ERRATUM

Suite à un problème technique, la police phonétique utilisée dans l'article de Marie-Noëlle Roubaud "*Du bon usage des notations phonétiques dans les corpus aphasiques*" est illisible. Nous sommes désolés de cet incident et nous republierons l'article dans sa totalité dans le numéro 18 de *Recherches sur le Français Parlé.*
L'article est d'ores et déjà disponible sur Internet à l'adresse :

http ://www.up.univ-mrs.fr/delic

20 SEP. 2002

Groupe Aixois de Recherches en Syntaxe

Centre des Sciences du Langage

RECHERCHES SUR LE FRANÇAIS PARLÉ

N° 17 - 2002

G.A.R.S.

2002

PUBLICATIONS DE L'UNIVERSITÉ DE PROVENCE

© PUBLICATIONS DE L'UNIVERSITÉ DE PROVENCE

29, avenue Robert Schuman - F - 13621 Aix-en-Provence cedex 1
Tél. + 33 (0) 4 42 95 31 91 – Fax + 33 (0) 4 42 20 28 04
Courrier électronique pup@up.univ-aix.fr
Catalogue complet sur internet www.up.univ-mrs.fr/wpup

Sommaire

Présentation 7
José DEULOFEU

Entre l'auxiliaire et le participe passé 11
Paul CAPPEAU

Note à propos de la construction dite du « datif partitif » 29
André VALLI

Détachement et linéarité 41
Mylène BLASCO-DULBECCO & Sandrine CADDEO

L'opposition de modalité en français parlé : étude macro-syntaxique 55
Frédéric SABIO

Les amorces de mots comme faits autonymiques en langage oral 79
Berthille PALLAUD

Du bon usage des notations phonétiques dans les corpus aphasiques 103
Marie-Noëlle ROUBAUD

Proposition pour un étiquetage en « séquences fonctionnelles » 117
Mireille BILGER & Estelle CAMPIONE

Comment présenter les « types de phrases » dans une grammaire de référence d'après corpus 137
José DEULOFEU

Résumés de thèses 169

Présentation

José DEULOFEU
Université de Provence

Le présent numéro de *Recherches sur le français parlé* s'inscrit dans les grandes lignes de recherche du groupe DELIC qui prolongent celles du G.A.R.S.

Un premier groupe d'articles prolonge l'entreprise de description du français à partir de corpus intégrant des données orales. Les deux grands domaines de la description (microsyntaxe et macrosyntaxe) sont abordés ainsi que leurs interrelations :

Dans le domaine de la microsyntaxe, Paul Cappeau propose, à partir de données orales, un classement des séquences qui peuvent figurer entre l'auxiliaire et le participe passé. La prise en compte de ce type de données lui permet à la fois de révéler de nouveaux faits (syntaxe de la négation) et de dégager les tendances structurales les plus remarquables qui se manifestent sur cette position.

André Valli à propos des constructions dites « datif partitif » souligne que le fait de retenir ou de ne pas retenir telles ou telles données a des conséquences dans la description syntaxique de la question : l'usage tel qu'il ressort de l'examen de corpus de langue parlée et de français écrit littéraire est en écart par rapport à l'usage puriste des grammairiens, et en revanche, conforme à l'usage qui était régulier dans les textes d'ancien et de moyen français.

Deux autres articles montrent que la description de certaines structures demandent que l'on articule finement les niveaux macro et micro syntaxiques.

Mylène Blasco-Dulbecco et Sandrine Caddéo montrent que des sous types de constructions présentant des dislocations doubles (pronoms forts + substantifs) peuvent être distinguées, à la fois par des propriétés micro (type de pronom) et par des propriétés macro (« éloignement » des éléments disloqués).

Frédéric Sabio propose de même une typologie de structures présentant des « contrastes » : *c'est pas l'appareil qui vaut cher. C'est les objectifs* ; *Ce ne sont plus les pays qui intéressent mais les gens. Pierre est venu, mais pas Marie* sur la base de la distinction entre propriétés d'ordre macro et propriétés d'ordre micro : certains contrastes imposent que la construction soit produite selon plusieurs unités macrosyntaxiques.

Deux autres articles abordent le problème de la description par une interrogation sur les données.

Bertille Pallaud s'interroge sur le statut des interruptions et des fragments de mots. Elle propose de les inclure dans la description dans la mesure où leurs occurrences répondent à des distributions contrainte et où « leur présence témoigne d'une réflexion ou d'une activité linguistique se traduisant par une interruption de morphèmes en cours d'énonciation ». C'est dire que l'articulation entre signifiant et signifié est présente dans ce domaine et qu'il est donc justiciable d'une étude proprement linguistique.

Marie-Noëlle Roubaud prolonge la réflexion du groupe sur les productions des locuteurs aphasiques, qui nous permettent de mieux cerner les limites du système. Elle aborde le problème de la présentation des transcriptions de telles données sous l'angle limité de l'emploi des signes API. « Le problème n'est pas simple : tout transcripteur, face à la recherche du mot par le locuteur aphasique, fait des hypothèses sur le morphème recherché ; que doit-il alors transcrire ? L'article expose ces difficultés et apporte des éléments de réponse pour l'usage des notations phonétiques ».

Un dernier groupe d'article propose de faire le lien entre l'activité de description linguistique et certaines applications pratiques qui lui sont liées.

La construction d'outils pour l'exploitation de corpus informatisés est une des nouvelles priorités du groupe. Au delà des concordanciers et des taggeurs, une question centrale pour aider la description syntaxique est de construire des « parseurs » adaptés aux données orales. Mireille Bilger et Estelle Campione proposent une étude préalable à la construction de tels outils. A l'inverse de l'approche dominante en TAL qui s'efforce de mettre au point des stratégies d'analyse partielle (ou shallow parsing) des énoncés, faisant apparaître des segments intermédiaires entre le mot et la « phrase », qui sont le plus souvent des suites de mots très courtes, elles proposent une stratégie d'analyse partielle non pas en segments minimaux, mais au contraire en « séquences » maximales, sans entrer dans le détail de leur composition.

Le groupe se pose la question de la présentation des résultats des recherches sous la forme synthétique d'une grammaire du français à partir de corpus. La question du métalangage et de la terminologie utilisée dans

un tel ouvrage est alors cruciale si l'on veut qu'il soit utilisable pour tous. Dans ce cadre, José Deulofeu présente une critique de la notion, en apparence indispensable, de « type de phrase » et montre la difficulté de la remplacer par une autre terminologie reflétant mieux l'état des descriptions linguistiques.

Entre l'auxiliaire et le participe passé

Paul CAPPEAU
Université de Poitiers

Résumé

Le but de cet article est de présenter un relevé et un classement des séquences qui peuvent figurer entre l'auxiliaire et le participe passé. Le travail a été conduit à partir de données orales. Ce choix tient à deux sortes de raisons : les exemples attestés sont nombreux et faciles à recueillir mais surtout un certain nombre de faits (comme ceux relevés sur la négation) semblent peu présents et moins variés dans les textes écrits. L'oral permettrait donc, semble-t-il, de proposer une description plus détaillée. On a supposé aussi que, comme pour d'autres descriptions, avec l'oral on accédait aux faits majeurs de la distribution, qui était ainsi moins « parasitée » par des phénomènes d'ordre stylistique (et donc plus atypiques). Bref, l'oral devrait donner une bonne vision de la langue.

1. Comment procéder ?

1.1. Où chercher des informations ?

Les débats sur l'ordre des mots ont conduit les grammairiens à s'intéresser aux éléments qui peuvent se placer entre l'auxiliaire et le participe. Mais ce point n'a jamais soulevé de débats passionnés. Ainsi, Oudin (1640) apporte un commentaire prudent et de l'ordre du constat (« nous posons librement ») :

> « Nous posons librement quantité de mots entre les auxiliaires et le participe, aux temps composés des verbes actifs, aux passifs, et réciproques : par exemple, *les affaires douteuses ont de tout temps grandement empêché les esprits : je suis à dire la vérité bien aimé de vous; je me suis sans sujet toujours estimé malheureux*. Si on ne pose qu'un seul adverbe, il le faut mettre entre lesdits auxiliaires et la composition ; *ont*

Recherches sur le français parlé, 17, 11-27.
©2002, *Université de Provence*

> *grandement empêché* : *je suis bien aimé* : *je me suis toujours estimé* : et toujours il faut que dans un plus grand nombre de dictions, ledit adverbe soit proche du participe, comme vous avez vu dans les premiers exemples. » (p. 201).

L'intérêt de ce passage réside notamment dans les exemples qui montrent des insertions de diverses sortes et dans une amorce de règle de placement des différents segments insérés. Brunot (1905-1953) permet de suivre quelques éléments des discussions qui ont porté sur ce point. Vaugelas, en conformité avec ses positions sur le rapprochement des termes solidaires comme le sujet et le verbe ou l'antécédent et le pronom relatif, plaide pour des insertions réduites entre auxiliaire et participe. Mais finalement, ce débat a peu retenu l'attention des grammairiens et Brunot observe que seul le COD[1] tend à être écarté de cet emplacement alors que les circonstanciels peuvent s'y maintenir, comme dans :

> puisque vous avez, par un effet de votre bonté et d'une générosité sans exemple, voulu faire un choix si peu digne de vous (Bussy-Rab., cité par Brunot)

Dans les grammaires contemporaines, quelle entrée rechercher dans un index lorsque l'on s'interroge sur les éléments qui peuvent prendre place entre l'auxiliaire et le participe passé ? *Auxiliaire* renvoie à divers points tels que le choix entre *être* et *avoir*, sous *participe*, on traite en général des problèmes d'accord, *place* ou *ordre des mots* orientent le plus souvent vers l'adjectif ou vers le sujet. Du côté de l'ordre des mots, toutefois, la « pêche » n'est pas totalement infructueuse. Des ouvrages déjà anciens tels que ceux de Blinkenberg (1928) ou de Boer (1954) contiennent bien quelques indications mais elles sont limitées à des adverbes particuliers ou aux compléments circonstanciels qui ne constituent qu'une partie des éléments rencontrés. C'est finalement dans les passages portant sur l'adverbe que l'on trouve fréquemment des indications sur ce problème : Martinon (1927) y fait une simple allusion, Nølke (2001) développe des analyses précises liées à la notion de portée.

---

[1] C'est là un changement notable avec la langue ancienne. Marchello-Nizia (1997) signale qu'au XIVe siècle le participe est fréquemment « séparé de l'auxiliaire par un régime direct ou indirect » (p. 407) et elle cite deux exemples tirés de Bérinus :

- quand vous avrez *femme* prise
- je leur ay *ma foy* plevie

1.2. Les « points » de repère

Pour l'inventaire (partie 2) tout autant que pour la combinatoire (partie 3), il semble utile de disposer de bornes stables pour pouvoir « calculer » la position des éléments qui apparaissent entre l'auxiliaire et le participe. Le recours à une forme composée de l'auxiliaire permet d'augmenter les points de repère et de délimiter deux zones distinctes :

avoir + zone A + *été* + zone B + *ppé*

On trouve assez peu de termes spécifiques de la zone A et de la zone B. Les plus probants seraient :

zone A : *pas* (et d'autres termes négatifs)
zone B : *très* (et variantes)

il n'a *pas* été très perturbé / ? il n'a été *pas* très perturbé
* il a *très* été perturbé / il a été *très* perturbé
il a été *vraiment* perturbé / il a *vraiment* été perturbé

En dehors de ces quelques cas très nets, les exemples attestés montrent que des tendances existent et que les éléments ne se distribuent pas au hasard dans la chaîne linéaire.

Lorsque la forme *été* n'est pas réalisée, on n'identifie plus qu'une seule zone dans laquelle les divers éléments peuvent se rencontrer, ce qui « écrase » quelque peu les différents emplacements posés.

1.3. Quel type de description ?

Ce travail s'appuie sur des corpus oraux de diverses provenances, du « tout venant »[2]. La description vise essentiellement à « débroussailler » les paradigmes identifiables et leur combinatoire. On ne s'est intéressé ni aux conditions externes qui peuvent influencer la réalisation de ces insertions (on n'a notamment pas porté un réel intérêt aux genres), ni à la valeur ou aux effets de sens que cette insertion révèle.

[2] Il s'agit de la banque de données orales CorpAix (un million de mots) qui rassemble des corpus très divers (entretiens, monologues, situations formelles, explications, récits, etc.).

2. La distribution observée

Les éléments qui apparaissent entre l'auxiliaire et le participe passé ne peuvent être appréhendés ni à partir des catégories, ni en tant que mots isolés.

L'éclairage par les catégories ne peut être envisagé car tous les éléments d'une classe ne présentent pas un statut syntaxique unique. Aussi, la description à partir des adverbes (que l'on trouve dans plusieurs des ouvrages cités auparavant) ne peut être que fragmentaire[3]. En effet, les formes rassemblées présentent des propriétés syntaxiques non homogènes. Les commentaires se présentent alors sous la forme d'une comparaison entre diverses unités et fournissent peu d'indications sur les regroupements en classes comme l'illustre le passage suivant :

> « Lorsque le verbe est à un temps composé, l'adverbe se place avant ou après le participe : *Il a lentement marché; Il a marché lentement*, mais les adverbes très courts ne peuvent pas toujours se placer avant le participe : on dira : *J'ai trop aimé*, mais on ne dira pas : *J'ai là aimé*, sans doute parce que, dans ce dernier cas, l'adverbe ne caractérise pas l'action, mais la localise, ce qui invite davantage à construire analytiquement : on ne dira pas non plus : *J'ai à Paris aimé.* » (de Boer, 1954, § 531).

Il n'est pas non plus possible de travailler sur des mots isolés car il existe des suites syntagmatiques qui doivent être appréhendées comme un seul bloc (on parlera de séquences pour les désigner). Ce point est sensible dans la description de la négation qu'on ne peut limiter à la forme *pas*.

Plusieurs séquences ont pu être identifiées. La négation, qui est étroitement liée au verbe, sera la première décrite. Nous avons ensuite regroupé deux séquences sensibles au lexique verbal (la séquence évaluative et la séquence aspectuelle). Puis, à partir des propriétés syntaxiques, nous distinguerons deux nouvelles classes : les séquences associées et les séquences régies. La combinatoire entre ces diverses blocs sera traitée dans la partie suivante (point 3).

2.1. La séquence négative

Cette séquence se réalise dans la zone A (c'est-à-dire entre l'auxiliaire et la forme *été*) :

---

[3] Nøjgaard (1992), à partir d'une vingtaine de tests, propose un classement des adverbes en 7 grandes rubriques qui se subdivisent à leur tour, ce qui montre assez le caractère hétérogène de cette pseudo catégorie.

vous n'avez **pas** été mélangés avec de futurs policiers (GRE)
si ces accords n'avaient **pas** été respectés (LIC 97)

On peut parler de négation de relation (Blanche-Benveniste, 1992) car c'est la relation entre le verbe et les éléments qu'il construit qui est affectée. Le passage en zone B (entre *été* et le participe) de la séquence négative conduit souvent à un rejet de l'exemple (aucun exemple réel de ce modèle ne se trouve d'ailleurs dans notre banque de données) :

* vous n'avez été **pas** mélangés

L'interprétation d'un exemple tel que :

s'ils avaient été **pas** respectés

est possible si *pas* est vu comme une modalité de terme (dans ce cas, *non pas* serait plus attendu et *ne* est forcément absent) ce qu'une suite à l'énoncé peut indiquer, ou si l'ensemble *pas* + *participe* est vu comme une mention[4] :

s'ils avaient été (non) pas respectés mais pourchassés

Il est utile de fournir un inventaire des éléments qui peuvent constituer cette séquence car elle présente une grande richesse de formes. Elle peut être constituée :

- d'un terme unique[5] (*pas, jamais...*) :

  je n'ai **jamais** eu envie de travailler pour une grosse société (Jeanne)

- des modifieurs de la négation qui se réalisent avant le terme négatif :

  qui ont **même pas** été à l'école (Portuga)
  il n'a **pratiquement plus** quitté son domicile (Avocat)
  c'est des gens qui ont **tellement pas** bougé depuis tellement longtemps (FR-PCR)

- d'un terme négatif et de son prolongement :

  j'avais **pas du tout** intégré le titre (Rapin)

- d'une séquence avec modifieur + négation + prolongement :

  il n'a **vraiment pas du tout** apprécié (énoncé inventé)

---

4 *cf.* la fin de cet échange entendu à la radio (qui serait plus à sa place dans ce travail si un participe était substitué à *bien...*) :

L1 mais s'il y a un truc qui est pas bien c'est normal d'en parler [...] L2 c'est pas pas bien c'est pas pas bien.

5 On ne s'intéressera pas à la présence ou à l'absence du premier segment (*ne*).

Les modifieurs permettent à la séquence négative de constituer une réponse autonome, ce qui n'est pas toujours possible pour un terme négatif isolé. C'est ce qu'illustre l'exemple de *pas* :

as-tu bougé de ton trou ? * pas / presque pas / pas du tout

Dans cette première partie, les éléments qui constituent la séquence négative peuvent être vus comme des termes qui renforcent, qui intensifient la valeur exprimée par le morphème négatif : *vraiment pas, pas du tout, etc.* Dans ce cas, la relation entre l'énoncé négatif s'oppose à l'énoncé positif :

Énoncé positif	**Énonces négatifs**
j'avais fait le lien	je n'avais **pas** fait le lien
	j'avais **absolument pas** fait le lien (Rapin)
	je n'avais **vraiment pas** fait le lien

On reviendra plus loin (partie 3) sur la réalisation d'une négation avec une autre séquence et sur les problèmes de portée (Larrivée, 2001) qui compliquent l'analyse lorsque plusieurs séquences se combinent.

2.2. Les séquences évaluatives et aspectuelles

Les deux types de séquences rassemblées dans cette partie sont très sensibles au lexique verbal, il s'agit en quelque sorte de modifieurs du verbe. Leur emplacement est toutefois distinct : l'une est plus liée au participe, l'autre à l'auxiliaire *été* (quand il est présent).

2.2.1. La séquence évaluative ou qualitative

Elle porte sur le processus verbal et apporte des indications de nature qualitative ou quantitative :

vous trempez un doigt de telle manière qu'il soit **largement** mouillé (FR-BES)
les différences ont été **un peu** gommées (Audrey)
la seconde s'est **relativement mal** passée (FR-Val)

On identifie précisément son point d'accroche quand la partie verbale comporte trois éléments. La séquence évaluative est alors placée dans le voisinage immédiat du participe passé, c'est-à-dire en zone B (entre *été* et le participe) :

le quartier avait été **passablement** touché (Birnie)
ce flux a été **complètement** coupé (MédSoc)
des pays + + latins qui ont été enfin **plus ou moins** influencés (Corse)

Cette séquence se place difficilement dans la zone A (entre l'auxiliaire et *été*) :

? ça a **un petit peu** été improvisé (ex inventé)
? il avait **très mal** été soigné (ex inventé)

Les rares exemples contraires rencontrés semblent limités à quelques termes particuliers (il faudrait ainsi creuser la syntaxe réelle de *beaucoup*) :

elle a **beaucoup** été suivie (Benot)
on a **tellement** été habitué depuis (PRCIEU)

Cette séquence ne semble pas réalisable en dehors de la partie verbale et notamment en tête d'énoncé :

? beaucoup elle a été suivie
? passablement le quartier avait été touché

La séquence évaluative porte directement sur le sens lexical du verbe. C'est pourquoi elle se place dans le voisinage immédiat du participe et connaît quelques limitations dans la combinatoire :

je suis bien / * mal / * le plus / * un petit peu /? tellement / arrivé
j'ai été * vachement / ? légèrement / très bien / logé

Avec une négation, la séquence évaluative est acceptée si elle reste en zone B (après *été*), comme le montrent les exemples attestés :

on (n') a pas été **très très bien** logés (Évacuation)
j'ai pas été **suffisamment** découragée (Avocat)

On ne rencontre pas ce bloc *négation + séquence évaluative* en zone A (avant *été*). Les exemples inventés peuvent ne pas heurter, ils n'en sont pas moins absents des données recueillies :

? on (n') a pas **très très bien** été logés
? j'ai pas **suffisamment** été découragés

De tels exemples posent des problèmes de frontières entre séquences qui seront évoqués dans la partie 3.

L'inventaire des éléments de cette séquence montre une assez grande variété. Elle peut comporter :

- un seul terme (*beaucoup*, *très*)

  moi je pense pas que ça ait te- que ça ait **tellement** changé (Portuga)
  tiens il a **bien** changé (CorpusAD)

j'avais **vachement** perdu (Portuga)
ils ont été **légèrement** modifiés depuis (Ciergerie - modifié)[6]

- une série de termes

des facteurs + avec lesquels j'ai **le plus** sympathisé qui font de la musique (CorpusMe)
il y a des cas qui t'ont **plus ou moins** touchée moralement (Infirm)

On peut trouver aussi un modifieur du terme évaluatif (qui permet de se déplacer sur une échelle graduable) :

je l'ai **un peu trop** malmené au sens + euh + de l'ironie (CorpusMe)
je l'ai **vraiment trop** malmené / je l'ai **bien trop** malmené / je l'ai **beaucoup trop** malmené
vs je l'ai **trop** malmené (évaluatif sans modifieur)

Certains termes sont compatibles avec des adjectifs (*très, vachement*...), d'autres sont plus liés à du verbal (*beaucoup*)[7] :

et ça ça m'a **vraiment** étonnée (Portuga) / c'est vraiment rouge / c'est vraiment joli
j'ai **beaucoup** aimé ce bouquin (Portuga) / ? c'est beaucoup joli / ? c'est beaucoup rouge / il est beaucoup malade

L'oral spontané permet de rencontrer des réitérations de cette séquence (généralement avec une valeur intensive) :

il nous a **beaucoup beaucoup** appris à aimer le chant (PR6)

Il paraît souhaitable d'inclure dans cette série des termes d'approximation, des commentaires[8] qui portent sur le choix du participe passé et se placent donc immédiatement avant lui. Ce serait le cas des séquences en gras dans :

elles sont toujours **disons** euh nettoyées à fond (Birnie)
on les avait **entre guillemets** assainies (POI 96)

Les « interruptions » de cette nature peuvent aussi être nettement plus développées :

---

6 L'exemple réel comporte une bribe qui pourrait gêner le lecteur non averti :
ils ont été un **légèrement** modifiés depuis (Ciergerie)

7 Sur ce point voir Goes (1999). On n'a pas cherché ici à creuser cette limite entre valeur verbale et adjectivale du participe. Il faudrait par exemple s'intéresser aux variations selon que la partie verbale compte deux ou trois éléments :
j'avais été **très** touchée par ce livre (Masque)
? ça m'a **très** touché

8 A l'écrit on pourrait trouver d'autres termes d'approximation tels que :
De juillet 2000 à février 2001, tous les partisans de la société civile avaient été **comme** gouvernés par ce mot magique : le *montada* que l'on peut traduire par « forum ». (*Libération*. 25-06-01. p. 3).

je m'étais euh **comment dites-moi le mot je le trouve pas** spécialisée (FR-BEL)

Une séquence évaluative peut combiner plusieurs des éléments décrits auparavant :

c'est un décor qui est euh **très très disons** varié (POI 96)

2.2.2. Les aspectuels

Les éléments regroupés dans cette classe portent sur le déroulement interne du procès. Ils peuvent paraître proches de certains compléments temporels mais ils ne possèdent pas les mêmes propriétés syntaxiques :

de Gaulle a **encore** perdu un cheveu (Histoire)
quand a-t-il perdu un cheveu ? * encore
* c'est encore qu'il a perdu un cheveu

de Gaulle a **hier** perdu un cheveu
quand a-t-il perdu un cheveu ? hier
c'est hier qu'il a perdu un cheveu

Cette classe regroupe des adverbes tels que : *toujours, encore, déjà, d'abord...* On n'a pas rencontré de séquence aussi complexe que dans les rubriques précédentes (la négation, la séquence évaluative). La séquence se réduit en général à un seul terme :

j'ai **toujours** vécu au bord de mer (Lemonde)
la tribune d'orgue qui est **presque** terminée (POI 96)

Mais on peut envisager des séquences plus étoffées :

j'ai **presque toujours** été soigné à cet endroit (ex inventé)

Cette séquence peut s'envisager aussi bien en zone A (avant *été*) qu'en zone B (après *été*) :

il a **déjà** été condamné pour coups et blessures (Avocat)
je crois que le bar il a été **déjà** vendu (GRE)

On peut toutefois noter que dans les exemples attestés, la séquence aspectuelle est presque toujours placée avant la forme *été* :

ça fait qu'on a **toujours** été - - écrasées (Femmes)
nous avons **encore** été euh dispersés (Feux)

On n'a pas rencontré d'énoncé négatif avec une séquence aspectuelle en zone B (après *été*) :

le contrat n'avait pas été encore signé (ex inventé)

Dans les questions cet emplacement est peut-être plus envisageable :

C N R S ATILF NANCY

pourquoi n'a-t-il pas été encore soumis en conseil d'État (POI-01)
il a pas été **déjà** condamné ? (ex inventé)

Le tableau suivant rappelle la répartition des deux séquences examinées :

Auxiliaire	Zone A	été	Zone B	Participe
	pas			
			évaluatif	
a		été	bien	reçu
	aspectuel			
a	pas encore	été		reçu
	aspectuel		**évaluatif**	
ai	toujours		beaucoup	dessiné

2.3. Les séquences associées

Un certain nombre de séquences seront analysées comme des associés. Cela signifie qu'elles entretiennent un lien syntaxique « lâche » avec le verbe, ce que l'on peut mesurer à travers un ensemble de propriétés. Elles ne sont pas en relation de proportionnalité avec une proforme (a), ne peuvent pas être extraites (b), ni constituer une réponse satisfaisante à une question (c) :

je suis **d'ailleurs** allée à à la bibliothèque (Portuga)
(a) je suis d'ailleurs [= * *alors*] allée à la bibliothèque
(b) * *c'est* d'ailleurs *que* je suis allée à la bibliothèque
(c) *comment* (*quand*, ...) suis-je allée à la bibliothèque ? *d'ailleurs

Mais ces particularités caractérisent un grand nombre de séquences qui possèdent des valeurs nettement différentes. On travaillera, dans cette partie, avec des formes verbales à trois termes qui permettent de disposer des zones A et B déjà présentées auparavant. Au moins deux séries de séquences associées peuvent être dégagées : les modalisateurs et les associés « libres ».

