

G.A.R.S.
Groupe Aixois de Recherches en Syntaxe

RECHERCHES
SUR
LE FRANÇAIS
PARLÉ

N° 15 – 1999

1999

Publications de l'Université de Provence
29, avenue Robert Schuman – 13621 Aix-en-Provence – Cedex 1

© Université de Provence
Service des Publications

Dépôt légal - 4^e Trimestre 1999
ISBN - N° 2-85399-444-9

Sommaire

Présentation, par José DEULOFEU	7
Berthille PALLAUD, Lapsus et phénomènes voisins dans la langue parlée. Problèmes d'identification	9
Marie-Noëlle ROUBAUD et Claude LOUFRANI, Éclairage des corpus de type aphasique par les grilles	41
André VALLI, Remarques sur le français parlé de locuteurs de la région de Marseille. Usage régional du français et "régionalisme"	59
Claire BLANCHE-BENVENISTE et Jean-Pierre ADAM, La conjugaison des verbes : virtuelle, attestée, défective	87
Éric CASTAGNÉ, Apports des données orales dans l'analyse des valences à l'infinitif des adjectifs recteurs	113
Modou N'DIAYE, Les constructions comparatives en <i>plus/moins... que</i>	137
José DEULOFEU, Questions de méthode dans la description morphosyntaxique de l'élément <i>que</i> en français contemporain	163
Paul CAPPEAU, Sujets éloignés. Esquisse d'une caractérisation des sujets lexicaux séparés de leur verbe	199
Philippe MARTIN, Prosodie des langues romanes : analyse phonétique et phonologie	233
Mireille BILGER, Coordination : analyses syntaxiques et annotations	255
Gabriel BÈS, La phrase verbale noyau en français	273



Présentation

Ce numéro contient des études qui s'inscrivent dans deux directions de recherches complémentaires de l'équipe du GARS.

La majorité d'entre elles sont consacrées à la poursuite du travail de description de la morphosyntaxe du français contemporain dans le cadre de l'Approche Pronominale. Des concepts descriptifs et des hypothèses formulés dans ce cadre sont ainsi mis à l'épreuve de données nouvelles, provenant, notamment, de l'examen de productions authentiques de français parlé.

Dans cette orientation, plusieurs articles s'intéressent à la relation description – données. Ils soulignent la nécessité de hiérarchiser les données avant d'entreprendre la description. B. Pallaud s'intéresse aux données généralement écartées de la description : les lapsus. Elle montre qu'une typologie sur des bases syntaxiques peut en être établie, éclairant les rapports entre le système de la langue et sa mise en œuvre par les locuteurs dans la situation d'énonciation. M.-N. Roubaud et C. Loufrani utilisent l'outil de l'analyse en grilles pour traiter les données particulières que sont les productions de locuteurs aphasiques. A. Valli revient sur le statut du régionalisme en syntaxe, dont la délimitation est essentielle si l'on veut travailler sur des données homogènes.

La question de la sous-estimation du système par les usagers, déjà soulevée en syntaxe par le GARS, est abordée en morphologie par C. Blanche-Benveniste, qui s'appuie sur le traitement informatique des données de J.-P. Adam pour distinguer entre le virtuel (permis par le système, sous-utilisé dans l'usage) et le défectif (possibilités limitées par le système lui-même). E. Castagné montre que la description des constructions *il est déjà ADJ à Vinf* est facilitée si l'on répartit les données en deux sous-ensembles sous-tendus par la distinction grammaticale première – grammaticale seconde, proposée par des études antérieures.

Du point de vue des outils descriptifs, M. N'Diaye distingue, par une étude minutieuse des proportionnalités pronominales, deux types de cons-

tructions comparatives, montrant qu'il reste encore à dire sur la grammaire de ces constructions.

Les autres articles explorent principalement la fécondité de l'hypothèse du recours à une composante macrosyntaxique pour l'étude du fonctionnement aussi bien de morphèmes (J. Deulofeu) que de relations : sujet / verbe (P. Cappeau), coordination (M. Bilger).

En complément de ces études portant sur des unités segmentales, P. Martin propose une analyse prosodique d'une production de français parlé en élargissant la problématique à une comparaison entre langues romanes.

Pour la seconde orientation, deux articles tentent d'établir un lien entre les études descriptives et les formalismes utilisés dans le traitement automatique des données linguistiques. Cette direction nouvelle s'est imposée devant les difficultés à rassembler rapidement à travers des transcriptions de corpus authentiques les données nécessaires à la validation empirique de nos descriptions morphosyntaxiques. L'extension à ces données des modèles utilisés pour le traitement syntaxique automatique des textes (tagging, parsing) suppose un dialogue entre informaticiens et linguistes descripteurs, chacun devant s'adapter aux exigences du travail de l'autre. Elle répond aussi au souci de poser, à plus long terme, la question du ou des formalismes qui seraient adaptés à nos descriptions.

M. Bilger propose d'engager la discussion du point de vue du linguiste par une réflexion sur la modélisation informatique, en vue d'un tagging, du traitement proposé par l'approche pronominale pour certains types de coordination. Ce traitement utilise la notion de "séquence nexus" — structurée par une relation macrosyntaxique préfixe – noyau — comme alternative à l'analyse par "gapping". La question cruciale du type d'annotation nécessaire à la représentation des relations macrosyntaxiques est ainsi abordée, au-delà du problème ponctuel de la coordination.

L'article de G. Bès vise à répondre, du point de vue de l'informaticien, à une préoccupation plus théorique : quel modèle, vérifiable par le calcul informatique, construire pour générer les séquences de mots constituant la "phrase noyau" en français ? Il renoue ainsi, en l'élargissant aux problèmes de formalisation, jusqu'ici non abordés, avec la question de la spécificité des règles de fonctionnement de la "séquence verbale", à partir de quoi a été construite la syntaxe de l'Approche Pronominale.

José DEULOFEU
Université de Provence

Lapsus et phénomènes voisins dans la langue parlée. Problèmes d'identification.

Les lapsus (des erreurs de langage, selon certains) ont fait l'objet de nombreuses études depuis un siècle et cela dans des disciplines aussi variées que la linguistique, la psycholinguistique et la psychanalyse. L'analyse des travaux dans ce domaine montre que si les efforts ont porté sur l'explicitation des lapsus, le problème même de leur identification a été rarement soulevé, comme si chacun se satisfaisait d'une reconnaissance implicite du lapsus¹. Or l'établissement des faits en ce domaine ne va pas de soi.

La question qui est posée dans notre étude concerne donc l'identification des lapsus *linguae*². Peut-on dire que les lapsus forment une catégorie homogène ? Quelles sont leurs caractéristiques ? Leur définition peut-elle être strictement linguistique ? L'identification des lapsus pose-t-elle des problèmes spécifiques ? Varie-t-elle en fonction des théories linguistiques et psychanalytiques ?

Il est évident que cette étude ne prétend pas répondre de façon approfondie à toutes ces questions. Elle vise seulement à en montrer une approche linguistique, approche permise par une méthode descriptive fondée sur les travaux de l'équipe du Groupe Aixois de Recherche en Syntaxe et par l'exploitation de corpus enregistrés dans le cadre de ces travaux.

1. *Lapsus* est un terme lié dès son origine à la notion d'erreur. Sa définition dans le *Trésor de la Langue Française* (1983, t. 10, p. 992) est la suivante : "faute que l'on commet par inadvertance soit en parlant (*lapsus linguae*), soit en écrivant (*lapsus calami*)".

2. Les *lapsus linguae* sont ceux que l'on produit en parlant lorsque, suivant l'expression choisie par Rossi et Peter-Defare (1998), "notre fourche a langué".

1. SYNTHÈSE SUR LES TRAVAUX DANS LE DOMAINE DES LAPSUS

1.1. QUELQUES REMARQUES SUR LES TRAVAUX ANTÉRIEURS CONCERNANT LES LAPSUS³

Meringer (philologue) et Mayer (psychiatre) sont considérés comme ayant été les premiers à avoir véritablement étudié les lapsus (1895, 1908, 1923) et constitué à la fin du XIX^e siècle un corpus de 8 800 lapsus (*lapsus linguae, calami et lectionis*); les *lapsus linguae* sont au nombre de 4 400. Il n'est pas une étude traitant des lapsus ou ratés de langage qui ne fasse référence aux travaux de ces auteurs allemands.

Freud connaissait ces recherches auxquelles il se réfère longuement puisqu'il emprunte à leur corpus un certain nombre d'exemples qui viennent étayer ses propres théories; il fait référence à ces travaux à plusieurs reprises dans son livre *Psychopathologie de la vie quotidienne* (1901) et analyse à son tour plusieurs de leurs exemples. Inutile de dire que Meringer protestera vigoureusement contre cet emprunt "redirigé". Comme Freud le dit lui-même, si les thèses de Meringer et Mayer lui ont semblé pendant un temps corroborer ses propres observations, le conflit est devenu certain après la parution d'écrits postérieurs de Meringer (Freud, 1901, p. 173).

En 1981, un numéro spécial de la revue *Linguistics* a été consacré aux lapsus (Cutler A., ed.): on y trouve une synthèse sur les travaux essentiellement anglosaxons et une réflexion sur les positions théoriques adoptées, les points de vue causal et explicatif, les difficultés méthodologiques spécifiques au recueil de ces données et leurs résultats descriptifs, etc. Il faut reconnaître que la controverse, entre certains linguistes et psychanalystes, sur le statut des lapsus, leur analyse et leurs causes, reste entière; et on ne voit pas pourquoi elle cesserait de l'être puisque leur objet et leurs objectifs diffèrent.

Cutler (1981), dans ce numéro de *Linguistics* tente de montrer que la controverse repose en grande partie sur une confusion entre les causes et les mécanismes. En effet, on est au niveau des mécanismes quand "on a clairement montré que les exemples traités par Freud pouvaient être soumis aux mêmes critères objectifs de classification que toutes les autres erreurs de langage" (Rossi et Peter-Defare, 1998, p. 14). Ce n'est pas

3. Pour une présentation plus systématique des études déjà faites dans ce domaine se reporter au livre récent de Rossi M. & Peter Defare E. (1998), mais aussi à ceux de Cutler (1982) et Goffman (1981).

pour autant que les explications avancées par Freud sur leurs causes, peuvent être ignorées ; or, chez certains linguistes, elles l'ont été largement. Il est évident qu'il y a intérêt⁴ à distinguer l'interprétation concernant les causes et l'interprétation au niveau des mécanismes.

Pour Cutler, lorsqu'on reconnaît que le débit des erreurs augmente en cas de fatigue ou sous l'empire de l'alcool ou d'une drogue, on est alors au niveau des causes et non au niveau des mécanismes. La fatigue et l'ébriété ne sont pas des mécanismes ; ce sont des états qui favorisent ou causent le déclenchement plus fréquent de mécanismes existants. "It is never suggested that fatigue and drunkenness are alternative mechanisms by which errors can occur, merely that such states can precipitate, or cause, the more frequent operation of the existing mechanisms" écrit-elle. Et les linguistes s'occuperaient de ces mécanismes, les causes leur restant inaccessibles, dans l'état actuel des connaissances du moins.

Les linguistes qui ont étudié les lapsus ont mis au rang des causes une grande diversité de phénomènes tels que :

- l'influence de la fatigue, de la hâte, de la distraction ou de produits toxiques dont on a constaté les effets sur la fréquence des lapsus (c'est ce que, on l'a vu, Cutler a mentionné).

- une pensée perturbatrice qu'elle soit consciente mais non dite ou qu'elle soit inconsciente⁵ (Meringer et Mayer⁶, 1895 ; Cutler, 1981 ; Arnaud, 1997). Ce point de vue a des conséquences méthodologiques importantes puisqu'il devient nécessaire pour une étude approfondie des lapsus d'interroger le locuteur sur "Ce à quoi il pense pendant qu'il parle" (Butterworth, 1982). Rossi et Peter-Defare (1997) qui fondent leur étude

4. Intérêt pour la rigueur et intérêt pour le style du débat.

5. Il est vraisemblable que Cutler (1982), Arnaud (1997) emploient le terme d'*inconscient* dans son sens courant c'est à dire de non-conscient. La psychanalyse en donne une définition à la fois plus précise et hors du champ de la linguistique : "La non pensée peut être la cause de la pensée : c'est ce à quoi nous sommes confrontés avec la notion d'inconscient" (Lacan (1968-1969), *D'un Autre à l'autre*, p. 9).

6. On notera que Freud, dans son chapitre sur les lapsus, soutient explicitement le point de vue de Meringer et Mayer : "Par sa théorie des images verbales 'nomades', qui sont situées au dessous du seuil de la conscience et qui ne sont pas destinées à être formulées en paroles, et par son insistance sur la nécessité de rechercher tout ce à quoi le sujet pense pendant qu'il parle, la conception de Meringer et Mayer, se rapproche singulièrement, il est facile de s'en rendre compte, de notre conception psychanalytique. Nous recherchons *nous aussi, des matériaux inconscients*, et de la même manière, à cette seule différence près que nous prenons un détour plus long, puisque nous n'arrivons à la découverte de l'élément perturbateur qu'à travers une chaîne d'associations complexe, en partant des idées qui viennent à l'esprit du sujet lorsque nous l'interrogeons" (Freud, 1901, p. 66).

sur des lapsus entendus dans la conversation courante ou relevés dans des dialogues et tables rondes diffusés sur les médias entre 1992 et 1996, choisissent, quant à eux, de ne pas interroger les locuteurs (même quand cela est possible) sur les raisons de leur lapsus : “Nous aurions en procédant différemment interrompu indûment le flux de la parole et faussé les résultats” (p. 17).

– l’anxiété et surtout certains types d’anxiété peuvent être à l’origine de lapsus. Ce sont les études sur les lapsus suscités artificiellement⁷ qui ont émis cette hypothèse selon laquelle les lapsus seraient “des révéléateurs des anxiétés” (Motley, 1985, p. 64).

– la contagion (une fois qu’une erreur a été faite, d’autres erreurs sont faites par les locuteurs qui participent à l’interaction) : Cutler souligne ce fait qui avait intrigué Freud à propos de l’oubli de nom (et pour lequel il a déclaré ne pas avoir d’explication 1901, p. 48). Goffman (1981, p. 204) ajoute à ce type de contagion qu’on pourrait dire “inter-individuelle”, une contagion intra-individuelle : “The occurrence of one imperfection increases the chance of another, and that in turn increases the chance of another ones — as if, indeed, there were such a thing as getting rattled”.

1.2. LES ETUDES SUR LES LAPSUS EN LANGUE FRANÇAISE

Rossi et Peter-Defare (1998) signalent un regain d’intérêt pour les lapsus en langue française depuis 1992. Cet intérêt se manifeste dans le fait que plusieurs équipes se lancent dans l’établissement et l’analyse de corpus (enregistrés ou non). Actuellement, en dehors de notre groupe, trois équipes⁸ étudient ces phénomènes.

7. C’est aux États-Unis et surtout durant la période de 1975 à 1990 que des linguistes (Bernard Baars et Michael Motley à San Francisco, Carl Camden de l’Université de Cleveland) ont étudié des lapsus suscités artificiellement en laboratoire. Les méthodes employées permettent de provoquer artificiellement des lapsus chez des sujets lisant des listes de paires de mots sur un écran. Ces méthodes qui se fondent sur les effets antérieurs du contexte visent à produire des contrepèteries en suscitant des indécisions cognitives. Ces études, selon leurs auteurs, ont pour objectif de combler une lacune que l’étude des lapsus naturels ne permet pas de combler : “déterminer la façon dont les composants linguistiques sont traités dans le cerveau”. À ce courant de recherche se rattachent les travaux de Pillon (1998) qui, toujours à l’aide de lapsus déclenchés artificiellement montre l’implication de la morphologie dérivationnelle des mots dans les processus de production de la parole.

8. Je mettrai à part les travaux sur les lapsus déclenchés artificiellement ; les études de Pillon (1998) à Louvain sont représentatives de ce courant de recherche.

– À l'Institut de Phonétique de l'Université de Provence une série de recueils et d'études ont abouti à plusieurs publications dont les principales sont celles de Rossi et Peter-Defare (1995, 1998) qui portent sur environ 4 000 exemples tous reconnus par le locuteur lui-même comme lapsus. Il est à noter que, Peter-Defare E., 1993 a réalisé la première étude systématique et linguistique dans ce domaine en Langue française.

– Arnaud (1997) a publié récemment une première analyse sur son propre recueil de lapsus qu'il qualifie de "naturels" (pour les distinguer des lapsus déclenchés en laboratoire, ou suscités artificiellement). La notation de ces 2400 exemples a été faite sur le mode systématique ("tout lapsus entendu a été noté", p. 308) et cela dans les secondes qui suivent leur production ("afin d'éviter la rapide dégénérescence de la trace mémorielle", p. 308).

– Fénoglio (1997) s'appuie sur un corpus de 25 lapsus relevés dans des entretiens enregistrés pour aborder l'énonciation en discours oral ; le lapsus apparaît comme une donnée d'articulation entre discours et parole.

1.3. DESCRIPTION LINGUISTIQUE DU LAPSUS

Dans le champ de la linguistique, les lapsus font, actuellement l'objet d'études cognitives ; ils constituent des preuves permettant d'étayer divers modèles de production et de perception du langage : accès au lexique et planification de la parole. Cette approche est représentée, en France, par celle d'Arnaud et celle de Rossi et Peter-Defare chez qui elle est particulièrement bien documentée.

Dans le domaine de la psychanalyse, les perspectives sont différentes, ne serait ce que parce que la psychanalyse n'est ni une théorie linguistique, ni une théorie cognitive. L'intérêt pour les lapsus tient justement à ce qu'ils révèlent du fonctionnement psychique humain et qui peut être retrouvé dans l'oubli, l'acte manqué, etc. (toute la table des matières du livre de Freud *Psychopathologie de la vie quotidienne* : 1899), les rêves et les psychopathologies. Les études de Freud sur les lapsus contribuent à son élaboration théorique sur l'inconscient. La psychanalyse est une théorie de l'inconscient ; ce n'est pas une théorie de la parole ou des langues.

Il n'est évidemment pas possible de décrire les lapsus sans faire référence aux hypothèses que l'on a sur leur formation. La description la plus courante, adoptée par certains linguistes, met en cause l'élaboration de l'énoncé sur les axes syntagmatique et paradigmatique. C'est le cas, par exemple, des études récentes sur le français (Rossi et Peter-Defage, 1998, et Arnaud, 1997) :

Selon le critère contextuel, chaque type de lapsus est i/ soit syntagmatique, si l'origine est présente dans le contexte de l'énoncé..., ii/ soit paradigmatique, si l'origine est absente de ce contexte et si l'erreur est provoquée par un phénomène associatif quelconque. (Rossi et Peter-Defare, p. 42)

Deux sortes de lapsus sont ainsi distingués selon que leur origine est attribuée à l'élaboration de la parole sur l'axe syntagmatique ou sur l'axe paradigmatique. Rossi et Defare (1998) établissent alors un premier classement fondé sur la distinction entre quatre types d'extension du lapsus selon que l'erreur porte sur le mot, la syllabe, le phonème ou le trait phonologique. Six types d'opération⁹ semblent expliquer la perturbation dont les lapsus résultent. Il est à noter que, selon la définition des lapsus dits "d'origine paradigmatique", leur identification n'est pas du même ordre que les autres types de lapsus. En fait, la catégorie "paradigmatique ou non contextuelle" n'est établie que sur des critères "négatifs" : le lapsus est déclaré "paradigmatique" lorsqu'il est impossible de trouver dans le contexte antérieur ou ultérieur l'origine de l'altération¹⁰. Ce type de lapsus a ainsi un statut de "reste".

Exemple n° 55 :

il se trouve qu'on (*dépasse*), qu'on **dépense** énormément de pognon [247].

Le lien sémantique entre "dépasse" et "dépense" va de soi car on dépasse son budget à "dépenser" trop d'argent, quant à la parenté phonologique elle est évidente. (55) est une substitution de mots qu'*aucune origine phonologique dans le contexte* ne peut expliquer ici. (Rossi et Peter-Defare, 1998, p 36).

Il semble très artificiel de vouloir affirmer l'existence de deux sources de lapsus alors même que les auteurs insistent sur la différence radicale qui existe entre ces deux sources. Elles n'appartiennent pas au même champ ; il ne peut donc pas y avoir de contradiction épistémologique à ce qu'elles puissent intervenir simultanément dans la production de la parole.

9. Amalgame, haplogie, omission, insertion, interversion et substitution.

10. Ce raisonnement est quelque peu circulaire puisque les auteurs posent en prémisses que les erreurs pouvant se *décrire* par des déplacements sur l'axe syntagmatique *s'expliquent* par ces perturbations sur l'axe syntagmatique et n'ont rien à voir avec une élaboration sur l'axe paradigmatique. L'insistance est forte sur cette hypothèse d'une origine contextuelle du lapsus ; elle se retrouve même dans les recommandations pour le recueil des données. "Nous n'avons retenu que les lapsus dont nous étions sûrs et pour lesquels nous pouvions obtenir un contexte suffisant où pouvait être identifiée l'origine éventuelle" (Rossi et Peter-Defare, 1998, p. 17).

D'ailleurs, peut-il y avoir une parole en dehors d'une élaboration simultanée sur les axes syntagmatique et paradigmatisque ? Évidemment, non.

On trouve chez Freud (1967, p. 70) la remarque suivante qui vient souligner la nécessité d'une approche plus complète et non divisée du lapsus :

Parmi les exemples que j'ai moi-même réunis, je n'en trouve guère où le trouble du langage se laisse réduire uniquement et exclusivement à ce que Wundt appelle "l'action par contact de sons". Je trouve presque toujours, *en plus de l'action par contact*, une action perturbatrice ayant sa source en dehors du discours qu'on veut prononcer, et cet élément perturbateur est constitué soit par une idée unique, restée inconsciente, mais qui se manifeste par le lapsus et ne peut le plus souvent être amené à la conscience qu'à la suite d'une analyse approfondie, soit par un mobile psychique plus général qui s'oppose à tout l'ensemble du discours.

Les hypothèses pour rendre compte des lapsus dits "paradigmatiques" sont diverses. Dans certains cas, les auteurs invoquent "le contexte de production du lapsus" comme dans l'exemple suivant :

15 Au pays du (*sommeil*), du **soleil** levant [208] ;

les auteurs font le commentaire qui suit :

En (15) aucune persévérance ni anticipation de consonne ne permet de rendre compte du lapsus ; en revanche, *sommeil* et *soleil levant* se trouvent quasiment dans une relation sémantique d'antonymie et le contexte de production du lapsus a pu faciliter la substitution des mots : il était 15 heures, l'heure de la sieste, et il faisait très chaud.

Si Rossi et Peter-Defare (1998) envisagent un recours éventuel à des indices liés aux circonstances de production des énoncés pour expliquer certains lapsus, Arnaud (1997), quant à lui, insiste beaucoup sur la nécessité de recueillir des informations précises sur les circonstances d'énonciation. Il précise même que cette "intrusion environnementale" peut être un véritable matériau inconscient puisque l'énonciateur n'a pas conscience de cette influence :

(9) (l'énonciateur mange des sardines en boîte) ; *elles sont bonnes ces olives* (au lieu de sardines). Il est écrit sur la boîte — chose dont l'énonciateur *n'avait pas conscience* — qu'il s'agit de sardines à l'huile d'olive. (p. 313)

Pour cet auteur, l'absence de conscience de "percepts" notée pour certains lapsus (autrement dit l'influence inconsciente d'unités lexicales présentes dans l'environnement du locuteur) est importante sur le plan théorique :

Ceci montre qu'une forme linguistique isolée peut entrer dans un processus d'accès lexical sans se signaler à la conscience. On sait qu'un des corollaires de la modularité est le caractère automatique et incoercible des processus de bas niveau. (p. 323)¹¹

Concernant les lapsus d'origine syntagmatique, Rossi et Peter-Defare (1998) signalent également qu'un certain nombre de lapsus sont "provoqués" par des matériaux à trouver dans un énoncé plus large, ce qui les a conduits à introduire les notions de contexte à court terme ou à long terme, notions qui ne sont cependant pas précisées. Par analogie, sans doute, avec les notions de mémoire à court terme et à long terme¹², on déduit des exemples analysés qu'il s'agit de l'énoncé proche voire immédiat (quelques mots avant ou après le lapsus) pour le court terme. Le contexte à long terme serait une partie plus vaste de la conversation englobant le lapsus. Lorsque "l'origine" du lapsus ne peut être retrouvée dans le contexte à court terme, elle peut être "trouvée" dans le contexte à long terme ; par exemple, pour le lapsus (57) "Vos (p)leurs, vos *f*leurs résistent bien [950]", Rossi et Peter-Defare (1997, p. 35) font le commentaire qui suit :

Si on se réfère au contexte à plus long terme, on apprend que le locuteur a *auparavant* hésité entre vos "fleurs résistent" et vos "fleurs prospèrent" ; on est alors en droit de considérer la substitution "fleurs => pleurs" comme un effet phonologique de l'initiale de "prospèrent".

2. POINT DE VUE MÉTHODOLOGIQUE POUR NOTRE RECUEIL DE LAPSUS

Depuis plus de 25 ans maintenant, l'équipe du GARS à Aix a développé pour ses études une méthode de transcription des corpus de français parlé contemporain (Blanche-Benveniste et Jeanjean, 1987). Les deux principales conventions de cette méthode ont été adoptées par certaines méthodes actuelles pour la transcription et l'édition de grand corpus (C. Blanche-Benveniste, 1997) : à savoir, une transcription orthographique et l'absence de ponctuation. Le souci d'établir un "document linguistique" a prévalu et avec lui, le refus d'apurer le texte de l'oral de ses soi-disant

11. Arnaud fait référence ici à la théorie modulaire de la production de la parole chez Levelt, 1989.

12. La mémoire à court terme ne dépasse pas trois minutes ; la mémoire à long terme qui correspond à un autre mécanisme physiologique s'étend au-delà de cette valeur.

“scories” même si la lecture de ces documents peut sembler, au premier abord incommode pour un lecteur non averti.

C'est l'accumulation des procédés caractéristiques des “avant-textes” de l'oral qui rend la lecture des transcriptions si incommode : retouches de toutes sortes, “hésitations”, “énoncés inachevés”, “ruptures de construction”, pauses fréquentes, etc. (Blanche-Benveniste et Jeanjean, 1987, p. 162)

Des indications prosodiques et des transcriptions phonologiques ne sont employées que dans certains passages (par exemple pour les énoncés ne pouvant recevoir une interprétation). Encore ne sont-elles pas portées dans le texte mais renvoyées en note (à l'exception des transcriptions d'enregistrement d'aphasiques à propos desquelles cette difficulté d'interprétation est si fréquente parfois que la lecture en serait trop incommode).

L'étude entreprise ici sur les lapsus se fonde sur deux types de recueils : l'un “à la volée”, l'autre est le résultat obtenu à partir d'un corpus enregistré et transcrit selon les conventions du GARS¹³ (corpus Bertuzzi 97). Dans les deux cas, les énoncés sont des propos tenus dans le cadre d'entretiens ou de conversations.

1. Recueil “à la volée” : les énoncés relevés n'ont pas été enregistrés. Ils ont été entendus au hasard de conversations ou d'entretiens écoutés à la radio¹⁴. Notés sur le champ, ces énoncés ne peuvent prétendre à la garantie d'une exactitude totale puisque la mémoire intervient forcément¹⁵. De plus non seulement, ce recueil n'est pas systématique (puisque'il est bien évident qu'il est impossible de noter tout lapsus entendu), mais il a évolué avec les réflexions issues de l'avancée du travail. En d'autres termes, si au départ cette activité de recueil semblait simple, l'analyse du corpus enregistré a révélé des phénomènes insoupçonnés au début. Cela ne manqua pas de modifier le recueil “à la volée” ; ces phénomènes passés inaperçus dans la vie quotidienne, devenaient perceptibles sinon faciles à relever. Si il y a une certitude méthodologique, c'est évidemment qu'un recueil “à la volée” de faits linguistiques ne peut être ni constant ni systématique.

13. Ce que Arnaud (1997) nomme recueils extensif (“à la volée”) et intensif (sur corpus enregistré).

14. Le relevé de lapsus est plus facile à faire lors d'écoutes d'émissions à la radio. Il n'est, en effet, pas facile de s'abstraire d'une conversation avec d'autres personnes.

15. Arnaud (1997, p. 308) qui établit une typologie des lapsus par substitution de mots souligne dans sa méthode de recueil le souci “d'éviter la rapide dégénérescence de la trace mémorielle”. On sait depuis Freud que les “infidélités” de la mémoire sont loin d'avoir pour seule origine “la dégénérescence de la trace mémorielle”.

L'étude présente est donc avant tout une tentative de construire son objet : et de répondre à la question : "Qu'est-ce qu'un lapsus dans le champ linguistique ?". Il reste que, comme on l'a vu, la très grande majorité des études sur le lapsus portent sur des recueils "à la volée".

Les quelques statistiques établies pour cette étude ont été faites à partir des 120 premiers exemples recueillis. D'autres exemples seront mentionnés dans le cours de cet article à titre d'illustration.

2. L'étude du corpus Bertuzzi (1997)¹⁶ s'est révélée essentielle pour cette recherche sur le lapsus ; ce n'est pas qu'elle en présente une quantité remarquable, mais la possibilité de lire et relire la transcription des paroles prononcées a permis d'éviter de construire une analyse sur des faits déjà définis ("un lapsus c'est ce que mon analyse étudie"). La lecture de l'énoncé permet de constater des phénomènes qui sont bien sûr perceptibles pour une écoute attentive et avertie mais souvent impossibles à noter. L'étude sur corpus permet surtout un relevé systématique de faits qui autrement paraissent rares (parce que discrets et rapides). C'est pourquoi, l'analyse de ce corpus a relancé en le modifiant le recueil "à la volée".

Il se trouve que ce corpus comporte 100 exemples de lapsus dont la diversité même m'a incitée à prolonger le recueil "à la volée" et à comparer ces phénomènes selon leurs conditions d'obtention.

Informations recueillies pour chacune des erreurs de langage

Cinq types d'informations ont été relevés pour chaque exemple recueilli.

1. L'exemple lui-même.
2. Le locuteur a corrigé ou non son lapsus (C/ non C).
3. Lapsus se traduisant par :
 - E : un mot entier orthographiable,
 - A : amorce d'un mot,
 - N : néologisme¹⁷,
 - O : omission d'un élément,
 - R : répétition d'un élément.

16. Le corpus Bertuzzi (1997) est la transcription intégrale de trois enregistrements d'une personne évoquant les conditions d'exercice de son métier de fleuriste à Paris. Cette étude a été faite dans le cadre d'une soutenance de maîtrise au Département de Linguistique française à l'Université d'Aix-Marseille I.

17. On trouve dans cette catégorie tout énoncé qui ne correspond pas à un mot du dictionnaire. Certains ne permettent pas une reconnaissance de morphèmes, d'autres sont des sortes de mots-valises ou des formules-valises.

Catégorie morpho-syntaxique de cet élément : nom, pronom, verbe ;
sujet, complément,

4. Renseignements sur le locuteur ; date du lapsus.

5. Informations contextuelles (linguistiques ou non) ; remarques

Exemple : 142¹⁸

si j'ai tant attendu pour *attendre* ce livre pour *écrire* ce livre

Corrigé :

E ; verbe Séquence Objet (lexique)

Écrivain ; 2 avril 1998

Contexte : France-Culture, entretien

3. ÉLÉMENTS D'ANALYSE

3.1. QUELQUES REMARQUES PRÉLIMINAIRES

3.1.1. LAPSUS, ÉCHEC AU REGARD D'UNE NORME ?

Les lapsus sont souvent assimilés à des erreurs que ce soit par rapport à une norme, un standard ou que ce soit par rapport à des données probables. Mais dans quel domaine linguistique (lexical, morphologique, syntaxique, intonatif ou discursif) peut-on repérer l'erreur ? L'étude de Colette Jeanjean (*Les ratés c'est fa- fabuleux*, 1984) sur les amorces¹⁹ et les répétitions dans le français parlé montre l'intérêt de distinguer les interprétations syntaxique et discursive. Si les amorces peuvent être analysées comme des erreurs discursives, sur le plan syntaxique, "il n'y a pas lieu de parler de ratés" (p. 174).

Il reste que l'attribution du qualificatif d'erreur à propos du lapsus sur le plan discursif est insistante. Cette erreur se traduit par la présence d'un élément surprenant ; il reste à préciser dans quel champ se situe cette surprise : sémantique, morpho-syntaxique, phonologique, etc.) :

- mais il te faut des *passages* [rires] des *massages* seulement du haut du corps (corpus Pallaud 1997, 126)

Dans cet exemple, le rapport au contexte signale l'insolite du terme *passage*. On remarque que le caractère surprenant du terme n'en serait pas plus accentué si on l'avait interprété comme *pas sages*, *passe-âge*, etc..

18. En caractères italiques, le lapsus et sa correction

19. Le terme de bribes recouvre deux phénomènes. Généralement l'équipe du GARS utilise le terme d'*amorces* pour désigner les bribes concernant les mots (C. Jeanjean, 1984). D'autres bribes correspondent en rhétorique à une anacoluthie (C. Loufrani, 1981).

Dans cet exemple (comme dans la plupart des cas où le lapsus porte sur un énoncé pluri-syllabique), le transcripateur peut se demander où arrêter son imagination interprétative puisque justement le lapsus remet en cause la cohérence textuelle²⁰ ? Ce qui est plausible, attendu en fonction du contexte, et ce qu'on peut savoir sur le locuteur, tout cela sert de guide au transcripateur. Dès cette étape du recueil des données, les hypothèses que l'on a sur la formation des lapsus vont influencer leur transcription.

Le locuteur qui parle est le plus souvent "conscient" de ses lapsus (ou tout au moins exprime cette reconnaissance²¹), les corrige, s'en amuse ou les regrette. Qu'il s'agisse du recueil "à la volée" ou du corpus enregistré, les résultats sont massifs : les lapsus corrigés sont beaucoup plus nombreux que ceux qui ne le sont pas :

- recueil "à la volée" : 72,5 % (87/120)
- corpus enregistré : 83 % (83/100)

Cependant, ces auto-corrrections n'ont pas toujours lieu comme le montrent les exemples suivants (corpus Pallaud, 1997, 125, 124) :

- L1 tu as entendu le Sud du Portugal est relié par le plus grand port d'Europe
- L2 tu veux parler d'un *pont* je suppose
- L1 oui bien sûr
- ma mère a dit surtout cache ce livre que Poupette ne le lise pas eh bien ce livre je l'ai [*li*] caché dans le lit

L'attribution du terme "erreur" cherche sa justification dans les auto-corrrections et marques de repentirs ou d'excuses diverses formulées par le locuteur. Il est à noter que souvent seules ces remarques explicites de correction signalent les hésitations du locuteur. En leur absence, ces énoncés ne seraient pas signalés comme comportant une irrégularité :

- et tu diras toutes ces *actions* oh flûte encore un lapsus + ces *occasions* que j'ai manquées c'est pas croyable
- j'ai vu euh j'ai *rencontré* une classe d'enfants de Bogota venus skier dans le Jura
- je vais aller à la *fac* euh à la *pharmacie*
- c'est un *film* c'est un *décor* excusez-moi que John Ford s'est approprié

20. On sait que le point de vue psychanalytique est différent puisque l'interprétation du lapsus (sa transcription donc) s'appuie sur les associations du sujet et se trouve limitée par elles.

21. Outre l'éclat de rires, un certain nombre d'énoncés signalent la reconnaissance par le locuteur tels que : *excusez-moi, pardon, encore un lapsus*, etc.

- il y a chez Gisèle Praxinos un sublime *amour* [rires] un sublime *humour* (corpus Pallaud, 1997, 70, 82, 139, 153, 136)

Le locuteur qui écoute ou qui transcrit est donc parfois le seul à percevoir ces "bizarreries". Il est donc clair que ses savoirs interviennent dans la détection de ces anomalies ; ce qui revient à dire que la constitution même d'un recueil de données dans ce domaine peut être biaisée (nous y reviendrons).

2.2. LAPSUS ET INTENTION

La deuxième notion à laquelle la plupart des études font appel pour caractériser le lapsus est l'intention ou plutôt, en l'occurrence, l'absence d'intention du locuteur : *le lapsus est une faute que l'on commet par inadvertance*. Sur le plan du discours quelque chose était attendu et/ou était dans l'intention du locuteur, et a été dit autrement. La perte de contrôle du dire est, on l'a vu, le plus souvent perçue par le locuteur lui-même et exprimée de diverses façons (pauses, marques d'hésitation, excuses, rires, etc.). Cet aspect du lapsus est souvent invoqué pour attester le lapsus :

Le lapsus linguae doit être considéré comme une déviation de l'intention du locuteur ayant pour résultat une modification intentionnelle de la forme, c'est-à-dire une altération d'une unité du niveau symbolique et non du niveau de la substance. (Rossi et Peter-Defage, 1998, p. 18).

Blanche-Benveniste (1990) propose de ne pas mettre à part les manifestations intentionnelles ; elles restent incluses parmi ces phénomènes de bredouillage qui jalonnent les énoncés parlés :

Sur l'axe paradigmatique²² se manifestent divers phénomènes : nous avons évoqué celui du bredouillage et celui de l'énumération. Nous proposons d'en évoquer d'autres. Au risque de choquer des habitudes bien ancrées, nous ne distinguerons pas entre les phénomènes qui paraissent créés par la volonté des locuteurs et ceux qui paraissent leur échapper ; nous traiterons de la même façon des phénomènes apparem-

22. La conception du paradigmatique qui est développée ici est proche de celle à laquelle se réfère Jakobson (1963) et se distingue de celle qu'a énoncée de Saussure (1922) : "Le rapport syntagmatique est in praesentia; il repose sur deux ou plusieurs termes également présents dans une série effective. Au contraire le rapport associatif unit des termes in absentia dans une série mnémonique virtuelle" (1985, p. 171). C'est pourquoi "le locuteur, restant sur cette place syntaxique du complément temporel, par hésitation, par mégarde, ou par souci d'expressivité, peu importe, a répété le même élément; nous dirons qu'il a utilisé l'axe des paradigmes" (Blanche-Benveniste, 1990, p. 19).

ment involontaires comme bredouillages hésitations, maladresses, reprises, et d'autres qui semblent intentionnels comme : répétitions intensives, variations stylistiques et autres. On verra que ce point de vue, s'il néglige la part d'intention du locuteur, sur laquelle nous ne pensons pas pouvoir porter de jugement, a l'avantage de suivre une ligne d'analyse grammaticale unifiée. (p. 20)

Le corollaire de cette intention prise en défaut chez le locuteur est le fréquent sentiment de honte qu'il exprime lorsqu'il s'aperçoit de son lapsus ou lorsqu'on le lui fait remarquer. Cette honte voire même la difficulté à admettre le lapsus ont été signalées déjà par Meringer et Mayer (1895). Freud a confirmé cette observation :

Les signes d'émotion qu'on suscite en prouvant à quelqu'un qu'il a commis un lapsus, et qui sont manifestement très voisins de la honte, ces signes sont significatifs. (p. 92)

Goffman (1981), dont l'approche est plus sociolinguistique, revient à plusieurs reprises sur ce sentiment de honte qui accompagne la perception par le locuteur de ses lapsus ; selon lui, ce sentiment est repérable dans les marques d'énonciation qui accompagnent la correction même de l'erreur (*oh flûte encore un lapsus, c'est pas vrai, excusez-moi, pardon, c'est la fatigue, c'est l'émotion, etc.*). Par la suite, si les études sur les lapsus ont souvent recours à l'absence d'intention pour spécifier le lapsus (ce qui évidemment est souvent problématique), elles ne mentionnent pas cette répugnance fréquente chez le locuteur à admettre ses anomalies de langage.

3. DIFFICULTÉS D'IDENTIFICATIONS DES LAPSUS

3.1. LAPSUS : "UN MOT POUR UN AUTRE", MAIS ENCORE FAUT-IL SAVOIR QUEL ÉLÉMENT EST INSOLITE

Dans cette catégorie, se trouvent réunies les productions qui peuvent recevoir une interprétation parce que la reconnaissance de morphèmes a été possible ; elles sont donc orthographiables. Cependant, on constate que, soit le locuteur lui-même récusé ce qu'il vient de dire, soit que le transcritteur (ou l'auditeur) trouve quelque chose d'insolite dans l'énoncé qui vient d'être dit²³. Cette catégorie d'exemples correspond à ce qu'on

23 On ne discutera pas ici de la justesse de l'interprétation. Disons que, dans ce champ linguistique, la justesse s'appréhende par le biais de la cohérence, de la vraisemblance et du probable qui sont les axes délimitant le champ de l'objectivité en matière de sémantique. On sait bien que ces prémisses peuvent être prises en défaut et que l'incohérent,

entend couramment par lapsus à savoir : “un mot pour un autre”. Ce type de lapsus est conçu le plus souvent comme une erreur sur le lexique nominal ou verbal et se trouve inclus dans ce que certaines études abordent comme erreurs de dénomination (Arnaud, 1997) :

(corpus Pallaud, 1997, 71, 79, 93 ; 15, 56)

- tu as les vœux de *Noëlle* euh de *Danielle*
- on vous a même fait remplacer un *cric* un *pneu* dans cette épreuve
- oh mon écharpe + je l'ai prêtée à Gisèle parce qu'elle avait froid et je l'ai gardée euh *elle* l'a gardée
- L1 on verra si Catherine Trautman va suivre les recommandations du parti socialiste et démissionnera de la mairie de *Bordeaux*
- L2 elle n'a pas encore conquis la mairie de Bordeaux que je sache
- L1 ben à la Libération c'est pour ça que j'ai fait trois *ans* de prison
- L2 trois ans ou trois *jours* ?
- L1 oui trois jours bien sûr

Les *trois quart* (91 sur 120) du total des lapsus relevés “à la volée” sont de l'ordre de “un mot pour un autre”. Sont compris dans ce décompte les sept énoncés (soit 7 sur 91 : 7,6 %) où on ne sait si on se trouve en présence d'un lapsus ou d'une omission et 4 lapsus de type “carambolage”²⁴.

Dans le corpus enregistré, ces lapsus de type “un mot pour un autre” sont au nombre de 33 ; ils constituent le *tiers* seulement des phénomènes relevés (33 sur 100). La proportion de ces lapsus pouvant être interprétés aussi comme des omissions est bien supérieure à celle qui a été trouvée dans le recueil “à la volée” (27 sur 33 : 81, 8%). Les proportions par rapport au recueil “à la volée” sont, donc, inversées.

Ces chiffres prouvent, s'il en était besoin, le biais qui peut être introduit par un recueil *à la volée* et le grand avantage qu'il peut y avoir à bénéficier de corpus enregistrés. Dans les recueils “à la volée”, les lapsus

l'in vraisemblable, l'improbable se produisent quand un Sujet parle. Si les effets sont négligeables pour une discipline à la recherche de faits statistiquement repérables, il n'en n'est pas de même pour la psychanalyse dont la “sémantique” ne vise qu'un sujet à la fois.

24. Les carambolages (*bumper-car errors*) signalés par Stemberger 1985 (cité par Arnaud, 1997) sont les cas d'interversion de termes ; les exemples présents dans le corpus Pallaud, 1997, 26, 35, 38, 92 :

- à Alexandrie il y a des avenues et des rues ce qui fait que pour moi *Alexandrie c'est New York* ou plutôt *New York c'est Alexandrie* mon lapsus n'est pas pour rien
- parler de la mort ça fait pas *les gens mourir*
- mais je le veux *moi pour tout* ah ah *tout pour moi*
- allô c'est *Babette Maman* euh [rires] c'est *Maman Babette*

du type “un mot pour un autre” sont très surestimés alors que toutes les hésitations se traduisant par des amorces (amorces corrigées et inachevées) se trouvent sous estimées.

Il est cependant des cas où il n'est pas possible d'affirmer que le lapsus porte sur tel élément plutôt que sur tel autre. Dans les exemples qui vont être donnés, on peut en effet hésiter sur la localisation du lapsus. Suivant les cas, la portée du lapsus n'est pas la même. Dans le recueil “à la volée”, on n'en trouve que cinq exemples ; il n'y en a aucun dans le corpus enregistré.

Dans l'énoncé suivant, sans interroger le locuteur il est impossible de dire si le lapsus porte sur *monsieur* corrigé plus tard en *madame* :

- restent *monsieur* Trautman euh monsieur Hue et *madame* Trautman

ou si l'erreur porte sur les noms propres (*Trautman* au lieu de *Hue*).

Le comble est atteint dans l'exemple suivant où le locuteur a du mal à sortir de l'ambiguïté de ce qu'il dit et où le lapsus peut porter sur plusieurs termes²⁵ (corpus Pallaud, 1997, 85) :

au XIX^e siècle il y a une *baisse* de l'intolérance un *accroissement* de la tolérance euh à force de croiser les chiffres je ne sais plus où j'en suis c'est un accroissement de l'intolérance auquel on assiste au XIX^e siècle

OU

au XIX^esiècle il y a une baisse de l'intolérance un accroissement de la *tolérance* euh à force de croiser les chiffres je ne sais plus où j'en suis c'est un accroissement de *l'intolérance* auquel on assiste au XIX^esiècle

OU

au XIX^esiècle il y a une baisse de *l'intolérance* un accroissement de la *tolérance* euh à force de croiser les chiffres je ne sais plus où j'en suis c'est un accroissement de l'intolérance auquel on assiste au XIX^esiècle

OU

au XIX^esiècle il y a une *baisse de l'intolérance* un accroissement de la tolérance euh à force de croiser les chiffres je ne sais plus où j'en suis c'est un *accroissement de l'intolérance* auquel on assiste au XIX^esiècle

On y reviendra plus loin, mais on peut déjà souligner que cette difficulté à localiser la portée du lapsus dans l'énoncé a pour conséquence de révéler l'équivocité de cet énoncé.

25. En caractères italiques, le lapsus et sa correction.

3.2. LAPSUS OU VARIATION ÉLABORÉE SUR L'AXE PARADIGMATIQUE ?

Un certain nombre de lapsus ont toutes les apparences d'un phénomène de liste c'est-à-dire une énumération sur l'axe paradigmatique et que la méthode de mise en grille élaborée par l'équipe du GARS présente ainsi (Blanche-Benveniste et Jeanjean, 1987, p. 170)²⁶ :

moi j'avais ma mère
mon père
ma grand-mère
ma marraine
quatre gendarmes sur le dos (corpus Baral, 29, 5)

Contrairement au phénomène de liste précédent, sans le recours à l'intention du locuteur, c'est-à-dire sans interroger le champ discursif, il est impossible d'attribuer le statut de lapsus à l'un ou l'autre des énoncés suivants, c'est-à-dire de les distinguer d'une énumération :

- c'est aussi cette manière *horizontale + orientale* plutôt qu'il a d'écrire ses textes
(Contexte : évocation d'un article concernant les écrits sur rouleaux en Chine ancienne)
- le but de cette réunion était de montrer que la justice a un rôle à jouer dans le *déclin* dans le *traitement* de la délinquance
(corpus Pallaud, 1997, 1, 66)

Dans certains cas, le locuteur envisage lui-même son lapsus comme une variante possible, quoique non intentionnelle au premier abord, de son discours : le lapsus est reconsidéré en une énumération.

- Giono disait les *misères* euh les *mystères* ne gagnent pas oui on peut dire aussi les *misères* ne gagnent pas à être montrées
- les Russes ont fait des plans *hysté-* oui des plans hystériques et *historiques* (corpus Pallaud, 1997, 19, 135)

Parfois, les circonstances de l'énoncé, associées à des phénomènes prosodiques discrets, permettent également de distinguer le lapsus là où on pourrait ne voir qu'une énumération, une variante discursive :

- C'est peut-être qu'on veut réparer quelque chose réparer ce qu'on a fait en *accouchant* soi-même + en *naissant* soi-même.
(Le locuteur est un médecin accoucheur; il tente de décrire les raisons de son choix professionnel) (corpus Pallaud, 1997, 12)

26. La présentation de la technique de mise en grilles se trouve notamment dans Blanche-Benveniste et Jeanjean (1987); Blanche-Benveniste, Bilger, Rouget et van Den Eynde (1990).

Ces commentaires, on le voit, sont indispensables ; ils sont notés le plus systématiquement possible dans les transcriptions de corpus faites au GARS ; là encore les recueils “à la volée” à cause de la l'étendue de la période au cours de laquelle cette activité de collecte se déroule sont plus sujets à des variations dans la notation des commentaires que les recueils constitués à partir de corpus.

Parfois, la transcription ne peut s'appuyer que sur des éléments prosodiques pour attester le lapsus et le distinguer de la variation énumérative ; c'est ce qui se passe dans les exemples suivants (une brusque variation d'intensité atteste que l'erreur porte sur l'énoncé en gras) :

- il faut changer l'âge de la *majorité non* de la *responsabilité pénale*
- est-ce que nous sommes parfaits ? bien sûr que *nous*²⁷ nous ne sommes pas parfaits (corpus Pallaud, 1997, 32, 143)

Comme il n'est pas toujours possible de connaître l'intention du locuteur, on ne peut pas toujours garantir la distinction entre lapsus et phénomènes de listes. Que les lapsus aient été recueillis “à la volée” ou tirés de corpus enregistrés, la difficulté est la même. Comme la plupart des lapsus relevés “à la volée” sont émis au cours d'émissions de radio, il est impossible d'interroger les locuteurs et de vérifier ce qui paraît essentiel au lapsus : l'intention que peut avoir le locuteur de dire autre chose.

En revanche, il est évident que lorsqu'on dispose de corpus enregistrés, on bénéficie d'un contexte large permettant plus de précisions sur ces phénomènes. Certains exemples, on l'a vu, montrent qu'il est parfois indispensable de recourir à des caractéristiques prosodiques pour attester le lapsus. Les conventions du GARS ne prévoient pas pour l'instant ce genre de notation mais cet inconvénient cesse d'être une difficulté à partir du moment où on peut consulter parallèlement l'enregistrement sonore. Or, les nouvelles technologies dans le domaine informatique permettent de proposer simultanément au linguiste la saisie numérique de l'enregistrement sonore et celle de sa transcription.

2.3. LAPSUS, RUPTURE DE CONSTRUCTION OU BRIBE ?

Un certain nombre de lapsus se présentent sous la forme de bribes corrigées c'est à dire d'amorces de construction avec “correction à gauche”. Par exemple dans l'exemple qui suit, une construction prépositionnelle est supplantée par une construction adverbiale. La réalisation du

27. La variation d'intensité et d'intonation conduisait à attendre : *bien sûr que non*.

complément temporel s'y prend à deux fois et cela de façon parfaitement adéquate. Simplement la première est vite abandonnée.

- alors ça revenait au même + parce que le matin je me levais *de plus tôt* pour prendre mon train (corpus Bertuzzi, 50, 7)

Les exemples qui suivent relèvent tous du même type : des lapsus sous forme de bribe :

- elle *avait* elle *connaissait* bien bien son travail (corpus Bertuzzi, 49, 10)
- euh moi *j'ai* pas *j'avais* pas de permis de conduire (corpus Bertuzzi, 41, 5)
- on les dépotait on enroulait *la + le* la terre²⁸ (corpus Bertuzzi, 35, 5)
- ah oui c'était une bonne place fleuriste *devant au* Père Lachaise (corpus Bertuzzi, 33, 11)
- Henry James + toute sa vie est *un une* oscillation entre l'Europe et les États-Unis (corpus Pallaud, 1997, 106)

Dans le corpus enregistré, on note la présence fréquente de bribes corrigées (27 % ; 27/100) alors que dans le recueil "à la volée" ces cas sont rares (5 % ; 7/120). Ces phénomènes se distinguent des bribes simples où les constructions ne sont interrompues que passagèrement²⁹, sans "correction à gauche" et se poursuivent par fragments successifs comme dans les exemples suivants (la présentation en grille adoptée par le GARS rend compte de cette interprétation) :

le	
le	
le	
l'acte fondateur	
je	
je	
je lui reproche	
après	quand on
	quand on regarde
je crois que	
je crois que beaucoup de gens savaient	

28. Le lapsus porte sur l'article *la* puis *le*.

29. La distinction entre les phénomènes de bribes corrigées ou non n'est pas toujours faite (Cappeau, 1998). Elles sont interprétées comme des piétinements à l'endroit de certaines places syntaxiques comme le déterminant (Blanche-Benveniste et Jeanjean, 1987; Cappeau, 1998) mais aussi le pronom personnel, le verbe, les prépositions, etc.

avec de
de
de
de gentils moments (corpus Pallaud 1997)

Dans le cas suivant, on constate que le locuteur qui entend l'énoncé hésite à considérer qu'il s'agit d'une rupture de construction (ce qui semble bien être le cas) et préfère répéter ce qu'il vient d'entendre³⁰. Ce faisant, il souligne l'insolite de cet énoncé, son caractère de lapsus. Si l'on opte pour un lapsus sur le temps du verbe (*suivrai* au lieu de *suive*), on écrira l'exemple :

- L1 oui on ira récupérer la voiture il faudra que je te suivrai jusque chez toi
L2 il faudra que je te suivrai ?

ou si l'on opte pour une omission :

- L1 oui on ira récupérer la voiture il faudra que / je te suivrai jusque chez toi
L2 il faudra que je te suivrai ? (corpus Pallaud, 1997, 107)

Dans l'exemple suivant, l'hypothèse d'une bribe corrigée s'impose d'emblée. On admet volontiers le lapsus sur le temps du verbe car cette hésitation n'enfreint pas les règles du système de la langue :

- il n'a pas d'effet direct qui *est* qui *serait* de nature à transformer la jurisprudence en situation non exceptionnelle (corpus Pallaud, 1997, 140)

Ce sont les cas des bribes et des interruptions de construction "découvertes" au cours de l'analyse de corpus enregistrés qui ont fortement modifié le relevé "à la volée" des lapsus. Elles ont également fortement contribué à amorcer la réflexion présente sur l'identification des lapsus.

2.4. LAPSUS : NÉOLOGISMES³¹

Certains passages des énoncés soulèvent des difficultés d'interprétation soit qu'ils permettent une reconnaissance approximative de mots, soit qu'ils n'autorisent qu'une reconnaissance de phonèmes. Ces difficultés conduisent à distinguer ce qui est du ressort du néologisme, ce qui suppose qu'une interprétation partielle et approximative est possible. Autre chose est le cas où une signification étant impossible, on en est réduit à rendre

30. Les deux locuteurs sont des enseignants à l'Université.

31. Goffman (1981) emploie le terme de "métathèse" pour désigner ces phénomènes et signale que, dans les malentendus (Garnes et Bond, 1975) aussi bien que dans les lapsus, les métathèses et interversions abondent.

compte de ce passage par une transcription phonétique. Cette deuxième solution est celle qui est également adoptée dans le cas de jargon.

Le néologisme est un “mot nouveau, le plus souvent créé par dérivation ou par composition, et qui, quand il n'est pas sérieux et destiné à s'intégrer au lexique de la langue, cherche à être frappant et humoristique voire burlesque” (Bacry, 1992). Tout le monde a pu lire sur les livres édités récemment la formule suivante où la *photocopie* se trouve estampillée de façon saisissante par l'accusation de *pillage*) :

Danger : le *photocopillage* tue le livre

ou bien (corpus Pallaud, 1997, 170)

L1 il faut pas que j'oublie le rapport que Guenz m'a donné hier

L2 ah oui que tu *synthèses*

L1 oui il faut que je *fasse une synthèse*

Un cas particulier du néologisme est le “mot-valise” qui est un terme nouveau (véritable création lexicale, absente du dictionnaire) obtenu à partir de deux mots ayant en commun une syllabe (ou quelques lettres, voire une seule lettre)³² : les *célibattantes* (*célibataires*, *battantes*) sont célèbres mais aussi le *bistroquet* (*bistrot*, *troquet*), la *foultitude* (*foule*, *multitude*) que bien des personnes utilisent comme s'il s'agissait d'un lexique standard.

Dans le cadre des études sur les lapsus, bon nombre de néologismes sont traduits orthographiquement, donc sous forme de mot-valises (Peter-Defage, 1993, Rossi et Peter-Defage, 1998, Arnaud, 1997 et Freud également [1901, p. 76]) :

– le meilleur *merrier* guerrier (le meilleur *guerrier*) (Defage, 1993, n° 70)

– *begleitdigen* (*begleiten*, *beleidigen*) (Freud, 1901, p. 77)

Pour certains linguistes, ces néologismes sont la preuve même de perturbations de l'émission de la parole, par anticipation ou persévération de phonèmes ou syllabes présents dans le contexte. La transcription de ces néologismes sous une forme orthographiée qui permet d'écrire ces morphèmes que le transcripateur identifie est destinée à rendre compte de ces mécanismes d'anticipation ou de persévération.

Le jargon est une séquence orale à laquelle il est impossible d'attribuer une signification et dans laquelle on ne peut identifier un morphème. Dans le cadre d'une transcription, on a alors recours à l'alphabet phonétique pour transcrire le passage concerné. On trouve souvent du jargon

32. Pour une étude détaillée du mot-valise cf. Grésillon (1984).

dans les transcriptions de personnes aphasiques (1981) ou d'enfants inintelligibles en cours de rééducation (Pasquale, 1989) et rarement dans d'autres cas. C'est pourquoi, cette terminologie est fortement connotée et peu employée en dehors du discours médical. Il reste que parfois dans des énoncés de personnes ne présentant pas de troubles aphasiques, on ne peut transcrire certaines de leurs paroles qu'à l'aide de l'alphabet phonétique. Par exemple, on n'éprouve pas de difficulté à transcrire l'énoncé suivant :

si j'avais pensé ça ça aurait pu être une marque de [*juvénisme*] de *juvénilité* (corpus Pallaud, 1997, 204)

En revanche, pour les exemples qui suivent, une transcription phonétique semble préférable (corpus Pallaud, 1997) :

- ma mère a dit surtout cache ce livre que Poupette ne le lise pas eh bien ce livre je l'ai [*li*] caché dans le lit (124)
- mais dans ce [*zyt*] euh ce *gîte* il n'y a pas d'ascenseur (163)
- je te fais des [*sol*] ou plutôt des *soles* [rires]³³ (contexte culinaire bien sûr) (102)

Nous avons rassemblé dans cette rubrique les lapsus qui n'ont pas pu recevoir une signification "standard"; ils constituent la deuxième grande catégorie des lapsus relevés "à la volée" (18 sur 120; 10 d'entre eux font l'objet d'une auto-correction).

- Pour certains de ces lapsus des solutions orthographiques "approximatives" (des néologismes donc) apparaissent d'emblée (corpus Pallaud, 1997) :

- j'ai l'impression que je vais *apprir* que je vais *apprendre* à vivre avec toi (28)
- je lui ai dit de ne pas faire la poussière sur les plans de travail et je crois que ça la *vâche* [rires] ça la *vexe* (*fâche*, *vexe*) (112)
- le test de *l'antinanisme* de *l'antinazisme* reste à faire (151)

Parmi ceux-ci, on trouve des cas où plusieurs solutions orthographiques sont envisageables :

- mais c'était quoi un (*matelot*, *mateleau*) à eau [rires] un *matelas* à eau (129)

La première solution conduit à ranger ce lapsus dans les cas de substitution, la deuxième le place dans les cas de mot-valise "matel(as) à eau" avec élision du phonème [a].

33. Pour le locuteur de cet énoncé, l'opposition [o], [ɔ], est pertinente.

- D'autres lapsus ne peuvent être transcrits que phonétiquement :

- j'ai fait une quantité de saignées au [byron] dans les murs de notre appartement (au burin) (23)
- les valises en carton sont les [vazil] de l'exil [rires] les valises de l'émigration (51)
- regarde-le avec son air [isyrite] (hirsute) (69)

Dans le corpus enregistré, ce type de lapsus est moins fréquent (8 sur 100), corrigé également une fois sur deux ; ils soulèvent les mêmes difficultés de choix de transcription :

- c'est une petite domination quelque part quoi du fait de parler + que les [parlân] + que les parents parlent le français pour les enfants c'est comme si ++ ils dominaient (corpus POI 97 - 97)
- alors tous les ans à une [sertif] à une certaine époque ils fêtaient le mur des Fédérés (corpus Bertuzzi, 13, 13)
- mais il fallait toujours soulever parce qu'en dessous c'était toujours beau ça [vwajet] déjà pas ça voyageait pas bien des fois (corpus Bertuzzi, 22, 3)

D'autres énoncés se présentent sous une forme que je propose d'appeler des "formule-valises", pour paraphraser la notion rhétorique de mot-valise qui caractérise de façon évidente, comme on l'a vu un certain nombre de néologismes. C'est le cas des exemples suivants (corpus Pallaud, 1997, 39, 40, 41) qui paraissent condenser en une seule formule deux énoncés tronqués correspondant à des locutions figées. Cet artifice, qui n'en est pas un puisqu'il s'agit de lapsus, révèle une ambiguïté qui se "déploie" à demi dans une concaténation de formules tronquées³⁴.

- cet auteur a une certaine conception sacrée à l'écriture
soit
cet auteur a une certaine (conception sacrée de, relation sacrée à) l'écriture
- j'ai mis du mal j'ai eu du mal mais j'y suis arrivé
soit
j'ai (mis du temps, eu du mal) mais j'y suis arrivé
- il faut *poursuivre* cette analyse à euh *poursuivre* cette analyse à d'autres domaines

34. Cette équivocité repérable dans ces formules-valises comme dans les néologismes suggère au psychanalyste qu'elle constitue le ressort du lapsus "réussi" dans une substitution de mot. Cette substitution fait écran à un autre énoncé que, on l'a vu, le locuteur évoque le plus souvent lui-même (à travers ses auto-corrections).

soit

il faut (poursuivre dans, étendre à) cette analyse à d'autres domaines

Dans le corpus enregistré, on ne relève que deux "formule-valises" du type :

- c'était pas joli joli + alors pour *s'en liquider* quand même on a fait des petites bottes de muguet (corpus Bertuzzi, 16,3)

soit la formule groupant *s'en débarrasser* et *les liquider* ; durant la nuit les bottes de muguet ont flétri trop vite, alors pour ne pas perdre toute la vente...

On remarquera que, dans ces exemples, l'affirmation de la présence d'un lapsus repose sur le postulat qu'un message ne doit pas être équivoque, que les impératifs de la communication imposent un message univoque : on ne peut dire une chose et son contraire. Pour la linguistique cognitive, ce postulat de cohérence du discours est le fondement même de l'objectivité de son approche (Kleiber, 1987). Les psychanalystes et certains linguistes soulignent au contraire l'équivocité foncière de la parole singulière et de la langue :

ce qui représente aux yeux de ces amateurs de langages formels, un défaut regrettable des langues qu'ils appellent "naturelles" se révèle n'être probablement pas un accident mais une caractéristique essentielle, condition même de leur évolution et de leur survie. (Fuchs 1996, p. 80)³⁵

Cette catégorie de lapsus suggère donc fortement que le lapsus n'est pas une substitution de mots, le glissement d'un terme à la place d'un autre mais bien plutôt un phénomène de condensation (dont la forme parfaite serait un mot faisant totalement écran à un autre), une fabrication de "chimères" linguistiques à l'instar du mot-valise ; l'effet de sens n'est pas un choix erroné entre deux significations mais plus vraisemblablement la somme de ces significations plurielles ("la surcharge incontrôlée de sens" comme le dit Fuchs).

Arnaud, à l'instar de Rossi et Peter-Defare, insiste sur la distinction entre les lapsus syntagmatiques (se traduisant le plus souvent par des néo-

35. Selon Fuchs, il y aurait néanmoins diverses sources d'ambiguïté : "Le risque de prolifération des ambiguïtés est à l'évidence encore plus grand dans le cas des usages quotidiens de la langue. Sans aller jusqu'à évoquer le travail de l'inconscient ou diverses pathologies qui semblent rendre les sujets incapables d'échapper à l'ambiguïté et à la surcharge incontrôlée du sens, on peut légitimement se poser la question de savoir si l'idée même d'une production linguistique totalement *univoque* au sens précis de 'dépourvue d'ambiguïté', est réaliste" (p. 80).

logismes ou des “formules-valises”) et les lapsus paradigmatiques. Il souligne l'importance des premiers dans le cadre d'une linguistique où une relation entre le *figement des locutions* et la production de la parole se trouve mise en avant :

Les lapsus syntagmatiques nous rappellent que la production d'énoncés ne se réduit pas à une série de dénominations de concepts, mais aussi à l'assemblage de spécifications grammaticales et au montage d'une chaîne. Il semble qu'un élément donné envoie de l'activation à ses collocants, même ceux qui ne font pas partie du message. C'est là un indice de la justesse d'un courant de pensée minoritaire en linguistique, et qui insiste sur le fait que la production repose en partie sur du préfabriqué (cf. Nattinger & De Carrico, 1992).

On a vu cependant, que certains des lapsus pouvaient apparaître comme des mots-valises si la transcription rendait compte de leur formation ou comme de simples substitutions de mots si l'interprétation le décidait. On ne peut arguer qu'il est fort improbable que *jute* ait été prononcé à la place de *gîte* ou même *vache* au lieu de *vexe* puisque le lapsus (erreur par inadvertance) oblige à mettre, par définition, la cohérence du texte entre parenthèses.

2.5. LAPSUS ET AMORCES

Le choix d'élargir l'étude sur les lapsus aux phénomènes d'amorces s'est trouvé renforcé par l'expérience de transcription de corpus enregistrés ; les phénomènes d'amorces soulèvent en effet des difficultés de reconnaissance des morphèmes qui ne sont qu'amorcés. Parfois, certaines amorces sont ininterprétables :

- il y avait plein de tombes des artistes + du cinéma des [ε] aviateurs tués à la guerre (Bertuzzi, 10, 9)

Certaines, au contraire, ne sont qu'un des choix possibles dans une multi-transcription³⁶

36. Le GARS (Blanche-Benveniste et Jeanjean, 1987) a proposé un procédé typographique simple pour rendre compte d'interprétations plurielles pour un même passage d'énoncé (ce qui n'est pas si rare même en tenant compte du contexte et en postulant l'absence de termes insolites ; autrement dit en prenant le parti pris d'une forte cohérence textuelle). Ce procédé consiste à encadrer ces multi-écoutes par deux barres obliques (exemples tirés de Bilger et al., 1997) :

- le décollage s'est très bien passé /*parce que, puisque*/ tu décolles pratiquement sur place (Anguenot 13,12)

- oui //œu-, le/ l'œuf se mélange entièrement (corpus Jacques, 35, 19)

D'autres laissent entrevoir qu'elles auraient pu ne pas être identifiées, être transcrites de façon phonétique et appartenir alors à la catégorie des néologismes :

- mais il y a eu des *palanq-* des *générations* de musiciens de jazz qui ont été sous son influence (corpus Pallaud, 1997, 30)
(on suppose que le terme amorcé est bien *palanquée*)

On trouve trois sortes d'amorces dont la distinction repose sur l'analyse de la place syntaxique occupée par ces amorces. Il est à noter que, si très souvent on est tenté de les interpréter (ce qui explique la grande fréquence des transcriptions orthographiques), on ne peut se targuer de certitudes à ce sujet. En revanche, la place syntaxique est approchée avec plus de certitude.

Oa-φ	amorce laissée inachevée
Oa- b	amorce corrigée
Oa- aa	amorce complétée

a. Amorces type Oa-φ ou amorces laissées inachevées (ce qui suit l'amorce occupe une autre place syntaxique; on n'est donc pas en présence d'un piétinement sur la même place syntaxique). Ces amorces sont des lapsus interrompus et non corrigés.

Dans le corpus "à la volée", on ne relève aucune amorce de ce type.

Dans les corpus enregistrés, ces amorces sont au nombre de 10

- quand on rentrait ben il était six heures six heures et demies + le temps de les *m-* on le mettait à l'eau (corpus Bertuzzi, 38, 10)
- + puis c'était *b-* toutes ces fleurs (corpus Bertuzzi, 43, 9)
- un cyclamen je sais pas ça *va-* + sept huit francs dix francs (corpus Bertuzzi, 43, 15)
- qu'est-ce que je me suis piquée les mains avec ces bon-sang de roses hein et *mé-* quand on faisait les couronnes (corpus Bertuzzi, 12, 16)
- cependant il y a longtemps hein et ça *m-* quand j'y pense je me dis (corpus Bertuzzi, 16, 16)
- il y a eu des artistes j'ai vu *m-* comment elle s'appelle une artiste là une qui est morte qui chantait (corpus Bertuzzi, 30, 9)
- c'est à voir le colombarium c'est connu il y en a pas beaucoup d'*au-* maintenant je crois qu'il y en a (corpus Bertuzzi, 32, 9)
- dans ce temps-là c'était bien bien bien bien on *fais-* maintenant il y a tellement d'artificiel que c'est plus pareil (corpus Bertuzzi, 34, 11)

-
- je fais des choix difficiles mais je //ai, les/ fais au regard de ma conscience (Capp, 43, 20).

