques les a tous travaillés*), les exemples (iv) et (vi) étant ou bien ambigus ou bien devant se représenter comme pour (v), mais non comme pour (i) à (iii). Les Propriétés de fléchage peuvent exprimer ces observations. Celles qui

suivent expriment respectivement et dans l'ordre, le fléchage non ambigu de (i) à (iii), le fléchage non ambigu de (v) et le fléchage, ici fixé comme ambigu, de (iv) et (vi).

< phn, {Ne[_pl,_]}, flèche(tous, [°X]) >.
 < phn, {le[_pl,_]cIn[_pl,_]}, flèche(tous, [le[_pl,_]]) >.
 < phn, {le[_pl,_] ET cIn[_pl,_] \}, flèche(tji, [°X, le[_pl,_]]) >.

On est ainsi arrivé aux limites à l'intérieur desquelles se situe cette présentation : l'utilisation du fléchage pour calculer la sémantique, le fléchage profitant des propriétés d'existence et de linéarité. Le Niveau 5, suivant et dernier, sera utilisé pour présenter une autre limite importante.

7.5. NIVEAU 5

Ce niveau 5 complète la description en incorporant des exceptions introduites par certains lemmes verbaux.

Le classement des lemmes verbaux présenté dans § 4.1 est rappelé par la suite.

	Exemples	Notation
ve		
1	arriver	1ve
2	s'agenouiller	2ve
va		
1	manger	1va
2	succéder	2va

Encore une fois, ce classement ne prétend pas être LE bon classement des lemmes verbaux, mais plus modestement UN classement possible qui, conjointement avec les Propriétés, est destiné à proposer une description des suites de la phrase verbale noyau en français écrit, description qui doit être jugée dans son ensemble. Ce classement a été construit en essayant de maximaliser l'ensemble des lemmes verbaux pour lesquels le classement est une partition — c'est-à-dire qu'ils sont associés à l'une ou à l'autre classe de la partition — et de rendre possible les formulations le plus générales possible dans les Propriétés.

Avec le classement proposé, et en admettant qu'il y a des lemmes verbaux qui, ne s'accommodant pas de la partition, rentrent dans deux classes différentes (p.ex. *monter* dans *1va* et *1ve*, *succéder* dans *2va* et *2ve*), on peut exprimer une pluralité d'observations, dont on signale par la suite quelques unes.

Les lemmes classés dans *1va* “n'exigent” pas l'auxiliaire *avoir*; selon *P+2*, ils admettent soit *êtref* soit *auxf*. Ils ne sont pas concernés par *P+1*, qui exige l'accord en nombre du participe des lemmes verbaux *ve* avec *êtref*. De telle manière que la spécification des exemples

Elles se sont égratignées
Elles se sont égratigné les jambes

ne pose pas problème. Les deux s'obtiennent à partir de la ligne basique VI, ici rappelée (cf. § 5).

1vapGN, êtreN1P, clrseN1P

Cette ligne exige l'accord en nombre et personne entre le clitique réfléchi et la forme *être* mais non l'accord en nombre entre le participe et la forme *être* (ceci est formellement dit par le fait d'utiliser *NI* pour la forme *être* et *N* pour *1vap*); par ailleurs dans § 4.1 le clitique réfléchi de 3^e personne (*se*) n'a pas de valeur pour le nombre et, de ce fait, il peut s'accorder avec le pluriel et le singulier.

Des lemmes verbaux comme *arriver*, *s'agenouiller*, *manger* (qui relèvent donc respectivement de *1ve*, *2ve* et *1va*) présentent des formes de participe avec des marques orthographiques pour les traits d'accord de genre et nombre, ce qui induit, pour exprimer les relations d'accord, l'utilisation de valeurs de traits *m* ou *f* pour *G* et *sg* ou *pl* pour *N*.. En revanche les formes de participe des lemmes de la classe *2va* comme *succéder* ou *ressembler* présentent une forme invariable de participe et, pour les traits d'accord, reçoivent la description [, ,] (cf. § 4.1). De ce fait, ces formes peuvent satisfaire les exigences d'accord avec d'autres formes qui, elles, ont comme valeurs *m*, *f*, *sg* et *pl*. P.ex. si l'on admet que *ressembler* relève de *2va* et de *2ve* (ce qui n'est pas le cas de *abonder*, qui relève de seulement de *2va*), on peut spécifier, sans modifier les Propriétés, les exemples qui suivent :

- i Elles ont ressemblé à Martine.
- ii Elles lui ont ressemblé.
- iii Elles t'ont ressemblé.
- iv Elles se sont ressemblé.

Les trois premiers s'obtiennent de la ligne basique XVI (*2vap*, *auxfP*) et le quatrième de la ligne basique I (*2vepN*, *êtrefNP*, *clrNP*), grâce à l'invariabilité de *ressemblé*, décrit comme [, ,], et à la possibilité qui s'en suit de pouvoir l'accorder aussi bien avec le singulier qu'avec le pluriel.

Le dernier exemple que l'on vient de présenter relève des cas cités dans [Grevisse 93, § 916b]. Dans le même ouvrage on trouve répertoriées d'autres situations qu'il faut envisager. Des lemmes verbaux tels que *dire, devoir, croire, pouvoir, permettre, entendre, penser, faire, laisser* [cf. Grevisse 93 §913, 914, 915], lemmes qui sont à classer comme *1vap*, ne présentent pas toujours une forme morphologiquement accordée avec une forme du pronom clitique *le*. On a ainsi, p. ex.

- i (Les sacs, Jacques) *les a laissés*.
- ii (Les enfants, Jacques) *les a laissé* (jouer).

Si l'on n'avait pour ces lemmes verbaux qu'une seule forme avec [*m, sg, _*], de par *P+9*, les Propriétés ne pourraient en rendre compte. Mais rien n'empêche d'associer disjonctivement ces formes aux traits [*m, sg, _*] et aux traits [*_, _, _*]; ainsi, p.ex. *laissé* aurait, disjonctivement, les deux types de traits d'accord. C'est en quelque sorte une solution symétrique de la précédente. Pour les situations de type *ressemblé* il s'agit de classer un même lemme dans deux classes différentes. Pour les situations de type *laissé*, on conserve le classement du lemme dans le seul type *1va* mais on attribue disjonctivement des traits différents à la forme *laissé*.