2.3.1. Les modalisateurs

On regroupera ici des éléments qui traduisent la position du locuteur par rapport à ce qu'il exprime tels que : *sans doute, peut-être...* :

il avait **peut-être** bu ou il était peut- être pas bien (Police)
des débats vous ont **sans doute** donné d'ores et déjà + une idée (avocats)

Ces éléments semblent plutôt bloqués en zone A (avant *été*) :

elle aurait **peut-être** été rejetée (Alsace)

On dispose de peu d'attestations pour vérifier cette hypothèse mais dans les couples suivants, les exemples (b) dans lesquels les modalisateurs passent en zone B (après *été*) sont moins satisfaisants :

(a) elle a **peut-être** été mal soignée / (b) elle a été **peut-être** mal soignée
(a) il a **sans doute** été touché / (b) il a été **sans doute** touché

Cette réticence à accepter les exemples de la série (b) tient probablement au fait qu'il est plus naturel que la modalisation porte sur l'ensemble du prédicat (*été* + *participe*) que sur la seule composante lexicale portée par le participe, alors que le processus verbal est déjà entamé.

Une nouvelle fois, la présence d'une négation clarifie les faits. L'énoncé est meilleur quand le modalisateur est placé en zone A (avant *été*) :

les dépenses n'ont **peut-être** pas été réduites (ex inventé)
?? les dépenses n'ont pas été **peut-être** réduites

Réalisées en tête d'énoncé, un certain nombre de ces modalisateurs apparaissent sous une forme longue (prolongés par *que*) :

peut-être qu'il a été mal reçu par certaines personnes (FR-Val)

2.3.1. Les associés libres

Dans cette classe, on a regroupé toutes les séquences qui peuvent être réalisées aussi bien en zone A (avant *été*) qu'en zone B (après *été*) :

- en zone A

il a **d'ailleurs** été confirmé dans sa position (POI-01)
elle avait **donc** été réalisée (POI 96)
Diane Lanster qui a **depuis** été porté d'ailleurs sur le petit écran (Masque)

- en zone B

certains ont été **d'ailleurs** sanctifiés par l'intelligentsia (POI-01)
nous avons été **donc** euh emmenés (Feux)
ça a été **même** - pris en accident de travail (Hold-up)

Il s'agit là de séquences associées à la construction dans son ensemble (pour lesquelles aucun lien particulier avec un terme ne peut être établi), qui possèdent des valeurs sémantiques diverses. Ce type d'associé n'est pas limité à un mot :

or issu de la campagne euh je n'ai pas **comme ça aurait pu être naturel**
choisi une profession euh qui touche de près à la nature (FR)

On rencontre aussi des associés couplés à une place de la construction. C'est le cas des formes comme « tous » et des séquences composées avec « même » (*lui-même, elle-même...*), associées au sujet ou au complément, qu'on trouve aussi dans les deux zones :

on en a **tous** soufferts (Barbara)
dans les vœux qu'il nous a **à tous** adressés (F3. 6-01-02)
nous avons **nous aussi** été confrontés au dossier (POI-01)
il a été **lui-même** écartelé ou partagé sur le Proche-Orient (POI-01)

Le tableau de synthèse ci-dessous rappelle les différentes séquences d'associées identifiées. On a tenu compte de la négation car elle crée une partition de la zone A ce qui affine la description :

Auxiliaire	**Zone A**		*été*	**Zone B**	**Participe**
		pas			
	modalisateur			°	
a	peut-être		été		reçu
	ass. libre	**ass. libre**		**ass. libre**	
a	(d'ailleurs)	(d'ailleurs)	été	(d'ailleurs)	reçu

2.4. Les séquences régies

On peut trouver entre l'auxiliaire et le participe passé des séquences régies qui présentent un ensemble de propriétés que l'on rappelle brièvement :

- être en relation avec une proforme

  j'ai **tout de suite** été envoyé au grand séminaire (FélicienM)

  proformes : *alors, quand*

- possibilité d'extraction

  il paraît que vous avez **il y a quelques années** euh + fait un pacte (Casane)
  c'est il y a quelques années **que** vous avez fait un pacte

- être la réponse à une interrogation

  on avait **tout** ramassé (Sagape)
  qu'avait-on ramassé ? tout

Ces séquences régies sont sélectionnées par le verbe, ce qui explique certaines impossibilités ou difficultés :

ils les ont aimablement remis (Ski)
? je suis aimablement passionné

Quand la catégorie s'y prête, une restriction est possible sur la séquence régie :

je n'ai que **très rarement** vu un élève venir vers moi (FLE)

Les éléments régis appartiennent à des paradigmes qui peuvent être identifiés par des proformes et dont les membres se rencontrent dans des quantités très inégales :

- la série P1 est limitée à *tout* et *rien*

  on avait **tout** ramassé (Sagape)
  sinon on (n') aurait **rien** pu voir (Sagape)

Ces deux termes ne peuvent apparaître que si le verbe construit un complément direct :

* on était apparu *cela* / * on était *tout* apparu / * on n'était rien apparu

- la série des temporels (en relation avec la proforme *quand*) est la plus représentée

  il a été **par la suite** restauré (POI 96)
  une question que je me suis **pendant longtemps** posée (Portuga)

- la série en relation avec la proforme *comment*, pour l'essentiel, comporte quelques adverbes en *-ment*

  avec ma cousine on a **longuement** hésité (Ecogla)
  cet homme qui était **gentiment** parti aux toilettes (Attentat)

De nombreux adverbes en *-ment* se rattachent plutôt à la séquence évaluative qu'à la rection de type *comment*. On peut s'en convaincre en observant que l'extraction est rejetée :

il y en a qui ont été **tragiquement** blessés (Birnie) / * c'est tragiquement qu'ils ont été blessés
ils sont **terriblement** exposés (Birnie) / * c'est terriblement qu'ils sont exposés

Il existe aussi quelques autres compléments régis que l'on rencontre assez rarement :

bon là nous avons euh **avec un éducateur de prévention sur le quartier** euh - fait un projet d'information collective (FR- BEL)

Ces séquences régies peuvent se situer aussi bien en zone A (avant *été*) qu'en zone B (après *été*). C'est cette deuxième position qui se retrouve plus souvent dans nos données :

j'ai **tout de suite** été envoyé au grand séminaire (FélicienM)
ça n'a pas été **tout de suite** concrétisé (Musique)
ce qui a été - **une fois pour toutes** approuvé (Castag)

3. La combinatoire

On s'intéressera, dans cette partie, à des cumuls de séquences. Kjærsgaard (1995) propose une description en cinq paradigmes qui permettent de gérer les divers cas de cumul. On fournit ici un échantillon des paradigmes (le paradigme C2 se situe après le participe passé) :

a	A	B	C1	C2
cependant	certainement	aussitôt	assez	en 1975
d'ailleurs	évidemment	déjà	beaucoup	sûrement

A partir de ce tableau, il est possible de prévoir l'ordre de succession des termes dans un exemple tel que :

Pierre n'a cependant (a) sûrement (A) pas souvent (B) été bien (C1) compris de tous (cité par l'auteur)

Notre description sera moins précise que celle-ci, on ne descendra pas au niveau de l'ordre des termes eux-mêmes et on n'indiquera pas précisément le contenu des paradigmes. On s'en tiendra à la distribution des séquences identifiées telle qu'on peut la repérer dans les exemples oraux.

Avant de décrire l'ordre de réalisation des séquences, il faut regarder de plus près l'interaction entre la séquence négative et certaines séquences associées. Dans certains cas, la frontière devient difficile à tracer entre séquence négative et séquence associée. La relation possible ou discutable entre énoncé positif et énoncé négatif permettra de présenter ce problème. Si l'on ajoute une négation dans certains énoncés qui comportent un évaluatif, deux lectures sont envisageables :

je suis très bien intégrée (Portuga) / je suis pas très bien intégrée

L'une en fait une modalité de terme [*pas très bien*] qui porte plus précisément sur le segment qui suit (et peut être renforcée ou remplacée par *non*), l'autre en fait une modalité de relation [*pas*] qui touche l'ensemble des éléments qui gravitent autour du verbe. Dans l'exemple suivant, c'est la modalité de terme qui est activée comme le montre le contexte droit :

tu as pas vraiment été préparé à ça - même pas du tout (GRE)

Il est, dans de nombreux cas, difficile de se prononcer sur le découpage d'un énoncé car les éléments situés à droite de la négation peuvent être lus comme des associés ou comme des modifieurs du morphème négatif. Le cas simple est celui où chaque séquence se réalise dans une zone différente :

j'ai **pas** été **suffisamment** découragée pour décrocher (Avocat)
[séquence négative] + été + [séquence évaluative]

Mais lorsque les éléments se suivent, certains énoncés se prêtent à deux interprétations (a) une modalité de terme qui ne s'oppose pas à l'énoncé positif, (b) une modalité de relation qui s'oppose à l'énoncé positif :

il est pas vraiment préparé à ça
a) il est [pas vraiment] préparé à ça
b) il est [pas] [vraiment] préparé à ça / il est vraiment préparé à ça

Cette différence de statut a été discutée par Damourette et Pichon (1940) qui ont observé la portée[9] de la négation dans des enchaînements tels que « pas toujours / toujours pas », « pas encore / encore pas » (§ 2970 et suivants). Ils considèrent que le premier terme affecte le second :

« ainsi *toujours* amplecte *pas* s'il est situé avant lui, et est, au contraire, amplecté par lui si c'est *pas* qui figure le premier. » (§ 2971)

La description de l'ordre de réalisations des différentes séquences rencontrées dans les exemples oraux sera présentée d'abord dans les énoncés positifs puis dans les énoncés négatifs.

Le calcul de la combinatoire est facilité par la séquence évaluative qui se place devant le participe (en zone B). Les séquences régies et aspectuelles sont rarement réalisées simultanément, c'est pourquoi une seule colonne leur est attribué pour faciliter la lecture du tableau :

Auxiliaire zone A			*été*	**zone B**	**participe**
	Séq. associées	*Séq. régie/aspectuelle*		*Séq. évaluative*	
a		toujours	été	un peu	méprisé
sont	quand même	maintenant	-	disons	implantées
sommes	je le disais tout à l'heure		-	beaucoup plus	attachés
ont	même	certaines fois	-		refusé
est	via le matériel	déjà			limité
avait	d'ailleurs			beaucoup	frappé

[9] Ils parlent d'amplexion c'est-à-dire « action d'un adverbe sur le terme auquel il se rapporte » (Glossaire).

(1) le Panier a toujours été euh ah un peu méprisé (PEL25B)
(2) elles sont quand même maintenant disons implantées depuis (Birnie)
(3) et comme en France nous sommes je le disais tout à l'heure beaucoup plus attachés à la réglementation (CH14-07-99)
(4) puisque les détenus ont même certaines fois re- refusé d'aller en promenade (Beaumettes)
(5) on est via le matériel déjà limité à pas plus de vingt vingt postes (FR)
(6) ce qui m'avait d'ailleurs beaucoup frappé au Liban (FR)

Des empilements de séquences plus complexes sont rares mais on en rencontre à l'oral où des phénomènes d'anticipation lié au mode de production peuvent intervenir. L'exemple suivant :

et surtout j'ai toujours été toujours alors moi personnellement très très soutenue dans le travail que je faisais (Sourd-1)

comporte les séquences suivantes : un aspectuel (*toujours*) réalisé avant et après *été*[10] + un associé libre (*alors*) + un associé libre couplé au sujet (*moi personnellement*) + un évaluatif (*très très*).

Pour les énoncé négatifs, il faut être attentif à l'empan de la séquence négative (la séquence régie n'a pas été retenue pour les exemples décrits) :

Auxiliaire	**zone A**		***été***	**zone B**	**participe**
	Séq. associé	*Séq négative*		*Séq évaluative*	
ai		pas	été	suffisamment	découragée
est		pas encore	-		venu
était	peut-être	pas forcément	-	très	fourni
est	encore	pas	-	totalement	aménagé
ai	pratiquement	rien	-		visité

(1) j'ai pas été suffisamment découragée pour décrocher (Avocat)
(2) il est pas encore venu en France (Avocat)
(3) le vocabulaire n'était peut-être pas forcément très fourni (Souvenirs)
(4) il est encore pas totalement aménagé (FR - BEL)
(5) j'ai pratiquement rien visité de l'Inde (FR- Voyages)[11]

Dans un exemple qui présente une relation satisfaisante entre énoncé positif / négatif comme :

ce n'est pas toujours réussi / c'est toujours réussi

---

[10] Cette réitération peut être vue comme un effet d'insistance. Mais on peut aussi considérer que la première occurrence est une anticipation.

[11] Un exemple comme :

j'ai pratiquement tout visité de l'Inde

permet de considérer que « pratiquement rien » ne constitue pas une séquence négative.

on identifiera [*pas toujours*] comme une séquence aspectuelle.

4. Conclusion

On aurait pu craindre que l'oral ne fournisse que peu de données exploitables soit parce que la succession entre auxiliaire et participe passé n'était pas rompue, soit parce que les insertions correspondaient à des « accidents » dans la production. Tel n'est pas le cas. Les faits recueillis ont permis, comme espéré, d'observer une grande richesse et une grande variété dans les séquences réalisées, de dresser un premier classement sur des bases syntaxiques et d'amorcer la description des cumuls de séquences.

Références

Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde, K. (1990). *Le français parlé. Études grammaticales*, Paris : Éditions du CNRS.

Blinkenberg, A. (1928). *L'ordre des mots en français moderne.* Copenhague : Munskgaard.

de Boer, C. (1954). *Syntaxe du français moderne,* Leiden : Universitaire Pers Leiden (2ème édition).

Brunot, F. (1905-1953). *Histoire de la langue française des origines à nos jours.* Paris : Colin.

Damourette J. & Pichon E. (1911-1940). *Des mots à la pensée : Essai de grammaire de la langue française.* Paris: D'Artrey, 7 tomes.

Kjærsgaard, P. (1995). *Fransk Grammatik I Hovedtræk.* Odense : Odense Universitetsforlag.

Larrivée, P. (2001). *L'interprétation des séquences négatives. Portée et foyer des négations en français.* Bruxelles : Duculot.

Marchello-Nizia, C. (1997). *La langue française aux XIVe et XVe siècles.* Paris : Nathan.

Martinon, P. (1927). *Comment on parle en français.* Paris : Larousse.

Nøjgaard, M. (1992). *Les adverbes français. Essai de description fonctionnelle.* Copenhague : Munksgaard.

Nølke, H. (2001). *Le regard du locuteur.* Paris : Kimé.

Oudin, A. (1640). *Grammaire française rapportée au langage du Temps.* Paris : Antoine de Sommaville.

Note à propos de la construction dite « du datif partitif »

André VALLI
Université de Provence

Résumé

Dans la littérature importante consacrée à l'examen des constructions datives du français qui mettent en jeu des noms « de parties du corps » et plus largement des noms qui relèvent de la possession inaliénable, des analyses assez différentes, parfois divergentes, ont été développées. Deux questions apparaissent au centre de l'intérêt :l'existence de 'sources' lexicales de ce datif pronominal et l'extension de la construction à des noms qui ne relèveraient pas de la possession inaliénable

Dans le passage en revue des études sélectionnées, j'essaie de montrer que la discussion de ces deux points soulève le délicat problème des données à retenir pour décrire un usage du français, observable dans les corpus de français parlé et de français écrit littéraire, qui ne coïncide pas avec l'usage puriste. Le fait de retenir ou de ne pas retenir telles ou telles données a des conséquences dans la description syntaxique de la question abordée : l'examen de corpus de langue parlée et de français écrit littéraire montre que, sur les points en jeu, l'usage décrit, en écart par rapport à l'usage puriste noté par les grammairiens, est conforme à l'usage qui était régulier dans les textes d'ancien et de moyen français.

1. Introduction

Dans la littérature volumineuse consacrée à l'examen des constructions datives du français qui mettent en jeu des noms de parties du corps et plus largement des noms qui relèvent de la possession inaliénable (Leclère, 1976, 1978 ; Spanoghe, 1995 ; Mélis, 1996 ; Lamiroy, 2001 ; Riegel *et al.* ; 1994, etc.) :

Recherches sur le français parlé, 17, 29-40.
©2002, *Université de Provence*

(1) Je lui ai serré la main

des analyses assez différentes, parfois divergentes, ont été développées, dont témoigne partiellement la multiplicité des étiquettes proposées pour rendre compte du phénomène : datif étendu, datif possessif, datif partitif, datif non lexical.

Deux questions apparaissent au centre de l'intérêt :

- l'extension de la construction à des noms qui ne relèveraient pas de la possession inaliénable (Herslund, 1983, 1988 ; Barnes, 1985 ; Herschensohn, 1992 ; Lamiroy, 2001) :

  (2) On lui a tué son chat

- l'existence problématique de « sources » lexicales de ce datif pronominal et de leur statut grammatical (Leclère, 1976 ; Olsson, 1984 : Barnes, 1985 ; Roorick, 1988 ; Herslund, 1983, 1988 ; Herschensohn, 1992 ; Spanoghe, 1995). L'acceptabilité d'un énoncé comme (3) est en débat :

  (3) On lui a secoué les puces à Paul

La discussion de ces deux points soulève le délicat problème des données à retenir pour décrire un usage du français observable dans les corpus de français parlé et de français écrit littéraire, qui ne coïncide pas avec l'usage puriste. Ce problème des données se présente sous deux aspects : la faiblesse — toute relative — des données des corpus oraux : le corpus G.A.R.S.– DELIC, dans l'état actuel, ne permet pas d'étudier dans le détail toutes les questions de syntaxe, en particulier cette construction du datif partitif ; sur le second point étudié, par exemple, les exemples d'un usage que nous pensons régulièrement attesté sont peu nombreux. En revanche, on relève des exemples de la construction en nombre important dans les différents corpus littéraires — la base de données Frantext et les corpus de textes romanesques contemporains — qui mettent en scène des personnages de la vie ordinaire. Si on ne perd pas de vue le caractère problématique du statut de telles données — on ne sait jamais vraiment si c'est la langue de l'auteur qui est notée ou bien celle qu'il prête à ses personnages, représentation plus ou moins manipulée langue de la langue parlée, force est d'admettre l'intérêt de telles données pour l'analyse linguistique.

Mon objectif dans cette note sera d'essayer de montrer que le fait de retenir ou de ne pas retenir telles ou telles données a des conséquences dans la description syntaxique de la question abordée : l'examen de corpus de langue parlée et de français écrit littéraire montre que, sur les points en jeu,

l'usage décrit, en écart par rapport à l'usage puriste noté par les grammairiens, est conforme à l'usage qui était régulier dans les textes d'ancien et de moyen français (Schøsler, 2001).

2. L'extension de la construction datif partitif à des noms qui ne relèvent pas de la possession inaliénable

La lecture des articles et ouvrages sur la question nous permet de distinguer deux types d'attitudes devant les données.

C'est sans doute Lamiroy (2001) qui adopte la position la plus puriste : tout comme dans l'étude de Spanoghe (1995), la relation datif *vs* génitif, qui caractérise en propre cette construction d'un datif partitif dit de type possessif ne s'observerait en français, selon elle, que dans la condition de stricte inaliénabilité :

(3a) Le médecin leur a examiné la gorge
(3b) ? le médecin a examiné leur gorge
(3c) Le policier a minutieusement examiné leurs passeports
(3d) ?* Le policier leur a minutieusement examiné le passeport.(: 84)

C'est pourquoi, selon Lamiroy (2001), contre l'évidence, un énoncé comme (3d), qui introduit dans cette construction dative un nom complément qui n'est ni un nom de partie du corps ni un nom qui relève de la possession inaliénable doit être est considéré comme inacceptable. Ce qui la conduit à poser que le français se montrerait plus restrictif que les autres langues romanes quant à l'emploi du datif qu'elle nomme, comme Spanoghe (1995), non lexical.

Pourtant dans un article déjà ancien, Leclère (1978) se montrait ouvert à l'extension de la construction, mais en fixant les limites de la relation entre la construction dative et la construction génitive :

(4a) Il me prend le bras
(4b) Il prend mon bras
(4c) * Paul lui a déchiré le sac
(4d) Paul lui a déchiré son sac
(4^{c}) Paul a déchiré son sac

Le caractère inacceptable de (4c), selon Leclère, résulterait seulement de l'emploi d'un déterminant défini ; l'acceptabilité de (4d) en revanche ne paraît pas être mise en doute.

C'est certainement Herslund (1983, 1988) qui se montre le plus ouvert quand il note que le syntagme nominal dont dépend le pronom datif dans les constructions du datif partitif ne dénote pas uniquement une partie du

corps, des capacités ou des états physiques ou moraux, mais peut aussi désigner des vêtements ou d'autres accessoires. Il en relève des exemples réguliers dans un corpus de textes littéraires français :

(5a) La moitié de l'humanité lui lèche les bottes. (Gary. Promesse. 18)
(5b) Elle me reconnaissait/ Elle est venue me flairer les souliers. (Courchay Malavel 30)
(5c) Il croit peut-être que c'est pour lui cirer les bottes ou rechercher ses fusils (Charles-Roux Elle 465)
(5d) C'est là qu'il continue depuis quelques années à massacrer les veuves de généraux et les orphelins d'ecclésiastiques qui lui tombent sous le crayon (à propos de Siné, texte publicitaire dans Wolinsky Lettre, i.f.)

On relève des exemples de cette extension du datif possessif dans notre corpus de français parlé ; en premier lieu, avec des noms de vêtement ou accessoire vestimentaire :

(6a) ils passaient ils lui arrachaient le ils lui prenaient le chapeau (sca59ab 15, 9)

(6b) tu peux pas marcher parce que tu te brûles les semelles (voyages 0 84, 3)

Cet emploi du datif s'observe également avec des noms d'institution humaine, le mariage, comme dans cet exemple :

(7) Monseigneur X il est allé lui bénir le mariage (gua15ab 13, 8)

Je citerai également l'exemple suivant, qui contient un nom d'une partie du corps mais qui relance une discussion à propos d'une autre restriction qui accompagnerait le verbe :

(8) puis ma marraine – elle soulève la robe et puis dès qu'on lui voit la cuisse et que l'élastique est à la cuisse on arrête on arrête la jarretière (pug2 29, 6)

L'hypothèse développée par Lamiroy (2001) concerne le type de verbe qui apparaîtrait de façon prototypique dans la structure analysée. Pour des raisons évidentes, selon Lamiroy qui reprend une interprétation assez consensuelle de la structure dative :

> « pour que le personnage indiqué par le datif puisse être d'une façon ou d'une autre impliqué, il faut que le verbe exprime une action susceptible d'affecter le possesseur » (2001 : 85).

Elle constate toutefois que parmi les langues romanes, l'espagnol relâcherait, ce sont ses propres termes, cette exigence puisque, à la différence du français, des verbes statiques ne seraient pas exclus dans les structures à datif possessif :

(9) F a. Je lui ai (maquillé + cassé + ?* vu) la figure
E b. Le he (pintado + roto + visto) la cara (Lamiroy 2001 : 89)

J. Herschensohn (1992) discute la même hypothèse :

« The dative clitics usually possess characteristics of animacy and affectedness ; the generally nonstative verbs imply intentional action by the subject « (1992 : 125).