- mais ils n'ont pas le droit de *jou- n-* au ballon ni de faire de la patinette (corpus Bertuzzi, 29, 11)
- à quatorze ans qu'il fallait faire tout ça hein + et ben *v-* je vous assure (corpus Bertuzzi, 47, 11)

b. Amorces type Oa- ϕ b ou amorces corrigées (la correction de l'amorce a lieu sur la même place syntaxique)

Il s'agit des amorces où le locuteur ne complète pas ce qu'il avait commencé à dire mais "*corrige*" et poursuit par un autre élément. Il s'agit donc de lapsus interrompus et corrigés :

Dans le corpus "à la volée", on trouve 11 exemples de ce type :

- mais il y a eu des *palanq-* des *générations* de musiciens de jazz qui ont été sous son influence (corpus Pallaud, 1997, 30)
- ah voilà le *mas-* le *bastardon* qui revient (corpus Pallaud, 1997, 30)

Mais aussi :

- d'après ce que *vous d-* d'après ce que *nous dit* la rumeur il est mort en prison (corpus Pallaud, 1997, 34)
- La France état libéral *ne veut rien d-* est une contradiction dans les termes (corpus Pallaud, 1997, 62)

Dans les corpus enregistrés, on relève 31 amorces de cette catégorie :

- et deux fois par semaine + le *lun-* le *mercredi* et le vendredi il y avait les marchands + de plantes vertes (corpus Bertuzzi, 38, 16)
- on voyait que des fleurs que des fleurs que des fleurs parce que c'est bien *ent-* c'est bien ϕ et puis les gens (corpus Bertuzzi, 33, 3)

Dans l'exemple suivant, si la place syntaxique est conservée (attribut), la catégorie morphologique n'est plus la même : on passe de l'amorce d'un nom commun (une *merveille*) à un adjectif (*beau*) :

on rêve tellement que c'est beau hein c'est *une me-* c'est *beau beau*
beau beau (corpus Bertuzzi, 30, 2)

Ces phénomènes ne sont donc pas rares (11 sur 120) dans le recueils "à la volée" mais encore plus fréquent dans les corpus enregistrés (31 sur 100).

c. Amorces type Oa- ϕ aa ou amorces complétées (le mot commencé et interrompu se trouve complété)

Par exemple, l'amorce se trouve complétée après la reprise du déterminant s'il s'agit d'un syntagme nominal ou du pronom personnel sujet s'il s'agit d'un verbe. Il s'agit donc des amorces où le locuteur *complète* finalement ce que, dans un premier temps, il n'avait qu'ébauché pour s'interrompre aussitôt. Elles sont considérées, dans notre étude, comme

des lapsus interrompus puis poursuivis. Deux exemples de ce type seulement se trouvent dans les lapsus relevés “à la volée”, alors qu'ils abondent dans les corpus enregistrés (18 sur 59 amorces).

- Dans le recueil “à la volée” :
 - mais si on fait une *e-* une *exception* pour lui (corpus Pallaud, 1997, 78)

De plus, dans l'exemple suivant (corpus Pallaud, 1997, 76) que nous avons classé dans les lapsus du type “un mot pour un autre”,

- vous avez un *organe* un *organisme* qui gère des intérêts très différents

On peut admettre également qu'il s'agit d'une amorce *complétée* qui sera alors transcrite comme suit :

- vous avez un *organ-* un *organisme* qui gère des intérêts très différents

- Dans les corpus enregistrés :
 - c'est un film de Spielberg X qui est qui a été fait sur la seconde guerre mondiale et justement l'univers des + des des camps concentrationnaires [Juder] non *Shi-* La Liste De *Shindler* voilà (corpus Portier [Cappada, 1997])
pour tout la *cou-* la *couronne* c'était pas drôle (Bertuzzi, 47, 8)

CONCLUSION

Les traces du constant travail d'élaboration au cours de la prise de parole sont nombreuses et diverses. Elles sont l'indice d'une activité qui réorganise en permanence l'agencement temporel de l'énoncé aussi bien que son sens. La transcription soignée de corpus oraux révèle que ces remaniements sont si nombreux qu'ils interrogent le transcritteur sur leurs causes et pas seulement sur leurs mécanismes. Ce fait a été souligné maintes fois et étudié dans de nombreux travaux.

The kind of phenomena considered here are far more widespread than it would seem from the examples given, and the fact they occur at all is more important than their experimental reproduction or their quantitative evaluation. (Bierwisch, 1981, p. 600)

La transcription de l'oral montre également que l'échange communicatif entre locuteurs se déroule le plus souvent dans une atmosphère de fréquents malentendus que les locuteurs signalent peu et questionnent encore plus rarement. La réflexion développée par Seleskovitch et Lede-

rer (1984) sur la traduction simultanée les conduit, en privilégiant la situation de communication, à insister sur la distinction entre signification et sens³⁷ et à affirmer que pour l'interprète qui traduit à la vitesse de l'oral, dans la très grande majorité des cas, la parole est univoque grâce au contexte (linguistique ou non) :

La traduction montre que les mots et les phrases sont perçues avec un seul sens dans la communication [...] le traductologue ne se heurte quasiment jamais aux ambiguïtés que pourchasse le linguiste et qu'interprète le psychanalyste. (p. 120)

C'est cette prégnance de l'atmosphère communicative qui fait que les lapsus qu'ils aboutissent ou non à la formation de néologismes sont, cependant, souvent ignorés dans la conversation courante, par les locuteurs et "redressés" si les éléments cotextuels et contextuels le permettent. La communication ne semble pas affectée par ces déformations. Claire Blanche-Benveniste et Colette Jeanjean (1986) signalent que, selon certains linguistes, il en est de même pour les achoppements, amorces et bribes qui jalonnent tout énoncé oral : seule la "séquence maximale" (la séquence sans les bribes, hésitations et retours en arrière) est retenue.

Nos savoirs, le contexte (contexte et cotexte) de l'énoncé les prérogatives de la communication conduisent à "ne pas entendre" ces surprises que sont les erreurs de langage ou tout au moins à les "corriger" pour que ce qui est dit demeure cohérent, plausible : on fait crédit au locuteur d'un savoir langagier et de cohérence.

Ces considérations conduisent à souligner la distance qui sépare l'auditeur engagé dans une conversation et le linguiste qui étudie ces échanges langagiers. On ne s'étonnera donc pas de constater que les linguistes qui étudient les lapsus soulignent que la détection de ces lapsus dans les énoncés nécessite d'adopter une *attention particulière* à l'énoncé que prononce le locuteur ; elle n'est d'ailleurs pas sans rappeler l'attention flottante que prêtent les psychanalystes aux analysants³⁸. Il s'agit d'une certaine manière de se retirer du registre de la communication c'est-à-dire de la compréhension du message transmis :

37. (p. 99) La *signification* correspond à la saisie des signifiés (c'est la connaissance linguistique) ; l'accès au *sens* suppose un raisonnement qui associe les signifiés de la langue à des connaissances autres que linguistiques. Il est le résultat non seulement des effets du contexte mais aussi des savoirs et des associations du locuteur qui écoute.

38. Cette attention flottante à l'égard de l'énoncé de l'analysant place le psychanalyste hors du champ de la communication et de la compréhension ; elle constitue le support de cette place qu'occupe le psychanalyste dans l'acte analytique.

Whoever is haunted by an interest in error is unable to take part in any given conversation or listen to any lecture without welcoming every mistake with acute interest and pleasure. Without these special, separate attention, the great majority of instances will escape him via his own correction mechanism. (Bierwisch, 1981, p. 600)

L'adoption d'une attention particulière n'est cependant pas suffisante pour garantir une rigueur dans le recueil des données de ce genre. Cutler, dans son introduction du numéro spécial de la revue *Linguistics* consacrée aux *Lapsus et erreurs de langage*, incite à la prudence quant à la validité et la fiabilité des recueils d'erreurs de langage. L'exhaustivité est loin d'être garantie et partant les statistiques sont hasardeuses car la détection même des erreurs de langage pose problème. Le chercheur est lui-même non un cueilleur neutre sans filtre mais un auditeur humain de langage. Les contraintes du système de la langue vont peser et introduire des biais dans la détection. Le chercheur comme tout auditeur construit activement du sens à partir des propos qu'il entend. Si les erreurs de langage émises ne sont pas rares, les "mal entendus" ou erreurs d'écoute (*slips of the ear*) ne sont pas rares dans la vie quotidienne mais moins signalés. Ils sont peu étudiés (Cutler cite deux études) et posent eux aussi des problèmes de détection. Il reste que leur existence même incite doublement à la prudence celui qui "détecte" une erreur de langage.

Un certain nombre de linguistes insistent sur la complexité de la perception auditive d'énoncés ; on n'entend pas vraiment tout ; on anticipe le sens grâce au contexte, ce que certains linguistes nomment la suppléance mentale. Lorsqu'elle fait défaut (comme c'est le cas pour les noms propres, les mots inconnus ou les nombres le plus souvent imprévisibles), on constate une défaillance auditive (Seleskovitch et Lederer, 1984, p. 138). Cette défaillance auditive est d'autant plus grande que le traducteur comme celui qui recueille "à la volée" des lapsus ne peut compter que sur sa mémoire pour transcrire ou traduire ce qu'il vient d'entendre. On sait que la profession de traductologue est un rude apprentissage de la mémoire. Pour les mêmes raisons, dans le cas du recueil de données entendues "chemin faisant", la fiabilité des données ne peut être assurée. De plus, comme le montrent la transcription de corpus, l'exhaustivité des phénomènes est loin d'être atteinte, ne serait-ce que dans leur variété même.

Les quelques chiffres obtenus dans notre étude suggèrent s'il en était besoin le biais qui peut être introduit par un recueil "à la volée" et le grand avantage qu'il peut y avoir à bénéficier de corpus enregistrés. Dans les recueils "à la volée", les lapsus du type "un mot pour un autre" sont

très surestimés alors que toutes les hésitations se traduisant par des amorces (amorces corrigées et inachevées) se trouvent sous estimées.

Berthille PALLAUD
Université de Provence



Bibliographie

- ARNAUD, P.J.L. [1997], "Les ratés de la dénomination. Typologies dans les lapsus", in Boysson et Thoinon (éds), *La dénomination*, PUL, Lyon, 307 p.
- AUTHIER-REVUZ, J. [1996], *Ces mots qui ne vont pas de soi. Boucles réflexives et non coïncidences du dire*.
- BIERWISCH, M. [1982], "Linguistics and language error", in Cutler, A. (ed.), *Slips of the Tongue and Language Production*, Moton Publishers, Amsterdam, pp. 583-626.
- BLANCHE-BENVENISTE, C. & JEANJEAN, C. [1987], *Le français parlé. Transcription et édition*, Didier Érudition, Paris.
- BLANCHE-BENVENISTE, C., BILGER, M., ROUGET C. & van Den EYNDE, K. [1990], *Le français parlé. Études grammaticales*, Édition du CNRS, Paris.
- BLANCHE-BENVENISTE, C. [1997], "Transcription et technologie", *Recherches Sur le Français Parlé*, 14, pp. 87-100.
- [1997], "Ponctuation et langue parlée", in *La ponctuation. Le Discours psychanalytique*, 18, pp. 73-109.
- BUTTERWORTH, B. (ed.) [1980], *Language production*, Academic Press, London.
- CUTLER, A. (ed.) [1981], *Slips of the Tongue and Language Production*, Moton Publishers, Amsterdam.
- CUTLER, A. [1981], "The reliability of speech error data", in Cutler, A. (ed.) *Slips of the Tongue and Language Production*, Moton publishers, Amsterdam.
- FÉNOGLIO, I. [1997], "La notion d'événement d'énonciation : le 'lapsus' comme une donnée d'articulation entre discours et parole", *Langage et société*, 80, pp. 39-71.
- FUCHS, C. [1996], *Les ambiguïtés du français*, Paris, Ophrys.
- GOFFMAN, E. [1981], *Forms of Talk*, Basil, Blackwell, Oxford.
- GRÉSILLON, A. [1984], *La règle et le monstre : le mot-valise. Interrogations sur la langue, à partir d'un corpus de Heinrich Heine*, Linguistische Arbeiten, Niemeyer.

- JAKOBSON, R. [1963], *Essais de linguistique générale*, vol. I, Éd. de Minuit, Paris.
- JEANJEAN, C. [1984], “‘Les ratés c'est fa- fabuleux’, Étude syntaxique des amorces et des répétitions”, *LINX*, 10, *Syntaxe et discours*, pp. 171-177.
- KLEIBER, G. [1987], “Connexité, cohésion, cohérence et lexique”, in F. Wiloand (éd.), *Enseignement/Apprentissage du lexique en langue étrangère*, Strasbourg, Université des Sciences humaines.
- KRISTOL, A.M. [1997], *Atlas linguistique audiovisuel du Valais romand*, Centre de dialectologie et d'étude du français régional de Suisse romande, Université de Neuchâtel.
- LOUFRANI, C. [1981], “Locuteur collectif ou locuteur tout court”, *Recherches sur le Français parlé*, 3, pp. 215-243.
- NATTINGER, J.R. & DE CARRICO, J. S. [1992], *Lexical Phrases and Language Teaching*, OUP, Oxford.
- MERINGER, R. und MAYER, K. [1895], *Versprechen und Verlesen : eine Psychologisch-linguistische Studie*, Göschen, Stuttgart.
- MERINGER, R. [1908], *Aus dem Leben der Sprache : Versprechen ; Kindersprache, Nachahmungstrieb*, Behr's Verlag. Ling, sémantique, lapsus, Berlin.
- [1923], “Die täglichen Fehler im Sprechen, Lesen und Handeln”, *Wörter und Sachen*, 8, pp. 122-140.
- MOTLEY, M. [1985], “Les lapsus”, *Pour la Science*, nov., pp. 62-65.
- PASQUALE, A. [1989], *Étude diachronique du langage d'un enfant inintelligible de 7;4 à 8;5*, Mémoire de maîtrise 1988-1989 en Sciences du langage, Université de Provence, Aix-Marseille I.
- PETER-DEFAGE, E. [1993], *Aspects phonologiques, syntaxiques et phonologiques de l'empan des erreurs de langage*, DEA de Phonétique, Université de Provence, Institut de Phonétique, Aix-en-Provence, 166 p.
- PILLON, A. [1998], “Morpheme units in speech production evidence from laboratory-induced verbal slips”, *Language and Cognitive Processes*, 13, 4, pp. 465-498.
- ROSSI, M. & PETER-DEFAGE, E. [1995], “Lapsus linguae : word errors or phonological errors ?”, *International Journal of Psycholinguistics*, 11, 1[30], pp. 5-38.
- [1998], *Les lapsus ou comment notre fourche a langué*, PUF, Paris.
- SAUSSURE, F. de [1972], *Cours de linguistique générale*(1922), Payot, Paris.
- SELESKOVITCH, D. & LEDERER, M. [1984], *Interpréter pour traduire*, Didier Érudition, Paris.
- Trésor de la Langue Française* [1983], Édition du CNRS, Paris.
- WEIL D. [1997], “Paraphrase et mi-dire dans la parole de femmes enceintes”, in Fuchs C. (éd.), *L'ambiguïté et la paraphrase. Opérations linguistiques, processus cognitifs, traitements automatisés*, Centre de publications de l'Université de Caen.

Éclairage des corpus de type aphasique par les grilles

Une analyse syntaxique des discours pathologiques d'origine aphasique ne peut se réduire à l'étude de fragments mais exige des ensembles plus vastes. Derrière l'émission apparente de ces productions se cachent des organisations comparables à celles de locuteurs ordinaires, mais plus complexes à analyser. Pour rendre compte des stratégies déployées pour compenser le déficit langagier, les outils d'analyse développés par le GARS¹, comme la mise en grille, sont commodes.

1. CHOIX DES CORPUS

1.1. CRITÈRES DE SÉLECTION

Certaines productions sont difficiles à aborder du fait :

- de leur faible débit² (37 mots par minute), des silences (jusqu'à 11 secondes) et de leur mode de production par syllabes isolées

pur râtre u sor tir dð - la fεz - mε bð - a apar mð ner go
lezotr ð vwa kizð pa fe zetyd pur pr.ldr le 31 (Patane, 4,
7³)

- de la présence de jargon

bè oui euh ils sont [pəsələsati] (Brémond, 99, 11)

[lalesələləsilələsila] (Brémond, 99, 11)

1. GARS : Groupe Aixois de Recherches en Syntaxe.
2. Le débit d'un locuteur ordinaire en situation de conversation est en moyenne de 150 mots par minute.
3. Le premier nombre correspond à la page du corpus d'où est extrait l'exemple et le suivant correspond au numéro de la première ligne de transcription.

- des réponses incongrues en *oui* ou *non*

Thérapeute : à quel endroit êtes-vous

Patient : oui (Löffler, 10, 1)

Thérapeute : ça fait longtemps que vous êtes parmi nous [...] quatorze jours

Patient : oui

Remarque : le patient n'est hospitalisé que depuis neuf jours (Löffler, 11, 5)

- de phénomènes massifs d'écholalie⁴

Thérapeute : ah non c'était hier

Patient : c'était hier ah (Löffler, 74, 9)

Thérapeute : pas plus que ça

Patient : pas plus que ça (Löffler, 108, 3)

Ces productions ne répondent pas aux exigences minimales qu'impose une analyse de discours : prise de parole d'au moins une minute, débit entre 50 et 100 mots par minute, sans silence démesuré, sans réponse en *oui* ou *non*, avec peu de phénomènes d'écholalie ou de jargon. Cependant une lecture attentive des corpus montre qu'à d'autres moments, ces malades sont plus productifs. L'intérêt d'une analyse de discours pathologique est donc de prendre ces fragments dans un ensemble plus vaste, pour restituer l'ensemble des capacités langagières du patient et présenter un tableau plus complet.

1.2. PRÉSENTATION DE LA SITUATION D'INTERVIEW

Les locuteurs, classés comme aphasiques⁵ par le milieu médical, sont en cours de rééducation dans un centre hospitalier spécialisé. Pendant les séances, ces malades rencontrent un thérapeute qui leur propose des exercices et qui les interviewe afin d'évaluer leurs difficultés. Certaines questions du thérapeute amènent quelquefois des réponses assez longues :

4. L'écholalie est un procédé par lequel le locuteur répète en partie ou en totalité ce que vient de dire le thérapeute.

5. Les locuteurs classés aphasiques sont repérés dans la bibliographie par leur prénom, le titre de l'extrait et la référence du corpus.

<i>Corpus</i>	<i>Question du thérapeute</i>	<i>Thème⁶ développé par le patient</i>
corpus 1 :	qu'est-ce que vous me disiez qui n'allait pas	le fauteuil électrique
corpus 2 :	comment [se fait la pêche]	la pêche
corpus 3 :	bien hier soir nous avons vu euh examiné ce qu'était une chauve- souris hum est-ce que vous pouvez m'expliquer là	la chauve-souris
corpus 4 :	comment fait-on une omelette dites-moi	l'omelette
corpus 5 :	comment vous faites les lasagnes dites-moi	les lasagnes
corpus 6 :	vous aviez une petite aphasie	le trouble du langage
corpus 7 :	racontez moi ce qui vous est arrivé	les aphtes

Notre démarche a pour objectif de mettre en évidence à la fois les régularités et les singularités de ces textes.

2. TECHNIQUE DE LA MISE EN GRILLE

2.1. FONCTIONNEMENT DE LA MISE EN GRILLE

La mise en grille a pour fonction de visualiser par écrit l'organisation du discours oral, qui se développe par fragments⁷. On met sur la même ligne tous les éléments en rapport de dépendance syntaxique : c'est l'axe des syntagmes (ou axe syntagmatique). Quand il n'y a plus de rapport de dépendance, on change de ligne, comme dans l'exemple ci-dessous où chaque construction verbale (même inachevée) occupe une ligne entière⁸ :

je fais une sauce tomate
je repasse le
je mets du gruyère râpé
 et *je repasse* tout au four
 et *je laisse cuire* à petit four pour (les lasagnes)

6. Ce thème servira de titre à chaque extrait cité dans l'article. Les corpus appartiennent à des mémoires de maîtrise (département de linguistique française, Université de Provence).

7. Des techniques récentes présentent les textes oraux en "alignement" sur CD-Rom. Pendant qu'on entend la parole du ou des locuteurs, le texte écrit et la prosodie défilent simultanément sur l'écran.

8. La juxtaposition et la coordination ne sont pas considérées comme des dépendances syntaxiques.

On aligne, dans une même “colonne” verticale, les éléments qui, dans une construction donnée, occupent la même place syntaxique (c'est l'axe des paradigmes) :

<i>je fais</i>	<i>une sauce tomate</i>			
<i>je repasse</i>	<i>le</i>			
<i>je mets</i>	<i>du gruyère râpé</i>			
et <i>je repasse</i>		<i>tout au four</i>		
et <i>je laisse cuire</i>		<i>à petit four</i>	<i>pour</i>	(les lasagnes)

Les éléments comparables de la valence verbale ont été alignés sans tenir compte de leur nature ou de leur lexique. Les places de construction verbale se trouvent ainsi sur l'axe syntagmatique, mais leur remplissage lexical se fait en plusieurs étapes sur l'axe paradigmatique.

Contrairement à l'écrit, l'oral ne filtre pas les ratés de la communication : les hésitations, les amorces, les bribes, les ébauches... Par respect des textes, toutes ces productions sont prises en charge dans les grilles, comme dans les exemples ci-dessous où apparaissent des hésitations et des recherches sur le lexique :

Le locuteur (L), ancien pêcheur professionnel, explique au thérapeute (T) la pratique de la pêche au râteau.

L	alors le râteau	c'était		<i>le</i>
		ça se pratique		<i>de nuit</i>
			et	

T			avec	<i>une lampe hein</i>

L				<i>non</i>
				<i>une</i>
				<i>une</i>
				<i>une petite lampe</i> (la pêche)

Le locuteur décrit la chauve-souris

ensuite	<i>la</i>			
	<i>la souris</i>			
	<i>la souris</i>			
la chauve-souris vous dites		elle a	<i>deux</i>	
			<i>deux propriétés chez vous</i>	
			<i>deux</i>	
			<i>deux propriétés</i>	
			<i>deux particularités</i>	(la chauve-souris)

Cette mise à la verticale révèle les recherches lexicales du locuteur, qu'une simple écoute ou une transcription linéaire ne mettrait pas aussi bien en évidence.

La dimension énonciative⁹ est très présente en particulier dans les commentaires des locuteurs sur leur propre production langagière. Voici des exemples présentés entre crochets :

elle devine que	si	la		
	si	la personne est à peu près pareille		
		[la personne enfin en temps normaux		
		la personne c'est plutôt la la bête]		(la chauve-souris)
et ensuite après est venue cette genre de dyslexie un peu non				
		[je ne sais pas si c'est comme ça que ça s'appelle		
		je sais plus]		(les aphtes)
rien que d'y penser qu' on va être en re-				
		on va être en retard pour le dire ça marche pas du tout		
		[vous comprenez]		(le trouble du langage)

2.2. RÉGULARITÉS OBSERVABLES DANS LES GRILLES

L'analyse en grille des textes pathologiques montre des régularités semblables à celles des textes ordinaires. Nous en avons dégagé quelques unes pour la syntaxe nominale, l'ordre des mots et les agencements de constructions verbales.

2.2.1. LA SYNTAXE NOMINALE

2.2.1.1. Le double marquage du sujet

Le sujet est repris deux fois : une fois sous forme de pronom et une fois sous forme nominale. Les exemples suivants font apparaître ces syntagmes nominaux à gauche de la construction verbale :

<i>les ultrasons</i>	<i>ils</i>			
	ce sont	en principe		
	c'est	en principe	le	(la chauve-souris)
<i>et le menton</i>	si on			
	si <i>il</i> ¹⁰ est étroit	la		(le fauteuil électrique)

9. Cette analyse du discours reste ouverte à d'autres domaines (pragmatique, sémantique...) que nous n'abordons pas ici.

10. Il existe une autre possibilité de mise en grille : *il* peut référer au *menton* mais aussi à *l'appui du fauteuil* et *la* peut être le locatif *là* ou l'amorce du syntagme nominal qui suit : *la boule en mousse*.

2.2.2. L'ORDRE DES MOTS¹¹

2.2.2.1. La place du sujet et des compléments nominaux

• sujet postposé

et ensuite après est venue *cette genre de dyslexie* (les aphtes)

• objets antéposés

quatre scanners j'ai eu (les aphtes)

la boule en mousse on avait changé (le fauteuil électrique)

2.2.2.2. La place des subordonnées

Les syntagmes en *parce que* et *puisque* ont tendance à se placer après la construction verbale :

mais la deuxième fois c'était moins *parce que* *mes my-*
les
les mycoses que j'avais dans la bouche
étaient un peu parties
 (les aphtes)

mais il est en location pour le moment
parce qu'on attend un système *avec* *les*
la langue
le palais
et *les dents vers le palais*
 (le fauteuil électrique)

elle a des ailes immenses *puisque elle a des ultras-*
sons qu'on peut appeler
 (la chauve-souris)

Les temporelles peuvent se placer avant la construction verbale :

et *quand*
quand *c'est plein*

et on la soulève
 et on la met
 on la vide à bord

et on recadre
 et puis on trie le poisson (la pêche)
 et *quand elle voit une bestiole qui la voit en face*

eh bien elle s'arrête
 (la chauve-souris)

11. Cf. Blanche-Benveniste, 1996.

et puis *une fois que c'est bien battu* vous mettez dans
dans les œufs (l'omelette)

2.2.3. LES AGENCEMENTS DE CONSTRUCTIONS VERBALES

2.2.3.1. Des dispositifs pseudo-clivés¹² en ce qu- ... c'est

ce qui m'a s-
surtout beaucoup fait souffrir *c'est* une¹³
une fibroscopie (les aphtes)
ce que je vous dis *c'est* que le jour que vous êtes ... (le trouble du langage)

2.2.3.2. Des structures en c'est

il se passe une chose *c'est* que il faut un certain temps pour faire redire quelque chose
(le trouble du langage)

2.2.3.3. Le double marquage des sujets en ça

ça c'est l'examen le plus terrible que j'ai pu passer [...]
ça c'est
c'est l'examen le pire que j'ai eu de toute depuis que je suis en maladie quoi
(les aphtes)

2.2.3.4. Des dispositifs d'extraction en c'est ... qu-

puis de pas parler justement *c'est*
c'est
ça devient
ça devient encore très dur aussi
c'est ça qui
qui
qui est énorme (le trouble du langage)

2.2.3.5. Des questions en qu'est-ce que

qu'est ce que j'ai eu (les aphtes)
qu'est-ce qu'on m'a fait comme examen (les aphtes)

L'accumulation de ces procédés est parfois difficile à la première lecture et traduit les dysfonctionnements de la langue. La mise en grille met en lumière différents procédés lexicaux ou syntaxiques ; elle révèle éga-

12. Pour l'étude des dispositifs pseudo-clivés, cf. Roubaud (1994 et à paraître).

13. Entre la production de *une* et celle de *une fibroscopie* a lieu un long commentaire du malade sur ce mot qu'il ne trouve pas : *comment on appelle ça maintenant eh l'on met un tube dans le parce on pensait que j'en avais dans l'estomac et tout bien entendu ça s'appelle une fibro. C'est le thérapeute qui va dire scopie et cette intervention achèvera la recherche du malade qui produira le mot attendu : "fibroscopie".*

lement comment la production s'organise. Elle permet de situer les difficultés des locuteurs au travers de grandes organisations.

3. TYPES D'ORGANISATIONS DISCURSIVES

Nous avons identifié trois grands types d'organisations : syntaxique, lexical et lexico-sémantique. Afin de rendre plus lisible la lecture de ces organisations, nous avons élaboré une convention de critères graphiques, présentés ci-dessous :

Convention de critères graphiques	
encadré et grisé :	thème remarquable
encadré :	unités syntaxiques
entre crochets :	commentaires du locuteur (métalangue)
majuscules :	interventions de l'orthophoniste ou d'un autre locuteur
gras :	répétitions ou structures remarquables
italique :	paroles rapportées
gras et italique :	syntagmes remarquables à gauche ou à droite de la construction verbale

3.1. ORGANISATION DE TYPE SYNTAXIQUE

Ce type d'organisation donne priorité à la syntaxe, le lexique faisant défaut. Dans les exemples suivants, le locuteur cherche à remplir un syntagme dans une place de construction verbale, le plus souvent sur la place complément. C'est la raison pour laquelle il accumule des constructions verbales (sujet, verbe) sans pour autant déboucher sur le lexique approprié.

Pour simplifier la lisibilité des exemples, nous ne donnerons qu'une partie de la production.

Exemple 1 : la grille vide (93, 5)

M. Bertrand, ancien médecin, est interrogé sur ses activités passées mais la recherche lexicale n'aboutit qu'à la production de jargon.

VOUS FAISIEZ DE LA MÉDECINE GÉNÉRALE

oui

oui

oui oh

eh la

la

la médecine hein c'est

HUM
 bon
 eh
 eh
 alors là

c'est [mpal]¹⁴

c'est euh
 euh
 t-
 trois ans de

VOUS AVIEZ FAIT TROIS ANS EN ALGÉRIE

non
 non
 non euh
 bon

c'est euh

non
 non
 non

il y en a de
 c'est eh
 par le beau [nɛmbe]
 Reims

Exemple 2 : les lasagnes

Maryse, en fin de reeducation, explique sa recette des cannelloni mais c'est le mot *lasagnes* qui lui vient, sans la satisfaire. Finalement, la recherche aboutira.

	c'est pas	des
	c'est	des
ET ÇA C'EST PAS LASAGNES		
non		
non		
		calle- callene callenoni
hein	c'est	ça
	[je crois mais le mot n'y est pas]	cannelloni
HUM CANNELLONI		
eh oui	c'est	ça

Exemple 3 : le trouble du langage

M. Bernard décrit les difficultés qu'il a à s'exprimer lorsqu'il rencontre d'autres personnes. Les négations qui ponctuent son texte traduisent sa détresse de ne pouvoir communiquer avec les autres. L'emploi en alternance de *pouvoir* et de *vouloir* indique

14. Les mots entre crochets sont du jargon transcrit phonétiquement.

les difficultés du locuteur à “dire” malgré sa volonté de “dire”. Dans cette grille, on note une très grande densité de lexique verbal par rapport au lexique nominal.

rien que d'y penser qu'on va être en re-		
on va être en retard pour le dire ça	marche pas du tout	
[vous comprenez]		
[c'est complètement	i-	
	i-	
	i-	
	idiot]	
	<i>c'est-à-dire que</i>	vous pou- pas
		vous voulez
		vous avez la conversation à dire
	mais là	vous y
		vous voulez dire
		vous
		vous voyez
		ils voudront dire <i>il y a ça qui va pas</i>
	mais au moment de le dire	c'est déjà passé
		on peut pas le dire

Ces trois exemples montrent que la production s'organise autour d'une armature syntaxique mais peut tourner à vide quand le lexique fait défaut. Les places de construction verbale sont marquées par des verbes de grande fréquence comme *c'est* (ex. 1 et 2) ou par des modaux comme *pouvoir* et *aller* (ex. 3) mais la recherche lexicale n'aboutit pas. Pour pallier ce manque du lexique nominal, le locuteur glose le plus souvent par *ça*.

3.2. ORGANISATION DE TYPE LEXICAL

Ce type d'organisation donne priorité au lexique, la syntaxe faisant défaut et, contrairement au type d'organisation syntaxique, cette production se développe par ramifications ou éclatement sur l'axe paradigmatique. Le lexique est souvent répété avec des effets d'emboîtement.

Exemple 4 : l'omelette

Emily raconte comment elle fait une omelette.

et puis *une fois que c'est bien battu* vous mettez dans
 dans les oeufs
 dans les pommes de terre
 dans l'omelette
 dans le
 dans
 dans
 dans des pommes de terre
 dans des oignons
 enfin des trucs

Exemple 5 : le trouble du langage

M. Bernard décrit les difficultés qu'il éprouve pour s'exprimer devant plusieurs personnes.

le jour que vous êtes en
 en
 plusieurs X
 plusieurs dans le coup
 le
 plusieurs
 vous êtes trois
 ou quatre eh ben au bout
 au bout de
 des cinq
 six minutes
 après on va vous laisser

Exemple 6 : les aphtes

Mme Dossier, âgée de 60 ans, explique au thérapeute qu'elle a souffert d'aphtes, ce qui a entraîné une perte de poids de plus de neuf kilos et une alimentation spécialisée.

on ne me donnait que des
 des trucs vitaminés là enfin
 [c'est peut-être pas le mot qui convient mais enfin disons]
 des
 DES PRODUITS
 des produits qui étaient
 et que je p-
 prenais avec une paille parce que XX¹⁵ j'avais

Dans ce type d'organisation, la syntaxe est pauvre. La recherche du mot concentre l'organisation du discours sur l'axe paradigmatique. C'est la fréquence et la longueur de ces recherches qui saturent les textes et les rendent difficiles à l'écoute.

3.3. ORGANISATION DE TYPE LEXICO-SÉMANTIQUE

La mise en grille, au-delà des organisations de type syntaxique ou lexical, permet aussi d'accéder à une dimension sémantique. Cette organisation, insensible à l'écoute et à la lecture, échappe au locuteur et à

15. Les X dans la transcription indiquent des syllabes incompréhensibles à l'écoute ; il y a autant de croix que de syllabes non identifiées.

l'interlocuteur. L'intérêt de la grille est de démontrer le mécanisme en révélant cette dimension lexico-sémantique.

Exemple 7 : la chauve-souris

Emily tente de caractériser la chauve-souris.

1. alors	la chauve- la chauve-souris	c'est une souris c'est une c'est un oiseau qui la	
2.		elle est elle est beaucoup plus grande hein elle a des ailes immenses	puisqu' elle a des ultras- sons qu'on peut appeler
3. et ces	ces ailes immenses	peut les avoir euh des propriétés euh nouvelles	parce que parce que elles
4.		ça ça lui permet d'a- d'arrêter son son travail	
5. c'est-à-dire qu'		elle peut s'arrêter de voler elle peut s'arrêter de chasser	pour ne pas faire du bruit
et puis et puis et puis après		elle peut s'arrêter ainsi de faire quelque chose quoi elle	

Dans cet exemple, l'organisation textuelle repose sur l'agencement de cinq structures syntaxiques, correspondant à autant de thèmes majeurs, que nous schématisons de la manière suivante (les chiffres renvoient à la grille ci-dessus) :

	Identification zoologique	Classification zoologique	Caractérisation zoologique
Construction 1	Dénomination de la catégorie animale	Animal terrestre Animal volant	Description
Construction 2	Propriétés de cet animal	Particularités physiques	Énumération
Construction 3		Propriétés et capacités	Propriétés
Construction 4			Capacités
Construction 5			

Exemple 8 : la pêche

M. F. explique la technique de pêche au râteau.

1.		ah ou	un gros bateau un vingt-quatre empan
2. et		je pêchais	<i>des oursins</i>
			<i>et</i>
			<i>et des Bernard l'hermite</i>
		et puis	je faisais le râteau
		AH QU'EST-CE QUE C'EST LE RATEAU	
3. alors	<i>le râteau</i>	c' était	le
		ça se pratique	de nuit
			<i>et</i>
			AVEC UNE LAMPE HEIN
			non
			une
			une
			une petite lampe
et			
et		ça consiste	à
			à
			à draguer
vous voyez		c'est une drague	
et			qu' on
			qu' on soulève
4. et	<i>quand</i>		
	<i>quand c'est plein</i>		
et		on la soulève	
et		on la met	
		on la vide à bord	
et			
et puis		on recadre	
et		on trie le poisson	

Dans cet exemple, l'organisation textuelle repose sur l'agencement de quatre structures syntaxiques, correspondant à autant de thèmes majeurs :

Construction	Dénomination de l'objet	Bateau	Description
1			
Construction 2	Caractéristiques de la pêche	Produits de la pêche Nom de la pêche	Énumération Identification
Construction 3	Caractéristiques de la pêche au râteau	Éléments de la pêche au râteau	Description
Construction 4		Pratique de la pêche au râteau	Explication

Exemple 9 : le fauteuil électrique

Le fauteuil de Jean-Baptiste, tétraplégique, a été donné en réparation (fauteuil n° 2). Un fauteuil de remplacement le fait souffrir (fauteuil n° 1).

1.	manier l'horreur qui est mon fauteuil électrique		
2.	je manie		avec le menton
	et le menton si on		
	si il est étroit la		
	la boule en mousse on avait changé		
3. mais	il est en location ...	parce qu' on attend un système avec	les la langue le palais et les dents vers le palais
	je pourrais	mettre la télévision actionner mon truc	
	mais	ça va être	compliqué
	mais	c' est pas	encore là
4. et	je perds	trois à quatre kilos de transpiration	

Dans cet exemple, l'organisation textuelle repose sur l'agencement de quatre structures syntaxiques et autant de thèmes majeurs, avec un effet d'ouverture et de fermeture autour de la souffrance du locuteur :

Construction	Dénomination de l'objet	Expression de la	Constatation
1	(fauteuil n° 1)	souffrance	
2	Propriétés du fauteuil n° 1	Particularités	Description
3	Propriétés du fauteuil n° 2	Particularités Possibilités	Propriétés
4	Retour à l'objet de départ (fauteuil n° 1)	Expression de la souffrance	Constatation

La grille lexico-sémantique, sur ces trois textes, permet de repérer une organisation qui structure le discours.

CONCLUSION

La mise en grille (syntaxique, lexicale ou lexico-sémantique) pratiquée sur des textes d'origine pathologique a montré que ces productions étaient analysables avec les outils du français oral. Mais cette démarche d'analyse est difficile à mettre en place parce que, dans les situations d'interviews,

les locuteurs ont plutôt tendance à produire des réponses brèves et il faut toute l'expérience du thérapeute pour amener le patient à en "dire plus".

L'analyse de morceaux étendus permet de dépasser les troubles liés au dysfonctionnement de la langue et de déboucher sur un travail de reconstruction, mettant en évidence les stratégies déployées par le malade et ses efforts pour compenser son trouble.

Marie-Noëlle ROUBAUD
Claude LOUFRANI
Université de Provence



Références bibliographiques

- BERNSTEIN, B. (1975). *Langage et classes sociales. Codes socio-linguistiques et contrôle social*. Paris, Ed. de Minuit.
- BLANCHE BENVENISTE, C. (1990), *Le français parlé. Études grammaticales*, Paris, CNRS.
- (1996), "Trois remarques sur l'ordre des mots dans la langue parlée", *Langue française*, 111, *L'Ordre des mots*, coordonné par Henning Nolke et Hanne Korzen, pp. 109-117.
- (1997), *Approches de la langue parlée en français*, Paris, Ophrys.
- DORZE, G. de (1985), "L'aphasie et le processus de lexicalisation", thèse de doctorat, Université de Montréal.
- LOUFRANI, C. (1988), "Un corpus de locuteur aphasique, originalité ou régularité ?", *Recherches sur le Français parlé*, 8, pp. 59-78.
- (1990), "Analyse du discours de locuteurs classés comme aphasiques : de l'art de tourner autour du mot", thèse de doctorat de linguistique française, Université d'Aix-Marseille I.
- LOUFRANI, C. et ROUBAUD, M.-N. (1990), "La notion d'approximation : langage ordinaire ou pathologique", *Recherches sur le Français parlé*, 10, pp. 131-142.
- NESPOULOUS, J.-L. (1980), "De deux comportements verbaux de base : référentiel et modalisateur. De leur dissociation dans le discours aphasique", *Cahiers de psychologie*, 23, pp. 195-210.
- ROUBAUD, M.-N. (1987), "L'approximation lexicale", mémoire de maîtrise, Département de linguistique française, Université de Provence.
- (1988), "Les mystères de l'approximation lexicale", mémoire de DEA, Département de linguistique française, Université de Provence.
- (1994), "Étude des énoncés pseudo-clivés dans le cadre de l'Approche Pronominale", thèse de 3^e cycle, Lettres et Sciences Humaines, Université de Provence.

— (à paraître), *Les énoncés pseudo-clivés en français contemporain*, Paris, Champion, (Coll. : INaLF).

ROUBAUD, M.-N. et LOUFRANI, C. (1993), “La syntaxe, c’est ce qui reste quand on a tout oublié’ : La syntaxe comme outil de description syntaxique”, *Recherches sur le Français parlé*, 12, pp. 85-113.

• Les locuteurs et le titre des grilles

Maryse (HOSNI, I)¹⁶ : les lasagnes

M. Bertrand (LÖFFLER, II) : la grille vide

Emily (NOLOT, II) : la chauve-souris, l’omelette

Jean-Baptiste (PATANE) : le fauteuil électrique

Mme Dossier (SPILKER, I) : les aphtes

M. F. (TRUONG, K.-L., II) : la pêche

M. Bernard (TRUONG, K.-Y., II) : le trouble du langage

L’ensemble de ces grilles sera présenté dans une édition de corpus aphasiques à paraître.

16. On retrouve le même corpus chez Loufrani (corpus V).

Remarques sur le français parlé de locuteurs de la région de Marseille. Usage régional du français et “régionalisme”

L'émergence et de le développement — récent en France — d'une linguistique appuyée sur de grands corpus est à l'origine d'une remise en question des études consacrées à ce qu'on appelle “la variation linguistique” en morphosyntaxe du français¹. Je m'interrogerai en particulier dans ces lignes sur le statut linguistique du français dit “régional” à Marseille.

1. FRANÇAIS “RÉGIONAL”, “RÉGIONALISME” ET VARIATION

1.1. UN NOUVEL INTÉRÊT POUR LA NOTION DE FRANÇAIS RÉGIONAL

Les études consacrées à la variation linguistique en syntaxe du français ont dégagé jusqu'à présent peu d'ouvrages descriptifs de référence. Quant au matériau lui-même, on y recense plus d'études conduites hors de l'hexagone — français du Québec² ou de l'Ontario³ — que d'études

1. Voir C. Blanche-Benveniste, 1997, “La notion de variation syntaxique dans la langue parlée”, *Langue française* 115, pp. 19-30.

2. Voir D. Sankoff, 1988, “Sociolinguistics and syntactic variation”, in F.J. Newmeyer (ed.), *Linguistics: The Cambridge Survey, vol. IV, Language : The Socio-cultural Context*. Cambridge, Cambridge University Press, pp. 140-161.

3. Voir E. Béniak et R. Mougeon, 1989, “Recherches sociolinguistiques sur la variabilité en français ontarien”, in *Le français canadien parlé hors du Québec*, Les Presses de l'Université Laval, pp. 53-74.

conduites en France métropolitaine, si on exclut les travaux du GARS⁴, dont l'objectif premier n'est pas l'étude de la variation mais la description des usages en français parlé. En revanche, comme l'écrit C. Blanche-Benveniste (1997), on peut attendre beaucoup des analyses conduites, à l'instar des études anglo-saxonnes récentes (Sinclair 1996, et Biber et alii 1996), en particulier une description de l'hétérogénéité des usages que manifestent les divers groupes de locuteurs, indépendamment des variantes liées au genre⁵ des productions examinées.

Assez souvent, les études sur la variation en français contemporain sont amenées à tenir compte des données historiques sur la variation ; en revanche, elles se montrent timides pour prendre en compte les données recueillies dans les études consacrées à telle ou telle variété de français régional d'hier comme d'aujourd'hui.

Il faut admettre la réticence que manifestent nombre de linguistes, en toute bonne foi, à l'égard de ces variétés de notre langue : je crois, comme D. Baggioni (1993) que

dans l'histoire de la langue française, la variété régionale [du français] est conçue comme phase transitoire dans le processus de changement de langue et les caractéristiques de cette variété essentiellement envisagées en termes de survivances substratiques.⁶

Je reprends la citation de Séguy (1951) qu'il donne à l'appui :

Le français de Toulouse, comme celui de tout le midi, est donc essentiellement caractérisé par la présence de substrats occitans [...]. Mais il saute aux yeux qu'il l'est moins que par le passé. La tendance générale et ancienne est à l'effacement des substrats au bénéfice des normes officielles.⁷

En revanche, je crois qu'il se trompe quand il émet le doute que les dialectologues qui se sont essayés à la description des français régionaux aient véritablement rompu avec cette vision des variétés régionales du français. J'en ai pour preuve la position claire exprimée par Tuailon (1983) :

4. Groupe Aixois de Recherche en Syntaxe, dirigé par C. Blanche-Benveniste, qui regroupe des chercheurs en linguistique française de l'Université de Provence et d'autres, rattachés.

5. Tous les exemples qui seront cités dans cet article renvoient au au genre de la conversation, qui présente, comme le rappelle C. Blanche-Benveniste, une grande variété interne (p. 26).

6. p. 26.

7. J. Séguy, 1951, *Le français parlé à Toulouse*, Privat, Toulouse, p. 101.

Quelques linguistes pensent qu'entre le dialecte immédiatement issu de la latinisation selon des évolutions propres à la région et le français officiel qu'ils appellent parfois *standard* mais qu'il vaut mieux appeler le *français commun* ou le *français tout court*, il existe une langue à part, produit de l'hybridation entre le dialecte et le français. Cette langue à part, différente à la fois du dialecte et du français serait le français régional, mais il n'en est pas ainsi. La locution *français régional* n'est qu'une commodité d'expression pour parler de l'ensemble des particularités linguistiques qui distinguent l'usage français d'un lieu ou d'une région donnée.⁸

On retrouve dans un article de Chambon (1996), où il dénonce une illusion d'optique en lexicologie française historique, une conception assez proche de cette notion de français régional : il montre qu'un mot comme *fourme* ne constitue pas un emprunt du français à l'occitan d'Auvergne, mais l'exemple d'une migration d'un mot qui a appartenu au français régional d'Auvergne depuis le XV^e siècle ; "en tant que régionalisme diachronique du français commun, *fourme* est un pur emprunt interne, d'une variété de français à une autre variété de français"⁹.

Le patient travail de documentation historique sur les variétés régionales du français qui est conduit par des dialectologues comme J.-P. Chambon en lexicologie ne connaît pas vraiment d'équivalent dans le domaine morphosyntaxique. Brun, déjà, regrettait¹⁰ de n'avoir pu, faute de documents, conduire une enquête historique sur le français régional de Marseille.

Dans ce domaine, il existe malgré tout une source de documentation importante : toute la série des ouvrages normatifs des XVIII^e et XIX^e siècles. L'intérêt de ces ouvrages n'est pas qu'ils mettent en évidence, à coup sûr, les traits lexicaux et morphosyntaxiques les plus saillants des français régionaux à cette époque. Comme l'a noté D. Baggioni (1993), et avant lui S. Branca-Rosoff (1987), à propos de l'ouvrage de Desgrouais¹¹, les auteurs ont fait l'inventaire des fautes de français commises par des locuteurs appartenant aux différentes couches de la bourgeoisie du temps. Brun, à la différence de Séguy, considérait également les recueils de ce

8. p. 19.

9. J.-P. Chambon, 1996, "Emprunts du français aux dialectes ou patois : une illusion d'optique ?", *Lalies* 17, *Actes des sessions de linguistique et de littérature*, Aussois, 2-7 septembre 1996, Presses de l'ENS Paris, pp. 33-53, 44.

10. p. 20.

11. Desgrouais, 1766, *Les gasconismes corrigés. Ouvrage utile à toutes les personnes qui veulent parler et écrire correctement, et principalement aux jeunes gens, dont l'éducation n'est point encore formée*, J.-J. Robert, Toulouse.

type réalisés par Gabrielli et Reynier pour la région provençale comme des ouvrages "qui s'inspirent d'un purisme étriqué"¹².

Il n'en reste pas moins que Brun a puisé dans l'inventaire de Gabrielli au moins autant que Séguy dans celui de Desgrouais. On réalise, quand on parcourt ces inventaires, qu'ils proposent des relevés très intéressants d'écart par rapport à la norme dont le classement est problématique et à cause de cela, touche de près à la question de la variation en syntaxe, comme nous le verrons quelques lignes plus loin.

Grâce à la bibliographie de ces ouvrages normatifs établie, il y a quelques années, par P. Rezeau (1991)¹³, il sera possible de procéder à un recouplement des inventaires, une fois que ceux-ci auront été reclassés sous les mêmes rubriques, ce qui constituera une base documentaire historique précieuse.

1.2. FRANÇAIS RÉGIONAL ET RÉGIONALISME LINGUISTIQUE

A. Kurt Baldinger (1957)¹⁴ a montré que trois traits, au moins, caractérisent les variétés régionales du français : si on peut y voir des archaïsmes incontestables, on peut y déceler également des traits d'évolution interne ainsi que des innovations par emprunt ; il est intéressant de noter que cette caractérisation est retenue dans l'étude de Brun¹⁵.

Ce qui peut être à l'origine de malentendu ou de difficulté, c'est l'importance relative de chacun des traits déclarés selon le domaine de la langue qu'on considère. Il faut reconnaître à l'étude de Brun le grand mérite d'avoir mesuré, pour chacun des domaines, phonétique, lexique, morphologie et syntaxe, la portée de l'explication par archaïsme, par création et, par interférence avec la langue en contact. Suivons-le dans le bilan qu'il dresse.

Il relève que les éléments de ce français régional peuvent être rangés "sous trois chefs principaux : archaïsmes, emprunts au provençal, innovations spontanées. Quelle est leur importance relative ?" demande-t-il ?¹⁶ Sa réponse est assez nuancée : s'il est tenté de voir un provençalisme dans

12. p. 11. Voir sur ce point Valli (à paraître).

13. Pierre Rézeau, 1991, "Bibliographie des travaux normatifs des XVIII^e et XIX^e siècles", *Parlers et cultures*, 9, Bulletin du GdR 009 "Atlas linguistiques, parlers et cultures des régions de France", pp. 58-64.

14. K. Baldinger, 1957, "Contribution à une histoire des provincialismes dans la langue française", *Revue de Linguistique Romane* 21, pp. 62-92.

15. Voir sur ce point Valli (à paraître).

16. p. 146.

l'usage de *de* partitif, dans l'emploi particulier des pronoms personnels : *nous se sommes rencontrés*, il reconnaît que dans beaucoup de cas, les faits existent en provençal, mais ils ne sont pas étrangers au français.

Il retient ensuite que les divergences les plus notables entre le français officiel et le français régional de Provence portent sur la phonétique et le vocabulaire : incontestablement, on note chez Brun une délimitation géographique des faits qu'il signale. Selon lui, même s'il faut admettre que le français de Provence comporte des variétés, et "qu'il n'est pas exactement pareil à lui-même, de Toulon à Digne et d'Arles à Grasse"¹⁷, il considère que la plupart des tours et expressions se retrouvent des départements du Var, des Bouches-du-Rhône, des Bases-Alpes et du Vaucluse. Tout au plus concède-t-il que le parler de cette capitale méditerranéenne [Marseille] apparaît plus nettement caractérisé que celui de l'arrière pays.

Dans la mesure où ces faits relevés sont géographiquement délimitables, ils répondent à "la définition de l'écart de langage qui constitue un régionalisme linguistique"¹⁸, selon G. Tuailon (1983); mais on ne peut sérieusement parler "d'idiotismes locaux" qu'à propos de variation morphologique dans le lexique, l'emploi de *huile* au masculin, les emplois particuliers des verbes *demeurer*, *rester*, ou de traits de prononciation : par exemple, les prononciations *doteur*, *acent*. Pour les faits de syntaxe, Brun n'a pas dégagé de régionalismes grammaticaux; il a deux formules claires sur ce point que je reprends : "Elles sont moins graves dans la morphologie et dans la syntaxe"¹⁹; il précise quelques lignes plus loin :

En morphologie et en syntaxe, le vulgaire parisien et le français de Provence s'écartent du français officiel sur les mêmes points. Ici comme là, remplacement de *lui* par *y* devant le verbe : *j'y dirai*. mêmes incorrections dans les formes verbales, dans l'utilisation des auxiliaires : *j'ai tombé*, *je m'ai trompé*, [...] même usage de *que* relatif ou ligature. Sans doute, il y a ici, en plus, quelques tournures empruntées au parler provençal. Mais le français des Marseillais ne présente, en morphologie ni en syntaxe aucune innovation proprement dite.²⁰

Enfin, il a cette formule qui donne tout son intérêt à l'étude de cette variété du français :

17. p. 18.

18. G. Tuailon, 1983, "Régionalismes grammaticaux", *Recherches sur le français parlé* 3, Publications de l'Université de Provence, pp. 227-239, 227.

19. p. 144.

20. *Ibid.*

Très souvent, on retrouve dans le français de Provence, le prolongement de certaines tendances linguistiques, qui se faisaient jour dans le français du XVI^e siècle, et qui, bannies par les grammairiens de la société polie, se sont épanouies et perpétuées dans le français populaire : c'est l'histoire de *que* relatif.²¹

On pourrait joindre à cette rubrique l'écart que constitue l'emploi chez de nombreux locuteurs méridionaux âgés d'une "forme forte de l'adjectif possessif", par exemple *ce chapeau est mien*²²; il s'agit selon Brun d'une vieille tournure française qui a été éliminée depuis le XVII^e siècle, mais qui reste vivante à Marseille, au début du siècle, sans doute, ajoute-t-il, parce que la formule existe aussi en provençal, comme l'attestent les observations de Gabrielli et Reynier.

1.3. DONNÉES HISTORIQUES DE RÉFÉRENCE POUR UN INVENTAIRE ACTUALISÉ DES TRAITS CARACTÉRISTIQUES DU FRANÇAIS PARLÉ DE LA RÉGION DE MARSEILLE

L'INVENTAIRE DE GABRIELLI

Cet inventaire présente, à la lecture, un aspect assez hétérogène; toutefois, par un certain nombre de remarques, il est intéressant dans la perspective d'une étude linguistique de la variation syntaxique en français.

a) On peut relever des remarques qui témoignent comme l'a écrit Brun, d'un purisme étriqué. Gabrielli voit une faute dans l'ordre des pronoms personnels compléments à l'impératif dans la phrase, *montre-nous la*, (38) quand la norme exige l'ordre complément direct – complément indirect, *montrez-la nous*. Il en voit une également dans le double marquage de certains compléments; soit pour la construction, *de là j'en conclus*, (48) sans doute une faute assez commune, mais c'est moins sûr dans, *je n'en ai besoin que de la moitié*, où le double marquage n'est que partiel. L'absence du *ne explétif* (71) constitue également une faute à ses yeux, *ce chien a peur que je le batte*, de même que l'absence du respect de la concordance des temps dans les énoncés avec une extraction, *ce sont les espagnols qui arrivaient* (52). Selon lui, le verbe du dispositif devrait apparaître à l'imparfait.

21. p. 146.

22. Brun, p. 50.

b) Gabrielli a relevé des fautes qui sont communes pour un locuteur de notre époque; ainsi, l'absence de la particule de négation *ne*, ou l'accord incorrect du participe du verbe *Faire* avec le complément du verbe à l'infinitif, *cette maison, je l'ai faite bâtir* (58). Dans d'autres cas, la faute ne paraît pas avérée. Que dire du double marquage du sujet dans un énoncé interrogatif comme, *où est-il votre père* (41), ou de la confusion entre deux formes de l'extraction, *C'est à Pierre à qui je parle*, du fait du double marquage du complément prépositionnel, dans la seconde version. On peut s'interroger également à propos de l'emploi jugé incorrect par Gabrielli de l'article défini avec valeur générique, *as-tu pris le café ce matin?*

c) Gabrielli a relevé un assez grand nombre d'écarts qu'il explique par l'interférence avec le provençal. On pourra retenir cet exemple d'extension de la construction dative, *elle se brûla le tablier* (36), que Brun a interprété comme une extension de la voix moyenne en français²³. Autre exemple, l'emploi du possessif singulier dans le cas où le possesseur est pluriel, *mes enfants sont allés chez sa tante* (39), et encore, l'emploi de *être* comme auxiliaire de lui-même, *je suis été mortifié* (51), ou la neutralisation de l'opposition de genre sur le pronom personnel sujet, en anaphore simple, *dites à mes cousines qu'ils me laissent tranquille* (41).

d) On retiendra également un certain nombre d'écarts intéressants, qui ne sont pas présentés comme des provençalismes; certains n'ont pas été repris dans l'étude de Brun. Je retiendrai, par exemple, certains emplois de *que* concurrent d'un relatif décliné, *au ton qu'il parle, l'armoire qu'on met les livres* (45), qui annoncent les faits étudiés par Deulofeu (1981), l'emploi de *que* "connecteur", *j'étais à Tarascon que j'attendais le train* (47)²⁴, l'emploi d'adjectifs dans les syntagmes nominaux dont la tête est constituée par les pronom *celui, celle*, *prenez les bonnes et laissez celles inutiles* (50), l'emploi de la particule négative verbale *pas* devant un adjectif ou un participe passé, *quelles chemises voulez-vous ? les pas repassées* (50), l'emploi du pronom interrogatif "objet direct" avec des verbes intransitifs, *Qu'est-ce que tu pleures ?*, ou avec un verbe transitif suivi de son objet, *Qu'est-ce que tu tires ma robe ?* (79).

23. p. 65.

24. Voir sur ce point l'étude de Deulofeu (1981).

L'INVENTAIRE DE BRUN

Brun exagère sans doute quand il reproche à un auteur comme Gabrielli de n'être "qu'un grammairien propre à redresser les fautes naturelles de [ses compatriotes]"²⁵, et "inspiré d'un purisme étriqué, où survivent les traditions rigoristes du XVII^e siècle". Toutefois, je ne crois pas que les remarques comme celles de Gabrielli portent, comme l'écrit brun, "sur des vétilles"²⁶. Je crois que le reproche que Brun adresse à Gabrielli en particulier, comme son livre en témoigne, c'est de ne pas avoir fait la différence entre fautes courantes et fautes remarquables, caractérisant l'usage de telle ou telle partie d'une population d'une région donnée. On peut soutenir qu'est présente dans l'étude de Brun une conception de l'écart par rapport à la norme du français qui annonce la notion de "faute typante" introduite par C. Blanche-Benveniste (1997), pour désigner des "infractions qui agissent comme des marquages sociaux, que certaines personnes ont appris très jeunes à éviter, sous l'influence des familles et de l'école"²⁷.

La démarche de Brun s'oppose à la tradition des ouvrages normatifs des XVIII^e et XIX^e siècles au moins sur trois points qui touchent successivement au recueil des données, à la délimitation sociologique des groupes d'informateurs, à la réflexion grammaticale qui organise l'inventaire.

Sa méthode d'observation est clairement exposée ; il a écouté, pendant des années, parler les gens, "surtout les femmes", et quant aux exemples, dans la tradition de Damourette et Pichon, "ils ont été notés autour de nous, à la rencontre"²⁸. On peut également retenir qu'il a procédé à des vérifications de représentativité des données recueillies en n'hésitant pas à soumettre ses exemples "à des marseillais que son enquête intéressait"²⁹. Et, s'il a eu recours aux manuels de provençalismes corrigés et autres dictionnaires provençaux, c'est uniquement pour savoir, écrit-il, "si ces constatations ont été faites avant nous".

La détermination qu'il appelle lui-même sociologique des locuteurs provençaux dont il souhaite décrire l'usage constitue, à l'évidence, un trait

25. p. 13.

26. *Ibid.* Il donne une liste de solécismes et de barbarismes qui ont en effet quelque chose de désuet ; par exemple : on ne doit pas dire *faire des chatouilles*, mais *chatouiller* ; on ne doit pas dire *lever le coude*, mais *hausser le coude*. *Matinal* est impropre ; il faut dire *Matineux*.

27. Claire Blanche-Benveniste, 1997, *Approches de la langue parlée en français*, Ophrys, Gap-Paris, p. 44.

28. p. 20.

29. p. 19.

original de l'entreprise. À la différence de Reynier, son étude ne concerne pas l'ensemble des classes sociales. Il écarte de son domaine "les classes instruites, haut négoce, professions libérales, ou mieux, les familles dont les membres, depuis deux ou trois générations ont reçu l'instruction secondaire"³⁰, qui pratiquent le français commun, pour retenir, au contraire, "les autres classes sociales, depuis la bourgeoisie moyenne des boutiquiers, des employés, jusqu'aux matelots, aux poissonnières et aux paysans..."³¹

Une autre originalité de la démarche de Brun est d'avoir cherché à sélectionner les faits de langue à retenir et d'avoir essayé de les présenter en les classant par rubriques grammaticales.

En effet, à la question qu'il pose "quels sont les faits à retenir ?", il apporte des réponses qu'il faut lire avec soin. S'il retient un certain nombre "d'idiotimes locaux", selon ses propres termes, dans le domaine lexical, *fatigué* pour *malade*, *bonnes manières* pour *bons procédés*, il reconnaît que les tournures syntaxiques appartiennent au fonds du français familier, par exemple, des tournures comme *qu'il lui a dit, pareil que lui*³², et il a cette conclusion qui nous intéresse directement : "En morphologie et en syntaxe, le vulgaire parisien et le français de Provence s'écartent du français officiel sur les mêmes points"³³.

Brun innove également dans la présentation de son inventaire. Ses remarques sont réparties selon un plan qui reprend l'organisation en catégories. On peut voir dans ce choix, une commodité, ou aussi bien une mention de la part de l'auteur d'une caractéristique des ces formes d'écarts : en effet, dans nombre de cas, le problème rencontré est lié à la catégorie de la forme : incontestablement, beaucoup des remarques concernent les catégories pronoms, articles, etc. La liste des rubriques constitue un intéressant inventaire des points de grammaire qui donnent lieu à variation. On relève des exemples de grammaire particulière d'une catégorie ; par exemple, l'emploi d'une forme forte du possessif en fonction d'adjectif attribut, qui n'appartient plus au français courant : *cette maison est mienne*³⁴. Les faits de morphosyntaxe verbale répertoriés sont nombreux. On retient, par exemple, des problèmes liés à la catégorie de l'aspect. Le verbe être se conjugue avec lui-même : *je suis été malade*, ou bien on emploie *avoir* au lieu de *être*, avec les verbes de mouvement : *il a*

30. p. 16.

31. *Ibid.*

32. Pour tous ces exemples, voir la p. 14.

33. p. 144.

34. p. 50.

tombé en descendant du tramway. On peut retenir également des exemples nombreux et fréquents de variation dans l'expression de la valence verbale ; les phénomènes d'alternance prépositionnelles sont courants : *il est marié à ma cousine, je l'ai perdu à quelque part*³⁵. À propos de tous ces faits, Brun a ce commentaire :

seules les parties fragiles [de la langue] sont attaquées, celles qui sont d'un maniement délicat, celles où la langue, au cours de son histoire, a hésité, celles où les grammairiens ont multiplié les prescriptions.³⁶

2. L'ENQUÊTE SUR LES PHÉNOMÈNES DE VARIATION DANS LE FRANÇAIS DE MARSEILLE CONTEMPORAIN

Si on a suivi notre développement, on admettra qu'on dispose, avec l'inventaire de Brun (et dans une moindre mesure, celui de Gabrielli), de l'amorce d'une réflexion linguistique à propos de la notion de "faute typante", et de listes des ces "fautes", en tous cas, celles qui sont anciennement installées dans le français parlé de la majorité de la population de la région marseillaise, dans les années 30 : l'ensemble de ces faits constituerait donc, en quelque sorte, un "registre grammatical", au sens où l'entendent Biber et alii (1996).

Dans la perspective d'une étude de la variation en syntaxe du français contemporain conduite dans le cadre d'une région de France, on peut s'interroger, en tout premier lieu, sur le point de savoir dans quelle mesure les mêmes types d'écarts se rencontrent dans l'usage, aujourd'hui, dans les différentes couches de la population.

Pour ce faire, nous disposons d'une base de données qui pourra servir à une investigation conduite selon les techniques d'une linguistique appuyée sur de grands corpus, qui se prête bien, comme le reconnaissent eux-mêmes Biber et alii³⁷, aux études sur la variation dans l'usage. On mesure l'intérêt de telles données, quand on sait que le point de départ de ces études consiste dans "l'identification des modèles syntaxiques réguliers, pour une population donnée et de ceux qui sont rares"³⁸.

Nous avons donc cherché à mesurer, pour servir d'exemple, la fréquence d'emploi d'un petit nombre des "fautes" relevées, en interrogeant

35. p. 69-72.

36. p. 145.

37. D. Biber, S. Conrad & R. Reppen, 1996, "Corpus-Based Investigations of Language Use", *Annual Review of Applied Linguistics* 16, p. 115.

38. *Id.*, p. 116.

le corpus du GARS de 400 000 mots³⁹, constitué de transcriptions d'entretiens avec des locuteurs d'origine géographique et sociale diverses, et de niveau d'instruction variés : le locuteur peut être un technicien originaire de l'est de la France, ou de la région parisienne, ou un syndic patron-pêcheur du quartier Saint-Jean à Marseille, aussi bien.

Comme nous allons le montrer, l'interrogation du corpus a révélé qu'un nombre particulièrement restreint des locuteurs apparaissant dans le "corpus du GARS" manifestaient ces usages, en particulier un petit groupe de locuteurs marseillais, en tous cas dans les conditions d'expression langagière des entretiens réalisés (entre vingt minutes et une heure trente, parfois plus). Un premier enseignement peut être dégagé donc, même s'il faut reconnaître que l'enquête effectuée ne retient pas un nombre important d'entretiens avec des locuteurs marseillais socialement différenciés : l'usage de certains écarts "typants", qui apparaissait majoritaire dans les années 30, selon Brun, apparaîtrait marginalisé aujourd'hui à Marseille et dans sa région⁴⁰.

Mais, si l'interrogation des corpus a révélé un certain nombre d'écarts typants, il faut également noter, comme l'ont observé Biber et alii⁴¹, que certains de ces écarts apparaissent, dans l'usage d'un certain nombre de locuteurs marseillais, en association avec d'autres constructions typantes, identifiant, au moins partiellement, le registre grammatical d'une variété de français qui signale le maintien, à une bien moins grande échelle, de la situation linguistique que décrivait Brun en 1930.

2.1. ATTESTATIONS DE LA PRÉSENCE D'ÉCARTS TYPANTS

Les écarts typants dont nous cherchons à déterminer la distribution dans notre corpus de conversations sont, pour la plupart, des points déjà étudiés et dont j'ai dressé la liste : il s'agit de faits bien connus en histoire de la langue⁴².

39. Il s'agit de la taille de ce corpus en 97. Depuis, ce corpus a atteint le million de mots.

40. Si l'on tient compte du fait qu'un certain nombre de locuteurs du corpus de GARS vivent dans la région Aix-Marseille et souvent en sont originaires.

41. p. 116.

42. Voir Steinmeyer (1979) et Valli (1995).

La première rubrique que nous aborderons est celle réservée, par Brun⁴³, après Gabrielli⁴⁴, à l'emploi de la forme partitive de l'article indéfini : *de, du, de la, de l', des*. Il y a trois observations.

La première concerne l'emploi de l'article indéfini devant les noms "massifs", je cite Gabrielli :

quand vous parlez d'une chose de manière indéfinie, c'est-à-dire, sans en déterminer la qualité ni la quantité, ou dans un sens partitif, faites précéder ce nom par *du, de la, des*, et non par *de*, comme cela n'est que trop ordinaire en Provence, où l'on dit : *je n'ai bu que d'eau à mon dîner, [...], Nous lui portons de raisin et de figue, [...], mets-y d'huile.*

La seconde vise l'emploi du partitif dans le cas où le nom est précédé d'un adjectif. La faute qui est dénoncée, dans ce cas, est l'emploi de la forme *des* au pluriel, au lieu de celui de la forme *de*, qui constitue encore la norme actuelle. Je cite encore Gabrielli : "employez toujours *de* si l'adjectif précède le nom : *j'ai trouvé là d'excellentes personnes* et non pas : *des excellentes personnes*".

La troisième concerne la forme des quantifieurs complexes, positifs et négatifs⁴⁵. La norme, assez largement observée, demande l'emploi des formes *beaucoup de, un peu de, ou, pas de, point de*. En Provence, Gabrielli et Brun notent que l'usage est au contraire l'emploi des mêmes quantifieurs, mais avec une forme longue du partitif : *du, de la, des*, au lieu de *de*. Voici les exemples que donne Brun de cet usage : *il est venu beaucoup des gens, il est tombé beaucoup de l'eau, il y avait assez du monde, bien du monde, tu m'as fait bien de la peine, [...], nous n'avons pas bu du vin, il n'avait pas du pain*⁴⁶.

2.1.1.

Sur le premier point, l'interrogation du corpus⁴⁷ révèle peu de choses. L'emploi du déterminant zéro, en dehors des constructions généralement admises (des locutions figées *avoir crainte que...* et de son emploi archaïque *avoir peine à comprendre*) ne se rencontre que dans de rares exemples ; ainsi avec le nom *peine* :

43. Respectivement aux pages 147 et 51-52.

44. pp. 59-62.