Les formes avec *être* présentent plusieurs "comportements syntaxiques" différents :

- i Verbe principal, dans p. ex., *Il est à Paris ; Il est content*.
- ii Verbe auxiliaire, dans p. ex., *Elle est partie, elle s'est rechauffée, elle s'est agenouillée, elle s'est donnée à Jacques; elle s'est donné un livre*.
- iii Verbe auxiliaire passif, dans p. ex., *Elle est aimée, elle a été aimée*.
- iv Verbe auxiliaire semi-passif, dans p. ex., *Les revues se sont bien vendues*.

Les trois premiers comportements sont largement entrés dans la *koiné* des descriptions linguistiques. On parle beaucoup moins du quatrième et très peu des formes soulignées dans l'exemple suivant, qui ne sont pas répertoriées par une étiquette bien établie

- v Jacques a été aimé par Marie et Pierre *l'a été* par Martine; Heureux, Jacques *l'est*.

Il existe toujours la possibilité de multiplier les entrées homomorphes et de leur associer des traits différents pour pouvoir les traiter syntaxiquement de manière différenciée. Or ce type de solution a un inconvénient majeur lorsque les comportements différents se manifestent dans des contextes différents. Si tel est le cas, comme c'est le cas pour les formes avec *être*, on peut laisser au contexte le soin de décider de quel compor-

tement il s'agit. C'est ce type de solution que l'on a proposé pour les formes avec *être* dans § 4 : grâce aux catégories et aux Propriétés proposées, on a des formes avec le trait *être* qui se distribuent dans la linéarité des suites de la phrase verbale noyau et qui, grâce au fléchage, permettent de calculer la sémantique résultante (cf. § 6).

Or, telle qu'elle est spécifiée au Niveau 1, l'analyse proposée présente des carences, qui peuvent s'illustrer par les trois exemples qui suivent :

- i Heureux, Jacques l'est devenu.
- ii Je suis été.
- iii Je suis eue.

L'exemple (i), de par les catégories proposées et $P\sim 2$, ne peut pas être spécifié (le lemme *devenir* est censé être associé au trait v). L'exemple (ii), de par $P2$, ne peut pas être spécifié ($P2$ exige une seule forme avec le trait *être*). L'exemple (iii) peut être spécifié, une forme *avoir* comme *aux* étant distinguée d'une forme *avoir* avec lva (on a ainsi, p.ex., *il a* [1va] *un livre*; *les revues, il les a* [aux] *eues* [1va]).

En relation aux exemples (i) à (iii), comme il a été proposé dans § 1.2, la question intéressante n'est pas de savoir si (ii) et (iii) sont ou non du français (iii) est considéré comme un usage sous-standard et (iii) est présenté comme une bizarrerie littéraire dans [Grevisse 93, § 742]). La question est de savoir comment accommoder la description à la décision de considérer (ii) et (iii) comme du français, (i) étant considéré comme du français par tout le monde.

Le cas de (i) est une erreur de formulation, que le formalisme permet de faire éclater au grand jour : la valeur de trait $-l$ est réservée aux clitiques qui, dans la terminologie de [Blanche-Benveniste et al. 84] ont une relation de proportionalité avec des adjectifs et des participes. Selon la description proposée, ces types de clitiques vont être attestés lorsqu'on aura une forme *être* comme noyau. Mais l'exemple (i) montre que les clitiques l' peuvent aussi être attestés lorsqu'on a un verbe de type copulatif comme noyau, et pas seulement une forme *être*. Ainsi si l'on incorpore le trait *copulatif*; {+,-}, la manière qui semble la plus simple de réparer l'erreur consiste à :

- supprimer $P\sim 2$
- ajouter la Propriété d'existence $exig(phn, [l'], [être], [copulatif])$.

Le cas (ii) relève du même type de traitement que celui proposé pour les accords du participe : ceux qui utilisent des exemples tels que *je suis été* se sont donné un *été* de type lve ; la forme *été* sera ainsi associée disjonctivement à [*être*] et à [lve].

Le cas (iii) permet d'illustrer une des limites de la description proposée. Si un exemple de type (iii) est attesté dans un corpus considéré comme présentant des énoncés bien formés, les catégories et les Propriétés proposées peuvent le spécifier, tout en respectant les deux exigences (i) et (ii) d'adéquation externe présentés § 2, selon lesquelles le système doit être vérifié. En effet, on respecte l'exigence (i) car l'exemple est attesté dans un corpus qui est considéré comme présentant des énoncés bien formés et on respecte l'exigence (ii), car elle demande que "toute catégorie de tout modèle spécifié par le système doit subsumer la catégorie maximale associée à au moins un mot du lexique des signifiants". Or on peut spécifier l'exemple (iii) avec la même suite basique [*lvapGN, êtrefN*] qui permet de spécifier (*je suis aimée*), les deux formes *aimée* et *eue* étant associées à des catégories maximales subsumées par *lvapGN*.

Ceci revient à dire que les exigences de § 2 sont plutôt faibles, car elles admettent la surgénération : on y admet qu'il y ait des signifiants qui puissent être associés à des catégories maximales subsumées par les catégories dans les modèles sans pour autant produire des énoncés bien formés; c'est le cas, pour certains Observateurs, de l'exemple (iii) ci-dessus. En fait, les exigences fixées au § 2 définissent des conditions nécessaires mais non suffisantes²³ de bonne formation.

Pour réduire la surgénération — et s'approcher ainsi des conditions nécessaires et suffisantes — il faut introduire des exigences supplémentaires sur les signifiants qui peuvent être associées aux catégories dans les modèles.