Comme Rooryck (1988), à qui elle emprunte l'exemple suivant, qu'elle considère comme un énoncé acceptable, elle note que les verbes statifs comme voir peuvent entrer dans cette construction :

(10) Je lui ai vu une jupe bizarre (p. 125)

Il me semble que l'exemple (8) extrait d'un récit oral montre mieux que (10) qu'il est sans doute nécessaire de préciser la contrainte sémantique sur les verbes entrant dans cette structure dative : la confrontation de (8) et (9), indique qu'un verbe d'aspect statif est compatible dans la structure selon l'aspect verbal, ponctuel, répétitif, sous lequel on l'envisage.

Il est donc important de noter qu'on rencontre dans notre corpus de français parlé des exemples relativement proches de ceux relevés par Herslund dans un corpus littéraire. Il est clair que, dans certaines de ses variétés, le français contemporain se montre moins restrictif dans l'emploi de cette structure dative que ne le soutient Lamiroy et se rapproche ainsi des variétés standard d'autres langues romanes.

Bien entendu, ce bilan est à rapprocher des résultats de l'étude conduite par Schøsler (2001) sur la distribution de cette structure dans un corpus de textes d'ancien et de moyen français.

Elle montre en effet que dans l'ancienne langue entrent dans cette structure dative des noms de parties du corps, qui relèvent typiquement de la possession inaliénable mais également des noms de choses qui n'entrent pas dans la sphère personnelle. Elle donne différents exemples de cette construction que Palsgrave présentait comme une caractéristique du français par opposition à la langue anglaise.

Le nom possédé prototypique selon Schøsler est un nom de vêtement ou du vocabulaire de l'armement :

(11) Dam Clins li fils Carduit le fiert en trespassant
si que l'elme li trencha e le cervele espand

[Dans Clins le frappe.../ tant qu'il lui coupa le heaume et répand le cerveau...Alia 1154] (Schøsler 2001 : 104)

Le nom possédé peut être moins prototypique, et ne représenter qu'un élément d'un ensemble où évolue la personne :

(12) Il se feri en l aigue jusc al col del cheval
Pui a traite l espee lor ancre lor tencha (Elie, 2319)

[il se jeta à l'eau...puis a pris l'épée et leur trancha leur ancre] (Schøsler 2001 : 104)

Comme on peut l'observer en comparant (11) et (12), on rencontre dans cette structure dative un déterminant défini (11) ou aussi bien un déterminant possessif (12) que Palgrave considérait comme redondant.

Schøsler adopte donc une attitude plutôt puriste quand elle note (p. 104) que ces constructions sont exclues du français standard moderne. Elle s'appuie, il faut le reconnaître sur une position traditionnellement exprimée par les grammairiens normatifs. C'est la norme que reprend Riegel *et al.* :

> « le datif de la totalité impliquée (ou partitif ou encore de la possession inaliénable) identifie le tout lorsqu'une de ses parties fonctionne comme objet direct ou indirect, avec à l'appui des exemples ne comprenant que des noms de parties du corps » (1994 : 227).

Au contraire, sur la base des éléments que nous avons réunis ici, on peut considérer que c'est seulement l'usage normatif qui restreint la structure du datif possessif à l'expression de la possession inaliénable ; dans l'usage littéraire ou dans la conversation, le français se montre moins restrictif dans l'emploi de cette structure comme en fait foi l'état de la question que dresse Mélis (1996).

3. La question des sources lexicales du datif partitif

La position sur ce sujet exprimée par Leclère dont l'étude est pourtant largement citée dans les articles et ouvrages que nous avons passé en revue ne fait pas l'objet d'un total consensus. Ce complément lui serait bien datif et non un complément du nom accusatif, mais il n'a pas un statut d'argument puis qu'il n'est pas sous-catégorisé par le verbe. D'autre part, il existerait selon Leclère une relation de « proportionnalité entre le clitique et une forme 'à N' » comme dans :

(13a) Paul lui a cassé le bras
(13b) Paul a cassé[le bras à Marie]
(13c) Paul a cassé le bras [à Marie]

Cette relation, s'établirait non pas avec le complément de noms (13b), mais avec un complément de verbe (13c). La position Blanche-Benveniste *et al.* (1984) est plus circonspecte. Ces auteurs voient dans un énoncé comme (13a) l'intervention d'un clitique dépendant de la présence du

complément valenciel direct, de lexicalisation restreinte : « il se réalise plus facilement comme pronom indéfini » :

(14) Je n'ai jamais cassé la jambe à personne

« que comme élément lexical avec préposition *à* » :

(15) ? j'ai cassé la jambe à Paul. (p. 78)

Dans son ouvrage consacré à la notion de datif en français, Herslund (1988) présente la structure du datif partitif du point de vue de la possession inaliénable, qui trouverait en français une expression aussi bien adverbale (*lui* + V) qu'adnominale (*son* + N). On trouverait selon lui dans ce domaine une certaine interférence entre datifs et génitifs. C'est pourquoi les deux constructions seraient évidemment à mettre en rapport avec des syntagmes nominaux complexes, *en de* ou *à*. Ainsi ces deux exemples que j'emprunte son l'étude :

(16) Le magistrat s'était contenté de serrer la main de Maigret. (Simenon Témoins 29) (cité par Herslund 1988 : 252)

(17) C'est que je veux changer les idées au Petit. (Courchay Malaveil 224) (cité Herslund 1988 : 253)

Si le statut adnominal des exemples comme (16) ne fait pas de doute à ses yeux, en revanche, les exemples comme (17), comportant la préposition *à*, ne lui paraissent pas aussi univoques. A la différence de Blanche-Benveniste *et al.*, il ne met pas en doute le caractère naturel d'énoncés comme (17) à structure dative à-NP, c'est leur statut grammatical qui fait difficulté :

> « en fait ces phrases ont l'allure de phrases trivalentes à OI lexical : on se trouve peut-être ici à mi-chemin entre une construction adnominale et une construction adverbale » (p. 253).

Cette dernière observation ne revêt toutefois dans l'étude de Herslund qu'un intérêt secondaire. Au contraire, ce qui apparaît au premier plan à propos de la structure, c'est :

> « l'idée que le datif partitif est l'effet de l'éclatement d'un syntagme nominal » (p. 253).

Herslund s'interroge sur les conditions de cet éclatement et sur le rapport entre la structure possessive et la structure dative résultante ; je résume sa démarche : Il voit dans (18 a)

(18) a. ... lui V ... (P) N...
b. ... V ... (P) son N ...

une structure partitive, où *lui*, qui dénote le tout, est extrait d'un syntagme nominal (18b) qui en dénote une partie. Comme on l'observe également dans l'éclatement de certains syntagmes prépositionnels,

(19) a. ... V ... sur N
b. ... lui V ...dessus

en faisant du régime de la préposition un dépendant du verbe, on l'intègre dans la sphère sémantique de celui-ci. En (18), le constituant qui dénote le tout devient membre primaire de la phrase, dépendant directement du verbe.

On pourrait faire le même commentaire à propos des articles que Herschensohn (1992) et Rooryck (1988) consacrent à la structure datif partitif. Le premier auteur ne manifeste aucun intérêt au phénomène de lexicalisation de cette forme de datif. Elle se borne à mettre en doute, à tort me semble-t-il, l'acceptabilité de l'exemple suivant de double marquage du complément datif, qu'elle emprunte à l'étude de Barnes :

(20) Le chiot lui a pissé dans ses laitues (?* à Paul) (Barnes 1985 : 154)

Au contraire, tout l'effort tend vers l'étude des sources de ce datif. La position est proche de celle de Herslund :

« I now will argue that they are cliticised from an NP or PP complement of the verb » (p. 128).

Suit une argumentation en quatre points pour faire la démonstration que les datifs « non lexicaux » doivent être co-indexés avec une catégorie vide dans les NP ou PP complément du verbe, qui aboutit à une évidence : ces datifs « non lexicaux » n'ont pas les mêmes propriétés que les datifs lexicaux. En particulier, ils ne peuvent figurer que dans une structure qui ne comprend pas le complément direct, en relation de solidarité avec l'OI datif :

(21a) Le chiot a pissé.
(21b) Le chiot lui a pissé dans ses laitues.*
(21c) * Le chiot lui a pissé. (Herschensohn : 129)

La position de Rooryck est plus ambiguë. Il note que Leclère, Bernes, d'une certaine façon, Herslund et surtout Olsson (1984) envisagent le datif non lexical comme un substitut pour un complément à-NP et malgré l'évidente acceptabilité d'énoncés comme :

(22) J'ai toujours vu à mon père une casquette, souvent à carreaux...(F. F. Bastide, La vie rêvée, p. 73, cité par Olsson1984 : 199)

il soutient que la définition du datif non lexical comme une pronom clitique qui peut être difficilement lexicalisé n'est pas mise en difficulté. Il

développe trois arguments pour montrer le caractère fondamentalement clitique de cette structure, et son statut marginal quand il est lexicalisé. L'analyse quantitative d'Olsson, qui montre, à partir d'un corpus littéraire, que le datif non lexical est beaucoup plus fréquent sous la forme d'un pronom que sous la forme d'un groupe prépositionnel lexicalisé, fournirait le premier argument : cette évaluation tendrait à prouver que la construction lexicalisée est marginalisée.

L'examen de critères formels apporterait le second argument : mais, comme dans l'étude de Herschnesohn, l'examen n'aboutit qu'à dégager des tests syntaxiques qui distinguent les datifs lexicaux et non lexicaux des verbes transitifs.

En troisième lieu, Rooryck s'appuie sur l'étude de Barnes pour discuter l'acceptabilité de ce qu'elle nomme à-NP non lexical ; il a recours à l'argument de l'analogie sémantique pour justifier l'inacceptabilité relative d'énoncés comme (23) :

(23) Elle lui a tiré dans le ventre. / * Elle a tiré dans le ventre à Jean (Rooryck, 1988 : 98)

Celle de (24) serait expliquée par une confusion avec la lecture adnominale de à-NP, qui appartiendrait au français populaire :

(24) ? On a cassé le bras à Jean (Barnes, 1985 : 34c)

En revanche, Rooryck n'a aucune explication pour rendre compte de l'acceptabilité de (22) que personne ne peut contester.

Je terminerai ce passage en revue des auteurs qui se sont intéressés à la question des sources du datif partitif par l'ouvrage de Spanoghe qui développe à propos de la structure de cette construction une position assez proche de celle de Herslund. Son analyse est que le datif dans ces exemples ne relève pas de la valence verbale mais de la valence d'un constituant nominal, ou plutôt, d'une solidarité de co-sélection entre O et 0I, ce qui rejoint l'analyse en prédication seconde, selon les termes de Herslund. Ce qui l'intéresse avant tout c'est l'examen de la relation entre la structure à déterminant possessif et la structure dative :

(25) On voyait ses côtes / On lui voyait les côtes (Spanoghe : 99)

et elle refuse de prendre en compte les énoncés dans lesquels l'OI non lexical prend une forme nominale :

(26) ça c'est la dernière phrase pour clouer le bec à Pitchoun (Cau 16)
(27) Pendant qu'il coupe les cheveux à Nanou, Monsieur Marcel lui passe la main sur la quéqette (Cau 132)

mais elle se montre intriguée par cet exemple qu'elle emprunte à une étude de Morin(1980), également cité par Barnes et Rooryck :

(23) On lui a tiré dans le ventre à ce pauvre garçon

qui pourrait être analysé comme un datif lexical.

Les données du français ancien et du français parlé de conversation consultables à partir des corpus qui sont à notre disposition apportent des éléments intéressants pour la description de cette construction du datif partitif qui apparaît quelque peu embrouillée quand on l'aborde avec un point de vue normatif synchronique.

Les exemples de cette structure qui sont présentés dans l'article de Schøsler (2001) attestent qu'en français médiéval le datif possessif est pronominal et très régulièrement lexical, quelque soit la nature de la relation de possession :

(29) si vus mes humes estes/ e obeir me vulez/ les chiefs as fiz vostre seigneur colpez (... si vous êtes mon vassal et voulez m'obéir vous coupez la tête aux fils de votre seigneur (Reis : 383343)

(30) s'il prend la terre au petit berengier (Charrois, ms B1)

Ce sentiment d'évidence quant au statut du datif partitif est confirmé par un énoncé comme (29) qui rend manifeste la relation de lexicalisation de la structure par le double marquage du complément :

(31) ...et quant il les orent pris il lor coperent les testes a toz (... ils leur coupèrent la tête à tous) (beauv, 49189)

Le français parlé de conversation offre de tels exemples de double marquage d'un datif partitif, à la fois pronominal et lexical

(32) je me souviens pas de lui avoir sauté au cou à ma mère (femmes)

comme il offre des exemples aussi réguliers de double marquage d'un complément natif « argumental » :

(33) je ne crois pas que je lui trouvais un défaut à ma mère (femmes)

4. Conclusion

On admettra donc que, selon les données retenues, le français contemporain se montre ou non assez différent du français médiéval dans la grammaticalisation de la possession au moyen d'une structure dative.

Sur le premier point étudié, nous avons vu que le français contemporain ne se montre restrictif dans l'emploi qu'il fait du datif partitif que dans sa variété normative : le français parlé de conversation prolonge

l'usage d'employer dans cette structure des noms qui ne relèvent pas de la possession inaliénable qui était régulier en français médiéval.

Sur le second point, la note que Riegel *et al.* consacrent dans leur grammaire au datif de la totalité impliquée (ou partitif ou encore de la possession inaliénable) résume assez honnêtement l'embarras des linguistes à propos de cette construction puisque la lexicalisation de ce datif pronominal y est envisagée, mais marginalement, comme le montre la présentation typographique, et par surcroît, de manière ambiguë. En effet, dans les exemples cités :

(34) Pierre lui serre la main (Pierre serre la main à quelqu'un)
Pierre lui saute au cou (Pierre saute au cou à/de quelqu'un)

la lexicalisation du clitique datif n'est envisagée que de façon restreinte au moyen d'un complément dont le statut hésite entre celui de complément de verbe ou de complément de nom.

Je pense que les nombreux exemples de lexicalisation de ce datif pronominal dans les corpus de français littéraire – celui d'Olsson et celui d'Herslund – ainsi que ceux qu'on relève dans les corpus de français médiéval et de français parlé contemporain devraient inciter les chercheurs à cesser de considérer ce phénomène de lexicalisation d'une structure qui serait avant tout pronominale comme secondaire ou marginal. Ce qui devrait les conduire à un réexamen du statut de ce complément à-NP non-argumental, en relation de co-sélection avec le complément accusatif.

Références

Barnes, B. (1985). A Functional Explanation of French Nonlexical Datives. *Studies in Language 9*, 159-195.

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J. & Van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au français.* Paris : SELAF. AELIA, CNRS.

Hershensohn, J. (1992). On the Economy of Romance Nonlexical Datives. In P. Hirscbülher & K. Koerner (eds.). *Romance Languages and Moderne Linguistic Theory* (pp. 123-135). Amsterdam : John Banjamins.

Herslund, M. (1988). *Le datif en français.* Paris : Bibliothèque de l'Information Grammaticale.

Labov, W. (2001). Qu'est-ce qu'un fait linguistique. *Marges linguistiques, 1.* [Traduction française de la version originale anglaise publiée chez l'éditeur Lisse-The Peter de Ridderd Press, sous le titre : « What is a linguistic fact ? »]

Lamiroy, B. (2001). Grammaticalisation et possession inaliénable dans les langues romanes et germaniques. In L. Schosler (éd.). *La valence, perspectives romanes et diachroniques* (pp. 82-98). Stuttgart : Frantz Steiner Verlag.

Leclère, C. (1976). Datifs syntaxiques et datifs éthiques. In G. Gross & J-C Chevalier (éds.), *Méthodes en grammaire française* (pp. 73-96). Paris : Klincksieck.

Leclère, C. (1978). Sur une classe de verbes datifs. *Langue française, 39*, 66-76.

Mélis, L. (1996). The dative in Moderne French. In W. Van Bell and W. Van Langendonck (eds.). *The dative. Descripitive Studies, Vol 1.* Amsterdam : John Banjamins.

Olsson, H. (1984). La construction Je lui trouve mauvaise mine en français moderne. *Studia Neophilologica, 56*, 183-213.

Riegel. M., Pellat, J.-C., & Rioul, R. (1994). *Grammaire méthodique du français.* Paris : Presses Universitaires de France (P.U.F.).

Roorick, J. (1988). Critères formels pour le datif non-lexical en français. *Studia Neophilologica 60*, 97-107.

Schøsler, L. (2001). La valence verbale dans une perspective diachronique : quelques problèmes méthodologiques. In L. Schosler (ed.). *La valence, perspectives romanes et diachroniques* (pp. 98-113). Stuttgart : Frantz Steiner Verlag.

Spanoghe, A.-M. (1995). *La syntaxe de l'appartenance inaliénable en français, en espagnol et en portugais.* Frankfurt am Main : Peter Lang.

Détachement et linéarité[1]

Mylène BLASCO-DULBECCO* & Sandrine CADDÉO**

*Université de Clermont II
**Université de Provence

Résumé

Les constructions qui font intervenir un pronom tonique, un pronom clitique et un syntagme nominal sur le modèle *lui, mon fils, il est d'accord* pose un problème d'analyse intéressant lorsque le groupe nominal est disjoint des deux autres : *lui, il est d'accord, mon fils*. Certaines contraintes qui concernent la classe lexicale du nom et sa détermination permettent d'écarter la thèse qui consisterait à y voir un phénomène de disjonction d'une construction contiguë. Il s'agit d'une construction originale dont les propriétés dépendent pour une grande part du type de pronom en jeu et de leur fonction. C'est également un bon exemple de dépendance s'établissant « à grande distance ».

1. Introduction

On caractérise les dislocations par la cooccurrence d'un pronom clitique et d'un élément coréférentiel selon deux modèles, soit à gauche du groupe verbal *le voisin, il est d'accord*, soit à sa droite, *il est d'accord, le voisin*. La prise en compte de la position des éléments par rapport au verbe recteur est essentielle pour la description. Les catégories en présence modifient également les analyses. Il ne s'agit pas tout à fait du même phénomène de dislocation avec des pronoms toniques (ou *non-clitiques*), de la série *lui : lui, il est d'accord / il est d'accord, lui*, qui, contrairement aux

[1] Nous remercions Claire Blanche-Benveniste pour sa lecture attentive et ses précieux conseils.

©2002, Université de Provence

lexèmes, se déplacent très librement autour du verbe (Blasco-Dulbecco, 1999).

Le schéma plus complexe, qui fait intervenir trois éléments, pronom tonique, pronom clitique et lexique, sur le modèle *il est d'accord, lui, le voisin/ le voisin, lui, il est d'accord*, est plus éclairant que les schémas plus simples à deux éléments. D'une part, il permet d'apporter des retouches à certaines règles de dislocation et d'apposition (Blasco-Dulbecco & Caddéo, 2001), d'autre part, il pose le problème de l'ordre des mots en d'autres termes que celui de la mobilité.

Lorsqu'il y a cooccurrence des trois catégories précitées, nous avons recensé six dispositions possibles des éléments :

1.	Lui	le N		il V			
2.	Lui			il V		le N	
3.		Le N	lui	il V			
4.		Le N		il V	lui		
5.				il V	lui	le N	
6.				il V		le N	lui

Pour raisonner sur les différentes dispositions linéaires que peuvent prendre ces éléments, il est nécessaire de faire les hypothèses suivantes :

- Certaines dispositions se calculent en « zones » successives, dont le point de référence est le verbe, par exemple verbale, préverbale, post verbale sur le modèle de skårup (1975), ou le nom. Les pronoms toniques de la série *lui* ont deux comportements différents, selon qu'ils sont préposés ou postposés à un SN lexical. Postposé à un SN lexical, *lui* a le trait [+individuel] : *le bureau, lui...* Préposé à un SN lexical, *lui* a le trait [+humain] : *lui, le voisin...* (Blanche-Benveniste *et al.*, 1987 : 49). Dans ce cas, il faudrait même parler de deux constructions bien distinctes (Caddéo, en préparation).
- D'autres dispositions répondent à des suites « contiguës/disjointes », avec des relations « à longue distance » entre par exemple, le pronom tonique et le SN lexical (disposition 2), celui-ci a des restrictions fortes sur les prédéterminants.

Nous souhaitons nous arrêter, dans ce travail, sur le cas particulier des séquences de type :

(1) est-ce qu'il a des draps mous ou durs parce que **lui il** est très maniaque hein **mon fils** mais il faut pas du tout lui céder (Pavray, 40, 3-6)

Dans ces constructions, il y a encore cooccurrence des trois éléments ; le syntagme nominal est détaché après le verbe et il entretient des relations référentielles avec le pronom tonique ou clitique - sujet ou objet - dont il est très éloigné (disposition 2 du tableau)[2].

D'un point de vue sémantique, il peut être tentant d'y voir une sorte de correction qui consisterait à compléter la séquence avec le pronom tonique dont la réalisation contiguë serait *lui, mon fils, il...* D'ailleurs, Apotheloz & Zay (1999) nomment ce phénomène « incidents de la programmation syntagmatique » :

(1. a) parce que **lui il** est très maniaque hein **mon fils**

lui **il** est très maniaque

mon fils

Mais ajoutons que lorsque le syntagme nominal est distant, il s'exerce un certain nombre de contraintes qui ne se retrouvent pas dans le cas de contiguïté. On peut parler d'un modèle syntaxique très rigoureux applicable à certains syntagmes nominaux uniquement, avec certains pronoms toniques seulement et dans un domaine lexical restreint. Il serait intéressant d'établir une comparaison avec les constructions du type *il... mon fils/ lui... mon fils*, mais nous privilégierons la comparaison avec les cas de contiguïté à trois éléments du type, *lui mon fils il...* Il s'agit, à partir de l'observation des données, d'ouvrir une discussion sur les problèmes posés par ce type de structures qui font intervenir une dimension rarement prise en compte, celle de « la distance syntaxique » (contiguïté/éloignement) entre les constituants de ce modèle.

2. Tendances distributionnelles

Dans la séquence *tonique+SN+clitique*, le syntagme nominal se laisse analyser comme un élément apposé de par la relation de dépendance qu'il

2 Ces exemples semblent caractéristiques du français parlé. Nous avons interrogé un corpus de presse qui représente dix années du Monde diplomatique et une vingtaine d'œuvres littéraires du XIXème et XXème siècles sans relever un seul exemple de ce type.

entretient avec le pronom tonique. Cette interprétation relève de deux constats :

- dans cette position, le tonique ne sélectionne que des noms ayant le trait [+humain]. C'est particulièrement intéressant à noter pour les pronoms anaphoriques de la série *lui*. Il est en effet impossible de rencontrer

  (2) ?**lui** *le bureau* **il** sera poussé contre le mur

- l'absence du tonique oblige à repenser le statut du syntagme nominal[3] qui apparaît alors comme un élément disloqué, et ce, quelle que soit sa position par rapport au verbe (Blasco-Dulbecco & Caddéo, 2001)

  - zone préverbale

    (3) **les savants ils** disent des choses vraies des fois et des fois ils disent des choses fausses (VFT 14)
    (4) **les filles on** voulait faire les gendarmes (Pel, 9, 10)
    (5) **les aînés nous** y sommes pas allés (Scalone, 2, 9)

  - zone post verbale

    (6) **ils** sont toujours là **les profs** toujours derrière (Boavi, 86,6)
    (7) **on** est vraiment une race bizarre **les chrétiens** (Caddéo, adulte, conversation)
    (8) **nous** étions assez vieux déjà **mon mari et moi** (SCA, 56)

Les effets sont cependant différents entre (3), (6) d'un côté et (4), (5), (7) (8) de l'autre. Le caractère complexe de *on* et *nous* oblige à faire des calculs inférentiels absents des cas avec *ils*. Ces particularités se maintiennent dans les exemples avec pronoms toniques. Les observations devront donc être mesurées en fonction des types de pronoms. La confrontation de *moi, toi, nous, vous, lui, elle, eux, elles* conduit à les présenter séparément pour ce qui est des contraintes observées sur la détermination du lexème et des fonctions syntaxiques.