45. Voir Valli (1995) où ce point est étudié un peu plus en détail.

46. À noter que cet usage est signalé dans d'autres régions de France que la Provence.

47 Il faut se souvenir que corpus réunit des transcriptions d'entretiens avec des locuteurs de différentes régions de France.

GENVE 114, 12	mais cette tapisserie on aurait eu de la peine à
GENVE 53, 12	-- la le Suisse en lui-même a de la peine à s'endetter +
FLICIENM 6, 11	+ je regrettais de lui avoir fait de la peine ça se trouverait dans les premiers articles ce que
FLICIENM 6, 6	+ les textes qui lui avaient fait de la peine mais peut-être qu'on pourra le faire ensemble
FGARCIN 92, 13	pas insister / et, / ça fait de la peine des fois moi je connais des animateurs qui persistent à
FEMMES 85, 7	++ poupées puis ça m'avait fait de la peine c'est pas ça c'était rien j'ai dit ça fait rien ++
FEMMES 23, 15	une nuit alors ça m'a fait peine de voir ma mère et ma marraine ++ qui faisaient les habits
CASTAG 3 12, 9	e+ moi j'ai eu de la peine à m'habituer à ce terme de social au début tu sais - cour vous
ARGOTB 3, 9	L1 elle me faisait de la peine la gadji XX
SCA59AB 22, 12	est nous que nous que nous avons eu de peine beaucoup ici les soldats aussi bien sûr + mais
SCA59AB 22, 11	c'est les soldats qui ont eu de la peine tandis que celle de quarante c'est nous que nous
SCA59AB 12, 3	ce pont transbordeur ça m'a fait une peine que vous pouvez pas vous faire une idée
SCA59AB 11, 3	ça m'a fait une peine terrible quand on l'a démolie + X on aurait dit qu'on
PR61AB 12, 3	et l'âne il avait pas beaucoup de peine pour passer vous - pour pour marcher vous comprenez

Comme on peut le voir, c'est l'usage du partitif qui paraît s'imposer ; avec le déterminant *zéro*, on ne relève que l'expression, *faire peine*, peut-être figée, et l'expression *avoir de peine*⁴⁸, qui semble continuer l'emploi que décrivait Brun, à moins qu'on considère que nous avons affaire à une forme disloquée du quantifieur *beaucoup de*. On retiendra que le second locuteur est une femme de pêcheur, fille d'émigré italien, analphabète en français.

2.1.2.

L'examen du corpus confirme le sentiment largement partagé à propos du second point : l'alternance *de/des* des formes de l'indéfini au pluriel dans une suite *verbe + de/des + adjectif + nom* signale une "faute commune" ; nous l'avons testée avec un des adjectifs antéposés le plus fréquent, la forme *grand*.

On peut observer que l'emploi de la tournure normée est bien représenté chez des locuteurs d'origines régionales différentes :

48. Il s'agit d'un locuteur du groupe des Marseillais, précisément.

- SAGAPE 3, 9 + ça m' a pas laissé euh + **de grands** souvenirs L1 ah ouais + L2
non XXX L1 XXX c' est joli
- FLICIENM 14, 14 que quand même une chose qui + a fait **de grands** progrès + les il y
a beaucoup plus de gens
- CORSE 31, 12 on faisait **de grands** feux dans le village on se réunissait + on
faisait il + il
- CHRABI 61, 3 il mérite **de grands** articles c' est - un écrivain L2 -vain X L1 X L2
c' est
- CASTAG4 17 euh on a **de grands** projets pour elle + elle aura + elle sera + elle -
même + h

Mais il est incontestable que l'emploi de la forme *des*, qui constitue une
“faute” en français est plus répandu :

- PUGET-VI 5, 11 + et on voit **des grands** ravins et on monte + surtout avec quinze
vaches c' est tr
- PUG_2VI 63, 1 ++ on vous sert dans **des grands** plats + bien présentés mais ++ c
' est tenu au chaud par d
- PUG_2VI 51, 5 + et après c' est **des grands** qui sont venus + et après nous après
on est reparti à Ruch
- JEANNE 8, 1 euh , eux / sont des **des grands** promeneurs + bien qu' ils
mangent un poulet tous les soir
- GENVE 85, 12 on a **des grands** trusts / de , des / des grands trucs comme U . B .
S . S . B . S L4
- GENVE 85, 11 est des il y a on a on a **des grands** trusts / de , des / des grands
trucs comme U . B . S .
- BIRNIE 8, 14 et malheureusement après ben il y a **des grands** domaines dévastés
+ il y a des propriétés

En revanche, on peut montrer que certains locuteurs appartenant à un
groupe d'informateurs du même quartier de la ville de Marseille, du
même âge, même type d'activité et niveau d'instruction identique, que
nous avons déjà remarqués, n'ont que l'usage qui constitue un écart :

- SCA59AB 43, 1 + oh ça et puis on fait **des grands** plats de de de chose + antipaste
+ nous appelons ça L1
- SCA59AB 25, 11 portai (en) t des tabliers et puis **des grands** châles avec une
grande une pointe derrière + XI
- PEL258B 35, 16 que on on faisait des v- des stores **des grands** stores avec les
grandes franges si vous êtes de

Les mêmes ont également cet emploi avec un autre adjectif, la forme
gros , fréquente en antéposition :

- SCA59AB 68, 12 elle portait pas les sabots elle avait **des gros** souliers + et elle était
comme ça et

FP257B 30, 6 vous savez ils ont **des gros** bâtons de quatre ou cinq mètres eh eh
L2 pour attraper

alors que les deux usages se rencontrent dans le groupe principal. Certains ont l'usage normé :

INFIRM 7, 6 euh / on a **de gros** gros problèmes cliniques + par rapport à leur
sida + on est
HISTOIRE 2, 10 comme il faut au moins ne dis pas **de gros** mots +

Chez d'autres, plus nombreux, on relève la "faute" selon les puristes, dont l'usage tend à se répandre à l'oral comme à l'écrit⁴⁹ :

PUG_2VI 30, 11 i on leur a mis **des gros** ballons à heu + - quatre mètres de
hauteur + ils +
MICROBIO 98, 10 + tu as **des gros** paquets comme si tu avais laissé tomber des
épingles hein da
MAON 71, 6 oui vous aviez **des gros** travaux
MAON 71, 5 oui vous aviez **des gros** travaux
CORPUSME 5, 8 + tout raturé en rouge avec **des gros** traits rouges comme comme un
devoir d ' école + euh
BENOT 48, 14 L1 il y avait un truc **des gros** tamis aussi il me semble qu ' il y avait
non qu ' ils ont fa

La situation est frappante avec l'adjectif *petit*; on note que l'emploi de *de*, dans le contexte sélectionné, ne se rencontre qu'à un seul exemplaire :

BENOT 4, 3 les rondelles sont découpées heu ça fait **de petits** transistors qui
après / ils , y / sont montés et

alors que l'emploi de la forme *des* est relevé à plusieurs reprises. Mais il faut noter que, dans un certain nombre de cas, l'emploi de *des* est normatif : par exemple, quand la suite "adjectif + nom" est un quasi non composé, *petits copains*, *petit bobo*, *petit carreau*, *petit conte*, *petit enfant*, *petit truc*, ou bien quand le nom est suivi d'un complément déterminatif, comme dans tous mes emplois signalés en gras.

VOYAGESO 55, 16 euh il fallait + + préparer **des petits sacs de plastique** + à l '
intérieur desquels tu mettais
VOYAGESO 32, 2 très peu non + alors ce sont encore des petits villages et les gens ont
un contact très facile i
PUGET-VI 38, 16 L1 c ' est des petits bateaux + des petits voiliers non + des bateaux à
voiles + c ' est tout petit

49. M. Grevisse, 1955, *Le bon usage*, 5^e édition, Duculot, Gembloux, § 330, Remarque.

PORTUGA 5 3, 3	c'est-à-dire que bon j' ai eu des petits copains portugais j' ai eu des petits copains français
PHARMACI 48, 12	qui ont l'oeil qui coule euh des petits bobos tu vois L2 des brûlures
FGARCIN 89, 15	+ on a des petits signes qu' on se fait aussi pour que ce soit mettre un
CORSE 20, 13	il y avait - également des magasins des petits magasins + qui offraient des services euh des +
CORPUSAD 13, 11	tu mimes / deux , de , de- / des petits trucs là + il y a + il y a + m- le monsieur qui est
BENOT 43, 7	L1 c' est des petits jeunes de Benot + pour s' amuser ++ alors Bob après il es
BENOT 21, 14	L2 tu y mets des petits carreaux L1 petits carreaux heu L2 aïe aïe aïe //
APOSTROP 150, 12	aimaient le créole et qui écrivaient des petits contes en en créole même euh enfin
APOSTROP 49, 10	enfin je leur lisais des petits contes et puis on ne comprenait pas le petit conte euh ++
ALSACEA 27, 9	L4 non c' est des petits poissons + et puis c' est un genre de petites sardines
ALCOOLIQ 94, 8	mes enfants sont mariés euh j' ai des petits enfants +
ADRIEN 13, 14	tu mimes / deux , de , de- / des petits trucs là + il y a + il y a + m- le monsieur qui est + q

On notera que, dans ce cas, le groupe de locuteurs marseillais donne, au contraire, une série d'exemples d'emploi régulier de *des* en présence d'un complément déterminatif :

SCA59AB 6, 4	et tout le long il y avait tout des petits autels + qui XX tout en marbre tout joli comme tout
SCA59AB 6, 2	+ il y avait des aut- tout des petits autels tout le long il y avait le grand autel / de , ø / f
GUA151AB 39, 8	ça c' est tous des petits cimetières qu' on avait organisés pendant la peste
PR60AB 23, 12	L3 on donne des petits sacs là L2 et nous autres ce papier gris qu' on / nous ,
PR60AB 20, 2	L1 il n' avait pas des petits refrains en Provençal ou L3 non et + vous vous l' avez
PR59B 3, 7	il y a eu des événements à Marseille mais des petits événements de quartier ou pas

2.1.3.

On admettra aisément que les résultats de l'interrogation du corpus, sur le point précédent, n'apportent pas des informations tranchées. Il en va un peu autrement lorsqu'on examine le troisième point.

En effet, la distribution des formes longue et courte des quantifieurs complexes, positifs et négatifs montre une distribution des usages plus marquée selon les groupes de locuteurs.

Étudions d'abord le cas des formes positives ce quantifieur : *beaucoup de (des), pas mal de (des)*. L'emploi de la forme normative *beaucoup de* est massif dans le corpus⁵⁰. En revanche, comme on peut le vérifier dans la liste d'exemples reproduite ci-dessous, l'emploi de la forme *beaucoup des* est marginal⁵¹.

- VOYAGESO 64, 4 + et il y a **beaucoup des gens** qui vivent du tourisme + mais +
mais c ' était
- VOYAGESO 54, 6 + moi j ' ai connu **beaucoup des gens** qui ont attrapé la rougeole
la rubéole des choses
- PUG_2VI 86, 4 heu samedi + il y avait *beaucoup des champignons* véreux parce qu '
il a beaucoup plu + et
- PHARMACI 87, 1 L2 mais mais il y en a beaucoup des tests de grossesse comme ça
vous les vendez
- PHARMACI 48, 9 à la pharmacie de suite euh on voit **beaucoup des choses euh**
mm des conjonctivites
- GENVE 81, 14 été en en ville + où il y a eu *beaucoup des gens* en ville + ils ont mis
un autre XXX

Rappelons la norme puriste⁵² : avec les quantifieurs *beaucoup, assez, pas mal, etc.*, les noms s'introduisent simplement par *de*; mais si ce nom est suivi d'un complément déterminatif ou d'une relative de même fonction, si on exprime une idée de partition d'extraction à partir d'un ensemble, il demande l'emploi de *des*. Tel paraît être le cas dans les énoncés signalés en gras; dans l'énoncé sans mention, *beaucoup* n'est pas quantifieur. Dans les énoncés en italiques, seulement, il semble que l'usage est non normatif, ce qui représente, au total, assez peu de chose.

Si nous considérons l'usage du groupe de locuteurs que nous avons isolé précédemment, nous pouvons remarquer que l'emploi de la forme *beaucoup de* apparaît avec une fréquence importante⁵³, en particulier avec le verbe *avoir* ou *il y a*, mais également avec des verbes comme *savoir* ou *connaître* :

50. On a relevé au moins 250 occurrences de l'emploi de la forme *beaucoup de* introducteur d'un complément de verbe ou d'un sujet.

51. De la petite liste d'exemples relevés, il faut retirer le quatrième exemple, qui est une construction "à double marquage du complément".

52. Pour un exposé grammatical de cette question, je renvoie à l'étude de Skårup (1994).

53. Au moins 40 occurrences, pour un corpus qui représente la dixième partie du corpus GARS.

- SCA59AB 56, 6 il venait une femme aussi qu ' elle avait **beaucoup de** petits alors elle allait aux comestibles
- SCA59AB 54, 7 c ' était Gênois parce qu ' il y avait **beaucoup de** Gênois à Marseille L1 oui +
- SCA59AB 8, 14 qu ' est-ce vous voulez oui je connais **beaucoup de** monde on s ' est connu on s ' est tous
- GUA151AB 7, 4 ça fait que je sais **beaucoup de** cho- mais là Gyptis et Protis je sais qu ' ils étaien

Mais l'emploi de la forme concurrente *beaucoup des* est également important, comme on peut le vérifier dans la liste d'exemples ci-dessous :

- SCA59AB 69, 11 il y avait des calfats il y avait **beaucoup de** dans le quartier **des** calfats ils étaient à la
- SCA59AB 69, 4 parce qu ' il y avait **beaucoup de** l ' eau vous savez elles brossaient le le chose
- SCA59AB 14, 7 y avait les belles femmes + il y avait **beaucoup des** femmes le les Italiens quand ils venaient
- FP257B 15, 11 la mairie vous filez tout droit il y a **beaucoup des** pêcheurs voilà qui peuvent vous renseigner
- *FP257B 15, 3 là vous comprenez je sais pas **beaucoup des** choses L2 mm mm L1 arrête L2 XXX L1 sois sage
- PR55AB 9, 14 - il y a beaucoup **beaucoup des** gens qui l ' ont fait mais enfin ça maintenant ça
- PR61AB 2, 15 L3 il y en avait **beaucoup des** perroquets non L2 il lui avait dit elle est belle

Dans tous ces exemples, sauf dans celui précédé d'une astérisque, l'apparition de la forme *des* ne paraît pas justifiée par le contexte, et constitue de ce fait un écart, observé à six reprises, alors que nous ne l'avons rencontré qu'à deux reprises dans l'ensemble du corpus du GARS, dix fois plus important en volume de transcriptions. On admettra donc que cet écart typant est largement présent dans l'usage du groupe de locuteurs marseillais que nous avons identifié.

Cette première observation paraît confirmée par l'examen de la distribution des formes concurrentes de quantifieurs complexes négatifs. Le relevé des formes concurrentes *pas de/ pas des* montre que la forme régulière est d'une fréquence importante⁵⁴ ; la forme *beaucoup des* est, proportionnellement, plus rare, comme l'atteste la liste suivante, où nous n'avons retenu que les suites *verbe + beaucoup de (des) + nom régime direct*, à l'exclusion des suites avec *être*, qui obéissent à une règle différente :

54. Plus de 200 occurrences pour l'ensemble du corpus du GARS.

- *SAGAPE 12, 10 (1) pas ils (1) avaient **pas des** places L2 comment tu as fait pour / entrer , rentrer /
- *SAGAPE 11, 16 L1 e- il y avait **pas des** Français alors L2 on les a pas beaucoup vus les Français -- n
- PUGET-VI 19, 14 L2 eh bé + ils font **pas des** galas tous les jours non + et + Sheila elle a bien un enfant i
- PORTUGA 9 1, 9 je veux dire des bons à **pas des** bons à rien puisque c ' étaient des gens qui travaillaient ma
- PORTUGA 1 6, 5 que comme généralement j ' en vois **pas des** Portugaises + ni des Portugais + et et
- PHARMACI 97, 4 L2 t ' en vends **pas des** médi- L1 non pas spécialement non de toute façon il y a pas d
- PHARMACI 67, 14 s rapports avec les médecins il y a **pas des** des L1 mm non non L2 des relations euh L1
- PHARMACI 67, 13 on donne quoi voilà L2 et vous avez **pas des** rapports avec les médecins il y a pas des des
- PHARMACI 51, 7 L2 et il y a **pas des** mises en garde sur les affiches publicitaires non
- PHARMACI 14, 14 ' on leur explique au moins il y a **pas des** critères des gens qui
- PHARMACI 7, 6 L2 et vous avez **pas des** consignes L1 on a beau leur ex- on a beau on / en plus on ,
- PHARMACI 3, 11 L2 mais les gens demandent **pas des** renseignements sur euh L1 il y en a qui demandent
- MICROBIO 21, 15 il faut **pas des** trucs trop gros hein ++ pour que la chaleur puisse bien
- MICROBIO 6, 11 il / Ø , n ' / y avait **pas des** aliments périmés avec des microbes enfin ce genre de chose
- HISTOIRE 14, 7 L3 X vous avez **pas des** histoires sur les + L4 sur les Russes + L3 X oui sur les Russes
- HISTOIRE 41, 15 par ex- s ' il n ' y avait **pas des** enfin si chez les gens il n ' y avait pas une certaine peur
- CORPUSG 35, 4 je vous demande **pas pas des** choses de lecture mais de ce qu ' ils disaient les gens
- CONTRA 8, 16 L1 est-ce que ça crée **pas des** situations un petit () peu difficiles vis-à-vis de ces ces
- CASTAG2 25, 14 nous ne leur laissons **pas des** documents en bon état ils pourront retrouver chez notre notai
- BENOT 31, 14 et puis on est revenu -- tu veux **pas des** pins ++ L2 oh des pins L4 pour planter +
- *BEAUMETT 7, 14 + on / ne , Ø / vous raconte **pas des** histoires on vous demande + allez voir vos détenus +
- APOSTROP 128, 16 ils n ' avaient / **pas des** rapports euh comme ont maintenant les enfants avec leurs
- ALSACEBC 20, 16 ils ont **pas des** + hivers aussi durs que nous hein
- ALCOOLIQ 102, 6 ou j ' ai un revolver et ça tire **pas des** trucs comme ça + et des fois

Dans la plupart des exemples de cette liste, qui sont extraits du corpus du GARS, dont on a seulement retranché les entretiens avec le groupe des locuteurs Marseillais, l'emploi de la forme *beaucoup des* est la norme⁵⁵, puisque la négation porte sur l'élément spécifique du nom — épithète ou complément — et non sur le complément lui-même. Dans trois exemples seulement, qui sont signalés par une astérisque, l'emploi de cette forme du quantifieur négatif n'est pas normatif.

Les performances des locuteurs du groupe Marseillais sont un peu différentes. L'emploi de la forme *beaucoup de* est bien représenté, dans les contextes définis ; mais surtout, il faut relever que l'emploi de la forme concurrente est plus important que dans l'ensemble du corpus du GARS ; dans la liste ci-dessous, nous avons signalé les “fautes” par une astérisque ; les autres exemples sont conformes à la norme⁵⁶.

- *SCA59AB 79, 13 L2 parce que on faisait **pas des** cadeaux comme maintenant comme on parlait pas des
- GUA151AB 35, 12 bon mais moi je vous demande pas **pas des** choses de lecture mais de ce qu' ils disaient
- *FP257B 26, 16 il y avait **pas des** jours où ils ne sortaient pas L1 non non
- FP257B 25, 4 est ce qu' il y avait **pas des** trad- enfin des traditions oui des espèces de traditions
- *FP257B 19, 7 est ce qu' ils n' avaient **pas des** des droits euh des privilèges ces ces pêcheurs enfin
- *FP257B 18, 13 L2 et euh il existait **pas des** prud' hommes L1 ah ben oui ça les prud' hommes toujours
- FP257B 15, 2 sur les pêcheurs il y avait **pas des** personnages euh un peu qui ressortaient euh
- FP257B 9, 1 L2 il y avait **pas des** histoires qui se racontaient L1 non non ça euh ça s'est p
- DUR58B 18, 13 L3 il y avait **pas des** des chansons ou des des slogans qui qui étaient repris L1
- DUR58B 17, 8 L3 autrement il y avait **pas des** chansons qui étaient euh créées dans le quartier par exem
- *PR61AB 13, 5 les joueurs de pétanque il y avait **pas des** histoires L1
- *PR61AB 5, 16 L2 il y avait **pas des** cabinets L1 et il y avait pas de cabinets et alors chacun
- *PR60AB 31, 3 on voyait **pas des** curés dans dans les rues à cette époque là et de notre temps
- PR60AB 20, 2 L1 il n' avait **pas des** petits refrains en Provençal ou
- *PR60AB 4, 5 il fallait les acheter on n' avait **pas des** sous ah euh +

55. Voir sur ce point la synthèse de Skårup (1994).

56. Pour la lisibilité de la liste, nous avons éliminé les exemples d'emploi de *pas des* avec le verbe *être* et les verbes *intransitifs en de*.

PR60AB 17, 2	L1 voilà L2 mais il y a pas des chansons qui sont restées
*PR6-B 12, 12 ai j' ai pas	+ L2 et sur la mer il y avait pas des superstitions _ _ L3 non non j'
PR6-B 4, 7 contines to	il y avait pas des chansonnettes des des petites euh contines L1 des
*PR59B 3, 8 l- l- la sépa	on ne voyait pas des curés dans la rue le moins possible parce que
PR59B 3, 7 mais des petits	mais il y avait pas de il y avait pas eu des événements à Marseille
MMEDURAN 18, 3 repris L1 non	il y avait pas des des chansons ou des des slogans qui qui étaient
MMEDURAN	L3 autrement il y avait pas des chansons qui étaient euh créées dans le quartier par exemple

On admettra assez facilement que, dans l'absolu, l'écart par rapport à la norme est plus fréquent dans ce groupe de locuteurs⁵⁷, proportionnellement plus important que dans l'ensemble des locuteurs du corpus du GARS.

2.2. ASSOCIATION D'ÉCARTS TYPANTS

Le développement qui précède, appuyé sur l'examen de quelques "fautes" classiques, paraît apporter l'indication selon laquelle le caractère "typant" des écarts par rapports aux normes courantes du groupe de locuteurs marseillais peu instruits ne se manifeste pas seulement dans la production de "fautes" proprement dites dont le statut n'est pas toujours aisé à déterminer, présentes également dans l'usage d'autres locuteurs du corpus, mais dans leur bien plus grande fréquence, chez ces locuteurs, dans le même genre de production linguistique : entretien avec un étudiant qui enquête sur la vie sociale dans un quartier de Marseille, entre les deux guerres.

Il s'agit là d'une indication d'une portée intéressante, si on se rappelle que Brun définissait en partie par ce trait, précisément, la particularité de la morphosyntaxe du français de Marseille à l'époque de son enquête.

C'est à une seconde indication de même portée, également révélée par l'interrogation du corpus que je voudrais consacrer ce second chapitre. L'examen du corpus révèle, en effet, dans l'usage du groupe de locuteurs marseillais peu instruits que nous avons identifié, la manifestation d'écarts "typants" qui sont absents dans l'usage des autres locuteurs du corpus.

57. Au moins 7 exemples.

2.2.1.

Le premier exemple que nous étudierons est l'écart concernant l'emploi des auxiliaires : l'emploi de *avoir* qui se substitue à celui de *être* devant les verbes dits de "mouvement", plus le verbe *rester*, est une question de linguistique historique liée à l'aspect verbal qui connaît une littérature importante⁵⁸. Brun donne des exemples de cet emploi en français de Provence, selon lui "conforme à celui du français populaire"⁵⁹ : *j'ai demeuré plusieurs jours chez lui, il a tombé en descendant du tramway.*

L'emploi, apparemment exclusif, de l'auxiliaire *avoir* avec les verbes de mouvement s'observe chez deux locutrices dans le groupe de marseillais, dans la construction intransitive des verbes *tomber* et *remonter* :

- PR6-B 13, 7 et il a remonté à bord l- la bombe a tombé dans le dans la cheminée
du bateau et il a été pulvérisé
- PR6-B 1, 16 nous autres ici des flammèches ils ont tombé de mais de côté quoi c '
était des flammèches
- PR54AB 5, 6 il a pris l ' ascenseur et il a monté au commerce là bas il s ' est plus
fait voir tant que les
- FR 54A alors il a monté à Paris - il a continué à faire ce fameux métier
- FR 54A quand j'ai passé la rue de la Renarde- et que j'ai monté là - je
tremblais de tous mes membres -

Chez les autres membres du groupe, l'usage de *avoir* est observé largement avec le verbe *rester*, en concurrence⁶⁰ avec celui de l'auxiliaire *être* :

- SCA59AB 72, 8 et il y avait + un type qui est resté au moins cinquante ans (Camelio)
+
- SCA59AB 39, 9 + nous avons resté quinze jours + chez ma cousine des cousins à
moi que nous a
- SCA59AB 19, 13 il était pas naturalisé rien mais il est resté ici + il était dans la rue
Bernard Deberre il avait
- SCA59AB 3, 5 + alors eh j ' ai resté comme ça + et comme + ça allait pas + ça allait
mal +
- GUA151AB 47, 2 ils en sont sauvés mais + ils ont resté bien entendu marqué
- GUA151AB 45, 7 la la variole il a resté une épidémie pendant plus de trois ou quatre
mois qu ' il a

58. En plus de l'ouvrage de Steimeyer, déjà cité, il faut retenir les articles de Sankoff et Thibault (1977) et de C. Blanche-Benveniste (1977).

59. p. 62. Voir également Valli (1995).

60. À première vue, cette concurrence dans l'emploi des auxiliaires ne paraît pas conditionnées sémantiquement selon l'aspect verbal : on ne retrouve pas, dans ces exemples, l'opposition du français classique entre l'indication d'un état permanent et celle d'un état transitoire.

- GUA151AB 42, 6 est ce que ce ce tableau a resté + on le sait pas L2 qu ' est ce tu voulais dire Nicole
- GUA151AB 31, 4 ils sont partis à la cathédrale et ils **ont** resté à la cathédrale qu ' ils jouaient la pastorale
- GUA151AB 12, 10 il y en a un qui a resté comme je vous dis aux Saintes Maries pour ça qu ' on fait
- FP257B 22, 5 où ils ont trouvé des maisons ils **ont** resté là bas vous comprenez
- FP257B 21, 11 ça moi je vous dis j '**ai** pas beaucoup resté parce que ma soeur elle m ' a L2 hébergée
- FP257B 2, 4 c ' **est** resté comme ça L1 ah non c ' était cet endroit ça a pas été touché
- FP257B 1, 3 alors j ' **ai** pas beaucoup resté voyez L2 alors vous êtes restée combien de temps pourtant

On notera avec intérêt que cette alternance dans l'emploi des auxiliaires avec les verbes de mouvement plus *rester* n'appartient pas à l'usage des autres locuteurs du corpus du GARS.

2.2.2.

Brun a retenu un écart important, en français de Marseille, dans le choix du modifieur attaché à l'adjectif⁶¹ ; il s'agit de l'emploi de *beaucoup*, modifieur attaché au verbe en français normé courant, attaché à l'adjectif ou au participe passé attribut. Brun cite comme exemple⁶² : *il fait beaucoup chaud*.

Le corpus du GARS révèle presque un milliers d'exemples d'emplois de la forme *beaucoup*, qui peut être, comme dans ces exemples empruntés à deux des locuteurs marseillais, adverbe modifieur d'intensité verbale :

- PR60AB 14, 5 + parce que moi mon père nous a beaucoup appris à aimer le chant
- PR61AB 13, 4 quelqu ' un qui avait beaucoup perdu alors ils déchiraient les billets de vingt sous da
- PR54AB 9, 13 ça se faisait beaucoup L3 oh oui certainement ça à Marseille ça se faisait je crois

ou bien, modifieur d'adverbe comparatif :

- PR60AB 23, 7 d ' abord il doit être beaucoup plus cher et que ces sacs L1 oui L2 en papier L3 on n '

ou encore, élément d'une forme complexe de quantifieur, standard ou non standard :

61. Je reprends ici un point développé dans une communication précédente (Valli, 1997).

62. p. 75.

- PR55AB 9, 14 - ça il y a beaucoup beaucoup des gens qui l'ont fait mais enfin ça maintenant ça ap
 PR55AB 3, 5 il y a beaucoup de charcutiers c'est tous des des gavots presque hein

ou pronom, occupant une place de complémentation verbale :

- PR60AB 34, 8 ils ils sont beaucoup L1 là + vraiment la curaille euh elle + elle tenait pas

enfin, plus curieusement, concurrent de l'adverbe *bien*, dans la fonction de modifieur d'adverbe temporel :

- PR60AB 10, 5 c'est avant L2 c'était beaucoup avant à Marseille qu'il y a eu Belsunce qu'il y a eu

Mais surtout, comme le révèle la liste suivante, *beaucoup* concurrence, dans de nombreux exemples, l'usage de *très*, modifieur adjectival. Cet emploi ne s'observe que dans le groupe de locuteurs marseillais, avec une fréquence assez importante :

- PR6-B 13, 9 parce que mon mari il était **beaucoup humain** hein alors il les menait les autres après qué dis
 PR6-B 1, 11 c'était socialiste c'était **beaucoup socialiste** quand même à cause c'est à dire que le maire
 PR60AB 33, 16 ah mais eux ils sont **beaucoup pieux** les Bretons L1 oui L2 oui c'est pour ça mais xxx
 SCA59AB 83, 9 + parce que mon père il était **beaucoup rigide** on dansait on s'amusait vous savez c'était
 SCA59AB 16, 8 on était **beaucoup connu** aussi parce que mon mari il était maître-maçon
 PEL258B 10, 5 vous en voyez mais c'est **beaucoup changé** L2 oui X L1 le Panier n'est plus ce qu'il
 GUA151AB 45, 5 mais la variole à l'époque il est **beaucoup mort** L1 puis alors il est + il en est mort tant
 GUA151AB 10, 8 parce qu'on commençait **beaucoup bonne heure** alors on allait faire la visite à Saint Vict
 FP257B 24, 16 il était **beaucoup estimé** voyez beaucoup estimé L2 et c'est que quand il y avait une élection
 FP257B 9, 14 je suis née en dix mais ça s'est passé **beaucoup avant** voyez alors euh XXX

Il faut noter, à ce propos, que les mêmes locuteurs ont également l'emploi de la forme *très* dans cet emploi, ainsi que dans l'emploi modifieur adverbial :

- PR60AB 45, 11 c'était **très mauvais** d'abord Georget il s'est tué à l'Amer Picon xx

- PR60AB 12, 1 les opérettes les il y avait de **très très bons chanteurs** dans le temps dans les opérettes
- PR60AB 11, 3 c ' était pas des imbéciles ils avaient **très bien compris** ce qui se passait et
- PR6-B 7, 7 vous comprenez ils savaient **très bien** à q- à qui ils s ' adressaient vaï ils étaient pas
- PR6-B 1, 15 / il restait , il est resté / l- d- **très longtemps** et il a perdu son chose en trente huit
- PR59B 3, 8 c ' est qu ' on voyait **très peu** de curés dans les rues et quand on voyait un curé
- PR59B 2, 6 sa femme elle avait été **très malade** elle avait eu un cancer au sein il était obligé de

2.2.3.

J'ai fait allusion plus haut à un écart typant signalé par Brun, après Gabrielli : il s'agit de l'emploi fréquent chez de nombreux locuteurs méridionaux âgés d'une "forme forte de l'adjectif possessif", par exemple *ce chapeau est mien*⁶³, vieille tournure française demeurée vivante à Marseille, au début du siècle.

L'interrogation du corpus du GARS nous révèle que cette construction d'un adjectif possessif n'est présente qu'à un unique exemplaire, précisément chez un des locuteurs marseillais du quartier Saint-Jean — le Vieux-Port de Marseille. Voici cet exemple :

- PR61AB 14, 9 et ton bateau eh il est **pas mien le bateau** L2 eh ben après il en a acheté deux alors

3. CONCLUSION

Comme on le voit, l'approche linguistique du phénomène de variation dans les usages du français conduite sur des corpus, relativement importants, de textes transcrivant des entretiens informels avec des locuteurs appartenant à tous les milieux de la société de la région de Marseille et d'ailleurs en France permet de dégager deux points intéressants : distinguer des traits de construction qui sont d'une fréquence rare de ceux qui se montrent d'une fréquence élevée; montrer ensuite qu'un certain nombre de ces traits, qui apparaissent associés soit au sein d'un genre de discours soit dans l'usage de certains groupes de locuteurs définis socialement constituent la base d'un registre linguistique.

63. Brun, p. 50.

Les observations de Biber et alii (1996) qui mettent l'accent sur la notion de fréquence des usages rejoignent celles de Brun : la morphosyntaxe du français de Marseille dans les années 30 ne se définissait par l'usage de traits de construction spécifiques, mais au contraire, par la fréquence élevée de traits communs aux usages populaires du français. La même considération sur les différences de fréquence dans l'usage des traits communs à différents groupes de locuteurs dans le corpus du GARS nous permet de d'identifier un groupe de locuteurs marseillais.

Il en va de même de leurs observations à propos de la notion de "registre" linguistique. Cette notion était centrale dans l'étude de Brun : le français parlé de la plus grande partie de la population de Marseille se caractérisait par l'association d'une certain nombre d'écarts typants du français populaire d'un usage fréquent et étendu à toutes les couches de la population qui n'avaient pas bénéficié de l'instruction secondaire.

Nous retrouvons, pour partie, une association de même nature sans l'usage du petit groupe de nos locuteurs marseillais.

Sur la base de cette étude qui ne constitue qu'un sondage qu'il faudra prolonger, en étendant l'enquête à d'autres traits de constructions, à d'autres groupes de locuteurs par l'enrichissement de notre corpus de référence et son extension systématique à d'autres régions que celle de Marseille, nous avons tenté d'apporter une contribution à l'étude du français régional de Provence qui s'inscrit dans le prolongement de l'étude d'A. Brun.

André VALLI
Université de Provence



Références bibliographiques

- BAGGIONI, D. [1993], "Historique du concept de français régional", *Travaux et Documents* 3, *La lexicographie variationniste en situation de contact*, C. Bavoux et M. Beniamino (éds), Publication de l'Université de la Réunion, pp. 17-39.
- BALDINGER, K. [1957], "Contribution à une histoire des provincialismes dans la langue française", *Revue de Linguistique Romane* 21, pp. 62-92.
- BENIAK, E. et MOUGEON, R. [1989], "Recherches sociolinguistiques sur la variabilité en français ontarien", *Le français canadien parlé hors du Québec*, Les Presses de l'Université Laval, pp. 53-74.

- BIBER, D., CONRAD, S. & REPPEN, R. [1996], "Corpus-based Investigations of Language Use", *Annual Review of Applied Linguistics* 16, p. 115.
- BLANCHE-BENVENISTE, C. [1997], *Approches de la langue parlée en français*, Ophrys, Gap-Paris.
- [1977], "L'un chasse l'autre. le domaine des auxiliaires", *Recherches sur le français parlé* 1, pp. 100-148.
- [1997], "La notion de variation syntaxique dans la langue parlée", *Langue française* 115. *La variation en syntaxe*.
- BRANCA-ROSOFF, S. [1987], "Matériaux pour une histoire des conceptions normatives dans la représentation de la langue XVIII^e-XIX^e siècles", Thèse de l'Université Paris VIII, Dir. J.-C. Chevalier.
- BRUN, A. [1930], *Le français de Marseille. Étude de parler régional*, Institut historique de Provence, Marseille.
- CHAMBON, J.-P. [1996], "Emprunts du français aux dialectes ou patois : une illusion d'optique ?", *Lalies* 17, *Actes des sessions de linguistique et de littérature*, Aussois, 2-7 septembre 1996, Presses de l'ENS, Paris, pp. 33-53.
- DESGROUAIS [1766], *Les gasconismes corrigés. Ouvrage utile à toutes les personnes qui veulent parler et écrire correctement, et principalement aux jeunes gens, dont l'éducation n'est point encore formée*. J-J Robert, Toulouse.
- DEULOFEU, J. [1981], "Perspective linguistique et sociolinguistique dans l'étude des relatives en français", *Recherches sur le français parlé* 3, pp. 135-195.
- GABRIELLI, C. de [1836], *Manuel de Provençal ou les Provençalismes corrigés à l'usage des habitants des Bouches-du-Rhône, du Var, des Basses Alpes, du Vaucluse et du Gard*, Aubin, et Camoin et Masvert, Marseille.
- GREVISSE, M. [1955], *Le bon usage*, 5^e édition, Duculot, Gembloux, § 330, Remarque.
- REYNIER [1829], *Corrections raisonnées des fautes de langage et de prononciation qui se commettent même au sein de la bonne société dans la Provence et quelques autres provinces du Midi*, Marseille.
- RÉZEAU, P. [1991], "Bibliographie des travaux normatifs des XVIII^e et XIX^e siècles", *Parlers et cultures*, 9, Bulletin du GdR 009, "Atlas linguistiques, parlers et cultures des régions de France", pp. 58-64.
- SANKOFF, G., et THIBAUT, P. [1977], "L'alternance entre les auxiliaires avoir et être en français parlé de Montréal", *Langue française* 45, pp. 43-56.
- SÉGUY, J. [1951], *Le français parlé à Toulouse*, Privat, Toulouse. p. 101.
- STEINMEYER, G. [1979], *Historische Aspekt des français avancé*, Droz, Genève.
- TUAILLON, G. [1983], "Régionalismes grammaticaux", *Recherches sur le français parlé* 3, Publication de l'Université de Provence, pp. 227-239.
- [1983], *Matériaux pour l'étude des régionalismes du français 1. Les régionalismes parlés à Vouvray*, Klincksieck, Paris.
- VALLI, A. [1993], "La problématique d'une enquête sur la variation en syntaxe dans le français parlé à Marseille", *Actes des Premières journées scientifiques du Réseau thématique de recherche, "Étude du français en francophonie"*, Université de Nice,

18-21 septembre 1991, *Inventaire des usages en francophonie : nomenclatures et méthodes*, John Libbey Eurotext, Paris-Londres, pp. 395-407.

- [1995], "Notes sur la variation linguistique en français", *Recherches sur le français parlé*, 13. pp. 91-109.
- [1997], "Quelques réflexions à propos de l'ouvrage de Brun, *Le français de Marseille. Étude de parler régional*", *Actes du colloque "De François Raynouard à Auguste Brun. La contribution des méridionaux aux premières études de linguistique romane"*, *Lenga*, 42. pp. 215-227.

La conjugaison des verbes : virtuelle, attestée, défective

Qu'est-ce que conjuguer un verbe ? — C'est le réciter avec toutes les différences dont il est susceptible, et dont nous rendrons compte dans la suite. (Restaut, 1745, p. 176).

Il nous est difficile de nous représenter la conjugaison des verbes du français autrement que par la représentation qu'en donnent les tableaux de conjugaison scolaires. Ces tableaux ont les vertus unifiantes des modèles grammaticaux. Dans les manuels de grammaire, les verbes qui servent de “modèles”, fléchis en modes, temps et personnes, sont disposés en autant de “lignes” qui s'inscrivent dans autant de “cases” composant les “tableaux”. Le compte peut varier selon ce qu'on entend par “conjugaison”. On y mettait naguère des conjugaisons “passives”, “pronominales”, voire “interrogatives” et “négatives”, qui ont généralement disparu des manuels récents. Dans *La Conjugaison pour tous* éditée en 1997 par Bescherelle, qui répertorie 12 000 verbes, les tableaux ont 8 cases à 6 lignes pour l'indicatif, 4 pour le subjonctif, 2 pour le conditionnel et 1 pour l'impératif ; le participe passé a 3 lignes, l'infinitif et le participe présent en ont 2. Au total : 97 lignes, qui représentent autant de formes verbales “modèles”. Le principe implique que tous les verbes du français peuvent se conjuguer selon ces 97 possibilités, à l'exception des “verbes défectifs”, présentés comme fondamentalement “anormaux”. Il est bien entendu que ces tableaux fournissent une conjugaison “virtuelle” et que cet ensemble de formes n'est pas attesté pour tous les verbes, ni en français écrit ni en français parlé.

Il nous a paru intéressant de chercher à connaître l'ampleur de la conjugaison réellement attestée en français parlé contemporain, à partir des corpus enregistrés. Nous avons été étonnés par les résultats : pour une grande partie des verbes utilisés, la conjugaison est extrêmement réduite, au point que cela mérite explication. Une même tendance à utiliser un nombre réduit de formes verbales a été constaté pour l'anglais, dans de nombreuses études récentes (G. Kennedy 1998 : 122). Cette réduction des capacités virtuelles des verbes, apparemment semblable dans ses résultats à celle des verbes défectifs, en est cependant fondamentalement différente dans ses principes. Elle semble pouvoir s'expliquer en partie par des raisons de fréquence d'emploi, de sémantisme ou de valeurs énonciatives. Mais l'essentiel en reste assez mystérieux.

Pour décrire cette "conjugaison effectivement attestée", nous avons défini certaines méthodes de classements morphologiques et nous avons développé un outil informatique d'interrogation des données orales¹. Pour interpréter les résultats, nous avons utilisé les observations des grammairiens et nous avons formulé quelques hypothèses d'explication. Il nous a semblé que les modèles de représentation de ces formes verbales avaient évolué au cours de l'histoire et qu'il serait intéressant d'en rendre compte, pour mieux comprendre les représentations actuelles.

1. DESCRIPTION

1.1. MÉTHODES DE DESCRIPTION

1.1.1. FORMES RETENUES ET FORMES ÉCARTÉES

Pour des raisons de méthode, nous ne retenons pas, dans la description proprement morphologique, les formes auxiliées qui figurent habituellement dans les tableaux. Les "temps composés", comme :

j'ai su, tu as su, j'avais su, j'aurai su, j'aurais su, que j'aie su, avoir
su,

sont considérés comme des combinaisons entre l'auxiliaire *avoir* (ou *être*) et la forme de participe passé, *SU*. Comme toutes les formes simples ont systématiquement un correspondant auxilié, la forme simple et l'auxiliaire

1. Jean-Pierre Adam, consultant pour le traitement informatique auprès de l'équipe CNRS de *CORPUS* à Aix, est responsable de la partie de recherche informatisée.

recevant les mêmes temps, modes et personnes, les combinaisons sont entièrement prévisibles² :

- 1^{re} personne du présent de l'indicatif : je sais / j'ai (su)
- 2^e personne du futur de l'indicatif : tu sauras / tu auras (su)
- 3^e personne du pluriel du subjonctif : qu'ils sachent / qu'ils aient (su)
- infinitif : savoir / avoir (su)
- participe présent : sachant / ayant (su)

Le nombre de temps et modes en sera donc diminué de beaucoup par rapport aux tableaux scolaires.

Pour des raisons d'usage, nous avons hésité à retenir les formes d'imparfait et de plus-que-parfait du subjonctif. Il faudrait certainement les retenir pour l'écrit, où l'on en rencontre beaucoup plus qu'il ne semblerait à première vue (cf. Barral 1980). Mais ces formes sont vraiment très rares dans la langue parlée actuelle, comme le notait déjà Martinon en 1927 :

- "On trouverait un très grand nombre de gens fort instruits qui n'emploient jamais ni un imparfait ni un plus-que-parfait du subjonctif, tout en connaissant parfaitement l'usage de ces deux temps" (p. 382).

Les imparfaits du subjonctif des verbes les plus fréquents, *qu'il le fît, qu'il le fût, qu'il l'eût dit, dussè-je*, ont peut-être conservé quelque chance d'apparaître dans le français parlé des conversations. On peut sans doute en rencontrer dans certaines formes de discours publics, bien que nous n'en n'ayons pas jusqu'ici d'exemples dans le corpus. Nous les avons, pour l'instant, éliminés de l'enquête, tout en conservant la possibilité technique de les recenser dans des tâches ultérieures.

Il n'en va pas de même pour le passé simple, encore bien utilisé aux formes de troisième et sixième personne : *on partit, ils partirent*, et même à la quatrième personne dans certaines prises de parole très appliquées, *nous partîmes*. Nous conservons donc intégralement le passé simple.

Nous avons dû écarter des formes, fréquemment utilisées en français parlé comme "phatiques", qui ont l'apparence de verbes mais n'en sont plus :

- n'est-ce pas, va, allons, allez, dis donc, dites donc, tu vois, voyons, voyez, tiens, tenez

dans des emplois comme :

2. Pour simplifier la tâche, nous ne mentionnons pas ici les formes surcomposées à deux auxiliaires, *je les ai eu sues par cœur*, ni les formes passives, *cela a été su*, pour lesquelles le raisonnement morphologique serait similaire.

- allons bon voyons te fâche pas (Cl 90,6,a)

Nous avons dû nous référer régulièrement aux contextes pour vérifier ces emplois et les écarter. Certaines décisions sont délicates : il est souvent difficile de savoir si *vous savez* est un emploi de verbe plein ou un emploi de phatique.

1.1.2. DÉSIGNATION DES FORMES CONJUGUÉES

Pour désigner les différentes formes prises par les verbes, sans avoir à transporter une terminologie trop lourde, nous utilisons un système de repérage plutôt qu'un vocabulaire grammatical. Nous prenons comme repères les formes du verbe *savoir* qui présente l'avantage d'offrir un maximum de différenciations morphologiques. C'est en effet un de ceux qui marquent nettement la différence entre indicatif et subjonctif présents :

- je sais, nous savons, vous savez
- que je sache, que nous sachions, que vous sachiez

et le seul qui ait une morphologie spéciale pour l'impératif, qui ne soit pas celle de l'indicatif (*savez*), ni du subjonctif (*sachiez*) :

- sache, sachons, sachez

Damourette et Pichon utilisaient ainsi les paradigmes du verbe *savoir*, qu'ils nommaient des " tiroirs ", pour classer toutes les formes de tous les verbes :

- le tiroir *savez*, le tiroir *sachiez*, le tiroir *saura*, etc.

Le système a été exploité par R.L. Wagner dans sa description des verbes de l'ancien français (1974 : 119-156). Nous ne reprenons toutefois pas le terme de " tiroir ". Pour désigner chacune des réalisations virtuellement disponibles, nous emploierons le terme de *poste de conjugaison*. Ce terme est moins ambigu que celui de *forme*, qui peut, sous une même apparence, correspondre à différentes unités morphologiques homonymes. Par exemple, une même forme *sais* représente deux postes de conjugaison distincts dans *je sais* et *tu sais*. Pour la commodité du repérage et du calcul informatique, chaque poste de conjugaison sera affecté d'un numéro.

1.1.3. CLASSEMENT

Une fois écartées les formes auxiliées, le classement est assez simple et il se retrouve, à peu près identique, chez de nombreux auteurs (cf. G. Guillaume 1929, H. Bonnard 1968, Curat 1991, Riegel 1994, Wilmet 1997). On compte six postes de conjugaison pour l'infinitif et les participes

(dont quatre pour les genres et nombres du participe passé); six postes pour le subjonctif présent (en considérant qu'il n'y a pas d'autre subjonctif, sinon celui qui est formé avec un auxiliaire); six postes pour chacun des cinq temps de l'indicatif et trois postes pour l'impératif (qui ne représente des formes distinctes que pour le verbe *savoir*) :

infinitif et participes	1 savoir	2 sachant	3 su	4 sue	5 sus	6 sues
subjonctif	7 sache	8 saches	9 sache	10 sachions	11 sachiez	12 sachent
indicatif						
présent	13 sais	14 sais	15 sait	16 savons	17 savez	18 savent
imparfait	19 savais	20 savais	21 savait	22 savions	23 saviez	24 savaient
passé simple	25 sus	26 sus	27 sut	28 sûmes	29 sûtes	30 surent
conditionnel	31 saurais	32 saurais	33 saurait	34 saurions	35 sauriez	36 sauraient
futur	37 saurai	38 sauras	39 saura	40 saurons	41 saurez	42 sauront
impératif		43 sache		44 sachons	45 sachez	

Le total affiché dans ce tableau est de 45 postes de conjugaison virtuellement disponibles pour les verbes français. C'est une sorte de moyenne ; en ôtant l'impératif, on en compterait 42 ; en comptant l'imparfait du subjonctif, on arriverait à 51³.

CALCUL DE L'UTILISATION EFFECTIVE

Un verbe qui serait effectivement attesté dans ces 45 postes de conjugaison utiliserait donc toutes ses virtualités. Un verbe qui serait attesté dans 18 postes de conjugaison sur 45 possibles serait un verbe utilisant 40 % de sa morphologie virtuelle. Avec 9 postes attestés, ce serait 20 % ; avec 5, environ 11 % et avec 3, à peu près 6 %.

Ces calculs n'ont évidemment qu'une valeur indicative. Nous les utiliserons, avec précaution, pour comparer les résultats.

3. En comptant les combinaisons de formes entre elles, on arriverait à un total plus important. Pour l'anglais, les calculs usuels, tenant compte des combinaisons, font généralement état de 224 formes verbales théoriques, dont 145 ne figurent jamais dans les corpus (Kennedy 1998, 122-3).

1.2. LA RECHERCHE INFORMATISÉE

1.2.1. LE CORPUS

Le corpus utilisé pour le dépouillement des formes verbales est une sous-partie du corpus du GARS. Il comporte environ 700.000 mots. Nous y avons recensé environ 580 verbes différents. Le dictionnaire de l'ensemble des formes verbales est donc : $580 \times 45 = 26.000$ formes. En y mettant le subjonctif imparfait, cela ferait 29 580. Ce dictionnaire sert au recensement et aux comptages des emplois des verbes dans le corpus et aussi à des extractions par le concordancier XCOR, en vue d'étude sur d'autres problèmes concernant les verbes (par exemple le calcul des valences verbales).

Le tableau suivant donne une indication sur la fréquence d'occurrences de ces verbes.

Nombre d'occurrences	0 à 9	10 à 49	50 à 99	100 à 499	500 à 999	+ de 1000
Nombre de verbes	143	240	68	83	16	16

1.2.2. REPRÉSENTATION DE LA CONJUGAISON

Nous n'avons pas intégré un programme construit sur la plus ou moins grande régularité des règles de conjugaison. Comme on le sait, les répartitions classiques en deux, trois ou quatre groupes de conjugaison ne suffisent pas à absorber tous les détails des réalisations dont le programme de reconnaissance doit tenir compte. Le nombre de sous-groupes à dégager en fonction des "irrégularités" est trop grand. Par exemple, pour les verbes du troisième groupe, le classement de Besche-relle (1997) comporte 65 sous-groupes, allant du sous-groupe 23 (constitué par le seul verbe *aller*) au sous-groupe 88 (*cuire, conduire, nuire*, etc.), qui en comporte 29. Cela fait 65 sous-groupes pour classer 326 verbes, soit environ 5 verbes par sous-groupe. À ce compte, il est plus économique de copier les listes de formes verbales plutôt que de programmer les classements par règles.

Nous avons utilisé une représentation inspirée des tableaux classiques des manuels scolaires, à partir de laquelle la programmation revient simplement à identifier le modèle de conjugaison d'après l'infinitif et à accoler au radical du verbe les différents éléments trouvés dans les tables pour créer la forme conjuguée. Cela ne correspond pas à un modèle idéal

d'analyse morphologique, tel que nous le proposerions pour rendre compte des règles de fonctionnement de la morphologie écrite ou orale du français (Cf. Eynde et Swiggers 1987). Mais c'est une représentation pratique pour reconnaître et analyser les formes verbales du corpus.

Voici la description des verbes du tableau 58 de Bescherelle (1997), qui comporte 50 verbes du type *défendre*, *descendre*, *rendre*, *tordre*, etc. Le tableau est décrit par un prédicat Prolog à 4 arguments :

- arg 1 : numéro du tableau de Bescherelle (58)
- arg 2 : numéro du paradigme (1 = indic présent; 2 = imparfait, ..., 10 = participe présent)
- arg 3 : modifications du radical "- dre" aux 6 personnes
- arg 4 : désinences correspondantes

Le radical est présenté sous les six formes que requièrent ses six personnes ; les désinences sont également présentées en séries de six, pour correspondre aux six personnes. Le signe ["], de double apostrophe, désigne une chaîne de caractères vide, par exemple l'absence de personnes 1, 3 et 6 à l'impératif. Nous nous donnons une facilité de notation pour les formes non finies, comme les participes présent et passé : c'est de les insérer également dans une série de six, comme s'il s'agissait d'une liste non-complète de personnes.

Nous avons adopté l'ordre de succession des formes de Bescherelle ; nous y intégrons comme lui, pour cette démonstration, l'imparfait du subjonctif.

tableau	paradigme	radical	désinence
58	1 ind. prés.	[d,d,"d,d,d]	[s,s,d,ons,ez,ent]
58	2 ind. impft.	[d,d,d,d,d,d]	{ais,ais,ait,ions,iez,aient}
58	3 passé simple	[d,d,d,d,d,d]	[is,is,it,imes,ïtes,irent]
58	4 futur	[dr,dr,dr,dr,dr,dr]	[ai,as,a,ons,ez,ont]
58	5 subj. prést.	[d,d,d,d,d,d]	[e,es,e,ions,iez,ent]
58	6 subj. impft.	[d,d,d,d,d,d]	[isse,isses,ït,issions,issiez,issent]
58	7 impératif	["d,"d,d,"]	["s,"ons,ez,"]
58	8 condit.	[dr,dr,dr,dr,dr,dr]	[ais,ais,ait,ions,iez,aient]
58	9 part. prés.	[d]	[ant]
58	10 part. passé	[d,d,d,d]	[u,us,ue,ues]

L'infinitif *tordre* est décomposé en Radical "tor" + terminaison "dre", à l'aide de la règle Prolog :

radical (58, Infinitif, Rad) : - append (rad, "dre", Infinitif), !.

Les radicaux intégrant un préfixe s'insèrent aisément en isolant le préfixe, sous la forme par exemple de (*re-tor-dre*). Les verbes à plusieurs formes radicales supplétives, comme *aller*, entrent aisément dans les tableaux :

tableau	paradigme	radical	désinences
23	1	{v,v,v,all,all,v}	{ais,as,a,ons,ez,ont}
23	2	{all,all,all,all,all,all}	{ais,ais,ait,ions,iez,aient}
23	3	{all,all,all,all,all,all}	{ai,as,a,âmes,âtes,èrent}
23	4	{iir,iir,ir,ir,ir,ir}	{ai,as,a,ons,ez,ont}
23	5	{aill,aill,aill,all,all,aill}	{e,s,e,ions,iez,ent}
23	6	{all,all,all,all,all,all}	{asse,asses,asse,assions,assiez,assent}
23	7	{"v","all,all,"}	{"a","ons,ez,"}
23	8	{ir,ir,ir,ir,ir,ir}	{ais,ais,ait,ions,iez,aient}
23	9	{all,"","",""}	{ant}
23	10	{"all,all,all,all,"}	{"é,és,ée,ées"}

Cette représentation permet d'insérer les verbes défectifs dans les mêmes tableaux, en éliminant les lignes correspondant aux formes absentes, ou de remplacer les radicaux ou terminaisons par des chaînes vides. Ainsi, pour transposer le tableau 39 que Bescherelle (1997) affecte au verbe *gêner*, la représentation est la suivante :

39	1	{gi,gi,gî,gis,gis,gis}	{s,s,t,ons,ez,eeent}
39	2	{gis,gis,gis,gis,gis,gis}	{ais,ais,ait,ions,iez,aient}
39	9	{gis,"","",""}	{ant,"","",""}

Le programme comporte environ 100 tables de ce type. Il produit un "dictionnaire", selon le format suivant :

Forme conjuguée	Catégorie	Infinitif	Code
Abandonne	v	abandonner	13
Abandonnes	v	abandonner	14
...
Abandonniez	v	abandonner	23
Abandonnaient	v	abandonner	24

1.2.3. ÉTABLISSEMENT DES COMPTAGES

Conceptuellement, le recensement des formes verbales du corpus obligerait à examiner un à un les 700 000 mots du corpus, à vérifier dans le dictionnaire si ce sont des formes verbales et, en ce cas, à incrémenter un compteur pour les mode/temps/personnes. Pratiquement, on utilise le dispositif d'indexation des formes du corpus et du dictionnaire, déjà programmé dans le logiciel de concordances XCOR.

1.2.3.1. PRINCIPE DE L'INDEXATION

L'index principal est celui du corpus : c'est un mécanisme auquel on présente une forme et qui restitue une liste des positions (pouvant aller de 1 à plusieurs milliers) où l'on trouve cette forme dans le corpus. Le corpus utilisé ici comporte 795 000 items, dont 727 136 formes lexicales, dont environ 100 000 formes différentes. L'index comporte donc 100 000 entrées. Le procédé de localisation des occurrences d'une forme donnée dans le corpus est très efficace et a l'avantage de ranger les formes dans l'ordre alphabétique. En voici l'explication technique (en petits caractères pour que le lecteur impatient puisse s'en dispenser).

Cette structure très volumineuse est construite en mémoire, le système Prolog se chargeant des échanges mémoire / disque quand c'est nécessaire.

Les 100 000 entrées sont organisées en une structure arborescente "n-aire". Chaque nœud correspond à une forme, pointe vers la liste des positions de ces formes dans le corpus et est relié par n branches aux autres nœuds. Pour chercher une forme, on part du nœud-racine. Si c'est la forme recherchée, on arrête là. Sinon, on continue le processus sur le prochain nœud, selon l'algorithme désignant l'une des n branches à suivre.

Lors de la création de l'index du corpus, on a ajouté à chaque nœud le nombre d'occurrences de la forme représentée par le nœud. On pourrait également y attacher d'autres informations, comme par exemple la catégorie ou les catégories qu'on jugerait bon de mentionner pour la forme.

Le dictionnaire des verbes est indexé de la même façon sur les formes conjuguées.

1.2.3.2. COMPTAGE

Le programme est extrêmement simple. Les 29 780 formes conjuguées du dictionnaire sont examinées. L'index du corpus indique immédiatement s'il y a des occurrences de la forme et fournit le nombre. Quand il y a au moins une occurrence, elle est notée dans une structure informatique en tableau.

1.2.3.3. PRÉSENTATION DES RÉSULTATS

Un premier tableau donne un relevé des conjugaisons attestées pour tous les verbes du corpus. Pour le verbe *partir*, la première colonne montre qu'il y a 1 003 attestations; la troisième colonne montre que 80 % de ces occurrences sont représentées par 5 postes de conjugaison, alors qu'il y en a 20 d'attestés en tout, ce que signale la colonne n° 4.

La colonne n° 5 note le taux de concentration morphologique : le rapport entre le nombre total de postes de conjugaison attestés (colonne 4) et le nombre de ceux qui constituent 80 % de ce total (colonne 3). Pour le verbe *partir*, il suffit d'un quart des postes de conjugaison (20 : 5 = 0,25, rapport noté dans la colonne 5) pour représenter 80 % des formes conjuguées.

La colonne 6 représente le taux d'exploitation morphologique du verbe. C'est le rapport du nombre de postes représentant 80 % de la conjugaison au nombre d'attestations total du verbe. Pour *partir* (5 : 1 003) x 100 = 0,5). Le verbe *venir*, qui a une fréquence comparable (1 024 attestations) a un taux d'emplois morphologiques plus élevé : 0,78.

1	2	3	4	5	6	7
nombre occurrences	forme de l'infinitif	nombre de postes faisant 80%	nombre total de postes attestés	taux de concentr (3 : 4)	taux exploi- tation (5 : 1) x 100	liste des postes faisant 80 %
1453	falloir	2	7	0,29	0,14	faut, fallait
1222	parler	5	22	0,23	0,41	parler, parle, par- lait, parlé, parlent
1024	venir	8	23	0,35	0,78	viens, venu, venir, viennent, viens, venait, venue, ve- naient
1003	partir	5	20	0,25	0,5	part, partie, partir, parti, partis

Le détail des postes de conjugaison, comportant le nombre d'attestations pour chacun des postes "privilegiés" qui font 80 % du total, est donné dans le tableau suivant :

infinitif	Nombre d'attestations total	nombre d'attestations de chacun des postes représentant 80% du total, avec leurs codes				
		part [n°15]	parti [n° 3]	partir [n° 1]	partie [n°5]	partis numéros 4, 25, 26, 27
partir	1003	233	233	205	106	76

La liste des autres postes morphologiques est indiquée dans le tableau suivant, avec leur nombre d'attestations et le numéro de code (éventuellement plusieurs numéros de code) :

partir	9	partent	[18,12]
	9	partez	{17,45}
	3	partant	[2]
	45	partait	{21}
	33	parties	[6]
	21	partaient	[24]
	10	pars	[13,14,43]
	9	partais	[19,20]
	3	parte	[7,9]
	3	partiez	{23,11}
	1	partira	[39]
	1	partirai	{37}
	1	partirais	[31,32]
	1	partiras	[38]
	1	partiront	[42]

2. RÉSULTATS

2.1. FIABILITÉ DES RÉSULTATS

Le dépouillement automatique des résultats réserve quelques surprises: erreurs d'identification qu'il faut corriger soit à la main soit en utilisant un étiqueteur morpho-syntaxique (taggueur). En voici quelques unes :

- *maintenant*, classé comme participe présent du verbe *maintenir*.
- *permissions*, classé parmi les formes du verbe *permettre*.
- *âger*, verbe fantôme, affublé d'un présent *âge* et d'un participe passé *âgé*.
- *les dépenses*, pris comme présent de *dépenser*.
- *une épouse*, pris comme forme du présent de *épouser*.
- *une faille*, pris comme subjonctif de *falloir*.
- 22 occurrences de *restaurant* prises pour le participe présent de *restaurer*.

Ces erreurs viennent de l'homographie entre des formes verbales et des formes non-verbales, et ressemblent à des calembours. Certaines peuvent être décelées au vu des tableaux, lorsqu'on s'aperçoit qu'un poste morphologique est excessivement représenté par rapport à ce qu'on pourrait attendre. En voici quelques exemples :

nombre d'occurrences	Infinitif	nombre de postes représentant 80%	nombre total de postes attestés	Codes	postes de conjugaison
3781	taire	1	4	[3]	[tu]
2823	plaire	1	9	[5]	[plus]
511	entrer	2	16	[7,9,13,15,4 3] et [4]	[entre, entrée]
219	placer	1	9	[7,9,13,15,4 3]	[place]
181	poindre	1	2	[15]	[point]
286	importer	2	4	[2]	[important]
133	envier	1	4	[7,9,13,15,4 3]	[envie]
255	livrer	2	8	[7,9,13,15,4 3] et [8,14]	[livre, livres]
179	courir	2	10	[14,43] et [1]	[cours, courir]
156	former	2	12	[7,9,13,15,4 3] et [8,14]	[forme, formes]

Certaines formes classées parmi les participes passés, [3,4,5,6] sont suspectes : les 80 % d'attestations de participes sur 3.781 occurrences de *taire* sont plutôt des occurrences du pronom personnel *tu*. Les très nombreux *du* sont des occurrences du groupe préposition + article. Les 80 % de supposés participes passés de *plaire* sont des manifestations de l'adverbe *plus*.

Des formes de première ou troisième personne (indicatif ou subjonctif) interfèrent avec des formes nominales : *place*, *point*, *envie*, *forme*, ou avec des formes de préposition, comme *entre*. Les deux formes *entre* et *entrée* qui composent apparemment 80 % des postes morphologiques du verbe *entrer*, doté ici de 511 occurrences, sont en fait majoritairement des occurrences de la préposition *entre* et du nom *entrée*.

Le système de classement a un autre inconvénient plus gênant, c'est l'homonymie entre les formes verbales. Une même forme graphique, isolée de son contexte et traitée en tant que mot graphique, peut appartenir à plusieurs postes de conjugaison. Afin de prévenir les conclusions trop hâtives qui pourraient en découler, le système indique régulièrement, par des chiffres, toutes les affiliations possibles pour une même forme graphique. Par exemple, pour le verbe *sauter*, on trouve :

saute : [13,15,7,9,43]

ce qui renvoie à plusieurs possibilités :

13 : je saute

15 :	il saute
7 :	que je saute
9 :	qu'il saute
43 :	saute

Tous les verbes du premier groupe sont touchés par cette homographie, régulièrement signalée par la même série de codes : [13,15,7,9,43] :

sauve, signale, signifie, sonne, souffle, souffre, souhaite, soupe, structure, supporte, supprime, termine, tire, tolère, etc.

Il est indispensable de recourir au concordancier pour identifier ces formes par leur contexte. Dans l'interprétation des résultats, nous en tiendrons compte et nous apporterons les correctifs nécessaires.

2.2. GRANDES TENDANCES

a. La répartition des emplois est variable et dépend en grande partie de la fréquence des verbes. Plus ils sont fréquents, plus ils utilisent un grand nombre de postes de conjugaison. *Etre, avoir, aller, dire, faire*, sont les verbes les plus amplement conjugués.

b. Les emplois ne sont pas répartis équitablement entre les différents postes de conjugaison. Certains représentent une concentration d'emplois bien plus grande que les autres. Par exemple, le verbe *chercher* a 224 attestations, réparties sur 12 postes de conjugaison :

chercher : 133 – cherche : 35 – cherché : 23 – cherchait : 10 – cherchez : 9 – cherchent : 7 – cherches : 2 – cherchai : 1 – cherchaient : 1 – cherchaient : 1 – cherchons : 1 – cherchée : 1 –

Les quatre postes les plus fréquemment remplis pour ce verbe, *chercher* (133), *cherche* (35), *cherché* (23), *cherchait* (10), représentent près de 90 % de l'ensemble.

c. Un grand nombre de verbes ne sont représentés que par deux, trois ou quatre postes de conjugaison. Ils n'exploitent donc qu'environ 5 % de leurs capacités morphologiques virtuelles. Ce sont surtout des verbes de moyenne et faible fréquence :

mériter : mérite, mérité, mériter
atteindre : atteint, atteindre, atteinte
chauffer : chauffe, chauffer, chauffée
piquer : pique, piqué, piquer
consister : consiste, consistait
pleuvoir : plu, pleut, pleuvait

suffire : suffit, suffira

remonter : remonter, remonté, remonte, remontait

On observe dans ces types de conjugaison très “réduits”, qui concernent un grand nombre de verbes, des régularités dont il sera intéressant d'essayer de tirer parti en les comparant aux autres.

2.3. LES VERBES LES PLUS FRÉQUENTS

Les verbes considérés comme les plus fréquents ont, pour le premier, *avoir*, 23 000 occurrences et pour le dixième, *venir*, environ 1 000 occurrences. Ils exploitent une grande partie des 45 postes de conjugaison, entre les trois quarts et la moitié :

	occurrences	nb de postes de conjugaison exploités
avoir	23133	35
être	26541	31
aller	3615	30
faire	6413	28
pouvoir	2261	27
voir	2578	26
devoir	3726	24
dire	5150	23
venir	1024	23
parler	1222	22

Voici l'ensemble des formes attestées pour *avoir*, *être*, *venir* (données par rang de fréquence) :

a, avait, ai, eu, as, avais, avez, avoir, avaient, avez, avaient, avons, aurait, ait, aura, avions, aurais, aviez, auraient, aient, ayant, aurai, auriez, auras, auront, aie, aurons, aurions, ayez, aurez, aies, eue, eues, ayons, eut

est, était, sont, suis, été, être, étaient, étais, soit, es, êtes, sommes, sera, serait, étions, étiez, étant, soient, serais, serai, seront, seras, sois, seraient, fut, serez, serons soyons, soyez, seriez, étés

vient, venu, venir, viennent, viens, venait, venue, venaient, venus, venez, vienne, viendra, venais, venant, venues, venez, viendront, viendrai, viendrait, viennes, venons, viendrez, vins

En considérant l'ensemble des verbes du corpus, on estimera que les verbes dépassant 20 postes de conjugaison sur 45 sont des verbes "largement conjugués". Ce sont régulièrement les verbes les plus fréquents.

On ne peut cependant pas établir une corrélation absolument stricte entre l'ordre de fréquence et la différenciation morphologique. Le tableau ci-dessus montre que *faire*, nettement plus fréquent que *aller*, a moins de postes morphologiques et que *dire* et *devoir*, nettement plus fréquents que *pouvoir* et *voir*, en ont également moins. En prenant tous les verbes qui montrent entre 24 et 20 postes différents, on verra aussi des décalages de ce type. *Parler* est comparativement moins différencié que *trouver* ; *vivre* est comparativement près de trois fois plus différencié que *partir*. On peut fournir une première explication de ces décalages, en fonction des types de productions orales dans lesquels se rencontrent ces verbes. Par exemple, *vivre* se rencontre surtout dans les récits autobiographiques, impliquant une expression différenciée de plusieurs nuances de passé. On peut donc s'attendre à rencontrer des formes morphologiquement distinctes, surtout pour le passé.

	occurrences	postes
trouver	639	24
parler	1222	22
arriver	822	22
rester	528	22
mettre	967	21
passer	925	21
sortir	581	21
demander	451	21
partir	1003	20
vivre	362	20

Mais la corrélation est globalement valide : tous les verbes qui attestent moins de 20 postes morphologiques ont moins de 1 000 occurrences

(l'inverse n'étant pas vrai). On peut donc maintenir la conclusion que la différenciation morphologique est liée à la fréquence des verbes.