Ces exigences, pour (iii) ci-dessus sont relativement simples à établir. Les formes *avoir* seront disjonctivement associées à [*lva, avoir*] pour les emplois comme noyau, et elles resteront associées à [*aux*] pour les emplois comme auxiliaire, et on incorporera l'exclusion suivante :

exclu(phn, [[avoir], [être]]).

Avec cette exclusion, on devrait éviter non seulement (iii), mais aussi les exemples qui suivent :

- * Je me suis eu un livre.
- * J'ai été eue.
- * Je me suis eue.

23. Ce qui, méthodologiquement, est parfaitement raisonnable (cf. [Bochenski 58], p. 209-212), en dépit d'une tradition qui semble s'être installée dans les discussions linguistiques selon laquelle est à écarter toute description présentant seulement soit des conditions suffisantes, soit des conditions nécessaires.

Les cas traités dans ce Niveau 5 relèvent, tous, des relations délicates entre entrées individuelles d'un lexique et Propriétés générales attribuées à des classes ou sous-classes d'entrées. Le choix d'essayer de factoriser des régularités en les exprimant au moyen des Propriétés entraîne l'acceptation, dans certaines limites, de la surgénération. Cette surgénération peut se réduire en structurant le lexique, la structuration, si besoin est, pouvant descendre au niveau des entrées individuelles. Ceci est formellement possible — il s'agit, dans tous les cas, d'utiliser des traits pour partitionner des classes — et les exemples traités ci-dessus montrent que le procédé peut être linguistiquement intéressant à utiliser. En développant son utilisation, on peut réduire de manière systématique la surgénération. Ainsi p. ex., tous les lemmes verbaux *Iva* n'admettent pas forcément une forme *lui* : on peut donc partitionner les lemmes *Iva* en ceux qui admettent cette forme, et les autres.

Ce niveau supplémentaire de la description exige notamment, d'une part, la spécification des exclusions entre les catégories exigées dans chaque ligne basique avec les catégories qui restent disponibles et, d'autre part, l'intégration au système des relations de proportionnalité de [Blanche-Benveniste et al. 84], afin de pouvoir décrire formellement le lien entre ce qui se passe à l'intérieur et à l'extérieur de la phrase verbale noyau. Or ce niveau supplémentaire — et exigeant — de description ne rentre pas dans les limites de ce travail.

8. MISE EN PERSPECTIVE

Le cadre proposé — 5P — et la description proposée — les Propriétés P2 — sont succinctement mis en perspective dans cette dernière section. Dans § 8.1 on situe le cadre d'ensemble par rapport à d'autres couramment pratiqués en essayant de noter convergences et divergences et en ayant, parmi les soucis principaux, celui de revenir sur la difficulté d'interaction entre linguistes, informaticiens et logiciens. Le point de vue adopté dans § 8.1 est ainsi de préférence méthodologique. Dans § 8.2 on revient sur le domaine visé de la description ici proposée sous la forme de Propriétés (P2), en essayant de la relativiser par rapport à d'autres approches et par rapport à des points essentiels qui restent en dehors.

8.1. LE POINT DE VUE MÉTHODOLOGIQUE

Le point méthodologique essentiel qui a guidé la structuration du cadre est l'objectif d'intégrer sérieusement la linguistique aux sciences du réel, dans lesquelles le progrès se fait par une interaction constante entre observation contrôlée, formulation d'hypothèses, calcul des conséquences de ces hypothèses, et test des conséquences obtenues par rapport à l'observé ou observable, aucune priorité temporelle ou causale n'étant prêtée à chacune de ces activités.

Dans une science du réel on ne prétend pas décrire tout le réel. Celui-ci est modélisé, c'est-à-dire qu'il est idéalisé de manière contrôlée pour l'observer selon le point de vue fixé. C'est la raison pour laquelle ici on ne parle pas du *français* mais d'un sous-ensemble de suites du français écrit observées dans des conditions censées avoir été suffisamment bien explicitées pour que n'importe qui puisse s'approcher avec assez d'exactitude de l'objet visé et puisse, par là même, vérifier la description proposée.

Dans une science du réel on n'exige pas que chaque hypothèse ou groupe d'hypothèses soit associé à un aspect particulier de l'objet étudié. C'est l'interaction entre les hypothèses dans leur ensemble qui permet de s'approcher de la connaissance de l'objet. C'est la raison pour laquelle on ne prête aucun avantage de "naturalité" aux différents éléments qui intègrent la description d'ensemble. Notre Observateur, p. ex., ne sait pas juger si les Catégories proposées sont naturelles, si la manière de penser la description du passif est élégante et ainsi de suite²⁴. Il ne sait faire que ce qu'on lui a dit de faire au § 2 et, dans l'expertise qui lui est assignée, ne figure pas la possibilité de juger l'élégance ou la naturalité des hypothèses proposées, dans leur ensemble ou séparément.

En revanche, dans une science du réel, on se donne comme exigence, ou tout au moins, comme objectif vers lequel il faut aller, la formulation des hypothèses au moyen de formalismes calculables pour en déduire

24. Notre Observateur n'a donc pas des intuitions sur les aspects partiels d'une description et notamment, sur des aspects partiels de la structure associée par une grammaire à un énoncé. Sur l'impossibilité de faire corrélérer une partie des hypothèses, isolées de l'ensemble auquel elles appartiennent, avec des aspects partiels de l'objet étudié, cf. [Bunge 69], p. 292 et *passim*, et [Sokal et Bricmont 97], p. 65. Ce choix méthodologique est donc à l'opposé de celui qui a prévalu en linguistique à partir de l'adoption des positions chomskyennes; celles-ci sont fidèlement rapportées dans [Ruwet 68]; on y trouvera (notamment p. 143-171 et p. 56-83) l'extraordinaire variété d'intuitions assignées à l'Observateur. La position ici adoptée est ainsi aux antipodes de [Katz 81], et en particulier, des deux derniers chapitres de cet ouvrage.

leurs conséquences. Dès que l'objet étudié est un peu complexe, on ne peut pas tirer les conséquences de ce qu'on dit sur lui sans un outil de calcul. C'est la raison pour laquelle on a cherché à décrire les Propriétés dans un formalisme calculable, qui devrait permettre de dégager toutes et seules les conséquences des Propriétés proposées.