2.1 Les pronoms toniques *moi* et *toi*

Le français ne permet pas de construire « directement » un sujet avec les pronoms toniques *moi* et *toi*. Il y a donc systématiquement cooccurrence

---

3 Cette observation ne concerne que les noms déterminés.

du tonique et du clitique, quelle que soit la détermination du nom. A côté de :

(9) que **moi la consommatrice j'**en ai marre de consommer il faut le faire[4] (Caddéo, Adulte, conversation)
(10) **toi le père de Marc tu** souhaites prendre la parole (exemple inventé)

il est impossible de trouver :

(9a) ?que **moi la consommatrice** en ai marre de consommer il faut le faire
(10a)?**toi le père de Marc** souhaites prendre la parole

Cette contrainte dépend également du caractère déictique des pronoms clitiques de même personne, on ne peut donc pas rencontrer de séquence de type *le N je / le N tu :*

(9b) ?que **la consommatrice j'**en ai marre de consommer il faut le faire
(10b) ? **le père de Marc tu** souhaites prendre la parole

Cette observation s'applique aussi lorsque la suite *tonique+lexique* est après le verbe ou entre le modal et le verbe recteur :

(11) **Je** ne peux pas, **moi, le père**, espérer des choses comme celles-là. (cité par Sandfeld, Duh.)
(11a) **Je** ne peux pas, le **père**, espérer des choses comme celles-là.

Ces observations rendent peu probables les séquences disjointes du type *moi... le N.* La disjonction du nom et du tonique pose des problèmes différents selon le type de pronom en jeu. Avec *moi*, les énoncés sont bizarres, quelle que soit la détermination du nom :

(12) ?que **moi j'**en ai marre de consommer **la consommatrice** il faut le faire
(13) ?**moi je** m'en fous **ta mère** comme de l'an quarante
(14) ?**moi je** tiens absolument **policier** à venir vous voir

Nous pensons avoir affaire ici à un emploi spécial de *moi je*, où les deux morphèmes sont liés ensemble et se distinguent d'autres emplois de *moi* détaché. Ce serait un *moi je* de l'énonciateur qui s'emploie en majorité avec les verbes de discours et ne donne aucun effet de contraste. En tant que morphème à deux éléments, *moi je* ne se laisserait pas « prolonger », à distance ou non, par un élément apposé – pas plus que *je* employé seul. La seule apposition possible interviendrait à l'intérieur du morphème, entre *moi* et *je :*

(15) que **moi la consommatrice j'**en ai marre de consommer il faut le faire (Caddéo, Adulte, conversation)

---

4 Il s'agit d'un énoncé exclamatif.

(16) tes problèmes de boulot tes problèmes avec ta femme tes problèmes de fric tes problèmes en général et en particulier **moi ta mère je** m'en fous comme de l'an quarante (Dialogue, Film « La crise »)

(17) vous les avez entendus euh les fonctionnaires dans la rue ils disent **moi facteur je** gagne sept mille francs par mois et c'est insupportable (Sinclair, 6, 15-16; 7, 1-4)

Avec *toi*, seuls les cas avec nom à déterminant zéro seraient acceptables. On obtient ce qu'on nomme plus traditionnellement un « vocatif ». Il s'agirait d'une adresse, une sorte d'interpellation :

(18) **toi tu** es absolument obligé **lecteur** de comprendre

(19) **toi tu** gagnes **facteur** sept mille francs par moi

Nous pourrions déduire que la contiguïté entre les pronoms toniques *moi* et *toi* et le nom détaché apparaît comme un critère syntaxique essentiel dans l'analyse de ces constructions. Cependant, nous avons le sentiment de ne pas avoir épuisé les possibilités distributionnelles de ces pronoms. Il faudrait leur consacrer une étude entière pour confirmer ou infirmer ces tendances.

Dans l'analyse des pronoms *nous* et *vous* d'une part et *lui, elle, eux, elles* d'autre part, nous dissocierons la position syntaxique de sujet et celle de complément. Les contraintes sur la détermination du nom et les effets sémantiques sont très différents.

2.2. Les pronoms toniques nous et vous

2.2.1. Le sujet

Les pronoms *nous* et *vous*, de par leur caractère complexe (*cf.* Blanche-Benveniste, 1987), sont très souvent développés par un ou plusieurs noms pour en préciser la valeur. Le paradigme des déterminants du nom détaché est ouvert, quelle que soit la position de la suite *tonique+lexique* :

- zone préverbale

(20) et après la manif il [R. Barre] avait il était intervenu à la télévision en disant que **nous les Provençaux on** avait fait une manifestation - folklorique (Nevchehirlian, 1987, 19,7)

(21) mais + **nous avocats nous** les côtoyons + et on sait que + tout homme a des bons côtés (Chervet, 1990, 47,5-7)

(22) elle va d'abord demander l'agrément du propriétaire est-ce que **vous propriétaire vous** acceptez que nous cédions - euh notre local - à -

une société que nous créons- et qui nous continuera dans notre activité (Castagné, 4, 1-3)

- zone post verbale

(23) **on** est fier de notre de notre histoire - **nous les Français** - eh eh mais euh et donc c'est bien de c'est bien de le garder ce patois (Elodie, 7, 10-16)

(24) est-ce que **vous** craignez pas **vous Parti Socialiste** d'être un peu déconnecté (Caddéo, Adulte, oral public T.V.)

La disjonction du nom semble réduire les possibilités. Dans les exemples avec *nous*, il n'y a presque que des cas du type *mon frère et moi*, c'est-à-dire un nom déterminé par un possessif coordonné au pronom *moi* :

(25) puis mes parents se sont mal entendus se sont toujours + pas déchirés + mais + ils sont toujours ensemble + seulement **nous on** en a pris **mon frère et moi** (Alcoolique, 104, 5-11)

Les cas sont beaucoup plus fréquents lorsque le pronom tonique est absent :

(26) enfin bon ça a pas été une année merveilleuse seulement **nous** en avons profité **mes soeurs et moi** (Corsin, 1984, 8,8-9)

(27) mais je me souviens euh - surtout de - des petites responsabilités qu'**on** avait **mon frère et moi** - en allant à l'école - par rapport par rapport à ma mère qui travaillait et mes deux frères aînés (Femmes)

La classe des noms de parenté est la plus représentée. Cette régularité dépend, peut-être du type discursif des corpus, les exemples étant tirés, pour la plupart, de récits de vie. Nous pourrions y voir également un premier modèle syntaxique qui ferait correspondre à l'éloignement du tonique et de la partie lexicale des restrictions sur la détermination du nom et sa classe lexicale. Ces caractéristiques réapparaissent en effet dans les constructions avec le tonique *lui* (*cf.* 2.3.)

2.2.2. L'objet

Les rares exemples avec *nous* tonique, que nous avons relevés échappent aux tendances distributionnelles de notre modèle syntaxique. Il s'agit de noms disjoints avec un défini et qui n'appartiennent pas à la classe lexicale des noms de parenté :

(28) bon on va passer - on va passer au deuxième point - bon on va parler de - de la parole - parce que bon **nous** ça **nous** intéresse **les apprentis linguistes** (Chervet, 1990, 10,10-12)

(29) les enfants connaissent la mort **nous** elle **nous** fait peur **les adultes** (Caddéo, Adulte, oral public T.V.)

L'exemple d'objet prépositionnel (29) est très intéressant. La fonction complément prépositionnel y est seulement marquée par le clitique alors que nous savons qu'en français les dislocations à droite de compléments prépositionnels font toujours apparaître, dans le cas des syntagmes nominaux, la préposition :

(30) mais j'ai dit moi je **lui** donne pas tort **à ma fille** (Demay 16,9)
(31) je me souviens pas **lui** avoir sauté au cou **à ma mère** (Lash 6,1)

Dans l'exemple (29a), le marquage prépositionnel du syntagme nominal semble impossible. Il bloque la relation de coréférence avec le clitique objet ou le tonique ; on dirait que la fonction complément prépositionnel ne peut être marquée par deux éléments, même si l'on supprime le pronom tonique détaché à gauche (24a) :

(29a) ?les enfants connaissent la mort **nous** elle **nous** fait peur **aux adultes**
(29b) ?les enfants connaissent la mort elle **nous** fait peur **aux adultes**

En revanche, la cooccurrence des trois éléments est possible si le pronom tonique détaché à gauche entretient une relation de double marquage avec le clitique, alors que le SN détaché à droite est comme une spécification lexicale:

(29c) les enfants connaissent la mort **à nous** elle **nous** fait peur **les adultes**

Il y a là un phénomène surprenant qui dépend forcément des caractéristiques « complexes » du pronom *nous*. Il ne semble pas compatible avec n'importe quel syntagme nominal.

Ce constat devrait permettre de s'interroger sur le statut du syntagme nominal détaché et sur sa dépendance au pronom tonique (*cf.* 2.). Nous y reviendrons après l'analyse des pronoms de la dernière série.

2.3. Les pronoms *lui*, *elle*, *eux*, *elles*

2.3.1. Le sujet

Deux constats ressortent de l'observation de nos données. Tout d'abord, quand la position sujet est remplie par le pronom tonique et le pronom clitique sur le modèle *eux, ils... nos profs*, la détermination du nom disjoint semble se restreindre au possessif. De ce fait, le lexique appartient souvent aux noms de parentés :

(32) et **lui il** est né en quarante-huit **mon mari** (Oral, Pavray)
(33) **lui il** avait encore une tante **mon mari** (SCA, 39, 5)

L'on peut voir ici des contraintes qui sont propres à cette structure, puisque nous savons qu'en français parlé, dans la construction disloquée avec un clitique sujet du type *ils... les profs*, le paradigme des déterminants est ouvert, bien que l'on puisse faire une hiérarchie des fréquences. Il est très fréquent de rencontrer des noms avec défini (*les profs*) et possessif (*ton père*) :

(34) **ils** sont toujours là **les profs** toujours derrière (Oral, Boavi 86,6)
(35) **il** était syndicaliste **ton père** (Oral, La Navale 16,2)

Les exemples avec noms accompagnés d'un démonstratif (*cette femme*) ou d'un indéfini (*un maçon*) sont plus rares[5] :

(36) quelle allure **elle** avait **cette femme** (Oral, Trignan 3,2)
(37) vous savez combien **il** m'avait demandé **un maçon**. (Oral, Cappeau 11,1)

Au vu de ces acquis descriptifs, on peut penser que c'est bien la discontinuité dans la séquence qui pose cette contrainte sur la détermination puisque dans les cas où le syntagme nominal est contigu au pronom, le paradigme des déterminants est également très divers : défini (*lui, le cochon*), démonstratif (*lui, ce garçon-là*) détermination zéro (*lui, Français*):

(38) le loup il a vu le petit cochon et **lui le cochon il** a vite couru (Oral, corpus enfants)
(39) **lui ce garçon-là il** réussissait il avait appris à lire (Oral, LAR-R00PRI002)
(40) pourquoi **lui Français il** aurait pas besoin d'un dico (Caddéo, Adulte, conversation)

Ensuite, comme pour les pronoms *moi* et *toi*, la cooccurrence des trois catégories grammaticales (pronom tonique, pronom clitique et syntagme nominal) paraît systématique, alors que la séquence *lui le N* est tout à fait possible. Nous n'avons pas rencontré d'exemples du type suivant (ex. 41), comme si la relation syntaxique entre le pronom tonique et le syntagme nominal posait problème dès lors qu'il y a discontinuité au regard des cas fréquents où le pronom tonique, immédiatement suivi de la partie lexicale, est sujet direct du verbe (ex. 42) :

(41) **lui** parlait que patois **mon père**

---

[5] Les noms à déterminant zéro doivent être considérés à part. Dans ce type de construction, il n'est pas possible de parler de dislocation, puisque le syntagme détaché ne pourrait avoir la fonction de sujet auprès du verbe :

je pense qu'**elle** doit être assez terrible **mère** (Oral, Caddéo, Adulte, conversation)

(42) on est derrière un bureau - **lui le malade** est là en face (193,112)

Nous pouvons dire que la structure syntaxique de type *lui, il... mon N* avec possessif est la structure privilégiée du français parlé dès que l'on envisage une relation entre deux éléments de type *lui... SN* dans laquelle le syntagme nominal serait détaché de l'autre côté du verbe. On peut donc parler de deux contraintes : l'une concerne le déterminant ; l'autre la présence obligatoire du pronom clitique dans cette structure où la distance syntaxique semble peu possible. Nous retrouvons le modèle syntaxique observé avec *nous*.

2.3.2. L'objet

Dans le cas d'un complément non prépositionnel, le constat est le même que dans la position sujet. Quand il y a les trois éléments, le paradigme de la détermination du nom semble se réduire au possessif (*lui aussi je le... son fils*) :

(43) comme je fais partie de la région centre on avait des relations ensemble **lui aussi** je **le** connais **son fils** puisque son fils il est militaire au sur la région de Blois je l'ai vu plusieurs fois (Orléans/gra014)

Nous aurions souhaité observer les cas de discontinuité avec un complément prépositionnel sur le même modèle que l'exemple avec *nous* déjà cité :

(29) les enfants connaissent la mort **nous** elle **nous** fait peur **les adultes** (Caddéo, Adulte, oral public T.V.)

Nous n'avons relevé aucun exemple avec *lui, eux* et les exemples construits non seulement ne donnent pas de résultats satisfaisants, mais encore se comportent différemment des exemples avec *nous* :

(44) ?les adultes aiment mon spectacle **eux** je **leur** fais peur **les enfants**
(45) ?**elle aussi** je **lui** donne pas tort **ma fille**
(46) ?**elle** je **lui** ai sauté au cou **ma mère**

Curieusement, le fait de marquer par la préposition la fonction auprès du pronom tonique (*à elle je lui... ma mère*) ou du syntagme nominal disjoint (*elle je lui... à ma mère*) ne permet pas de construire des exemples plus « familiers ». Intuitivement, nous aurions tendance à rejeter les cas suivants :

(44a) ?les adultes aiment mon spectacle **à eux** je **leur** fais peur **les enfants**
(44b) ?les adultes aiment mon spectacle **eux** je **leur** fais peur **aux enfants**
(44c) ?les adultes aiment mon spectacle **eux** je **leur** fais peur **les enfants**

Notons que les exemples (44a) et (44c) ne sont pas possibles alors qu'ils sont les seuls acceptables avec le pronom *nous*.

Est-ce la cooccurrence des trois catégories (pronom tonique, pronom clitique, syntagme nominal) ou la disjonction du nom qui bloquent la réalisation de la construction prépositionnelle ici ? Pour tenter de répondre à cette question, trois remarques. Tout d'abord, un problème de compatibilité existe entre le tonique et le lexique. Nous accepterions très facilement le même énoncé sans élément lexical *eux, je leur fais peur* ou sans tonique *je leur fais peur aux enfants*. Dès qu'il y a le tonique et le groupe nominal, l'énoncé devient « bizarre » : *?eux, je leur fais peur, aux enfants*.

Ensuite, les cas où le pronom tonique et le syntagme nominal sont contigus sur le modèle *je leur... à eux, les enfants*, ne sont pas plus fréquents : aucun exemple en français parlé et seulement deux dans le journal « Le monde Diplomatique » sur plus de dix ans d'archives :

(47) En septembre, lorsqu'ils apprennent que, à l'initiative des autorités universitaires françaises et des mêmes Chinois qui organisèrent le mouvement travail-études, s'ouvre à Lyon, dans le vieux fort Saint-Irénée, un Institut franco-chinois, et que cet institut ne **leur** est pas destiné, **à eux, les sans-ressources**, mais qu'on y attend des boursiers recrutés spécialement en Chine, ils décident de lancer le mouvement, célèbre dans les annales de la République populaire sous l'appellation de Lida Yundong, (...) (Ecrit, Le Monde Diplomatique, MD8908:17)

(48) Mais puisqu'on ne **leur** en demande pas compte, **à eux, les sans-grade**, puisqu'ils se découvrent nombreux, éberlués et ravis dans l'aigreur, puisque celui à qui ils donnent leurs voix focalise l'attention par ses éructations, puisque sur toute chose ses avis sont recherchés avec une déférence craintive, ils se sentent libérés des pesanteurs de la morale sociale qui entravait leur ego. (Ecrit, Le Monde Diplomatique, MD9203:8)

Ces énoncés semblent cependant plus satisfaisants, comme s'il était plus naturel d'envisager la dislocation d'un pronom tonique « prolongée » par un SN non marqué syntaxiquement. D'ailleurs les éléments « associés » à droite du *type je **leur** fais pas confiance **les femmes*** ne se rencontrent pas en français parlé.

Ce constat d'incompatibilité du pronom tonique et du lexique contigu ou non dans ce type de construction conduit à penser :

1. que pour la fonction complément prépositionnel, et pour ce type de pronoms, la cooccurrence des trois catégories ne s'impose pas ;

2. il n'est pas possible d'avoir simultanément un pronom tonique détaché à gauche, sans fonction syntaxique, et un syntagme nominal détaché à droite du verbe en relation de double marquage avec le pronom clitique.

Nous avons, en effet, observé que les seuls cas de cumul possibles en français parlé étaient de type :

(49) **le couvre feu** on **l'**avait établi **le couvre feu**
(50) **ce pain-là** j'**en** mange pas **de ce pain-là**

Dans ces exemples, on peut dire que l'élément disloqué avant le verbe n'a pas de fonction syntaxique auprès du verbe recteur et qu'il sert de cadre préalable à l'énoncé. L'élément disloqué après le verbe est toujours un prolongement lexical et syntaxique de la rection du verbe marquée par le clitique. Les deux statuts sont compatibles dans la même construction verbale, mais, contrairement aux exemples qui posaient problème précédemment, il s'agit de deux syntagmes nominaux disloqués identiques.

On peut donc dire que, dans les structures qui nous intéressent, le cumul n'est pas possible car il s'opère avec deux catégories grammaticales différentes.

3. Statut du syntagme nominal disjoint

Ces différentes observations posent le problème du statut syntaxique du syntagme nominal disjoint. Il peut avoir deux statuts : celui d'élément apposé ; celui d'élément disloqué, en relation avec le pronom clitique.

Dans le premier cas, le nom n'est pas concerné par la fonction de l'élément tête du constituant, il est dans une relation de dépendance.

Dans l'autre, il peut marquer, lui aussi, une position syntaxique, au même titre que le pronom clitique et que le pronom tonique, mais seulement pour la position sujet puisque pour les positions compléments, le SN n'est jamais introduit par une préposition.

Pour cette fonction, nous pouvons alors considérer que dans le modèle *lui il... mon N*, le SN détaché ne peut pas être analysé comme un élément disloqué en « reprise » syntaxique et référentielle du pronom clitique. L'impossibilité de marquer la fonction prépositionnelle à la fois par le pronom clitique et le syntagme nominal dans les cas avec *nous* apparaît

comme un indice renforçant cette hypothèse. Le syntagme nominal est un élément apposé qui a, comme support, le pronom tonique[6] :

(51) [lui] il avait encore une tante **mon mari** (SCA, 39)

La classe lexicale du syntagme nominal détaché est un indice supplémentaire. Nous savons que, dans cette position, le pronom ne peut sélectionner que des noms qui ont le trait [+humain] (Blanche-Benveniste *et al.*, 1987 ; Caddéo, 2000).

(2) ?**lui le bureau il** sera poussé contre le mur

Cette restriction semble observable dans le cas de discontinuité. Nous n'avons pas rencontré d'exemples du type :

(2a) ?**lui il** sera poussé contre le mur **le bureau**

Si l'augmentation du volume d'exemples confirme cette tendance, nous conclurions que c'est une contrainte imposée par le pronom, alors identifiable comme support.

4. Conclusion

Les distributions observées pour la structure *lui il... SN* conduisent à s'interroger sur les contraintes distributionnelles des différents pronoms toniques de type *moi*, *nous*, *lui* dans ce type de constructions, sur la détermination et les caractéristiques sémantiques du syntagme détaché et sur le rôle de la distance syntaxique. La cooccurrence du pronom et du clitique est comme systématique dans certains cas. On peut alors parler d'une contrainte d'éloignement qui bloquerait les appositions sur un pronom tonique.

Le modèle syntaxique identifié oblige à repenser l'analyse par dislocation ou par apposition puisque se pose la question du statut à accorder au SN détaché. Les contraintes relevées pour la séquence *lui il V... SN* et les tendances observées pour les dislocations nous incitent à proposer une analyse pour certaines constructions ambiguës.

Dans la fonction objet prépositionnel, le constat est tout autre et impose de distinguer le pronom *nous* et le pronom *lui*. Non seulement il semble difficile de cumuler une dislocation de type « double marquage » à droite et une apposition à gauche, dans un contexte de discontinuité, mais

---

6 Le support est mis entre crochets.

encore les catégories grammaticales jouent un rôle important pour l'analyse du phénomène. Nous avons une certitude : ce n'est pas une disjonction d'éléments solidaires.

Références

Apotheloz, D. & Zay, F. (1999). Incidents de la programmation syntagmatique : reformulation micro- et macro-syntaxiques. *Cahiers de linguistique française, 21*, 11-34.

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J. & Van den Eynde, K. (1987). *Pronom et syntaxe. L'approche pronominale et son application à la langue française*. Paris : SELAF (2e éditon Augmentée).

Blanche-Benveniste, C. (1987). Le pronom on : propositions pour une analyse. In J. Bonnamour (Ed), *Mélanges offerts à Maurice Molho, linguistique, vol III*, 15-30.

Blasco-Dulbecco, M. (1999). *Les dislocations en français contemporain. Étude syntaxique*. Paris : champion.

Blasco-Dulbecco, M. & Caddéo, S. (2001). Apposition et dislocation : la séquence pronom + lexique + clitique. *Recherches sur le français parlé, 16*, 125-149.

Caddéo, S. (2000). *l'apposition détachée : description syntaxique de l'apposition nominale détachée dans divers registres de la langue parlée et de l'écrit en français contemporain*. Thèse de doctorat, Université de Provence, Aix-en-Provence.

Caddéo, S. (en préparation). Lui, le propriétaire, le propriétaire, lui : deux constructions bien distinctes.

Skårup, P. (1975). Les premières zones de la proposition en ancien français- essai de syntaxe des positions. *Études romanes, 6*.

L'opposition de modalité en français parlé : étude macro-syntaxique

Frédéric SABIO
Université de Provence

Résumé

Cet article décrit certains énoncés qui expriment une opposition entre modalité affirmative et négative, éventuellement accompagnés d'un connecteur comme *mais* ou *et*. Par exemple : *c'est pas l'appareil qui vaut cher. C'est les objectifs* ; *Ce ne sont plus les pays qui intéressent mais les gens*. (énoncés oraux). Dans un premier temps, nous donnons une description assez générale des usages qui apparaissent fréquemment dans les productions orales, en proposant une typologie des principales organisations rencontrées. Dans la seconde partie de l'article, nous poursuivons la description en nous fondant sur l'opposition entre micro-syntaxe et macro-syntaxe.

1. Introduction

Cet article décrira certains énoncés qui expriment une opposition entre modalité affirmative et négative, éventuellement accompagnés d'un connecteur comme *mais* ou *et*. Par exemple :

C'est pas l'appareil qui vaut cher. C'est les objectifs. (oral)[1]
Ce ne sont plus les pays qui intéressent mais les gens. (C13, 231, 19)
Les irlandais, ils ont l'habitude de ce genre de choses ; les roumains pas. (radio)

---

[1] Pour rendre ces premiers exemples plus lisibles, nous les accompagnons d'une ponctuation. Dorénavant, les exemples issus de corpus oraux seront transcrits sans aucune ponctuation. Les données utilisées dans cette étude proviennent en grande partie du corpus d'énoncés recueilli par M-L. Hennequin, que nous remercions. Nous avons analysé pour cette étude un peu plus de 200 énoncés à contraste.

Recherches sur le français parlé, 17, 55-77.
©2002, *Université de Provence*

Je suis rentrée hier soir de Lisbonne et ma valise non. (oral)
Il faut qu'il soit très propre et pas sale. (Mazo, C55, 9, 16)
Ils ont le droit mais pas tout le temps. (Levy, C48, 8, 14)

L'étude concernera essentiellement l'oral, et se limitera aux séquences qui ne comportent qu'un verbe constructeur, à l'exclusion des énoncés qui introduisent chaque versant de l'opposition modale à l'intérieur d'une construction verbale autonome :

il vient il vient pas ça dépend (oral)
j'ai pas fait d'études j'ai fait de la luge (C46, 93, 1)

Ces derniers exemples, qui ne seront pas décrits ici, se laissent en effet segmenter en plusieurs constructions autonomes, chacune organisée autour de son propre verbe recteur :

[j'ai pas **fait** d'études] [j'ai **fait** de la luge]

La description sera fondée sur l'opposition, exposée dans Blanche-Benveniste *et al.* (1990), entre deux modes différents d'expression de la modalité négative ; on distinguera :

- Les modalités dites « de relation », qui se trouvent intégrées à la syntaxe des constructions au moyen de certaines marques dont la distribution apparaît fortement contrainte : ainsi, dans *il ne travaille pas à Aix*, les éléments *ne* et *pas*, qui signalent une modalité de relation, encadrent nécessairement le verbe tensé de la construction verbale.
- Les modalités « de terme » qui se manifestent par la présence du terme modalisateur (*pas*, *non*, *non pas*) juste avant l'élément affecté sémantiquement par la négation : ainsi, la séquence *pas à Aix* constitue une modalité de terme dans *il travaille à Paris et pas à Aix*.