2.4. LA CONCENTRATION DES EMPLOIS

Certains verbes sont utilisés avec une grande fréquence sur quelques postes de conjugaison et très peu sur les autres. Pour mesurer cette concentration, nous avons calculé systématiquement pour chaque verbe quels postes de conjugaison (classés par ordre de fréquence) constituent 80 % des occurrences. Par exemple, pour *permettre* (140 occurrences), 5 postes sur 14 attestés occupent 80 % des emplois :

permet, permis, permettre, permettait, permettant

Les 9 autres, qui n'occupent que 20 % de l'ensemble, sont :

permettent, permettra, permets, permettez, permettais, permette, permettrai, permettront, permit

Voici des indications semblables pour quelques verbes dotés d'une centaine d'occurrences. Le premier chiffre indique le nombre de postes de conjugaison occupant 80% des emplois (la liste en est fournie après les chiffres) et le second indique le nombre total de postes de conjugaison attestés :

ouvrir	6 sur 15	ouvert, ouvre, ouvrir, ouverte, ouvertes, ouverts
envoyer	5 sur 15	envoie, envoyé, envoyer, envoyait, envoyée
danser	5 sur 11	danse, danser, dansé, dansait, danses
mener	5 sur 10	menait, mène, mener, mené, menais
fermer	5 sur 10	fermé, ferme, fermée, fermés, fermées
crier	4 sur 10	criait, crié, crie, criaient
paraître	3 sur 12	paraît, paraître, paru

Un des plus "concentrés" est sans doute le verbe *dépendre*, 107 occurrences, qui attestent 5 postes de conjugaison (*dépend, dépendait, dépendre dépendant, dépendons*), mais dont 80 % des occurrences sont représentées par un seul poste, *dépend*, qui se trouve régulièrement dans le syntagme *ça dépend*. Le verbe *interdire* se manifeste sous 5 formes, mais 80 % des emplois sont occupés par le participe passé *interdit*. De la même façon, toutes les occurrences du verbe *présumer* appartiennent à un seul et même poste de conjugaison, *présume*, qui fait partie du syntagme *je présume*. La très forte concentration montre ici des emplois fixés sur quelques postes privilégiés en nombre très réduit. Il faudrait sans doute les sortir de la conjugaison ordinaire pour en faire des cas à part : *ça dépend, c'est interdit, je présume*.

2.5. LES VERBES À CONJUGAISON RÉDUITE

Une partie des verbes de moyenne fréquence ont 80 % de leurs emplois représentés par trois postes de conjugaison, apparemment fondamentaux :

3 ^e personne	<i>SAIT</i>
infinitif	<i>SAVOIR</i>
participe passé	<i>SU</i>

C'est le cas pour une série de verbes comme :

chercher (3 sur 12)
essayer (3 sur 12)
perdre (3 sur 15)
gagner (3 sur 11)
continuer (3 sur 12)
paraître (3 sur 12)
rencontrer (3 sur 14)
préparer (3 sur 7)
construire (3 sur 7)
calmer (3 sur 7)
terminer (3 sur 6)
ramener (3 sur 6)
fatiguer (3 sur 6)
profiter (3 sur 7)
filer (3 sur 7)
piquer (3 sur 7)
chauffer (3 sur 6)
gronder (3 sur 5)
repasser (3 sur 5)
réaliser (3 sur 5)
transporter (3 sur 5)
mériter (3 sur 7)
transmettre (3 sur 5)
équilibrer (3 sur 6)
épouser (3 sur 5)
lâcher (3 sur 5)
aboutir (3 sur 5)
repousser (3 sur 5)
redoubler (3 sur 4)
confirmer (3 sur 4)
disposer (3 sur 4)
justifier (3 sur 4)

Certains verbes ne sont attestés que pour le participe passé, l'infinitif et la troisième personne du présent, ou seulement deux de ces postes (le chiffre indique le nombre d'occurrences dans chacun des postes cités) :

dégager	3 dé gagée 2 dé gager 1 dé gagé
détruire	7 détruit 3 détruite 2 détruire
justifier	2 justifier 2 justifiée 1 justifie 1 justifiées
redevenir	3 redevient 1 redevenir 1 redevenue

Cette restriction dans la conjugaison rappelle celle qu'on voit dans les verbes défectifs, souvent réduits aux mêmes trois postes de conjugaison :

clore	– il clot la séance – l'incident est clos
frire	– il frit – c'est frit
moudre	– il moud – c'est moulu
transir	– ça me transite – je suis transi

Pourtant, même si les formes attestées sont identiques dans les deux cas, les verbes à conjugaison très restreinte ne sont pas assimilables aux verbes défectifs. Leur conjugaison virtuelle semble être disponible et un Français pourrait sans doute citer une dizaine de formes pour le verbe *dégager* : imparfait, futur, participe présent, etc., alors qu'il ne le pourrait pas pour *clore*, *frire*, *moudre*, *transir*. Placés devant l'obligation de fournir d'autres postes de conjugaison pour ces verbes défectifs, la plupart des Français se trouvent confrontés à l'indicible (cf. Morin 1995) : impossible d'inventer la moindre forme d'imparfait ou de subjonctif pour *frire* ou *transir*. Comme l'explique Picabia (1998), à partir de *frire*, *ça frit*, on n'arrive pas à trouver la bonne "consonne latente" qui permettrait de produire la forme longue du radical, comme on le fait si facilement pour passer de *détruire*, *il détruit* à *nous détruisons*, *je détruisais*. Même si le verbe *détruire* n'est utilisé effectivement dans le corpus qu'à l'infinitif et au participe passé, la connaissance grammaticale que nous en avons n'est pas celle d'un verbe défectif. Il faut donc essayer de comprendre comment s'explique cette réduction dans la conjugaison attestée, tout en se démarquant des explications fournies pour les verbes défectifs.

3. INTERPRÉTATION DES RÉSULTATS

3.1. FACTEURS PROPREMENT MORPHOLOGIQUES

Les résultats montrent qu'on ne peut pas interpréter la limitation des formes conjuguées par une quelconque difficulté morphologique qu'éprouveraient les locuteurs à en produire davantage. Ces arguments vaudraient si la conjugaison était particulièrement réduite pour des verbes comme *résoudre*, *luire*, *fuir*, *coudre* ou autres réputés difficiles. Mais, bien au contraire, la réduction de la conjugaison s'observe pour les verbes du premier groupe réputés les plus faciles. Martinon l'avait déjà bien observé :

- “Il y a d'ailleurs une infinité de verbes en *-er* qui en fait ne sont usités qu'à l'infinitif ou encore au participe ou au présent” (1927 : 317).

3.2. IMPORTANCE DES INFINITIFS ET DES PARTICIPES PASSÉS

La fixation sur l'infinitif et le participe passé se comprend facilement. Le participe passé apparaît dans toutes les formes auxiliées, ce qui lui assure une fréquence importante :

il les aura justifiées – nous les aurions justifiés – les avoir justifiés

Le participe passé contribue par là à l'expression des temps et aspects des verbes, mais il a aussi pour effet que la morphologie proprement dite de ces verbes n'est, en ce cas, pas sollicitée. Dans *il les aura justifiées*, le verbe *justifier* n'est pas fléchi en temps ni personne. Il n'y a ici de conjugaison que celle de l'auxiliaire et on pourrait imaginer un état de la langue où ce système serait généralisé et dans lequel il n'y aurait plus de morphologie verbale que pour les auxiliaires.

L'infinitif apparaît dans les combinaisons avec les verbes modaux :

ils vont les justifier – vous pouvez le justifier – on devrait le justifier –

Ici non plus, la morphologie du verbe *justifier* n'est pas sollicitée et seuls les verbes modaux, *vont*, *pouvez*, *devrait*, sont en ce cas morphologiquement fléchis.

Le rôle assuré par les auxiliaires et les verbes modaux semble donc avoir supplanté, pour une partie des emplois, le recours à une morphologie verbale différenciée.

3.3. FACTEURS SÉMANTIQUES ET ÉNONCIATIFS

Il est un peu factice de vouloir chercher un impératif pour les 12 000 verbes du français. Les grammairiens, l'ayant reconnu depuis longtemps, ont souvent expliqué qu'il ne saurait y avoir d'impératif pour *devoir*, *pouvoir*, car "le pouvoir et le devoir ne se commandent pas" (cf. Martignon 1927, Curat 1991, Wilmet 1997).

Une autre limite est celle des verbes à emploi "impersonnel", non seulement les "verbes météorologiques" comme *il pleut*, *il neige*, mais aussi les verbes à sujet fortement restreint comme *ça urge*, *ça dépend*, etc. Il est possible que certains verbes, qui ont par ailleurs une morphologie virtuelle développée, soient exploités préférentiellement, dans certaines situations de communication, à la troisième personne du présent.

Il existe aussi, sans doute, une restriction de conjugaison inhérente au sens de verbe de parole ou de jugement comme *je crois*, *je trouve*. *S'exclamer* et *s'écrier* semblent bien être limités à quelques formes de troisième personne, au présent et au passé simple, à la façon des *ait* et *inquit* latins (Traglia 1950).

D'autres facteurs sémantiques ou énonciatifs, que nous n'avons pas identifiés, pourraient probablement expliquer la grande restriction dans la conjugaison de certains verbes.

3.4. FACTEURS HISTORIQUES ET INSTITUTIONNELS

Habités à des tableaux de conjugaison scolaires très homogènes, au regard desquels tout manquement est une défektivité, nous aurions facilement tendance à trouver "normal" que toutes les cases de tous les verbes soient équitablement remplies. Mais c'est peut-être une vue moderne, influencée par les pratiques scolaires, qu'une enquête historique amènerait à relativiser.

Les cases des tableaux de conjugaison des grammairiens semblent avoir été de plus en plus remplies au fil du temps, comme si les auteurs de manuels avaient tenu à doter tous les verbes d'une morphologie virtuelle maximale. Du coup, la distance entre la conjugaison virtuelle et la conjugaison effective en a été minimisée et sous-estimée. Une petite incursion dans l'histoire de la présentation des verbes défectifs peut en donner une illustration frappante.

3.5. ÉVOLUTION DU TRAITEMENT DES DÉFECTIFS

Les verbes “défectifs” sont, en principe, ceux auxquels il manque des formes de conjugaison, ne serait-ce qu'une ou deux par rapport à l'ensemble canonique. Les raisons du manque peuvent être doubles : soit que les formes virtuelles nécessaires n'existent pas, soit qu'elles existent mais que, pour différents motifs, l'usage n'en soit pas recommandé. Les grammairiens sont malheureusement peu explicites sur ce point :

Qu'est-ce qu'on appelle verbes défectueux ? – Ce sont ceux auxquels il manque certains tems ou certaines personnes que l'usage n'admet pas : tels sont *querir* dont on ne se sert qu'à l'infinitif, au prétérit et aux tems composés, *frîre* qui ne se dit pas aux trois personnes du pluriel du présent de l'indicatif et quelques autres dont nous parlerons dans la suite. (Restaut, 1745, p. 234)

On appelle verbe défectif tout verbe qui n'est pas usité à certains temps ou à certaines personnes : *Gésir, clore, ouïr, falloir*. (E. Robert et F. Chazallet, 1910, *Nouvelle Grammaire française, cours moyen*. Paris : Figoreau, p. 200)

De grandes différences semblent s'être installées dans les règles d'usage, les grammaires actuelles s'efforçant de remplir au maximum les cases prévues pour la conjugaison, quitte à fournir des formes très peu attestées, alors que les grammaires plus anciennes étaient plus disposées à reconnaître qu'il y avait des “trous” dans les usages. Sans pouvoir pour l'instant fournir une documentation suffisante, nous pouvons citer quelques exemples de verbes nettement défectifs, dont les grammairiens modernes ont élargi le champ de conjugaison.

– *Clore*

En 1745, Restaut ne lui donnait à l'indicatif présent que les personnes du singulier, *je clos, tu clos, il clôt*, pas de pluriel et pas de subjonctif (p. 93). Ces lacunes sont confirmées par de Wailly en 1803. En 1997, Bescherelle lui donne une troisième personne du pluriel, *ils closent*, et un subjonctif présent complet, *que je close*, etc. Le Goffic ajoute, sans commentaires, *nous closons*.

– *Éprendre*

Il est mentionné par la plupart des grammaires du XX^e siècle comme un verbe qui se conjugue, sur le modèle de *prendre* (Grevisse 1986, Le

Goffic 1997), alors que, au XVII^e siècle, les observateurs du bon usage avaient tendance à l'écartier⁴:

Epris est un participe dont le verbe n'est point en usage. On ne dit pas éprendre, et encore épris ne se dit guères qu'en Poésie. M. Dépreaux néanmoins l'a dit à propos en prose. Je fus sincerement épris de tant d'admirables qualitez. (N. Andry de Boisregard, Réflexions sur l'usage présent de la langue françoise, 1689, p. 197)

En 1997, Bescherelle et Le Goffic indiquent simplement qu'il se conjugue comme *prendre*, c'est-à-dire apparemment sans aucune lacune.

– *luire*

Les grammaires actuelles donnent un passé simple au verbe *luire* : *luisit*, déclaré "rare" mais donné cependant comme vivant par Le Goffic. Bescherelle (1997) le donne, sans commentaire, comme aligné sur *cuire*. Les grammaires plus anciennes déclaraient que *luire* n'avait pas de passé simple (Restaut en 1745, Robert-Chazallet en 1890).

– *Vaincre*

Il est présente actuellement comme un verbe bien complet, qui aurait seulement une orthographe difficile au présent, *je vaincs, tu vaincs, il vainc* (Bescherelle 1997, Grevisse-Goosse 1986, Le Goffic 1994). Ces formes étaient tout simplement déclarées hors d'usage aux XVII^e et XVIII^e siècles :

Vaincre n'est pas d'usage au singulier du présent de l'indicatif ; on ne dira pas, par exemple, comme le dernier Traducteur de l'Imitation : l'accoutumance au bien se vainct par l'accoutumance au mal, mais on doit dire l'accoutumance au bien se surmonte par l'accoutumance au mal. (de Boisregard, 1689, p. 697)

Restaut fait la même remarque en 1745 :

Ce tems n'est guere d'usage au singulier. (p. 328)

– *vêtir*

En 1997, Bescherelle et Le Goffic lui donnent un présent complet : *je vêts, tu vêts, il vêt, nous vêtons, vous vêtez, ils vêtent*. En 1745, Restaut déclarait :

Le singulier de ce temps n'est point en usage. (p. 312)

4. Je dois ces citations à M. Seijido, qui fait une thèse sur le grammairien Andry de Boisregard.

ce que confirmait de Wailly en 1803 :

Dans *vêtir*, le singulier du présent indicatif, *je vêts, tu vêts, il vêt*, n'est guère en usage. (1803, p. 86)

Il semble donc bien que la notion de défectivité a évolué au cours de l'histoire de la langue et que les grammairiens classiques consentaient davantage que ceux de maintenant à déclarer hors d'usage certaines formes de conjugaison, effectivement peu utilisées par leurs contemporains.

De cette évolution dans le traitement des verbes défectifs, on peut facilement en inférer une autre, valide pour tous les verbes du français : les tableaux de conjugaison ont été de plus en plus remplis, de sorte qu'ils ont laissé de moins en moins de place à la différence entre conjugaison virtuelle et conjugaison effective. On le voit bien pour un verbe comme *vaincre*, qui n'a semble-t-il jamais été déclaré défectif, mais que les grammairiens classiques estimaient "peu en usage" aux formes du singulier du présent. Cette notion a disparu. Il ne reste plus que deux extrêmes. Un verbe doit être soit défectif soit complet. Il ne saurait être "réduit". Les grammairiens contemporains ont "rempli" le verbe *vaincre* pour le rendre "normal". Nous ne pouvons donc plus nous fier à notre instinct qui nous pousserait à juger un peu gênants *vaincs, vainc*.

Les tableaux de conjugaison contiennent ainsi quantité d'exemples qui nous font perdre confiance en notre jugement grammatical : avons-nous vraiment, comme semble le dire Bescherelle (1997), une connaissance grammaticale du verbe *messeoir* qui nous ferait accepter comme formes virtuellement disponibles *ils messièrent, ils messeyaient* ? Acceptons-nous *mouvriions, échoyait, closions, pâtrez, joignissions, trayions*, toutes formes données comme "normales", sur le même plan que *chante, chantait* ? Les tableaux scolaires, bien remplis, se sont substitués aux jugements de l'usage. Nous sommes, du coup, très surpris lorsque nous constatons les importants décalages entre la conjugaison attestée, la conjugaison virtuelle et la conjugaison des tableaux scolaires.

4. CONCLUSION

Dans les corpus de français parlé que nous avons analysés, seuls les verbes de très haute fréquence ont une large conjugaison effective. Les verbes de moyenne fréquence ont des emplois très fortement concentrés sur certaines formes plus que sur d'autres. La conjugaison des verbes de faible fréquence tend à se fixer sur trois postes de conjugaison

privilegiés : le participe passé, l'infinitif et la troisième personne du présent de l'indicatif.

Une bonne partie des verbes défectifs en sont réduits également à ne fonctionner que pour ces mêmes trois postes de conjugaison. Mais nous faisons l'hypothèse qu'il y a, entre les deux types, une grande différence. Alors que les verbes à conjugaison effective réduite conservent l'ensemble de leur conjugaison virtuelle, les verbes défectifs n'en ont plus. Les locuteurs ne peuvent plus, en ce cas, puiser dans le réservoir de la conjugaison virtuelle pour "retrouver" des formes. N'ayant plus du tout ce réservoir de conjugaison virtuelle, les verbes défectifs, bloqués sur un très petit nombre de formes (de une à trois), ne sont plus vraiment des verbes. *Transir, traire, quérir, clore*, par exemple, accompagnés de leurs quelques formes disponibles, sont les équivalents, en morphologie, de ce que sont les formes figées en syntaxe.

Le phénomène de la conjugaison effective réduite, qui s'explique sans doute par un grand nombre de facteurs divers, est une situation morphologique normale. Il est normal qu'il y ait un grand décalage entre la conjugaison virtuellement disponible et la conjugaison effective. Mais, selon l'hypothèse que nous formulons, ce phénomène aurait été progressivement masqué par les grammaires scolaires, soucieuses de présenter des tableaux bien remplis et de minimiser le décalage entre ces tableaux et les usages effectivement attestés. Or il s'agit de savoir si ce que nous appelons "conjugaison" est un modèle fondé sur les formes verbales utilisées ou si c'est seulement la liste "récitée", comme le disait déjà Restaut en 1745, des formes du verbe qu'on enseigne à l'école. La liste "récitée" est sans doute, surtout de nos jours, beaucoup plus extensive que l'ensemble des formes virtuelles dont disposent les locuteurs.

La langue parlée permet, mieux que la langue écrite, d'approcher ce type de décalage, qui joue sans doute un rôle important dans de nombreux autres fonctionnements linguistiques. Il serait possible d'en tirer des conséquences pour la présentation générale des formes grammaticales du français, comme on l'a déjà fait pour d'autres langues⁵.



5. "Most English verb forms do not seem to be frequent enough to warrant pedagogical attention until quite advanced stages of the second language acquisition process. Learners' time might well be better spent on acquiring a bigger vocabulary or paying more attention to sociopragmatic aspects of competence [...]" (G. Kennedy, 1998 : 284).

Bibliographie

- ADAM, Jean-Pierre, à paraître, "Exploitation informatique des corpus de langue parlée au GARS. Questions de méthode dans la linguistique sur corpus", in M. BILGER (éd.), *Actes du Colloque de Perpignan, Collecte, organisation et traitement des corpus linguistiques*, 7-9 mai 1998.
- BARRAL, Marcel, 1980, *L'imparfait du subjonctif. Étude sur l'emploi et la concordance des temps du subjonctif*. Paris : Picard.
- BERGON, Anne, 1996, La défectivité verbale, Mémoire de DEA, Département de Linguistique française.
- BESCHERELLE, 1966, *L'Art de conjuguer. Dictionnaire des 8.000 verbes usuels*. Paris : Hatier.
- , 1997, *La Conjugaison pour tous. Dictionnaire de 12.000 verbes*. Paris : Hatier.
- CSECSY, M., 1968, *De la linguistique à la pédagogie. Le verbe français*. Paris : Hachette (coll. Le Français dans le Monde, BELC).
- CURAT, Hervé, 1991, *Morphologie verbale et référence temporelle en français moderne. Essai de sémantique grammaticale*. Genève-Paris : Droz.
- DAMOURETTE, J. et PICHON, E., 1914-1940. *Des Mots à la Pensée. Essai de Grammaire française*. Paris : D'Artrey.
- EYNDE, van den, Karel et BLANCHE-BENVENISTE, Claire et 1970, "Essai d'analyse morphologique du verbe français. Présentation d'hypothèses de travail", *Orbis*, 19, 404-29.
- EYNDE, van den, Karel et SWIGGERS, P. (éds.), 1987, *Morphologie du verbe français*. Numéro Spécial de *I.T.L.*, 77-78.
- FOLEY, J., 1979, "Theoretical Morphology of the French Verb", *Linguisticae Investigationes, Supplementa*, vol. 1.
- GERTNER, M.H., 1973, *The Morphology of the Modern French Verb*. Janua Linguarum, Series Practica, 204. The Hague : Mouton.
- GRABOWSKI, E. and MINDT, D., 1995, "A corpus-based learning list of irregular verbs in English", *ICAME Journal* 19 : 5-22.
- GREVISSE, M., 1986, *Le bon usage, Grammaire française*. 13^e édition refondue par M. GOOSSE. Louvain : Duculot.
- ISAAC, Luc, 1985, *Calcul de la flexion verbale en français contemporain*. Genève-Paris : Droz.
- KENNEDY, Graeme, 1998, *An Introduction to Corpus Linguistics*. London-New York : Longman.
- KJELLMER, G., 1992 "Grammatical or nativelike ?", in G. LEITNER (ed.) *New Directions in English Language Corpora. Methodology, Results, Software Developments*. Berlin : Mouton de Gryter.

- LE GOFFIC, Pierre, 1997, *Les formes conjuguées du verbe français*. Paris : OPHRYS (coll. L'essentiel).
- MARCELLO-NIZIA, Christiane, 1997, *La langue française aux XIV^e et XV^e siècles*. Paris : Nathan (coll. Université).
- MARTINON, Philippe, 1927, *Comment on parle en français. La langue parlée correcte comparée avec la langue littéraire et la langue familière*. Paris : Larousse.
- MORIN, Yves Charles, 1995, "De l'acquisition de la morphologie : le cas des verbes défectifs du français", in H. BAT-ZEEV SCHYLDKROT et L. KUPFERMAN (éds), *Tendances récentes en linguistique française et générale. Volume dédié à David GAATONE*. Amsterdam-Philadelphie : Benjamins, 295-310.
- PICABIA, Lelia, à paraître, "De la défectivité de certains verbes français", XXII^e Congrès de Linguistique et de Philologie Romanes, Bruxelles, juillet 1998.
- RESTAUT, 1745, *Principes généraux et raisonnés de la Grammaire Française [...]*, Paris : chez Ph. Lottin.
- RIEGEL, Martin, PELLAT, Jean-Christophe et RIOUL, René, 1994, *Grammaire méthodique du français*. Paris : PUF.
- RIGAULT, André, 1971, *Grammaire du français parlé*. Paris : Hachette.
- SELJIDO, Magali, en cours, Thèse sur les *Réflexions sur l'usage présent de la langue française*, par Andry de Boisregard, 1689, Département de Linguistique Française, Université de Provence.
- TRAGLIA, Antonio, 1950, *La Flessione verbale latina. Trattato di morfologia storica*. Torino : Società Editrice Internazionale.
- WAGNER, Robert-Léon, 1974, *L'ancien français*. Paris : Larousse (Coll. Langue et langage).
- WAILY, de, 1803, *Principes généraux e particuliers de la langue française, Suivis d'un Abrégé de Versification*. Paris : chez H. Barbou.
- WILMET, Marc, 1997, *Grammaire critique du français*. Louvain-Paris : Duculot-Hachette.

Apports des données orales dans l'analyse des valences à l'infinitif des adjectifs recteurs

Dans les dix dernières années, les linguistes ont complété de plus en plus leurs descriptions par des observations sur les corpus de langues parlées. B. Altenberg (1991) a présenté une bibliographie comptant plus de 600 références d'études sur l'anglais parlé. J. Ambrose (1996), qui propose une bibliographie faisant suite à celle de Cl. Blanche-Benveniste et de C. Jeanjean (1987), a rassemblé plus de 1000 références publiées entre 1986 et 1995 d'études prenant en compte le français parlé. Les travaux entrepris sur les productions réelles ont permis d'identifier des structures qui avaient été considérées jusque-là comme peu importantes (par exemple, pour le français parlé, Cl. Blanche-Benveniste et alii [1990], les articles de la revue du GARS *Recherches Sur le Français Parlé*; et par exemple, pour l'anglais parlé Stenström et Svartvik [1994]). Ces travaux ont également permis d'observer un certain nombre de comportements langagiers. Cl. Blanche-Benveniste (1984 et 1990) a observé une disparité entre d'une part les modèles syntaxiques élaborés pour décrire la grammaire du français et d'autre part l'usage que font les locuteurs de la syntaxe. Concernant ces relations entre les locuteurs et leur grammaire, elle a fait une hypothèse : il faudrait distinguer deux sortes de savoirs : "l'un, qui est indiscutable, et semble acquis par les francophones avant leur entrée à l'école primaire; un autre, aléatoire, qu'ils acquièrent plus ou moins bien, tout au long de leur vie, et parfois jamais. L'un correspond à ce que nous appelons la 'grammaire première' et l'autre à la 'grammaire seconde'" (Cl. Blanche-Benveniste et alii, 1990). D. Biber et alii (1996), qui pensent que c'est surtout dans ce domaine que la linguistique sur corpus a eu le plus d'impact, concluent par l'idée que la linguistique sur corpus a certainement permis d'établir à partir de la mise

en évidence de faits nouveaux que les descriptions doivent se faire en distinguant plusieurs registres de langue :

Research that was previously infeasible — considering the association patterns across lexical, grammatical, and textual levels, as well as the interplay of linguistic and non-linguistic characteristics — has become possible with corpus-based techniques. [...] Thus, both for understanding how users exploit the resources of the language and for teaching language based on real communication and appropriate use, corpus based research promises to be an increasingly valuable and productive resource for applied linguists.

À partir de l'exemple des valences des adjectifs recteurs comportant un infinitif, comme par exemple *je suis curieux de le voir* ou *la vérité est douloureuse à dire*, je voudrais montrer comment une étude sur corpus avec observation de la langue parlée peut aider à mieux comprendre l'usage que l'on fait de sa grammaire. Ces structures se prêtent particulièrement bien à de telles études de par les multiples types et de par la forte fréquence de certains de ces types comparée à celle d'autres.

Je m'attacherai à analyser l'utilisation des différents types de valences à l'infinitif dans la rection d'un adjectif. Cette description sera fondée sur un corpus de plus de 800 exemples oraux et écrits. Ce corpus a été constitué à partir de plusieurs sources : un relevé des exemples littéraires présentés dans la *Syntaxe du français contemporain* de Kr. Sandfeld et dans la *Grammaire française* de K. Togeby; et un relevé systématique effectué dans environ 20 heures de corpus de français parlé (soit près de 350 000 mots) du Groupe Aixois de Recherches en Syntaxe (GARS), et dans plusieurs numéros des quotidiens *Le Monde* et *Libération*.

En insistant particulièrement sur l'importance de la fréquence des emplois à l'oral, je tenterai de montrer que l'analyse de ce type d'infinitifs est incomplète dans les études traditionnelles, que les types et les formes de valences à l'infinitif dépendent souvent de plusieurs sortes de savoirs, et que l'on ne peut pas mettre sur le même plan des constructions qui dépendent d'un savoir grammatical différent même si elles ont un "air de famille" apparent. Pour ce faire, je commencerai par introduire brièvement dans le premier paragraphe la notion de nexus. Puis, dans le deuxième paragraphe, consacré aux nexus à l'infinitif avec relation complément, je rappellerai la présentation traditionnelle des données en termes normatifs et reprendrai la description en y intégrant des exemples oraux pour tenter de proposer un autre classement qui distinguerait les tournures "avec formule courte" *une langue difficile (à apprendre)* et celles "sans formule courte" *une affection difficile à combattre*. Le troi-

sième paragraphe, qui traitera des tournures avec relation sujet, confirmera l'intérêt de différencier les tournures intégrant un infinitif complément réductible à une proforme *l'élève y est apte, à se débrouiller* et celles qui comportent un nexus sans infinitif réductible *nous sommes les seuls à les avoir*. Dans le quatrième paragraphe, j'essaierai de montrer l'intérêt qu'il y a à analyser *les mouettes c'est joli* comme une formule courte de *les mouettes c'est joli à voir*, et les limites de cette notion. Et dans le cinquième et dernier paragraphe, j'aborderai la question des infinitifs sujets.

1. LA NOTION DE NEXUS

O. Jespersen (1971a et 1971b) et par O. Eriksson (1993) ont proposé d'admettre l'existence de deux combinaisons possibles entre deux éléments. D'une part, la *jonction* serait le type de relation qui unit un épithète à un nom :

- a) il me donne une rose rouge, il m'en donne une

Et d'autre part, dans des énoncés comme :

- b) je trouve les roses belles, je les trouve belles
c) j'aime bien Pierre quand il sourit, j'aime bien ça (Pierre quand il sourit)

la valence ne serait pas constituée par deux constituants syntaxiquement autonomes, mais par la relation qui unit les deux termes, *les roses et belles, Pierre et quand il sourit*. La grammaire traditionnelle parle dans le cas de b) d'attribut de l'objet.

Je reprendrai cette analyse en nexus pour analyser des exemples tels que :

- d) il est long à partir
e) il est facile à convaincre

Je considérerai qu'en d) l'adjectif *long* est recteur et que sa rection est constituée non pas par un complément à l'infinifitif mais par un nexus, c'est-à-dire par la relation "sujet" qui unit l'infinifitif *partir* et le pronom *il*. En e) l'adjectif *facile* est recteur et sa rection est constituée par la relation "complément" qui unit l'infinifitif *convaincre* et le pronom *il*.

2. NEXUS À L'INFINITIF AVEC RELATION COMPLÉMENT : UNE LANGUE DIFFICILE À APPRENDRE

2.1. PRÉSENTATION DES DONNÉES

Les grammaires traditionnelles s'accordent à différencier deux emplois des nexus à l'infinitif avec une relation complément :

a) Certaines tournures impliquent des relations “verbe transitif + complément de type *le, la, les*”. Elles sont normatives :

- 1) c'est une langue *qui est très difficile à apprendre* (corpus Portugais 24, 6)
- 2) *la vérité elle est trop douloureuse à dire* (LIC 89-A, Chatti, 28, 8)
- 3) j'ai des échecs derrière moi *ces échecs sont lourds à porter ils sont lourds à surmonter* (LC 88, Lovico, 10, 15)
- 4) *le vieux fusil c'est valable à voir* (corpus Giacomi, 222, 28)
- 5) *ce que je veux vous dire est vraiment très difficile à dire* (hôtesse d'accueil)

b) D'autres comportent des relations “verbe + complément prépositionnel de type *y / à ça, en / de ça, avec ça, sur ça*”. À l'exception de la tournure *quelqu'un difficile à vivre*, elles ne sont pas normatives :

- 6) Je n'ai pas sur moi *la somme nécessaire à prendre seulement le tramway* (Miomandre, in Sandfeld § 172)
- 7) elle élevait par instants ce petit manchon au-dessus de ses yeux, afin de se ménager *le jour nécessaire à bien voir le tableau* (P. Bourget, in *ibid.*)
- 8) *la tendresse nécessaire à supporter à deux le lourd fardeau* (Const.-Weyer, in *ibid.*)
- 9) *la route est mauvaise à marcher* (Gervais, in *ibid.*)
- 10) *c'est pas facile à répondre* (TF1, B. Tapie)
- 11) c'est *une: très bonne question* + pertinente mais *très difficile à répondre* (LIC94, étudiant comorien, 1, 8)
- 12) c'est *un insecte difficile à se débarrasser* (France Inter, Michel le jardinier)
- 13) Estoril est *un circuit difficile à dépasser* (TF1, A. Prost)
- 14) *c'est facile à marcher avec* (M6, Film)

La liste des tournures normatives, avec des verbes à complément non prépositionnel, est ouverte :

- le, la, les aimer, apprendre, combattre, dérouler, dire, dissiper, écouter, émouvoir, évaluer, gérer, obtenir, porter, poser, surmonter, tenir, visiter, voir ...

En revanche, celle des tournures non normatives semble plus limitée. La majorité des exemples relevés présente une seule occurrence :

- prendre le tramway avec cette somme
- bien voir le tableau avec ce jour
- supporter à deux le lourd fardeau avec cette tendresse
- marcher sur la route
- se débarasser de cet insecte
- dépasser (quelqu'un, une voiture) sur ce circuit
- marcher avec (ça)

Si l'on considère l'occurrence la plus fréquente, *une question difficile à répondre*, l'ensemble des relations semble contraint. Je n'ai observé aucune variation lexicale dans les différents paradigmes, ni sur l'adjectif :

- une question difficile à répondre
- (?) facile
- (?) dure
- (?) aisée

ni sur le verbe :

- une question difficile à répondre
- (?) à réagir

ni sur le nom :

- une question difficile à répondre
- (?) un appel
- (?) une sollicitation

De plus cette occurrence présente une relation de type *y répondre* et il est difficile d'imaginer d'avoir une relation de type *lui répondre* :

- (?) quelqu'un difficile à répondre
- (?) une personne
- (?) des gens

L'incorporation de données orales au corpus va permettre de proposer un classement différent de ces tournures, fondé sur plusieurs critères : syntaxiques, sémantiques, lexicaux, niveaux et registres des textes, fréquence et variété des occurrences.

2.2. PROPOSITION D'UN CLASSEMENT EN DEUX TYPES

Reprenons les exemples 1) à 5), déjà présentés au paragraphe 2.1 et rangés ci-dessous dans la liste “type A”, et comparons-les aux exemples 15) à 19) introduits dans la liste “type B” :

Type A :

- 1) c'est *une langue qui est très difficile à apprendre* (corpus Portugais 24, 6)
- 2) *la vérité elle est trop douloureuse à dire* (LIC 89-A, Chatti, 28, 8)
- 3) j'ai des échecs derrière moi *ces échecs sont lourds à porter ils sont lourds à surmonter* (LC 88, Lovico, 10, 15)
- 4) *le vieux fusil c'est valable à voir* (corpus Giacomi, 222, 28)
- 5) *ce que je veux vous dire est vraiment très difficile à dire* (hôtesse d'accueil)

Type B :

- 15) j'ai autour de moi euh donc un exemple d'une angine d'un début d'angine qui est quand même *une affection euh + as- assez- assez difficile à combattre* enrayée avec des essences de plantes que l'on trouve dans le commerce (LB88, Hiss, 20, 9)
- 16) vous savez *il est drôlement dur à poser hé le le grillage de deux mètres* (corpus Maçon)
- 17) *elle le savait faible et facile à émouvoir* (du Gard, *Les Thibault*, I 55-56, in Sandfeld § 209)
- 18) *Le coût de la défense du franc au mois de juillet, encore difficile à évaluer*, pourra atteindre près de 10 milliards (*Le Monde*)
- 19) Mais *le sentiment d'échec reste difficile à dissiper* dans le cas de la Bosnie (*Le Monde*)

Les énoncés des types A et B ont les mêmes propriétés syntaxiques et sémantiques : par exemple, en 1) et en 15), l'adjectif *difficile* (à) est respectivement recteur des nexus *une langue ... à apprendre* et *une affection ... à combattre*, il est sémantiquement compatible avec les infinitifs *apprendre* et *tenir*, et porte donc sur la relation qui unit les infinitifs et les noms. Il faut pourtant relever une différence entre les énoncés des deux types. Les tournures de type A font fortement penser aux tournures sans infinitif suivantes :

- c'est une langue qui est très difficile
- la vérité elle est trop douloureuse
- ces échecs sont lourds
- le vieux fusil c'est valable
- ce que je veux vous dire est vraiment très difficile

En revanche, il est mal aisé de vouloir mettre en relation les tournures de type B avec les tournures sans infinitif :

- un début d'angine est une affection assez difficile
- il est drôlement dur le grillage de deux mètres
- elle le savait facile
- le coût de la défense du franc est encore difficile
- le sentiment d'échec reste difficile

Récemment, L. Barblan (1989), à partir d'expériences qu'il a proposées à plusieurs groupes d'enfants d'âge différent, a montré l'importance d'un corpus en tenant compte dans ses analyses de l'usage et de la fréquence des énoncés, et de la mise en situation de l'énoncé : il a ainsi posé comme exceptionnelle la neutralité contextuelle qu'avait imposée C. Chomsky à ses énoncés-type. Il s'est aperçu que la commutation d'un des termes nominaux, adjectifs ou infinitifs modifiait certaines relations sémantiques. Dans son étude, L. Barblan a montré par exemple que les jeunes enfants ont des difficultés à comprendre les tournures de type B. Même si l'on évite les exemples avec les verbes de perception comme *voir*, *entendre* et *toucher* qui peuvent les perturber, les enfants font des erreurs sur l'interprétation de l'agent de l'infinitif. Par exemple, confrontés à une question comme :

- une bulle de savon c'est dur à tenir ?

les enfants répondent "non" dans un premier temps parce qu'ils rattachent *dur* à *une bulle* comme dans la construction à double-marquage du sujet en *ce* :

- une bulle de savon c'est dur ?

Puis, dans un second temps, ayant détecté une qualification de l'infinitif, ils se reprennent en répondant "oui".

S'inspirant de la notion de "prévisibilité lexicale" développée par Z.S. Harris (1970c) pour qui la séquence présupposée est la séquence appropriée, Z. Vendler (1968) propose l'analyse suivante : selon les types d'usage, dans la structure *easy problem*, le verbe *to solve* serait présupposé; dans la structure *good shoe*, il est probable que le verbe à envisager soit *to wear*, mais il serait tout à fait possible, pour un explorateur affamé, d'envisager le verbe *to eat*. Appliquée aux tournures équivalentes

en français, cela donne : (a) lorsque le verbe à l'infinitif n'est pas lexicalement prévisible, il doit être réalisé lexicalement; la valence relationnelle de l'adjectif apparaît avec le verbe à l'infinitif; par exemple, dans *elle le savait facile à émouvoir*, on ne peut pas avoir *elle le savait facile* sans que l'adjectif ainsi que l'énoncé tout entier change de sens; ici, l'infinitif doit être réalisé lexicalement; (b) lorsque le verbe à l'infinitif est hautement prévisible, l'adjectif devient métonymiquement compatible avec le nom, et l'infinitif peut ne pas être réalisé lexicalement; par exemple, dans *une langue qui est très difficile à apprendre*, on peut avoir *une langue qui est très difficile* sans aucun problème d'interprétation.

Je propose de nommer ce même problème que l'on a désigné comme "adjectif compatible avec le nom", "infinitif réalisé lexicalement zéro quand hautement prévisible" sous le nom de "tournure à formule courte", et celui que l'on a désigné comme "adjectif non compatible avec le nom", "infinitif non hautement prévisible" sous le nom de "tournure sans formule courte"¹. Les adjectifs qui interviennent le plus fréquemment dans ces tournures, qu'elles soient avec ou sans formule courte, sont : *agréable, amusant, beau, bon, commode, compliqué, difficile, délicat, douloureux, dur, épuisant, facile, important, impossible, intéressant, joli, long, lourd, magnifique, pénible, prêt, simple, subtil, terrible, valable*.

Les "tournures avec formule courte" sont relativement fréquentes aussi bien dans la langue littéraire que dans les corpus de français parlé. Le relevé des exemples littéraires intègre plus de 80 lexèmes adjectifs différents tels que *abject, admirable, adorable, affreux, agréable, amer, amusant, âpre, avide, bête, charmant, comique, cruel, curieux, dangereux, délicat, délicieux, désespérant, divertissant, doux, dramatique, drôle, effrayant, élégant, émouvant, enivrant, ennuyeux, épouvantable, étonnant, étrange, exquis, extraordinaire, fastidieux, fatigant, faux, gai, gênant, glacial, gros, grotesque, heureux, hideux, honteux, indispensable, injuste, inquiétant, instructif, insupportable, interminable, laid, lent, maladroit, malaisé, mélancolique, merveilleux, nécessaire, négligent, nostalgique, pesant, pis, plaisant, précieux, rafraîchissant, redoutable, scabreux, succulent, triste, troublant, utile, vraisemblable* :

20) Car voilà *ce qui est abject à dire* et qu'il faut pourtant que j'avoue (Henri Lavedan *Duel* III.4, in Sandfeld § 210)

21) *la salle vide*, sonore comme un caveau et *glaciale à voir* (Daudet *Trente ans de Paris* 184, in *ibid.*)

1. Sur les formes longues et courtes d'une valence, voir C. Jeanjean (1984), Cl. Blanche-Benveniste (1992), et É. Castagne (1995 et 1998).

- 22) *Elle était* divertissante et *instructive à voir* (de Curel. Sauvetage du Grand-Duc, in Sandfeld § 211)

Cette variété est remarquable quand on la compare à celle observée dans les corpus de français parlé et dans la presse écrite contemporaine qui ne comptent qu'une vingtaine de lexèmes adjectifs différents qui appartiennent tous à la liste des adjectifs les plus fréquents. Ainsi, l'adjectif qui intervient le plus fréquemment dans le corpus de français parlé est *difficile* et représente en moyenne 15 % des emplois alors que, dans le relevé d'exemples littéraires, il s'agit de *facile*, qui ne représente que 5 % des emplois. Il faut remarquer aussi que certains adjectifs n'apparaissent que dans des "tournures avec formule courte", qui sont relativement contraintes :

- 23) *ce sac est lourd à porter* (in L. Picabia, 1978)
 24) *ce tissu est doux à toucher* (in *ibid.*)
 25) *cette pente est raide à gravir* (in *ibid.*)
 26) *cette voiture est nerveuse à conduire* (in *ibid.*)

Par exemple, en 23), on voit mal avec quel autre verbe à l'infinitif on pourrait associer sémantiquement *un sac lourd* sinon avec *porter*, *soulever* ou d'autres verbes du même champ sémantique. En 26), on voit mal avec quel autre verbe à l'infinitif on pourrait associer sémantiquement *une voiture nerveuse* sinon avec *conduire*. Il est à noter également que les tournures non normatives présentées en 2.1. sont massivement² de type "avec formule courte".

Les "tournures sans formule courte" sont beaucoup plus rares que les premières. Dans les corpus de français parlé³ et dans la presse écrite contemporaine, elles représentent en moyenne moins de 10% des tournures avec relation complément. Les exemples sont restreints à un petit nombre de lexèmes adjectivaux, en l'occurrence *difficile*, *dur*, *facile*, *agréable* et *beau* parmi lesquels *difficile* et *dur*, les plus fréquents, représentent à eux seuls plus de 80 % des emplois :

Corpus de FP :

- 15) j'ai autour de moi euh donc un exemple d'une angine d'un début d'angine qui est quand même *une affection euh + as- assez:*

2. Dans notre corpus, seuls les exemples 12) et 14) ne paraissent pas avoir de "formule courte".

3. Il est remarquable que la production orale des "tournures sans formule courte" est souvent accompagnée de bribes, hésitations, répétitions, amorces ou corrections. Les "tournures avec formule courte" sont massivement produites à l'oral sans ces phénomènes de mise en place du discours.

assez difficile à combattre enrayée avec des essences de plantes que l'on trouve dans le commerce (LB88, Hiss, 20, 9)

- 16) vous savez *il est drôlement dur à poser hé le le grillage de deux mètres* (corpus Maçon)
- 27) *notre atmosphère est trop difficile pour eux à tenir* (corpus Bilgal, 8, 146)

Presse écrite :

- 18) *Le coût de la défense du franc au mois de juillet, encore difficile à évaluer*, pourra atteindre près de 10 milliards (*Le Monde*)
- 19) Mais *le sentiment d'échec reste difficile à dissiper* dans le cas de la Bosnie (*Le Monde*)
- 28) *La dureté de la concurrence internationale, déjà difficile à gérer en période de prospérité*, devient insupportable en période de récession et de chômage. (*Le Monde*)

Les grammaires que nous avons consultées en citent peu d'exemples et rarement attestés. Dans les études plus spécialisées (par exemple, R. Kayne (1974-1975), H. Huot (1981), N. Ruwet (1982), K. Baschung (1991)), il y a un exemple qui revient très souvent : *Jean est facile à contenter (ou joindre)*⁴. Quand on trouve un corpus d'exemples attestés, comme dans la monographie de Kr. Sandfeld, il s'agit d'exemples qui sont exclusivement extraits de la langue littéraire, qui sont en nombre restreint et dont le taux ne dépasse pas 15% des emplois de ce type :

- 17) *elle le savait* faible et *facile à émouvoir* (du Gard, *Les Thibault*, I 55-56, in Sandfeld § 209)
- 29) *Cette France difficile, presque impossible à aimer* (Paul Bourget, André Cornélis, 207, in *ibid.*)
- 30) *L'argent* n'est pas *difficile à gagner, mais à garder* (Frondaie, Appass. I. 5., in *ibid.*)

Même s'il ne compte guère qu'une quinzaine de lexèmes adjectivaux différents tels que bon, curieux, difficile, dur, effroyable, facile, impossible, long, lourd, magnifique, malaisé, périlleux, prêt, propre, satisfaisant, le relevé confirme une variété lexicale des adjectifs dans un corpus d'origine littéraire. Ainsi, l'adjectif qui intervient le plus fréquemment, tous corpus confondus, dans les "tournures sans formule courte" est l'adjectif difficile. Il représente jusqu'à 50 % des emplois dans les corpus de français parlé et seulement environ 20 % des emplois dans la langue littéraire.

4. L'exemple le plus fréquent dans les études anglaises est *John is easy to see*.

2.3. ANALYSE

Il semble que le fonctionnement syntaxique particulier aux valences que demandent les adjectifs de ce type, qu'il s'agisse de l'ordre des mots ou du type relationnel de cette valence, n'est pas aisément maîtrisé par les locuteurs, comme le démontre le nombre relativement important d'études consacrées à cette question comme C. Chomsky (1969; 1972), J. Cambon et H. Sinclair (1974), L. Solan (1978), L. Barblan (1989). Dans son étude, L. Barblan montre que, pour tenter de comprendre, les locuteurs semblent s'aider alors non pas d'une propriété syntaxique, mais d'une propriété sémantique mieux dominée : la possibilité d'une "formule courte"⁵. Si l'on confronte les deux emplois que nous venons de distinguer et si l'on prend en compte que les tournures non normatives sont massivement de type "avec formule courte", on est amené à faire l'hypothèse que les "tournures à formule courte", dont le fonctionnement est acquis avant la scolarisation, appartiennent à la "grammaire première" et que les "tournures sans formule courte", dont le fonctionnement est mal maîtrisé par les locuteurs, appartiennent à la "grammaire seconde". L'analyse en formule courte pourrait avoir des conséquences intéressantes et non prévues jusqu'ici. On peut déjà penser qu'il y aurait beaucoup d'avantages à analyser "les mouettes c'est joli" comme une formule courte de "les mouettes c'est joli à voir"⁶ (cf. *infra*).

3. TOURNURES AVEC RELATION "SUJET" :
 UN ÉLÈVE APTÉ À SE DÉBROUILLER, NOUS SOMMES
 LES SEULS À LES AVOIR

Je distinguerai deux sortes d'emplois, les infinitifs en position complément (qui peuvent être réduits à une forme pronominale) :

- l'élève est apte à se débrouiller (l'élève y est apte)
- je suis très heureux de vous rencontrer (j'en suis très heureux)

et les infinitifs intégrés à un nexus (qui se caractérisent notamment par la difficulté de se voir ramener à une forme pronominale)⁷ :

- nous sommes les seuls à les avoir
- (?) nous y sommes les seuls

5. Au sens harrissien et au sens du GARS.

6. Voir D. Willems (1977).

7. Pour plus de précision sur la distinction syntaxique entre les infinitifs en position complément et les infinitifs intégrés à un nexus, voir ma thèse.

- ils sont obligés de les frapper
- (?) ils en sont obligés

Je me limiterai aux emplois des infinitifs introduits par *à* et par *de*.

3.1. LES INFINITIFS COMPLÉMENTS : *L'ÉLÈVE Y EST APTE, À SE DÉBROUILLER; J'EN SUIS TRÈS HEUREUX, DE VOUS RENCONTRER*

Plusieurs études psycholinguistiques ont montré que les enfants comprennent les structures relationnelles de type agentif, sans doute parce qu'ils posséderaient la structure où le SN doit être interprété comme le sujet de l'infinitif. Par exemple, L. Solan (1978) a obtenu avec des enfants de 3 à 5 ans d'excellents résultats dans l'interprétation des structures avec les adjectifs de type *eager*.

Les adjectifs qui interviennent le plus fréquemment dans ces tournures sont d'une part les adjectifs adroit *à*, apte *à*, attentif *à*, enclin *à*, habile *à*, prêt *à*, propice *à*, les participes accoutumé *à*, adapté *à*, aguerri *à*, amené *à*, appliqué *à*, contraint *à*, habilité *à*, habitué *à*, obligé *à* et d'autre part les adjectifs capable de, certain de, heureux de :

- Adj. à V-er :
 - 31) à partir de là le moniteur juge — si *l'élève est apte + à se débrouiller* (LIC 92-6, Guillo, 8, 16)
 - 32) ben oui *on est prêt à y penser aux victimes + on est prêt à dialoguer* (corpus Lorgeoux, 40, 14)
 - 33) selon les *ministères* que nous avons + *qui sont plus ou moins euh + enclins à à faire des des réformes ou de ne pas en faire ou à en faire dans le mauvais sens + euh + on on obtient des choses absolument. atroces* (corpus Les Baumettes, 18, 6)
 - 34) il faut faire des sacrifices *je suis prête à les faire* (LC 88, Lovico, 10, 11)
- V-é à V-er :
 - 35) on peut: supposer qu'*il est habitué à se faire voler sa mallette* (LIC 92-8, Villel, 20, 1)
 - 36) *il faut. être: assez aguerri à: marcher. à grimper euh + des sites /tout de même, XXX/ qui sont accidentés* (LIC 92-4, Blosse, 15, 15)
 - 37) *ils sont obligés à dépasser le à dépasser leur(s) différence(s)* (LIC 92-6, Lions, 11, 12)
 - 38) puisqu'*ils sont euh je dirais accoutumés à + à faire les tâches quotidiennes* (LIC 94, l'expert, 8, 7)

- Adj de V-er :

- 39) vous savez monsieur Julio *je suis très très heureux de vous rencontrer* (corpus Histoires)
- 40) voilà monsieur Julio *je suis très content de vous avoir rencontré* (corpus Histoires)
- 41) tu sais euh après un an de travail dans une pharmacie *n'importe qui est capable de délivrer une ordonnance* (corpus Pharmacie)
- 42) *on (n') est pas du tout étonné de voir qu'un jour + ils se sont révoltés* (corpus Les Baumettes)

3.2. LES NEXUS À L'INFINITIF : NOUS SOMMES LES SEULS À LES AVOIR, ILS SONT OBLIGÉS DE LES FRAPPER

Plusieurs études psycholinguistiques ont montré que les enfants comprennent les structures relationnelles “sujet”. Pourtant je voudrais mettre en évidence plusieurs traitements en relation avec les différents types de valences que nous avons présentés et que ces études ne prennent pas en compte.

Le relevé que j'ai effectué semble montrer que les nexus à l'infinitif de ce type, dans lesquels l'infinitif ne peut être réduit à une forme pronominale sont aussi fréquents que les infinitifs compléments vus dans la paragraphe 3.1.

Les grammaires latines et les grammaires historiques du français ne citent aucun exemple. À ma connaissance, A. Haase est le premier grammairien à donner dans sa *Syntaxe française du XVII^e siècle* quelques rares exemples de ce type. Il faut remarquer qu'il y a eu évolution des tournures avec infinitif introduit par *de* ou avec gérondif vers nos tournures :

- 43) *je suis libre à disposer de moi* (Corneille, D. Sanche , I, 3, 131, in Hase § 124B)
- 44) *Je vous trouve plaisante à me parler ainsi* (Molière, *les Femmes savantes* , V, 2, 1584, in *ibid.*)
- 45) *Je deviendrais suspect à tarder davantage* (Corneille, *Cinna*)

Les grammaires du français contemporain se limitent souvent à des exemples non attestés, et à un petit nombre de lexèmes adjectivaux comme *bon à, lent à, long à, le premier à, le seul à* ou *bon de, bête de, gentil de*. Kr. Sandfeld présente dans sa monographie une liste d'une trentaine de lexèmes adjectifs différents tels que *alerte à, ardent à, avide à, bon à, commode à, court à, docile à, énergique à, franc à, fidèle à, haute à, impuissant à, ingénieux à, inlassable à, juste à, lent à, long à, paresseux à,*

perspicace à, ponctuel à, préférable à, régulier à, savant à, seul à, subtil à, timide à, utile à, vif à, les superlatifs tels que *le premier à, le dernier à, le seul à*, ainsi que les adjectifs *aimable de, bon de, drôle de, fou de, gentil de* :

- Adj à V-er :

- 46) Ses moustaches russes -- indiquaient *une nature spontanée, ardente à s'épancher* (Maurois, in Sandfeld § 168)
- 47) Comme *elle est* soumise aujourd'hui et *docile à poser cette même tête sur l'épaule de M. Termonde* ! (P. Bourget, in *ibid.*)
- 48) Parmi *les femmes de la Halle les plus énergiques à défendre le réactionnaire Pitou*, on eût trouvé beaucoup de celles qui -- avaient marché sur Versailles (Drumont, in *ibid.*)
- 49) *Il est alerte à saisir les occasions de gagner de l'argent* (Dict. de l'Académie française, in *ibid.*)
- 50) Parlant peu, sec et rogue, n'admettant ni réplique ni contestation, *vif à se défendre* (Lav. Le Montyon de la légende, in *ibid.*)

- Adj de V-er

- 51) *Ce garçon est très aimable d'entrer ainsi chaque jour* (Zola, in Sandfeld § 272)
- 52) *Que vous êtes gentil, mon petit Édouard, de penser toujours à moi* ! (Jaloux, in *ibid.*)
- 53) *Nous fûmes bien bons de nous donner tant de peine* (Béraud, in *ibid.*)
- 54) *Est-elle drôle de tout me raconter* ! (Rolland, in *ibid.*)
- 55) Tiens, *je suis fou de t'avoir parlé* (Deval, in *ibid.*)

Dans les corpus de français parlé ou dans la presse écrite contemporaine, les tournures de ce type semblent aussi fréquentes que dans la langue littéraire. Mais la liste des adjectifs utilisés fréquemment semble restreinte aux adjectifs *bon à, long à, obligé de, contraint de* et surtout aux superlatifs *le premier à, le dernier à, le seul à* qui à eux seuls représentent environ 90 % des emplois de ce type :

- Adj à V-er :

- 56) alors élève à ce moment-là tu es tu peux être chauffeur tu peux être mécanicien tu *tu es bon pour ainsi dire à tout faire* (LC 88, Magnan, 5, 4)
- 57) il y a eu *un début de championnat qui a été un petit peu dur à se mettre en marche* (TF1, T. Roland)

- 58) est-ce que c'était + plus précisément parce que: *elle était seule à : à assumer: financièrement aussi les problèmes* (corpus Barall, 2, 8)
- 59) *nous on était les premiers élèves à ++ à descendre à bien visiter le chantier à bien tout voir* (LIC 94, le tunnel 14, 8)
- 60) à Marseille nous avons effectivement /un, Ø/ bataillon de marins-pompiers exceptionnel(s) + *nous sommes les seuls à les avoir* car Paris possède les pompiers et les provinces ont ce que l'on appelle les sapeurs-pompiers (corpus Birnie, 1, 12)
- V-é de V-er :
 - 61) donc *on est obligés de la + transporter* quoi (corpus Alsace)
 - 62) *j'ai été obligé de partir plus tôt* c'était déjà très tard (corpus Bus)
 - 63) on comprend qu'*ils sont obligés de les frapper* pour les faire rentrer dans les cellules et pour faire rétablir le calme (corpus Les Baumettes)
 - 64) La première fois, *le gouvernement avait été contraint de durcir le crédit* au bout de quelques semaines (*Le Monde*)
 - 65) le secrétaire général de l'ONU a averti qu'*il serait alors contraint de proposer des "remaniements majeurs"* pour permettre à la machine onusienne de continuer de tourner (*Le Monde*)

3.3. ANALYSE ET HYPOTHÈSE

Tous les locuteurs utilisent le fonctionnement syntaxique que demandent les adjectifs de ce type avec les deux types de valences que nous avons dégagés. Pourtant, à l'exception des cas avec quelques rares adjectifs et participes passés tels que *bon à, long à, le premier à, le dernier à, le seul à, et obligé de, contraint de*, les locuteurs n'utilisent pas les adjectifs qui peuvent régir un nexus à l'infinitif introduit par *à, de* dans une relation "sujet". Il n'y a là non plus rien d'étonnant quand on sait que la majorité des ouvrages de référence ne présentent pas ce type de valence dans l'éventail proposé des constructions possibles de ces adjectifs. On est amené alors à faire l'hypothèse suivante : dans le premier cas, les adjectifs ont une construction dont le fonctionnement est connu et acquis bien avant la scolarisation, qui appartient à ce que nous avons appelé "grammaire première"; en revanche, dans le second cas, les adjectifs peuvent proposer plusieurs types de constructions, dont au moins un, en l'occurrence le nexus à l'infinitif, semblerait mal connu et appartiendrait à la "grammaire seconde".

4. LIMITES DE LA NOTION DE "FORMULE COURTE" : C'EST JOLI CES MOUETTES

Je vais maintenant détailler l'étude des formules courtes dont on peut retrouver les formules longues.

À partir d'un relevé systématique dans les productions orales spontanées, j'ai distingué deux types de tournures : les "tournures à formule courte" et les "tournures sans formule courte". Les formules courtes sont relativement fréquentes et concernent aussi bien les nexus à l'infinitif avec relation complément présentés au paragraphe 2 que les tournures avec relation sujet traitées au paragraphe 3 :

- relation complément :
 - 66) on essaie d'extraire de ce dossier ce qu'il y a d'intéressant et de le mettre en scène on le met en scène avec des mots *des mots qui doivent être simples* parce que ceux à qui on s'adresse en principe sont des gens qui n'aiment pas le compliqué et vous verrez que toutes les personnes anciennes les personnes âgées qui ont beaucoup de compétence parlent avec des mots simples (corpus Avocat) (= simples à comprendre)
 - 67) il y a trois méthodes en gros + on peut mettre de l'alcool à quatre-vingt dix + et laisser cinq minutes + bien imbiber et tout + c'est *une méthode un peu longue* + on l'utilise + de préférence quand ... (corpus Microbio) (= longue à faire, à utiliser)
 - 68) mais c'est *une parenthèse utile* (corpus Holding) (utile à faire)
- relation "sujet" :
 - 69) de l'autre côté là-bas + on on plante des arbres on fait des allées on fait des parcs + ici on ne fait rien + je lui ai dit + alors il m'a dit que *je n'étais pas raisonnable* (corpus Bus) (= raisonnable de dire ça)
 - 70) il y a l'éléphant qui va voir Noé qui lui demande + euh Noé on voudrait bien se marier + Noé il dit écoutez je peux pas + je peux pas absolument pas célébrer un + un mariage contre nature + et la petite souris elle s'avance en baissant la tête et puis elle dit + maintenant c'est trop tard *vous êtes obligé* (corpus Histoires) (= obligé de le célébrer)

Dans la valence de l'adjectif, il peut y avoir deux types de formes en concurrence.

On peut avoir une forme à valeur [+ individuelle], de type *je, tu, il* :

- 69) de l'autre côté là-bas + on on plante des arbres on fait des allées on fait des parcs + ici on ne fait rien + je lui ai dit + alors il m'a dit que *je n'étais pas raisonnable* (corpus Bus) (= raisonnable de dire ça)

- 70) il y a l'éléphant qui va voir Noé qui lui demande + euh Noé on voudrait bien se marier + Noé il dit écoutez je peux pas + je peux pas absolument pas **célébrer un** + un mariage contre nature + et la petite souris elle s'avance en baissant la tête et puis elle dit + maintenant c'est trop tard **vous êtes obligé** (corpus Histoires) (= obligé de le célébrer)
- 71) et c'est bien vrai **l'Alsace** c'est + **elle est belle** hein (corpus Alsace) (= belle à voir)

On peut avoir la forme *ça*, à valeur [-individuelle], qui représente soit un terme de la relation :

- 72) on a visité Lisbonne + puis + **c'est très joli** hein + il y a beaucoup de statues (corpus Alsace) (= joli à voir, à visiter)

soit toute la relation, c'est-à-dire le nexus qui unit la forme nominale et le verbe à l'infinitif non réalisé lexicalement :

- 73) on essaie d'extraire de ce dossier **ce qu'il y a d'intéressant** et de la mettre en scène (corpus Avocat) (= intéressant à relever, noter, à extraire)
- 74) vous verrez que toutes les personnes anciennes les personnes âgées qui ont beaucoup de compétence parlent avec des mots simples **plus c'est simple mieux c'est** (corpus Avocat) (= simple à comprendre)
- 75) **les crèches** oui il faut payer une certaine somme oh c'est pas **c'est pas cher** (corpus Alsace) (= cher à payer)
- 76) elles sont pas venues ici pour apprendre le français c'est plutôt euh + euh + la façon de boire quoi + et si euh encore ce serait du bon vin + mais alors **le vin de table vraiment + le le pinard là + c'est affreux** (à boire) (corpus Alsace)
- 77) oui + alors **le résultat c'est pas + excellent** (à voir) (corpus Alsace)
- 78) parce que vous les matériaux où vous allez les prendre parce qu'il m'a dit que **c'était cher** (à payer) (corpus Maçon)
- 79) donc on voit toujours euh + un nuage de de de mouettes au-dessus de la prison des Baumettes moi je disais comme c'est **c'est joli ces mouettes** + mon fils m'a dit + oui elles viennent à cause des pouelles (corpus Baumettes) (= à voir)
- 80) (une personne à qui on demande si l'un des membres de sa famille est avocat) **c'est assez particulier** (à situer, considérer, prendre en compte) **moi** bon euh parce que mon père était militaire et ma mère est institutrice et je suis métisse (corpus Avocat)
- 81) **c'est beau ça + c'est très joli ça** (corpus Alsace) (= à voir)

- 82) on m'avait dit que *c'était cher la la clôture* (corpus Maçon) (= à payer)

Par exemple en 79), la forme *ça* correspond à la relation qui unit le SN *ces mouettes* et le verbe à l'infinitif non réalisé lexicalement *à voir*.

Toutefois on peut se demander s'il est raisonnable de poser tous les énoncés proposés ci-dessus comme des formules courtes. Par exemple, en 75), le recours au verbe semble à peine acceptable. La construction *les crèches c'est pas cher à payer* a ici un côté artificiel. De plus, le nombre des verbes lexicalement réalisables à cette position semble extrêmement limité : il semble difficile d'avoir un autre verbe que *payer*.

J'ai relevé un seul exemple (tous types de relations confondus) où le locuteur fait une erreur sur le type de forme en produisant une forme courte alors que les caractéristiques de ce type de structures ne le permettent pas. Toutefois, il faut remarquer que le locuteur s'en aperçoit immédiatement et se reprend :

- 83) oh *mon chéri il n'est pas difficile euh je veux dire qu'il n'est pas difficile à vivre* (F2, judoka française aux J.O. d'Atlanta)

5. LES INFINITIFS SUJETS

5.1. PRÉSENTATION DES DONNÉES

Si les infinitifs compléments sont relativement fréquents, les infinitifs sujets sont très rares⁸. Pourtant plusieurs linguistes comme R. Kayne, (1974-1975) et, en partie, N. Ruwet (1982) ont essayé de rendre compte du lien rectionnel dans les tournures avec relation complément en proposant une "transformation par montée de l'objet" qui présente comme sources des tournures avec infinitif sujet (*convaincre Jean est facile*) ou des tournures impersonnelles qui sont des équivalents des infinitifs sujets (*il est facile de convaincre Jean*). Même si cette proposition semble pouvoir satisfaire l'interprétation immédiate, elle semble cependant pouvoir être mise en doute. La faible fréquence des infinitifs sujets révèle le caractère peu naturel de cette proposition qui accorde un privilège aux tournures avec infinitif en position sujet, ou aux tournures impersonnelles, pour en faire le centre d'une explication.

8. Voir L. Honoré (thèse en cours).

Si j'ai trouvé peu d'exemples d'infinitifs en position sujet, j'ai obtenu un relevé plus important de structures infinitives en double-marquage avec une forme de type *ça* en position sujet. La structure infinitive se trouve parfois en position antéposée :

- 84) *dire qu'elle est trop laxiste c'est faux* (corpus Avocat)
 85) *aller se développer dans des tas d'organes bon c'est très grave* (corpus Microbio)

mais le plus souvent en position postposée :

- 86) *est-ce que véritablement c'est normal de violer d'abord une femme c'est pas normal c'est /anormal, pas normal/ est-ce que c'est normal de violer une enfant c'est encore moins normal est-ce que c'est normal de tuer quelqu'un d'autre ce n'est pas normal* (corpus Avocat)
 87) *est-ce que véritablement c'est normal de violer d'abord une femme c'est pas normal c'est /anormal, pas normal/ est-ce que c'est normal de violer une enfant c'est encore moins normal est-ce que c'est normal de tuer quelqu'un d'autre ce n'est pas normal* (corpus Avocat)
 88) *... qui font + un millimètre un millimètre et demi deux millimètres trois millimètres en vingt-quatre heures quand même c'est rare d'atteindre beaucoup plus hein* (corpus Microbio)
 89) *c'est un peu cruel hein de lui faire ça hein* (corpus Apostrophe)
 90) *c'est pas indiscret de vous demander votre âge* (corpus Histoires)

Dans ces exemples, l'infinitif comporte souvent une réaction avec indéfini ou quantifieur : *des tas de* (85), *une enfant* (86), *quelqu'un* (87), *beaucoup plus* (88).

Quand on nous présente des exemples attestés, il sont souvent très littéraires et paraissent peu naturels⁹, comme :

9. Il s'agit là d'un fait qui nous paraît très intéressant. En effet, lors de notre travail de thèse, nous étions arrivé à la conclusion que l'infinitif ne semblait pouvoir être régi dans une valence autonome qu'en position de sujet :

— *quid est agrum bene colere ? — bene arare* (Caton, Agr. 61, i) (qu'est-ce bien cultiver une terre ? — bien labourer)

— *turpe est mentiri* (mentir est laid)

en position complément, on avait des formes supplétives qui sont le gérondif, l'adjectif verbal en *-ndus* et le supin.