Mais dans 5P *outil de calcul* n'est pas identifié avec *outil d'analyse ou de génération automatique d'énoncés ou de textes* et encore moins, avec logiciels des industries de la langue susceptibles de devenir marchandise. On peut donc calculer, et de manière rigoureuse, sur les Propriétés sans que, pour autant, on puisse obtenir par là même un analyseur automatique efficace pour être intégré dans tel ou tel logiciel commercialisable.

Le point précédent est une conséquence directe d'une idéalisation de l'objet langage humain différente de celle proposée par Saussure et qui a eu cours jusqu'à et y compris Chomsky. La langue (compétence) reste distincte de la parole (performance) mais la langue "n'est pas utilisée dans la parole". Un peu plus techniquement, cela veut dire que *les Propriétés (P2) ne sont pas la source déclarative utilisée dans les Processus (P5) d'analyse et de génération automatiques*. La grammaire n'est donc pas utilisée dans la performance et ceci pour la simple raison qu'il n'y a pas de grammaire dans 5P.

La notion de grammaire telle qu'elle est pratiquée dans les modèles dit calculables ou formels, recouvre une double fonction : (i) outil descriptif et (2) source déclarative des algorithmes d'analyse ou de génération automatiques²⁵. Le linguiste, après avoir appris la syntaxe des formules à déclarer, et acquis une idée de ce que ces formules sont censées dire (cf. ci-après § 8.2) doit écrire ces formules en vue de décrire la langue dont il s'occupe; par ailleurs ces formules sont la source qui est censée être utilisée par l'algorithme de traitement. Comme il est dit dans [Pollard et Sag 94, p.12] la fonction d'une grammaire "is to play the role[...] of information that fits directly into a model of processing".

Le problème avec cette manière de voir les choses est que le formalisme déclaratif est conçu sous l'exigence que la machine puisse le traiter efficacement, particulièrement dans les analyses automatiques. De ce fait on a essayé d'injecter le maximum d'information de tout type dans les entrées lexicales en essayant d'y exprimer des phénomènes de plus en plus complexes, ce qui a conduit à ce qui semble bien être l'impasse actuelle : aucun de ces formalismes supposés calculables n'est aujourd'hui utilisé

25. Il s'agit, dans la formulation vraisemblablement la plus claire et extrême, de la "strong competence hypothesis" de [Bresnan et Kaplan 82], Introduction § 3.

dans des tâches effectives de traitement effectif de textes réels, et l'acquisition de ces formalismes, toujours changeants, est de plus en plus coûteuse par rapport à ce qu'on peut obtenir en les utilisant.

En revanche et de manière symétrique, les outils d'analyse automatique qui commencent à être aujourd'hui efficacement utilisables dans des tâches, certes spécialisées, mais effectives sur des textes réels, ne sont pas associés à des sources déclaratives calculables²⁶.

26. On fait référence dans le texte ci-dessus aux travaux issus du *shallow parsing* (cf. une définition de cette notion dans [Abney 92] et [Frederici et al. 96]). L'objectif central de ces travaux est d'analyser automatiquement de vastes corpus avec des exigences sur la sortie plus modestes que celles de grammaires dites formelles pour ce qui est des structures obtenues (on obtient de descriptions syntaxiques partielles, parfois présentées comme une étape dans l'obtention de structures plus riches, cf. [Frederici et al. 96]), mais des exigences plus fortes sur l'effectivité de l'analyse automatique sur des textes réels, la satisfaction de ces exigences étant dûment testée et statistiquement évaluée (sur l'application au français, notamment sur la reconnaissance et extraction du sujet et de l'objet, cf. [Aït-Mokhtar et Chanod 97a] et [Aït-Mokhtar et Chanod 97b]). Mais si dans ces travaux on peut évaluer les résultats obtenus par un analyseur confronté à des textes réels, on a beaucoup de mal à cerner le formalisme déclaratif — s'il y en a — à partir duquel il devrait être possible de décrire toutes et seules les suites censées être analysées avec les structures qui leur sont associées. Ce problème est étroitement lié aux types de règles utilisées et à la présentation des analyseurs. Dans [Hindle 94], § 3.4.1, on utilise des "rules actions" qui sont en fait des instructions pour effectuer des opérations; les règles de [Tapanainen et Järvinen 97] s'organisent autour des deux opérations fondamentales de *supprimer* et *sélectionner* qui sont exécutées selon le résultat d'explorations locales dans les textes. Dans [Karlsson et al. 95], chap. 2, les informations à donner à la machine ont parfois un aspect déclaratif et parfois sont explicitement des heuristiques, ce qui est explicitement reconnu et présenté comme un avantage (p. 43). L'incrémentalité du processus est obtenue par Aït-Mokhtar et Chanod par l'utilisation de transducteurs qui opèrent en cascade, chaque étape de l'analyse pouvant non seulement ajouter des informations à celles obtenues dans l'étape précédente mais aussi les modifier [Aït-Mokhtar et Chanod 97a], § 2.2. Ceci n'est pas sans conséquences sur la conception de l'information linguistique qui semble être utilisée. Dans [Aït-Mokhtar et Chanod 97a], § 2.4, on lit que "This [l'incrémentalité du processus d'analyse] has a strong impact on the linguistic character of the work. The ordering of the linguistic descriptions is in itself a matter of linguistic description : i.e. the grammarian must split the description of phenomena into sub-descriptions, depending on the available amount of linguistic knowledge at a given stage of the sequence". On a ici le sentiment qu'il n'existe plus la limite entre informations déclaratives permettant de spécifier les propriétés des énoncés d'une langue et machine permettant d'analyser ces suites pour obtenir un résultat quelconque : les "contraintes" dans [Aït-Mokhtar et Chanod 97b], § 3.1 sur les sujets potentiels dans un énoncé français, présentées en langue naturelle, ne semblent pas avoir été incorporées au système sous la forme de propriétés et déclaratives et calculables; ce sont plutôt des guides sur les objectifs que la machine doit atteindre. Par ailleurs la question de la discrimination de la grammaticalité des suites n'est par perçue comme une question intéressante, les parseurs acceptant explicitement des suites agrammaticales, cf. [Karlsson et al. 95], p. 37, [Hindle 94], § 2.1, et [Aït-Mokhtar et Chanod 97a], § 2.5. On arrive ainsi à la conclu-