Nous défendrons dans cet article l'hypothèse selon laquelle la distinction entre modalité de relation et de terme peut s'enrichir d'une interprétation macro-syntaxique : nous montrerons notamment que dans les énoncés qui présentent un contraste fondé sur la modalité de relation (*il ne vit pas à Aix mais à Paris*), les termes, positif et négatif, du contraste semblent être nécessairement réalisés au sein d'une même unité de macro-syntaxe ; en revanche, dans les contrastes bâtis sur une modalité de termes, il est fréquent que l'une des deux modalités forme une séquence macro-syntaxique qui jouit d'une certaine autonomie structurelle : ainsi, dans *il vit à Aix et pas à Paris*, la séquence négative *pas à Paris,* bien qu'elle soit dépendante du verbe recteur en terme de relations micro-syntaxiques, se

laisse analyser comme une séquence macro-syntaxique « détachée » du reste de l'énoncé.

Dans un premier temps, nous donnerons une description assez générale des usages qui apparaissent fréquemment dans les productions orales, en proposant une typologie des principales organisations rencontrées. C'est dans la seconde partie que nous analyserons les constructions contrastives en recourant à la distinction entre micro-syntaxe et macro-syntaxe.

2. Propriétés syntaxiques des constructions à contraste

2.1. Typologie

Les données peuvent être commodément réparties en 5 types principaux, en fonction du genre de modalité présent (modalité de terme ou de relation), du type de dispositif utilisé (extraction ou autre), et de certains faits qui touchent à la disposition linéaire des éléments du contraste[2].

Précisons qu'on n'aura pas recours à l'analyse par ellipse : les énoncés

ce ne sont plus les pays qui intéressent mais les gens
c'est pas l'appareil qui vaut cher c'est les objectifs

ne seront pas décrits comme la réduction d'une construction comme :

ce ne sont plus les pays qui intéressent mais [ce sont] les gens [qui intéressent][3]
c'est pas l'appareil qui vaut cher. C'est les objectifs [qui valent cher]

2.1.1. Les contrastes qui impliquent une modalité négative de relation, en dispositif direct

On dit qu'une construction syntaxique est exprimée sous dispositif direct lorsqu'elle ne présente aucun phénomène de focalisation tel que l'extraction ou d'autres types de clivage : *il vit à Paris* est bâti sous dispositif direct, à l'inverse de *c'est à Paris qu'il vit* ou *là où il vit c'est à Paris* qui réalisent respectivement le dispositif d'extraction et de pseudo-clivage.

---

[2] La classification reprend l'analyse proposée dans Blanche-Benveniste *et al.* (1990).

[3] Voir Cappeau et Bilger (1995) pour une argumentation.

On désigne par « modalité négative de relation » les modalités qui, dans le français normatif, s'expriment par l'encadrement du verbe tensé au moyen des marques *ne... pas*. Il a souvent été noté qu'avec ce type de structure, la portée sémantique de la modalité est parfois difficile à cerner avec certitude[4]. Dans :

je ne suis pas allé à Paris parce qu'on me l'a demandé

on peut envisager que la négation puisse porter soit sur la relation entre le verbe et *à Paris* (= *je ne suis pas allé à Paris, et ce parce qu'on me l'a demandé*), soit sur la relation entre le verbe et la séquence causale (= *je suis allé à Paris, mais pas parce qu'on me l'a demandé*).

Les contrastes qui font intervenir une modalité de relation sont très fréquents, à l'oral comme à l'écrit. Par exemple :

on ne se drogue pas pour mourir par refus de vivre comme le pensent les autres mais pour vivre une autre vie (Cronert, C74, 10, 5).
son pardessus ne recouvre point comme il le pense – une tenue de ville – mais une tenue de sport (C37, 77, 12)

Tous les énoncés rencontrés dans le corpus qui fonde cette étude font usage du connecteur *mais*. *Et* est exclu, tout comme l'absence de connecteur (qui sera désormais désigné par « connecteur ø »).

Dans le corpus analysé, les contrastes s'opèrent sur divers types de compléments, mais jamais sur les sujets, qu'ils soient antéposés ou postposés au verbe recteur :

*Marie n'est pas venue mais Serge[5]
*Marie mais Serge n'est pas venue

En outre, le corpus ne recèle pas d'exemple du type *ne sont pas arrivés les Anglais mais les espagnols*, dont le caractère grammatical n'est pas à exclure selon nous.

2.1.2. Les contrastes qui impliquent une modalité de terme en dispositif direct

A la différence de la modalité de relation, les modalités de terme sont placées juste avant la séquence sur laquelle elles portent sémantiquement : par exemple, il y a une modalité négative de relation dans *il ne viendra pas demain*, mais une modalité négative de terme dans *il viendra mais pas demain* ou dans *il viendra dans la semaine, pas demain :* en d'autres

---

4 Muller (1991).

5 Jeanjean (1980) avait analysé ce type d'impossibilité.

termes, la modalité négative de terme n'est jamais fondée sur la négation du verbe recteur, ce que l'on constate aisément à travers l'impossibilité d'utiliser le *ne* pré-verbal (**il ne viendra dans une semaine mais pas demain*, **il ne viendra dans une semaine, pas demain*). Dans cette structure, le négateur utilisé est variable : *pas*, *non* ou *non pas* (*il viendra dans une semaine, et pas demain*, *il viendra non demain mais la semaine prochaine*, *il viendra non pas demain mais la semaine prochaine*).

a. distinction entre les modalités de terme et de relation

La distinction entre les deux types de modalité négative est en certains cas difficile à établir avec certitude : aucune difficulté lorsque l'élément *ne* est réalisé (modalité de relation, attelée au verbe) ni lorsque le négateur prend la forme *non pas* ou *non* (qui caractérisent la modalité de terme). L'analyse paraît également assurée lorsque la structure de l'énoncé n'est pas compatible avec la présence de *ne* : on posera la réalisation d'une modalité négative de terme dans :

il a été pas gentil

dans la mesure où est exclue la possibilité de trouver

*il n'a été pas gentil

En revanche, plus délicate est l'analyse d'une construction comme

il est pas gentil

dans laquelle *pas* connaît un statut grammatical ambigu : il peut soit instancier une modalité de relation de laquelle l'élément *ne* serait absent :

[il] [(n')est pas] [gentil]

soit être analysé comme une négation de terme ayant un statut de modalisateur adjectival :

[il] [est] [pas gentil] [6]

Certains énoncés à contraste posent un problème de description similaire. Dans :

on buvait pas dans des verres mais dans des bols (Junqua, C53, 3, 22)

on a tout autant la possibilité de restituer le *ne* antéverbal (*on ne buvait pas dans des verres mais dans des bols* : contraste sur une modalité de

[6] Blanche-Benveniste *et al.* (1990). Remarquons que la morpho-syntaxe de l'anglais ne présente pas d'ambiguïté de ce type : une construction à modalité de relation comme *he is not a fool* resterait distincte de la construction à modalité de terme correspondante : *he is no fool*.

relation), que d'établir une commutation entre *pas* et *non* ou *non pas* (*on buvait non pas dans des verres mais dans des bols :* contraste sur une modalité de terme). Afin de ne pas obscurcir l'exposé, les exemples ambigus ont été écartés de cette étude.

b. disposition linéaire des termes du contraste

Parmi les contrastes fondés sur une modalité de terme, il convient de distinguer ceux qui présentent en premier lieu l'élément positif de l'opposition (+), et en second lieu l'élément négatif (-), comme :

> va le dire +[à d'autres] mais –[pas à moi] (Demay, C74, 15, 4)
> je côtoie +[des bonhommes] et –[non pas des nanas] (Martin, C60, 5, 15)
> on parle +[de la mentalité] –[pas de la jeunesse] (C15, 37, 4)

et ceux qui donnent en premier l'élément négatif du contraste :

> j'ai traité ce livre euh –[non pas du côté des conquérants] mais +[du côté du peuple] (...) (Enfru, C58, 17, 5)
> tout cela est vécu –[non par les possédants] mais +[par le plus grand nombre] (C68, 4, 13)

Lorsque le contraste va du positif vers le négatif (contrastes [+ -]), il y a une commutation possible entre les connecteurs *mais*, *et* et *ø*. L'élément négatif le plus fréquent dans les productions orales est de loin *pas*.

Lorsque le contraste présente en premier lieu l'élément négatif (contrastes [- +]), l'usage de *et* est exclu : c'est *mais* qui apparaît de façon très majoritaire. Sa présence paraît systématique lorsque l'élément négatif est introduit par *non pas* ou *non*, si bien que l'on peut prévoir la cooccurrence régulière de *non pas... mais*. L'absence de connecteur, très minoritaire, a été relevée dans quelques exemples où l'élément négatif est introduit par *pas* comme :

> Ce n'est pas un film, ce sont des rencontres. Pas avec des personnages qui se croiseraient dans une petite ville de province. [Ø] Avec des personnes. (presse écrite)
> le retournement des morts c'est une coutume qui se pratique dans les plateaux malgaches pas dans tout Madagascar [Ø] uniquement autour de Tananarive (C16, 88, 16).

Si dans les contrastes [+ -], c'est le négateur *pas* qui domine, on relève que dans les contrastes [- +], l'emploi de *non pas* est relativement fréquent, y compris à l'oral. *Non pas* ne saurait donc être considéré comme un tour réservé à l'écrit cérémonieux. Les contrastes qui impliquent une modalité de terme en dispositif direct sont d'ailleurs les seuls où l'emploi de *non pas* soit attesté dans le corpus constitué pour cette étude : *non pas* paraît ainsi

être réservé à cet emploi contrastif[7], si bien que les énoncés dans lesquels il figure et qui ne développent aucun contraste tendent à être interprétés comme structurellement inachevés :

> c'est-à-dire que je j'ai énormément de mal - non pas à me réveiller je me réveille – immédiatement je me lève – et je me dis (...) (Gris-gris pour le quotidien)

Les contrastes bâtis sur une modalité de terme en dispositif direct (que ce soit l'élément positif ou négatif qui soit placé en tête) peuvent porter aussi bien sur des éléments pleinement valenciels (sujets, et compléments que la grammaire scolaire décrirait comme « obligatoires » ou « essentiels ») que sur des éléments de simple rection (les « circonstants »). On relève une proportion notable de séquences en *pour* (*pour eux / pour cela / pour faire cela*) et de séquences proportionnelles à *ainsi*, *comme ça :*

> dans le cadre de la photo les gestes sont importants pour leur dimension affective – pas pour ce qu'ils peuvent m'apprendre sur la personne (c39, 23, 13)
> il prêcherait non pas pour sa paroisse mais pour la mienne (Castagne, C63, 19, 10)

c. Quelques contraintes

Les syntagmes à fonction de sujet antéposé connaissent un comportement particulier. D'une part, les contrastes [- +] sont toujours impossibles :

> *non pas Marie mais Sophie est venue
> *non pas Marie est venue mais Sophie

d'autre part, les contrastes [+ -] ne semblent guère pouvoir apparaître sous la forme

> *Sophie et non pas Marie est venue

mais exigent le rejet de la séquence négative à droite du verbe recteur :

> Sophie est venue et non pas Marie

Ce type d'énoncé, qui a surtout été relevé dans la langue écrite, ne paraît pas très fréquent dans la langue parlée quotidienne :

> Elle veut le garder, pas lui (presse)
> Les Américains ont fait ce film, et pas nous. (presse)
> certains sont plus traditionalistes mais pas l'ensemble (Balian, C47, 8, 8)

---

7 A part quelques emplois rares à l'oral, du type :

> alors non pas que je veuille entretenir les malades dans leurs maladies... (Gris-gris pour le quotidien)

Ce phénomène de rejet de la partie négative du contraste ne touche pas uniquement les éléments sujets mais peut affecter divers types de compléments, dont la version positive est antéposée à la construction :

avec ma mère on avait des parties de rire (...) mais jamais avec euh mon frère (C28, 59, 10)

d. Les contrastes à un seul terme réalisé

Soit un énoncé à contraste sur un terme du type [+ -] (*je passerai demain mais pas ce soir*) ou [- +] (*je passerai non pas ce soir mais demain*). Il est possible de ne pas réaliser le premier des deux éléments du contraste, pour obtenir : *je passerai ~~demain~~ mais pas ce soir* ou *je passerai ~~non pas~~ ~~ce soir~~ mais demain.*

Ce type d'énoncés, où un seul des termes de l'opposition modale est instancié, est d'une grande fréquence dans les données recueillies :

Le chef du gouvernement explique qu'il veut débattre... **mais à quatre** : Juppé-Léotard contre Jospin-Hue. (presse écrite)
mon père il il savait lire écrire **mais pas énormément** (C28, 67, 6)
Je l'instruis **mais sans succès** (Giraudoux, Ondine)
on lui donne un un un ordre il va l'exécuter **mais sans sans sans avoir le fond méchant** (Gaucel, C72, 9, 14)

Ces constructions ont la particularité d'induire le sémantisme contrastif de manière implicite ; on pourrait sans difficulté opérer un remplissage lexical de la partie non réalisée : *il veut débattre non pas à huit mais à quatre, je l'instruis avec persévérance mais sans succès...*

Dans le corpus étudié, ces contrastes « à un seul terme » affectent des types de rection assez divers : temporelle, rection de type *ainsi*, *comme ça*, *pour cela*, *avec lui*, rection nominale... Il existe cependant quelques contraintes[8] : la plus forte est l'impossibilité d'exprimer les contrastes à un seul terme avec les syntagmes qui instancient une place valencielle dont la réalisation est obligatoire ; ainsi en va-t-il des sujets, qu'ils soient antéposés ou postposés au verbe recteur :

*sont arrivés ~~les anglais~~ mais pas les espagnols

Il en est de même pour certains compléments réputés non effaçables :

*il ressemble ~~à son père~~ mais pas à sa mère
*il ressemble ~~non pas à sa mère~~ mais à son père
*il ira ~~à aix~~ mais pas à Paris
*il ira ~~non à Aix~~ mais à Paris
*il a mis ~~une semaine~~ et pas deux jours

---

8 Voir Bilger (1998) à propos des énoncés en *et*.

2.1.3. Les contrastes de modalité sous dispositif d'extraction

Dans les productions orales, l'expression du contraste se trouve fréquemment associée à l'extraction en *c'est... qu- :*

c'est pas à nous qu'il faut le demander c'est au syndicat des forains (C3, E36, 13)
c'est pas à Cassé-le-Vivien que je suis né c'est à Calais (C46, 54, 2)
c'est pas en une heure et demie qu'il peut changer c'est en cinq heures (C89, 137, 12)

C'est le verbe de dispositif *c'est* qui s'adjoint les marques négatives : (*ne*)... *pas*. Ce type de négation est à assimiler aux modalités de terme.

Plusieurs types sont à distinguer :

a) Le premier type concerne les énoncés qui réalisent deux fois l'élément d'extraction *c'est*, et dont l'extraction est réalisée une fois sous forme complète, la seconde sous forme courte[9]. Le contraste de modalité peut aller du négatif vers le positif, ou du positif vers le négatif, soit :

c'est pas X [forme longue], c'est Y [forme courte] :

c'est pas lui qui m'a répondu c'est monsieur (C3, E54, 5)
c'est pas moi qui l'ai fait c'est c'est un un copain (Arlaud, C62, 1, 2)

c'est X [forme longue], c'est pas Y [forme courte] :

c'est elle qui a passé l'examen c'est pas moi (Castagne, C63, 4, 16)
c'est vous qui instituez des gosses c'est pas les femmes (C15, 155, 8)

Les données attestent que la structure qui domine très nettement est celle qui présente en premier le terme négatif de l'opposition : les énoncés qui placent l'élément positif en tête sont quasiment absents du corpus.

b) Dans un second type, très rarement attesté à l'oral, la forme courte de l'extraction se trouve réalisée en premier lieu : c'est pas X [forme courte], c'est Y [forme longue] :

c'est pas l'encadrement c'est pas ce petit encadrement c'est la la partie euh mur quoi de la porte qui était habillée de bois (Roméas, C42, 27, 10)

---

9 Rappelons que les énoncés qui comportent deux constructions verbales successives *c'était pas moi qui le qui le dirigeais c'était lui qui me qui me dirigeait* (Mazo, C55, 12, 7) sont exclus du champ de cette étude. L'opposition entre formes longues et formes courtes de dispositifs est due à Blanche-Benveniste (1992).

c) Le troisième type comprend les énoncés qui ne réalisent qu'un seul *c'est* de dispositif (toujours sous forme longue) : soit la partie positive du contraste est réalisée en premier lieu, comme dans :

> c'est ceux de derrière qui sont bons pas ceux de devant (oral)

Soit c'est la partie négative qui figure en tête :

> ce ne sont plus les pays qui intéressent mais les gens (C13, 231, 19)

A la différence de l'oral, la langue écrite cérémonieuse semble préférer cette structure à celle qui fait usage de la forme courte de l'extraction (type 1), peut-être en raison du fait que l'école nous a enseignés qu'il fallait faire « des phrases complètes » pour bien écrire, et qu'en outre, la forme *c'est* est réputée manquer d'élégance, *a fortiori* si elle apparaît deux fois dans le même énoncé.

d) Un dernier type apparaît très rare : les éléments du contraste sont tous les deux disposés sous forme de liste paradigmatique avant le verbe recteur :

> Ce n'est donc pas le décalage horaire mais l'angoisse des dépenses s'ajoutant à celle de plonger dans cette ville phénomène qui perturbe la première nuit à Tokyo. (presse)

Les exemples de contrastes qui s'organisent autour de l'extraction sont assez nombreux dans le corpus pour que l'on puisse dégager quelques tendances :

- Les deux types de contraste à extraction qui apparaissent le plus rarement dans le corpus ont une caractéristique commune : l'élément positif et négatif du contraste sont tous deux placés avant leur verbe recteur (Vr) :

> c'est [lui] et [pas elle] qui **viendra**
> Vr
> c'est [lui] mais c'est [pas elle] qui **viendra**
> Vr

Il semble que le fait d'exprimer conjointement chacun des termes du contraste avant le verbe soit un procédé très peu usité, auquel les locuteurs n'ont pas spontanément recours.

A l'inverse, les contrastes de modalité que l'on trouve le plus fréquemment associés à l'extraction ont en commun la distribution suivante : l'un des éléments du contraste, puis le verbe recteur, et enfin le second élément du contraste :

c'est [lui] qui **viendra** c'est [pas elle]
Vr
c'est [pas lui] qui **viendra** c'est [elle]
Vr

- Tous types de contrastes à extraction confondus, la tendance est de placer le terme négatif en tête. Les exemples écrits trouvés dans la presse paraissent d'ailleurs confirmer cette tendance.

- La dernière remarque concerne la fonction syntaxique que remplissent les éléments du contraste : les sujets y sont très fortement majoritaires. Ce fait est d'autant plus remarquable que, comme on l'a rappelé *supra*, l'expression d'un contraste de modalité sur la position sujet connaît de sévères restrictions par ailleurs.

  Plus généralement, il semble que l'expression du contraste par le biais de l'extraction concerne presque exclusivement les éléments nettement valenciels.

2.1.4. Les contrastes qui font intervenir un dispositif pseudo-clivé

Les énoncés de ce type restent beaucoup moins nombreux que ceux qui utilisent l'extraction :

> ceux qui ont des stratégies ce ne sont plus les individus ce sont les gènes (Messaoudi, C42, 7, 8)
> on a un peu l'impression que ceux qui peuvent perdre quelque chose ce sont les occidentaux et certainement pas les soviétiques dans cette rencontre (Grigoriadi, C41, 10, 4)

Comme on l'a noté à propos des contrastes bâtis sur l'extraction, on relève dans le cas du pseudo-clivage la prééminence des contrastes sur les sujets.

2.1.5. Les contrastes qui font intervenir une modalité « autonome »[10]

Dans ces structures, l'élément du contraste réalisé en toute fin de construction est toujours suivi par un élément de modalité, positif (*oui*, *si*) ou négatif (*pas*, *non*)[11] :

> De baths, elles n'en ont pas vu la couleur, mais des coups, ça oui. Tous les jours (M.Cl, 12 / 89)

---

[10] Une analyse de ces énoncés a été proposée dans Bilger (1984) et Blanche-Benveniste *et al.* (1990).

[11] On trouve, de manière très minoritaire, d'autres modalités comme *jamais* ou encore des modalités développées sous forme de Groupes Nominaux : *pas un mot*.

L1: mais on (n')y produit plus de café alors
L2: du café si un tout petit peu (Ferrulo, C71, 9, 10)
les irlandais ils ont l'habitude de ce genre de choses les roumains pas (radio)
je suis rentrée hier soir de Lisbonne et ma valise non (oral)

Ces structures interdisent l'usage du *c'est* extractif :

*Paul est venu c'est Marie non
*Paul est venu Marie c'est non

Les énoncés en *non* sont clairement plus fréquents que ceux en *pas*. Les éléments analysables comme des sujets dominent nettement, particulièrement avec la modalité négative.

A nouveau, il est intéressant de rappeler la relative rareté à l'oral des énoncés comme :

certains sont plus traditionalistes mais pas l'ensemble (Balian, C47, 8, 8)

comparé aux structures à « modalité autonome », dont il est très facile de trouver des exemples oraux attestés :

certains les ont bien surmontés d'autres non (Bonnet, C62, 11, 9)

2.2. L'usage des connecteurs[12]

La possibilité de trouver tel ou tel connecteur dans les énoncés contrastifs ne dépend pas exclusivement de la distinction entre modalité de terme et modalité de relation ; d'autres facteurs tels que la présence d'un dispositif ou encore l'ordre linéaire des termes du contraste interviennent également. Malgré la pluralité des paramètres qui président à la sélection des connecteurs, il semble que leur distribution soit en partie prédictible. Nous en donnons un aperçu en dégageant plusieurs paradigmes calculés à partir des possibilités de commutation observées.

Avant de donner le détail des paradigmes, on peut faire une double remarque : *mais* semble toujours pouvoir être utilisé sans contrevenir à la grammaticalité des énoncés contrastifs (ce qui ne veut pas dire que son occurrence soit toujours très naturelle, comme on le verra plus bas). On relève en outre une relation d'implication dans l'usage des connecteurs : dans tous les cas où *et* est possible, *ø* semble l'être également ; par exemple,

c'est toi qui as fait ça **et** pas lui
c'est toi qui as fait ça **ø** pas lui

[12] On laisse de côté les structures les moins fréquentes, pour lesquelles on a trop peu d'attestations.

En revanche, *ø* peut être possible sans que *et* le soit :

c'est pas Paul qui travaille à Aix ø c'est Michel
*c'est pas Paul qui travaille à Aix et c'est Michel

L'étude de la distribution des connecteurs dans les énoncés à contraste amène à distinguer les paradigmes suivants :

2.2.1. Paradigme [*mais / et / ø*]

Ce paradigme est attesté pour les oppositions bâties sur une modalité de terme dans lesquelles l'élément positif est placé en tête.

L'exemple le plus clair est celui des énoncés sous dispositif direct :

Paul travaille à Aix **mais / et / ø** pas à Paris
Paul travaille à Aix **mais / et / ø** pas Michel

Cependant, dès lors que les énoncés apparaissent sous extraction, *ø* devient de loin le plus fréquent dans nos relevés, alors que *mais* semble utilisé de manière extrêmement marginale :

c'est Paul qui travaille à Aix **ø / et / (mais)** pas Michel
c'est Paul qui travaille à Aix **ø / et / (mais)** c'est pas Michel

2.2.2. Paradigme étendu

Dans le cas des contrastes qui font intervenir une modalité autonome en fin de construction, le paradigme des connecteurs paraît pouvoir s'étendre à des éléments comme *tandis que, alors que, en revanche...* :

Paul travaille à Aix **mais / et / ø / tandis que** Michel non
Paul ne travaille pas à Paris **mais / et / ø / tandis que** Michel oui

2.2.3.Paradigme restreint à [*mais*]

Mais apparaît invariablement dans les contrastes dont le terme négatif précède le terme positif, en l'absence du marqueur de dispositif *c'est*.

Dans nos exemples, les énoncés bâtis sur une modalité de relation font systématiquement usage de *mais*[13] ; comme le note Muller (1991), la non réalisation du connecteur se traduirait ici par une totale absence de démarcation entre le terme négatif et le terme positif du contraste :

Paul ne travaille pas à Aix **mais** à Paris
Paul travaille non pas à Aix **mais** à Paris

[13] Nous n'irons pas toutefois jusqu'à considérer qu'un énoncé comme *je bois pas de l'eau, du vin* est à exclure comme agrammatical : on se borne à constater son absence dans le corpus.