- 91) *Bien rosser et garder rancune est aussi par trop féminin !*
(Beaumarchais, in M. Riegel et alii, 1994)
- 92) *Vaincre les êtres et les conduire au désespoir est facile*
(A. Maurois, in Grevisse)
- 93) *Admirer la pensée de Proust et blâmer son style serait absurde* (J. Cocteau, in *ibid.*)

Je n'ai relevé aucun exemple en position sujet dans la presse, plus précisément dans un numéro entier du journal *Le Monde*, et seulement deux occurrences dans les corpus de français parlé que nous avons consultés, en l'occurrence dans le discours d'un avocat :

- 94) lorsque vous gagnez de l'argent c'est très agréable et *dire le contraire serait faux* (corpus Avocat)
- 95) *faire étalage de ses richesses n'est jamais très bon notamment au regard du fisc* (corpus Avocat)

Le relevé des tournures impersonnelles de type *il est Adj de V-er* a été pauvre (moins d'une dizaine) dans les corpus de français parlé :

- 96) *il est beaucoup plus avantageux pour euh un individu d'avoir affaire à une personne plutôt qu'avoir affaire à dix personnes* (corpus Avocat)
- 97) *il est strictement interdit de parler au chauffeur* (corpus Histoires)
- 98) *il est très fortement conseillé + de bien faire tenir la lamelle sur la lame* (corpus Microbio)
- 99) *il est très important d'avoir des conseils juridiques* (corpus Holding)
- 100) pour des raisons pratiques + avant de lancer l'opération de fusion proprement-dite nous nous apercevons -- que + il c'est *il est très important -- euh de + céder les parts des trois ...* (corpus Holding)
- 101) est-ce qu'*il est prévu de + limiter cette + opération à ces huit mois qui sont prévus* ou est-ce que d'ores et déjà on pense + l'étendre (corpus Alsace)

Mais il a été plus fructueux dans les colonnes du *Monde* :

- 102) *Il lui sera toutefois difficile de dissiper les malaises* d'une rentrée sociale dominée par l'inquiétude qu'inspirent la dégradation continue de l'emploi et la montée de la grogne syndicale (in *Le Monde*)
- 103) Mais son ministre de l'intérieur a mis la barre si haut qu'*il va lui être difficile de trouver une solution de compromis.* (in *ibid.*)

- 104) "L'influence des volcans sur les changements climatiques reste un domaine scientifique très mal connu. *Il est donc difficile de se prononcer*", reconnaît Harvey Weiss. (in *ibid.*)
- 105) Le prieuré est en partie occupé par *de petites maisons dont il est difficile de déloger les locataires*. (in *ibid.*)

5.2. ANALYSE

Dans l'usage des infinitifs régis, la distinction établie entre les infinitifs compléments et les infinitifs sujets peut d'abord étonner quand on sait que la majorité des ouvrages de référence présentent l'infinitif sans distinction de position, mais elle est intéressante par l'hypothèse que l'on peut faire : les valences autonomes à l'infinitif en position complément sont acquises avant la scolarisation et appartiennent à la "grammaire première"; en revanche, les valences autonomes à l'infinitif en position sujet, dont le fonctionnement est mal maîtrisé, appartiennent alors à la "grammaire seconde".

6. CONCLUSION

Si 1) l'on admet la différence de statut, dans le savoir grammatical des Français, accordé aux deux sous-types de nexus à l'infinitif avec relation complément que nous avons appelés "tournures avec ou sans formule courte", si 2) l'on considère l'usage d'une structure peu naturelle comme l'infinitif sujet majoritairement restreint aux écrits littéraires et au discours oral spontané d'un avocat professionnel, et si 3) l'on prend en compte la restriction des combinatoires que l'on constate à l'oral et la dispersion lexicale que l'on observe à l'écrit, alors on est conduit à reconnaître qu'il est difficile d'établir un statut syntaxique sans tenir compte de la nature des données. Toutes les données ne peuvent pas être mises sur le même plan. Comment les présenter dans les descriptions ? On ne peut pas se contenter d'un classement intuitif par "registres". On ne peut pas se satisfaire d'un classement selon l'opposition "infinitif régi/nexus à l'infinitif", ni selon l'opposition "relation 'sujet'/relation complément", ni selon l'opposition "position sujet/position complément". On est conduit alors à réviser plus ou moins l'approche des grammaires fondée sur les catégories.

Sans recours aux données orales, ni à la notion des savoirs grammaticaux, ni à celle des restrictions lexicales, on risque de proposer des exemples non représentatifs d'un emploi naturel, comme c'est le cas pour

l'infinitif sujet. On peut même s'aventurer à proposer des traitements incohérents et inadaptés en ne prenant pas en compte la notion de nexus. Les transformationalistes considèrent des structures appartenant à la grammaire seconde comme sources d'une structure appartenant à la grammaire première. Dans les présentations de grammaires formelles que l'on peut lire aujourd'hui, on aimerait savoir si la prise en compte des différents statuts des données pourrait avoir des répercussions sur l'analyse.

On voit l'importance que peuvent avoir les paramètres qui définissent respectivement des usages distincts comme par exemple la source des exemples, la fréquence des emplois ou le type de locuteur. L'apparition de telle ou telle tournure, même parfois non normative, n'est pas le fait du hasard. Au contraire, il semble s'agir de phénomènes savamment organisés par l'intuition grammaticale et lexicale des locuteurs.

Éric CASTAGNÉ
Université de Provence



Bibliographie

- ALTENBERG, B., 1991, "Amplifier collocations in spoken English". In S. Johansson and A.-B. Stenström (eds) *English in speech computer corpora : Selected papers and research guide*. Berlin, Mouton, pp.127-147.
- AMBROSE, Jeanne, 1996, *Bibliographie des études sur le français parlé*, Paris, Didier Érudition.
- BARBLAN, Léo, 1989, *Apports de la pragmatique, de la sémantique et de la syntaxe à la maîtrise du langage. Une étude psychogénétique "difficile à conclure"*, Thèse de Doctorat, Université de Genève.
- BASCHUNG, Karine, 1991, *Grammaires d'unification à traits et contrôle des infinitifs en français*, Clermont-Ferrand, Adosa.
- BIBER, D., & HARED, M., 1992, "Literacy in Somali : linguistic consequences", in *Annual Review of Applied Linguistics*, Cambridge University Press.
- BIBER, D., CONRAD, S., & REPPEN, R., 1996, "Corpus-based investigations of language use", in *Annual Review of Applied Linguistics*, Cambridge University Press.

- BLANCHE-BENVENISTE, Claire & JEANJEAN, Colette, 1987, *Le Français parlé. Transcription et édition*, Paris, Didier Érudition.
- BLANCHE-BENVENISTE, Claire, 1984, "Connaissance naturelle et connaissance secondaire de la grammaire", *Foreign Languages Teachers Association*, Chicago.
- , 1990, "Grammaire première et grammaire seconde : l'exemple de EN", in *Recherches Sur le Français Parlé*, 10, pp.51-73.
- , 1992, "Les énoncés réponses", in *R.S.F.P.*, 11, pp.57-85.
- BLANCHE-BENVENISTE, Claire, BILGER, Mireille, ROUGET, Christine & van den EYNDE, Karel, 1990, *Le Français parlé. Études grammaticales*, Paris, Éditions du CNRS.
- BLANCHE-BENVENISTE, Claire, DEULOFEU, José, STEFANINI, Jean & van den EYNDE, Karel, 1987, *Pronom et Syntaxe. L'approche pronominale et son application à la langue française*, Paris, SELAF (1^{re} éd. : 1984).
- CAMBON, J. & SINCLAIR, Hermine, 1974, "Relations between syntax and semantics, are they easy to see ?", in *British Journal of Psychology*, 64, 1, pp. 133-140.
- CASTAGNE, Éric, 1995, *Analyse syntaxique des structures nominales intégrant une forme verbale de type "une étude à faire". Étude contrastive avec l'espagnol, l'italien et le portugais*, Thèse doctorat nouveau régime, Université de Provence.
- , 1998, "Utilité de la notion de métonymie pour l'analyse syntaxique des structures du type *une communication intéressante à écouter*", in *Orbis Linguarum*, Legnica, Nauczycielskie kolegium jezykow obcych, pp. 113-120.
- CHOMSKY, C., 1972, "Stages in language development and reading exposure", in *Harvard Educational Review*, 42, pp. 1-33.
- ERIKSSON, Olof, 1993, *La Phrase française*, Göteborg, Acta Universitatis Gothoburgensis.
- HALLE, Morris, BRESNAN, Joan & MILLER, George A. (eds), 1978, *Linguistic Theory and Psychological Reality*, Cambridge, MIT Press.
- HARRIS, Zellig S., 1970, *Papers in structural and Transformational Linguistics*, Dordrecht, D. Reidel.
- (a) [1964] "Elementary transformations";
- (b) [1965] "Transformational Theory";
- (c) [1969] "The two Systems of grammar : Report and Paraphrase".
- HONORÉ, Laurence, thèse en cours sur l'infinitif à l'Université de Provence.
- HUOT, Hélène, 1981, *Constructions infinitives du français. Le subordonnant DE*, Genève, Droz.
- JEANJEAN, Colette, 1984, "Toi quand tu souris. Analyse sémantique et syntaxique d'une structure du français peu étudiée", in *Recherches sur le Français Parlé*, 6, pp. 131-165.

- JESPERSEN, Otto, 1971a, *La Philosophie de la grammaire* (trad. française de A.-M. Léonard), Paris, Éditions de Minuit (1^{re} éd. : 1924, London, George Allen & Unwin Ltd).
- , 1971b, *La Syntaxe analytique* (trad. française de A.-M. Léonard), Paris, Éditions de Minuit (1^{re} éd. : 1937, London, George Allen & Unwin Ltd).
- KAYNE, Richard, 1974-1975, "French relatives QUE", in *Recherches Linguistiques*, 2, pp. 40-61, et 3, pp. 27-92.
- PICABIA, Lélia, 1978, *Les Constructions adjectivales en français. Systématique transformationnel*, Genève, Droz.
- RIEGEL, Martin, PELLAT, Jean-Christophe & RIOUL, René, 1994, *Grammaire méthodique du français*, Paris, P.U.F.
- RUWET, Nicolas, 1982, *Grammaire des insultes et autres études*, Paris, Le Seuil.
- SANFELD, Kr., 1978, *Syntaxe du français contemporain, Tome III : L'infinitif* (3^e éd.), Genève, Droz.
- SMITH, Neil & WILSON, Deidre, 1980, *Modern Linguistics. The Result of Chomsky's Revolution*.
- SOLAN, L., 1978, "Acquisition of Tough Movement and Object Deletion", in *Linguistics*, 4, pp. 127-143.
- STENSTRÖM, A.-B. & SVARTVIK, J., 1994, "Imparsable speech : Repeats and other non-fluencies in English". In N. Oostdijk and P. de Haan (eds) *Corpus-based Research into Language*. Amsterdam, Rodopi, pp. 241-254.
- TOGEBY, Knud, 1982-1985, *Grammaire française* (5 volumes), Copenhague, Akademisk Forlag.
- VENDLER, Zeno, 1968, *Adjectives and Nominalizations*, La Haye-Paris, Mouton.
- WILLEMS, Dominique, 1977, "Recherches en syntaxe verbale : quelques remarques sur la construction absolue", in *Travaux linguistique*, 5, pp. 113-125.

Les constructions comparatives en *plus/moins... que*

Dans la présentation générale du numéro 5 de la revue *Faits de langues*¹ (p. 8), Laurent Danon-Boileau et Mary-Annick Morel relèvent que “l'idée que Jean est plus grand que Paul peut se traduire [...] de deux façons. Ou bien par un verbe, et l'on aboutit alors à ‘Jean dépasse Paul, question taille’, ou bien en se servant de Paul après coup pour fixer un ordre de grandeur à la propriété que l'on reconnaît à Jean. Ce qui conduit à quelque chose comme ‘Jean est grand, à l'échelle de Paul’”. Plusieurs articles contenus dans ce numéro montrent que ces deux façons d'exprimer la comparaison se retrouvent dans de multiples langues.

D. Creissels fait observer qu'en tswana² (p. 42), on peut exprimer une comparaison de supériorité

— soit en recourant à un verbe comme “*go feta*”, signifiant “dépasser” pour établir la relation de supériorité entre les éléments comparés, et préciser par un complément introduit par la préposition “*ka*” (ayant généralement un sens d'instrument, de moyen ou de manière) la caractéristique sur laquelle porte la comparaison.

Gaborone o feta Kanye Ka botona.
Gaborone il dépasse Kanye par grandeur
Gaborone est plus grand que Kanye.

— soit au moyen de deux phrases dont la première affirme une caractérisation se rapportant au premier terme de la comparaison et, la

1. Marie-Annick Morel et Laurent Danon-Boileau (1995), “Présentation générale”, *Faits de langues*, n° 5, pp. 1-12.

2. Denis Creissels (1995), “L'expression de la comparaison dans une langue africaine : l'exemple du tswana”, *Faits de langues*, n° 5, p. 48.

seconde, présentée comme une expansion de la première, exprime la comparaison en prenant le second terme comme repère d'évaluation.

Pitse e tona go na le tonki.
Cheval il est grand être avec âne
Le cheval est plus grand que l'âne.

Le tahitien également semble offrir ces deux possibilités d'exprimer une comparaison d'inégalité. Michel Aufray et Jacqueline de la Fontinelle³ montrent que dans cette langue, "pour prendre la mesure de deux choses" ou "évaluer deux états", on recourt à une comparaison envisagée comme un dépassement en employant le verbe "*hau*" signifiant "dépasser", "dominer", "l'emporter sur".

E hau oia ia'u te `àravirhi.
(acquis) surpasser il moi habileté.
Il est beaucoup plus habile que moi.

Dans cette construction, les deux termes sont des arguments de la valence verbale ("il/moi"), et le registre de la comparaison ("habileté"), fourni après coup, est extérieur à cette valence.

On pourrait également exprimer une comparaison de supériorité en commençant d'abord par poser l'existence de la propriété sur laquelle porte la comparaison, puis l'appliquer au premier terme de la comparaison et enfin énoncer l'existence d'un repère que représente le second terme de la comparaison sans fournir de réflexion à son sujet.

E Mea rahi a'e te pahi i te poti.
(prédication) chose grande a'e — navire — canot
Un navire est plus grand qu'un canot.

Michel Aufray et Jacqueline Fontinelle glosent cet exemple comme suit :

Il existe une grandeur — je l'applique à un objet existant ("le navire")
— il y a un objet repéré ("le canot") mais je ne fais aucun commentaire à son sujet.

Dans cette seconde construction, c'est le premier terme et la propriété sur laquelle porte la comparaison qui sont regroupés. Le second terme de la comparaison est réalisé en dehors de la valence verbale.

3. Michel Aufray et Jacqueline de La Fontinelle (1995), "La comparaison dans les langues d'Océanie : rupture, rapprochement, éloignement", *Faits de langues* n° 5, pp. 53-56.

Ces deux types de constructions comparatives existent également en latin où l'on peut exprimer le comparatif soit par la particule *quam* soit par l'ablatif.

Émile Benveniste montre dans son ouvrage *Noms d'agent et noms d'action en indo-européen*⁴ que l'ablatif latin s'explique à partir de l'ablatif indo-européen qui permettait d'exprimer un type de comparatif dans lequel l'un des termes en comparaison est pris comme repère à partir duquel s'effectue la quantification. Des exemples sont fournis en hittite ou en darde.

- Hittite

– *Ana erinmes – ka erinmes – ia mekki.*
 À tes fantassins mes fantassins sont nombreux.
 Mes fantassins sont plus nombreux que les tiens.

- Darde

– *Aïto-ta bâlô thum.*
 moi toi – de grand suis (moi, par rapport à toi, suis grand)
 "je suis plus grand que toi".

En latin, l'ablatif indique le terme de référence pris comme modèle, comme étalon de la propriété, à partir duquel est évaluée la caractéristique de l'objet comparé. Dire d'un objet qu'il est "plus clair que la lumière" (*luce clarior*), c'est dire qu'il possède la qualité de clarté à un degré supérieur à celui de ce qui est clair par excellence c'est-à-dire la "lumière".

La construction avec la particule *quam*, qui met en liaison deux énoncés, insiste sur le parallélisme des termes comparés qui sont appréciés l'un et l'autre par rapport à la propriété sur laquelle porte la comparaison. Comme l'indique Charles de Lamberterie⁵, l'opération mentale consiste à poser une notion et "à mettre en balance deux choses auxquelles cette notion pourrait s'appliquer". Ainsi, si l'on réalise l'idée que "Pierre est plus grand que Paul" par la construction avec *quam*, on choisit d'apprécier *Pierre* et *Paul* par rapport à la qualité de grandeur. L'ablatif ne serait acceptable que si *Paul* constituait un étalon de la grandeur, un géant connu de tout le monde.

Ces deux façons d'exprimer une comparaison d'inégalité que l'on relève dans différentes langues ne s'observent pas directement en français.

4. Emile Benveniste (1948), *Noms d'agent et noms d'action en indo-européen*.

5. Charles Lamberterie (1995), "L'apport d'Emile Benveniste", *Faits de langues*, n° 5, p. 17.

Certes, les locutions prépositives comme “par rapport à”, “à côté de” pourraient permettre de réaliser une construction du genre :

Par rapport à Paul, Pierre est grand.
À côté de Paul

construction dans laquelle l'un des termes de comparaison est évalué à partir de l'autre terme pris comme repère. La réalisation d'une telle construction ne pose aucune difficulté si on a affaire à la position “sujet”. Cependant, elle est très peu commode pour la position “objet” par exemple.

- ? Par rapport à Paul, nous parlerons plus à Pierre.
? À côté de Paul
- ? Par rapport à Paul, nous parlerons plus de Pierre.
? À côté de Paul

Pour ces positions, il paraît nécessaire de recourir à la locution “plus... que”.

- Nous parlerons plus à Paul qu'à Pierre.
- Nous parlerons plus de Paul que de Pierre.

Ainsi, pour exprimer l'inégalité, le français ne semble disposer que d'une seule structure grammaticalisée en “plus/moins... que”. Nous faisons l'hypothèse qu'en réalité derrière cette seule forme morphologique “plus... que”, il y a deux structures syntaxiques différentes permettant d'exprimer les deux relations conceptuelles de comparaison observées jusqu'ici.

1. STRUCTURE SYNTAXIQUE DE LA RELATION PRIVILÉGIANT N1, N2 N'ÉTANT QU'UNE RÉFÉRENCE (RELATION I)

Prenons pour exemple l'énoncé :

- Je parlerai plus à Paul qu'à Pierre.

On peut décrire la forme de la relation entre les deux termes comparés ainsi que leur différence de statut par les critères de l'interrogation, de la proportionnalité, du jeu de modalités et de l'extraction⁶.

6. Cf. Claire Blanche-Benveniste et alii (1984), *Pronom et syntaxe. L'approche pronominale et son application à la langue française*, pp. 64-67.

1.1. L'INTERROGATION

Les éléments contenus dans le syntagme “plus à Paul qu'à Pierre” répondent à deux questions : “À qui ?” et “Comment ?” (ou “Combien ?”).

Dans le cas de la relation 1, seul “à Paul” répond à la première question. Il est seul à occuper la position de valence d’“objet indirect”. La réponse à la seconde question serait fournie par tout le reste du syntagme “plus qu'à Pierre” occupant une position de rection de manière. L'accent est mis sur *Paul* et sur l'appréciation faite à son propos. *Pierre* est présenté comme une référence permettant de fournir cette appréciation.

1.2. LA PROPORTIONNALITÉ

La présence de la relation est vérifiée par la possibilité de réduire *Paul* à un élément pronominal clitique.

- Je lui parlerai plus qu'à Pierre.

Une telle possibilité démontre l'autonomie de “à Paul”. On peut également établir une relation de proportionnalité entre la séquence “plus qu'à Pierre” et des compléments de type “ainsi”.

- Je parlerai à Paul plus qu'à Pierre.
 ainsi
 de cette manière
 à ce point

L'énoncé pourrait alors être glosé comme suit :

Concernant la manière dont je vais parler à Paul, je lui parlerai plus qu'à Pierre.

Cette proportionnalité indique que “plus qu'à Pierre” est à interpréter comme un élément de rection, donc sur un tout autre plan que “à Paul”.

1.3. LE JEU DE MODALITÉS

La séquence “plus qu'à Pierre” pourrait être soumise à une négation de la manière suivante :

- Je ne lui parlerai pas plus qu'à Pierre.

On pourrait également avoir un contraste du type :

- Je lui parlerai mais plus qu'à Pierre.
- Je lui parlerai et plus qu'à Pierre.

Nous savons que dans ce type de construction que Mireille Bilger⁷ analyse comme ayant un effet de surenchère, il y a un premier élément à réalisation zéro et auquel *mais* et *et* relie la séquence subséquente. Cet élément non réalisé mais quand même présent dans la structure a une valeur d'“indétermination”. On pourrait le gloser ici respectivement par *peu* et *beaucoup* comme suit :

- Je lui parlerai peu mais (quand même) plus qu'à Pierre.
- Je lui parlerai beaucoup et (même) plus qu'à Pierre.

1.4. L'EXTRACTION

La différence de statut entre N1 et N2 pourrait enfin être vérifiée par la possibilité d'extraire “plus qu'à Pierre” par le dispositif “c'est... que”.

- C'est plus qu'à Pierre que je lui parlerai.

Ces différents tests que nous venons d'appliquer montrent bien que les deux termes de comparaison sont rattachés différemment au verbe. Ils ont des fonctions syntaxiques différentes, le premier étant un complément de type “objet” et le second, contenu dans la séquence introduite par “plus... que” étant un complément de type réaction de manière.

2. STRUCTURE SYNTAXIQUE DE LA RELATION PRIVILÉGIANT LE RAPPORT ENTRE N1 ET N2, TOUS LES DEUX ÉTANT PLACÉS AU MÊME NIVEAU

Les critères de la proportionnalité, de l'interrogation et de l'extraction pourraient, ici également, nous permettre de décrire la structure syntaxique de cette relation.

2.1. LA PROPORTIONNALITÉ

En effet, pour cette seconde relation, *Paul* et *Pierre* seraient proportionnels aux indéfinis *l'un* et *l'autre*. L'énoncé se gloserait ainsi :

- (Entre Paul et Pierre), je parlerai plus à l'un qu'à l'autre.

On a également une proportionnalité des deux termes avec le pronom *leur*.

7. Cf. Mireille Bilger (1983), *Étude distributionnelle de la coordination par ET*, Thèse de troisième cycle, Université de Provence.

- Je leur parlerai plus à l'un qu'à l'autre.

L'accent est mis cette fois sur le rapport entre *Paul* et *Pierre*. La comparaison est articulée sur les termes en contraste et non sur les valeurs quantifiées qui leur seraient liées et dont l'une fonctionnerait comme repère ainsi que nous l'avons vu dans la relation précédente.

L'énoncé tire plus vers une réalisation avec "plutôt... que" sans que pour autant l'effet de quantification cesse d'exister.

- Je parlerai plutôt à Paul qu'à Pierre.

2.2. L'EXTRACTION

Cette seconde relation donne un regroupement des constituants différent de celui qu'on a dans la première relation. Pour cette première relation, en effet, on avait observé que le dispositif "c'est... que" ne prenait que la séquence introduite par "plus... que" et contenant N2 ("c'est plus qu'à Pierre que..."). Pour la deuxième relation par contre, N1 et N2 sont unis dans un même syntagme qu'on pourrait extraire globalement au moyen de ce dispositif :

- C'est plus à Paul qu'à Pierre que je parlerai.

Ainsi, on a un regroupement qu'on pourrait schématiser comme suit :

— SV / PLUS À PAUL QU'À PIERRE /

Alors que pour la relation 1, on aurait :

- (... à Paul) SV / plus qu'à Pierre /

2.3. L'INTERROGATION.

Dans une telle relation, N1 et N2 répondent tous les deux à la question "à qui ?". Ils occupent ainsi une même position de valence. La réponse à la question "comment ?" qui serait "plus à l'un qu'à l'autre" ne fournit pas une position syntaxique mais un mode d'organisation des éléments de valence que sont *l'un* et *l'autre*. La structure "plus... que" organise ainsi ces éléments de valence avec un effet de sens comparatif. Elle n'occupe pas une place syntaxique en dehors de ces éléments ; on est proche d'une construction avec "soit... soit" par exemple.

Je parlerai soit à Paul soit à Pierre.

Je leur parlerai soit à Paul soit à Pierre.

Comme “soit ... soit”, la locution “plus ... que” a ici le statut d'un joncteur de liste⁸. Les éléments qu'elle lie sont deux réalisations lexicales d'une seule et même place syntaxique. La place en question est celle d'un complément d'“objet indirect”. Dans une configuration en grille, ces éléments seraient représentés en liste comme suit :

je parlerai plus à Paul
qu'à Pierre

Il s'agit ainsi d'une relation paradigmatique, celle qu'on retrouve dans le phénomène dénommé habituellement “coordination” et que permettent en fait, en dehors des conjonctions de coordination, une bonne partie d'éléments de liaison étiquetés comme “adverbes”, “prépositions” ou parfois “conjonctions de subordination”.

3. RÉALISATIONS ORIENTÉES VERS UNE SEULE INTERPRÉTATION

Les deux relations que nous venons d'observer et que permet une construction avec “plus... que” sont le plus souvent difficiles à discriminer. Il existe cependant des cas où la construction s'oriente vers une lecture plutôt que vers une autre.

3.1. LORSQUE LES ÉLÉMENTS EN CONTRASTE SONT DES ADJECTIFS COMME DANS :

Il est plus beau que gentil.

Ici, il paraît difficile d'avoir:

? Il l'est plus que gentil.

Autrement dit, on n'a pas une place de valence qu'occuperait le clitique “I” proportionnel à “beau” et une place de rection qu'occuperait la séquence “plus que gentil”.

D'un point de vue sémantique, la “beauté” n'est pas évaluée par rapport à un élément repère qui serait la “gentillesse”. Il s'agit plutôt d'une confrontation entre la “beauté” et la “gentillesse”, les deux notions étant

8. Cf. Modou Ndiaye (1989), *L'analyse syntaxique par joncteurs de liste*, thèse nouveau régime, Université de Provence.

mises en balance. Les termes comparés sont sur une même position syntaxique. On a une structure de liste⁹.

3.2. LORSQU'ON A AFFAIRE À UNE POSITION "OBJET DIRECT" À ÉLÉMENT "DE" COMME DANS :

- Je vous souhaite plus de chance que de courage.
- Nous achèterons plus de journaux que de livres.
- Elle avait plus de soucis que de joie.

Il paraît difficile d'avoir :

- ?Je vous en souhaite plus que de courage.
- ? Nous en achèterons plus que de livres.
- ?Elle en avait plus que de joie.

Pour ce type de réalisations, la réédition de l'élément "de" semble imposer une lecture selon laquelle les termes comparés sont mis en parallèle et donc situés sur une même place syntaxique.

4. RÉALISATIONS À DOUBLE INTERPRÉTATION

En dehors de ces cas, l'ambiguïté de la construction en "plus ... que" est observable en différentes positions de valence ou de rection.

— POSITIONS DE VALENCE

POSITION "OBJET DIRECT"

Il aime plus le beurre que le fromage.

Pour la première relation, cet énoncé correspond à la réalisation :

- (Le beurre) il l'aime plus que le fromage.

Il permet le contraste :

- Il l'aime mais pas plus que le fromage.

et répond à la question : "À quel point aime t-il le beurre ?".

9. Cette notion de "structure de liste" appliquée aux constructions comparatives avait déjà été abordée dans la thèse citée ci-dessus et a bénéficié des remarques pertinentes de D. Willems.

Par ailleurs, la séquence “plus que le fromage” est proportionnelle à des éléments comme “ainsi”, “à ce point”.

- Il l'aime plus que le fromage.
 ainsi
 à ce point

D'un point de vue sémantique, la priorité est portée sur le premier terme de comparaison (“beurre”) à propos duquel est faite une appréciation. Le second élément de comparaison (“fromage”) sert de repère à cette appréciation.

Pour la deuxième relation, l'énoncé répond à la question :

- (Entre le beurre et le fromage) lequel aime t-il le plus ?

La priorité n'est pas accordée au premier terme mais plutôt au rapport entre les deux termes. L'énoncé correspond à une réalisation avec “l'un”, “l'autre”.

- Il aime plus l'un que l'autre.

On a un parallélisme des deux termes comparés, qui sont situés sur une même place d'“objet”.

POSITION D'“OBJET PRÉPOSITIONNEL”

Nous parlions plus de Paul que de Pierre.

Pour la relation 1, l'énoncé répond à la question.

- Combien parlions -nous de Paul ?

Il répond à la réalisation :

- De lui, nous parlions à cette fréquence.
 ainsi
 plus que de Pierre

Il permet les contrastes :

- Nous parlions de lui beaucoup (certes) mais pas plus que de Pierre.
- Nous parlions de lui peu mais (quand même) plus que de Pierre

Pour la relation 2, l'énoncé répond à la question :

- (Entre Paul et Pierre) de qui parlions-nous le plus ?

Il correspond alors à la réalisation.

- Nous parlions plus de l'un que l'autre.

Paul et Pierre sont ainsi tous les deux proportionnels à l'interrogatif "de qui ?". Ils occupent donc la même position syntaxique.

POSITION SUJET :

- Il mange plus que moi.
- Elle travaille plus que son mari.

S'agissant de la relation 1, on pourrait établir une relation de proportionnalité entre la séquence "plus N2" et l'interrogatif "combien ?". Ces énoncés répondent ainsi respectivement aux questions :

- Combien mange t-il ?
- Combien travaille t-elle ?

La séquence pourrait être extraite au moyen de "c'est... que".

- C'est plus que moi qu'il mange.
- C'est plus que son mari qu'elle travaille.

On pourrait également la soumettre à une négation.

- Il ne mange pas plus que moi.
- Elle ne travaille pas plus que son mari.

ou à un contraste de modalités comme suit :

- Il mange beaucoup mais pas plus que moi
- Il mange (certes) peu mais (quand même) plus que moi.
- Elle travaille beaucoup mais pas plus que son mari.
- Elle travaille peu mais (quand même) plus que son mari.

Sémantiquement, la séquence "plus que N2" permet d'exprimer une quantification sur l'assertion faite à propos de N1 ; N2 est pris comme une référence, un étalon. C'est cette relation qu'on aurait dans :

- Il mange plus qu'un porc.

Quant à la relation 2, les deux énoncés répondent respectivement aux questions :

- (Entre Pierre et moi) qui mange le plus ?
- (Entre elle et son mari) qui travaille le plus ?

La construction privilégie la confrontation des deux termes. Le second est ainsi à placer au même niveau que le premier terme. On a une sorte de coordination différée. C'est la portée de "plus" qui empêche de réaliser le second terme immédiatement après le premier comme cela est observé pour les autres positions. On n'a pas un tel obstacle dans des exemples comme :

1. – Vinrent plus de femme que d'hommes.
– Plus de femmes que d'hommes vinrent.
– Plus de femmes vinrent que d'hommes.
2. – Plus de robes que de chaussures ont été achetées.
– Plus de robes ont été achetées que de chaussures.
– Ont été achetées plus de robes que de chaussures.

POSITIONS DE RECTION :

Rection locative :

- Le climat est plus doux à Paris qu'à Rome.
- Il pleut plus à Paris qu'à Rome.

Pour la relation 1, ces deux énoncés contenant une rection locative répondent respectivement aux questions :

- À quel degré le climat est-il doux à Paris ?
- À quelle fréquence pleut-il à Paris ?
Combien

On a, alors, correspondant à ces deux énoncés, les réalisations :

- (À Paris) Le climat y est plus doux qu'à Rome.
- (À Paris) Il y pleut plus qu'à Rome.

La séquence "plus qu'à Rome" pourrait aussi être niée.

- Le climat n'y est pas plus doux qu'à Rome.
- Il n'y pleut pas plus qu'à Rome.

Pour la relation 2, les deux énoncés répondent respectivement aux questions :

- (Entre Paris et Rome) où le climat est-il plus doux ?
- (Entre Paris et Rome) où pleut-il plus ?

Ainsi, *Paris et Rome* occupent tous les deux la rection locative indiquée par *où*. D'un point de vue sémantique, l'accent est mis sur la confrontation entre les deux termes comparés et non plus seulement sur le premier terme. Les deux énoncés correspondent alors aux deux réalisations :

- Le climat est plus doux ici que là.
- Il pleut plus ici que là.

Rection temporelle

- Il est plus riche maintenant qu'autrefois.

Pour la relation 1, cet énoncé répond à la question :

- À quel degré est-il riche maintenant ?

Il correspond à la réalisation :

- Maintenant, il est plus riche qu'autrefois.

On a les contrastes suivants :

- Maintenant, il est certes riche mais pas plus qu'autrefois.

Ainsi, seul le premier élément comparé (*maintenant*) occupe la position de réaction temporelle. Le second (*autrefois*) est situé dans la séquence introduite par *plus* et fonctionnant comme un complément proportionnel à un groupe comme “à ce degré”, “à ce point”.

- (Maintenant) Il est riche plus qu'autrefois.
à ce degré
à ce point

Ainsi d'un point de vue syntaxique, la séquence “plus qu'autrefois” ne serait pas différente d'une séquence comme “plus que son frère” ou “plus que ça” ayant cette même proportionnalité.

- (Maintenant) Il est plus riche qu'autrefois
que son frère
que ça

Sémantiquement, l'opération de pensée consiste à exprimer un jugement se rapportant au premier terme de comparaison, et à se servir du second terme comme référence de quantification de ce jugement. C'est cette hiérarchie sémantique qui permet une relativation du genre :

- Maintenant qu'il est plus riche qu'autrefois (il peut se permettre...)

Il faut souligner que pour cette relation 1, ce premier terme peut facilement ne pas être réalisé ni lexicalement ni pronominalement comme dans :

- Il est plus riche qu'autrefois.

Il est cependant présent comme élément du temps de l'énonciation. C'est cela qui explique la difficulté d'avoir un énoncé comme :

- Il est plus riche qu'aujourd'hui.¹⁰

10. Un tel énoncé ne serait possible que dans un contexte tel que : “sur cette photo, il est plus riche qu'aujourd'hui”.

où l'élément non réalisé (*maintenant*) qu'il faudrait alors interpréter à partir du verbe avec un effet de présent serait en contraste avec un élément de référence exprimant l'actuel (*aujourd'hui*).

Pour la relation 2, l'énoncé répond à la question "Quand est-il plus riche ?".

Les termes comparés occupent tous les deux la position de rection temporelle marquée par *quand*. "Plus... que" permet de les relier avec un effet de sens comparatif. La priorité est ainsi mise sur la confrontation des deux termes et non plus sur l'évaluation du premier terme. Cette relation interdit le déplacement de *maintenant* en début d'énoncé, de même que le remplacement de *autrefois* par un élément non temporel, procédures qui aboutiraient toutes les deux à la première relation.

Ce parcours des différentes positions syntaxiques contrôlables à partir du verbe prouve bien que l'ambiguïté que l'on peut relever dans les comparatives étudiées ici n'est pas un simple fait de contexte mais qu'elle est inscrite dans la forme "plus ... que" qui recouvre deux structures syntaxiques distinctes.

5. RELATIONS DANS LA COMPARATIVE TRONQUÉE

La comparative tronquée correspond à la construction dans laquelle le second terme de comparaison n'est pas réalisé. Elle est d'une fréquence remarquable dans les exemples réels de constructions comparatives que nous avons récoltés et qui proviennent pour l'essentiel des corpus de français parlé du GARS¹¹ et de la presse écrite (*Le Monde*). Il s'agit d'énoncés du genre :

- ouais je me vois plus euh – plus dans dans un bureau et faire faire de la dé (Nouguier 60, 3)
- L'ersatz ce vieux mot justement d'origine allemande, serait plus compréhensible au commun des bacheliers. (*Le Monde*, 16.09.97, p. 34)

La question est de savoir si l'ambiguïté de la structure comparative telle que nous avons essayé de l'expliquer demeure avec ce type de construction ou si la troncation tend à privilégier l'une quelconque des deux relations. Il faudra, pour cela, déterminer le terme de l'énoncé avec lequel le second élément non réalisé est en relation, et préciser la nature de la relation. Nous établirons l'interprétation fournie par la construction

11. Corpus oraux et transcrits du Groupe Aixois de Recherche en Syntaxe.

elle-même par l'examen de quelques énoncés isolés avant de vérifier les possibilités offertes par une analyse en contexte.

5.1. DIFFICULTÉS D'INTERPRÉTATION

5.1.1. LES TERMES EN RELATION

Dans un énoncé isolé de son contexte, il est souvent difficile d'identifier le terme avec lequel l'élément non réalisé est en relation. Il peut, en effet, s'agir de n'importe lequel des termes de valence ou de rection figurant dans l'énoncé et parfois même d'un terme de rection temporelle non réalisé. Ainsi :

- 1^o – il faudrait qu'on y passe plus de temps (Nouguier, 95, 11)

peut correspondre soit à :

- il faudrait qu'on y passe plus de temps (que vous)
que vous ne l'avez fait
que vous ne le ferez
que nous ne l'avons fait

l'élément non réalisé étant en relation avec le terme P0

soit à :

- il faudrait qu'on y passe plus de temps (qu'à Marseille)

l'élément non réalisé étant en relation, cette fois, avec la rection locative.

soit à :

- il faudrait qu'on y passe plus de temps qu'autrefois

l'élément non réalisé étant en relation avec une rection temporelle elle-même non réalisée et qui serait *maintenant*.

- 2^o – il y a plus de SDF de cinquante cinquante-cinq ans soixante (Nouguier 102, 7)

pourrait correspondre aussi bien à :

- il y a plus de SDF de cinquante ans (que de SDF jeunes)

qu'à :

- il y a plus de SDF de cinquante ans (qu'autrefois)

- 3^o – c'est pour ça qu'en fait on travaille plus (Nouguier, 73, 8)

pourrait correspondre à :

- c'est pour ça qu'en fait on travaille plus (que nos concurrents)

ou bien à :

- c'est pour ça qu'en fait on travaille plus (qu'autrefois)

Cependant, il est difficile de mettre l'élément non réalisé en relation avec une valence adjectivale ou une réaction adverbiale. En effet, l'énoncé :

- 1° – on me dit que je suis plus tolérante (Nouguier, 10, 13)

renvoie soit à une relation avec le sujet :

- on me dit que je suis plus tolérante (que lui).

ou à une relation avec la réaction temporelle :

- on me dit que je suis plus tolérante (qu'autrefois)

mais il ne peut renvoyer à :

- on me dit que je suis plus tolérante (que gentille).

- 2° – Paul parle plus calmement

renvoie soit à

Paul parle plus calmement (que Jean).

soit à :

- Paul parle plus calmement (qu'autrefois).

mais, il ne peut renvoyer à :

- Paul parle plus calmement (qu'intelligemment)

Ainsi, dans un énoncé hors contexte, le terme avec lequel l'élément non réalisé est en relation n'est pas identifiable avec précision. Parfois cependant quelques indices comme, par exemple, la nature du verbe pourraient orienter vers un choix plus précis :

- Paul est devenu plus sage.

Le verbe *devenir* semble orienter vers une relation avec la réaction temporelle non réalisée (*maintenant*) plutôt que vers l'élément sujet.

- Paul est devenu plus sage maintenant (qu'autrefois).

5.1.2. LA NATURE DE LA RELATION :

Il y a une égale difficulté à déterminer la nature de la relation. En effet, dans ce type de construction, l'interprétation tire plus vers la relation 2, c'est-à-dire celle dans laquelle l'accent est mis sur la confrontation

des termes en comparaison plutôt que sur la quantification du premier terme par le second pris comme repère. Si l'on considère, par exemple, l'énoncé :

- les échanges internationaux ont augmenté plus vite (*Economi*, 4, 15)

quel que soit le terme avec lequel l'élément non réalisé est en relation¹², on peut observer que cette relation se déroule sur l'axe paradigmatique. En effet, on peut recourir au dispositif "c'est ... que", qui, comme on le sait, ouvre un paradigme, pour rendre les interprétations attendues comme suit :

1) L'élément non réalisé est en relation avec le terme sujet. Cela correspondrait à l'énoncé :

- Les échanges internationaux ont augmenté plus vite que la production mondiale.

L'interprétation attendue, dans ce cas, serait :

- Ce sont les échanges internationaux qui ont augmenté plus vite (non la production mondiale)

2) L'élément non réalisé est en relation avec le terme temporel contenu dans le verbe et paraphrasable, par exemple, par *maintenant*. Cela correspondrait à l'énoncé :

- Les échanges internationaux ont augmenté plus vite maintenant qu'autrefois.

En ce cas, l'interprétation attendue serait :

- C'est maintenant que les échanges internationaux ont augmenté plus vite.

Ainsi, dans la comparative tronquée, l'élément non réalisé n'apparaît pas comme une expansion du quantifieur avec lequel il formerait un complément de degré comme cela est le cas dans la relation 1. L'énoncé ci-dessus ne correspond pas à :

- Les échanges internationaux ont augmenté à ce rythme.

En d'autres termes, il ne répond pas à la question :

- À quel rythme les échanges internationaux ont-ils changé ?

12. Il s'agit, en général, d'un terme de valence ou de rection figurant dans l'énoncé ou sinon d'un terme de rection temporelle non réalisé.

On pourrait faire les mêmes observations concernant l'énoncé suivant :

- Le climat est plus doux à Paris.

dont l'interprétation correspondrait à :

- C'est à Paris qu'il fait plus doux.

et non à :

- Le climat est doux à Paris à ce degré.

Autrement dit l'énoncé n'est pas une réponse à la question :

- À quel degré le climat est-il doux à Paris ?

5.2. INTERPRÉTATION ORIENTÉE PAR LE CONTEXTE

5.2.1. LES TERMES EN RELATION

Dans un énoncé en contexte, il n'y a aucune difficulté à repérer le terme avec lequel l'élément non réalisé est en relation. Sunniva Whittaker a montré dans un intéressant article¹³ (p. 169) qu'il y a presque toujours des indices permettant d'inférer "l'étalon omis". Dans les constructions qui l'ont intéressé le plus, et qui sont celles où "l'étalon ouvre un espace mental", cet auteur observe que "la phrase contenant la construction comparative tronquée contient en règle générale des indications concernant le type d'espace mental dont il est question". Il s'agit souvent d'un complément circonstanciel qui établit le premier espace mental. Ainsi, dans des énoncés comme :

- Avec la réintroduction du divorce, les raisons d'adopter cette dernière procédure [...] étaient plus impérieuses encore. (*Histoires des passions franç.*, 1, p.339, cité par Sunniva, p. 170)
- Depuis 20 ans ses recherches se sont orientées vers un objectif plus fondamental. (*Sciences hum.*, nov., 1990, p. 37 ; cité par Sunniva, p. 171)

l'élément non réalisé ouvre un espace temporel parce qu'il faut l'interpréter à partir, respectivement, des expressions temporelles : "avec la réintroduction" et "depuis 20 ans".

Parfois, cependant, d'autres éléments de l'énoncé peuvent permettre cette identification. Dans les exemples :

13. Sunniva Whittaker (1995), "À la recherche de l'étalon", *Faits de langues*, n° 5, pp.165-174.

- Recours face aux duettistes Giscard et Chirac, incarnation d'une nouvelle droite moins arrogante et plus ouverte. (*Libération*, 5-6.12.1992, p. 7 ; cité par Sunniva, p. 173)
- le romantisme a peut-être rendu les prostituées encore plus nécessaires. (*Histoires des passions franç.*, 1, p. 339, cité par Sunniva, p. 172)

c'est l'adjectif "nouvelle" dans le premier énoncé et le substantif "romantisme" dans le second qui indiquent que l'élément non réalisé est temporel.

Comme on le voit, ces interprétations ne s'appuient pas seulement sur des argumentations grammaticales à proprement parler, mais aussi sur des indices lexicaux ou sur une série d'inférences.

Si l'on s'intéresse à des exemples dans lesquels on a des éléments reactionnels autres que temporels, on s'aperçoit que, le plus souvent, la présence du complément reactionnel réalisé n'est pas une condition suffisante pour l'interprétation du second complément non réalisé. En général, il est nécessaire que ce dernier soit déjà exprimé dans un énoncé antérieur. C'est ce que l'on peut observer dans les exemples suivants :

- Sur l'ensemble de la saison, la jeune américaine (Marion Jones) mène devant la jamaïcaine trois victoires à deux sur 100 m. Pire, Marion Jones est plus impressionnante sur 200 m et mène trois à zéro dans le face à face. (*Le Monde*, 16.09.1997, p. 23)
- Pour les hommes, l'exigence d'avoir à tout prix un enfant biologique semble s'être accrue considérablement ([...]. Pour les femmes, cette exigence d'une filiation biologique est moins forte. (*Le Monde*, 16.09.97, p. 17)

En effet, dans le premier exemple, la dernière phrase est interprétée comme suit :

- [...] Marion Jones est plus impressionnante sur 200 m que sur 100 m.

à partir de l'indication contenue dans la phrase précédente. En l'absence d'une telle indication, l'élément reactionnel "sur 200 m" n'aurait pas suffi pour exclure l'interprétation :

- Marion Jones est plus impressionnante que son adversaire sur 200 m.

De même, dans le second exemple, le deuxième élément de comparaison non réalisé est interprété comme correspondant à la séquence "pour les hommes" déjà exprimée dans la phrase précédente. On n'aurait pas une telle précision si on se limitait au cadre strict de la dernière phrase contenant la structure comparative. En effet, la séquence reactionnelle "pour les femmes", malgré par ailleurs sa charge sémantique, n'aurait pas permis d'écarter l'interprétation du genre :

- Pour les femmes, cette exigence d'une filiation biologique est moins forte qu'avant.

Ainsi, en l'absence d'indication suffisamment nette dans le contexte antérieur, il peut y avoir parfois quelques difficultés d'interprétation à partir du texte même comme dans l'exemple suivant.

- L'asile constitutionnel ne devrait concerner, selon les spécialistes, que quelques dizaines de personnes chaque année. L'asile territorial sera accordé par le ministre de l'intérieur sans aucune voie de recours efficace et rapide. Quant à l'asile classique, le projet en limite l'accès en proposant une procédure simplifiée et **moins** protectrice pour les demandeurs issus de pays jugés sans risques. (*Le Monde*, 16.09.97, p. 10)

Faut-il comprendre :

- [...] moins protectrice pour les demandeurs issus de pays jugés sans risques **que pour ceux issus de pays à risques.**

interprétation plausible parce qu'au début de l'article, il est fait mention que l'asile classique n'était réservé qu'aux personnes menacées par leur état, ou

[...] procédure simplifiée et moins protectrice pour les demandeurs issus de pays jugés sans risques **que la procédure d'avant.**

Pour les compléments temporels examinés par Sunniva, il y a lieu de remarquer que si on a affaire à un terme temporel ponctuel, il ne pourrait suffire pour l'interprétation du second terme de comparaison non réalisé. C'est le cas dans un exemple comme

- En 1997, Merlène Ottey était un peu plus tranquille.

pour lequel, on peut hésiter entre :

- Merlène Ottey était un peu plus tranquille qu'avant 1997.

et

- Merlène Ottey était un peu plus tranquille qu'après 1997.

ou peut-être même :

- Merlène Ottey était un peu plus tranquille que Marion Jones.

Il faut alors recourir au contexte antérieur pour retrouver les éléments nécessaires à une interprétation exacte. Ce contexte est le suivant :

- Aux jeux olympiques d'Atlanta, elle avait terminé deuxième du 100 mètres derrière l'Américaine Gail Devers [...]. Aux Mondiaux de 1993, elle s'était inclinée d'un millième de seconde devant Gail Devers. (*Le Monde*, 16.09.1997, p. 23)

On a donc l'interprétation :

- Merlène Ottey était un peu plus tranquille en 1997 qu'aux Jeux Olympiques ou aux Mondiaux de 1997.

Les compléments temporels qui permettent d'interpréter le second élément de comparaison non réalisé sans qu'il soit nécessaire de sortir de la construction elle-même correspondent en général à des expressions temporelles non ponctuelles, c'est-à-dire couvrant une certaine étendue comme cela est le cas dans les exemples de Sunniva. Ils sont du genre :

- Depuis + événement ou date
- Avec + événement
- Désormais

Il y a lieu d'observer qu'avec certains de ces compléments temporels, la restitution du second terme de comparaison paraît difficile. Sa non réalisation semble être liée à la nature même de ces compléments plutôt qu'à un souci d'éviter une redondance.

- ? Depuis 20 ans ses recherches se sont orientées vers un objectif plus fondamental **qu'avant**.

Il y a également lieu de constater que la construction peut parfois empêcher la restitution du premier terme même, qui, comme on l'a déjà signalé, lorsqu'il correspond à un complément temporel peut ne pas être réalisé.

- quand j'étais beaucoup plus jeune + et j'ai : je rigolais tout ça quand il essaya (Nouguier, 63, 12)

Dans cet exemple, le premier terme correspondrait à *avant* ; il est contenu dans le verbe sous forme de passé. Et le second terme correspond à *maintenant*. Cependant, la présence de *quand* empêche de faire une glose du genre :

- * Quand j'étais beaucoup plus jeune **avant** que maintenant

Seul le second terme pourrait être restitué.

- Quand j'étais beaucoup plus jeune que maintenant

Le complément temporel peut n'avoir aucune étendue mais détenir une sorte de polarité à partir de laquelle l'élément non réalisé sera interprété.

- Dans l'après-midi [...], le vent du sud-ouest deviendra plus soutenu. (*Le Monde*, 16.09.97, p. 26)

- Ce sera encore une belle journée. Les nappes de brouillard seront un peu plus fréquentes au lever du jour. (*Le Monde*, 16.09.97, p. 26)

Ces exemples sont interprétés à partir des oppositions habituelles : matinée/après-midi ; lever du jour/reste de la journée.

- Dans l'après-midi [...], le vent du sud-ouest deviendra plus soutenu **que dans la matinée**.
- Ce sera encore une belle journée. Les nappes de brouillard seront un peu plus fréquentes au lever du jour **que dans le reste de la journée**.

Lorsque l'élément non réalisé est un terme de valence, il y a toujours un indice formel dans le contexte antérieur permettant de l'interpréter.

Valence sujet :

- Traitant des **emplois jeunes**, Alain Madelin les a qualifiés "d'ersatz d'emplois" [...]. Voilà qui va sans doute beaucoup parler aux plus jeunes, les plus concernés. "L'ersatz", ce vieux mot justement d'origine allemande, serait **plus** compréhensible au commun des bacheliers. (*Le Monde*, 16.09.1997, p. 34)

Le contexte antérieur fournit l'indice "emplois-jeunes" comme second terme de la comparaison :

- [...] serait plus compréhensible au commun des bacheliers qu'aux jeunes les plus concernés que le terme "emploi-jeune".

Valence objet direct :

- [...] Mais je ne sais si **cette éventualité** a même effleuré l'esprit de notre couple qui a trouvé beaucoup **plus** simple de piocher dans le sac à malices des techniques dérivées de l'assistance médicale à la procréation.

L'interprétation serait :

- [...] qui a trouvé beaucoup **plus** simple de piocher dans le sac à malices des techniques dérivées de l'assistance médicale à la procréation **que de recourir à cette éventualité**.

Valence objet indirect :

- En 1967, Aragon me dédicça *Blanche ou l'oubli* avec ces mots à contresens du titre de son livre : "Pour nos souvenirs inoubliables, ton ami Louis."
"D'aucun de mes écrits, je ne suis **plus** fier", écrira Aragon. (*Le Monde*, 24.09.97, p. 9)

On a l'interprétation :

- D'aucun de mes écrits, je ne suis **plus fier que de Blanche ou l'oubli**

En l'absence d'indice formel permettant d'identifier l'élément non réalisé, celui-ci est interprété comme correspondant à un temporel mis en parallèle à un premier terme temporel contenu dans le verbe et également non réalisé.

- Ayant senti que la discussion irait à l'échec, Alain Juppé avait demandé à ses ministres de la fonction publique de ne parler que d'aménagement du temps de travail : il proposait d'ouvrir les services publics **plus** largement et d'accorder des réductions horaires aux agents qui accepteraient de travailler à des heures atypiques (le soir ou le samedi). (*Le Monde*, 16.09.1997, p. 19)

Interprétation :

[...] il proposait d'ouvrir les services publics **plus** largement **qu'avant**

La présence dans le verbe d'un terme temporel sans réalisation morphologique est prouvée par l'existence d'exemples du genre :

- La fraîcheur se dissipera plus rapidement **que les jours précédents**. (*Le Monde*, 16.09.1997, p. 26)
- Les héritiers étaient deux fois moins nombreux **que lors de la précédente réunion de famille, le 30 août**. (*Le Monde*, 16.09.97, p. 20)

Dans ces exemples, le second terme de comparaison qui est temporel est mis en contraste avec un premier terme non réalisé qui est nécessairement temporel.

5.2.2. LA NATURE DE LA RELATION

Comme pour les exemples isolés, la relation entre les termes de comparaison est une relation de type 2, c'est-à-dire celle où le second terme (ici non réalisé) n'est pas pris comme étalon. La relation met en parallèle les deux termes plutôt qu'elle n'évalue le premier par le second. Dans les énoncés suivants, on peut vérifier cette relation aussi bien par le dispositif "c'est... que" que par la construction pseudo-clivée.

- 1° - L'ersatz ce vieux mot justement d'origine allemande, serait plus compréhensible au commun des bacheliers. (*Le Monde*, 16.09.97, p. 34)
- **C'est l'ersatz ce vieux mot justement d'origine allemande qui** serait plus compréhensible au commun des bacheliers.

- Ce qui serait plus compréhensible au commun des bacheliers **c'est** l'ersatz ce vieux mot justement d'origine allemande
- 2° – ouais je me vois plus euh+ plus dans un bureau et faire faire de la di (Nouguier, 60, 3)
 - **C'est** dans un bureau **que** je me vois plus
 - **Là où** je me vois plus **c'est** dans un bureau
- 3° – moment – là bon ben le : le placement est plus un rapport euh qualité-sécurité – donc là c'est pas un obj (*Prévoyance*, 5, 6)
 - **Ce qu'il** est **c'est** plus un rapport qualité-sécurité
 - Marion Jones est plus impressionnante sur 200 m (*Le Monde*, 16.09.97, p. 23)
 - **C'est** sur 200 m **que** Marion Jones est plus impressionnante.
 - **Là où** Marion Jones est plus impressionnante **c'est** sur 200 m.

Ces tests qui rendent l'interprétation que ces énoncés ont dans leur contexte prouvent bien que les termes en comparaison sont en relation paradigmatique, c'est-à-dire en relation 2.

Modou N'DIAYE
Université Cheikh Anta Diop, Dakar



Bibliographie

- AUFREY, Michel et DE LA FONTINELLE, Jacqueline (1995), "La comparaison dans les langues d'Océanie : rupture, rapprochement, éloignement", *Faits de langues*, n° 5, pp. 51-61.
- BILGER, Mireille (1983), *Étude distributionnelle de la coordination par ET*, Thèse de troisième cycle, Université de Provence.
- BLANCHE-BENVENISTE, Claire, DEULOFEU, José, STEFANINI, Jean & EYDEN, Karel van den (1984), *Pronom et syntaxe. L'approche pronominale et son application à la langue française*, Paris, SELAF.
- BENVENISTE, Émile (1948), *Noms d'agent et noms d'action en indo-européen*, Paris, éd. Maisonneuve.

- BONNARD, Henri (1968), "Que de que", *Le français dans monde*, n° 59.
- CAPPEAU, Paul & SAVELLI, Marie-josée (1995), "Corrélation ne vaut pas comparaison", *Faits de langues*, n° 5, pp.175-182.
- CREISSELS, Denis (1995), "L'expression de la comparaison dans une langue africaine : l'exemple du tswana", *Faits de langues*, n° 5, pp.41-50.
- GAATONE, David (1996), "Subordination, subordinées et subordinants", dans C. Muller (éd.), 1996.
- JONAS, Pol (1973), *Les systèmes comparatifs à deux termes en ancien français*, Bruxelles, Éditions de l'Université de Bruxelles.
- LAMBERTERIE, Charles (1995), "L'apport d'Émile Benveniste", *Faits de langues*, n° 5, p. 17.
- MOREL, Marie-Annick & DANON-BOILEAU, Laurent (1995), "Présentation générale", *Faits de langues*, n° 5, pp. 1-12.
- MULLER, Claude (1996), *La subordination en français*, Paris, Armand Colin.
- NDIAYE, Modou (1989), *L'analyse syntaxique par joncteurs de liste*, thèse nouveau régime, Université de Provence.
- NOJGAARD, Morten (1992), *Les adverbes français. Essai de description fonctionnelle*, t. I, Copenhagen.
- RIVARA, René (1990), *Le système de la comparaison. Sur la construction du sens dans les langues naturelles*, Paris, Minuit.
- WHITTAKER, Sunniva (1995), "À la recherche de l'étalon", *Faits de langues*, n° 5, pp. 165-174.

Questions de méthode dans la description morphosyntaxique de l'élément *que* en français contemporain

1. PROBLÈMES D'ANALYSE POSÉS PAR CERTAINS EMPLOIS DE *QUE*

Cet article voudrait soulever quelques points de méthode en vue de l'étude du morphème *que* en français contemporain. L'opportunité de cette démarche tient à l'apparition d'études, comportant des propositions d'analyse nouvelles, notamment Le Goffic (93), Moline (95), Muller (96), Boone (96). J'avais moi-même proposé (Deulofeu [88]), dans le cadre théorique de l'Approche Pronominale, des éléments pour une étude descriptive du fonctionnement syntaxique de *que*. Ces diverses études développent des conceptions très différentes du statut morphologique de *que* et de son fonctionnement syntaxique. J'essaierai de comprendre les raisons de ces divergences. Je montrerai qu'aucune de ces analyses n'est pleinement satisfaisante et je m'appuierai sur cet examen critique pour dégager des préalables méthodologiques à une approche cohérente de la description morphosyntaxique de *que*.

Je limiterai la discussion à trois types d'emplois. Les exemples qui les illustrent sont classés provisoirement sous des rubriques héritées de la tradition. Chaque type d'exemple comporte une attestation écrite trouvée dans divers ouvrages normatifs et une attestation rencontrée dans des transcriptions de français parlé ou entendue dans des conversations.

A) **Circonstant universel** : *que* suffit à introduire une subordonnée circonstancielle exprimant des rapports sémantiques variés, là où la

norme actuelle exigerait le relais d'un connecteur (*pour, de telle manière que, puisque, alors que*) :

• **manière-degré**

- (1a) il se rengorgeait que c'était un plaisir (Saint-Simon)
 (2b) c'est laid que ça en peut plus (Oral, 1981)

• **temps :**

- (2a) la vie est trop courte et la mort nous prend que nous sommes encore
 pleins de nos misères (Sévigné)
 (2b) j'ai pris l'appartement que c'était cultivé et j'ai continué à le faire¹
 (Giudic. 6, 12)

• **but :**

- (3) viens que je t'embrasse (Académie)

• **conséquence :**

- (4a) j'ai une tendresse pour mes chevaux qu'il me semble que c'est moi-
 même quand ils souffrent (Molière in Grevisse)
 (4b) ça nous a fait des frais que là les sous ils sont encore partis (Ni. 4,
 36)

• **cause :**

- (5a) vous êtes donc brouillés que vous ne vous saluez plus ? (Hugo)
 (4b) mon oncle a même failli tirer sur une feuille qu'il croyait que c'était
 un gibier (BU. 4, 10)

B) Subordonnant à valeur de coordonnant : On a ici une simple équivalence avec un *et* ou un *mais* :

- (6a) et la raison bien souvent les pardonne que l'honneur et l'amour ne les
 pardonne pas (Mol., Littré)
 (6b) son cœur croit aimer aussi parfois qu'il n'en est rien (Mol., Littré)
 (6c) il a été habiter à côté de chez Rosalie que Rosalie elle savait pas
 (Kneip, 4, 11)

Ces emplois à valeur de coordonnant non seulement sont vivants, mais encore seraient en train de gagner du terrain dans l'usage "populaire". Ainsi, A. Sauvageot [66] affirme :

l'emploi de l'élément *que* en fonction de simple conjonction de coordination s'est étendu considérablement dans le langage familier ou teinté

1. Comprendre : "alors que les fleurs qui s'y trouvent étaient cultivées".

de vulgarité : *on entendait pas un bruit, pas une mouche voler. Qu'on aurait cru qu'il n'y avait pas une mouche à la maison* [p. 41]

C) *que* introduit une principale

On citera d'abord le cas classique et normatif de la "subordination inversée" : la proposition introduite par *que* est "ressentie" comme une principale, alors que la principale par la forme est ressentie comme subordonnée. Ces cas sont attestés en français de conversation :

- (7a) il me le demanderait à genoux que je ne cèderais pas (Thérive)
- (7b) je n'avais pas fait vingt pas que la pluie se met à tomber (Batany, in Dupré [72])
- (7c) et la race humaine sera éteinte que du poisson il y en aura encore (Nony, 18,16)

Batany commentant Thérive mentionne un autre emploi, qui, bien qu'attesté dans la "langue populaire" lui apparaît hors système : il s'agit du *que* devant une construction qui n'est ni coordonnée ni subordonnée à une autre. Il s'agirait d'un "gasconisme" :

- (8) pour lors que vous avez renversé cette dame et que vous auriez dû me prévenir

Brunot [1922], au contraire, ne voit pas dans de tels usages une influence étrangère mais le simple développement extrême du système français² :

- (9) "Au dernier stade, il sert d'introductif à des propositions [...] : *que c'est comme un bouquet de fleurs*. C'est du style gendarme sans doute, mais le tour est très répandu dans la langue parlée" [p. 704]

Les difficultés soulevées par l'analyse de ces exemples peuvent se résumer en deux problèmes :

- le premier est celui d'une conjonction de subordination, qui peut, néanmoins introduire des principales et des coordonnées.
- le second est moins simple à formuler. Dans les cas où *que* semble indiscutablement introduire une subordonnée circonstancielle, par exemple :

- (2a) la mort nous prend que nous sommes encore pleins de nos misères

on peut s'étonner qu'il puisse introduire à lui tout seul un complément circonstanciel de verbe. On s'attendrait, comme pour les groupes nomi-

2. Sachant que Brunot a été le premier à réaliser des enregistrements de français parlé, c'est là une affirmation à prendre au sérieux : s'il avait jugé que ce tour était un gasconisme, il l'aurait sans doute exclu de son ouvrage de référence.

naux, à voir le lien de dépendance s'exprimer par une préposition ou tout autre élément de relation. La conjonction *que* est en effet un élément qui permet à une construction verbale d'entrer dans la réaction d'un verbe, mais qu'on la conçoive comme un "complémentiseur" ou un "translateur" à la manière de Tesnière, ou encore comme un marqueur modal, à la suite de A. Boone [96], elle n'exprime ni ne marque ce lien de réaction lui-même. Il faut donc expliquer pourquoi les subordinées en *que* peuvent être construites directement comme circonstants d'un verbe. C'est le problème du circonstant sans relateur.

En fait, les problèmes naissent du caractère confus de la catégorie de conjonction de subordination qui est au centre de l'analyse traditionnelle. Les traitements récents de *que* remettent en cause divers aspects de cette analyse. Parmi les approches nouvelles, j'ai choisi d'en confronter trois qui épuisent les possibilités logiques de résoudre les problèmes en termes morphosyntaxiques : jouer sur la morphologie, la syntaxe, ou les deux à la fois.

- Deulofeu [88] remet en cause le principe du parallélisme catégorie morphologique / relation syntaxique en posant que la conjonction *que* est à la fois un marqueur de subordination et de parataxe.
- Le Gotlic [93] remet en cause le statut "morphologique" de *que* : il ne l'analyse pas comme une conjonction, mais comme un adjectif relatif indéfini.
- Moline [95] remet en cause à la fois l'analyse morphologique de *que* et l'analyse par subordination : *que* est analysé comme un relatif qui peut marquer un lien de subordination aussi bien que de parataxe.

2. QUELQUES PROPOSITIONS RÉCENTES D'ANALYSE

Je propose de situer cette confrontation dans un cadre distinguant clairement les divers aspects de l'analyse morphosyntaxique.

2.1. LES COMPOSANTES DE L'ANALYSE : CATÉGORIES MORPHOLOGIQUES, SYNTAXE INTERNE DES CONSTRUCTIONS, SYNTAXE EXTERNE DES CONSTRUCTIONS

Une description explicite du fonctionnement morphosyntaxique des constructions introduites par *que* en A, B, C ci-dessus — désormais désignées par le terme de *que-Phrase*- comporte trois aspects :

- la détermination de la **catégorie morphologique** du morphème *que* :
Ainsi, dans nos exemples, il est légitime de se demander si le morphème introducteur *que* est de type conjonction ou relatif.
- la détermination de la **syntaxe interne** de la *que*-Phrase :
Quelle est l'organisation interne des *que*-Phrases ? S'agit-il de constructions verbales ordinaires simplement marquées par un *que* qui leur est extérieur ? S'agit-il au contraire de constructions dont une place est affectée d'une opération particulière : relativisation, extraction...
- la détermination de la **syntaxe externe** de la *que*-Phrase :
Quelle est la relation syntaxique qui unit l'ensemble du constituant introduit par *que* avec le contexte ? On peut se demander par exemple si le constituant est régi par le verbe de la principale ou s'il est construit par simple parataxe.

Dans une perspective descriptive, ces trois types d'études doivent être menées indépendamment et faire l'objet chacune de justifications distinctes.

2.2. UNE SOLUTION SYNTAXIQUE : DES CONJONCTIVES À SYNTAXE EXTERNE DIFFÉRENTE. CONJONCTIVES RÉGIES ET "ASSOCIÉES"

Dans Deulofeu [88], j'ai montré, dans un cadre descriptif inspiré de l'approche pronominale, que, sur la base de l'examen de leurs propriétés syntaxiques, ces *que*-Phrases peuvent avoir un double fonctionnement.

- Certaines fonctionnent comme des constitutants "régis" par un verbe : exemples (1) et (2) ci-dessus (pour le détail, voir : [88] pp. 93 à 97).
- d'autres sont simplement "associées" à une construction verbale : exemples (3) à (5) pour les "circonstanciellles", 6 (subordination à valeur coordonnante), 7 (subordinations inversées). Les "associés" de l'approche pronominale peuvent être considérés comme des éléments construits en "parataxe" ([88], pp.97 à 103).

Cette solution présentait l'avantage de ne pas faire intervenir de relations *ad hoc*. Par rapport aux solutions habituelles, elle renonçait à la notion mal définie de subordination au profit de celle de "rection", définie par des critères syntaxiques explicites. Je montrais alors que le groupe des "circonstanciellles" en *que* — exemples 1 à 5 — est hétérogène sur le plan de la syntaxe externe : il comporte des régies et des associées.

J'analysais donc *que* comme un introducteur d'une que-Phrase pouvant soit entrer dans la rection d'un verbe, soit être construite par simple parataxe (dans le cas de l'associée). Des références à d'autres conjonctions aptes à ce double fonctionnement (*parce que, bien que*) fournissaient une justification indépendante à cette analyse. Le problème de la prétendue "conjonction de subordination" impliquée dans des coordinations était résolu.

Le problème de l'absence de relateur était également résolu pour tous les cas où la que-Phrase fonctionne comme associée. Une catégorie "associée" à une construction n'a, en effet, pas besoin d'outil de liaison pour fonctionner. On peut, de ce point de vue, comparer le comportement des que-Phrases à celui des groupes nominaux associés :

il serait pas venu que je lui aurais fait un procès
 il serait pas venu un procès tout de suite
 il a pris un magasin que ça a été encore une grosse catastrophe
 il a pris un magasin encore une catastrophe

LES INSUFFISANCES DE LA SOLUTION

1. Mon travail ne discutait pas la syntaxe interne des constructions : je tenais pour acquis la nature de conjonction de *que*.

2. Des problèmes de syntaxe externe restaient à résoudre ou à approfondir

- Pour les constructions analysées comme régies :

Je n'avais pas discuté le fait que ces constructions ne présentaient pas toutes les propriétés habituellement attribuées aux constructions régies. N'ayant rien dit de l'absence d'élément de relation, je n'avais pas résolu le problème de la que Phrase régie sans relateur. Je n'avais pas non plus rendu compte du fait que, contrairement aux autres constructions régies, les que-Phrases n'étaient pas antéposables :

(12a) il est venu quand j'étais à Paris
 (12b) quand j'étais à Paris il est venu
 (2b) on a pris le logement que c'était cultivé
 (2b') * que c'était cultivé on a pris le logement

- Les constructions associées

La relation d'association est définie de façon négative : est associé ce qui n'est pas analysable comme régi par un verbe. On l'a vu, l'opposition régi/associé rappelle tout à fait l'opposition hypotaxe / parataxe de la tradition. Or la parataxe ne nous dit rien de positif sur les rapports entre les deux membres qu'elle unit.

2.3. UNE SOLUTION MORPHOLOGIQUE : QUE ADVERBE INDÉFINI EN FONCTIONNEMENT "INTÉGRATIF"

2.3.1. PRINCIPES DE LA SOLUTION DE LE GOFFIC [93]

Le Goffic propose de renoncer à l'analyse en conjonction et d'analyser *que* comme un relatif-indéfini. Un des fonctionnements syntaxiques possibles de ce type de morphèmes consiste, en effet, à entrer dans la composition de constructions "intégratives" adverbiales (des relatives "sans tête" ou "sans antécédent", selon les terminologies les plus courantes), sur le modèle de :

(13) il est parti (quand il l'a voulu + où il a voulu)

Le *que* des exemples 1 à 7 sera donc analysé comme un adverbe relatif-indéfini (comme *quand*, *où*...). Les *que*-Phrases de ces mêmes exemples sont analysées comme des intégratives. Du point de vue de la syntaxe externe, ces intégratives sont tout simplement, et dans tous les cas, des compléments accessoires du verbe "principal", comme l'intégrative en *quand* de (13). Examinons les arguments qui sont donnés pour justifier ces analyses.