On arrive ainsi à la situation diagnostiquée dans [Blache 98c, p. 15, texte où nous aurions dit *système descriptif censé être calculable* à la place de *théorie*, cf. ci-après § 8.2] :

Nous sommes donc en quelque sorte confrontés à la quadrature du cercle : utiliser une théorie linguistique est très complexe, ne pas en utiliser est trop restrictif et aucune des deux démarches n'est réellement satisfaisante si on prend en compte les paramètres fondamentaux que sont la *cohérence*, la *généralité*, l'*efficacité* et la *réutilisabilité*, mais aussi la gestion du développement des descriptions.

Les fonctions qui sont censées être assurées par une grammaire calculable sont, dans 5P, analytiquement assignées aux Propriétés (P2) d'une part, aux sources déclaratives qui seront utilisées par les algorithmes d'analyse et de génération automatiques dans les Processus (P5) d'autre part. Ces sources déclaratives des algorithmes dans les Processus (P5) ne sont pas les Propriétés (P2), mais elles doivent être le résultat d'une fonction qui a comme entrée les Catégories et les Propriétés, si l'on veut se limiter à calculer jusqu'au fléchage, ou les Catégories, les Propriétés et, en plus, les Fonctions sémantiques qui calculent à partir du fléchage, si l'on veut obtenir les représentations sémantiques. Ceci donne une autonomie réelle à la formulation des Propriétés qui d'une part, peuvent être exprimées sans qu'on ait forcément trouvé la solution de comment les implanter efficacement dans un analyseur automatique et qui peuvent être exprimées de manière très analytique et avec différents niveaux d'exigence et de généralisation.

On peut maintenant revenir à la question soulevée au tout début de § 1.2 sur la difficulté d'interaction entre linguistes, informaticiens et logiciens. Malgré des tentatives timides, on a parfois le sentiment que, dans l'organisation des colloques, le recrutement des sociétés savantes, Conseils scientifiques, les Commissions de spécialités, les Départements des Langues ou d'Informatique, ou autres instances avec pouvoir de décision, on s'accommode souvent volontiers d'une espèce d'alibi pernicieux selon lequel on peut, en toute bonne conscience, soit faire de la linguistique, soit faire du traitement automatique, chaque activité étant conçue comme autonome par rapport à l'autre.

Le choix méthodologique directeur de 5P est clair : on ne peut pas de nos jours faire de la linguistique sans un formalisme calculable dans

sion que si les grammaires dites formelles sont inapplicables à des textes effectifs (compétence visée mais non calculable et sans performance effective), les travaux issus du *shallow parsing* produisent des analyses à partir des textes réels sans une information déclarative distincte (performance effective sans compétence).

lequel exprimer les propriétés des langues naturelles, formalisme qui doit permettre de tester les descriptions et les hypothèses qui y sont formulées par rapport à des données systématiquement représentées et observées dans des conditions spécifiées.

Ceci ne veut pas dire :

- qu'on ne puisse pas exploiter des trésors de matériel descriptif présentant de précieuses régularités observationnelles, sous-jacentes aux phénomènes directement observés et formulées dans un langage peu formalisé ou non entièrement formalisé, mais formalisable
- que le formalisme de calcul doive être identifié à des algorithmes d'analyse ou de génération automatique, ni moins encore, à un modèle de grammaire dite formelle
- qu'il soit nécessaire d'aboutir à une représentation sémantique calculable exprimée dans la syntaxe des formules d'une logique calculable.
En revanche ceci veut dire que l'on est confronté à un triple défi :
- par rapport à toute observation intuitive, le défi qu'elle devienne intégrée à un formalisme de calcul afin de pouvoir inférer ses conséquences et les tester
- par rapport à tout traitement automatique, le défi de référer le traitement à des Propriétés déclaratives du langage traité, y compris — peut-être surtout — quand le traitement automatique échoue
- par rapport à toute importation de formalisme logique, le défi de vérifier sa capacité expressive sur un échantillonnage des problèmes proches de ceux qui concernent l'utilisation effective des langues naturelles et, tout particulièrement, le défi de montrer une possibilité raisonnable de construire les formules de la représentation sémantique, avec leur syntaxe particulière, à partir des structures syntaxiques des énoncés de la langue naturelle.

Est-ce que 5P propose un terrain non éclectique d'entente où intégrer des traditions différentes et continuer à relever ces trois défis, en proposant des passerelles entre linguistes, informaticiens et logiciens ? On le souhaite, mais il n'y a que la pratique, et donc le temps, qui pourra arbitrer.

Dans l'alinéa qui suit on aborde encore un point méthodologique, mais cette fois davantage en relation avec le domaine visé de la description.

8.2. LE DOMAINE VISÉ DE LA DESCRIPTION

Tout au long de la présentation de la description proposée, grand soin a été pris de toujours parler d'une description possible du domaine

observé et jamais d'une théorie ou d'une explication de ce domaine, en insistant sur le fait que le mécanisme proposé est fait de pièces dont chacune semble relativement simple, mais que le tout interagit de manière plus complexe, et qu'il est impossible de juger des pièces isolément. Il est sans doute vrai que les catégories peuvent être décrites différemment, de même que les Propriétés, que rien n'a été fait pour éviter la redondance éventuelle de certaines Propriétés²⁷.