C'est encore *mais* que l'on trouve dans les contrastes [-+] dont le premier terme est placé sous extraction :

c'est pas Michel **mais** Paul qui travaille à Paris
c'est pas à Paris qu'il vit **mais** à Aix

2.2.4.Paradigme [*mais /ø*]

Ce paradigme est compatible avec les énoncés dont le terme négatif apparaît avant le terme positif, lorsque ce dernier est accompagné du marqueur de dispositif *c'est* ; l'usage de *ø* domine très nettement dans nos relevés :

c'est pas Paul qui travaille à Aix **ø** / (**mais**) c'est Michel
celui qui travaille à Aix c'est pas Paul **ø** / (**mais**) c'est Michel

2.3. Synthèse

Nous résumons dans le tableau suivant les principales observations qui précèdent[14].

	Place de rection affectée par le contraste contraintes observées	**Type de connecteurs relevés**
Modalité de relation *Il ne travaille pas à Paris mais à Aix* Modalité de terme – [- +] *il travaille non pas à Paris mais à Aix*	*Sujet antéposé ? Sujet postposé	[*mais*]
Modalité de terme – [+ -] *il travaille à Aix mais pas à Paris*	Rejet à droite de la partie négative du sujet	[*mais, et, Ø*]
Modalité de terme – 1 seul terme réalisé *il travaille mais à Aix* *il travaille mais pas à Paris*	*sujet *compléments valentiels à réalisation obligatoire	[*mais*]
Dispositif d'extraction [-+] *c'est pas à Paris qu'il travaille, c'est à Aix*	sujets fréquents	[*mais, Ø*] : *Ø* majoritaire *mais* très minoritaire

[14] Les structures qui apparaissent très peu fréquentes dans notre corpus ne sont pas mentionnées. Le signe * signale les structures présumées impossibles, et ? les structures non représentées dans le corpus.

Dispositif d'extraction [-+] *c'est pas à Paris qu'il travaille, mais à Aix*	Sujets fréquents	[*mais*]
Dispositif d'extraction [+-] *c'est à Aix qu'il travaille, c'est pas à Paris*	Sujets fréquents ces structures sont moins fréquentes que celles qui présentent l'ordre [-+]	[*mais*, *et*, *Ø*] : *Ø* majoritaire ; *mais* très minoritaire
Dispositif d'extraction [+-] *c'est à Aix qu'il travaille, pas à Paris*		
contraste à modalité autonome *Damien travaille à Aix, mais Paul non*	Sujets fréquents	Paradigme de connecteurs étendu

3. Organisation micro- et macro-syntaxique dans les constructions à contraste

Les observations qui précèdent seront à présent interprétées selon la distinction théorique entre micro-syntaxe et macro-syntaxe[15].

La micro-syntaxe est une grammaire de dépendance qui associe une structure rectionnelle (sujet, objet) à des unités telles que le nom ou le verbe, sur la base des propriétés constructrices des catégories. L'unité d'extension maximale qui puisse être dégagée à ce niveau est le groupe de construction.

Le niveau macro-syntaxique, lui, rend compte de l'organisation des énoncés en « grandes unités » désignées « unités macro-syntaxiques ». Pour Blanche-Benveniste *et al.* (1990), ces unités pourraient être présentées comme des :

> « unité(s) globale(s) apparentée(s) à ce qu'on appelle la « phrase » ou supérieure au niveau de la phrase » (p. 290).

C'est sur la notion d'unité macro-syntaxique que nous souhaiterions revenir à présent : nous ferons l'hypothèse que les énoncés à contraste posent de manière cruciale la question de l'extension de ces unités. Ainsi, certaines constructions à contraste semblent s'organiser en une seule unité macro-syntaxique alors que d'autres se présentent comme une succession de plusieurs unités.

15 Blanche-Benveniste *et al.* (1990).

3.1. Organisations en deux unités macro-syntaxiques

Parmi les énoncés qui présentent une opposition de modalité, certains semblent pouvoir s'analyser comme une séquence de deux Unités macro-syntaxiques distinctes.

C'est le cas des contrastes bâtis sur une modalité de terme en dispositif direct de type [+-], de la plupart des énoncés qui s'organisent autour de l'extraction ou du pseudo-clivage, ou encore de ceux qui réalisent une modalité autonome en fin de construction.

Toutes ces structures partagent en effet les caractéristiques suivantes :

3.1.1. Réalisation de manière non connexe

Les deux éléments du contraste sont susceptibles d'être réalisés de manière non connexe. Rappelons que, dans le cadre de l'approche pronominale, on parle de « liste paradigmatique » pour décrire le fait qu'une même position syntaxique est réalisée par plus d'un constituant : ainsi dans

ils connaissent Paul et Sophie

la suite *Paul et Sophie* est une liste, composée de deux groupes nominaux, qui instancie la position objet du verbe recteur. Le caractère paradigmatique de la liste est clairement mis en évidence si l'on adopte une disposition verticale :

ils connaissent Paul
et Sophie

La présence d'un contraste de modalité ne change rien à l'analyse : c'est encore l'axe paradigmatique qui est à l'œuvre dans :

ils connaissent Paul
mais pas Sophie

Comme on le voit avec les deux exemples précédents, les listes paradigmatiques les plus simples se caractérisent par leur caractère connexe, puisque les termes listés, réalisés de manière contiguë, y forment une séquence linéaire.

Dans le cas des structures ici décrites, les listes peuvent apparaître de manière non connexe ; en d'autres termes, chacun des deux termes listés se réalise de part et d'autre du verbe constructeur :
dans le cas du contraste sur un terme de type [+ -], la non connexité s'observe essentiellement sur les sujets, forcément réalisés à distance l'un de l'autre :

certains sont plus traditionalistes mais **pas l'ensemble** (Balian, C47, 8, 8)
ces gens-là ont commencé à nous raconter **pas tous** (Bonnet, C62, 10, 1)

Dans le cas de l'extraction, on a relevé plus haut que les listes connexes (du type *c'est pas lui mais c'est elle qui viendra*) paraissent très peu fréquentes dans les productions orales ; les listes non connexes sont la règle :

c'est **pas lui**	qui m'a **répondu**	c'est **monsieur**
c'est **lui**	qui m'a **répondu**	c'est **pas monsieur**
c'est **lui**	qui m'a répondu	**pas monsieur**
1er élément du contraste	Verbe recteur	2e élément du contraste

Le même type de distribution s'observe bien sûr pour les listes qui instancient une position de complément :

c'est pas demain qu'il viendra c'est la semaine prochaine

C'est encore la distribution non connexe qu'on observe dans les contrastes à modalité autonome, qui concernent très fréquemment les sujets[16] :

Bruno	doit être parti	mais **Indurain** pas
je	suis rentrée hier soir de Lisbonne	et **ma valise** non
1er Verbe		2ième élément

3.1.2. Sélection des connecteurs

Le second indice qui nous conduit vers une interprétation en deux unités macro-syntaxiques concerne la sélection des connecteurs ; on a relevé *supra* que dans toutes les structures dont il est question, *et* et *ø* sont tous deux présents dans le paradigme des formes possibles :

il était bon pour rester à l'arsenal [**mais / et / ø**] pas pour aller en mer (Scappini, C49, 13, 9)

A cet égard, on a vu plus haut que les énoncés à extractions font majoritairement recours à *ø*, au détriment de *mais*.

[16] Ces constructions à modalité autonome introduisent un degré de complexité supérieur aux autres structures décrites, car l'unité macro-syntaxique finale doit être analysée comme une séquence de deux éléments : préfixe + noyau ; ex.: *Paul* [préfixe] *non* [noyau] (voir Blanche-Benveniste *et al.*, 1990, pour une analyse détaillée). Ce phénomène de « double projection » paradigmatique, ajouté aux remarques faites *supra* sur la compatibilité de ces constructions avec des connecteurs comme *en revanche*, illustre le statut syntaxique particulier de ces énoncés à contrastes.

Dans le cas des contrastes qui font intervenir une modalité autonome, on a décrit un paradigme de connecteurs « étendu » ; on relève par exemple l'usage de *tandis que*, dans :

il va pas être sans arrêt (...) à essayer de mordre quelqu'un **tandis que** le berger allemand oui (Gaucel, C72, 8, 13)

La possibilité de trouver ce type de connecteur, qui a la réputation de pouvoir réunir deux « phrases » verbales, est un bon indice de la forte disjonction macro-syntaxique qui semble exister entre les deux séquences.

3.1.3.Usage de la ponctuation

Troisième caractéristique, la segmentation forte par la ponctuation est fréquemment attestée dans les énoncés écrits ; voici quelques exemples recueillis dans la presse, dont chaque séquence de ponctuation est disposée entre crochets :

[Ils « tiennent les murs » dans la rue.] [Pas les filles.]
[Bien des amis auraient rompu pour moins que cela.] [Eux non.]
[Voyez-vous, depuis toujours, ce sont les filles cadettes qui ont la charge de leurs parents jusqu'à leur mort.] [Pas les garçons (...)]
[Mais ce n'est pas cet aspect de la question qui nourrit la polémique ou déclenche les railleries :] [c'est le fait d'affirmer que des cinéastes en pleine activité, comme l'étaient Glauber, Rocha et Léon Hirszmann, sont morts d'une sorte de dégoût, d'une mélancolie, provoqués par la situation culturelle de leur pays.]
[En privé, le jeune Turc de l'Elysée a défini ainsi sa mission :] [« combattre l'idée que le Premier ministre est inamovible ;] [mais en douceur, car il est très susceptible »]

L'observation de séquences conversationnelles offre une illustration supplémentaire : il apparaît que les deux éléments de la liste ne sont pas toujours produits par le même locuteur de manière enchaînée ; dans cet énoncé par exemple, un premier locuteur produit la partie positive de la liste, et un second la partie négative :

L1: c'est en bois
L2: ouais ouais – pas en fer hein (Tridon, Lic91, 4, 11)

Parfois, la liste est donnée dans son entier par le même locuteur mais à l'intérieur de deux tours de parole différents :

L1 : où il faut pas de *e* j'en mets beaucoup il m'arrive
L2 : en parlant **on** les prononce
L1 : dans ma lancée je
L2 : tandis que **eux** non (C15, 74, 12)

L1: **avant** c'était plus tranquille avant c'était plus familial c'était
L2 : tout le monde y allait pour s'amuser
L1: le père la mère tout ça avec les fille tout ça tout ça hé allait au bal
L2: **maintenant** non [cité par Blanche-Benveniste]

On conclut que dans ces constructions verbales, la liste paradigmatique qui exprime le contraste se réalise comme une succession de deux unités macro-syntaxiques :

il a travaillé à Paris	pas à Aix
il a travaillé	mais pas à Paris
c'est pas à Paris qu'il a travaillé	c'est à Aix
c'est à Aix qu'il a travaillé	c'est pas à Paris
c'est à Aix qu'il a travaillé	pas à Paris
Damien a travaillé à Aix	Paul non
Unité MS	**Unité MS**

3.2. Organisations en une seule unité macro-syntaxique

Les autres types d'énoncés contrastifs ne présentent pas les propriétés de disjonction macro-syntaxique que l'on vient de relever.

3.2.1.Connexité

Contrairement aux énoncés que l'on propose d'analyser comme une séquence de deux unités macro-syntaxiques, certaines structures contrastives n'admettent pas la disposition non connexe des termes listés : il en va ainsi des contrastes bâtis sur une modalité de relation, pour lesquels les deux éléments du contraste doivent former une liste connexe (dont les deux termes sont produits en succession), toujours située après le verbe recteur. Cette contrainte de connexité post-verbale est particulièrement nette dans le cas des contrastes sur la position sujet. Impossible en effet d'accepter une tournure comme

*Victor ne viendra pas mais Marie

qui contrevient à ce principe de connexité, puisque les deux candidats au contraste se trouvent séparés par le verbe recteur ; tout aussi inacceptable est l'énoncé

*Victor mais Marie ne viendra pas

qui contrevient à la nécessité de la position post-verbale.

Pour prendre un exemple de contraste sur un complément, on peut aisément passer de :

il n'a pas parlé à Paul mais à Marie

à la version pronominale

? il n'a pas parlé à lui mais à elle

mais on refusera la réalisation sous forme clitique, qui voudrait établir un contraste entre un élément situé avant le verbe (clitique) et un second élément placé à droite (non clitique).

*il ne lui a pas parlé mais à elle.

Comme on l'a vu plus haut, il suffirait que l'on adopte une structure contrastive compatible avec une présentation non connexe des termes de la liste pour que l'opposition entre un clitique et un élément non clitique soit recevable ; ce serait le cas si l'on établissait un contraste à modalité autonome en fin de construction, analysable comme une suite de deux unités de macro-syntaxe :

il ne lui a pas parlé mais elle si
il ne lui a pas parlé mais à elle oui

Enfin, pour des raisons similaires, on ne peut pas passer de

il n'a pas plu à Aix mais à Paris

à

*à Aix il n'a pas plu mais à Paris

Ce type d'impossibilité est d'autant plus remarquable que les compléments de type locatif ont la réputation d'être assez peu contraints au plan linéaire. Or dans ce tout dernier exemple, l'antéposition du terme positif du contraste, en conférant à la liste son caractère non connexe, crée une structure totalement agrammaticale.

3.2.2.Connecteurs

La seconde observation concerne l'usage du connecteur : contrairement aux structures compatibles avec chacun des trois connecteurs (*et* / *mais* / *ø*), certaines constructions appellent invariablement l'usage de *mais* : les contrastes construits sur une modalité de relation relèvent de ce type, puisqu'il paraît exclu de passer de :

ce texte n'est pas l'œuvre d'une personne **mais** l'œuvre d'un certain nombre de gens

à

*ce texte n'est pas l'œuvre d'une personne **ø** l'œuvre d'un certain nombre de gens
*ce texte n'est pas l'œuvre d'une personne **et** l'œuvre d'un certain nombre de gens

Il en va de même des contrastes sur un terme de type [-+], comme

tout cela est vécu non par les possédants mais par le plus grand nombre

3.2.3. Ponctuation

Enfin, les structures à l'étude semblent assez mal se prêter à la disjonction au moyen de la ponctuation ; on ne peut guère imaginer l'usage d'un point, dans l'exemple :

? Ce texte n'est pas l'œuvre de d'une personne. Mais l'œuvre d'un certain nombre de gens.
? Tout cela est vécu non par les possédants. Mais par le plus grand nombre.

Même si des études précises restent nécessaires, on peut penser que ce type d'énoncés à contrastes seraient invariablement produits en une seule séquence prosodique.

L'ensemble de ces propriétés (connexité obligatoire de la liste, expression nécessaire du contraste au moyen de *mais*, segmentation difficile en plusieurs blocs graphiques ou prosodiques) nous amènera à conclure que, dans ce type de construction, la totalité des éléments de la liste fonctionne au plan macro-syntaxique comme une unité cohésive, dont les deux éléments ne peuvent pas être disjoints. On interprétera donc comme formant une unité macro-syntaxique unique, les listes contrastives présentes dans :

- les énoncés qui présentent un contraste de modalité de relation ; par exemple :

il ne le prend pas une fois de temps en temps mais tous les jours
Unité Macro-Syntaxique

- les énoncés fondés sur une modalité de terme de type [-+] qui présentent la suite (*non*) *pas*... *mais* :

Tout cela est vécu non par les possédants mais par le plus grand nombre
Unité Macro-Syntaxique

4. Conclusion

Cette étude a permis de souligner l'existence de deux types macro-syntaxiques liés à l'expression du contraste de modalité : le premier type a pour caractéristique de réunir les deux termes du contraste au sein d'une unité macro-syntaxique unique ; l'autre type réalise chacun des termes du contraste au sein d'une unité macro-syntaxique spécifique.

On propose de recourir à la notion de segmentation macro-syntaxique pour caractériser les constructions syntaxiques dont tous les éléments régis ne sont pas réalisés à l'intérieur d'une même unité macro-syntaxique.

Ainsi, il y a segmentation macro-syntaxique en trois unités successives dans :

il est venu pas hier avant hier : [il est venu] [pas hier] [avant hier].

mais il n'y a pas de segmentation dans :

il n'est pas venu hier mais avant-hier

ni dans :

il est venu non pas hier mais avant-hier

De même, on décrira un fait de segmentation macro-syntaxique dans

[c'est pas Paul qui est venu] [c'est Damien]

mais non dans

c'est pas Paul mais Damien qui est venu.

Dans certains cas, la segmentation macro-syntaxique peut être conçue comme une « option » de l'énonciateur, qui peut choisir de segmenter en plusieurs groupes des éléments qui auraient pu être produits dans une séquence macro-syntaxique unique, comme le fait de passer de

[elle ne viendra pas demain]

à

[elle viendra] [mais pas demain] (énoncé segmenté).

Dans d'autres cas en revanche, la segmentation macro-syntaxique apparaît davantage comme une contrainte de langue ; par exemple, en dispositif direct, il n'est possible de réaliser un contraste sur le sujet que de manière segmentée :

[Marie est venue] [et pas Paul]

De fait, l'opposition macro-syntaxique entre construction segmentée et non segmentée semble particulièrement opératoire pour décrire les contrastes qui mettent en jeu les syntagmes à fonction sujet : on pourrait

certes décrire ces constructions, et les contraintes qu'elles subissent, en termes purement micro-syntaxiques, en évoquant par exemple la nécessité d'une postposition nécessaire dans le cas des listes bâties sur l'élément sujet. Mais la notion macro-syntaxique de « segmentation » nous semble plus puissante au plan de la description, puisqu'elle permet d'appréhender du même coup certains faits liés à l'usage des connecteurs, ou à la structuration prosodique en plusieurs tronçons successifs.

Enfin, la notion de segmentation macro-syntaxique permet de décrire l'organisation des constructions verbales sans avoir à recourir à la notion de phrase, concept dont on connaît le caractère composite et – tout particulièrement dans le cas des énoncés oraux – peu opératoire[17].

Références

Béguelin, M.-J. (éd) (2000). *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. Bruxelles : De Boeck – Duculot.

Berrendonner, A. (1993). La phrase et les articulations du discours, *Le français dans le monde, février / mars*, 20-26.

Bilger, M. (1984). ET, quoi de neuf ?, *Recherches sur le français parlé, 6*, 82-107.

Bilger M. (1998). Le statut micro et macro-syntaxique de Et. In M. Bilger, K. van den Eynde & Françoise Gadet (Eds), *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp. 91-102). Leuven-Paris : Peeters.

Blanche-Benveniste, C. (1992). A propos des énoncés sans verbes : les énoncés réponses, *Recherches sur le français parlé, 11*, 57-85.

Blanche-Benveniste, C. & Jeanjean, C. (1987). *Le français parlé. Édition et transcription*. Paris : Didier-Erudition.

Blanche-Benveniste C., Bilger M., Rouget C. & van Den Eynde K. (1990). *Le français parlé. Études grammaticales*. Paris : Édition du CNRS.

Cappeau, P. & Bilger, M. (1995). J'ai une douleur dans la cuisse mais pas là. Analyse d'un cas de contraste, *Recherches sur le français parlé, 13*, 33-43.

Jeanjean, C. (1980). *Les formes sujet de type nominal. Étude sur le français contemporain*. Thèse de troisième cycle, Université de Provence, Aix-en-Provence.

Muller, C. (1991). *La négation en français. Syntaxe, sémantique et éléments de comparaison avec les autres langues romanes*. Genève : Droz.

[17] Blanche-Benveniste & JeanJean, (1987) ; Berrendonner (1993) ; Béguelin (2000).

Les amorces de mots comme faits autonymiques en langage oral

Berthille PALLAUD
Université de Provence & CNRS

Résumé

Les amorces ou fragments de mots sont des événements langagiers d'une grande fréquence à l'oral. Elles sont facilement repérables dans les corpus oraux de toutes origines. Qu'elles soient ou non accompagnées de corrections ou qu'elles soient simplement « complétées » par le locuteur, elles constituent une forme linguistique orale du fait autonymique. Leur présence témoigne d'une réflexion ou d'une activité linguistique se traduisant par une interruption de morphèmes en cours d'énonciation. On a vu (Pallaud, 1999) que ces phénomènes constituent plus de la moitié des achoppements de langage identifiés dans un corpus enregistré de français contemporain. Lorsque les conventions de transcription prévoient de conserver ces traces d'élaboration des énoncés que sont les amorces, les nouvelles technologies permettent un recensement suffisamment large pour qu'il soit possible d'en proposer, pour la première fois, une description raisonnée. L'étude entreprise porte sur le corpus de français parlé (Corpaix) recueilli depuis 25 ans par le G.A.R.S. et se propose, en particulier, de vérifier la pertinence du classement en trois catégories des amorces identifiées (corrigées, complétées et inachevées) au regard de plusieurs paramètres comme la place syntaxique touchée par l'amorce et surtout le fonctionnement de correction ou de réparation (terme employé par Martinie, 2001) selon cette place syntaxique. Notamment, il apparaît que l'amorce n'est pas toujours le lieu même de la rectification mais plutôt l'indice d'une correction qui va affecter un élément antérieur de l'énoncé.

Recherches sur le français parlé, 17, 79-101.
©2002, *Université de Provence*

1. Introduction

Martinie (1999), à l'instar de Jeanjean (1984) regroupe sous le terme d'amorces les fragments de mots ou de syntagmes comme :

> et le bâtonnier se chargera de : attention parce que lui c'est un barbot je vous l'ai dit combien de fois (dans Martinie, 1999)

Dans notre étude, le terme d'amorce est conforme à l'emploi qui en est fait dans les conventions de transcription du Groupe Aixois de Recherche en Syntaxe (G.A.R.S.) reprises par l'équipe DELIC à Aix : il est réservé aux ratés dans la production de morphèmes (ce que Martinie appelle des fragments). Les fragments de syntagme sont plus souvent identifiés sous le terme de *bribes* (Cappeau, 1998).

Parmi les phénomènes de bribes dans le langage parlé, les amorces de mots constituent un ensemble de phénomènes univoques d'interruption de morphèmes dans les recueils de français parlé transcrits selon les conventions du G.A.R.S.. Elles sont en effet signalées dans le texte de transcription par une marque typographique : le trait d'union collé au fragment d'énoncé interrompu. Il n'est donc pas nécessaire d'attribuer à l'amorce une fonction ou un statut cognitif. Elle est identifiée d'abord par sa « matérialité » :

> un bon film s'il est bon il **mar-** eh ben il ne marchera pas (Radio, 02-04 1998, 144)

Dans la très grande majorité des cas, ces phénomènes étant identifiés comme des « réductions » de mots qui ne sont qu'amorcés, leur transcription est orthographique et atteste ainsi de cette interprétation. C'est en tout cas la règle de transcription qui est donnée :

- on transcrira : « c'est **mau-** c'est de très mauvais goût »
- et non : « c'est **mo-** c'est de très mauvais goût »

Il reste cependant que si cette interprétation est hautement probable, elle ne peut être posée avec certitude. Le locuteur avait peut-être commencé à dire ***c'est moche***, terme qu'il a refusé et interrompu. Côté transcripteur, les amorces demeurent une interprétation et côté locuteur, elles ne sont pas à l'abri des effets de « lapsus ». Une perspective qui adopte pour prémisse l'encadrement strict de la transcription par deux hypothèses selon laquelle il est fait crédit au locuteur qu'il est cohérent et que ses propos sont pertinents au regard de la réalité extérieure, tend à sous-estimer les phénomènes de lapsus c'est-à-dire ce qui n'est pas probable (Pallaud, 1999 et 2001).

L'analyse descriptive des amorces dans les énoncés suppose, ici, ces hypothèses de cohérence et congruence.

Lorsque le contexte ne permet pas le recours à la cohérence, ces fragments de morphèmes soulèvent des difficultés de reconnaissance. Parfois même, certaines amorces ne sont pas interprétables :

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

D'autres, au contraire, ne sont qu'un des choix possibles ce qui est explicité, dans les transcriptions du G.A.R.S., par le recours à une multi-transcription

> oui /**l'oeu-, le**/ l'oeuf se mélange entièrement (corpus Jacques, 35, 19)

Il reste qu'il est très souvent possible d'interpréter les amorces, ce qui explique la grande fréquence des transcriptions orthographiques proposées pour ces fragments de morphèmes. Mais, même dans les cas où on ne peut interpréter avec certitude toutes les amorces, leur place syntaxique est connue plus sûrement. comme le soulignait Jeanjean :

> « Pour assigner une valeur syntaxique à ce genre d'amorces, on n'a pas besoin de reconstituer un quelconque lexème » (1984 : 171).