2.3.1.1. Morphologie de *que* et syntaxe interne de la *que*-Phrase

La démonstration repose essentiellement sur le raisonnement par analogie : on repère dans la langue des cas clairs de *que* adverbe relatif-indéfini et l'on suggère, sur la base de paraphrases, que les emplois de type 1 à 7 sont des extensions des emplois indiscutables. Observons la démarche :

On part d'emplois où la forme *que* commute indiscutablement avec des adverbes, en l'occurrence des adverbes de quantité. Il peut donc légitimement être analysé comme occupant la position syntaxique d'un adverbe de degré :

que Marie est jolie ! cf. Marie est très jolie
qu'il me gêne cf : il me gêne beaucoup
si vous saviez *que* d'ennuis il a eus (cf. combien d'ennuis il a eus)

Le Goffic propose de retrouver cet adverbe dans des constructions qu'il analyse comme "intégratives" :

"au total *que* adverbe indéfini de degré a des emplois exclamatifs résiduels, mais des emplois intégratifs nombreux et importants [...] (Le Goffic [93], § 367)" comme d'autres intégratifs (*quand*), *que* permet de constituer des compléments accessoires de phrase. *Que* ayant une

valeur de base quantitative, l'intégrative directement formée "identifie des valeurs quantitatives" :

il danse que il en peut plus

(paraphrasé par : il danse à un degré auquel il n'en peut plus) (*id.*, § 285)

Le Goffic étend alors cette analyse à d'autres emplois :

... les intégratives en *que...* sont utilisées avec des valeurs dérivées : les deux structures sont mises en relation, non pas au titre d'une égalité de degré, mais d'une manière vague, sous l'angle d'une identité de conditions d'existence indéterminées, le contexte se chargeant de spécifier dans une certaine mesure :

P1 dans lesquelles circonstances (conditions) P2 :

la vie s'achève que l'on a à peine ébauché son ouvrage

il était à peine arrivé qu'il partait

vous me le diriez que je ne vous croirais pas

viens ici que je t'embrasse

tu ne partiras pas d'ici que tu ne nous aies chanté quelque chose

(§ 286)

venez-vous ? qu'il fait froid (§ 367)

2.3.1.2. Syntaxe externe

La présence du *que* semble suffire à garantir le statut de complément ou de subordonnée de la que-Phrase. Aucun autre argument n'est donné pour confirmer ce statut. La syntaxe externe de ces constructions ne fait pas l'objet d'une étude détaillée.

2.3.2. SOLUTION DES PROBLÈMES

Il est indéniable que le recours à l'analyse par adverbe intégratif rend compte, pour les exemples analysés, de distributions qui apparaissent autrement comme aberrantes.

Le problème des conjonctives sans connecteur est résolu très simplement : les intégratives adverbiales, au contraire des conjonctives, se construisent directement, comme les adverbes, avec les verbes qui les régissent. Si la que-Phrase est une intégrative, il est normal que l'on ait le paradigme (14) :

(14) il est parti quand il le voulait

il l'a mis où je souhaitais

il l'a fait comme je le voulais

il danse que il en peut plus

la vie s'achève que l'on a à peine ébauché son ouvrage

2.3.3. DIFFICULTÉS SOULEVÉES PAR LA SOLUTION

Le Goffic ne se propose pas dans l'ouvrage de 93 de démontrer en détail les hypothèses qu'il avance, il en présente essentiellement la cohérence théorique, dans un ouvrage qui couvre l'ensemble de la syntaxe de la phrase. Dans la perspective descriptive limitée que je me suis fixée, il est au contraire indispensable de rassembler les éléments dont il faudrait, à mon avis, disposer pour choisir, sur des bases empiriques et non spéculatives, entre la position de Le Goffic et l'analyse en conjonction et conjonctive. Si elle est indispensable, l'entreprise n'est pas simple. Une des raisons essentielles est que nous ne sommes pas sûrs des données. Nous manquons d'exemples réellement attestés et parfois l'intuition de grammaticalité fait défaut.

2.3.3.1. *Le statut catégoriel de que*

Pour prouver que *Que* est un adverbe intégratif, il faut montrer que ce *que* possède des propriétés formelles à la fois communes avec les adverbes au fonctionnement intégratif indiscutable : *quand, où, comment, pourquoi, combien...* et distinctes de celles de la conjonction.

Dans le cas des adverbes relatifs indiscutables, on peut appuyer l'analyse en proforme adverbiale sur les propriétés intrinsèques que sont les traits-sémantico-syntaxiques [+ Temps], [+ Loc], [+ manière], dégagés par affiliation paradigmatique avec des adverbes de la même série : *alors, là, ainsi...*

Mais pour un *que* adverbe circonstant neutralisé, on voit mal, sur la base d'un simple rapprochement lexical : "dans les quelles circonstances", sans l'appui d'une proforme équivalente, à quel paradigme l'affilier et par quel trait le caractériser. On aura du mal à caractériser positivement par un trait sémantico-syntaxique cette supposée pro-forme.

On peut alors proposer d'examiner des propriétés combinatoires pour comparer *que* aux autres adverbes intégratifs.

- formes renforcées :

Quand et *où* peuvent être renforcés dans certains usages par *est-ce que* ou *c'est que* :

(15) je vais jamais où (c'est + est-ce) qu'il passe ses vacances

(16) quand (c'est + est-ce) qu'il est venu j'étais pas là

Or cela ne paraît pas possible pour les *que* de nos exemples :

(17) * la mort nous prend que (c'est + est-ce) qu'on a à peine ébauché
notre ouvrage

- (18) * viens! que c'est qu'on va manquer le train
 (19) * il n'était pas encore là que c'est que l'autre est parti

• fonctionnement syntaxique limité

Que serait le seul adverbe intégratif à avoir un fonctionnement interrogatif contraint. Seule la construction avec négation est naturelle dans un registre soutenu, mais l'interprétation est différente de l'intégrative :

- (20) ? que le disais-tu ?
 (21) que ne le disais-tu ?

Ce serait aussi le seul à ne pas se construire avec l'infinitif :

- (22) j'ai trouvé (comment faire + où aller + quand partir + *que partir
 [au sens de dans quelles conditions partir])

On conclura de ces observations que l'on a aucun argument direct clair en faveur de l'analyse de *que* comme adverbe indéfini.

2.3.3.2. *Difficulté d'établir des propriétés
 justifiant une syntaxe interne d'intégrative*

La syntaxe interne de l'intégrative se distingue de celle d'une simple conjonctive par le traitement particulier qui affecte l'une de ses positions syntaxiques, traitement que l'on retrouve dans la relative, dans l'interrogative et dans l'extraction. Ce traitement ne consiste pas seulement dans le fait que cette position soit réalisée par une catégorie morphologique particulière, celle des pronoms dits "interrogatifs-relatifs" ou encore pronoms *qu-*. Il existe en effet des constructions verbales dont un complément est réalisé par un pronom *qu-* et qui pourtant ne se distinguent pas fondamentalement, sur le plan de l'organisation syntaxique interne, d'une construction verbale "ordinaire" :

- (23) il est allé au cinéma avec qui ?

Et d'un autre côté, on peut soutenir avec de bons arguments que la position à traitement particulier apparaît sans la présence d'un pronom *qu-*, par exemple dans les relatives du type (24), où *que* peut être analysé comme une conjonction-complémenteur :

- (24) l'homme que j'ai vu - l'accident qu'on vous a parlé

ou encore dans les constructions à complément antéposé affecté d'une intonation de fin d'énoncé étudiées dans Sabio [92] :

- (25) à la caisse ! ils se payent

Il existe donc bien des caractéristiques proprement syntaxiques, indépendantes des réalisations morphologiques, des constructions dont l'intégrative est un sous-type. Ces caractéristiques ont été résumées en grammaire générative par l'idée que la position concernée aurait été l'objet d'un "mouvement"³ par rapport à une position canonique. Dans un modèle "item et arrangement", on parlera plutôt, comme en syntaxe "constructionnelle" de construction "filler-gap" (cf. Fodor [83]). L'approche pronominale analyse ces constructions comme des dispositifs syntaxiques particuliers de la construction verbale, qui possèdent des propriétés syntaxiques communes largement indépendantes de l'appartenance catégorielle de la place "disposée".

Au-delà de la terminologie, ce qui importe, c'est de reconnaître l'autonomie de la syntaxe et de la morphologie. La morphologie peut certes servir d'indice qu'un énoncé a une structure que j'appellerai désormais à "place (syntaxique) distinguée". Mais cet indice fait parfois défaut, comme dans les exemples précédents. Il faut alors des arguments syntaxiques indépendants pour prouver le statut d'intégrative de la que-Phrase.

a) Un premier argument de ce type en faveur de l'analyse en intégrative s'appuie classiquement sur le fait que l'on ne puisse pas trouver dans la que Phrase un complément de même type que celui que représente l'intégratif. Pour le cas des intégratives supposées en *que*, compte tenu de l'idée que la place occupée par l'intégratif est un circonstant passe partout, il n'est pas simple de trouver une réalisation lexicale qui pourrait être considérée comme redoublant le circonstant passe-partout. Il est cependant possible de montrer que certaines incompatibilités attendues ne sont pas observées.

Considérons le cas des subordinations inversées :

(26) il est à peine arrivé que l'autre se met à l'injurier

Dans la logique de l'analyse, *que* identifie des "moments" de la principale et de l'intégrative. Il tient donc lieu de circonstant de temps dans cette

3. On désigne d'ordinaire ce mouvement par le terme "mouvement de qu-". Mais comme on vient de le voir, il est dangereux de préjuger que ce mouvement est lié à la présence d'une catégorie morphologique particulière. Pour exprimer la caractéristique syntaxique indépendamment de ses réalisations morphologiques il faut donc parler simplement de mouvement, en spécifiant éventuellement le type syntaxique de mouvement par rapport à d'autres possibles : le terme de "chopping rule" utilisé par J. Ross permettait de bien distinguer les caractéristiques proprement syntaxiques des constructions en question.

intégrative. Il devrait dès lors être incompatible avec d'autres circonstants de temps. Or on peut construire :

(27) il était à peine arrivé que juste à ce moment là l'autre est parti

De même :

(28) il me le donnerait que même dans ces conditions je n'en voudrais pas

où *dans ces conditions* devrait redoubler indument la fonction attribuée à *que*.

Toujours selon l'analyse de Le Goffic, dans un énoncé tel que :

(29) je ferai des heures en plus que ça compensera

Que devrait occuper la fonction d'un circonstant de type : "dans ces circonstances".

Considérons alors l'énoncé suivant, relevé dans une observation participante :

(30) la semaine prochaine je ferai des heures en plus que ça compensera
comme ça (Oral, oct. 80)

comme ça y représente un circonstant très général, bien proche de "dans ces conditions". On pourrait retrouver ce circonstant général assez librement me semble-t-il dans les énoncés comme :

(31) viens que je puisse signer ce papier

(32) viens que comme ça je puisse signer ce papier

(33) tu viens ? que comme ça on sera pas en retard

Dans tous ces cas, *que* ne semble pas "redoubler" un complément du verbe, ce qui ne confirme pas l'analyse en intégrative.

Pour certains énoncés, cependant, la situation est moins claire. C'est le cas des énoncés où *que* produit un effet de "degré" :

(34) il danse que c'est merveilleux

Si *que* exprime une identification de degré paraphrasable par :

(35) il danse à un degré (point) et à ce point c'est merveilleux,

on ne devrait pas pouvoir construire sans redondance de fonction les énoncés suivants :

(36) ? il danse que à ce degré c'est merveilleux

(37) ? il donne au pauvres que plus que ça on pourrait pas

(38) ? il danse que, (à ce point d'élégance + aussi bien) c'est impossible

Ces énoncés paraissent effectivement moins naturels que (29)-(33). On aurait donc là un argument pour analyser au moins les que-Phrases à effet quantitatif comme des intégratives.

Les faits ne sont pas facile à établir, compte tenu du caractère vague des paraphrases que l'on peut faire, mais il ressort donc que, pour un nombre important d'exemples, aucune contrainte syntaxique ne semble peser sur la non réalisation d'une place de construction de la *que* phrase. L'analyse en intégrative ne peut donc être généralisée.

b) possibilité d'enchassement du verbe de l'intégrative

Les intégratives, comme les interrogatives et relatives, se prêtent à des effets de "rection à distance" : l'élément *qu-* peut être complément d'un verbe enchassé (en gras) de la *qu-* Phrase et non de son verbe principal :

- (39) il l'a fait quand on lui avait demandé qu'il le **fasse**
- (40) il l'a fait de la manière qu'on lui avait demandé qu'il le **fasse**
- (41) il l'a fait comme on lui avait demandé qu'il le **fasse**

Cet effet d'enchassement ne me semble pas possible pour toutes les que-Phrases analysées comme des intégratives. Considérons l'énoncé suivant comportant une relative à verbe recteur enchassé :

- (42) il a fait son travail dans les conditions où on lui avait demandé qu'il le **fasse**

On devrait pouvoir construire en parallèle une intégrative en *que* de même sens :

- (43) ? il a fait son travail qu'on lui avait demandé qu'il le **fasse**

Ce dernier énoncé, s'il est acceptable, n'a pas l'interprétation de l'énoncé (42). Il peut tout au plus avoir une interprétation causale, sans effet de rection à distance : "parce qu'on lui avait demandé qu'il le fasse".

La gêne que l'on ressent devant l'énoncé :

- (44) ? Marie était encore là qu'on m'a dit que Pierre était déjà parti

reflète sans doute le fait que l'on perd l'équivalence avec :

- (45) Marie était encore là que Pierre était déjà parti, me dit-on

ce qui s'expliquerait par la difficulté de donner à (44) une lecture avec recteur enchassé.

Sur ce point, les énoncés à effet quantifieur semblent avoir encore un comportement plus proche des intégratives :

- (46) il danse que je pense qu'on peut pas mieux
- (47) il danse qu'on dit que c'est merveilleux

c) l'usage de *le plus*

Les phrases à place distinguée sont des contextes favorables pour l'utilisation du constituant à valeur superlative *le plus* (48-50), dont l'usage est restreint dans les constructions verbales ordinaires (51-52) :

- (48) c'est le jambon que j'aime le plus
- (49) je vais aller en vacances où je m'amuse le plus
- (50) le judo j'ai pas aimé quand je l'ai le plus pratiqué mais après
- (51) ? il m'a dit qu'il aimait le plus le jambon.
- (52) ? j'aime le plus le jambon

Les intégratives supposées en *que* ne semblent pas constituer des contextes favorables à l'emploi, de *le plus* :

- (53a) ? il a gardé mes enfants qu'il a gardé le plus les tiens

est peu acceptable. Il ne peut, en tout cas, être une paraphrase du possible :

- (53b) il a gardé mes enfants dans les conditions dans lesquelles il a plus gardé les tiens

Les *que* à effet de justification paraissent particulièrement rebelles à cet emploi :

- (54a) ? travaille ici! qu'on est le plus éclairé
- (54b) travaille ici ! que c'est là qu'on est le plus éclairé⁴

d) *que* peut ne pas introduire une construction verbale

Une intégrative a par définition une structure de construction verbale puisque le pronom intégratif exerce une fonction de complément non essentiel de verbe. Or il est possible d'observer des cas où *que* n'introduit pas une construction verbale :

- (55) mais il m'a pas foutu dehors moi Bellier qu' à elle oui (A.L. 14, 28)
- (56) et lui ben il a toujours cherché à me rendre service que elle jamais (oral, 1981)

Ces emplois ne peuvent donc être analysés comme des intégratives.

e) Dans certains emplois, *que* introduit bien une construction verbale, mais certaines caractéristiques la rendent impossible à analyser comme une intégrative.

4. L'impossibilité d'emploi de *le plus* avec les *que* de degré, n'est pas révélatrice puisqu'il pourrait s'agir d'une incompatibilité entre deux éléments appartenant au même paradigme.

- On rencontre un subjonctif à valeur injonctive :
 - (57) elle me disait de prier et de me recommander au bon Dieu que là où ils sont qu'ils retournent Rossi, bande
- une forme interrogative partielle :
 - (58) non je veux dire on a une affection pour l'endroit + que ça fait combien de temps que vous êtes là Monsieur M (Tomeï, 10, 4)
- *que* peut aussi introduire des constructions verbales coordonnées par et :
 - (59) mais le sens c'est comme la violette que il y en a des blanches et puis il y en a des roses (OC)
 - (60) et le quinze août ça a été le Débarquement mais le Débarquement tous les jeunes se sont fait tuer pour nous défendre combien de morts il y a eu + que je suis resté enfermée dans une cave pendant dix à quinze jours et les gens à côté m'ont pas donné un verre d'eau (Marseille, femme 50 ans)
- il peut aussi introduire un ensemble constitué de constructions en parataxe ou coordonnées, comportant éventuellement des incises et du discours direct. Dans les exemples suivants, les crochets marquent les limites du groupe introduit par *que* :
 - (61) (maintenant, un ouvrier) il va payer une assurance sur son petit salaire + que [là à l'usine on nous l'enlevait c'était fini mais le patron il en payait déjà un peu il en payait la moitié là-dessus] (Richaud, 21, 5)
 - (62) on se voyait **que** [maintenant à Saumaty euh il y des matins moi il y a des collègues que tant + je descends pour ainsi dire toutes les nuits + eh bien tant de trois ou quatre jours je les vois pas alors d'un coup oh Jo comment ça va+ et euh] **que** [l'autre entrée automatiquement c'était petit on se voyait tous les matins]

On ne voit pas, dans les exemples 59-62 de quel verbe les divers *que* pourraient être compléments. Ni quel type de complément ils pourraient représenter.

2.3.3.3. *Les problèmes posés par le fonctionnement syntaxique externe de l'intégrative*

A. Limitations d'emploi par rapport aux autres intégratives

- détachement à gauche impossible

Les intégratives adverbiales ont pour l'ensemble un fonctionnement diversifié. On les trouve par exemple en position antéposée détachée :

- (63) où j'habite il fait beau
 (64) quand je viens il fait beau

avec éventuellement une reprise pronominale :

- (65) quand tu viens j'aime ça
 (66) comment il a fait je m'en souviens plus
 (67) Pourquoi il a fait ça je peux pas l'approuver

Ce n'est pas le cas pour les intégratives en *que*

- (68a) la vie s'achève qu'on a pas terminé son ouvrage
 (68b) * qu'on a pas terminé son ouvrage la vie s'achève
 (69a) je m'en vais qu'il va faire nuit
 (69b) * qu'il va faire nuit je m'en vais
 (70) ? qu'on a été éliminés c'est rare
 (71) ? qu'on a été éliminés tu l'as jamais connu toi

Ces deux derniers exemples ne peuvent, en tout cas, pas être interprétés comme : “les circonstances dans lesquelles on a été éliminés...”.

Là encore, les *que*-Phrase à effet de degré ne sont pas concernées par le test : les éléments exprimant le degré ou la manière ont des antépositions très contraintes, quelle que soit leur catégorie.

- Impossibilité d'emploi comme complément direct de verbes

Les intégratives peuvent être utilisées comme compléments directs de verbes :

- (72) j'ai pas aimé (comment + pourquoi) tu l'as fait

Cet emploi n'est pas possible pour *que* :

- (73) ? j'ai pas aimé que tu es parti

Si cet énoncé est acceptable, la *que*-Phrase ne recevra jamais l'interprétation : “les circonstances dans lesquelles tu es parti”, mais celle de “le fait que tu sois parti”, révélant par là un *que* conjonction.

- Emploi des prépositions

Les intégratifs acceptent des prépositions faibles (*de, à, en*) et fortes (*pour, par...*)

- (74a) il se rappelle de (quand+pourquoi+où-comment) il a fait ça
 (74b) il a été surpris par comment j'ai procédé
 (74c) il s'est préparé pour quand je reviendrai

que intégratif n'est compatible ni avec les faibles ni avec les fortes

- (75a) * il se rappelle de que il a fait ça (interprété comme : il se rappelle des circonstances dans lesquelles il a fait ça).
 (75b) * il s'est préparé pour que on va l'accueillir (les conditions dans lesquelles...)

Dans le cas des prépositions faibles, cette contrainte pourrait être un effet de règles "locales" affectant la forme *que* quelle que soit son statut morphologique. Ainsi le *que* conjonction n'est pas non plus compatible avec les prépositions faibles :

- (76a) * il se plaint de que tu sois venu
 (76b) il se plaint de ce que tu sois venu

Mais *que* conjonction est compatible avec des prépositions fortes telles que : *pour*, *sans* etc. Il reste donc surprenant que la forme *que* ne puisse être employée avec préposition forte dans les énoncés :

- (77) * il a été étonné par que on m'a renvoyé (les conditions dans les quelles)
 (78) * il s'est préparé pour que on va s'embarquer (les conditions dans lesquelles)
 (79) * pour que ils nous ont renvoyés je t'en parlerai demain (alors que l'on peut contruire : pour comment il faut faire je te le dirai demain)

Dans tous ces exemples *que* présente un fonctionnement différent de celui des autres intégratifs. En revanche, si on l'analyse comme une conjonction, les distributions sont régulières.

B. L'intégrative en que a des emplois inconnus des autres intégratives

Si l'on prend en compte d'autres exemples que ceux étudiés par Le Goffic, on s'aperçoit que la solution par intégrative pose de nouveaux problèmes :

- *que* n'est pas subordonné à un verbe

Dans des exemples comme :

- (80) oh ! que je vous avais pas vu
 (81) heureusement ! que ça commençait à bien faire

la construction n'est pas complément d'un verbe et sa relation au contexte est hors syntaxe grammaticale. Je ne vois pas d'intégrative adverbiale prototypique qui pourrait fonctionner ainsi :

- (82) oh ! je te reconnais bien là
 (83) ? oh ! où je te reconnais bien

- *que* est combinable avec le coordonnant *et*

Les corpus nous offrent des exemples de *que* combiné avec *et*, exemples signalés par Batany (voir (8) plus haut) :

- (84) j'ai été réfugié tout et je suis allé à Pelissanne et remarque de Pélissanne nous sommes allés à Riez et de Riez nous sommes venus à Miramas **et que** depuis que nous sommes à Miramas je crois que là vraiment nous sommes dans un endroit que ça va bien (Arquier, 1, 11)
- (85) et ils les ont mis dans le camion et ils sont partis et moi je suis resté et quand je suis sorti je suis resté bête que j'ai dit merde ils sont partis quest-ce que c'est et voilà comment je m'ensuis sorti *et que* si j'avais été dans le camion je serais plus de ce monde voilà (Béziade, 13, 12)

Dans ce cas, évidemment, l'analyse en intégrative complément non essentiel ne peut plus se justifier : la que-Phrase n'a pas seulement la valeur, mais bien le statut d'une coordonnée et non d'une subordonnée. Aucune intégrative ne peut être coordonnée par *et* avec le verbe dont elle dépend.

L'ensemble de ces données, pour difficiles qu'elles soient à établir et à manipuler, m'invite à conclure à la prudence. Au vu de leurs propriétés de syntaxe interne et externe, beaucoup de ces que-Phrases sont impossibles à analyser comme des intégratives. Des études empiriques complémentaires seront nécessaires pour cerner les emplois qui pourraient l'être. C'est sans doute le cas de la que-Phrase à valeur de degré, où les propriétés syntaxiques confirment une analyse morphologique intrinsèque justifiée en adverbe de degré.

2.3.3.4. *Problèmes posés par l'analyse de la syntaxe externe de l'intégrative*

J'ai signalé plus haut que la seule relation utilisée par Le Goffic pour rendre compte de la syntaxe externe des que-Phrases était celle de complément accessoire. Cette option aboutit à donner la même analyse syntaxique à des segments manifestant pourtant des propriétés formelles très différentes.

Pour certains d'entre eux, le recoupement de nombreuses propriétés prouve leur dépendance syntaxique par rapport à un verbe, notamment :

il danse que il en peut plus
la vie s'achève que l'on a pas terminé son ouvrage

où, par exemple, les commutations : *il danse ainsi, la vie s'achève ainsi,*

soutiennent l'idée d'une analyse par complément non essentiel, pour d'autres au contraire :

il me le dirait que je ne le croirais pas
on croit parfois aimer qu'il n'en est rien
est-tu malade ? que le médecin est venu

le seul argument que l'on peut invoquer pour analyser la que-Phrase comme dépendante est l'argument circulaire de la présence de *que*.

Le Goffic relève d'ailleurs, pour certaines conjonctives et des relatives, ce paradoxe de constructions verbales introduites par de prétendus "subordonnants" qui sont syntaxiquement indépendantes :

"une subordonnée en *si bien que* s'utilise facilement comme une énonciation indépendante § 290" "Les groupes pronominaux en *quoi* peuvent s'employer derrière ponctuation forte comme des énoncés quasi indépendants :

il dina; après quoi, il s'en alla tranquillement"

Cette observation pourrait s'appliquer à beaucoup des que-Phrases étudiées précédemment. À travers le terme de "quasi-indépendantes", Le Goffic soulève, sous une autre forme, le paradoxe des subordonnées à valeur coordonnante. Son cadre d'analyse ne lui permet pas de le résoudre.

Le Goffic aurait pu proposer une solution en utilisant la relation de parataxe, qu'il présente ainsi. § 341 :

la parataxe aux contours mal définis (= subordination non marquée, c'est-à-dire, en pratique non marquée autrement que par la prosodie)

Mais Le Goffic ne distingue pas la relation de parataxe et son système de marquage, il n'envisage donc pas que des liens de parataxe puissent unir des éléments marqués par des morphèmes tels que : *si bien que*, *à quoi* ou *que*. Accepter la notion de parataxe avec lien lui permettrait de résoudre, en termes syntaxiques, tout en restant dans son cadre d'analyse, le paradoxe des subordonnées fonctionnant, selon sa propre intuition, comme des énoncés quasi- indépendants". Mais là encore le lien nécessaire établi, *a priori*, entre composante morphologique et composante syntaxique externe ne permet pas d'envisager une telle hypothèse : les conjonctions de subordination ne peuvent que subordonner.

2.4. UNE SOLUTION MORPHOLOGIQUE ET SYNTAXIQUE : SUBORDINATION SYNTAXIQUE VS PRAGMATIQUE

2.4.1. PRINCIPES

Moline [94] résout les problèmes en jouant à la fois sur une réanalyse de *que*, variante de l'analyse de Le Goffic, et sur une réorganisation de la composante syntaxique, option non retenue par Le Goffic. Le modèle est donc puissant et potentiellement capable de fournir une solution à l'ensemble des problèmes. Le principe de la solution peut être résumé ainsi : les constructions en *que* régies par le verbe de la principale sont des relatives sans tête, les constructions en *que* "paratactiques" sont toutes des relatives "continuatives" ou "de liaison". On entrevoit d'ores et déjà que cette hypothèse met Moline à l'abri des difficultés soulevées, dans la solution de Le Goffic, par les cas de subordonnées à valeur coordonnante. Mais là aussi, il faut voir de près comment l'hypothèse est justifiée.

2.4.1.1. Statut de que : forme invariante d'une variable syntaxique

Comme Le Goffic, Moline réduit donc considérablement le rôle du que "conjonction" ou "complémenteur" : "nous montrerons tout au long de cette étude que, dans de nombreux emplois, *que*, habituellement considéré comme une conjonction, est en fait analysable comme la réalisation d'une forme invariante de relatif neutralisé" (Moline [94]), p. 48). La conclusion essentielle pour l'analyse est que *que* "occupe diverses fonctions dont celle de circonstanciel neutralisé (que glosable : dans certaines conditions)" (*ibid.*, p. 46).

2.4.1.2. Syntaxe interne de la que Phrase

Syntaxe interne de la que-Phrase et relatif sont en fait circulairement définis. Moline définit la structure syntaxique relative ainsi :

"la relative est une expression ouverte, où le relatif représente une inconnue (une variable libre)"(p. 42) "la relative est une fonction propositionnelle" (p. 45).

Par rapport à Le Goffic, qui utilisait la seule structure de la relative "sans tête" ou "intégrative" et la seule catégorie des relatifs indéfinis, la solution de Moline consiste à poser deux structures de syntaxe interne pour analyser nos emplois.

- relative libre, équivalent de l'intégrative :

il danse que c'est une merveille
il ne vient pas qu'il n'apporte un bouquet de fleurs

- relative "continuative" ou "de liaison" :

il n'était pas encore arrivé que l'autre partait
il faut partir qu'il fait nuit

Cette différence de structure se refléterait dans les propriétés logiques (variable liée/variable libre) ainsi que dans les paradigmes des morphèmes introducteurs. Comme dans l'approche classique, on pose deux paradigmes distincts : celui des introducteurs de propositions, où la variable est liée : les "interrogatifs-indéfinis" et celui des introducteurs de "proposition ouverte" : les relatifs.

2.4.1.3. Syntaxe externe de la que-Phrase

L'hypothèse de Moline est la suivante : non seulement la relative est reliée en syntaxe externe à un constituant recteur par des relations de subordination syntaxique variées (épithète, attribut, prédicat) mais encore elle peut l'être par des relations qui ne relèvent plus de la syntaxe mais bien de l'organisation pragmatique du discours : elle peut être dite en situation de subordination pragmatique.

Pour établir un lien de subordination syntaxique entre deux constructions verbales P1 et P2, les critères utilisés sont essentiellement sémantiques :

"Les circonstants 'modifiant', en leur imposant des conditions de validité supplémentaires, les unités syntaxiques auxquelles ils s'appliquent. [...] Pour que P2 puisse être considéré comme un circonstant [donc subordonnée syntaxique] par rapport au syntagme verbal définissant le prédicat de P1, il faut s'assurer que P2 appartient à ce prédicat :

En 12 et 13 :

- (12) je viendrai quand j'en aurai envie
(13) je travaille maintenant qu'il fait beau

les groupes en italique apportent une restriction au prédicat de P1. De ce fait ils constituent de véritables circonstants.

Les énoncés 14 et 15 :

- (14) je viendrai, quand j'en aurai envie
(15) maintenant qu'il fait beau, je travaille

sont au contraire composés par la succession de deux séquences énonciatives [E1 et E2] et nous ne considérons pas comme des circonstants au sens strict les éléments détachés qui constituent de mini- énonciations

et qui confèrent à l'énoncé E qui les contient un caractère complexe”
[Moline 94, p. 19.]⁵

Ces éléments détachés sont reliés à la construction verbale principale par une relation de subordination pragmatique, définie comme suit :

“Nous définirons E1 et E2 comme étant des mini énoncés entrant dans la composition de l'énoncé global E. [...] E1, énonciation principale est une unité minimale, le plus souvent énonçable isolément, quelle que soit la nature syntaxique des éléments qui la composent. C'est une sorte de noyau autour duquel s'articule le reste du discours. E2, énonciation secondaire n'existe qu'en ce que E1 lui confère un tel statut, l'inverse n'étant pas vrai. La structure énonciative (pragmatique) [...] peut être schématisée de la manière suivante : E = E2, E1.” (*ibid.*, p. 22)

“La subordination syntaxique et la subordination pragmatique relèvent de deux niveaux d'analyse différents. Les subordonnants peuvent apparaître dans les deux cas, avec la même forme.” (*ibid.*, p. 26)

C'est sur la base de ce cadre que Moline va établir la syntaxe externe des constituants comportant un *que*. L'idée est que chacun d'entre eux peut être relié au contexte soit par une subordination syntaxique, soit par une subordination pragmatique.

- subordination syntaxique

il danse qu'on peut pas mieux
je ne partirai pas que tu ne m'aies donné raison

On rappelle que, dans ce cas, du point de vue interne, l'analyse proposée est celle de la relative “sans tête”.

- subordination pragmatique

1. forme E = E2 E1

– les “subordinations inversées”

il viendrait que je ne le recevrais pas
il n'était pas plus tôt arrivé que l'autre partait

– les effets consécutifs

j'ai travaillé travaillé qu'à la fin tout a bien marché

5. L'explication de la quasi-indépendance un peu paradoxale de certaines subordonnées relevée par Le Goffic, trouve ici une explication : les constructions en questions sont indépendantes sur le plan syntaxique, mais elles sont dépendantes sur le plan pragmatique. Cette solution “énonciative” serait d'ailleurs possible dans le cadre de Le Goffic.

2. forme E = E1 E2

– les effets “justificatifs” :

il faut partir qu'il se fait tard

On rappelle que l'analyse interne de la que-Phrase est, dans ce cas, celle d'une relative fonctionnant comme “continuative ou de liaison”.

2.4.2. SOLUTION DES PROBLÈMES

Les cas de subordination à valeur coordonnante ne posent plus de problème : ils constituent dans ce cadre des cas réguliers de subordination pragmatique.

Pour le problème de la subordonnée sans relateur, il faut distinguer :

- Les emplois relevant de la subordination syntaxique.
Le problème est résolu, comme pour Le Goffic, par l'analyse en relative sans tête. Une relative sans tête n'a, on le rappelle, pas besoin d'introducteur pour être circonstant de prédicat.
- Les emplois relevant de la subordination pragmatique.
Il s'agit d'une relation entre énonciations et non entre constructions. La grammaire de la langue n'impose donc pas de mot de relation. La relative en subordination pragmatique, qu'elle soit introduite par *que* ou par tout autre relatif n'a pas besoin de mot de liaison pour fonctionner :

(86a) je suis allé à Paris que j'ai passé des heures en bibliothèque

(86b) je suis allé à Paris où j'ai passé des heures en bibliothèque

Outre la solution de ces problèmes, l'analyse fournit une explication des contraintes de position des que-Phrases.

Les relatives en subordination pragmatique “continuatives” sont par définition post-posées à la construction avec laquelle elles enchainent. La position strictement à droite des que-Phrases de ce type suit alors beaucoup plus naturellement ici que dans l'analyse “en relative sans tête” de Le Goffic. On rappelle en effet qu'il était difficile dans son cadre d'expliquer pourquoi la relative sans tête en *que*, contrairement à celles en *quand*, *où*, *comment*., était limitée à la position droite. L'analyse de Moline présente donc à première vue les avantages de celle de Le Goffic, sans les inconvénients relevés en syntaxe externe.

2.4.3. DIFFICULTÉS SOULEVÉES PAR L'ANALYSE

2.4.3.1. Morphologie et syntaxe interne des constructions

Comme dans le cas de Le Goffic, nous sommes en droit d'attendre de Moline la démonstration par les propriétés syntaxiques que tel ou tel emploi de *que* Phrase a bien une syntaxe interne de construction verbale "à place distinguée" et non de construction verbale "ordinaire". Examinons l'argumentation de Moline sur l'exemple de la subordination inversée.

Soit à montrer que dans la construction suivante :

il n'était pas arrivé qu'il repartait

que est une forme de relatif neutralisé réalisant une place "temporelle" dans une relative continuative. L'argumentation pp. 245-246 présente deux types d'arguments syntaxiques en faveur de l'analyse de la *que*-Phrase comme relative.

Un premier argument repose sur la comparaison avec un fonctionnement indiscutable de *que*-Phrase conjonctive :

(5) il est à peine arrivé qu'il repart déjà

(12) je crois qu'il part

"[...] à la différence d'un morphème en tête d'une relative continuative, *que* complétif n'est pas susceptible d'entrer en relation paradigmaticque avec *et* ou *mais* (cf. *je crois et il est parti)".

Un deuxième argument intéresse la syntaxe interne de la *que*-Phrase :

"[...] *que* a bien une fonction dans la proposition en tête de la quelle il se trouve : il anaphorise la repère temporel établi par la première séquence. Il est d'ailleurs difficile d'insérer un complément temporel dans la seconde proposition :

?? elle était partie depuis longtemps qu'on la cherchait encore (à ce moment là + à huit heures)

?? il est à peine arrivé qu'à cinq heures il repart déjà

L'acceptabilité est cependant meilleure lorsque d'une part, ce complément est situé en tête de la seconde proposition, et que, d'autre part, son sens reste imprécis :

il est à peine arrivé, qu'à ce moment-là, il repart déjà"

Sur ce point la note 19, p. 247 précise : "cf. cependant l'exemple d'Altaire 82 : le scrutin était à peine ouvert qu'à neuf heures trente le verdict était déjà tombé"

Les deux arguments donnés ne sont guère convaincants. L'argument de syntaxe externe est affaibli par le fait que l'on présuppose que le seul emploi possible pour une *que*-Phrase conjonctive est celui de la complé-

tive. C'est évidemment quelque chose qu'il faut établir. On pourrait dire que le fait de ne pas entrer en paradigme avec une *et*-Phrase ne concerne que les *que*-Phrases conjonctives régies, pas celles que l'on pourrait analyser comme construites par parataxe :

(87) on faisait comme ça à cette époque que maintenant on le fait plus

(88) on faisait comme ça à cette époque *et/mais* maintenant on le fait plus

L'argument portant sur la syntaxe interne est affaibli par les exemples mêmes que présente l'auteur: le *que* de subordination inversée paraît bien compatible avec des circonstants de temps.

Mais pour Moline, la faiblesse des arguments proprement syntaxiques n'est pas déterminante. Elle considère qu'elle est compensée par la solidité des arguments sémantiques donnés par ailleurs. Dans une perspective descriptiviste, on est cependant en droit d'exiger une argumentation syntaxique indépendante. Or, nous avons déjà constaté à propos de la solution de Le Goffic que cette argumentation syntaxique indépendante est impossible à établir pour ce type d'exemples. Ni l'une ni l'autre approche ne montrent, sur la base de propriétés formelles, qu'il s'agit d'une construction différente d'une construction verbale ordinaire introduite par un *que* conjonction.

2.4.3.2. Syntaxe externe

- Analyse syntaxique et interprétation pragmatique ne sont pas distinguées

La notion hybride de subordination pragmatique empêche de poser clairement la question des types de relations syntaxiques qui peuvent intéresser les constituants. En effet, la notion de subordination pragmatique recouvre deux choses :

1. une relation syntaxique définie négativement entre une construction verbale et un constituant dont on ne peut montrer qu'il lui est subordonné syntaxiquement. Il s'agit donc d'une reconnaissance implicite de la notion de parataxe avec lien.

2. L'interprétation pragmatique de cette relation de type paratactique. De fait, les sous cas de subordination pragmatique ne représentent rien d'autre que les interprétations pragmatiques possibles de la situation de parataxe avec lien entre deux énoncés.

La question n'est même pas posée de savoir si l'on peut distinguer sur les bases formelles des sous types syntaxiques de parataxe avec lien. L'analyse juxtapose donc une relation de subordination syntaxique définie en termes formels et sémantiques, et divers types de subordinations prag-

matiques définis négativement en termes formels (le constituant n'est pas subordonné syntaxiquement) et positivement en termes d'interprétation. Une analyse descriptive doit faire apparaître le chaînon manquant : l'analyse des sous-types syntaxiques que recouvre la notion confuse de subordination pragmatique.

L'analyse des structures en termes de E2 E1 est ainsi révélatrice des ambiguïtés entre analyse des formes et analyse des contenus pragmatiques. Il est indiscutable que du point de vue de l'analyse pragmatique dans (89) et (90) la que-Phrase représente "l'assertion principale du message donné par le locuteur" et doit être analysée comme E1 :

- (89) Marie n'était pas arrivée que Paul était déjà parti
 (90) et alors j'ai pris la boucherie que nous avons très bien réussi

alors que dans (91) :

- (91) allons-y ! qu'il va faire nuit

la structure s'analyse pragmatiquement avec une que-Phrase en position E2 d'assertion secondaire, justifiant un acte d'injonction pragmatiquement principal.

Mais du point de vue des formes, on doit constater qu'aucune des que-Phrases ne peut fonctionner comme un énoncé autonome. L'une comme l'autre doit apparaître après un premier énoncé présent dans le discours :

- (92) ? que Paul était déjà parti
 (93) ? que nous avons très bien réussi
 (94) ? qu'il va faire nuit

En dépit d'interprétations pragmatiques différentes, ces que-Phrases ont une syntaxe externe identique. L'analyse de Moline ne permet pas d'exprimer ce fonctionnement syntaxique commun.

- L'analyse ne rend pas compte de toutes les données

Pas plus que l'approche de Le Goffic, celle de Moline, ne résout les problèmes posés par deux types d'exemples mentionnés plus haut.

a) On ne voit pas quelle fonction le *que* relatif pourrait exercer lorsque la que-Phrase n'a pas la structure d'une construction verbale, notamment dans le cas des constructions sans verbe et des "gapping" (voir exemples 55 à 62)

b) Les que-Phrases construites avec *et* (exemples 84 et 85) posent également problème. L'analyse par "relative continuative" ne peut en rendre compte : aucune relative continuative explicitement marquée ne peut être introduite par *et* :

- (95a) * il allait arriver et quand soudain surgit un homme
 (95b) * il arriva à Paris il fit de nombreuses démarches il rencontra un homme et à qui, d'ailleurs, son Père l'avait adressé

- Des problèmes nouveaux se posent :

L'analyse de *que* comme relatif neutralisé occupant une position syntaxique de variable soulève deux problèmes supplémentaires touchant à la cohérence interne de l'analyse :

1. Comment expliquer que cette forme soit la seule forme de relatif continuatif qui ne puisse fonctionner comme interrogatif ?

2. Comment expliquer que ce relatif-variable neutralisé puisse exercer toutes les fonctions à l'exception de celle de sujet ? Cette difficulté ne se pose pas dans l'approche de Le Goffic où le *que* est un adverbe indéfini donc incompatible avec la position sujet.

2.5. LA SOLUTION SYNTAXIQUE RÉVISÉE : INTRODUCTION D'UNE COMPOSANTE MACROSYNTAXIQUE

Les solutions à base de *que* relatif-indéfini ne sont pas convaincantes sur l'ensemble des exemples étudiés. Certains emplois parmi les plus vivants sont même franchement impossibles à analyser ainsi. Cette situation me conduit à la conclusion qu'il faut reprendre l'analyse par *que* conjonction et examiner, à la lumière du débat précédent, si une fois révisée, elle ne permettrait pas d'offrir une alternative crédible à l'analyse par relatif. Un aboutissement possible de la démarche étant qu'il faille articuler les deux solutions pour venir à bout de la description.

Pour la clarté de l'exposé je distinguerai le cas des *que*-Phrases "régies" de celui des *que* Phrases "associées". Je commencerai par le second.

- Domaine des associés et composante macrosyntaxique.

L'hypothèse d'une syntaxe articulée en deux composantes: microsyntaxe et macrosyntaxe (voir Blanche-Benveniste et alii [90]), permet de structurer positivement le domaine des associés et d'y reconnaître des unités différentes sur la base de propriétés formelles. Ainsi la composante macrosyntaxique permet d'analyser les énoncés en unités telles que: noyau, préfixe, suffixe :

- (96) Il a beau dire (préfixe) pas question (noyau) car ce serait une erreur (suffixe)

Le noyau peut fonctionner de façon autonome dans le discours (97a), à la différence d'autres unités comme les préfixes (97b) ou les suffixes (97c) :

- (97a) pas question
 (97b) * il a beau dire
 (97c) * car ce serait une erreur

Les suffixes supposent la présence d'un premier noyau pour apparaître et ne peuvent donc enchaîner directement avec un préfixe :

- (98) * il a beau dire car ce serait une erreur

Du point de vue de la composition interne, les unités macrosyntaxiques ne sont pas tenues à être réalisées par des catégories grammaticales particulières : ainsi un noyau peut être réalisé comme une tournure nominale (97a), une tournure adverbiale : tant mieux ! voire une simple interjection : bof !

Dans ce cadre là, on peut analyser *que* comme un marqueur de suffixe dans un très grand nombre d'exemples qui font difficulté. Cette analyse convient aussi bien aux cas où le suffixe est une construction verbale qu'à tous les exemples vus plus haut (55 à 62) où il introduit d'autres types de constituants.

Le noyau sur lequel s'appuie le suffixe en *que*, en tant qu'unité macrosyntaxique, peut être lui-même composé d'une simple interjection ou un adverbe comme en (80, 81).

Ce que l'on ne rencontre jamais, c'est une construction introduite par *que* en début absolu d'interaction ou de séquence discursive⁶ :

- (99) ? tiens ! que Paul arrive (au sens de voilà que Paul arrive)
 (100) ? à propos que j'ai pas pu passer à la banque

6. Les exemples de pseudo-principales signalés en C ne sont pas des contre exemples à l'analyse par suffixe. Certes les *que*-Phrases des exemples 8 et 9 ou de l'exemple suivant entendu au début d'une prise de parole :

que tu pourrais pas prendre une fourchette

ne s'appuient pas sur des segments linguistiques, mais ils ne sont pas pour autant auto-suffisants. La présence de *que* signale au moins que le contenu de la *que*-Phrase doit être interprété en rapport avec un élément du contexte, par exemple un geste maladroit de l'interlocuteur dans l'exemple précédent. Une construction verbale sans *que* n'imposerait pas une telle référence extérieure. Il est donc possible d'absorber ces exemples, particulièrement rebelles à l'analyse par dépendance syntaxique, dans une extension de la composante macrosyntaxique prenant en compte des références au contexte non verbal, telle que le propose par exemple Berrendonner [1990].

On remarquera en outre que ces constructions ne fonctionnent pas directement avec des éléments exclusivement préfixes :

(101) ? il a beau pleurer *que* je l'écouterai jamais

alors *que* la paraphrase avec un premier noyau est bonne :

(102) il pleurerait *que* je l'écouterais pas

De même l'énoncé :

(103) ce livre qu'il est très beau

ne serait interprétable *que* si le premier terme était un noyau nominal à intonation injonctive :

(104) (donne-moi) ce livre ! qu'il est très beau

Le groupe nominal ne peut être un préfixe. L'énoncé (103) n'est pas un équivalent de :

(105) ce livre il est très beau

La construction introduite par *que* a toutes les propriétés d'un suffixe. *Que* peut donc être considéré comme une conjonction marqueur de suffixe parallèlement à son fonctionnement microsyntaxique d'introducteur de *que*-Phrase régie. On peut attribuer à ce morphème un signifié. Il est le plus souvent interprétable comme marquant des suffixes à valeur argumentative au sens très large : il souligne une conclusion, une justification, une objection voire une évaluation formulée à propos d'un noyau énoncé précédemment. Il indique parfois simplement que le suffixe est pris dans le même ensemble discursif que les éléments qui précèdent. L'interprétation ne sera jamais, en tout cas, celle d'une simple constatation ou d'une affirmation directe. Le détail de ces interprétations reste à étudier, mais on trouvera d'excellentes analyses sur ce point dans Moline [94].

L'analyse par suffixe et lien macrosyntaxique apporte le chaînon manquant évoqué plus haut : la notion confuse de subordination pragmatique est analysée du point de vue du signifiant comme une structure macrosyntaxique noyau-suffixe et du point de vue du signifié devra l'être en types d'interprétations pragmatiques qui restent à préciser à partir d'emplois attestés. Une telle analyse met à l'abri de la constitution de sous-types syntaxiques, là où il y a seulement influence du contenu. Ainsi, les exemples interprétés comme un *alors que* exprimant le contraste ne doivent pas cette valeur à *que*, mais bien au parallélisme de structure des énoncés articulés. De même, les effets de subordination inversée peuvent être mis au compte des décalages chronologiques créés par les adverbes mis en relation par la structure :

(106) il n'est pas encore là que Pierre s'en va déjà

L'interprétation de la que Phrase suffixe peut être explicitée à peu près ainsi : “Compte tenu du fait qu'il n'est pas encore là et des attentes contextuelles, ce n'est pas une simple constatation de dire que Pierre s'en va déjà”.

Si l'on intègre à l'analyse le cas des que Phrases liées par *et*, on pourra préciser la notion de “marqueur de suffixe”. Dans le cas où *et* se combine avec *que*, on doit considérer que c'est *et* qui précise la relation noyau-suffixe. Quel est alors le statut de *que* ? Il conviendrait de distinguer entre marqueur de lien macrosyntaxique (*et*, *mais*, *parce que*, ...) et marqueur du statut d'un énoncé. *Que* serait donc un marqueur de statut indiquant que l'énoncé qu'il introduit doit nécessairement enchaîner sur une première énonciation, ce qu'assure le fonctionnement suffixe. *Que* marquerait ainsi un type de modalité énonciative : une énonciation nécessairement seconde dans le discours, parce qu'effectuée à propos et à la suite d'une autre.

Restent alors les cas de que-Phrase régie.

- Le cas de l'interprétation en terme de degré :

il a dansé qu'on peut pas mieux

On peut y voir un *que* adverbe intégratif de degré : c'est le cas pour lequel l'analyse en intégrative paraît la plus argumentable, comme on l'a vu plus haut.

- Le cas plus général de la relation perçue comme un circonstant généralisé interprétable comme : “dans des conditions telles que”,

Je proposerai encore une analyse en *que* marqueur de suffixe. Cette hypothèse demande des vérifications empiriques précises, mais je la propose comme une alternative à l'analyse en que-Phrase relative dans l'espoir que de la confrontation naîtra une solution bien argumentée. Je rappelle que le problème posé par ces constructions, c'est que l'on comprend mal pourquoi la relative sans tête que l'on pense y trouver est limitée à cet emploi après le verbe, contrairement aux mêmes relatives en *quand*, *comment*, etc. Mon hypothèse consiste à mettre cette contrainte au compte de la composante macrosyntaxique. Supposons que la que-Phrase soit un suffixe. Sa position est immédiatement explicable. Mais alors pourquoi l'effet de rection ? L'effet de rection tiendrait au fait que ce suffixe est en quelque sorte greffé sur un noyau précédent non encore “achevé”. L'idée est que les constructions verbales centres de noyaux peuvent être développées de deux façons différentes :

– soit par une catégorie prévue par la grammaire de rection et selon les contraintes de cette grammaire :

- (107) la mort nous surprend (ainsi + alors que nous n'avons pas tout fait)
 (108) je l'ai laissé (malade + comme ça)

– soit par une unité macrosyntaxique interprétable comme une assertion “de première main”, ce que peut confirmer la possibilité d'un même effet de sens sans la présence de *que* :

- (109a) la mort nous surprend que nous n'avons encore rien fait
 (109b) la mort nous surprend nous n'avons encore rien fait
 (110a) je l'ai laissé qu'il respirait à peine
 (110b) je l'ai laissé il respirait à peine

Dans le cas des éléments non régis présenté plus haut, le suffixe enchaîne après un noyau “achevé” qui porte ses propres modalités :

- (111a) il est malade. qu'il reste couché
 (111b) tu es malade ? que tu restes couché

Dans le cas de l'effet rectionnel, il n'y a qu'une seule modalité couvrant le noyau et le suffixe :

- (112a) je l'ai laissé qu'il était malade
 (112b) tu l'as laissé qu'il était malade ?

Deux types d'exemples me paraissent apporter une justification indépendante à cette hypothèse :

1. Les pseudo-relatives en *il y a... que*

que permet de construire des variantes intéressantes de constructions à présentatif. Pour ces constructions, la “partie prédicative”, d'ordinaire réalisée sous forme de relative (113a), peut se présenter soit comme une construction verbale ordinaire sans référence à l'élément présenté(b), soit comme une énonciation de forme quelconque (c à d). Elle peut aussi être réalisée comme une construction verbale non introduite par *que* (e et f)

- (113a) il y a des gens à qui on peut pas faire confiance
 (113b) il y a des gens que on peut pas leur faire confiance
 (113c) il y a des gens que tu m'as compris
 (113d) il y a des gens que pardon !
 (113e) il y a des gens on peut pas leur faire confiance
 (113f) il y a des gens tu m'as compris

Si l'on veut rendre compte des formes variées que peut prendre la partie prédicative, sans vider la notion de relative de tout sens, on peut proposer une analyse de ces énoncés articulant microsyntaxe et macrosyn-

taxe. On poserait que la tournure en *il y a* peut être construite soit au moyen de la structure grammaticalement prévue de relative (*à qui, où*) soit par la greffe d'une unité macrosyntaxique suffixe. La construction grammaticale et la greffe produisent généralement le même effet de sens, mais la greffe permet d'exprimer des contenus plus variés que la construction grammaticale. Ainsi on voit mal comment une relative pourrait donner l'équivalent des greffes suivantes :

- (114) il y a des gens (que) plus tu les vois moins tu as envie de leur parler
 (115) il y a des gens (que) quand ils se mettent en colère il faut s'en aller

Ainsi, qu'il s'agisse de déterminer un nom ou un prédicat verbal, on pourrait dire que la "macrosyntaxe y va quand la microsyntaxe ne peut ou ne veut y aller".

2. Un autre cas où l'on voit apparaître des constructions en *que* là où on ne les attendrait pas, c'est dans des cas de recherches lexicales sur une position syntaxique. La construction est d'abord réalisée selon la catégorie grammaticale attendue (adjectif ou adverbe dans les exemples ci-dessous), puis elle est relexicalisée sous la forme d'une que Phrase en principe non autorisée par la grammaire :

- (116) L1 il faudrait qu'il vienne des filles valables
 L2 c'est quoi valables
 L1 valables + des filles à peu près intéressantes pas intéressantes
 physiquement + que tu peux discuter avec eux et tout (AG)
- (117) sinon c'est la routine je t'ai dit tu as des interventions qui marquent plus que d'autres que tu te remémore de temps en temps en toi ou bien que tu en discutes avec ceux qui ont fait l'intervention avec toi tu te rappelles l'accident à tel endroit (Col, 12, 1)
- (118) elle a dit j'aimerais faire le maçon mais moi j'en suis sûr c'est impossible parce que si vous prenez un aggro comme ça que ça serre là-dessus (AG).

Si le premier exemple peut encore être analysé comme une relative non standard compte tenu des caractéristiques du locuteur, le deuxième exemple est prononcé par un interne en médecine qui ne construirait certainement pas d'emblée une relative du type :

c'est des interventions que tu en discutes...

L'hypothèse d'une que-Phrase suffixe greffée sur la liste de réalisations lexicales et constituant une extension des catégories que l'on devrait y trouver est beaucoup plus vraisemblable. De même, le troisième exemple

est plus naturel ainsi qu'avec une que-Phrase directement construite avec le verbe :

(119) ? si vous prenez un agglo que ça serre là-dessus

Ce que je propose donc, c'est que l'on puisse greffer un suffixe, avec ou sans *que*, sur une place d'une construction syntaxique pour constituer une unité de discours. Cette analyse dispense du recours ad-hoc à la notion de rupture de construction pour rendre compte des énoncés précédents. Elle s'appuie seulement sur un type particulier d'interférence entre composantes syntaxique et macro-syntaxique : la greffe d'un suffixe sur une construction.

Au bout du compte, le choix me semble être entre une solution qui propose d'analyser ces cas de rection étendue comme une relative au comportement très idiosyncrasique et une solution qui propose de mettre les contraintes observées au compte de l'interférence entre structure grammaticale et structure macrosyntaxique.

Je reconnais que ma proposition se présente plus sous la forme d'un programme que d'un résultat démontré. Je suis cependant persuadé que la recherche de faits nouveaux, notamment issus de l'observation de corpus authentiques, dans la perspective de mesurer la portée empirique des deux hypothèses, amènera un progrès important dans l'étude de la difficile question du statut de *que* en français.

3. LES OPTIONS MÉTHODOLOGIQUES EXPLICITES ET IMPLICITES QUI GUIDENT L'ANALYSE

3.1. LE RAPPORT ENTRE LES DONNÉES ET L'ANALYSE

Il faudrait éviter que le choix des données détermine les analyses proposées. Or une conclusion qui pourrait se dégager de cette étude, c'est finalement que la solution par conjonction s'impose surtout si l'on prend en compte des données authentiques non standard, voire régionales, comme le suggèrent les références de mes exemples. Si l'on s'en tient aux exemples standard, la solution par relatif neutralisé reste la meilleure. Je tiens à souligner que ce n'est pas le cas et cela pour au moins deux raisons :

- les deux solutions sont en compétition pour le cas de la "subordination inversée" qui reste tout à fait standard.

- les emplois qui paraissent régionaux ont une extension en général plus grande dès lors que l'on fait des recherches précises. Ainsi le *que* marquant l'opposition, qui figure ici dans des exemples prononcés par des locuteurs originaires de la région de Marseille, est un phénomène attesté dans le parler spontané d'autres régions. Tout ce que l'on peut dire, c'est que ces emplois sont difficiles à cerner pour des universitaires linguistes dont les intuitions sont guidées par la norme.

Il me semble donc possible de travailler sur des données "tout venant" à charge pour celui qui choisit cette option de fournir des attestations indiscutables. De ce point de vue, il reste encore un travail de collecte de données important à faire pour étudier le fonctionnement de *que*.

Les exemples authentiques me paraissent en outre essentiels pour une étude qui articule structures grammaticales et structures discursives. On se rend compte que si l'on travaille avec la seule intuition, ou avec des exemples pris au vol, une réduction des possibilités d'emploi de *que* apparaît. On ignore systématiquement tous les cas où *que* ne construit pas une construction verbale canonique. L'analyse préalable sélectionne en quelque sorte les données. On a vu que l'examen d'exemples provenant de transcriptions ou d'observations minutieuses étendait le champ des possibles, augmentant ainsi les opportunités de validation empirique des hypothèses.

3.2. INTERACTION ENTRE LES COMPOSANTE DE L'ANALYSE : FONCTIONNEMENT EN PARALLÈLE OU PAR MODULES AUTONOMES ?

Un présupposé m'a semblé important à mettre en question, c'est celui de la conception qu'a le descripteur des rapports entre composantes de la description morphosyntaxique. On a vu que Le Goffic s'appuyait sur une conception stricte du parallélisme : un même type de morphèmes = un même type de syntaxe interne = un même type de syntaxe externe. Ainsi la présence de certains morphèmes introducteurs garantit que la construction est subordonnée. Moline désolidarise la syntaxe externe du reste en proposant que des constructions par ailleurs régies puissent fonctionner comme paratactiques : cela veut dire que la présence d'un morphème ne garantit plus le fonctionnement externe de la construction qu'il introduit. Mais Moline continue d'accepter le parallélisme entre morphologie et syntaxe interne : les relatives sont introduites par des relatifs, les conjonctives par des conjonctions. Dans ce cas encore, la présence d'un morphème va garantir, sans besoin d'autre justification, l'analyse d'une construction.

J'ai moi-même proposé de poser une autonomie de principe entre ces composantes. Au bout de ma description, il ne reste plus qu'un cas de détermination réciproque entre morphologie et syntaxe. Il s'agit, comme le montre le tableau suivant du cas où un morphème de type pronom *qu-*⁷ garantit la présence d'une construction à place distinguée (relative, interrogative, extraction...)

	syntaxe interne		syntaxe externe	
	place distinguée	c.v. ordinaire	rection	parataxe
pronom <i>qu-</i>	+	-	+	+
conjonction	+	+	+	+

Il serait à mon avis important de vérifier en détail si une telle corrélation existe vraiment. Certains emplois qui semblent possibles pourraient la mettre en doute. Il paraît ainsi difficile de trouver, dans (120-121), un verbe par rapport auquel *où* occuperait une place distinguée par "relativisation" :

- (120) c'est un endroit où plus tu travailles moins on te paye
- (121) c'est un truc où quand tu tapes dessus ça se met en route

Si d'autres exemples de ce type se rencontraient, on aboutirait à la conclusion d'une autonomie totale des composantes de la description, ce qui renforcerait l'idée que chaque aspect de la description d'une construction : morphologie, syntaxe interne et syntaxe externe doit faire l'objet d'une analyse argumentée indépendante. Une façon d'en finir avec la position privilégiée que les arguments de type morphologique occupent encore dans l'argumentation syntaxique.

José DEULOFEU
Université de Provence



7. Les formes *qu-* dans l'emploi *in situ* sont déjà des contre-exemples. La discussion portera sur les formes *qu-* antéposées.

Références bibliographiques

- BLANCHE-BENVENISTE, Cl., et Alii [1990], *Le français parlé. Études grammaticales*, Paris, éd. CNRS.
- BOONE, A. [1996], "Les complétives et la modalisation", *Dépendance et intégration syntaxique*, Muller éd., Tübingen, Niemeyer.
- BRUNOT, F. [1922], *La pensée et la langue*, Paris, A. Colin.
- DEULOFEU, J. [1988], "La syntaxe de que en français parlé et le problème de la subordination", *Recherches sur le Français parlé*, 8.
- DUPRÉ, P., *Encyclopédie du Bon Français*, Paris, Trévisé.
- FODOR, J. [1983], "Phrase structure parsing and the island constraint", *Linguistic and Philosophy*, 6.
- LE GOFFIC, P. [1993], *Grammaire de la phrase française*, Paris, Hachette.
- MOLINE, E. [1994], *Constructions subordonnées en que : diversité ou polysémie*, Thèse de Doctorat, Université Toulouse Le Mirail.
- MULLER, Cl. [1996] "La conjonction *que* : rection vs. dépendance immédiate et concurrence avec *que* pronominal", *Dépendance et intégration syntaxique*, Muller éd., Tübingen, Niemeyer.
- SABIO, F. [1992], "Les constituants antéposés en français parlé : analyse de deux types syntaxiques et prosodiques", *RSFP*, 11.
- SAUVAGEOT, A. [1962], *Français écrit, français parlé*, Paris, Larousse.

Sujets éloignés.

Esquisse d'une caractérisation des sujets lexicaux séparés de leur verbe

Les analyses qui portent sur la place du sujet se situent, pour l'essentiel, dans le cadre de l'ordre des mots, c'est-à-dire en terme d'antéposition ou de postposition par rapport au verbe. En français, c'est évidemment la seconde position qui a suscité le plus de commentaires. Mais cette orientation conduit, finalement, à négliger certaines des particularités du sujet lorsqu'il précède un verbe. Pour les sujets lexicaux, deux dispositions sont envisageables. Soit le sujet précède immédiatement le verbe :

la chambre est toute simple sans télé — très ordinaire — (POI 97)

soit une séquence s'intercale entre ces deux éléments et interrompt quelque peu l'enchaînement¹ :

et **la personne** *peu à peu* a craqué (POI 97)

on voit bien que **le général** *au fond* était contre tout ce qui pouvait faire écran si j'ose dire (POI 95)

Les sujets "éloignés" du verbe constituent un point sur lequel le locuteur, même linguiste, a peu d'intuition (cet arrangement est-il seulement attesté à l'oral ?). C'est aussi un phénomène sur lequel les indications des grammairiens sont rares. On les rencontre dans deux chapitres : la construction des phrases et la ponctuation.

Brunot (1905-1953) illustre la première tendance. Il aborde à quelques reprises ce problème dans une rubrique sur le "rapprochement

1. Dans la plupart des exemples cités, le sujet sera écrit en gras et l'insertion en italiques.

des mots” qui permet, à travers des commentaires sur le style et la clarté de l'expression, d'envisager une certaine évolution de ce phénomène dans la langue. Dans le tome III de *l'Histoire de la langue française* (qui porte sur la période 1600-1660), il signale ainsi que le pronom sujet tend à ne plus être séparé du verbe. Le sujet lexical est traité de façon semblable et, sous l'influence de Vaugelas, les phrases avec des éléments intercalés (adjectifs, déterminatifs) entre le sujet et le verbe se raréfient. Mais cette tendance n'a pas pleinement abouti si l'on en croit la remarque finale de Brunot :

ce n'est qu'après Vaugelas, et à la suite de ses remarques sur la netteté du style que ces phrases suspendues disparaîtront peu à peu. Encore en faisons-nous de très analogues. (p. 658)

Dans le tome IV (période 1660-1715), le commentaire est légèrement modifié. La formulation laisse entrevoir que cet usage est différent du nôtre :

Le nom sujet continue toujours à pouvoir être séparé de son verbe, et personne n'imagine d'y mettre obstacle (p. 1083)

La remarque laisse aussi entendre que cet usage n'aurait pas survécu. L'exemple cité à l'appui relève certes d'une syntaxe qui est probablement peu usitée aujourd'hui, mais plus à cause du cumul d'éléments insérés que de l'insertion elle-même :

Le duc de Nevers, dont la vie étoit glorieuse par la guerre et par les grands emplois qu'il avoit eus, quoique dans un âge un peu avancé, faisoit les délices de la cour. (Princesse de Clèves)

Le tome VI (langue du XVIII^e siècle) confirme la lecture du passage précédent. Le rapprochement du sujet et du verbe s'est stabilisé et devrait donc avoir disparu de nos jours :

La séparation du sujet et du verbe par une phrase circonstancielle, à la manière latine, est définitivement écartée, (p. 1951)

Mais la comparaison des deux versions d'une phrase — l'exemple original en (a), l'énoncé corrigé par l'Académie en (b) — laisse entrevoir que la disparition évoquée frappe plus une catégorie d'insertion (i.e. *une phrase circonstancielle*) que le procédé d'insertion lui-même, qui reste toujours utilisé dans la version (b) :

- (a) Mais Darius, comme il étoit religieux et plein de douceur, répondit...
(*Quince-Curce*)
- (b) Mais Darius, religieux et plein de douceur comme il l'estoit...

En dépit de ces fluctuations, c'est finalement la remarque du tome III qui semble la plus adaptée à la situation contemporaine. Même si les restrictions sur les catégories intercalées se sont accentuées, il est manifeste que les "phrases suspendues" ont perduré.

À l'écrit, la question du sujet éloigné interfère aussi avec le problème de la ponctuation². Girault-Duvivier (1811) développe une argumentation qui justifie l'emploi de la virgule entre le sujet long et le verbe. Ce signe est paré d'un double statut — respiratoire et syntaxique — comme en témoigne ce passage :

si l'étendue d'une proposition excède la portée ordinaire de la respiration, il faut y marquer des repos par des *Virgules* placées de manière qu'elles servent à y distinguer quelques-unes des parties constructives (p. 341)

À l'appui de cette position, il cite l'exemple :

Le plaisir de soulager un infortuné, est un remède sûr contre la peine que nous fait sa présence.

C'est là un changement notable avec l'usage contemporain qui, généralement, bannit la virgule dans cette position. On se trouve toutefois aux confins (voire même en dehors) de ce travail puisqu'il n'y a pas insertion d'un constituant mais introduction d'un outil de "repos", jugé nécessaire lorsque le sujet est trop long.

1. PRÉSENTATION DES DONNÉES

Notre attention, dans ces pages, portera sur les sujets "éloignés" du verbe auquel ils sont liés. Une première difficulté tient à la collecte des exemples pertinents. Il existe certes un certain nombre de cas clairs :

- À l'écrit. Si l'on prend au sérieux la ponctuation, il est bien sûr facile de repérer les sujets que nous appelons éloignés, lorsque les éléments insérés sont encadrés par des signes typographiques couplés (deux virgules, des parenthèses, des tirets) :

Les critiques, lorsqu'elles ne cherchent qu'à dénigrer, ne m'intéressent pas. Longtemps, j'ai eu du mal à les accepter. Aujourd'hui j'en ai fait mon deuil. En revanche, **les enjeux internes, oui,** ont beaucoup d'importance. (Presse – *Marianne*. 29-09-97)

2. L'insertion prend alors l'apparence d'un signe de ponctuation.

Il suffit que **Timsit** (*admirable dans ce film, grâce à une composition en tiroirs*) regarde Chabat avec des yeux de chien battu [...] (*Le Monde*, 11-12-97)

• À l'oral. Pour certaines formes sujets "compactes" (i.e. qui constituent le groupe sujet à elles seules comme le relatif *qui*, le pronom *lui*, etc.), tout constituant³ qui se réalise avant le verbe sera interprétée comme une insertion (hormis bien évidemment les clitiques) :

ils sont responsables de tout un tas de touristes **qui franchement** se ruent sur leurs euh bateaux (POI 97)

il faut revenir à une lecture républicaine et démocratique de nos institutions **qui hélas** sont aujourd'hui interprétées dans leur sens le plus monarchique (POI 95)

c'est certain que **lui en tout cas** me parle d'une manière enfin — j'espère qu'il me parle euh je l'imagine en tout cas (GRE 97)

Les énoncés retenus pour ce travail se situeront en dehors de ces cadres "clairs", ce qui requiert quelques explications complémentaires.

1.1. LES EXEMPLES SÉLECTIONNÉS

Le choix de travailler sur l'oral appelle un commentaire particulier. En effet plusieurs travaux ont montré que l'oral utilisait avec une certaine parcimonie les sujets lexicaux⁴. Il semble alors paradoxal de recourir à des données orales pour le travail envisagé (même si l'on a pris soin de travailler sur des corpus qui comportent de nombreux sujets lexicaux). Néanmoins, un tel choix est justifié à double titre. D'une part, la variété des séquences insérées est plus large à l'oral (elle inclut en particulier des éléments liés au discours spontané, habituellement gommés de l'écrit, cf. le point 2.1. ci-après); d'autre part, la description gagne, dans un premier temps, à porter sur des données brutes (i.e. moins maîtrisées et non retouchées, caractéristiques plus simples à obtenir à l'oral). De fait, comme dans de nombreux autres cas, l'abondance de certains tours à l'écrit rend finalement opaque leur distribution. Lorsque de multiples combinaisons se présentent, il semble plus difficile de faire ressortir les contraintes. L'oral offre un panorama plus sélectif ce qui permet de dégager plus nettement les lignes de force de la distribution.