On peut donner au mot *théorie* un sens un peu plus précis de celui qu'il reçoit souvent, en particulier dans les écrits de linguistique. On aurait une *théorie* de l'objet *Z* lorsque, à partir d'un ensemble de formules (hypothèses générales) exprimées dans un formalisme calculable, on peut, conjointement avec des formules_x (hypothèses descriptives) qui décrivent des aspects_x de l'objet *Z* déduire des conséquences qui décrivent des aspects_y, différents des aspects_x, de l'objet *Z*, les aspects_x et les aspects_y devant être confirmés par l'observation de *Z*. Si tel est le cas, on dit souvent que la théorie *explique* les aspects_y, et si on y ajoute une composante temporelle, que la théorie *prévoit* ou *prédit* les aspects_y.

Ce schéma est prévu dans 5P dans les Projections (P3) et les Principes (P4), rapidement présentés dans § 1.2. On n'ajoutera grand chose d'autre ici; on notera, que, encore une fois, les *P3* et les *P4* restent distincts de *P2*, et qu'ils doivent être calculés de telle manière qu'en formulant des *P4* on puisse, conjointement, avec les *P2* et les *P3*, soit déduire d'autres *P2'*, distincts et complémentaires des *P2*, soit modifier la formulation des *P2* afin d'obtenir d'autres *P2'* qui seraient confirmés par d'autres données, celles-ci s'ajoutant à celles qui confirment *P2*.

L'utilisation de *théorie* est fréquente en linguistique avec un sens moins précis que celui résumé ci-dessus. Si dans [Chanod 95, p. 6] la question "Qu'apporte le TALN à la théorie linguistique (et non l'inverse)" est posée donnant à *théorie* le sens qu'on vient de préciser, en revanche on se réfère souvent à HPSG comme une *théorie*. Citons comme exemples [Emirkanian et Bouchard 95, p. 11], [Abeillé 97, p. 10], [Sag et Miller 97, § 2]; ceci n'est pas étonnant étant donné la présentation de HPSG dans [Pollard et Sag 94].

27. On fait la conjecture que les *P2* proposées sont *cohérentes* (aucune suite de catégories ne pouvant à la fois être un modèle et ne pas être un modèle), et formellement *complètes* (toute suite de catégories pouvant ou non être associée à un modèle satisfaisant les *P2*). En revanche, on ne fait pas la conjecture que toute Propriété est *indépendante* de toutes les autres. On utilise ici respectivement les notions de *consistency*, *completeness* et *independence* de [Partee et al. 90], § 8.5.

Dans [Pollard et Sag 94, p. 6-7] il est explicitement reconnu le besoin de formalisation calculable signalé ci-dessus :

An informal theory is one that talks about the model in natural language [...]. But as theories become more complicated and their empirical consequences less straightforwardly apparent, the need of formalization arises. In cases of extreme formalization, of course, the empirical hypotheses are cast as a set of axioms in a logical language, where the modelling structures serve as the intended interpretations of expressions in the logic.

Le besoin est reconnu, même si le texte ajoute :

This does not mean that the empirical hypotheses must be rendered in a formal logic as long as their content can be made clear and unambiguous in natural language [...], but in principle they must be capable of being so rendered. Unless these criteria are satisfied, an enterprise purporting to be a theory can not have any determinate empirical consequences.

La difficulté que nous avons pour considérer que HPSG est une théorie vient de deux motifs :

- i Nous n'avons pas réussi à calculer, à partir de deux descriptions [Sag et Miller 97] et [Abeillé 97] du domaine qui nous occupe toutes et seules les suites susceptibles d'être spécifiées par ces descriptions et les structures qui leur sont associées
- ii Nous ne percevons pas dans les écrits consultés de HPSG, une piste permettant de fixer les conditions pour tester une *théorie* HPSG (et non simplement une description écrite dans le système descriptif HPSG)²⁸.

28. Les propositions descriptives dans HPHG sont certes parfois présentées sous forme de structures de traits qui, elles devraient être en principe calculables. Mais, déjà complexes en elles-mêmes, avec des opérateurs de concaténation et d'unification, lorsqu'elles sont effectivement présentées avec des principes universaux et l'adaptation de ces principes à des situations particulières non entièrement définies de manière formelle, plus des relations d'héritage dans le lexique, elles deviennent impossibles à calculer. C'est ce que l'on observe dans [Sag et Miller 97], où sont présentées successivement les solutions pour traiter les clitics dans différents contextes. Par ailleurs, pour pouvoir considérer HPHG comme une théorie, au sens fixé ci-dessus, ce système descriptif devrait pouvoir incorporer des limites de principe à la possibilité actuelle de multiplier les entrées homomorphes. Des observations analogues peuvent être faites à propos du traitement de ce domaine en grammaire générative par des transformations — opérations à ma connaissance jamais formalisées, que ce soit dans [Roulet 69], [Kayne 77], pour la tradition chomskienne, ou dans [Gross 68] avec une conception différente des transformations. Dans aucun de ces travaux, il n'est possible de calculer toutes et seulement toutes les conséquences des descriptions proposées.

Les carences de la description ici proposée sous forme de *P2* relèvent de deux sources. L'une est constituée par les limites de l'adéquation externe du système, l'autre par les limites des modules internes à 5P dans leur formulation actuelle.

Les limites descriptives ont été signalées § 2 et dans la discussion des Niveaux 1 à 5. Une insuffisance ici est l'expression de l'accord, avec des phénomènes relativement bien repérés mais mal exprimés en traitement automatique, où on a l'impression que l'on s'est limité à exprimer ce que l'on peut traiter avec les traits classiques de genre, nombre et personne et le mécanisme de l'unification. L'introduction de *on*, le traitement du *vous* singulier pour l'adjectif et pluriel pour le verbe, des impersonnels, de *ce*, du genre du possessif (où on voit mal le masculin de *son* dans *son adorée artiste*) passe par la possibilité de donner une solution formelle à l'expression de l'accord, ceci sans compter l'accord dans la coordination. Par ailleurs, l'identification des problèmes descriptifs liés aux chaînes adverbiales a été à peine entamée avec la description proposée pour quelques cas de *toujours* et des formes minimales de la négation.