2. Fréquences des amorces dans un recueil de corpus de 47 000 mots

Un recensement du nombre total des amorces sur un recueil de corpus de 46 649 mots[1] (Tableau 1) montre que ces phénomènes (197 au total) viennent loin derrière les pauses courtes silencieuses[2] ou remplies[3]. Les amorces sont six fois moins nombreuses.

---

[1] Cet ensemble est composé de 12 corpus recueillis par moi-même dans le cadre du Corpus de Référence de Français parlé (Pallaud, 2001, Tours).

[2] Les pauses silencieuses courtes d'une durée de 200 ms environ. Cette appréciation n'est qu'approximative dans ces corpus: des intervalles mélodiques ont pu être identifiés comme des pauses.

[3] Ces pauses s'accompagnent d'un *euh* « d'hésitation ».

	Nb d'événements	Nb de mots moyen entre deux éléments [4]
Pauses courtes	1123	41, 5
Pauses moyennes	48	971, 8
Pauses remplies *euh*	1151	40, 5
Amorces lexicales	**197**	**236, 8**

Tableau 1. Amorces lexicales et pauses (silencieuses et remplies) dans 12 corpus de français parlé.

Comme la taille des corpus varie d'un corpus à l'autre, une analyse des résultats corpus par corpus s'imposait. Elle a été faite sur *le nombre moyen de mots entre deux amorces*. Ce paramètre a été choisi car il permet d'obtenir facilement une valeur comparable d'un corpus à l'autre : le nombre d'événements (ici les amorces) tous les 1000 mots.

L'analyse des amorces selon les corpus (Tableau 2) montre une grande hétérogénéité entre les locuteurs puisque les valeurs peuvent varier d'un facteur 10. En termes de nombre d'amorces tous les 1000 mots, les valeurs extrêmes vont de 0, 9 à 9, 5 selon les locuteurs. En moyenne, on peut dire qu'on en trouve environs 4 tous les 1000 mots soit pour un débit moyen de 200 mots par minute une amorce toutes les 75 secondes.

Corpus	Nb moyen de mots entre deux amorces	Nb d'amorces tous les 1000 mots
1	284	3, 5
2	188	5, 3
3	814	1, 2
4	105	**9, 5**
5	327	3
6	158	6, 3
7	293	3, 4
8	301	3, 3
9	194	5, 15
10	154	6, 5
11	1128	**0, 9**
12	805	1, 2
Moyenne (écart-type)	395 (330, 11)	4, 11 (2, 6)

Tableau 2. Répartitions des amorces selon les corpus.

[4] Ce chiffre est obtenu par la formule suivante: Taille du corpus/ nombre d'éléments observés dans le corpus.

La variabilité observée dans la production des amorces suggère qu'elle reflète fortement une composante individuelle. Le phénomène d'amorce semble être une marque d'énonciation qui rend compte de caractéristiques subjectives. Il resterait à vérifier qu'à cette différence inter-locuteurs ne se rajoute pas une différence intra-locuteur. En d'autres termes, il reste à établir si l'amorce de mot est une composante stable du « style » langagier du locuteur. C'est l'hypothèse que suggèrent ces résultats mais elle reste à vérifier.

3. Place des amorces parmi les erreurs d'écoute

Le Projet « Corpus de Référence de Français Parlé[5] » confié à notre équipe prévoyait des enregistrements d'entretiens dans différentes régions de France. Ces enregistrements ont été transcrits par des transcripteurs confirmés et une partie a été vérifiée par moi-même. L'analyse porte sur les modifications que j'ai apportées à ces transcriptions (Pallaud, sous presse). Les erreurs d'écoute ont été relevées et analysées à partir de cet ensemble de 12 corpus (46 649 mots).

Les éléments modifiés lors de la vérification (les relances mises à part) sont au nombre de 487. Au total, 26 amorces lexicales seulement seront concernées dans ces modifications soit 5, 3% en moyenne des modifications apportées (tableau 3). La place des amorces dans ces modifications est donc faible. Peut-être est-ce dû au fait que les amorces apparaissent en des moments où la parole du locuteur trébuche (ce qui pourrait être signalé par d'autres indices énonciatifs). De ce fait, l'amorce réclame de la part du transcripteur une écoute centrée à la fois sur les phonèmes et les morphèmes. L'attention du transcripteur est donc alertée, ce qui pourrait expliquer le faible pourcentage de rectifications dans cette catégorie.

---

[5] Pour cette étude, l'analyse porte sur les modifications apportées lors de la phase de vérification à 12 corpus enregistrés en septembre 1999 dans le cadre du Projet Corpus de Référence de Français parlé conduit à Aix par l'équipe CNRS ESA 6060 puis par l'équipe DELIC. Cinq villes du Nord-Est, du Centre et du Sud-Ouest de la France sont concernées.

Catégories de modifications	Nb d'amorces			%	Total des éléments rectifiés6
	Ajout.	*Modif.*	*Suppr.*		
Élucidation d'éléments incompréhensibles	1	0	0	3, 3%	31
Éléments oubliés	7	0	0	3, 4%	203
Éléments en trop	0	0	1	3, 8%	26
Éléments remplacés	5	3	3	**8, 9%**	123
Multitranscriptions insérées	4	0	1	**15, 6%**	32
Multitranscriptions supprimées	0	0	1	1, 3%	72
Total	**17**	**3**	**6**	**5, 3%**	**487**

Tableau 3. Place des amorces parmi les différentes catégories de modifications proposées lors de la phase de vérification.

L'examen détaillé de ces modifications peut être envisagé sous l'angle de l'enrichissement ou de la réduction du texte initial.

La majorité des rectifications apportées au texte initial (360/581 soit 62%) contribue à enrichir le texte que ce soit :

- en élucidant des passages restés incompréhensibles pour le premier transcripteur (n=31) ;
- en rajoutant des éléments oubliés (relances, pauses, lexique) (n=297) ;
- en proposant des multitranscriptions supplémentaires (n=32).

La part de l'enrichissement du texte par des amorces dans ces secteurs est tout aussi importante si on tient compte de l'étude détaillée des 11 éléments remplacés et des 5 multitranscriptions proposées. Au total, 17 cas[7] de modification impliquant des amorces correspondent à un enrichissement soit 65%. Dans le domaine des amorces, les modifications se traduisent également par un enrichissement majoritaire du texte.

De plus, l'implication des amorces dans ces modifications n'est pas la même selon les catégories de modification. Si le phénomène d'amorce est peu concerné dans la catégorie « Multitranscriptions supprimées » (1, 3%), il en va tout autrement pour la catégorie « Multitranscriptions insérées » où les amorces sont concernées dans 15, 6% des 32 cas. Les chiffres sont cependant trop faibles pour qu'on puisse en tirer des conclusions argumentées.

---

6 Le total des rectifications, qui comprend aussi les relances, les pauses courtes et longues, est de 581 (Pallaud, 2001).

7 On a donc ajouté les cas d'amorces non perçues relevées dans les éléments remplacés et les multi-transcriptions proposées.

3.1. Parmi les éléments remplacés[8]

Dans 5 cas sur 11, l'élément proposé pour en remplacer un autre est une amorce non perçue lors de la transcription. Trois d'entre elles sont des solutions presque homophones.

- Amorces non perçues (n=5)

  des sculptures qui représentaient un peu (caricature, *caricatur-*) d'une manière un peu caricaturelle (Belfor5b, 2, 20)
  je (m'investissais, *m'investis- tissais*) beaucoup dans le milieu associatif (BordeauxC7b, 9, 5)
  parce que (les possibilités, *s'il y a des pos-*) il y a des possibilités (StrasbourgC2, 1, 12)

- Amorces non retenues (n=3)

  si je peux - (l-, *le*) le faire je le fais (TroyesPR101, 8, 12)
  je leur ai fait un proto (avec la fonctionna-, *et il a fonctionné*) (Stg41c, 6, 21)
  c'était génial (c-, *c'est*) il y a avait trois volets (TroyesPR102, 11, 1)

- Amorces différentes proposées (n=3)

  mesurer (l'hydro-, *l'hygro-*) (l(hydrométrie, *l'hygrométrie*) (Stg41c, 4, 12)
  vous aviez d'ailleurs (ab-, *al-*) (Belfor5a, 11, 4)
  il y a déjà trois (muci-, *munici-*) municipalités (BordeauxC7b, 7, 4)

3.2. Parmi les multi-transcriptions proposées

La majorité (4 cas sur 5) correspond à un enrichissement du texte par une amorce. Dans 3 cas sur 5, les multitranscriptions proposées sont des amorces homophones d'éléments présents dans la transcription.

- Amorces non perçues (n=4)

  vous en vous euh /au fond, *vous f-*/ dans une journée (Belfor5b, 7, 10)
  si /la, *l'a-*/ l'assistance sociale (Belfor5a, 6, 14)
  et /j'ai, *j'ai-*/ /j'ai, *j'ai-*/ j'aimais pas quoi (Belfor5a, 2, 16)
  ils emploient les mots /qui, *qu'i-*/ qu'il faut (Belfor5c, 4, 8)

- Amorces mises en concurrence (n=1)

  j'aime j'aimais bien passer des disques /*j'aim-*, mais/ j'adorais écouter des disques (BordeauxC7b, 2, 27)

---

8 La solution retenue est en italique.

4. Place des amorces parmi les erreurs de langage

Deux recueils de, respectivement, 120 et 100 erreurs de langage oral avaient été constitués. Selon que ces erreurs étaient recueillies « à la volée » ou relevés dans un corpus transcrit (corpus Bertuzzi 97).

Bien que, comme nous l'avons vu, la fréquence des amorces puisse varier beaucoup d'un corpus à l'autre, cette étude (Pallaud, 1999) avait montré le biais qui peut, vraisemblablement, être introduit par un recueil « à la volée »et la nécessité qu'il peut y avoir à bénéficier de corpus enregistrés si on veut étudier la place de ces phénomènes dans l'élaboration des énoncés oraux. Dans les recueils « à la volée », les erreurs du type « un mot pour un autre » sont, sans doute, très surestimées alors que les hésitations se traduisant par des amorces (amorces inachevées, corrigées et complétées) se trouvent sous-estimées. En effet, 59% des erreurs de langage dans le corpus enregistré (59/100) sont des amorces de mot alors que dans les recueils « à la volée » ce pourcentage tombe à 10, 8% (13/120).

On trouve trois sortes d'amorces (tableau 4) dont la distinction repose sur leurs effets au lieu syntaxique occupé par ces amorces. Y a-t-il réduction définitive sur cette place syntaxique auquel cas, l'amorce n'est pas complétée ni corrigée mais simplement laissée inachevée ? Y a-t-il piétinement sur la place syntaxique et une simple reprise et poursuite du morphème inachevé ; ou plutôt une correction lexicale ? Ces trois catégories d'amorces sont donc :

- amorce laissée inachevée
- amorce corrigée
- amorce complétée

Amorces	**recueil « à la volée »**	**corpus Bertuzzi 97**
amorces complétées	2 (5, 4%)	18 (30, 5%)
amorces inachevées	0 (0%)	10 (16, 9%)
amorces corrigées	11 (84, 6%)	31 (52, 5%)
Total	**13 (100%)**	**59 (100%)**

Tableau 4. Types d'amorces parmi les erreurs de langage notées « à la volée » (N=120) et relevées dans le corpus Bertuzzi 97 (N=100)

Ces résultats (tableau 4) permettaient la conclusion suivante : qu'il s'agisse d'un recueil « à la volée » ou d'amorces relevées dans un corpus, les amorces corrigées semblent les plus nombreuses. Nous verrons que cette répartition varie selon les corpus et la taille des corpus.

4.1. 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 :

mais si on fait une **e-** une **exception** pour lui (corpus Pallaud, 1997, 78)
pour tenir la **cou-** la **couronne** c'était pas drôle (Bertuzzi, 47, 8)

Elles ont été considérées, dans notre étude, comme des lapsus interrompus puis poursuivis. Deux exemples de ce type seulement (soit 5% des amorces) se trouvent dans les lapsus relevés « à la volée », alors qu'ils abondent dans les corpus enregistrés (18 sur.59 amorces soit presque un tiers).

4.2. 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 peuvent être considérées comme des lapsus interrompus et non corrigés :

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)

Dans le corpus « à la volée », on ne relève aucune amorce de ce type, alors que dans les corpus enregistrés, ces amorces inachevées constituent 17% des cas d'amorces (elles sont au nombre de 10).

4.3. 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 :

mais il y a eu des **palanq-** des **générations** de musiciens de jazz qui ont été sous son influence (corpus Pallaud, 1997, 30)
et deux fois par semaine + le **lun-** le **mercredi** et le vendredi il y avait les marchands + de plantes vertes (corpus Bertuzzi, 38, 16)

Dans l'exemple suivant, si la place syntaxique est conservée (une place d'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 de correction sur l'élément touché par l'amorce représentent la quasi totalité des amorces relevées « à la volée » alors que, dans le corpus enregistré, leur pourcentage n'est que légèrement supérieur à 50%

5. Les phénomènes d'amorces dans le recueil Corpaix[9]

L'étude entreprise sur le recueil de corpus Corpaix porte sur 1000 amorces. Nous n'avons pas retenu un certain nombre d'amorces. En effet, certaines d'entre elles tiennent à la situation d'interlocution puisqu'elle sont provoquées par l'interruption du deuxième locuteur (L2) :

> euh en dehors de de la pharmacie enfin L1 quels *prod-* L2 la parapharmacie L1 oui (Pharmaci 50, 2)

Ces amorces-là ne sont pas traitées dans cette analyse. Ont été écartées également les prononciations réduites du pronom *il(s)* dont le phonème [l] est si souvent élidé (Gadet, 1997) qu'il est impossible de déterminer avec certitude quand la prononciation réduite [i] correspond à une amorce du pronom *il(s)*.

Certaines amorces sont répétées par le locuteur mais elles ne sont comptabilisées qu'une seule fois dans l'étude :

> quand ma mère elle y allait moi *j'a- j- j-* je me faisais pas voir (Pavrayst 32, 1)

Ce total de 1000 amorces (tableau 5) a été identifié sur 60 corpus différents. On retrouve les mêmes types d'amorces distinguées lors de l'étude sur les erreurs de langage (Pallaud, 1999). Il s'agit des amorces complétées, inachevées et corrigées, telles qu'elles ont été définies précédemment.

---

9 Le recueil Corpaix auquel a été adjoint en 1997 un logiciel concordancier Xcor (produit par J-P. Adam) a été établi sur la base d'un ensemble de corpus recueillis dans le cadre du G.A.R.S. et de l'équipe ESA CNRS 6060. Cet ensemble, dans sa version d'alors, comporte des doublons. Si il a été possible de les repérer, il demeure impossible de référer le nombre d'amorces à la taille du recueil.

Parmi les nombreuses amorces inachevées (216), une partie non négligeable (17, 6% ; soit 38), n'a pu recevoir une interprétation et donc une catégorie grammaticale. La rupture syntaxique qui permet d'identifier une amorce inachevée est présente quand, après l'amorce, l'élaboration de l'énoncé sur l'axe syntagmatique se poursuit. Ce qui suit l'amorce est le début d'un nouveau syntagme nominal ou verbal comme dans les exemples suivants :

on a souvent *tendan-* à à analyser un peu les choses (Femmes 10, 9)
enfin je veux dire ça c'était *dif-* **et** il y avait entre (euh, eux) (Mmeduran 4, 3)
certaines oui quand ça concerne le *médica-* **par exemple** l'aspirine (Pharmaci 51, 9)

A l'inverse, les amorces corrigées nécessitent pour être reconnues que l'erreur de sélection se manifeste dans un piétinement sur la même place syntaxique. En particulier, elles se distinguent des phénomènes de coordination traditionnelle en ceci qu'elles « superposent » les éléments sur une même place là où la liste coordonnée énumère et « déroule » ses éléments[10]. Un exemple d'amorce corrigée est le suivant :

c'est un tare pour un *intellect-* euh pour quelqu'un qui écrit (Chrabi, 29, 6)

Un phénomène de coordination et donc d'amorce inachevée aurait été :

c'est un tare pour un *intellect-* **et** pour quelqu'un qui écrit

En l'absence de connecteurs et sans l'aide du contexte, le phénomène de liste coordonnée est impossible à établir : l'amorce corrigée est alors la seule hypothèse possible.

Amorces	**Nombre (Corpaix)**	**Pourcentage (Corpaix)**	**Corpus Bertuzzi 97**
Amorces complétées	607	59, 9%	30, 5%
Amorces inachevées	216	28, 8%	16, 9%
Amorces corrigées	177	11, 3%	52, 5%
Total	**1000**	**100%**	**100%**

Tableau 5. Répartition d'amorces relevées dans le recueil Corpaix et le corpus Bertuzzi97

La comparaison avec les amorces relevées dans un seul corpus (le corpus Bertuzzi97) montre qu'il y a probablement une grande variabilité d'un corpus à l'autre c'est-à-dire d'un locuteur à l'autre. En effet, sur un grand nombre de corpus (60 locuteurs au total), les amorces les plus

[10] L'analyse en grille proposée par le G.A.R.S. ne distingue pas ces deux types de phénomènes. La pluralité de réalisation lexicale (appelée *liste*) est représentée comme des alignements sur l'axe paradigmatique (Blanche-Benveniste, 1990 : 43).

nombreuses sont celles qui sont complétées, presque deux tiers, alors que, dans notre étude précédente, les amorces corrigées étaient les plus nombreuses (tableau 4). Viennent ensuite les amorces inachevées et corrigées qui représentent respectivement 29% et 11%).

On peut donc dire que, si majoritairement les locuteurs complètent ou corrigent leurs amorces lexicales, seules 3 amorces sur 10 sont laissées à l'état de fragment.

5.1. Amorces et constituants syntaxiques

Une étude a été entreprise pour évaluer si les trois types d'amorces (complétées, inachevées ou corrigées) touchaient de façon similaire, quant à la classe syntaxique, l'élément affecté. Des amorces ont été classées selon les quatre grands constituants[11] syntaxiques : Syntagmes sujet, objet, verbal, et associé (Tableau 6).

Constituants syntaxiques	Amorces complétées	Amorces Inachevées	Amorces corrigées	Total
Objet	306 (51%)	87 (48, 8%)	86 (48, 8%)	479 (50, 2%)
Verbe	188 (31, 3%)	74 (41, 6%)	73 (41, 5%)	335 (35, 1%)
Sujet	92 (15, 4%)	13 (7, 4%)	16 (9, 1%)	121 (12, 8%)
Associé	14 (2, 3%)	4 (2, 2%)	1 (0, 6%)	19 (1, 9%)
Total	**600 (100%)**	**178 (100%)**	**176 (100%)**	**954 (100%)**

Tableau 6. Places syntaxiques où se produisent les amorces

Un certain nombre d'amorces n'a pu être identifié (46 au total) et donc recevoir une appartenance syntaxique. C'est particulièrement le cas pour les amorces inachevées où cela se produit dans 17, 6% des cas (38/216). Définir le statut syntaxique d'un élément amorcé n'a pas été possible également dans quelques rares cas pour les amorces complétées (1, 1% ; 7/607) et un seul cas d'amorce corrigée (sur 177). Ces amorces dont le statut syntaxique n'a pu être précisé se réduisent toutes à un seul phonème. On observe cependant que ce n'est pas là la seule raison de cette absence d'identification. Les contextes antérieur et postérieur à une amorce contribuent fortement, sinon à l'interprétation de l'élément, du moins à l'établissement de sa catégorie grammaticale. Pour les amorces inachevées où, par définition, le contexte qui suit signale sinon toujours une rupture

[11] Concernant, les sujets et les compléments (Objet) les constituants correspondent ici aux « places de rection » (Blanche-Benveniste, 1990 : 42) que permettent d'identifier les pronoms.

syntaxique au moins une liste « coordonnée », ce qui précède l'amorce est alors primordial.

- Amorces complétées (Syntagmes nominal et verbal)

  si je dis les écrivains *m-* maghrébins ça les intéresse pas (Chrabi 42, 8)
  comme vous m ' *a-* vous m'avez pas dit (Maon 20, 9)

- Amorces inachevées (Syntagme verbal)

  elle lui a dit que *c'é-* petit à petit ils lui ont levé le travail (Benot 28, 9)
  quand j'ai *f-* euh enfin /le, un/ prof nous donne un exercice (Aidesd 3, 3)

Ce qui frappe au premier abord dans les résultats du tableau 6, c'est que la moitié des amorces, qu'elles soient complétées, inachevées ou corrigées, sont des éléments appartenant au constituant Objet (complément de verbe ou de nom). La catégorie Verbe regroupe 42% des amorces inachevées et corrigées, et 31% des amorces complétées. Pour ces dernières, la catégorie Sujet est plus sollicitée (15% contre 7 et 9% pour les autres amorces).

Les amorces corrigées et inachevées ont, donc, une répartition remarquablement semblable dans les divers constituants syntaxiques. Si on regroupe ces deux catégories et qu'on compare leur répartition à celle des amorces complétées, on constate qu'elles diffèrent significativement[12]. Comparativement aux amorces inachevées et corrigées, les amorces complétées concernent moins la catégorie du Verbe et plus celle du Sujet. Autrement dit, les amorces sur le Sujet seront plus souvent complétées tandis que sur le Verbe les amorces seront plus souvent inachevées ou corrigées. Dans la catégorie Objet, la répartition selon les catégories d'amorces est la même. Ces résultats sont encore plus nets si on considère les éléments à droite du verbe c'est-à-dire qu'on adjoint à la catégorie Objet, les verbes à l'infinitif, les participes passés et présents (Tableau 7).

---

12 Test du χ^2 ;χ^2=15, 94; p<0, 001.

Lieu de l'amorce	Amorces complétées	Amorces inachevées	Amorces corrigées	Total
Objet (moins les pronoms)	300 (51, 2%)	84 (48, 3%)	86 (49, 4%)	470 (50, 4%)
Éléments à la droite du verbe	75 (12, 8%)	29 (16, 6%)	25 (14, 3%)	129 (13, 8%)
Verbe	112 (19, 1%)	45 (25, 9%)	47 (25, 9%)	204 (21, 9%)
Éléments avant le verbe	98 (16, 7%)	16 (9, 2%)	16 (9, 2%)	130 (13, 9%)
Total	**585 (100%)**	**174 (100%)**	**174 (100%)**	**933 (100%)**

Tableau 7. Répartition des amorces avant ou après le verbe

Qu'ils soient complétés, inachevés ou corrigés, lorsque les éléments amorcés sont à la droite du verbe, leur pourcentage est le même : 63%. L'interruption a lieu majoritairement dans le contexte qui suit le verbe et les trois catégories s'y trouvent également représentées.

Une amorce sur 4 touche le verbe et sera corrigée ou restera inachevée. Les amorces complétées n'affectent un verbe que dans un cas sur 5. En revanche, elles concernent la catégorie sujet dans deux fois plus de cas que les amorces inachevées ou corrigées.

5.2. Les catégories grammaticales touchées par les amorces

Les catégories grammaticales des éléments amorcés sont très variées et la distribution des amorces n'est pas également répartie dans ces catégories. Quatre amorces sur cinq portent sur du lexique qu'il soit verbal ou nominal. Le reste des amorces affecte des « petits mots » tels que des déterminants, prépositions, relatifs, pronoms, négation.

Amorces	Nombre d'amorces	Pourcentage
Objet Nom	261	54, 5%
Objet adjectif et adverbe	140	29, 2%
Autre Objet	78	16, 3%
Total	**479**	**100%**
Verbe infinitif	79	23, 6%
Verbe participe passé et présent	52	15, 5%
Autre Verbe	196	58, 5%
Négation verbe	8	0, 2%
Total	**335**	**100%**
Sujet Nom et adjectif	39	32, 2%
Sujet pronom	63	52, 1%
Autre Sujet	19	15, 7%
Total	**121**	**100%**

Tableau 8. Répartition des amorces selon les catégories grammaticales

Cette grande prépondérance des amorces affectant le lexique se retrouve dans les catégories Objet (83, 7%) et surtout les Verbe (97, 6%). Ce n'est que dans les Sujets que cette prépondérance n'est plus retrouvée : seul le tiers des amorces concerne le lexique. La moitié des amorces sur des éléments sujets portent sur des pronoms.

Dans les catégories Objet, les amorces concernent pour moitié le lexique nominal et pour 30% à part égale des adverbes et des adjectifs.

Dans les catégories Verbe, les verbes infinitifs constituent un quart des amorces, les participes passé et présent (ce dernier très rare) 16%. Le reste, à part 6 cas de négation amorcée, est composé des verbes auxiliaires et verbes conjugués.