Il a aussi paru plus prudent, pour cette première approche, d'opérer une sélection des exemples sur les bases suivantes :

3 La suite de ce travail précisera ce que recouvre cette notion de constituant.

4. Blanche-Benveniste (1994), pour une description récente.

- Seul le sujet lexical a été examiné. Il n'est en effet pas certain que d'autres réalisations (*qui, lui, etc.*) possèdent les mêmes propriétés.
- Le sujet lexical n'est pas repris par un clitique.
- Parmi les insertions, les relatives, compte tenu des nombreuses ambiguïtés auxquelles elles donnent lieu, ont été écartées.

Ces trois critères conduisent à ne pas conserver des exemples comme :

ils sont responsables de tout un tas de touristes **qui franchement** se ruent sur leurs euh bateaux (POI 97)

mon père tous les ans il nettoie la citerne (LIC 95)

en fait euh **Bernard de Clervaud** dont je vous ai parlé tout à l'heure sera aussi à cette entrevue (POI 96)

1.2. CRITÈRES DE DÉLIMITATION

En dépit de cette première sélection, le repérage des énoncés pertinents nécessite quelque précaution. Il convient en effet de vérifier si la suite linéaire qui précède le verbe est constituée de une ou de deux unités syntaxiques distinctes. En l'absence de paramètres tels que la ponctuation ou l'intonation, deux critères syntaxiques ont été utilisés pour sélectionner les exemples. Ils permettront de vérifier si la séquence repérée est effectivement située hors des limites du sujet. Il s'agit de :

a) l'extraction en *c'est... qui*. Elle permet de tester la cohésion des syntagmes candidats à être sujets. Quand aucun fractionnement n'est possible, c'est le syntagme dans sa totalité qui est traité comme le sujet :

la jeune fille en question était la serveuse qui la veille était dans cette crêperie (POI 97)

* **c'est** la jeune fille **qui** en question était la serveuse

c'est la jeune fille en question **qui** était la serveuse

Lorsqu'une scission est envisageable, ce test permet d'identifier comme sujet le segment encadré et de le dissocier de la séquence qui y est accolée. Ce deuxième constituant sera analysé comme une insertion :

le contenu *en tout cas* pourrait être très compromettant pour la la survie de la religion (POI 95)

c'est le contenu **qui** *en tout cas* pourrait être très compromettant

b) la proportionnalité à une proforme. Il s'agit de vérifier si la partie dégagée précédemment peut être en relation avec une forme autonome telle que *celui-ci* ou *lui* :

le contenu en tout cas pourrait être très compromettant
celui-ci en tout cas pourrait être très compromettant
lui en tout cas pourrait être très compromettant

Ces deux critères ne constituent toutefois pas des tests irrécusables. Ils laissent parfois subsister une marge d'incertitude. Les problèmes qui se manifestent sont dus à la portée de la séquence insérée. Deux difficultés majeures interviennent :

1. La double interprétation possible d'une construction

Deux solutions sont en concurrence selon le point de rattachement de la séquence "insérée" : au seul sujet ou à l'ensemble de la construction (seul ce dernier cas entre dans le cadre de ce travail). Les tests envisagés ne permettent pas, à eux seuls, de trancher entre ces deux situations. L'exemple ci-dessous en fournira une illustration :

bon le classement de la façade *au dix-neuvième siècle* va entraîner une restauration dans les années mille huit cent quarante-cinq mille huit cent cinquante (POI 96)

Une première interprétation conduit à analyser comme sujet l'ensemble constitué par la partie lexicale et la séquence en italiques. C'est ce bloc-là qui répond aux deux critères envisagés. On ne peut pas alors parler d'insertion entre le sujet et le verbe :

c'est le classement de la façade au 19^e siècle **qui** va entraîner une restauration
celui-ci va entraîner une restauration

Cette propriété se manifeste en particulier avec les nominalisations qui reconvertisent les places de construction du verbe original. C'est ce que l'on retrouve dans l'exemple ci-dessous :

classer + la façade + au 19^e siècle -> le classement de la façade au 19^e siècle

Dans une deuxième interprétation, la séquence en italiques sera interprétée comme un complément, de type *alors*, construit par le verbe *entraîner*. Cela conduit à ne retenir pour sujet que la partie lexicale qui précède et à considérer qu'il y a bien éloignement du sujet et du verbe. Cette vision est confirmée par les réponses aux deux critères utilisés :

c'est le classement de la façade **qui** *au 19^e siècle* va entraîner une restauration
celui-ci *au 19^e siècle* va entraîner une restauration

Le complément correspond à ce que la tradition identifie, de façon sommaire, comme un complément mobile à valeur temporelle. Cette difficulté se posera lorsque l'insertion est une séquence régie (cf. 3).

2. Le cadre dans lequel les tests sont pertinents.

Dans le cas de l'apposition, où l'insertion entretient un lien privilégié avec le syntagme sujet, les manipulations, parce qu'elles sont inefficaces à l'intérieur des constituants, fournissent des résultats plus que douteux. Il convient dès lors d'appréhender autrement ces énoncés (i.e. sans passer par des tests conçus pour opérer au niveau de la construction) pour pouvoir les intégrer dans notre corpus de travail. Ce dernier type d'exemples sera discuté ultérieurement :

- le héros *lui-même* met euh met pas mal de temps avant de pouvoir lui-même voir cette énergie (POI 96)
- ? c'est le héros **qui** *lui-même* met pas mal de temps
- ? **celui-ci** *lui-même* met pas mal de temps

1.3. DONNÉES ET ANALYSES

Les exemples retenus ont été puisés, pour l'essentiel, dans les quatre banques de données suivantes : POI 95 interviews d'hommes politiques, POI 96 guides sur le terrain, POI 97 résumé de films et récit de faits divers et LIC 95 regroupement plus hétérogène, soit plus de 200 000 mots. Sur ce corpus ainsi constitué divers éclairages seront envisagés. En prenant appui sur la syntaxe, nous proposerons un classement des différentes sortes d'insertions repérées. Sur la base du lien syntaxique que l'insertion entretient avec son entourage, nous distinguerons trois grands groupes : les associés (rassemblement assez disparate de séquences qui ne dépendent pas syntaxiquement d'un terme de la construction), les séquences régies par le verbe et celles placées dans la dépendance d'un nom. Nous chercherons alors à vérifier s'il existe une régularité liée à la provenance des exemples. Mais se satisfaire du relevé et du classement des différentes séquences qui s'intercalent entre le sujet et le verbe ne peut fournir qu'un résultat somme toute incomplet. Il apparaît ainsi trop réducteur d'isoler une construction verbale et de prétendre la décrire sans la relier au contexte dans lequel elle s'insère. Au-delà d'un simple regroupement formel, nous tenterons de faire émerger quelques propriétés de ces différents sujets "éloignés".

2. L'INSERTION D' ASSOCIÉS

Pour être classées dans les associés, les insertions doivent présenter les trois caractéristiques suivantes :

a) ne pas être proportionnelles à une proforme. La relation entre *en fait* et *ainsi* (ou une autre proforme) est jugée non satisfaisante, ce qui explique le rejet de l'exemple fabriqué :

- et euh euh Dieu *en fait* euh satisfera ses désirs (POI 96)
- * et Dieu *ainsi* satisfera ses désirs

b) ne pas pouvoir être extraites entre *c'est... que* :

- vous savez que nous avons signé un accord que la C.G.T. *d'ailleurs* n'a pas signé ni Force Ouvrière (POI 95)
- * *c'est d'ailleurs que* la C.G.T. n'a pas signé

c) ne pas être sensibles aux modalités du verbe (interrogation, contraste, etc.)

- on voit bien que le général *au fond* était contre tout ce qui pouvait faire écran (POI 95)
- * où le général était-il contre tout ? au fond
- * on voit bien que le général *au fond* mais pas de ce côté-là était contre tout...

2.1. LES DIFFÉRENTS CAS

À l'intérieur de ces séquences qui échappent à toute dépendance syntaxique, on peut repérer au moins trois réalisations catégorielles différentes qui correspondent à des fonctionnements particuliers.

2.1.1. LES MARQUEURS DE SEGMENTATION

À l'oral, les locuteurs recourent assez régulièrement à des marqueurs de segmentation récurrents (tels que *quoi*, *donc*, *hein*...). Ces éléments, généralement monosyllabiques, quelquefois regroupés sous l'appellation vague et commode de "particules" se caractérisent par une valeur sémantique quelque peu incertaine (il serait bien délicat de leur trouver un synonyme). Leur emplacement peut être soit assez spécialisé (*quoi* se réalise souvent en fin de construction ou de syntagmes⁵), soit plus libre (*hein*) :

5. Authier-Revuz (1995) écrit que "*quoi*, marque la clôture d'un mouvement méta-discursif d'incertitude, de réserve" (p. 128).

il faut avoir lu le bouquin pour comprendre le film *quoi* (POI 97)
 elles sont *en fait* elles sont elles sont destinées à à rester à l'intérieur de
 la maison (POI 97)

Le court passage suivant (a) permet de repérer des positions multiples qu'occupe l'élément *bon*. Les multiples "points de chute" de ces particules laissent bien évidemment attendre qu'elles puissent se trouver, notamment, entre le sujet et le verbe (b) :

- (a) les les gendarmes étaient là ceux qui avaient passé la nuit *bon* étaient assez frigorifiés car *bon* en février il fait il ne fait pas bien chaud — et je leur ai proposé de leur de leur apporter un thermos de café *bon* (POI 98)
- (b) **leur longueur** *bon* n'est pas rigoureusement la même à la syllabe près (LIC 95)

La densité d'emploi ainsi que le choix des termes varient selon les locuteurs. Un même marqueur peut ainsi revenir de façon récurrente. Voici l'exemple de *hein* pris chez le même locuteur :

il peut arriver que les les bateaux euh *hein* chavirent — c'est bizarre
 hein — (LIC 95, 71)
 c'est (là, la) (ville, Ø) la le petit village où la Sainte Vierge *hein* a donc euh euh a vu enfin a eu l'apparition de l'ange qui (est venu, venait) lui annoncer hein qu'elle allait être mère bon — (LIC 95, 72)
 il faut vraiment se rincer parce que l'eau euh *hein* est poisseuse (LIC 95, 73)

La présence de ces particules illustre un fonctionnement propre à l'oral à double titre. D'une part, ces termes qui scandent la production sont peu employés à l'écrit⁶. D'autre part, on ne retrouve pas d'équivalent à l'écrit. Il serait vain de rechercher un correspondant typographique factice (tel qu'une virgule) qui, à cet emplacement, est proscrit par la norme contemporaine. Les productions écrites présentent en quelque sorte une déficience puisqu'elles ne peuvent manifester dans cet emplacement un

6. Même dans des écrits qui cherchent à "mimer" un oral, plutôt caricatural d'ailleurs, ces éléments semblent peu attestés. L'énoncé cité, qui n'a pas valeur d'exemple, illustre que la forme employée (*là*) et sa distribution sont peu en phase avec la description précédente :

Qu'esc'y a ? Balise pas ! Si ça se trouve, tous les adjoints en ont une photocopie... Une chacun... Sois pas naïf... Je suis sûr que le Bertrand Machin, *là*, celui de la serviette, il est cent fois plus emmerdé que toi ... Tu sais quoi ? C'est un coup à se faire du blé... (Philippe Carrese. *Trois jours d'engatse*, p. 77, Fleuve Noir)

phénomène attesté à l'oral. Les jeunes scripteurs (de 10 ou 11 ans environ) qui commencent à employer la virgule dans leur productions spontanées n'en font, en général, pas usage à cet endroit-là.

L'oral dispose donc d'un procédé original : un indice de suspension dans l'enchaînement entre le sujet et le verbe. Mais l'apparition de ce marqueur ne semble pas corrélée avec l'ampleur du sujet invoquée par Girault-Duvivier (1811). Dans les exemples recueillis le sujet est souvent constitué d'un groupe réduit composé d'un déterminant et d'un nom. D'autres pistes ont été proposées, qui vont être maintenant examinées.

Camproux (1958) signale, dans certains usages du Gévaudan, des tours qui rappellent le problème présenté, même si les éléments de rupture entre le sujet et le verbe sont différents :

je le sais mon père que est parti (cité p. 342)

Dans un tel environnement, il faut traiter *que* comme une sorte de *phatique*. Camproux explique cette "fracture" entre les constituants par un souci d'expressivité : le verbe parce qu'il est isolé, séparé du sujet, ressort mieux.

Pour Fernandez (1994), l'organisation de l'information par les locuteurs devient le principe d'explication central. D'après elle, ces "particules énonciatives" ne doivent pas être envisagées comme un remplissage passe-partout d'une pause mais bien comme des balises qui servent à marquer "le bornage explicite d'un thème". L'analyse s'appuie sur le fait que, dans un énoncé sans rupture, le sujet grammatical remplira, selon le contexte, aussi bien le rôle de thème que celui de propos. La particule insérée établirait alors explicitement que le sujet a une valeur thématique. Qu'apporte une observation syntaxique à cette piste ?

Pour prendre en compte la valeur thématique d'un constituant, il est nécessaire d'élargir le contexte et d'observer comment est traitée la séquence lexicale. Apothéloz (1995) considère que dans les cas où la reprise pronominale est possible, le choix d'un sujet lexical au lieu du pronom est fondé sur des critères informatifs. L'anaphore nominale doit proposer un autre regard, un changement de point de vue. Le lexique et le clitique sujets rempliraient donc deux rôles distincts : le clitique traduit la continuité dans l'enchaînement des sujets, la poursuite de la même perspective alors que le lexique a plutôt tendance à marquer une rupture, une réorientation. Les sujets éloignés condenseraient alors deux fonctionnements distincts qui œuvrent dans le même sens : la catégorie lexicale explicite une rupture avec les sujets qui précèdent, l'insertion permet de faire mieux émerger le syntagme réalisé en position sujet.

Dans le passage suivant, le sujet éloigné reprend un syntagme qui possède déjà une “histoire syntaxique”⁷. Ce groupe est apparu quelques instants auparavant en position sujet. Les deux occurrences sont séparées par plusieurs constructions verbales. Se trouvent ainsi réunies les deux composantes envisagées : rupture avec le contexte immédiatement antérieur (où s'enchaînent des sujets *ils*) et reprise d'un groupe (à valeur thématique). L'insertion de *en fait* assure la mise en valeur de ce sujet “particulier” :

alors euh c'est l'histoire d'un couple qui au départ euh vit euh aux États-Unis où tout se passe bien et bon lui euh **le mari** travaille dans dans = un hôpital et je sais plus euh ce que fait la femme mais enfin c'est pas ce qui importe et euh ils ont une petite fille donc euh tout tout se passe très bien leur vie de couple leur vie de famille et ils ont des amis et euh oui ce qu'il faut préciser c'est que **le mari en fait** est est musulman — et euh mais aux États-Unis ça ça leur pose pas de problème (POI 97)

Le passage suivant reproduit le même schéma : le marqueur de segmentation opère sur du lexique déjà utilisé une dizaine de lignes au-dessus. On retrouve, à nouveau, ce double fonctionnement : rappel d'une occurrence antérieure et rupture avec le contexte immédiat.

eh bien les les **les corps** ont été retrouvés donc les derniers corps à savoir ont été les corps euh des enfants puisque les enfants sont les plus légers ils ont ils ont eu du mal à revenir à la surface donc ils ont été retrouvés [...] donc euh voilà — donc quoi rajouter euh donc que **les corps bon** sont revenus petit à petit (POI 97)

Dans l'exemple ci-dessous, l'emploi de *donc* (avec sa valeur de validation) vient encore renforcer le fonctionnement décrit :

alors donc **cette cette frontière** qui avait qui avait on a on est on est toujours à la frontière allemande (Ø, et) tchécoslovaque — **cette frontière donc** avait quand même quelque chose de d'extraordinaire (LIC 95)

Dans cette situation, les exemples relevés acceptent très facilement le couplage du lexique avec un sujet clitique, ce qui traduit la valeur “thématique” des formes lexicales employées :

le mari en fait il est est est musulman
voilà — donc quoi rajouter euh donc que **les corps bon ils** sont revenus petit à petit

7. Blanche-Benveniste (1995).

Mais pour pouvoir décrire ce fonctionnement régulier, deux contraintes doivent être posées :

a) Il est nécessaire d'écartier certaines productions qui regorgent d'éléments de segmentation. Les locuteurs les utilisent en effet de façon si dense qu'aucune régularité n'apparaît et que les deux caractéristiques signalées n'apparaissent plus. En voici un exemple :

et à ce moment-là *bon ben* le le placement est plus un rapport euh qualité-sécurité — donc là c'est pas un objectif *en fait* euh régulier d'épargne c'est quoi c'est euh de réaliser un bon coup financier *en fait* alors *bon* dans le contexte économique actuel je crois que (là, la) les taux d'intérêts *donc* baissent partout — et il y a un certain scepticisme *en fait* de toutes les personnes que je rencontre (LIC 95)

Cette observation limite quelque peu les conclusions de Fernandez (1994) qui ne semble pas envisager de “trous” dans sa distribution ou d'analyses différenciées selon les occurrences des “particules”.

b) Certains termes tels que *en fait* se rattachent à deux ensembles : les marqueurs de segmentation ou les éléments de validation (cf. ci-après). Ils possèdent donc deux fonctionnements distincts. C'est pourquoi un exemple tel que :

en fait euh d'après certains euh historiens et d'après euh ce que pensait aussi ce religieux Aliénor d'Aquitaine aurait favorisé euh euh certaines violences occasionnées par le roi il va lui demander donc de favoriser la paix la paix et euh euh Dieu *en fait* euh satisfera ses désirs et un an après allait naître une fille (POI 96)

dans lequel le sujet ne reprend pas un syntagme déjà utilisé dans le contexte antérieur ne constitue pas une exception ou une anomalie par rapport à la description mais illustre un autre cas d'éloignement du sujet. Une description syntaxique minutieuse conduit, de fait, à réduire l'importance de certains fonctionnements posés comme généraux et constants dans les approches informatives. D'où, certainement, l'importance de “pousser à bout” la syntaxe, de décrire au mieux les énoncés pour en comprendre les mécanismes.

2.1.2. LES ÉLÉMENTS DE VALIDATION

Par cette désignation bien générale on souhaite regrouper des associés réalisés sous une forme lexicale non verbale (adverbe, syntagme) tels que :

auparavant donc euh le le bouquin et le film *de de ce côté-là* nous tiennent en haleine (POI 97)

elles s- sont toujours en bisbille avec Theodora parce que Eleonor *forcément* est p- pas super jolie (POI 97)

et alors au fond ce qui est intéressant c'est que on voit bien que le général *au fond* était contre tout ce qui pouvait faire écran si j'ose dire (POI 95)

Le fonctionnement de ces associés se distingue nettement du cas précédent. Les séquences insérées jouent un double rôle :

- elles possèdent le fonctionnement des éléments qui portent non sur le contenu de l'énoncé mais sur l'énonciation elle-même. Les fameux "adverbes de phrase" prendraient place ici.
- elles semblent entretenir avec le sujet lexical une relation étroite liée, semble-t-il, à leur voisinage.

Il est assez difficile, sur des exemples isolés, de montrer que ces éléments présentent de légères variations de sens selon la position dans laquelle ils se trouvent. Toutefois quelques indices étayent l'hypothèse avancée :

- on trouve dans cet emplacement les adverbes paradigmatiques décrits par Nølke (1993) qui présentent un lien étroit avec le lexique auquel ils s'accolent :

la gouvernante *aussi* ressemble beaucoup beaucoup à Eléonor (POI 97)

- le cumul de deux termes produit des résultats différents selon le classement des éléments insérés⁸.

- La succession de plusieurs éléments de validation est possible :

donc bon le la première approche est difficile déjà euh — par la langue (POI 97)

et *donc bon ben* euh ils apprennent à faire connaissance en fait (POI 97)

- La cohabitation d'un terme de segmentation et d'un élément de validation ne provoque pas de réticence :

et *donc finalement* ils s'engueulent et elle prend la voiture elle s'en va quoi (POI 97)

donc ensuite forcément ils ont été euh condamnés (POI 97)

8. Dans ce paragraphe, les exemples attestés conduisent à observer le cumul en dehors de l'emplacement entre le sujet et le verbe.

Cette suite semble acceptable quand on la place entre le sujet et le verbe, quel que soit l'ordre de succession retenu :

chaque guide *donc justement* reste avec son groupe
chaque guide *justement donc* reste avec son groupe

Il est probable que ces éléments de validation s'organisent en paradigmes, ce qui permet de rendre compte des possibilités d'empilement et de non cooccurrence constatées⁹. Si l'on rattache *aussi* et *d'ailleurs* à deux paradigmes différents, on pourra justifier de leur juxtaposition dans l'exemple forgé ci-dessous :

la C.G.T. *aussi d'ailleurs* n'a pas signé
la C.G.T. *d'ailleurs aussi* n'a pas signé

Le non cumul de *visiblement* et *au demeurant* pourrait être expliqué par leur appartenance au même paradigme :

? monsieur Pasqua *visiblement au demeurant* est démangé par l'envie
d'y aller

Si l'un des termes occupe une autre place, l'énoncé ne suscite pas les mêmes réserves :

au demeurant monsieur Pasqua *visiblement* est démangé par l'envie
d'y aller

Cette contrainte semble indiquer que l'élément de validation, dans cet emplacement, entretient une relation privilégiée avec la séquence qui précède. Cette relation n'est pas d'ordre syntaxique, puisque l'élément appartient aux associés. Pour autant la combinatoire n'est pas libre ni indifférente.

Nølke (1993) décrit les différentes positions que peuvent occuper certains adverbes. Pour un énoncé comme :

Pierre, cependant, a vendu sa voiture (cité, p. 59)

9. La coordination est un critère difficile à maîtriser dans ce cadre. En effet, si elle passe bien dans l'exemple :

mais l'élection présidentielle est aussi et d'abord un choix de personne
qui se fait sur le fond (POI 95)

Elle semble moins naturelle entre le sujet et le verbe :

? mais l'élection présidentielle aussi et d'abord est un choix de personne

Il devient dès lors délicat de s'en servir comme test d'appartenance au même paradigme.

Il considère que le sujet est thématique mais possède un certain degré de focalisation. La mise en relief opérée ici ne repose pas sur un matériau morphologique supplémentaire. Elle est toutefois manifeste dès que l'on replace l'énoncé dans un contexte élargi. Un autre emplacement de l'adverbe pourrait aller, selon Nølke, jusqu'à détruire la cohérence textuelle.

Il s'appuie sur deux séries d'arguments :

a) d'ordre intonatif

Il observe alors l'effet que l'insertion provoque sur la courbe intonative. L'intonation plate sur l'adverbe pourrait aider à la perception de l'intonation terminale sur le sujet.

b) d'ordre paradigmatique

Les éléments qui peuvent remplir cet emplacement partagent un même usage : ils sont spécialisés dans la focalisation (tels que *lui*, *pour sa part*...). D'un point de vue lexical ou catégoriel, on trouvera dans cette position des connecteurs et des factifs (*heureusement*, *naturellement*...), et des termes qui possèdent une valeur oppositive : des connecteurs adversatifs (*par contre*, *au contraire*...) ainsi que des concessifs (*pourtant*, *cependant*...). Seuls les modaux (*peut-être* dans certains emplois, ...) y sont rares.

Si l'on élargit le contexte, le fonctionnement des sujets éloignés par ces éléments de validation apparaît plus divers que dans le cas précédent.

Ainsi, le groupe sujet peut être une séquence nouvelle qui n'a pas d'existence dans le contexte antérieur. L'exemple suivant illustre bien cette possibilité. Ce passage se trouve en début de corpus. L'insertion de la séquence *en tout cas* permet de faire émerger le caractère focalisé du sujet et d'accentuer l'opposition établie entre le film et le livre qui sont présentés :

euh bah je vais te parler du film La Maison Du Diable de Robert Wise
— qui a été tourné euh peut-être il y a longtemps et euh **le livre** *en tout cas* a été écrit en 1959 donc c'est un film en noir et blanc (POI 97)

On trouve aussi ce phénomène en début de prise de parole de locuteur. La séquence sujet a déjà été utilisée à de multiples reprises dans le contexte qui précède. Une double rupture va être marquée grâce à l'emploi d'une séquence lexicale (vs la poursuite du pronom) et l'insertion d'un terme de validation. Ainsi marque-t-on une opposition singulière (en partie factice et en quelque sorte rhétorique) qui repose sur des outils syntaxiques atténués :

- L1 oui il [le général de Gaulle] croyait que les peuples demeurent euh les peuples demeurent parce que il disait c'est comme le cèdre du Liban ou comme l'olivier de Platon qui a deux mille cinq cents ans euh ça demeure euh on ne peut pas effacer un peuple mais une nation avec un état pour la défendre il n'est pas sûr que ça demeure et il voulait lutter pour la pérennité de la France et il voulait prendre des dispositions qui rendent la pérennité de la France irréversible et je crois que euh c'est la raison fondamentale pour laquelle il était contre la supranationalité et contre tout ce qui tendait à confondre la France avec les autres et à supprimer la légitimité d'un pouvoir national français
- L3 et puis le **général de Gaulle** aussi avait le sens de la formule formules qui ont toujours fait mouche je vous cite encore je cite vos notes (POI 95)

Le lexique est parfois récupéré du contexte antérieur. Mais son caractère inédit tient à de multiples changements apparus entre les deux occurrences (changement de position syntaxique, de déterminant...).

et après le le petit enfant il a cherché euh dans son placard à jouets euh et il a trouvé **une tente une tente** qu'il a qu'il a donnée euh qu'il a mis sur la euh à côté de l'indien et puis l'indien il avait peur il voulait pas trop toucher et puis euh et il avait peur alors il voulait pas dormir dedans alors le le petit garçon il a eu une idée et il l'a mis dans son placard il a refermé la clef il a ouvert son placard et **la tente bien sûr** était devenue réalité — (POI 97)

L'exemple suivant illustre ce double mécanisme de récupération et transformation. Pour cela, il est nécessaire de prendre un contexte relativement large. Le sujet lexical qui apparaît à la fin de l'extrait récupère des éléments fournis au début du passage dans des termes proches. Mais plus que le phénomène de reprise, c'est le remodelage qui est important. On passe d'une construction verbale à un SN en position sujet qui va permettre d'activer une opposition entre personne physique (au début du passage) et l'image de l'individu (à la fin de l'extrait). Les éléments lexicaux en place dès le début de l'extrait ne permettraient pas de jouer sur cette distinction. L'insertion de l'associé (*effectivement*) ainsi que la nouvelle présentation du lexique dans un schéma syntaxique différent rendent explicite cette opposition :

il y a sûrement des choses qui tiennent à monsieur Balladur lui-même — *monsieur Balladur a fondé son image* sur deux affirmations — la première c'est je gouverne et je ne m'intéresse pas à l'élection présidentielle

- L4 et vous y croyez

L2 la deuxième — c'est — je ne vous fais pas de promesses mais je réussis or qu'est-ce qu'on constate on constate qu'il ne gouverne plus mais qu'il ne pense qu'à l'élection présidentielle et on constate qu'il ne réussit pas et donc je pense que **l'image de monsieur Balladur effectivement** est en train de se modifier dans le pays (POI 95)

Dans le cas des éléments de validation, le sujet éloigné accepte moins facilement le couplage avec une forme clitique (il faut bien évidemment replacer ces exemples dans leur contexte pour juger de leur caractère maladroit) :

- ? et Dieu *en fait* il satisfera ses désirs
- ? et la tente *bien sûr* elle était devenue réalité

Finalement, malgré la diversité apparente des situations observées, le statut focalisé du sujet semble constituer le point commun aux énoncés rassemblés dans cette partie. Une double conclusion se dégage : l'organisation avec élément de validation se distingue bien des cas avec marqueur de segmentation; la structure informative de l'énoncé complète de façon pertinente la description syntaxique de ce dernier type d'exemples.

2.1.3. LES COMMENTAIRES

Dans cette rubrique, on regroupera toutes les interventions que les locuteurs produisent pour commenter les propos tenus, justifier ou situer le choix lexical effectué, etc. Ce type d'intervention¹⁰ est fort répandu dans les productions orales et se réalise de multiples façons. Dans les exemples relevés pour ce travail, elles se présentent, à la différence des autres insertions, sous forme de construction verbale :

donc l'île *comme je te disais tout à l'heure* a été très touchée (POI 97)
 mais les les ins- les institutrices — *comme on disait à l'époque* parlaient
 quand même le français (POI 97)

Le fonctionnement est celui des parenthèses : la séquence s'intercale dans une construction verbale sans lui être syntaxiquement reliée, malgré l'aspect trompeur de certains morphèmes introducteurs (*parce que*, *comme*, etc.); sémantiquement ces séquences apportent une information décalée par rapport au reste de la construction. L'un des aspects remarquables de ces insertions tient à leur envergure. Certaines séquences inter-

10. Il s'agit en général de ce que Authier-Revuz (1993) identifie comme des "boucles du dire" i.e. des "ouvertures sur l'activité des énonciateurs aux prises avec les mots" (p. 101).

rompent parfois longuement la construction verbale sans que le locuteur présente des difficultés à retrouver le fil de son discours :

toute personne qui est en situation de — difficulté quant à son logement
je parle surtout pendant l'hiver c'est ce qu'on appelle les sans domicile fixe est aussitôt accueillie euh reconnue euh aidée (POI 95)

le mari part à peine que là euh l'esprit frappeur *parce que finalement on on nous révèle que c'est un esprit frappeur* casse toutes les vitres les unes après les autres et euh et se encore une fois se jette sur elles (POI 97)

Plusieurs situations différentes se rencontrent dans ce groupe. Le lexique peut avoir déjà été mentionné dans le contexte antérieur. Le commentaire vient en quelque sorte valider a posteriori le choix lexical déjà effectué et confirmé par cette deuxième occurrence :

après il est venu des institutrices — euh pas forcément du non pas d-pas de la commune [...] mais les les ins- **les institutrices** — *comme on disait à l'époque* parlaient quand même le français (POI 97)

À l'autre extrême, le lexique sujet peut n'avoir pas été utilisé dans le contexte antérieur. Il est introduit pour la première fois (déterminant indéfini). Le couplage avec un clitique serait alors peu heureux :

au bout d'un moment euh une des une des — une des sœurs je crois ou ou une des s- une belle-sœur *je sais plus trop le lien de famille euh* — essaie de l'aider (POI 97)

? **une belle-sœur je sais plus trop le lien de famille euh** — elle essaie de l'aider

Mais le caractère “nouveau” du lexique peut tenir essentiellement au commentaire que le locuteur porte sur le choix des termes qu'il utilise. On peut, en détournant la formulation de Apothéloz et Reichler-Béguelin (1995), dire que le locuteur intervient dans le processus de désignation. La mention dans le contexte antérieur ne joue alors qu'un caractère accessoire :

le mari part à peine que là euh **l'esprit frappeur** *parce que finalement on on nous révèle que c'est un esprit frappeur* casse toutes les vitres les unes après les autres (POI 97)

? **l'esprit frappeur** *parce que finalement on on nous révèle que c'est un esprit frappeur* il casse toutes les vitres

les — deux deux gendarmes en en tenue en treillis *disons en quelque sorte avec euh de grands sacs s'apprêtaient donc à aller examiner les les lieux* (POI 98)

Le fonctionnement de ces insertions paraît conduire, à première vue, à une situation inverse du cas précédent. Le sujet est suivi d'une séquence à l'intonation généralement plate, qui d'une certaine façon le renvoie au second plan. Dans les exemples cités, l'insertion est longue, le sujet n'est pas repris par un pronom. Il n'y a plus ici de morphème ou de séquence proéminente destiné à servir de repère intonativement saillant.

On mesure ici que l'explication à partir de la structure informative semble plus fragile et ne constitue pas un critère de regroupement suffisamment fort des énoncés.

2.2. RÉPARTITION DANS LES CORPUS

La partition opérée à l'intérieur des séquences associées qui prennent place entre le sujet et le verbe trouve un appui dans notre relevé d'exemples. En effet, les trois types présentés ne sont pas répartis de façon équitable dans les corpus.

Ainsi le recueil POI 95 qui regroupe des interventions d'hommes politiques à la radio ou à la télé fournit des phénomènes parenthétiques originaux qui tiennent :

a) à la situation d'interlocution dans laquelle étaient placés les participants. C'est pourquoi l'on y trouve des renvois précis aux locuteurs :

notre seul pouvoir et je les cite est de fermer les équipements dangereux (POI 95)

vous lui prêt- enfin — **les historiens** et vous (vous, Ø) abondez dans ce sens lui prêtent un grand dessein européen la réconciliation de la chrétienté (POI 95)

d'abord le **ministre de l'Éducation** — *contrairement à ce que vous croyez* ne règle pas l'emploi du temps de toutes les écoles de France (POI 95)

b) à la longueur de certaines insertions. La transcription permet de mieux mesurer l'amplitude de ces coupures qui n'affectent pas, en apparence, la continuité de la production. Rien ne manifeste une difficulté du locuteur à retrouver le fil syntaxique de son énoncé. Le sujet notamment n'est pas repris (sous forme clitique ou lexicale) à la fin de l'insertion :

en 1984 tout le monde parlait d'eurosepticisme — d'euroessimisme et aujourd'hui même si **les opinions publiques** — *parce que l'Europe précisément a pénétré dans leur vie professionnelle et dans leur vie privée* - sont plus soucieuses plus incertaines il n'empêche que l'Europe a progressé (POI 95)

La proportion de commentaires est plus élevée dans ce volume. Les éléments de segmentation, à l'inverse, ne tiennent qu'une place marginale. Les locuteurs de POI 95 multiplient les parenthèses (qui servent à préciser ou à multiplier les points de vue, à prendre appui sur le lexique pour en modifier le rôle, ...). Ce type d'insertions entre le sujet et le verbe pourrait être une caractéristique propre à des locuteurs professionnels qui doivent redoubler de prudence dans des interventions publiques. Sa rareté dans les autres corpus utilisés le rend intéressant comme critère attaché à un genre précis :

dès lors que **des — hommes** *ce sont des hommes /de la circonstance, XXX/ ont été déjà sanctionnés punis par de la prison (POI 95)*

d'abord **monsieur Pasqua** lui-même — *qui confond un peu bien qu'il s'efforce de s'en défendre sa fonction ministre de l'Intérieur et sa fonction de président de l'association pour les primaires et c'est un problème quand même en république — que de trouver le même homme euh jouer euh le lobbyiste dans un cas et le ministre de l'intérieur dans l'autre — a dit lui-même voyant les réticences les réactions des grands juristes y compris des juristes qui ne sont pas proches de la gauche (POI 95)¹¹*

En définitive, on obtient un discours qui, sur le plan syntaxique, comporte de multiples décrochages dans la linéarité et qui multiplie les parenthèses et les digressions¹² :

ce que j'ai fait dans ma commune *je peux vous dire évite je crois à n'importe quel citoyen d'être ce soir alors sauf si vraiment il passe à travers les mailles du filet mais d'être ce soir en situation de détresse absolue (POI 95)*

Les volumes POI 96 (guides sur le terrain) et POI 97 (résumés de films) offrent une répartition plus équilibrée des différentes sortes d'insertions identifiées.

11. Certes, l'exemple sort des limites tracées. Mais il permet de mesurer l'ampleur de ce type d'insertions.

12. Une autre caractéristique, qui sort des limites fixées au départ de ce travail, concerne le nombre d'insertions après le relatif *qui*. Elles sont nettement plus fréquentes dans le volume POI 95 (Hommes politiques) :

il y a quelque chose de plus **qui peut-être** peut se nouer autour de — la candidature de Jacques Delors enfin si elle est avérée (POI 95)

2.3. PREMIER BILAN

L'analyse présentée a permis de distinguer trois sortes d'associés qui peuvent "s'insinuer" entre le sujet et le verbe. Dans les pages précédentes, on a essayé de dégager les différences de fonctionnements et d'observer ce que les approches en terme de structure informative apportaient comme regard complémentaire à l'analyse syntaxique. Avant de poursuivre cet examen des constituants qui s'insèrent entre le sujet et le verbe, il semble utile de mettre l'accent sur deux caractéristiques que l'on ne retrouvera pas par la suite : les insertions ont pu être observées in situ, il faudra à l'avenir prendre en compte leur cheminement syntaxique dans des contextes plus larges (comme cela a déjà dû être envisagé pour le sujet); le sujet isolé du verbe par un associé n'est pas affecté de façon interne par cette insertion. Il s'agit là d'aspects qui feront l'objet de discussions dans la partie suivante.

3. LES SÉQUENCES RÉGIES PAR LE VERBE

Pour être considérées comme régies, les séquences "insérées" doivent répondre de façon positive aux trois critères suivants :

a) être proportionnelles à une proforme

je pense que le Parti communiste *à un certain moment de notre histoire* a joué un rôle national in- contestable (POI 95)

je pense que le Parti communiste **alors** a joué un rôle incontestable

b) pouvoir être placées entre *c'est... que*

Jean de Berry *au quatorzième* l'a agrandi et transformé (POI 96)

c'est au quatorzième **que** Jean de Berry l'a agrandi et transformé

c) être sensibles aux modalités du verbe

ma propre mère *quelquefois* s'est tue devant des gens (POI 97)

ma propre mère *quelquefois* s'est tue devant des gens **mais pas toujours**

Les séquences intercalées appartiennent à la réaction au sens large. Aucun exemple de complément valencielle (en dehors des clitiques, bien évidemment) n'a été recueilli. Le seul cas original concerne un complément lié au superlatif :

non mais — raisonnons globalement — jusqu'à aujourd'hui — l'Europe — *de tous les grands continents* est le seul — à être un continent nain un continent mineur dans le domaine de la sécurité — (POI 95)

Les groupes intercalés sont proportionnels à des proformes comme *là, ainsi, alors*, etc. qui permettent de fixer leur valeur sémantique :

et ces clients-là *petit à petit* sont devenus des amis (LIC 95)

et justement ces deux-là en question — *un jour* ont reconnu le gardien de la maison de redressement qui leur faisait subir toutes ces tortures et tous ces viols — (POI 97)

Pour rendre compte de l'insertion d'associés entre le sujet et le verbe, il a fallu établir un classement formel des associés et observer certaines caractéristiques du sujet. Lorsque l'élément intercalé est un constituant régi, il convient en outre de prendre en compte les occurrences antérieures, "l'histoire syntaxique" de cette séquence. En effet, ces constituants rectionnels insérés ont généralement été utilisés dans le contexte antérieur (Blanche-Benveniste, 1997). Ils peuvent être récupérés soit sous la même forme :

c'est l'histoire d'une f- d'une femme qui vit *en Amérique* avec son mari et euh la vie euh *là-bas en Amérique* est très bien ils sont heureux et tout (POI 97)

ça n'était pas possible de leur permettre de s'installer *en France* ce qu'ils n'auraient pas manqué de faire puisque le niveau de vie *en France* était beaucoup plus élevé qu'en Algérie (POI 95)

soit sous une apparence modifiée qui traduit le changement opéré lors de la deuxième mention. C'est alors une partie du lexique qui est reprise ou reformulée :

et quand il y aura plus de message ben *le compte à rebours* sera fini et euh mh donc les vaisseaux *à la fin du compte à rebours* (euh, Ø) détruisent euh toutes les grandes villes (POI 97)

Certaines propriétés des éléments régis semblent liées à la position qu'occupe le complément¹³ :

13. Le travail sur ces séquences est délicat puisqu'il remet en cause des "évidences" d'analyse qui laissent entendre que la place de certains compléments n'affecte pas leur statut syntaxique. Les analyses en thème et rhème ont bien montré que les présupposés étaient différents.

a) La plus générale, qui concerne toutes les séquences régies, tient à la proportionnalité avec certaines proformes. En position initiale ou finale, le complément entre dans un paradigme complet :

dans les jours qui suivent on a appliqué les consignes
alors on a appliqué les consignes
quand on a appliqué les consignes ?
 on a appliqué les consignes *quand* ?

Intercalé entre le sujet et le verbe, le complément entre dans un paradigme réduit qui exclut la relation avec les interrogatifs :

le conseil général de l'époque *à partir de mille huit cent trente-trois* avait décidé d'y installer une sorte de musée charentais (POI 96)
 le conseil général de l'époque *alors* avait décidé d'y installer un musée
 * le conseil général de l'époque *quand* avait décidé d'y installer un musée

b) Dans le cas de cumuls de compléments, les restrictions sont plus lourdes pour la position d'insertion :

avant ton arrivée *mardi* mes amis sont venus
 * avant ton arrivée mes amis *mardi* sont venus
mardi mes amis avant ton arrivée sont venus

c) La position initiale permet à la séquence de porter sur plus d'une construction verbale comme dans :

puis *en quelques minutes* le jour tombera et les lampes à pétrole s'allumeront

Lorsque le complément est réalisé à droite du verbe il porte sur un seul processus verbal, il n'est pas partageable :

puis le jour tombera *en quelques minutes* et les lampes à pétrole s'allumeront
 puis le jour tombera et les lampes à pétrole s'allumeront *en quelques minutes*

Placé entre le sujet et le verbe, la séquence temporelle présente cette particularité d'être enfermée dans une construction verbale précise :

puis le jour tombera et les lampes à pétrole *en quelques minutes* s'allumeront

D'autres caractéristiques conduisent à poser deux types distincts dans les insertions de constituants régis.

3.1. L'INSERTION JUXTAPOSÉE

L'apparente redondance de cette dénomination indique que, dans ce premier cas, le constituant placé entre le sujet et le verbe ne semble pas changer d'interprétation s'il est placé ailleurs dans l'énoncé :

et justement **ces deux-là en question** — *un jour* ont reconnu le le gardien de la maison de redressement (POI 98)

et justement *un jour* **ces deux-là en question** ont reconnu le le gardien de la maison de redressement

L'interprétation du sujet ne paraît pas non plus affectée par la place qu'occupe ce complément. Qu'il soit avant ou après le sujet, la différence d'interprétation de l'énoncé est difficile à percevoir :

alors **l'individu ensuite** euh — a cherché à nous nous enfermer dans un local quelconque (POI 97)

alors *ensuite* **l'individu** euh — a cherché à nous nous enfermer dans un local quelconque

Dans ce type d'insertion, le complément peut être décrit comme étanche à double titre : son sens n'est pas affecté, de façon flagrante, par sa position et il ne "contamine" pas le SN sujet auprès duquel il est placé.

3.2. L'INSERTION SOUDÉE

À l'inverse, le complément rectionnel peut montrer une grande sensibilité à l'emplacement qu'il occupe. La position du complément modifie parfois son interprétation et sa portée. La démonstration reposera sur deux points :

1. Le changement de sens du complément rectionnel

Certaines rections de type temporel ou locatif font mieux ressortir la variation de sens liée à la position. Divers arguments peuvent être avancés :

a) La position intermédiaire (entre le sujet et le verbe) est compatible avec deux interprétations là où la position initiale ne retient qu'une valeur. On le voit nettement en travaillant sur des paires d'énoncés.

Dans une première version, *ici* peut renvoyer :

- soit à la situation générale du locuteur (*la ville où il parle*), cette interprétation s'accompagne généralement d'un décrochage intonatif,
- soit être liée au démonstratif et fonctionner comme équivalent de *ce mur-ci*.

et **ce mur** *ici* est tout à fait célèbre pour les personnes qui font des études d'architecture (POI 96)

Dans la deuxième formulation, l'interprétation localisante (*dans cette ville*) est privilégiée et s'impose d'emblée, sauf contexte particulier qui activerait, avec difficulté toutefois, la seconde possibilité :

et *ici* **ce mur** est tout à fait célèbre

b) Il est connu que les enchaînements attendus sont différents selon la position qu'occupe la séquence "régie". Dans la première formulation, la séquence *au-dessus* a une valeur locative. Le prolongement de l'exemple pourrait s'appuyer sur l'alternance *au-dessus* / *au-dessous* avec conservation du même sujet :

au-dessus les ébrasements et les voussures sont maçonnés en avant du mur

au-dessus les ébrasements sont maçonnés en avant du mur *au-dessous* ils sont moins travaillés

Dans le deuxième exemple, la position qu'occupe cet élément locatif ouvre une liste implicite des points architecturaux observés. L'enchaînement conduira plutôt à modifier le syntagme sujet et à ne pas activer l'opposition *au-dessus* / *au-dessous* :

les ébrasements et les voussures *au-dessus* sont maçonnés en avant du mur (POI 96)

les ébrasements et les voussures *au-dessus* sont maçonnés en avant les contreforts sont moins travaillés

c) Certains exemples permettent d'approcher ces variations sémantiques liées à la place du "complément". Ainsi la séquence en "quand..." présente un caractère de probabilité plus ou moins fort selon sa position dans l'énoncé. Dans la première version la construction en *quand* est proche d'une construction en *si*. L'existence du SN sujet *la lecture* est alors incertaine, hypothétique¹⁴ :

Mais **cette lecture**, *quand elle a lieu*, se borne quasiment à un ou deux spécimens de la production saisonnière. (Revel. *Le voleur dans la maison vide*. p. 45)

14. Comme dans l'exemple vu en (a), une deuxième interprétation, temporelle cette fois-ci, est envisageable. Elle semble toutefois moins immédiate.

Le deuxième exemple s'écarte de cette interprétation. Le complément a une valeur temporelle et dans l'énoncé l'existence de *cette lecture* n'est pas contestée, ni discutée :

Mais quand elle a lieu **cette lecture** se borne quasiment à un ou deux spécimens

2. L'influence de la rection sur le sujet

Le dernier exemple peut servir de transition entre les deux parties de l'argumentation. La fluctuation dans la valeur du complément affecte indirectement l'interprétation du SN sujet (*cette lecture*), puisque, comme on l'a vu, l'existence du référent est plus ou moins bien établie. D'une certaine façon, le sujet est donc atteint par le groupe rectionnel. Ces faits sont explicables dans deux cadres distincts :

- soit il s'agit d'un problème de portée. Le découpage syntaxique est inchangé quelle que soit la position du complément rectionnel. La dissociation est alors nette entre le statut syntaxique du complément (rattaché au verbe) et sa portée sémantique (due à un lien étroit avec le sujet). L'inconvénient de cette approche réside dans la notion même de "portée" qui s'avère plutôt vague (Daneil, 1998).
- soit il faut traduire dans l'analyse syntaxique les liens particuliers observés. C'est la solution qui sera privilégiée dans les pages suivantes, avec comme question centrale : faut-il remettre en cause le découpage syntaxique de l'énoncé, là où a priori il y a insertion d'un groupe rectionnel entre le sujet et le verbe ?

Les modifications dans le fonctionnement et l'interprétation de la séquence rectionnelle selon sa position incitent à explorer l'hypothèse que la structure syntaxique a été modifiée. Dans l'exemple :

il ne voulait pas le croire euh bien que **les sondages dans les deux ou trois jours précédents** euh l'eussent annoncé il se refusait à le croire (POI 95)

l'insertion permet en fait de placer le sujet dans une sorte de processus différent de celui exprimé par le verbe. La séquence *dans les deux ou trois jours* semble plus en accointance avec le SN *les sondages* qu'avec le processus verbal.

Comment traduire syntaxiquement cette présentation ? L'une des solutions serait de réanalyser le sujet : ce ne serait pas la séquence lexicale brute (limitée au seul SN) mais la séquence lexicale étoffée (composée de l'ensemble SN + insertion). Pour cela, il convient de bousculer quelque

peu certains cadres habituels et d'envisager une relation à moyen terme entre :

- le complément de nom (*les sondages des 2 ou 3 derniers jours*) qui réalise une sorte de fusion, de dépendance complète,
- et le complément de verbe (*dans les deux ou 3 derniers jours, les sondages...*) qui n'atteint pas le sujet¹⁵.

L'insertion (*les sondages dans les 2 ou 3 derniers jours*) ouvre la possibilité d'une nouvelle relation qui s'écarte d'un modèle de type binaire¹⁶ peut-être trop réducteur.

Cette hypothèse permet de mieux traiter les modifications qui atteignent le SN. C'est une sorte de sujet "dynamique", aspectuel qui est utilisé. D'une certaine façon, l'élément inséré reporte sur le sujet son sémantisme, il fournit l'angle sous lequel doit être envisagé le groupe qui précède. Ainsi, ce n'est pas le *beau-frère* en tant qu'individu mais le beau-frère transformé par son ivresse qui *pique une crise* :

le lendemain matin euh au réveil — le **le beau-frère** *encore sous l'effet de l'alcool* pique euh une une crise en apprenant que euh que sa sa sœur n'est pas mariée (Mystery Train)

3.3. CONCLUSION PARTIELLE

L'analyse des insertions qui appartiennent à la rection met à jour de nombreuses caractéristiques originales. Elle conduit d'une part à s'intéresser au "passé syntaxique" de ce constituant et à rechercher sous quelle forme il a pu être introduit dans le contexte antérieur. D'autre part, certains exemples possèdent deux interprétations qui correspondent à deux analyses différentes.

Aussi a-t-il fallu poser deux types d'insertion : la juxtaposition, qui laisse intacts le sujet et la rection placés côte à côte, et la fusion qui fabrique un sujet complexe dans lequel les deux groupes enchaînés forme

15. Chervel (1998) fournit un éclairage original sur le problème étudié dans ces pages. À travers les versions latines, il observe les constructions syntaxiques qui sont favorisées dans les copies d'élèves. Il est conduit à identifier une sorte de "modèle" de la prose classique qui tend à occuper l'emplacement entre le sujet et le verbe :

Caracalla pour diminuer l'horreur de son action, mit Geta au rang des Dieux (Montesquieu) exemple cité par Chervel

Comme il l'indique, on connaît finalement assez mal la distribution des compléments qui ne peuvent occuper cet emplacement et doivent forcément être placés en tête.

16. On pense aussi à l'opposition explicative / déterminative, qui pourrait être utilisée pour décrire certains exemples. Là encore, cette alternative est-elle suffisante ?

un constituant original, dans lequel le SN est affecté de façon interne, un peu à l'image des nexus.

4. LES SÉQUENCES GREFFÉES À UN NOM

Les énoncés examinés dans cette partie se rattachent à l'apposition. Il est bien évidemment exclu de s'intéresser en détail à cette notion particulièrement rebelle¹⁷. Comme on l'a déjà signalé, les tests de reconnaissance du sujet sont peu adaptés pour traiter de ce phénomène de syntaxe interne. Les cas les plus facilement repérables sont ceux dans lesquels "l'insertion"¹⁸ est proportionnelle à la série marquée par le pronom *lui* sur le modèle de :

et ce **recueil** — *le Gaspard De La Nuit* (n'a paru, n'apparut)
qu'après sa mort — a été tiré à quelques exemplaires et n'a eu aucun
retentissement (LIC 95)

et ce **recueil** *lui* n'a paru qu'après sa mort

4.1. LA SUITE LEXIQUE + PROFORME

On s'en tiendra à des exemples où le pronom présente une forme "simple" telle que *lui, eux, elle*¹⁹. Le sujet possède généralement une étendue réduite (c'est-à-dire que le syntagme se compose, au maximum, d'un déterminant et d'un nom). L'ordre d'apparition des deux composantes de l'ensemble est contraint :

en contrat C.D.D. les vacataires sont euh des C.D.D. mais sur une
journée hein on y reviendra **les C.D.D.** *eux* ont un contrat plus ou
moins long donc un un contrat euh de -- un mois pour les plus courts

17. Cette dernière partie ne fera qu'effleurer un domaine complexe qui ne pouvait pas être ignoré dans les facteurs d'éloignement du sujet. La littérature sur l'apposition est abondante. On renvoie à Wilmet (1997) et à Neveu (1998) pour une discussion sur son extension. Caddéo (thèse en cours) en dresse un inventaire dans les corpus de français parlé.

18. Le codage des énoncés reste semblable : l'insertion i.e. l'apposition est en italiques, la séquence qui précède est en gras. Les exemples "délicats" où la frontière entre insertion et commentaire demanderait un examen minutieux n'ont pas été recherchés :

le chef d'orchestre **donc mari de Charlotte Rampling** est reparti (POI 97)
l'image des partis quels qu'ils soient **partis de gauche partis de droite** a
certainement besoin d'être réhabilitée en France (POI 95)

19. Un terme comme *lui-même* semble posséder d'autres valeurs et d'autres fonctionnements que ceux décrits ici.

six mois pour euh ceux qui sont présents le le plus longtemps ici sur le parc (POI 96)

Dans le schéma le plus répandu, deux éléments sont présentés conjointement ou traités en parallèle. Le recours à la proforme va ainsi jouer d'une opposition et servir à singulariser l'un des deux²⁰. Dans l'exemple ci-dessous, dans le couple présenté (Paul Abadie père / Paul Abadie fils), le pronom *lui* permet de singulariser le second²¹ :

les amateurs d'art vont finalement hum presque regretter la restauration qui a été exécutée quelques années auparavant par l'architecte **Paul Abadie père** – finalement **Paul Abadie fils** *lui* arrive et veut rendre la cathédrale son unité romane (POI 96)

Le même mécanisme se retrouve dans l'opposition entre une mère et sa fille²² :

elle est [la mère] **elle** est Américaine mais euh **sa fille** *elle* est Iranienne est Iranienne selon la loi euh – la loi euh qui est établie en Iran (POI 96)

C'est, en règle générale, le deuxième terme qui s'adjoint la proforme selon le schéma suivant :

1 ^{er} terme	2 ^e terme + lui
<i>Bertrand</i>	<i>Rimbaud lui</i>

et **Bertrand** est l'inventeur d'une espèce de de prose rythmée très particulière que l'on retrouvera presque pas chez Baudelaire – un peu chez Rimbaud par moment(s) mais très discrètement parce que **Rimbaud** *lui* veut se démarquer complètement de la versification traditionnelle donc il faut pas qu'on ait l'impression de pseudo vers (LIC 95)

L'autre disposition (1^{er} terme + *lui* et 2^e terme) est bien moins représentée dans les corpus :

pourquoi – parce que les fournisseurs eux – **les fournisseurs** *eux* se se gargarisent là on leur ils ils désossent leur viande sans que **le professionnel** la voie (LIC 95)

20. Nolke (1997) parle de pronom focalisateur pour décrire de tels emplois. Ce n'est pas le pronom lui-même qui est focalisé. Il fonctionne comme un index qui pointe sur l'élément focalisé (le sujet dans ce cas).

21. Dans les exemples suivants, les deux éléments qui sont opposés seront présentés en gras.

22. La proforme *elle* dans l'énoncé *sa fille elle est Iranienne* se prête à deux interprétations. Mais, dans cet exemple, l'écoute de l'enregistrement permet d'affirmer, sans hésitation, qu'il n'y a de double marquage sur le sujet.

je crois que le contrat phare de de référence c'est le contrat Afer – hein euh qui est une filiale du groupe Victoire – donc **ce contrat-là Afer lui** arrive réellement souvent en tête **et notre contrat donc Lybra** qui s'appelle en fait Préviretraite donc sur les les bouquins financiers on /les, le/ retrouvera sous ce vocable-là arrive donc dans les cinq premiers contrats –

Cette distinction rejoint un peu les problèmes d'anaphore et de cataphore. Il semble plus habituel, dans les exemples recueillis, de s'opposer à un élément déjà introduit dans le contexte antérieur que de s'opposer, par anticipation, à un SN.

Cette disposition des éléments n'est pas propre aux sujets humains et se retrouve dans les contextes où doit être marquée une singularisation :

et puis – – le style externe – *par contre lui* – – est basé sur les mouvements les techniques – et – – donc X – – di- différent des du **du style interne** (LIC 95)

c'est un calcaire très fin assez blanc et euh qui est surtout destiné à la sculpture **le gros œuvre de la façade lui** était réalisé principalement avec du calcaire Achaie (POI 96)

Depuis le mois de mai, tous les mardis, pendant une heure à une heure et demie, **le rituel** est presque immuable. **L'atmosphère, elle**, est incertaine (*Le Monde*, 25-09-98)

Bien évidemment, des schémas plus étoffés existent (avec combinaisons des formes décrites séparément). Ainsi la proforme peut s'ajouter à une séquence lexicale :

ensuite monsieur euh **Léon Tief** — *un économiste célèbre aussi* — *lui* nous parle d'un certain paradoxe c'est le paradoxe de de Léon Tief (LIC 95)

De même, la proforme peut apparaître simultanément avec un associé décrit en 2.2. (élément de validation) qui sert à renforcer la saillance du terme singularisé :

et c'est là où euh **la femme elle de son côté** euh n'accepte pas la situation (POI 96)

Le terme de validation (*de son côté*) permet de focaliser sur un sujet thématique, la proforme (*elle*) intervient pour singulariser un syntagme qui forme couple avec un autre élément.

4.2. RÉPARTITION DANS LES CORPUS

Ce tour présente une répartition remarquable dans les données que nous avons exploitées. Les volumes POI 96 (Guides) et POI 97 (Résumés de films) comportent un nombre élevé d'appositions, alors qu'elles sont quasiment absentes des enregistrements d'hommes politiques POI 95. Un seul exemple, d'un type différent de celui retenu ici, a été relevé dans ce dernier volume :

votre directeur de campagne *Bruno Mégret* a présenté cette semaine le détail de votre campagne électorale (POI 95)

Une grande régularité apparaît dans l'utilisation de cette tournure²³. Les attestations sont fréquentes dans des passages narratifs. Ce contexte favorable d'apparition peut-être formulé en terme de "genre". Dans les corpus utilisés les locuteurs présentent de nombreux personnages (résumés de films, récits de faits historiques). Cela explique certaines caractéristiques des exemples trouvés. Beaucoup évoquent des personnages fictifs ou réels. Mais le schéma grammatical ne se réduit pas au trait "humain". Plusieurs exemples proviennent du volume LIC 95. Ils ont été puisés dans les passages narratifs des corpus rassemblés.

Il est frappant d'observer que des phénomènes parfois proches tels que l'apposition et certains associés à valeur parenthétique possèdent des distributions aussi différentes. Leur répartition pourrait vraisemblablement servir d'indicateur pour caractériser les genres. Ce type (lexique et pro-forme) se rencontre peu en début de texte. Il sert à opposer deux éléments.

5. CONCLUSION

La réalisation du sujet présente une variété de fonctionnements encore assez mal décrites. Ce travail a cherché à faire ressortir quelques particularités des sujets au travers des insertions qui se réalisent entre le sujet et le verbe. Trois regards complémentaires ont été envisagés. À partir d'un cadre syntaxique général (l'opposition entre rection et associés) a pu être dressé un inventaire des différentes sortes d'insertions et de quelques-unes de leurs propriétés. Le travail sur des contextes larges a permis d'observer les propriétés d'emploi de ces sujets éloignés. Plusieurs fonctionne-

23. Cette observation porte sur l'ensemble des appositions observées et pas seulement sur le type examiné dans ces pages.

ments ont été présentés. Ainsi, certains associés favorisent la prééminence du sujet, servent à le focaliser. Certains compléments réactionnels tendent à "s'intégrer" au sujet et à le modifier de façon interne. Enfin, la répartition des diverses sortes d'insertion a été mise en relation avec les corpus utilisés. On a ainsi pu constater que la distribution des diverses organisations décrites ne se fait pas de façon aléatoire dans les corpus mais correspond bien à des genres différents. Une classe d'associés insérés (les éléments de validation) et les réactions se retrouvent dans tous les corpus. Les appositions et les parenthèses (autre classe d'associés) sont utilisées de façon sélective. Les passages narratifs constituent un contexte favorable à l'apparition des appositions. Inversement, les interviews d'hommes politiques en sont presque entièrement privées. Ce sont plutôt les commentaires (sous forme de construction verbale) qui y abondent. Ces résultats, même s'ils demandent à être affinés et plus amplement vérifiés, illustrent la variété de fonctionnement des sujets.

Paul CAPPEAU
Université de Poitiers



Bibliographie

- APOTHÉLOZ, Denis et REICHLER-BÉGUELIN, Marie-José (1995). "Construction de la référence et stratégies de désignation", *TRANEL*, 23, pp. 227-271.
- (1995). *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*, Genève, Librairie Droz.
- AUTHIER-REVUZ, Jacqueline (1995). *Ces mots qui ne vont pas de soi. Boucles réflexives et non-coïncidences du dire*, Paris, Larousse.
- BLANCHE-BENVENISTE, Claire (1994). "Quelques caractéristiques grammaticales des 'sujets' employés dans le français parlé des conversations", dans YAGUELLO, Marina, (éd.), *Subjecthood and subjectivity. The status of the subject in linguistic theory*, Paris, Ophrys.
- (1995). "Le semblable et le dissemblable en syntaxe", *RSFP*, 13, pp. 7-32.
- BRUNOT, Ferdinand (1905-1953). *Histoire de la langue française des origines à nos jours*, Paris, Colin.

- CAMPROUX, Charles (1958). *Étude Syntaxique Des Parlers Gévaudanais*, Paris, PUF.
- CADDÉO, Sandrine (en cours). "L'apposition : analyse syntaxique de l'apposition détachée dans la langue écrite et divers registres de la langue parlée", Thèse en cours, Université de Provence.
- CHERVEL, André (1998). *La culture scolaire. Une approche historique*, Paris, Belin.
- DANELL, Karl Johan (1998). "La portée comme phénomène linguistique", *Le français moderne*, LXVI-1, pp. 1-26.
- FERNANDEZ, M. M. Jocelyne (1994). *Les particules énonciatives*, Paris, PUF.
- GAPANY, Joël et ZAY, Françoise (1995). "Les relatives parenthétiques : problèmes de reconnaissance et de délimitation", *Scolia*, 5, pp. 31-64.
- GIRAULT-DUVIVIER, Ch-P. (1811). *Grammaire des grammaires*, Paris, A. Cotelle.
- NEVEU, Franck, (1959). *Étude sur l'apposition. Aspects du détachement nominal et adjectival en français contemporain, dans un corpus de textes de J.-P. Sartre*. Paris, Honoré Champion.
- NØLKE, Henning, (1993). *Le regard du locuteur. Pour une linguistique des traces énonciatives*, Paris, Kimé.
- (1997). "Anaphoricité et focalisation : Le cas du pronom personnel disjoint" dans De MULDER, Walter, TASMOVSKI-De RICK, Liliane, VETTERS, Carl (éds), *Relations anaphoriques et (in)cohérence*, Amsterdam, Editions Rodopi B.V., pp. 55-67.
- WILMET, Marc (1997). "L'apposition : une fonction à réestimer", dans KLEIBER, Georges et RIEGEL, Martin (éds), *Les formes du sens. Études de linguistique française, médiévale et générale offertes à Robert Martin à l'occasion de ses 60 ans*, Louvain-la-Neuve, Duculot.

Prosodie des langues romanes : analyse phonétique et phonologie

LE PARLÉ ET LE LU

Il n'y a pas si longtemps, les techniques courantes d'analyse du signal de parole n'autorisaient que l'étude de productions de parole enregistrées avec beaucoup de soins, en studio, par des locuteurs bien entraînés. Les corpus recueillis étaient le plus souvent dépourvus de toute spontanéité, et les descriptions de la prosodie qui en résultaient ne pouvaient que refléter implicitement ces conditions de travail imposés au phonéticien et au phonologue. Plus d'un commentateur critiquait d'ailleurs volontiers l'aspect artificiel des énoncés soumis à la description prosodique, au point de remettre en question leur validité pour la parole "réelle", c'est-à-dire spontanée.

Depuis, les techniques de traitement du signal ont évolué. Les algorithmes d'analyse sont devenus plus résistants aux bruits, plus rapides, et surtout plus faciles d'emploi. Ils peuvent être utilisés sur un ordinateur personnel, avec des enregistrements réalisés sur cassette ordinaire. La recherche devient ainsi indépendante des multiples contraintes, techniques et administratives, imposées parfois par les laboratoires de phonétique naguère incontournables pour ce genre d'études. Les corpus peuvent être spontanés et enregistrés dans des conditions peu contraignantes. Les observations et les modèles théoriques qui en résultent auront d'emblée une portée plus large, incluant la parole spontanée.

Cette évolution a déclenché une nouvelle vague de recherche sur l'intonation, qui, sans doute enthousiasmée par la liberté offerte par les nouvelles technologies, semble la plupart du temps ignorer les travaux

existants, au prétexte qu'ils relèvent de corpus lus et ne sauraient présenter la moindre validité. Tout se passe comme si on admettait que les locuteurs, dans une activité de discours spontané, ne se servaient plus des systèmes phonologique, morphologique, syntaxique de la langue qu'ils utilisent dans d'autres conditions de discours. En vérité, comme le souligne C. Blanche-Benveniste (1997), en parole spontanée les locuteurs se servent différemment des sous-systèmes de la langue, et ceci est vrai aussi pour les phénomènes prosodiques. Du reste on peut montrer que la plupart des faits prosodiques observés dans la parole spontanée trouvent une explication dans la mise en œuvre de la grammaire prosodique, qui se manifeste dans la lecture oralisée. C'est dans cette optique que cet exposé se propose de rappeler les principes généraux de la grammaire prosodique, en les appliquant à différents aspects spécifiques à quatre langues romanes, italien, espagnol, portugais et français.

OUTILS D'ANALYSE

Traditionnellement, l'analyse prosodique d'un corpus oral implique la transcription, la segmentation et l'observation des phénomènes prosodiques liés à chaque segment. Ces opérations effectuées à partir d'un enregistrement magnétique sont longues et fastidieuses. Ainsi par exemple pour la transcription, le déroulement de la bande d'enregistrement se faisant en temps réel, l'opérateur doit constamment revenir en arrière dans l'enregistrement pour atteindre une saisie mot à mot, certains mots difficiles à décoder devant être répétés plusieurs fois.

On présente ici quelques caractéristiques d'un logiciel développé et optimisé pour l'analyse prosodique de grands corpus. Ce logiciel d'analyse, appelé WinPitch [WinPitch, 1996], permet d'effectuer toutes ces opérations beaucoup plus facilement qu'avec les méthodes traditionnelles.

– RÉÉCOUTE

La première étape du travail d'analyse consiste en la conversion de l'enregistrement magnétique en format numérique. Les enregistrements magnétiques du corpus sont transformés en une série de fichiers informatiques. Le chargement d'un fichier par le programme réalise immédiatement l'analyse acoustique du matériel sonore, qui est représenté par une série de courbes présentant l'évolution de la fréquence fondamentale (mélodie), de l'intensité, de l'onde sonore et du spectre (spectrogramme).

L'opérateur travaille alors avec une durée de fenêtre réglable (1 à 10 secondes), qu'il peut positionner à volonté en un point quelconque du fichier son, de durée maximum de 12 minutes.

Une première originalité du système consiste en la possibilité de réécouter un segment de son particulier par une commande beaucoup plus simple que celle d'autres logiciels de traitement sonore. L'opération revient à dessiner un bloc graphique sur l'écran avec la souris, bloc qui définit la durée du segment que l'on veut réécouter. Chaque fois que le curseur est placé à l'intérieur du bloc, le segment sonore est réécouté, la rapidité d'exécution résultant de ce que le même bloc ne doit pas être redessiné pour entendre un autre segment, mais simplement glissé vers une nouvelle position sur l'écran.

– RALENTISSEMENT

La transcription d'un corpus spontané présente parfois des difficultés de saisie de certains segments, qui pourraient être résolues plus facilement si l'enregistrement était ralenti. C'est précisément ce que permet le logiciel, en réalisant à la demande un ralentissement qui peut aller jusqu'à cinq fois le temps réel. Ici également la commande est très simple. Une fois le degré de ralentissement choisi (il est également possible d'accélérer), un simple clic de souris à l'intérieur du bloc définissant le segment sonore à saisir réalise la reproduction du son à vitesse réduite. Des ralentissements de deux à trois fois le temps réel s'avèrent ainsi judicieux, le passage de l'écoute en temps réel à l'écoute ralentie ne dépendant que de la position enfoncée ou non de la touche "Contrôle" du clavier de l'ordinateur.

– SEGMENTATION

L'intérêt principal d'un logiciel comme WinPitch ne réside pas seulement dans la transcription des données, mais surtout dans la visualisation qui est faite de phénomènes qu'on n'est pas toujours entraîné à percevoir et à transcrire, comme la mélodie, l'intensité, la présence de pauses, etc. Le programme rend visible ce qui est difficile de percevoir : le mouvement mélodique sur telle ou telle syllabe accentuée est-il montant ou descendant ? La pause perçue à tel endroit existe-t-elle réellement ou est-elle seulement virtuelle ? Et si tel mouvement mélodique montant est bien visible sur l'écran, sur quel segment phonologique porte-t-il ?