Même dans le cadre d'une adéquation externe exigeante par rapport à des textes effectifs, les questions précédentes, à l'exception des celles relevant de la coordination, semblent pouvoir être traitées dans le formalisme proposé pour les *P2*, et pouvoir être testées, de même que les Propriétés d'existence et de linéarité des Niveaux 2 à 5, par le GénMod dans sa version actuelle²⁹. Mais la formulation actuelle des modules internes de 5P introduit un autre type de limites : celles qui résultent de la problématique d'incorporer des descriptions sur des domaines limités (cf. les notions de *shallow parsing* et de microgrammaire) aux descriptions des suites correspondant à des énoncés.

On a remarqué dans § 7.5 la surgénération possible des Propriétés (*P2*) proposées et la possibilité de la réduire. Dans § 6 on a esquissé la possibilité de calcul de la sémantique. Si l'on se réfère aux travaux du GARS, le domaine visé concerne principalement ce qui est traité dans l'approche pronominale en termes de règles d'ordre et de co-occurrence, de constitution de séries et d'exclusions dans les séries. Tout ce qui concerne la notion de proportionalité et les relations des clitiques dans la phrase verbale noyau avec les autres constituants — domaine qui est aussi

29. Le GénMod a déjà été formellement spécifié [cf. Bès 98b] mais il n'est pas implanté. La spécification formelle a permis de calculer "à la main" [cf. Bès 98c] les lignes basiques de § 5, mais le test des descriptions dans les Niveaux 2 à 5, par l'obtention des lignes basiques, tâche fastidieuse et longue à effectuer sans un outil informatique, n'a pas été effectué.

visé dans les descriptions HPSG —, et des clitiqes avec les entrées lexicales, ne peut pas être abordé avec les éléments descriptifs présentés § 3.

Ceux-ci sont des points clefs. La proportionalité dans l'approche pronominale semble être une notion particulièrement féconde, car elle permet de calculer beaucoup d'information à partir de peu d'information; en effet elle "permet de restreindre le corpus d'observation aux phrases pronominales, qui sont représentatives pour la syntaxe fondamentale" [Blanche-Benveniste et al. 84, p. 29].

Pour aborder ces questions on doit se donner en 5P, pour améliorer les Propriétés (P2)³⁰ au moins trois éléments supplémentaires. L'un est la possibilité d'accéder à une information lexicale plus riche que celle qui est nécessaire dans les catégories proposées, information qui doit incorporer, parmi d'autres, l'information sémantique associée à l'entrée. L'autre consiste à formaliser l'insertion d'un modèle dans un autre, et en particulier, à transporter le fléchage, s'il est nécessaire pour le calcul de la sémantique, de l'un sur l'autre. Cette insertion est effectuée dans 5P par la règle de substitution du GénMod, opération non présentée car non utilisée dans la description proposée. Enfin, le troisième élément supplémentaire est le calcul effectif de la représentation sémantique à partir du fléchage et des entités qui fléchent dans un modèle déterminé. Mais ce sont, comme dirait Shéhérazade, des thèmes pour d'autres histoires.

Gabriel BÈS

GRIL

Université Blaise-Pascal (Clermont II)



30. On n'a pas évoqué dans ce travail les processus effectifs d'analyse automatique dans P5 (sur ces points cf. [Hagège 98] et [Blache 98a et 98b]) ni le test d'adéquation externe du fléchage par le biais de la représentation sémantique que celui-ci permet de construire, ce qui exige l'enrichissement des Protocoles.

Bibliographie

- [Abeillé et al. 97] ABEILLÉ, Anne, Danièle GODARD et Philip MILLER. *The Major Syntactic Structures of French*. Esslli Summerschool, Aix-en-Provence, 1997.
- [Abeillé 97] ABEILLÉ, Anne. "La compréhension des auxiliaires français". Dans [Abeillé et al. 97], préalablement paru dans *Langages*, 1996 (122), p. 32-61.
- [Abney 92] ABNEY, Steven P. "Parsing by chunks". Dans R.C. BERWICK, S.P. ABNEY et C. TENNY [ed.]. *Principle-based Parsing : Computation and Psycholinguistics*. Dordrecht. Kluwer Academic Publishers, 1992, p. 257-278.
- [Aït-Mokhtar et Chanod 97a] AÏT-MOKHTAR, Salah et Jean-Pierre CHANOD. "Incremental Finite-State Parsing". Dans *Proceedings of ANLP'97*, Washington, avril 1997.
- [Aït-Mokhtar et Chanod 97b] AÏT-MOKHTAR, Salah et Jean-Pierre CHANOD. "Subject and Object Dependency Extraction Using Finite-State Transducers". Dans *ACL/EACL'97 Workshop on Automatic Information Extraction and Building of Lexical Semantic Ressources for NLP Applications*, Madrid, juillet 1997.
- [Auroux 81] AUROUX, Sylvain. "Falsification et induction". Dans *Diogène*, 1981 (XX-2), p. 282-307.
- [Bès 98a] BÈS, Gabriel G. *Protocoles, Propriétés, Processus*. Rapport de Recherche, GRIL, 1998.
- [Bès 98b] BÈS, Gabriel G. *Le générateur de modèles*. Rapport de Recherche, GRIL, 1998.
- [Bès 98c] BÈS, Gabriel G. *Le calcul des suites basiques de la phrase verbale noyau*. Rapport de Recherche, GRIL, 1998.
- [Bès, Blache et Hagège, en préparation] BÈS, Gabriel G., Philippe BLACHE et Caroline HAGÈGE. "5P : A Paradigm for the Formalisation of Natural Language Descriptions and their Automatic Processing".
- [Blache 98a] BLACHE, Philippe. "Parsing Ambiguous Structures Using Controlled Disjunctions and Unary Quasi-Stress". Dans *Proceedings of COLING'ACL'98*.
- [Blache 98b] BLACHE, Philippe. "Proof Nets for Controlling Ambiguity in Natural Language Processing". Dans *Proceedings of ICTAI'98*.
- [Blache 98c] BLACHE, Philippe. *Des principes et des propriétés pour une nouvelle approche syntaxique en TALN*. Projet de recherche. LPL-CNRS (ESA 6057), Aix-en-Provence, 1998.
- [Blanche-Benveniste 75] BLANCHE-BENVENISTE, Claire. *Recherches en vue d'une théorie de la grammaire française; essai d'application à la syntaxe des pronoms*. Paris, Honoré Champion, 1975.

- [Blanche-Benveniste 96] BLANCHE-BENVENISTE, Claire. "De l'utilité du corpus en linguistique". Dans *Revue française de linguistique appliquée*, 1996 (1-2), p. 25-42.
- [Blanche-Benveniste et al. 84] BLANCHE-BENVENISTE, Claire, José DEULOFEU, Jean STEFANINI et Karel VAN DEN ÉYNDE, *Pronom et syntaxe; l'approche pronominale et son application au français*. Paris, Selaf, 1984.
- [Bochenski 58] BOCHENSKI, I.M. *Los métodos actuales del pensamiento*; 2^e éd. Madrid, Rialp, 1958. Traduction à l'espagnol de Raimundo Drudis Baldrich.
- [Bresnam et Kaplan 82] BRESNAM, Joan et Ronald M. KAPLAN. "Introduction : Grammars as Mental Representations of Language". Dans BRESNAM, Joan [ed.]. *The Mental Representations of Grammatical Relations*. Cambridge, Mass., The MIT Press, 1982.
- [Bunge 69] BUNGE, Mario. *La investigación científica ; su estrategia y su filosofía*. Barcelona, Ariel, 1969. Traduction à l'espagnol de Manuel Sacristán.
- [Chanod 95] CHANOD, Jean-Pierre. "Introduction à la table ronde TALN et Linguistique". Dans *Actes TALN-95*, Marseille, juin 1995, p. 4-6.
- [Coppieters 97] COPPIETERS, René. "Quelques réflexions sur la question des données : corpus et intuitions". Dans *Recherches sur le français parlé*, 1997 (14), p. 21-41.
- [Emirkanian et Bouchard 95] EMIRKANIAN, Louisette et Lorne H. BOUCHARD. "De la description linguistique à l'implantation de la syntaxe du français". Dans *Actes TALN-95*, Marseille, juin 1995, p. 7-15.
- [Frederici et al. 96] FEDERICI, Stefano, Simonetta MONTEMAGNI et Vito PIRELLI. "Shallow Parsing and text Chunking : A View on Inderspecification in Syntax". Dans *Proceedings ESSLLI'96 Workshop on Robust Parsing*. Prague, 1996.
- [Granger 92] GRANGER, Gilles-Gaston. *La vérification*. Paris, Odile Jacob, 1992.
- [Grevisse 93] GREVISSE, Maurice. *Le bon usage ; grammaire française*. 13^e édition, refondue par André Goosse. Paris, Duculot, 1993.
- [Gross 68] GROSS, Maurice. *Grammaire transformationnelle du français ; syntaxe du verbe*. Paris, Larousse, 1968.
- [Habert et al. 97] HABERT, Benoît, Adeline NAZARENKO et André SALEM. *Les linguistiques de corpus*. Paris, Colin, 1997.
- [Hagège 98] HAGÈGE, Caroline H. *Les algorithmes d'analyse par feuilles*. GRIL, Rapport de recherche, 1998.
- [Hagège et Bès 98] HAGÈGE, Caroline et Gabriel G. BÈS. "Da observação da propriedades linguísticas à formalizaõ numa gramática do processamento da lingua". Dans *Actes du III Encontro para o processament computacional da lingua portuguesa escrita e falada*. Porto Alegre, Brazil, 3-4 novembre 1998.
- [Hindle 94] HINDLE, Donald. "A Parser for Text Corpora". Dans A. Zampolli [ed.]. *Computational Approaches to the Lexicon*. New York, Oxford University Press, 1994, p. 103-130.

- [Hintikka 94] HINTIKKA, Jaakko. *Fondements d'une théorie du langage*. Paris, PUF, 1994.
- [Karlsson et al. 95] KARLSSON, Fred, Aro VOUTILAINEN, Juha KEIKKILÄ et Arto ANTILA [ed.]. *Constraint Grammar : A Language-Independent System for Parsing Unrestricted Text*. Berlin..., Mouton-de Gruyter, 1995.
- [Katz 81] KATZ, Jerrold J. *Language and Other Abstract Objects*. Totowa, New Jersey, Rowman and Littlefield, 1981.
- [Kayne 77] KAYNE, Richard S. *Syntaxe du français ; le cycle transformationnel*. Paris, Seuil, 1977.
- [Kerbrat-Orecchioni 80] KERBRAT-ORECCHIONI, Catherine. *L'énonciation ; de la subjectivité dans le langage*. 2^e éd. Paris, Colin, 1980.
- [Pollard et Sag 94] POLLARD, Carl et Ivan A. SAG. *Head-Driven Phrase Structure Grammar*. Chicago..., The University of Chicago Press, 1994.
- [Roulet 69] ROULET, Eddy. *Syntaxe de la proposition nucléaire en français parlé ; étude tagmémique et transformationnelle*. Bruxelles, Aimav, 1969.
- [Ruwet 68] RUWET, Nicolas. *Introduction à la grammaire générative*. Paris, Plon, 1968.
- [Sag et Miller] SAG, Ivan A. et Philip MILLER. "French Clitic Movement without Clitics or Movement". Dans [Abelle et al. 97], p. 1-67.
- [Sokal et Bricmont 97] SOKAL, Alan et Jean BRICMONT, *Impostures intellectuelles*. Paris, Odile Jacob, 1997.
- [Tapanainen et Järvinen 97] JÄRVINEN, Timo et Pasi TAPANAINEN. "A Non-Projective Dependency Parser". Dans *Proceedings of ANLP'97*, Washington, avril 1997.
- [Valli 95] VALLI, André. "Notes sur la variation linguistique en français". Dans *Recherches sur le français parlé*, 1995(3), p. 91-109.