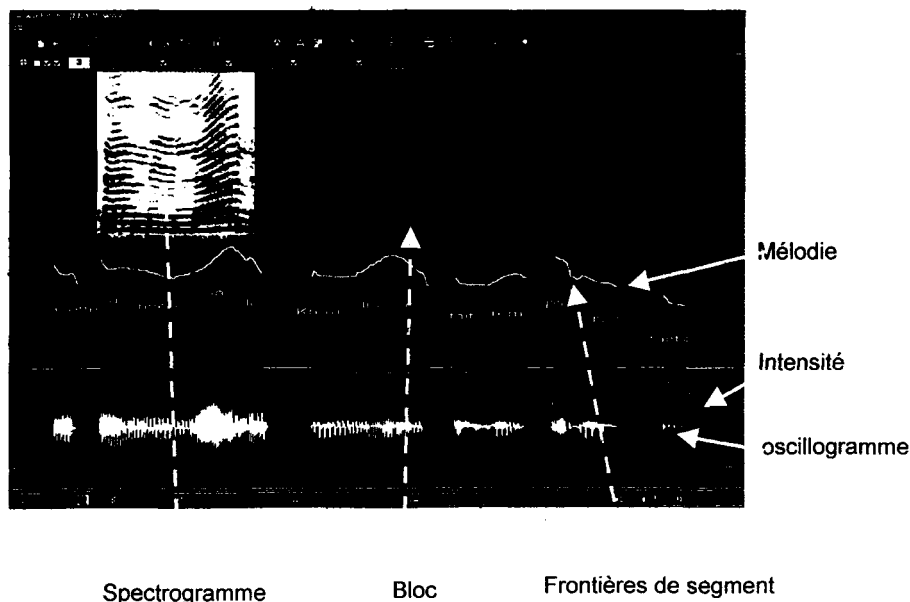
Pour répondre à toutes ces questions, il faut opérer la segmentation du corpus en unités plus petites, qui peuvent être des groupes rythmiques, des syllabes, ou des réalisations de phonèmes. Cette tâche est rendue très facile avec WinPitch, qui met à la disposition de l'utilisateur plusieurs outils. Outre le bloc permettant une écoute normale ou ralentie d'un segment pour en établir les frontières, une commande simple permet d'obtenir l'affichage d'un spectrogramme à bande large ou étroite. Une fois les caractéristiques principales des spectrogrammes des voyelles, des consonnes sourdes et voisées, des fricatives et des occlusives connues de l'opérateur, la segmentation devient très aisée et sûre. Elle se note par un clic de souris à l'endroit de la frontière détectée, qui fait apparaître une frontière graphique verticale. Une commande supplémentaire permet la vérification de l'opération, une touche clavier réalisant l'écoute du segment sonore limité par deux frontières verticales et déterminé par la présence du curseur entre ces frontières. En cas d'erreur, la frontière peut être tout aussi aisément déplacée ou effacée.

– ÉTIQUETAGE

Après chaque opération de segmentation, l'utilisateur peut ajouter sur l'écran la transcription orthographique et/ou phonétique. Les séquences de texte (syllabique, lexicaux ou autres) peuvent de plus être positionnées à volonté le long de la courbe mélodique, par exemple, ou le long d'une des autres courbes disponibles. Lorsque le fichier est rejoué, un curseur indique la position temporelle des sons au fur et à mesure de leur écoute, et les textes éventuellement ajoutés s'allument au gré du déroulement, à la façon d'un texte de karaöke.

– COMMENTAIRE

En plus du texte, il est également possible d'ajouter un nombre quelconque de commentaires, liés à des marqueurs graphiques positionnés sur l'écran par l'opérateur. Une fois la segmentation d'un fichier corpus achevée, les résultats peuvent être retranscrits sous forme de texte, les courbes mélodiques et d'intensité apparaissant sur la forme de colonnes de chiffres en Hz et en dB, chaque ligne correspondant à un instant déterminé, espacé du précédent d'une valeur temporelle programmable (par exemple 25 millièmes de seconde). Ce texte comprend également les parties de transcription orthographique et/ou phonétique, et l'ensemble peut être facilement importé dans d'autres programmes de traitement, comme le tableur Excel.



– PHONOLOGIE DE L'INTONATION

Les bases de la phonologie de l'intonation exposées succinctement ici s'appliquent à quatre langues romanes, français, italien, espagnol et portugais. Dans cette description, l'intonation, manifestée acoustiquement par des variations de hauteur mélodique, d'intensité et de durée syllabique, assume trois fonctions principales dans les système linguistique : 1) l'indication de la modalité déclarative ou interrogative de l'énoncé et leurs variantes, 2) la division éventuelle en thème et propos, et 3) l'indication d'une structure prosodique organisant de manière indépendante les unités accentuelles.

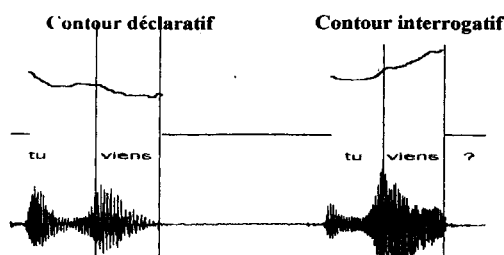
– MODALITÉ

La fonction d'indication de la modalité de l'énoncé est bien connue. Si on accepte l'existence de deux modalités de base, la déclaration et l'inter-

rogation (l'ordre apparaissant alors comme une variante de la déclaration), le système est simple : un contour mélodique descendant, corrélé à une modalité d'énoncé déclarative, s'oppose à un contour montant, corrélé à une modalité interrogative. Phonétiquement, le contour déclaratif descendant s'appuie sur un noyau syllabique long, sur la partie voisée de la syllabe accentuée finale de la phrase, et présente une faible chute mélodique, alors que le contour interrogatif a une plus grande variation mélodique, mais sur la partie voisée de la syllabe finale plus courte que dans le cas déclaratif.

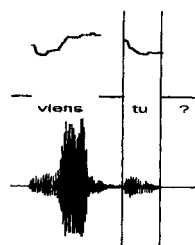
Dans les deux cas, on retiendra que phonologiquement, ces deux contours de modalité s'opposent à d'autres contours à l'intérieur de la phrase par leur longueur (c'est-à-dire la combinaison de leur variation mélodique et de leur durée).

En français, la dernière syllabe accentuée se confond avec la dernière syllabe prononcée, et les contours de modalité se trouveront donc sur la même syllabe. Nous verrons plus loin que dans les autres langues romanes, ce ne sera plus le cas si la dernière syllabe prononcée n'est pas accentuée. Le contour déclaratif est alors placé sur la syllabe accentuée, et le contour interrogatif sur la dernière syllabe, qu'elle soit accentuée ou non.



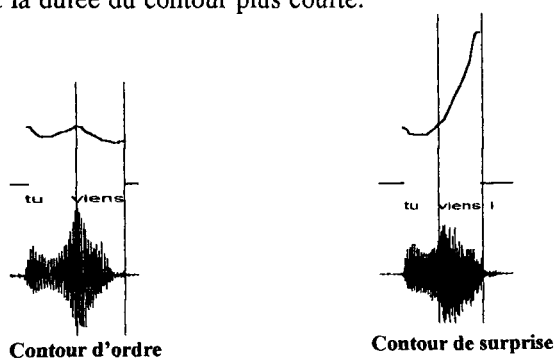
(Les contours correspondent aux variations mélodiques entre les deux barres verticales)

Comme toute autre entité phonologique, le contour de modalité est susceptible d'être neutralisé lorsque d'autres marques existent dans la phrase ou dans le contexte. C'est le cas dans l'exemple ci-contre, où la modalité interrogative est indiquée par l'inversion sujet verbe.

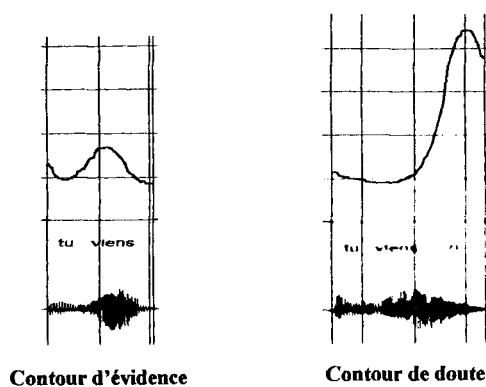


– VARIANTES DES CONTOURS DE MODALITÉ

Les contours de base déclaratif et interrogatif peuvent présenter des variations corrélatives de l'ordre pour la déclaration, et de la surprise pour l'interrogation. Dans les deux cas, le mouvement mélodique est amplifié, et la durée du contour plus courte.



À ces deux variantes d'ordre et de surprise, s'ajoutent l'évidence, variante de la déclaration, et le doute, variante de l'interrogation, manifestées toutes les deux par un mouvement mélodique caractéristique montant descendant imbriqué dans le mouvement général du contour.



La description phonologique des variantes de la modalité utilise les traits +/- Montant, +/- Ample et +/- Circonflexe selon le tableau:

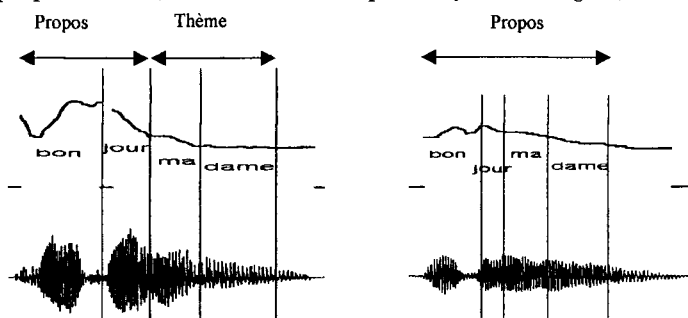
	Déclaratif	Interrog.	Ordre	Surprise	Évidence	Doute
Montant	-	+	-	+	-	+
Ample	-	-	+	+	+/-	+/-
Circonflexe	-	-	-	-	+	+

D'autres analyses récentes font état d'autres classes possibles, obtenues en catégorisant différentes caractéristiques de l'acte énonciatif (Cresti, Martin, Moneglia 1998).

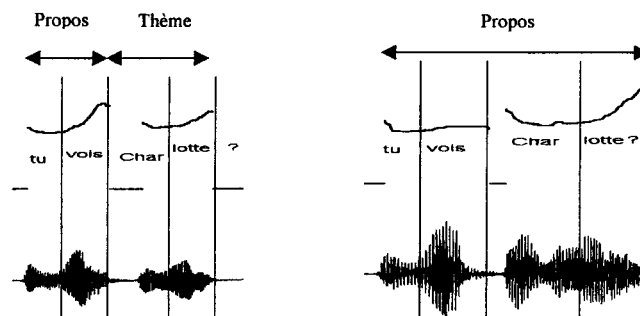
- THÈME-PROPOS

La division prosodique de la phrase en thème et propos a été souvent décrite et est bien documentée. Dans une des formes que prend cette division, le thème peut être défini comme la deuxième partie de la phrase qui suit le propos, et qui reprend des informations déjà connues dans le processus d'énonciation. Ainsi, à la question *que fait Jules à Aix ?*, la réponse *il enseigne à Aix* avec un contour mélodique descendant sur la dernière syllabe accentuée du propos *enseigne* et un contour plat sur la dernière syllabe accentuée du thème *à Aix* réalise une division de la phrase qui implique qu'en effet, la partie thème contient des informations déjà connues, et simplement rappelées dans la phrase.

En discours spontané, la partie thème renvoie souvent à un élément du contexte présent dans la situation de communication, mais pas nécessairement énoncée explicitement. Un énoncé aussi courant que *bonjour madame* est souvent accompagné d'un contour mélodique divisant la phrase en propos (*bonjour*) et thème (*madame*), le référent du thème étant déjà connue par la situation ou le contexte. Si la situation telle qu'appréciée par le locuteur ne contient pas cette information, les contours mélodiques seront cette fois montant et descendant, marquant l'absence d'une division propos thème (cf. deuxième exemple, *bonjour les dégâts*).



Les variétés interrogatives présentent les mêmes caractéristiques, mais cette fois le deuxième contour finissant le thème sera montant.



Dans le premier tracé, le pronom *tu* est coréférent de *Charlotte* (équivalent de *Charlotte, est-ce que tu vois [cette maison là-bas] ?*), dans le second *Charlotte* est objet de *tu vois* (équivalent de *est-ce que c'est Charlotte que tu vois ?*).

– STRUCTURE PROSODIQUE

Le concept de structure prosodique repose sur l'hypothèse suivante : il existe, indépendamment de toute autre structure impliquant des unités du discours et de la phrase, une structure qui organise hiérarchiquement des unités minimales prosodiques (des mots prosodiques), liées aux syllabes accentuées de la phrase. Partant, cette structure est indiquée par un mécanisme impliquant des marqueurs existant dans les mots prosodiques.

Les parties perceptivement les plus importantes dans la phrase étant les syllabes accentuées, et la phrase la plus simple possible étant composée d'un mot unique d'une seule syllabe, nécessairement accentuée, l'unité minimale prosodique ou mot prosodique apparaît comme une unité de nature prosodique comportant un et un seul segment correspondant à une syllabe accentuée. Il y a donc autant de mots prosodiques dans une phrase que de syllabes accentuées, d'autres manifestations prosodiques comme les accents d'insistance et d'emphase étant donc surajoutées au mot prosodique.

Si une telle structure organise les mots prosodiques, il existe nécessairement un mécanisme, de nature prosodique, qui en assure l'indication sous la forme de marqueurs à déterminer. Une fois de plus, un mot prosodique pouvant correspondre à une seule syllabe, ce marqueur peut se retrouver dans le segment prosodique relatif à cette syllabe accentuée, dans les manifestations prosodiques de l'accent, mais éventuellement aussi

dans des positions nécessairement repérables même si le mot est réduit à une seule syllabe, c'est-à-dire le début et la fin du mot prosodique.

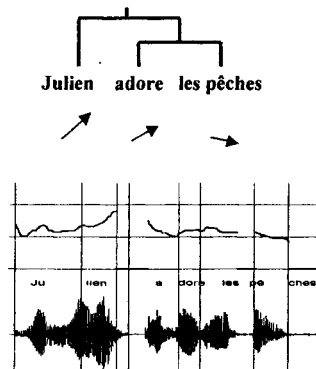
En observant les manifestations de ces marqueurs à l'intérieur d'une structure prosodique connue, on doit pouvoir en déterminer le fonctionnement, et donc la grammaire prosodique de la phrase.

Reste alors à déterminer la structure prosodique d'une phrase donnée dans une énonciation donnée. Les techniques habituelles cherchent à prendre en quelque sorte la structure syntaxique en défaut pour forcer la structure prosodique à se manifester. Elles utilisent des phrases syntaxiquement ambiguës (type *un joueur de football américain*), des phrases dépourvues de marqueurs syntaxiques (énumérations avec hiérarchie, comme les numéros de téléphone). La lecture oralisée est également de nature à révéler l'organisation prosodique de la phrase, le style de la lecture consistant précisément à réaliser une adéquation entre syntaxe et intonation. En discours spontané, on n'a évidemment pas accès à la structure prosodique de manière aussi sûre.

Le résultat de cette recherche portant sur les énoncés lus permet de découvrir, pour le français, les mécanismes suivants, partant des configurations simples aux structures plus complexes, mais en se limitant ici au seul propos, donc aux phrases dépourvues de thème :

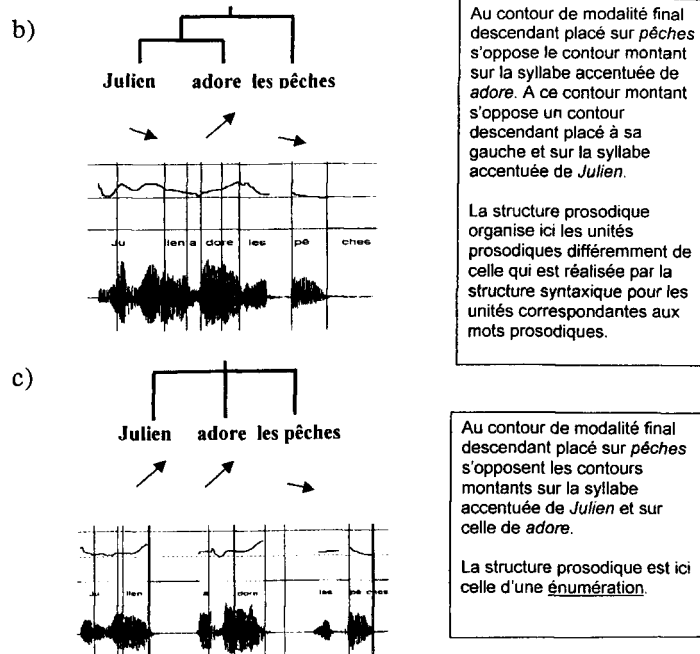
1. Une phrase à un seul mot prosodique n'a pas de structure prosodique, et le contour placé sur la seule syllabe accentuée n'indique donc que la modalité ;
2. Une phrase à deux mots prosodiques ne présente qu'une seule structure possible : la première syllabe accentuée du mot qui ne porte pas le contour de modalité ne peut que se différencier de celui-ci ;
3. Avec trois mots prosodiques, trois configurations sont possibles :

a)



Au contour de modalité final descendant placé sur *pêches* s'opposent, en deux niveaux différents, le contour montant de la syllabe accentuée de *Julien*, et le contour montant, mais de moindre amplitude de variation, placé sur la syllabe accentuée de *adore*.

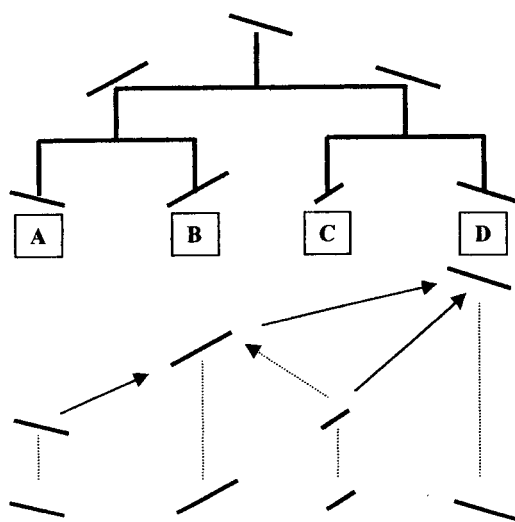
La structure prosodique est ici **congruente** à la structure syntaxique.



Ce mécanisme de base se généralise à un nombre plus grand de mots prosodiques organisés en plusieurs niveaux, en utilisant les règles de contour suivantes :

- contraste de pente mélodique envers le contour de droite terminant un groupe prosodique donné. Ainsi, la syllabe accentuée d'un groupe comprenant deux mots prosodiques portera un contour montant si le groupe, et donc le deuxième mot prosodique, présente un contour descendant, et un contour descendant si le groupe se termine par un contour montant ;
- contraste d'amplitude de variation mélodique et/ou de durée avec les contours de même pente situés à gauche, assurant l'indication de l'appartenance des groupes prosodiques à des niveaux différents dans la structure.

En examinant les mécanismes d'expansion de la structure prosodique, on conclut facilement que la grammaire de la structure prosodique indique une dépendance à droite : un contour donné dépend du contour qui le domine situé à sa droite. Schématiquement, ce mécanisme, peut se représenter de la manière suivante :



Les flèches en trait plein indiquent un contraste de pente mélodique marquant la dépendance à droite du contour, celle en pointillé un contraste d'amplitude de variation mélodique. On voit que par expansion le contour final d'un groupe prosodique se transmet toujours à droite : c'est le mot prosodique terminal du groupe qui hérite du contour dominant à chaque niveau de la division d'un groupe.

– LE CONTOUR EN PÉNULTIÈME POSITION

Tout comme les autres unités phonologiques, les contours prosodiques sont susceptibles d'être neutralisés dans leur fonctionnement, et, partant, leur réalisation modifiée. C'est le cas en particulier de l'avant dernier contour d'une phrase déclarative, qui, dans sa fonction d'indication de la structure prosodique, est toujours montant puisqu'il s'oppose à un contour à droite nécessairement descendant. C'est effectivement ce qu'on observe, aussi bien dans les corpus lus que les productions spontanées.

– PARENTHÈSE

Les parenthèses présentent des propriétés prosodiques semblables à celles de ses propriétés syntaxiques (Blanche-Benveniste, 1997). Elles peuvent par exemple apparaître seules prosodiquement, c'est-à-dire se terminer par un contour de modalité descendant bien que placées à l'intérieur de la phrase, donc à l'intérieur de la séquence de contours mélodique indiquant la structure mélodique sans la modifier. D'autres variantes, décrites également comme parenthèses, font par contre intégralement partie de la structure prosodique, de la même manière qu'un syntagme prépositionnel par exemple.

RELATION SYNTAXE INTONATION

Dans le cas général, hors des structures syntaxiques ambiguës ou absentes, le rapport pouvant exister entre les structures prosodique et syntaxique d'une même phrase, peut faire l'objet de deux hypothèses contradictoires :

1. Les structures prosodique et syntaxique sont congruentes, c'est-à-dire quelles organisent les unités correspondantes selon la même hiérarchie. Les unités prosodiques et syntaxiques correspondantes sont donc organisées de la même façon ;
2. Les structures prosodique et syntaxique sont indépendantes l'une de l'autre : on peut associer n'importe quelle structure prosodique d'une séquence de mots prosodiques donnée quelle que soit la structure syntaxique de la phrase.

L'observation des données montre que la nature du rapport entre les structures syntaxique et prosodique doit se situer entre ces deux extrêmes (Martin, 1987). En fait, on peut montrer que, à une structure syntaxique donnée, le locuteur associe une structure prosodique choisie dans le stock de structures organisant un nombre donné de mots prosodiques. Ces structures possibles s'obtiennent par une combinatoire respectant les contraintes suivantes :

1. Le nombre de syllabes d'un mot prosodique ne peut excéder 6 (7 ? La valeur de ce paramètre dépendant de la vitesse d'élocution) ;
2. Les syllabes accentuées évitent les collisions d'accent (deux syllabes accentuées sur deux syllabes successives) ;

3. Les structures prosodiques admissibles respectent la condition de non collision syntaxique (Martin, 1987).

La première de ces conditions rend compte des différents schémas d'accentuation d'une phrase en français, dont la variabilité est due au caractère non obligatoire de la réalisation de l'accent lexical.

La deuxième rend compte des déplacements d'accent sur une syllabe plus à gauche en cas de collision d'accent.

La dernière condition est moins connue, et porte sur l'impossibilité de regrouper par la prosodie deux unités syntaxiques appartenant à des groupes différents, sauf s'il s'agit de la dernière unité de la phrase.

4. De plus, une condition d'eurythmie intervient en ce que

- a) le locuteur choisit la structure la plus eurythmique possible entre plusieurs structures qui satisfont les conditions ci-dessous ;
- b) sinon, un effet d'équilibrage rythmique tendra à la réalisation plus lente de groupes prosodiques à petit nombre de syllabes, par rapport à une réalisation plus rapide de groupes ayant un plus grand nombre de syllabes situés à des niveaux correspondant dans la structure prosodique.

Dans *Antonella a bu du vin* la solution congruente est aussi la plus eurythmique, puisque le nombre de syllabes du premier groupe prosodique (*Antonella* 4) est égal à celui du deuxième (*a bu du vin*, 4). Une solution non congruente à la syntaxe, [*Antonella a bu*][*du vin*] partage la phrase en deux groupes de 6 et 2 syllabes respectivement, ce qui se traduit par une accélération rythmique sur le premier groupe prosodique et un ralentissement sur le second.

ET LE FRANC, C'EST PARLÉ ?

Cette grammaire prosodique, succinctement exposée, rend compte d'un grand nombre d'observation faite sur la prosodie du français parlé. Ainsi, beaucoup d'exemples cités dans Blanche-Benveniste, C. (1997) utilisent la marque de division propos thème :

Vous le saurez bien assez tôt il nous disait toujours

/ —

À la caisse ils se présentent

/ —

D'autres, avec une organisation syntaxique disloquée, illustrent le mécanisme de contraste de pente :

Et les femmes ici elles sont comme ça



Le gars là il a rien il y comprends rien



Au centre ville d'accord mais ailleurs rien



SYSTEMES PROSODIQUES DES LANGUES ROMANES

– LA POSITION DU CONTOUR DE MODALITÉ LE CONTOUR DE MODALITÉ DANS LA DIVISION PROPOS THÈME

Le contour de modalité est positionné, en français, sur la dernière syllabe accentuée, qui se trouve être aussi la dernière syllabe prononcée de la phrase. Il en va de même si la phrase est divisée en thème et propos, le contour de modalité est alors placé sur la dernière syllabe accentuée du propos, qui se confond avec la dernière syllabe.

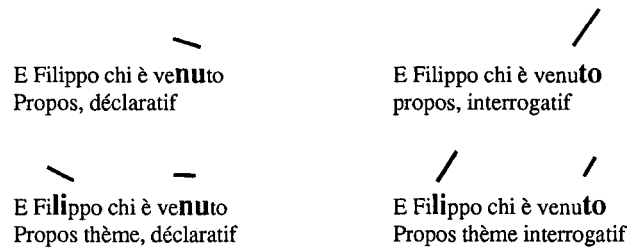
Dans les autres langues romanes, la dernière syllabe accentuée n'est pas nécessairement en position finale. La position de contour de modalité est alors distribuée selon les configurations suivantes :

1. dans la phrase sans division propos thème, le contour de modalité déclaratif est positionné sur la syllabe accentuée finale, et le contour interrogatif sur la dernière syllabe (que celle-ci soit accentuée ou non) ;

2. dans la phrase divisée en propos thème, les contours de modalité déclaratif et interrogatif sont tout deux placés sur la syllabe accentuée finale du propos, et également, mais sous forme neutralisée (donc phonétiquement plate ou faiblement montante) sur la syllabe accentuée finale du thème dans le cas déclaratif, et sur la syllabe finale, accentuée ou non, dans le cas interrogatif.

Il y donc une différence de configuration dans les cas interrogatif et déclaratif, différence non apparente en français du fait de la position

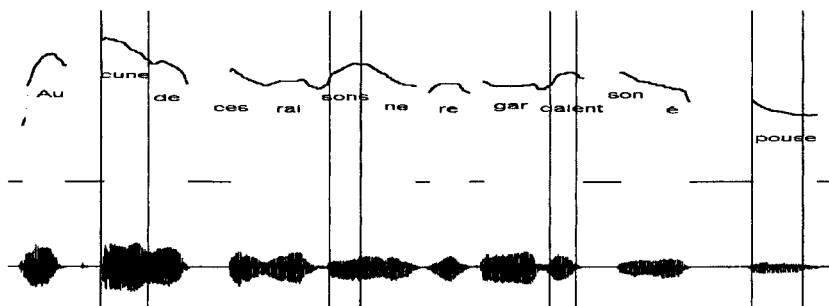
nécessairement finale de la dernière syllabe accentuée. Des exemples qui illustrent ces configurations apparaissent ci-dessous :



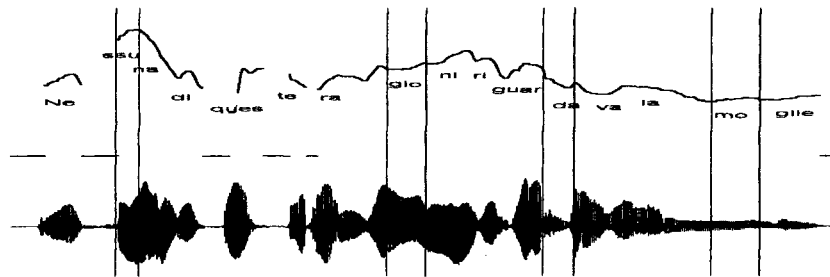
– L'INDICATION DE LA STRUCTURE PROSODIQUE

Si les caractéristiques théoriques qui régissent les rapports entre les structures syntaxique et prosodique sont valables pour toutes les langues romanes considérées, ainsi que les contraintes propres aux structures prosodiques (collision d'accent, collision syntaxique, eurythmie et longueur maximale d'un mot prosodique), chaque système aura son mécanisme propre, mettant en jeu des marques prosodiques aux réalisations phonétiques spécifiques (tout comme le système phonologique vocalique commun à l'italien et à l'espagnol se voit réalisé différemment sur le plan phonétique).

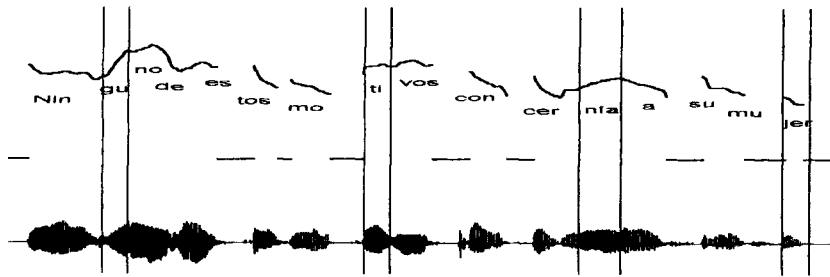
Considérons une structure prosodique de type $[[A B] [C D]]$ raisonnablement susceptible d'être réalisée en lecture soignée comme congruente à une structure syntaxique similaire $((A B) (C D))$. Les phrases suivantes sont extraites du corpus Eurom4 (Blanche-Benveniste, 1997).



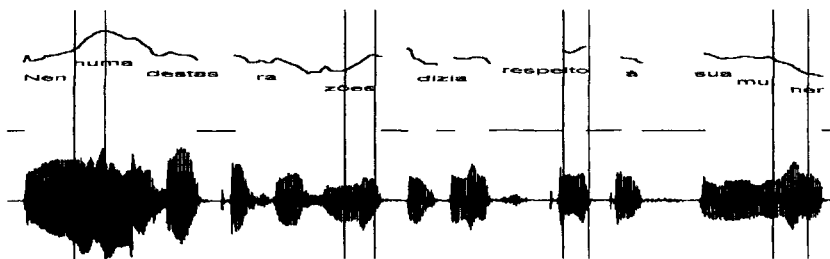
Français : aucune de ces raisons ne regardaient son épouse



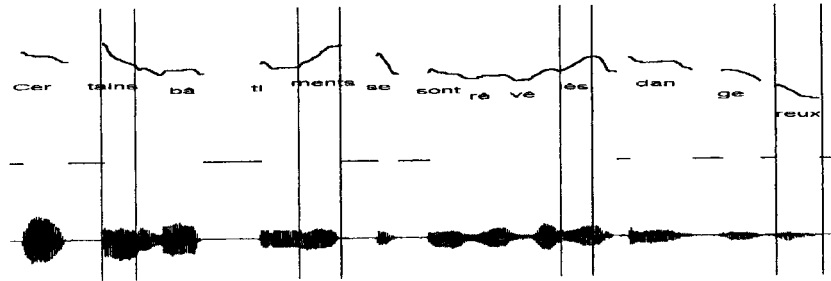
Italien : *Nessuna di queste ragioni riguardava la moglie*



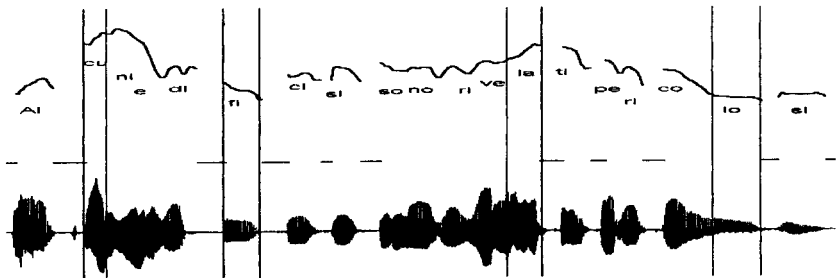
Espagnol : *Ninguno de estos motivos concernía a su mujer*



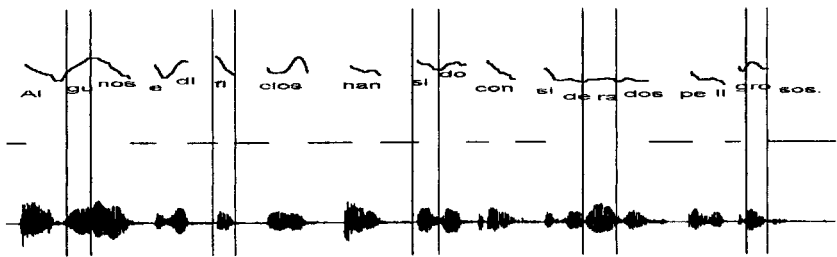
Portugais : *Nenhuma destas razões dizia respeito à sua mulher*



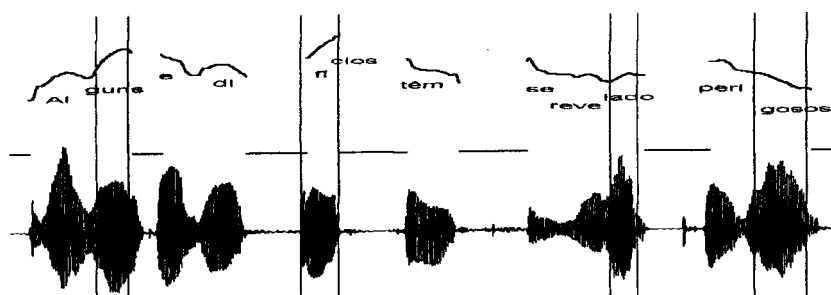
Français : *Certains bâtiments se sont révélés dangereux*



Italien : *Alcuni edifici si sono rivelati pericolosi*



Espagnol : *Algunos edificios han sido considerados peligrosos*

Portugais : *Alguns edifícios têm se revelado perigosos*

Pour découvrir les contours principaux, et le mécanisme qui les génère, examinons les corrélations de la structure prosodique avec des structures de type (SN SV) et (SP SN SV), en admettant par hypothèse que la congruence prosodie syntaxe est effectivement réalisée dans les données de lecture.

On trouve alors les résultats suivants :

1. Le contour de premier niveau, qui divise la phrase en 2 groupes prosodiques correspondant aux frontières SN SV ou SP SN SV, est manifesté :
 - a) en français : par une contour montant situé sur la dernière syllabe accentuée de SN. Ce contour est neutralisé est pourra être réalisé par une faible variation mélodique montante ou descendante si SN ne comprend qu'un seul mot prosodique (un seul accent) ;
 - b) en italien : par 2 variantes de contour en distribution complémentaire
 - montant sur la syllabes accentuée finale de SN
 - plat ou légèrement descendant sur la dernières syllabe accentuée de SN et montant sur la dernière syllabe, si celle-ci est distincte de la syllabe accentuée
 - c) en espagnol, en portugais : contours similaires à ceux de l'italien.

Certaines réalisations de l'italien montrent une variation montante descendante sur la même syllabe accentuée en position finale. Les réalisations portugaises sont semblable, le contour montant semble parfois placé sur l'avant dernière syllabe phonologique (comme dans *anos*), qui se trouve être en fait la syllabe accentuée en position finale, du fait de la dévocalisation de la dernière syllabe.

À l'inverse du français, la première syllabe accentuée de l'italien de l'espagnol et du portugais portent toujours un contour montant, dans les phrases de type SP SN SV ou SN SV (en français ce contour dépend de la structure du groupe prosodique qui correspond à SN).

Cette grammaire prosodique est généralisable aux autres langues romanes examinées, à l'exception du contour placé sur la première syllabe accentuée, qui est toujours montant.

CONCLUSIONS

Dans des phrases lues, pour lesquelles l'hypothèse d'une congruence entre les structures syntaxique et prosodique semble valide, la distribution et la réalisation des marques prosodiques indiquant les mêmes structures prosodiques est étonnamment semblable en italien, portugais et espagnol (cette conclusion s'étend du reste au roumain, cf. Martin, 1998). C'est donc, comme dans d'autres domaines comme la syntaxe et la morphologie, le français qui fait bande à part dans l'intonation de ces langues romanes.

Une des raisons tient à la position en syllabe finale de la syllabe accentuée, une autre à l'opposition de pente mélodique envers le contour dominant à droite dans la structure prosodique. Dans le groupe des systèmes prosodiquement semblables du point de vue de la description phonologique des contours indiquant la structure syntaxique, on doit évidemment trouver des différences dans les réalisations phonétiques de ces contours, et dans le rythme des mots prosodiques, défini par les alternances de syllabes inaccentuées courtes et de syllabes accentuées longues.

Des expérimentations complémentaires devraient lever les dernières incertitudes sur le mécanisme de l'indication de la structure prosodique, notamment en ce qui concerne ces problèmes de rythme.

Philippe MARTIN

Université de Toronto



Références bibliographiques

- BLANCHE-BENVENISTE, C. et VALLI, A., éd. "L'intercompréhension : le cas des langues romanes", *Le français dans le monde*, Hachette, Paris, janvier 1997.
- BLANCHE-BENVENISTE, C. [1997], *Approches de la langue parlée en français*, Ophrys, Paris, 164 p.
- [1997], "EuRom4. Méthode d'enseignement simultané des langues romanes", CD-ROM, La Nuova Italia, Firenze, 1998.
- CRESTI, E., MARTIN, Ph., MONEGLIA, M. [1998], "Intonazione delle modalità naturali rappresentative : analisi e sintesi", *Le IX Giornate di Studio del Gruppo di Fonetica Sperimentale*, Venezia, Dec. 1998.
- MARTIN, Ph. [1978], "Questions de phonosyntaxe et de phonosémantique en français", *Linguisticae Investigationes*, II, 93-126.
- [1987], "Prosodic and Rhythmic Structures in French", *Linguistics*, 25, 925-949.
- [1997], "Sentence Intonation in 4 Romance Languages", in *Intonation : Theory, Models and Applications*, Botinis et al., ed., ESCA, Athens 1997, pp. 227-230.
- [1998], "L'intonation du roumain, musique romane ?", *Journées de Phonétique 1998*, Université Western Ontario, mai 1998.
- WinPitch (1996) <http://www.winpitch.com>

Coordination : analyses syntaxiques et annotations

Comme le signale B. Habert et alii (1997 : 71), “l'utilisation de par-seurs pour la constitution de corpus arborés (annotés syntaxiquement) suppose encore une coopération étroite entre linguistes et informaticiens”. D'autre part, on sait que, au moment de délimiter des “unités” et de statuer sur leurs relations, les points de vue entre linguistes sont rarement concordants, et en ce qui concerne la coordination, il n'y a, semble-t-il, même pas de consensus de représentation.

En me situant dans une problématique en relation avec le “traitement automatique du langage” dont je ne suis pas spécialiste, je voudrais revenir sur le fonctionnement de certaines tournures coordonnées afin de pouvoir mener quelques réflexions et faire des propositions dans ce domaine. Ainsi, après le rappel des analyses proposées pour rendre compte des phénomènes de coordination, seront abordés deux points qui semblent essentiels pour ce type d'approche : les cas d'ambiguïtés, plus ou moins prévisibles, et la nécessité de poser différents niveaux d'analyse.

1. BREF ÉTAT DE LA QUESTION

Si le terme de “coordination” s'applique aussi bien aux constituants de la phrase qu'aux différents phénomènes de jonction interphrastique, il est à noter que seule la coordination touchant aux constituants semble avoir reçu un traitement privilégié de la part des grammaires, et ce sera encore le cas de cet article.

1.1. L'analyse des coordinations fait généralement appel à des opérations renvoyant à l'effacement, à la réduction, à l'ellipse. Ce type de trai-

tement, qui consiste à ramener le maximum de coordinations de constituants à des coordinations de phrases avec effacement¹, renvoie à une longue tradition et ne semble pas être remis en question. Cf., par exemple, C. Fuchs (1996: 176) qui parle de l'ellipse du sujet "Pierre" dans :

1. Pierre est venu et reparti aussitôt.

et de l'ellipse du verbe dans :

2. Pierre a applaudi et Marie aussi.

Cf., également, G. Girard (1994: 43) qui souligne que : "*parler de coordination c'est parler d'ellipse*".

Ces analyses ont cependant donné lieu à diverses critiques dont la plus importante tient au caractère non contrôlé de ces mêmes effacements (cf. d'ailleurs le traitement proposé pour l'exemple 1 qui signale l'ellipse du sujet mais ne dit rien de l'auxiliaire "être").

1.2. À côté de ces positions "logicistes", il en existe d'autres qui semblent accepter un plus grand nombre de coordinations de constituants ou de sous-constituants, sans ellipse.

C'est le cas, par exemple, de A. Abeillé et D. Godard (1996) qui, dans le cadre des grammaires formelles, et en particulier dans le cadre de la HPSG, proposent de traiter un exemple tel que :

3. Paul a parlé à Marie et compris son erreur

sans faire appel à un auxiliaire "nul ou manquant"² qui serait à situer dans le second membre de la coordination, mais en posant une simple coordination de séquences de compléments. Les auteurs signalent d'ailleurs que "l'on peut coordonner des séquences parallèles de compléments qui ne forment pas des constituants", comme dans cet exemple :

4. Paul donnera un livre à Jean et un sou à Marie

sur le même modèle que les coordinations de constituants.

1. À l'exception, bien entendu, de celles que l'on ne peut interpréter comme la coordination de deux phrases, comme : "Pierre et Marie forment un joli couple" que l'on interprète dans ce cas comme une coordination à l'intérieur du syntagme.

2. Le terme "nul", pour rendre compte de l'éllision d'une forme, est en revanche utilisé chez Bouchard et alii (1996) dans le cadre de la GSG.

C'est également notre position. Le traitement proposé³ dans le cadre de l'Approche Pronominale⁴ a pour avantage de rendre compte de manière unifiée, en faisant appel à la notion de "liste paradigmatique", de l'ensemble des coordinations, sans recourir aux procédures d'effacement citées précédemment.

Dans cette approche, le verbe est considéré comme une unité lexicale nucléaire qui définit la forme de sa réaction; celle-ci est alors posée comme une variable abstraite ayant pour représentation prototypique une proforme, mais pouvant être virtuellement représentée par toutes les formes lexicales ou semi-lexicales proportionnelles à cette même proforme.

Dans ce sens, on pose que chaque verbe n'a qu'une seule place de réaction d'un type donné, mais que cette place peut recevoir des remplissages lexicaux successifs. Autrement dit, on va pouvoir instancier lexicalement cette place autant de fois que l'on veut, notamment dans le cas des coordinations. Ainsi, l'énoncé suivant fait apparaître une réalisation en liste du paradigme complément de type "ça" du verbe "il y a":

- il y a des congrès
des sardines
des crabes
des fleurs
des algues
toutes sortes de poissons (carr. 3, 5-8)

Cette notion de liste nous permet de bien différencier le phénomène relationnel créé par "la coordination" et les opérations qui s'y effectuent par le biais de la signification des joncteurs. Ainsi, les listes, qu'elles fassent apparaître ou non un joncteur⁵, renvoient au même phénomène; cependant, l'absence ou la présence du joncteur, de même que sa forme (*et*, *et...et*, *ni*, *mais*, *ou*, *soit...soit*, ...), organise la liste avec des effets sémantiques spécifiques (cf. Ndiaye (89)). En voici quelques exemples:

— effet "additif" avec ou sans "et":

- on assiste à un mouvement où à la fois l'église
l'école
et les familles se désengagent
d'une mission (R.6,9-11)

3. Cf. Bilger (1983, 1984, 1985).

4. Cf. Blanche-Benveniste et alii (1984).

5. À ce sujet, il est intéressant de rappeler que, dans le TALN, on peut donner au signe de ponctuation comme "la virgule" un rôle de coordonnant, cf. Habert (1997:35).

- dans les mairies
dans les postes
dans les lieux publics on a des affichages (Th.,4, 15-16)

— effet “distributif” avec “et ... et” :

- et là on marche et en zigzaguant
et en regardant par terre (Héb.1, 14-15)

— effet “alternatif” avec le connecteur double “soit”, ou avec “ou” :

- cette synthèse vocale va lire soit les mots
soit les phrases
soit les paragraphes entiers (Com.6, 11-12)
- le problème c'est pas le manque de diplôme
ou d'emploi (Th. 9, 7)
- il y avait des gros rubans rouges qu'ils avaient ou sur le chapeau
ou sur le machin (C.9, 7-8)

La répétition lexicale peut s'accompagner d'un changement de modalités, comme dans :

- on met euh de l'huile d'olive
de l'ail
mais pas de coriandre

sans que cela modifie l'analyse par “liste”, puisque la source pronominale atteste toujours d'une forme unique comme ici l'interrogatif “*Que met-on ?*” et de la possibilité d'avoir plusieurs assertifs, en fait, autant d'assertifs que de modalités différentes.

2. PRÉLIMINAIRES : ANALYSE DE QUELQUES SÉQUENCES

2.1. LISTE PARADIGMATIQUE SUR UNE PLACE SYNTAXIQUE

Les constituants fonctionnels majeurs tels que les valences (sujet et complément), les éléments de simple réaction, la séquence “verbe” peuvent aisément se réaliser sous forme de liste. En voici quelques exemples, tirés de corpus oraux, avec ou sans le joncteur “et” :

— *valence sujet : avec “et”* :

- 5 voyages vers le Pôle Sud
et un séjour au Groënland ne l'ont pas guéri (tlrm 2465)

— *valence sujet post-posée sans "et"* :

- à cet ordinateur classique viennent se greffer des éléments
des appareils adaptés aux D.V.
(Com. 2, 11-12)

— *valence complément avec "et"* :

- vous avez réussi à associer le roman
la biographie
et l'essai historique (Troy. 3, 4)
- j'ai dit toi
toi
et toi
et moi (Mat. 25, 4)

— *valence complément sans "et"* :

- elle est vraiment spontanée
naturelle (As.5, 6-7)
- est-ce qu'il est possible d'utiliser
avec le scanner des documents photographiques
des dessins
des graphes (Com. 6, 1-2)
- je n'essaie pas d'appâter l'animal
que ça fasse un gros minou
un gros chat-chat (Bou. 10, 15-16)

— *rection avec "et"* :

- nous sommes financés par le Conseil Général
par le Conseil Régional
et par les communes qui adhèrent à l'association
(Th.2, 15-16)

— *séquence "verbe" avec "et"* :

- l'Algérie est
et restera française
- il faudrait que le public se mette à la place du razeteur
et se dise si moi j'étais... (Cam. II.18, 16)
- c'est un système avec un ressort
qui quand tu le glisses dans une prise s'écarte
et bloque (Esc.2, 15)
- alors le scanner permet d'agrandir
et de réutiliser le document dans leur propre travail
(Com. 5, 5)

- j'essaie de siffler avec ma sonnette
et de la faire marcher (Steph. 5, 2)

— séquence “verbe” sans “et” :

- au dessus il y a des espèces de petits pics qui vont se relever
se rabaisser (Com.3, 5-6)

2.2. LISTE PARADIGMATIQUE SUR PLUSIEURS PLACES SYNTAXIQUES

Cette propriété paradigmatique de réitération peut se manifester sous un aspect simple comme précédemment ou sous un aspect plus complexe ; c'est notamment le cas quand le phénomène de liste concerne une relation syntaxique, autrement dit, quand sont concernés au moins deux paradigmes syntaxiques à la fois pour un seul et même verbe ; c'est le cas du verbe “j'ai pris”, “on va faire” et “(qu')il ait” dans les exemples suivants :

- j'ai pris un coup de corne à la main
et deux coups de corne dans le bras (Cam. II.29, 3-5)
- on va faire ici une haie de peupliers ça abritera du haut
et ici ou des pins
ou des peupliers (MB, 83)
- alors le tout c'est que (le présentateur se plante pas)
que le matin il ait une cravate rouge
et le soir une cravate bleue (R., 134, 10)

La réitération simultanée d'au moins deux ensembles paradigmatiques différents produit plusieurs effets sémantiques, et en particulier, *la distributivité* :

- j'ai pris **respectivement** un coup de corne à la main
et deux coups de corne dans le bras
- ou :
- j'ai pris **et** un coup de corne à la main
et deux coups de corne dans le bras

Nous retrouvons ce même effet distributif dans les exemples suivants :

- on est tous cantonnés dans notre coin **les journalistes d'un côté**
et les acteurs de l'autre (Gel. 8, 7)
- on porte **les femmes** très souvent **des robes à pois**
les hommes **des cravates à pois** (Mar. 3, 6-7)
- L 3 (il fallait) quatre personnes pour maintenir (le cochon)
L 1 **une à la corde**
une à la jambe dessous
une à la jambe dessus
et moi avec le crochet (Mat. 22,16)

2.3. QUELQUES PROBLÈMES DE PORTÉE

Nous ne traiterons pas ici des ambiguïtés qui sont dues au mode de production, car en ce qui concerne l'analyse du phénomène peu importe que la liste soit interprétée comme une "vraie" énumération ou comme une recherche de vocabulaire, ou encore comme un phénomène de retouche. L'exemple suivant:

- quel regard vous portez maintenant sur vos confrères Américains
sur les médias Américains (thrill. 7, 13)

pourrait aisément se prêter à ce problème d'interprétation⁸.

En fait, nous allons essentiellement traiter des ambiguïtés renvoyant à la possibilité d'avoir aussi bien une interprétation paradigmatique que syntagmatique des éléments, ainsi que de celles dues au problème de "portée".

2.3.1. PARADIGME OU "IMBRICATION" SYNTAGMATIQUE

Ce phénomène semble toucher essentiellement la syntaxe nominale qui comporte plusieurs "rections adjectivales" pouvant s'imbriquer les unes dans les autres (cf. Cl. Blanche-Benveniste 95). Dans l'exemple suivant, cité par Sabio (97):

- une femme énorme dégueulasse (Lefebvre, 7, 8)

On peut effectivement avoir deux interprétations, une de type paradigmatique :

- une femme énorme
dégueulasse

ou une de type syntagmatique :

- ((une femme énorme) (dégueulasse))

On a le même problème pour l'exemple suivant :

- Gisèle elle c'est une femme simple quoi qui a des enfants qui doit être
qui doit avoir à peu près le même âge que Simone quoi qui a 2 enfants
(As. 5, 1-3)

dans lequel la liste "définitoire" peut s'interpréter comme étant composée à la fois de la forme adjectivale "simple" et de plusieurs relatives :

8. Nous ne nous étendrons pas non plus sur les ambiguïtés qui concernent les projections du type : "je la vois parfaitement et toi aussi" dans lesquelles "toi" peut être projeté soit sur le paradigme "sujet" soit sur le paradigme "complément".

- a – Gisèle elle c'est une femme simple
 qui a des enfants
 qui doit être
 qui doit avoir à peu près le
 même âge que Simone
 qui a 2 enfants (As. 5, 1-3)
- quoi
 quoi

ou, autre interprétation, être composée seulement de relatives, il s'agit alors du syntagme "une femme simple":

- b – Gisèle c'est une femme simple
 qui a des enfants
 qui doit être
 qui doit avoir à peu près le même
 âge que Simone
- quoi
 qui a 2 enfants

L'exemple qui suit pose le même problème :

- le livre typique qui plaît énormément aux Américains et qu'ils lisent
 dans l'avion en voyageant c'est euh des suspens (Thrill. 2, 11-12)

Première interprétation possible :

- a – le livre typique
 qui plaît énormément aux Américains
 et qu'ils lisent dans l'avion en voyageant c'est euh des suspens
 (Thrill. 2, 11-12)

ou, autre interprétation, sans aucun doute la plus vraisemblable :

- b – le livre typique qui plaît énormément aux Américains
 et qu'ils lisent dans l'avion en voyageant c'est euh des suspens

2.3.2. Le cas des "appositions"

Certaines formes, de part leur sémantisme, vont pouvoir s'interpréter comme étant proches d'une sorte "d'apposition", dans le sens où elles pourraient englober les réalisations lexicales qui les précèdent ou les suivent. C'est le cas notamment de "tout ça"⁹, dans l'exemple suivant :

– interprétation comme simple élément de la "liste" :

- dites voir tout ça
 faire des gâteaux
 faire des crêpes
 acheter des bonbons moi je pourrais le faire (MB)

9. Cf. Bilger (1989).

lequel la liste renvoie à une non-singularité due à la formule “à la fois”, pose ce type de problème, doit-on l’interpréter ainsi :

- il rencontrerait à la fois (l’amour et la raison)
et (la politique)

ou :

- il rencontrerait à la fois (l’amour)
et (la raison et la politique) (Troy. 2-3)

2.3.4. LE RATTACHEMENT DES EXTENSIONS

Certains prolongements vont pouvoir s’interpréter comme concernant un seul terme de la liste ou comme concernant l’ensemble. De fait, les apparitions à un endroit donné de certains éléments ne rendent pas toujours compte de leur “portée” réelle, ainsi dans l’exemple suivant :

- à cet ordinateur classique viennent se greffer des éléments
des appareils adaptés aux D.V.
(Com. 2, 11-12)

la séquence “*adaptés aux D.V.*” (Déficients Visuels) peut concerner soit seulement “*des appareils*” soit l’ensemble de la liste constituée de “*des éléments et des appareils*”. Ce type d’ambiguïté ne peut, bien entendu, avoir lieu que dans le cas où les éléments de la liste partagent les mêmes caractéristiques morphologiques.

En guise de conclusion transitoire, et pour revenir à la problématique de l’analyse automatique, il est intéressant de noter que plusieurs auteurs semblent d’accord pour signaler que le traitement de la coordination oblige à multiplier considérablement le nombre de règles; dans ce sens, beaucoup reconnaissent la nécessité d’avoir des annotations “à profondeur variable” et de poser différents niveaux de complexité dans l’analyse, (cf., par exemple, I. Castellon et alii (1998) qui mentionne au moins trois niveaux de grammaire).

3. PROPOSITIONS DE REPRÉSENTATION POUR UN PARSING

L’objectif de cet article est de proposer une convention d’annotation qui correspondrait le plus à nos analyses. Pour ce faire, je réduirai l’annotation aux parenthésages, en éliminant au maximum les étiquettes pour ne conserver que les découpages en constituants fonctionnels.

3.1. RAPPEL DE QUELQUES PRINCIPES

1. Tout terme en relation de construction possède la propriété de pouvoir être réitéré lexicalement.
2. Les éléments “coordonnables” et “coordonnés” se situent sur l'axe paradigmatique, ils n'occupent qu'une seule et même place syntaxique en tant que constituant ou sous-constituant.
3. La liste ainsi formée est équivalente à “un seul élément syntaxique”, elle est d'ailleurs proportionnelle à une seule proforme.
4. Cette propriété de réitération peut s'appliquer à plusieurs paradigmes à la fois, dans une même construction verbale.

3.2. LES LISTES “PRÉVISIBLES”

Si l'on se limite aux constituants majeurs de la construction verbale, on peut déjà prévoir un certain nombre de possibilités, telles que :

1°) tout terme “sujet”, “objet”, “rection” et “verbe” peut être réitéré :

- (S et S) V O Rc
- S V (O et O) Rc
- S V O (Rc. et Rc.)
- S (V et V) O Rc.

2°) le terme “sujet” peut avoir une séquence (verbe+objet) ou une séquence (verbe+objet+rection) réitérée :

- S (V O) et (V O) Rc.
- S (V O Rc.) et (V O Rc.)
- Paul téléphonerà à Marie (ce soir)
et écrira à Gisèle demain

3°) toute séquence (sujet+verbe) peut avoir une séquence (objet+rection) réitérée :

- S V (O Rc.) et (O Rc.)
- Paul téléphonerà à Marie ce soir
et à Gisèle demain

4°) tout (verbe) peut avoir une séquence (sujet+objet) ou une séquence (sujet+objet+rection) réitérée :

- S V O et (S O)
- S V O Rc. et (S O Rc.)
- Paul téléphonerà à Marie (ce soir)
et Max à Gisèle (demain)

5°) Enfin, la réitération du terme “sujet”, “objet” et “rection” peut être modalisée :

- S V O Rc. et (S+mod.)
- S V O Rc. et (O+mod.)
- S V O Rc. et (Rc.+mod.)
- Paul téléphonera à Marie ce soir **et Max aussi**
- Paul téléphonera à Marie ce soir **et à Gisèle non**
- Paul téléphonera à Marie ce soir **et demain aussi**

Bien entendu, chaque type de réitération se caractérise par un certain nombre de contraintes et de règles que nous ne rappellerons pas ici¹⁰.

3.3. TERMINOLOGIE ET NIVEAUX D'ANALYSE

La terminologie que nous nous proposons d'utiliser dans cet essai de parsing est la suivante: SéqS, SéqV, SéqO, SéqRection; elle a été choisie pour être une des "moins marquées" et pour faciliter la démonstration. Elle nécessite cependant quelques éclaircissements.

- Dans une première étape, et comme nous l'avons déjà vu, on identifie les constituants fonctionnels notamment par l'équivalence qu'ils entretiennent avec une proforme; à noter que, bien entendu, cette opération de proportionnalité ne fonctionne pas pour la “Séquence Verbe”. Les parenthèses délimitent ainsi les constituants en question.
- Les Séquences S, O et Rection peuvent être constituées d'éléments susceptibles d'être traités dans une seconde étape. En effet, chacun de ces constituants peut se réaliser, par exemple, sous la forme d' :
 - une extraction en “c'est...que”,
 - une coordination
 - ou encore comporter :
 - des relatives,
 - des que-phrases,
 - des appositions,
 - etc.
- La Séquence Verbe comporte pour sa part: le verbe conjugué ou l'auxiliaire, les pronoms clitiques “sujet et compléments”, les particules de négation, mais aussi les modaux, les participes et infinitifs ainsi que les adverbes.

10. De même, nous ne mentionnerons pas les mises en facteur possibles concernant les sous-constituants tels que les auxiliaires d'aspect ou encore les verbes modaux.

3.4. QUELQUES APPLICATIONS

Quand le phénomène de liste ne concerne qu'une seule place syntaxique, la segmentation semble poser peu de problème, en voici 3 exemples :

— **position sujet :**

a. *l'église l'école et les familles se désengagent d'une mission*

– Premier niveau d'analyse :

(l'église l'école et les familles) SéqS

(se désengagent) SéqV

(d'une mission) SéqO

– La forme de la “SéqS” peut être traitée ultérieurement selon un deuxième niveau d'analyse :

- (l'église l'école et les familles) : séquence “sujet” réalisée sous la forme d'une liste, présence du joncteur “et”, énumération avec effet de pluralité.

— **position objet :**

b. *cette synthèse vocale va lire soit les mots soit les phrases soit les paragraphes entiers* (Com.6, 11-12)

– Premier niveau d'analyse :

(cette synthèse vocale) SéqS

(va lire) SéqV

(soit les mots soit les phrases soit les paragraphes entiers) SéqO

– La forme de la “SéqO” peut être traitée ultérieurement selon un deuxième niveau d'analyse :

(soit les mots soit les phrases soit les paragraphes entiers) : séquence “objet” organisée en liste, “attelée” par “soit” connecteur double, effet alternatif.

— **position recton :**

c. *dans les mairies dans les postes dans les lieux publics on a des affichages*

– Premier niveau d'analyse :

(dans les mairies dans les postes dans les lieux publics) SéqRection

(on a) SéqV

(des affichages) SéqO

- La forme de la SéqRection s'analyse comme :

(dans les mairies dans les postes dans les lieux publics) : séquence
 “rection loc.” organisée en liste, absence du joncteur “et”, effet
 d'énumération.

Dans le cas où plusieurs paradigmes syntaxiques sont concernés par le phénomène de réitération, le système de parenthésage devra quelque peu se différencier. Cela va dépendre à la fois de la séquence mise en facteur commun et du statut que l'on va pouvoir donner au regroupement.

Ainsi, dans le cas d'une mise en facteur de la “SéqS” seule, ou accompagnée de la “SéqV”, nous pouvons proposer le même type de parenthésage :

- exemple de réitération des paradigmes “verbe” et “objet”; seule la “SéqS” est en facteur commun :

“Paul parlera à Marie et comprendra son erreur”

(Paul) SéqS
 (parlera) SéqV
 (à Marie) SéqO
 (et)joncteur
 (comprendra) SéqV2
 (son erreur) SéqO2

- exemple de réitération des paradigmes “objet” et “rection”; la “SéqS”, accompagnée de la “SéqV”, est en facteur commun :

“Paul livrera un livre ici et un disque dans cet immeuble”

(Paul) SéqS
 (livrera) SéqV
 (un livre) SéqO
 (ici) SéqRection.
 (et) joncteur
 (un disque) SéqO2
 (dans cet immeuble) SéqRection 2

En revanche, quand les paradigmes “sujet” et “objet”, ou “sujet” et “rection”, sont réitérés et quand seule la séquence “verbe” est mise en facteur commun, on a tout intérêt, du moins pour éviter tout problème lié aux accords, à proposer une autre autre solution.

On pourrait analyser l'exemple : *“Paul parlera à Jacques et Marie à Gisèle”* ainsi :

– Premier niveau :

(Paul) SéqS
 (parlera) SéqV
 (à Jacques) SéqO
 (et) joncteur
 (Marie à Gisèle) SéqNexus¹¹.

Cette séquence “nexuelle” serait également à analyser à un autre niveau, celui de la macro-syntaxe, comme étant formée de :

(Marie) SéqPréfixe projetée sur SéqS
 (à Gisèle) SéqNoyau projetée sur SéqO.

L'argument qui justifierait en quelque sorte la création d'un tel constituant “SéqNexus” pourrait se trouver dans le fait que l'on a affaire ici à une structure syntaxique particulière que l'on a tout intérêt à signaler. Comme nous l'avons déjà vu, le même contenu de sens aurait pu se présenter de façon liée sous la forme :

– *Paul et Marie parleront respectivement à Jacques et à Gisèle*

(Paul et Marie) SéqS
 (parleront respectivement) SéqV
 (à Jacques et à Gisèle) SéqO

On pourrait également faire appel à “ce nexus” pour rendre compte des énoncés tels que :

– *il y en avait qui s'habillaient en blanc et d'autres non*

Le parenthésage proposé serait :

(il y en avait qui) SéqS
 (s'habillaient) SéqV
 (en blanc) SéqRection
 (et) joncteur
 (d'autres non) SéqNexus modalisé

Traité dans une deuxième étape, ce regroupement s'analyserait comme suit :

(d'autres) SéqPréfixe projetée sur SéqS
 (non) SéqNoyau projetée sur la modalité de la SéqV.

Il semble intéressant d'avoir la possibilité de recourir, dans le cadre des annotations, à un “constituant” de ce type. On pourrait d'ailleurs poser que sur toute construction verbale, au sens large, ce type de séquence peut

11. Cette proposition de terminologie nous a été suggérée par Claire Blanche-Benveniste, et renvoie également aux travaux de O. Eriksson (1993).

venir se greffer. Et, selon sa constitution, on pourrait déjà en prévoir les effets :

– effet distributif.

ou – effet + ou - contrastif.

Ceci permettrait, par exemple, d'intégrer dans ce cadre descriptif, un énoncé limite du type :

- l'économie de la région elle est essentiellement basée sur le tourisme *et dans les campagnes l'agriculture* (Port. 2, 2-3)
 - (l'économie de la région) SéqS
 - (elle est essentiellement basée) SéqV
 - (sur le tourisme) SéqO
 - (et) joncteur
 - (dans les campagnes l'agriculture) SéqNexus, à effet distributif

La SéqNexus pouvant s'analyser comme :

(dans les campagnes) SéqPréfixe projetée sur "région", complément de nom du sujet
(l'agriculture) SéqNoyau projetée sur "tourisme", SéqObjet .

Pour conclure, je voudrais rappeler que cet article s'inscrit dans le cadre de la demande de coopération entre informaticiens et linguistes, demande signalée au début et à laquelle les membres du GARS essaient de répondre. Les propositions avancées ici doivent être interprétées comme une première ébauche de réflexion sur le type de représentation que pourraient avoir les tournures coordonnées.

Mireille BILGER
Université de Provence



Bibliographie

- ABEILLE, A. & GODARD, D., 1996, "La complémentations des auxiliaires français", *Langages*, n° 122, Larousse, 32-61.
- BILGER, M., 1983, *Étude distributionnelle de la coordination par ET*, Thèse de 3^e cycle, U. de Provence.

- , 1984, "Et, quoi de neuf ?", *RSFP*, 6, 81-108.
- , 1985, "Pour une nouvelle analyse des coordinations dites par gapping", *Queste*, 2, 175-191.
- , 1988, "Les réalisations en 'et tout (ça)' à l'oral", *RSFP*, 9, 97-111.
- , 1998, "Le statut micro et macro-syntaxique de 'et'", in *Analyse Linguistique et approches de l'oral, Recueil d'études offert en hommage à Claire Blanche-Benveniste*, éd. Bilger & van den Eynde & Gadet, Leuven-Paris, Peeters, col. Orbis/Supplementa, t. 10.
- BLANCHE-BENVENISTE, C., 1995, "Le semblable et le dissemblable en syntaxe", *RSFP*, 13, 7-33.
- BLANCHE-BENVENISTE, C., DEULOFEU, J., STEFANINI, J. & EYNDE, K. van den, 1984, *Pronom et Syntaxe. L'approche pronominale et son application au français*, Paris, Didier-Érudition.
- BLANCHE-BENVENISTE, C., BILGER, M., ROUGET, C. & EYNDE, K. van den, 1990, *Le français parlé, études grammaticales*, Paris, CNRS.
- BOUCHARD & EMIRKANIYAN & LABELLE, 1996, "Traitement de l'ellipse du verbe dans les coordonnées en GSG", *Travaux de linguistique*, 32.
- CASTELLON, CIVIT & ANTERIAS, 1998, "Syntactic Parsing of Unrestricted Spanish Text", in *First International Conference on Language Resources & Evolution*, vol. 1, éd. Rubio et alii, Granada, 603-609.
- DELECHÈLE, G., 1994, "Relations inter-énoncés : de la subordination à la coordination", *RANAM*, 27, "Syntaxe et parataxe", 29-41.
- DUBOIS, S., 1995, "Structural processes in enumeration", in *Evolution and Change*, n° 4, Cambridge University Press.
- ERIKSSON, O., 1993, *La phrase française. Essai d'un inventaire de ses constituants syntaxiques*. Suède, Acta Univeritatis Gothoburgensis, Romanica Gothoburgensia, XLII.
- FUCHS, C., 1996, *Les ambiguïtés du français*, Ophrys, Paris.
- GIRARD, G., 1994, "To/-ing : coordination, subordination ou autre type de liaison ?", *RANAM*, 27, "Syntaxe et parataxe", 43-56.
- HABERT, B., NAZARENKO, A. & SALEM, A., 1997, *Les linguistiques sur corpus*, Armand Colin, Paris.
- LE GOFFIC, 1993, *Grammaire de la phrase française*, Hachette, Paris.
- N'DIAYE, M., 1989, *L'analyse syntaxique par joncteurs de listes*, Thèse nouveau régime, UP.

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

leux, 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 reconnues 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 précedence : il y a interaction entre hypothèses et observations, mais pas d'exigence de précedence 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

exignd. ([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 $\text{exclu}(\text{phn}, [[\text{\u00e9}t\text{r}\text{e}f], [\text{aux}f], [1\text{vaf}], [2\text{vaf}], [1\text{vef}], [2\text{vef}]])$.

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 :

$\text{exclu}(\text{id}, [[\text{cat}1, \text{catn}], [\text{cat}2, \text{catm}], [\text{cat}3]])$.

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 $\langle [\text{cat}1, \text{catn}] \text{ et } [\text{cat}2, \text{catm}] \rangle$, $\langle [\text{cat}1, \text{catn}] \text{ et } [\text{cat}3] \rangle$, $\langle [\text{cat}2, \text{catm}] \text{ et } [\text{cat}3] \rangle$ 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 $\langle \text{id}, \{ _ \}, \text{précède}(\text{cat}1, \{ \text{cat}2 \}) \rangle$.
- ii $\langle \text{id}, \{ \{ \text{catm} \} \setminus \{ \text{catn} \} \}, \text{précède}(\text{cat}1, \{ \text{cat}2 \}) \rangle$.

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(êref, [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



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'à 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_j 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_j — la représentation sémantique qui sera associée à un modèle — disons modèle_j — dont le modèle_j 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_{nv}* modèle de la phrase verbale noyau, *m-phv* modèle de phrase verbale, on aura

((La °fille) m-sn °(la °regarde) m-phv_{nv})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 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 *Iva*, une forme *être* et un clitique nominatif, alors, si la forme verbale *être* est *été*, le participe passé *Iva* 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 *Iva* 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 et le clitique nominatif s'accordent en nombre, la forme *être* et le clitique nominatif s'accordent en personne (ex : *sont-elles aimées ?*).

ou si la forme verbale *être* est *être* et il y a un clitique réfléchi, la forme *être* 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 minime : elle consiste à l'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 Pa1 à Pa9 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 P12 et P13 du Niveau 2 restent sans modification. Les Propriétés P11, P14, P15, P16 et P17 sont modifiées de la manière qui suit :

P11[Niveau 3-vnpu]	< phn, { _ }, précède(f, { cln, été, p, pas }) >.
P14[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'infinitif.

- 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 $exig(phen, [[l'], [^etre], [copulatif]])$.

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 corrélér 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 pas 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 ou 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 descriptiops 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 ANTTILA [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.

Maquette et mise en page
D. BLANC
13100 AIX-EN-PROVENCE
tél. 04 42 21 58 25 - fax 04 42 21 02 54

Imprimé sur les Presses
de l'A.N.R.T.

Dépôt légal - 4^e Trimestre 1999
ISBN - N° 2-85399-444-9