Dans les catégories Sujet, à part les pronoms qui constituent la moitié des amorces, 16% sont des « petits mots » (des déterminants ou des relatifs *qui, quel*) et le reste (le tiers) sont des sujets lexicaux (nom et adjectif).

Sur la totalité des amorces, celles qui concernent des « petits mots » ou des pronoms sont de loin les moins nombreuses (18% ; 168/925). Que cela soit dû à un lien entre la taille du mot et le phénomène d'amorce reste à vérifier. Cela suggère, en effet que les mots plurisyllabiques puissent être plus fréquemment l'objet d'achoppements que les mots monosyllabiques.

5.3. Comment les amorces sont-elles complétées ou corrigées ?

La plupart des amorces sont complétées ou corrigées après que le déterminant, le quantifieur, la préposition si c'est un nom ou le sujet (pronom, relatif) si c'est un verbe aient été **repris**. Le locuteur tend, en effet, à reprendre l'énoncé antérieur depuis le début du syntagme :

- Déterminants

  on dirait on dirait on dirait **une** mer- **une** merveille (Guarraci 41, 8)
  alors **l'**a- **l'**agent de police attend que le télégraphiste arrive (Pavrayst 11, 8)
  il y a **certains** é- + **certains** élèves qui montrent leur (s) note (s) (Aaidesd 8, 10)

- Préposition et déterminant

  ils font entre autre de comment dire de la **de l'**expériment- **de l'**expérimentation sur les logiciels (Exprienc 4, 15)

- Pronom sujet

  en lui disant combien **je** re- + **je** regrettais de lui avoir fait de la peine (Flicienm 6, 11)
  ça p- p- **ça** peut poser des problèmes avec la population (Voyageso 69, 9)

De façon générale, on constate que la reprise a lieu depuis le début du syntagme mais une étude plus approfondie serait nécessaire pour préciser la nature et la fréquence du début de syntagme dans ces reprises : la préposition, le déterminant, le connecteur ...

- L'élément repris peut même être répété

  pour que l'eau elle aille pas **dans** l- **dans dans** la propriété (Maon 3, 12)
  pendant trois mois dans une société **qui** utili- **qui qui** utilisait des sous-produits (Jeanne 1, 8)

- Certaines amorces, moins fréquentes, sont complétées après qu'une séquence plus longue soit **reprise** :

  il y en avait et dans la rue - **comme on l'**ap- **comme on l'**appelait la rue là (2) (Scalone2 13, 15)
  mais **il y a des** en- **il y a des** élèves qui ont plus ou moins de volonté (Aides 2, 3)

Il est moins courant qu'une amorce lorsqu'elle porte sur le lexique nominal (19 cas) ou verbal (4 cas de verbes avec sujet) soit complétée par reprise simple du début du lexique :

c'est en fonction du sujet L1 et des **interlo- interlocuteurs** aussi (Portugal 4, 15)

tu as une ordonnance avec des **mé- médicaments** (pharmaci 8, 10)
nous n'**av- avons** eu de de bons rapports + pas fréquents (Flicienm 6, 11)
on le **re- re- revendique** encore une fois sur le cahier (Navale 14, 11)
alors que ils **co- commencent** à s'engastrer les autres (Argota 34, 8)
donc là ça ça **co- correspond** à la compétence (Metall 15, 5)

Il est très rare qu'une amorce soit simplement complétée par la fin du lexique sans qu'il y ait au moins reprise du début du lexique. Seules 3 amorces sur les mille recensées, correspondent à ce cas. Nous avons choisi de les transcrire comme suit :

la caserne du boulevard de Strasbourg qui est dans une **an- euh -cienne** bâtisse (Birnie 2, 16)
oh + ça ne te dit **certai- + -nement** + rien + mais (Corse 28, 11)
il faut passer des **é- + -lectro** + + bien travailler à l'école (Puget-VI 2, 8)

5.4. Les phénomènes d'incises à l'occasion des amorces

Les amorces s'accompagnent parfois de phénomènes d'incises diverses (tableau 9). Cette sous-catégorie n'est pas négligeable puisqu'elle constitue plus de 41% des amorces complétées (254/607), le quart des amorces inachevées (56/216) et 16% des amorces corrigées (29/177). L'amorce interrompue s'accompagne de deux sortes de phénomènes. Tantôt, elle permet l'ajout d'éléments de façon parenthétique ; tantôt, elle s'accompagne d'une reprise d'éléments antérieurs qui seront alors modifiés. Parfois, les deux phénomènes se produisent simultanément. Dans tous ces cas, l'amorce n'est donc pas le lieu-cible de la réparation mais en constitue le premier pas. Elle anticipe le second pas au cours duquel seront ajoutés ou corrigés un ou plusieurs éléments.

Incises et modifications	**Amorces complétées**	**Amorces inachevées**	**Amorces corrigées**	**Total**
Incises énonciatives	100 39, 4%	56	27 93, 1%	183 54%
Incises qui modulent	81 31, 9%	0	2 6, 9%	83 24, 5%
Incises parenthétiques	29 11, 4%	0	0 0%	29 8, 5%
Modifications antérieures	44 17, 3%	0	0 0%	44 13%
Total	**254 100%**	**56**	**29 100%**	**339 100%**

Tableau 9. Incises et modifications du contexte antérieur aux amorces.

Cette sous-catégorie d'amorces est particulièrement la marque d'une activité autonymique puisque les amorces concernées ne sont pas le lieu d'une modification mais annoncent une ré-élaboration sur ce qui a été énoncé antérieurement.

5.4.1. Les éléments simplement ajoutés

Les éléments simplement ajoutés (incises énonciatives, parenthétiques ou qui modulent le texte) sont les plus nombreuses (87% ; 275/339) (tableau 9). Les travaux de Duez (1997) et d'Authiers-Revuz (1995) ont souligné la fonction énonciative de certains éléments présents dans les énoncés. Parmi ceux-ci figurent les pauses silencieuses et remplies ainsi que certains éléments « associés énonciatifs » [13] et connecteurs. Il semble utile de distinguer ces incises « énonciatives » des autres ajouts apportés à l'énoncé.

- Ces incises « énonciatives » peuvent être une répétition de l'amorce avant qu'elle ne soit complétée ou corrigée, une pause silencieuse brève ou remplie, ou un élément à fonctionnement associé comme *enfin, quoi*
  - Répétition de l'amorce

    liaison **hé- hél-** hélicoïdale (Metall 9, 8)
    pour euh décoder analyser les données **informat- in-** informatisées (Metall 2, 5)
  - Une pause brève ou remplie

    gagner cent dollars par un pour une semaine c'est ç- + ça veut rien dire (Voyageso 5, 11)
    deux ans de service huit mois d'occupa- **heu** d'occupation (1) (Scalone2 20, 13)
  - Ajout d'un élément à fonctionnement associé

    donc euh après euh les vacan- **enfin** vacances d'été je me suis euh retrouvé à la rue (Eexprienc 1, 8)
    à condi- + **ah oui** + mais c'était on chauffait pas comme maintenant (Alsacebc 21, 14)

  Ces incises énonciatives sont les plus nombreuses (54% ; 183/339) et constituent les seuls ajouts pour les amorces inachevées ainsi que la quasi totalité pour les amorces corrigées. L'interruption définitive d'un élément de l'énoncé est suivie parfois (dans 8% des cas ; 18sur 216) d'un connecteur, ce qui introduit un phénomène de liste. Ce n'est jamais le cas pour les amorces complétées ou corrigées. Dans le cas

[13] La liste est courte de ces éléments associés « énonciatifs » (« *L'interprétation porte sur l'attitude énonciative et non sur ce que dit l'énoncé porté par la construction verbale comme pour* normalement, au fond, franchement. » Blanche-Benveniste, 1990 : 79) que nous avons relevés à la suite des amorces: *enfin, bon, ben, oh, ah, eh, oui, non, non mais, ma foi, té, vé.*

d'amorces complétées, ces marques énonciatives représentent un peu moins de la moitié des constructions incidentes (39, 4% ;100/254). La catégorie d'incises non « énonciatives » semble donc réservée aux amorces complétées

On trouve ces différents ajouts « énonciatifs » dans les trois types d'amorces (Tableau 10) mais dans des proportions différentes. Si les amorces inachevées et corrigées se comportaient comme les amorces complétées, les répétitions de l'élément amorcé constitueraient presque la moitié des ajouts « énonciatifs », or ce n'est pas le cas. Cependant les effectifs dans les amorces corrigées et inachevées sont insuffisants pour qu'on puisse en tirer des conclusions précises.

Incises « énonciatives »	**Amorces complétées**	**Amorces inachevées**	**Amorces corrigées**	**Total**
Amorce répétée	45 (45%)	4 (7, 1%)	6 (22, 2%)	55
Pause brève	21 (21%)	20 (35, 7%)	12 (44, 4%)	53
Pause remplie	17 (17%)	9 (16, 1%)	4 (14, 8%)	30
Élément méta-énonciatif	17 (17%)	23 (43, 4%)	5 (18, 5%)	45
Total	**100 (100%)**	**56 (100%)**	**27 (100%)**	**183**

Tableau 10. Incises énonciatives après des amorces complétées, inachevées ou corrigées.

- Dans d'autres cas, tout aussi nombreux, les incises trouvées après les amorces sont des éléments qui viennent modifier sans remplacer ce qui a précédé l'amorce. Ces ajouts sont à peu près réservés aux amorces complétées (n=81) puisqu'on n'en trouve que 2 cas dans les amorces corrigées. Ils représentent un peu moins du tiers (31, 9%) des incises et modifications produites par le locuteur à l'occasion de ces amorces complétées (tableau 10).

Des adverbes, des adjectifs, des verbes, des noms, des pronoms et des déterminants constituent l'essentiel (80, 2%, 65/81) de ces ajouts modifieurs.

et après une fem- + **une jeune** femme que je connais pas dit (Pug_2VI 29, 2)
on a p- **non je crois qu** 'on n'a pu faire que celle-là (Vidal 15, 3)
pardon monsieur je voudrais savoir le Br- le **chemin de** Bruxelles (Histoire 3, 1)
tu vois + après nous on li- on **les** livre comme ça (Benot 3, 13)
au Panier elle était connue + elle vendait tout pa- tout **le** Panier (1) (Scalone2 28, 4)

ah pas toujours elles vont aussi à l'o- + **pas** à l'opéra (31) (Puget-VI 14, 16)

Ces ajouts complètent, nuancent ou même rectifient l'énoncé qui vient d'être dit. En particulier, les verbes ajoutés sont des verbes faiblement recteurs :

je crois, je pense, je considère que, avoir pu, vouloir, essayer, c'est que, ce qu'on appelle
en français donc nous enrich- **je considère que** nous nous enrichissons la langue (Chrabi 35, 2)

- Un certain nombre d'incises ne « modulent » pas le contexte antérieur à l'amorce mais s'insèrent comme de véritables parenthèses autonomes dans ce contexte antérieur repris. C'est « un morceau qui vient se loger à l'intérieur d'un énoncé, comme un parasite qui n'aurait pas de relation syntaxique avec son hôte » (Blanche-Benveniste, 1997 : 121). Les incises sont donc des « constructions incidentes sans éléments métalinguistiques » (Authier-Revuz, 1995). Ces ajouts ne sont observés que dans le cas d'amorces complétées et représentent 11, 4% des ajouts ou modifications (tableau 10) :

prends le lui là bas s- **Carole** sur le la machine il doit y en avoir un regarde (Mmeduran 8, 10)
alors ça fait que nous av- **nous y sommes allés et** nous avons pas bougé de mon quartier (2) (Scalone2 38, 7)
à ce stade là + vous n'a- **il n'y a pas eu de lutte** + vous n'avez porté que des éléments (Avocat 12, 7)

5.4.2. Les éléments antérieurs modifiés

Si dans les exemples précédents, l'amorce permettait l'insertion d'éléments « énonciatifs » ou incidents, oubliés ou insérés, les amorces qui vont être décrites annoncent des modifications sur certains éléments antérieurs. La plupart du temps, la modification consistera en un remplacement. Le locuteur a dit un mot pour un autre, arrête le cours de son énoncé au milieu d'un mot qui le suit, reprend une partie de son énoncé et corrige un élément. Ce ne sont pas des amorces corrigées puisqu'en fait elles seront complétées après qu'une partie de l'énoncé antérieur ait été remplacé. Rien ne s'opposerait à ce que ce type de phénomènes soit observé également dans le cas des amorces corrigées ou inachevées mais, en fait, il n'y en a aucun. Ces cas (au nombre de 44) sont réservés aux amorces complétées et représentent 17, 5% des incises ou modifications produites par le locuteur à l'occasion de ces amorces complétées (tableau 9).

Les éléments amorcés dans ce cas sont variés mais les trois quart (35/44) sont des verbes (15) ou des noms (20). Les éléments antérieurs modifiés sont encore plus variés ; les trois quart (36/44) sont des débuts de syntagme (déterminants, quantifieurs, préposition, sujets pronoms ; 23) ou des verbes (13) dont plus de la moitié subit une modification temporelle. Nous avons vu que lorsque les amorces sont le lieu de la réparation et sont corrigées, ces éléments sont très rarement impliqués (10/177). Voici les différents types d'interruption et de modifications constatés dans ce recueil de corpus :

- Le syntagme verbal est interrompu
  - le sujet du verbe est modifié (fréquent) :

    parce qu'il était sévère il il a vécu comme il **il** vi- **on** vivait à son pays (Scalone2 2, 2)
    je connais beaucoup de **monde** qui ha- beaucoup de **gens** qui habitent en France (Portuga3 1, 14)
  - le temps du verbe est changé

    quand tu **as** des mi- quand tu **as eu** des milieux colorés dedans (Microbio 40, 3)
- Le syntagme nominal est interrompu
  - le déterminant est modifié

    dès qu'on a fait **cette** infection oppor- dès qu'on fait **une de ces** infections opportunistes (Fgarcin 23, 7)
    on peut mettre jusqu'à dix dix kilos + on + heu + **une** pu- + **la** purée d'olives dans un scourtin (Pug_2VI 3, 12)
    s'il est **au** qua- + s ' il est **du** quatorzième siècle (Pug_2VI 14, 14)
    comment on reconnaît un Belge **dans** un marchand de chau- **chez** un marchand de chaussures (Histoire 3, 14)
    c'est environ à + **trois** kilo- **deux** kilomètres et demi trois kilomètres de Valpajola (Corse 8, 9)
  - L'adverbe est interrompu et son modificur est corrigé

    les personnes qui me touchent **toute** part- **tout** particulièrement à cœur (Birnie 1, 8)

Parfois, la modification antérieure consiste à ôter une élément présent dans l'énoncé précédent. Cette catégorie est rare (9 cas) et réservée, elle aussi, aux cas d'amorces complétées :

ils sont **trop** nombr- ils sont nombreux (Vidal 2, 4)
mais souvent c'est que + tous les **petits** en- + tous les enfants c'est le désir qu'ils ont de jouer (Puget-VI 28, 9)

6. Conclusion

Les amorces de mots sont des phénomènes relativement fréquents à l'oral puisqu'en moyenne il s'en produit quatre tous les mille mots. Une étude précédente (Pallaud, 1999) a montré qu'ils constituaient la moitié des achoppements qui se produisent dans les énoncés oraux. La fréquence des amorces varie beaucoup d'un locuteur à l'autre et probablement d'un moment à l'autre pour un même locuteur. Cette hétérogénéité inter- et probablement intra-individuelle dans les hésitations et ratés lors des prises de parole a été déjà mentionnée notamment par Lickley (2001) qui passe en revue les différents facteurs susceptibles d'intervenir sur la production de ces phénomènes.

Les amorces de mots sont également concernées par les erreurs d'écoute relevées lors de la vérification des corpus transcrits. Les modifications apportées dans ce domaine comme dans celui des autres erreurs contribuent surtout à enrichir les textes transcrits.

Les amorces de mots sont présentes dans les recueils de « lapsus » « à la volée » mais dans des proportions différentes de celles qu'on trouve dans les corpus enregistrés. Si on retrouve bien les amorces complétées et corrigées, les amorces inachevées sont absentes des recueils « à la volée ».

L'étude de ces phénomènes dans un grand corpus montre que si un tiers des amorces est inachevé, le reste est constitué par les amorces complétées, de loin les plus nombreuses (60%) et par les amorces corrigées (11%).

La moitié des amorces quelle que soit leur catégorie concerne les catégories Objet mais plus largement encore (63%) les éléments qui se trouvent à droite du verbe. Le tiers des interruptions de mots se produit sur des verbes alors que 13% seulement des éléments amorcés sont des sujets. Les trois types d'amorces (complétées, inachevées et corrigées) sont également concernés lorsque les éléments touchés sont à droite du verbe. En revanche, les amorces sur la catégorie Sujet seront plus souvent des amorces complétées tandis que, sur la catégorie Verbe, les amorces seront plus fréquemment inachevées ou corrigées.

C'est le lexique nominal ou verbal qui est, dans la grande majorité des cas, affecté par ces phénomènes d'interruption. Seuls 16% des amorces touchent des « petits mots » (déterminants, pronoms, quantifieurs).

La façon dont les amorces sont complétées ou corrigées montre que le locuteur tend, dans la plupart des cas, à reprendre l'énoncé interrompu

depuis le début du syntagme qu'il soit nominal ou verbal. C'est le cas également pour les bribes ou fragments de syntagme (Martinie, 1999).

Les amorces s'accompagnent de phénomènes d'incises diverses. Si la grande majorité de ces incises (87%) est constituée d'éléments simplement ajoutés, d'autres incises s'accompagnent de modifications dans l'énoncé à gauche de l'amorce. Parmi les éléments simplement ajoutés, les incises « énonciatives » (pauses silencieuses, ou remplies, éléments méta-énonciatifs) sont les plus nombreuses et se retrouvent pour les trois types d'amorces. les autres types d'incises sont réservées aux amorces complétées. Le tiers des incises vient moduler le contexte à gauche de l'amorce par l'insertion d'adverbes, d'adjectifs, de noms, de verbes, de pronoms ou de déterminants. D'autres ajouts moins nombreux (11%) sont de véritables incidentes parenthétiques à fonctionnement d'associé. Les cas de correction à gauche ne sont pas rares (17%).Ont été comptabilisés dans cette catégorie les cas, rares, où le locuteur rectifie sont énoncé en supprimant un élément déjà prononcé.

Références

Authier-Revuz J. (1995). *Ces mots qui ne vont pas de soi. Boucles réflexives et non coïncidences du dire*. Paris : Larousse.

Blanche-Benveniste C., Bilger M., Rouget C. & van Den Eynde K. (1990). *Le français parlé. Etudes grammaticales*. Paris : Edition du CNRS

Blanche-Benveniste C. (1997). *Approches de la langue parlée en français*. Paris-Gap : Ophrys.

Cappeau P. (1998). Quelques mots sur quelques bribes liées au genre. In M. Bilger, K. Van den Eynde & F. Gadet (Eds) *Analyse linguistique et approches de l'oral. Recueil d'études offert en hommage à Claire Blanche-Benveniste* (pp. 301-311). Paris : Peeters Leuven.

Duez D. (1997). La signification des pauses dans la production et perception de la parole. *Revue Parole 3-4*, 275-299.

Gadet F. (1997). *Le français ordinaire*. Paris : Colin.

Jeanjean C. (1984). Les ratés c'est fa- fabuleux. Étude syntaxique des amorces et des répétitions. *LINX 10, Syntaxe et discours*, 171-177.

Martinie B. (1999). *Étude syntaxique des énoncés réparés en français parlé*. Thèse de Doctorat, Université Paris X, Paris.

Pallaud B. (1999). Lapsus et phénomènes voisins dans la langue parlée : problèmes d'identification. *Recherches sur le français parlé, 15*, 1-33.

Pallaud B. (2001). Les lapsus : des pierres dans le champ linguistique, *Actes du Colloque de Cerisy-la-Salle sur « Linguistique et psychanalyse »*, 1-8 septembre 1998 (sous presse).

Pallaud B. (sous presse) Erreurs d'écoute dans la transcription de données orales. *Revue Parole.*

Du bon usage des notations phonétiques dans les corpus aphasiques

Marie-Noëlle ROUBAUD
Université de Provence

Résumé

Les notations phonétiques sont plus nécessaires dans les corpus de locuteurs aphasiques que dans les corpus de locuteurs ordinaires. Elles traduisent ce que le transcripteur n'est pas en mesure de restituer sur un plan orthographique. Mais l'étude et l'analyse des corpus montrent qu'il faut faire un usage restreint de ces notations : tout ce qui peut l'être doit être orthographié. Mais qu'est-ce qu'un mot ou un fragment de mot orthographiable ?. Il faut distinguer les jargons, qui seront notés phonétiquement, des amorces et déformations phonémiques, qui ont un traitement particulier. Le problème n'est pas simple : tout transcripteur, face à la recherche du mot par le locuteur aphasique, fait des hypothèses sur le morphème recherché ; que doit-il alors transcrire ? L'article expose ces difficultés et apporte des éléments de réponse pour l'usage des notations phonétiques.

1. Introduction

Le G.A.R.S.[1] a une longue expérience de transcription de corpus qui l'a amené à établir des conventions de transcription. Ces conventions ont peu changé car les principes adoptés lors de leur élaboration restituent le texte écouté avec un maximum de fidélité, sans aucun toilettage.

Trois principes régissent les transcriptions effectuées par le G.A.R.S. :

---

[1] Les travaux du G.A.R.S. sont désormais poursuivis par l'équipe DELIC (DEscription Linguistique Informatisée des Corpus).

Recherches sur le français parlé, 17, 103-116.
©2002, *Université de Provence*

1. la transcription se fait en orthographe standard, sans aucun trucage orthographique, les grands dictionnaires servant de référence ;
2. elle ne comporte aucun signe de ponctuation, aucune majuscule de début de phrase ;
3. elle comprend quelques notations phonétiques en note de bas de page.

2. Typologie

1.1. Les notations phonétiques correspondent à des morphèmes interprétables

C'est le cas des morphèmes prononcés rapidement et dont la déformation est banale :

(1) enfin maintenant[1] je vais marquer ce qui a à mettre dans le cartable et puis vous allez sortir (corpus parodie enfants, Mor,IV,11,15)

---

[1]prononcé [mE)nα)]

Pour les mots de grande fréquence, comme *maintenant* (1), *voilà, et puis, peut-être*, qui subissent des raccourcissements caractéristiques ([mE)na)], [ϖλα], [επι], [πτΕτ]) quel que soit le locuteur qui les produit, le transcripteur a souvent intérêt à les signaler dans la présentation du corpus, supprimant ainsi les notes de bas de page. Mais la prononciation de certains morphèmes, à un certain moment du corpus, peut être utile à garder en note de bas de page :

(2) il a fallu lui acheter un chauffage d'appoint à bain d'huile et également[1] une bouilloire (corpus adulte, Lav,38,15)

---

[1] prononcé [εγαZ↔μα))]

Chez les enfants, la déformation d'un morphème peut s'expliquer parfois par la présence de la liquide /r/. Dans l'exemple (3), l'enfant produit le phonème /m/ au lieu de la suite consonantique /b/ + /r/ et dans l'exemple (4), il semble que le /r/, présent dans *réseaux* et dans *très*, parasite la production du morphème *tension* , que le locuteur adulte produira ensuite sans aucune erreur :

(3) tout le nord a été ravagé les arbres ravagés euh euh les poteaux d'électricité tombés les toits des des maisons arrachés euh notre euh

dix dix millions d'arbres[1] dix millions d'arbres sont abattus (corpus enfants, Ver, II, 21,6)

---

[1] prononcé [αρμ]

(4) suite à la tempête alors c'est surtout des problèmes notamment de de réseaux de lignes très haute tension[1] euh très haute tension (corpus adulte, Hen,34,9)

---

[1] prononcé [τρα)σ□)]

C'est le cas également des morphèmes qui en remplacent d'autres (cas des lapsus) :

(5) la phrase est toute faute[1] - il y a qu'une se- il y a qu'un seul mot dans tout le - qui est juste (corpus parodie enfants, Mor,IV,5,11)

---

[1] pour « fausse »

C'est encore le cas des morphèmes que le transcripteur ne peut pas orthographier, parce qu'ils lui sont inconnus (noms propres, mots étrangers). Dans l'exemple (6), le transcripteur identifie un nom propre mais ne sait pas l'écrire, d'où la notation phonétique de bas de page :

(6) un jour - il y avait un - il y avait quel- euh il y avait quelqu'un qui s'appelait - Jacques [ ][1] - qui fai- qui était hum - qui était turbulent comme lui (corpus parodie enfants, Mor,II,3,2)

---

[1] [σατρΕρ]

Dans l'exemple (7), le transcripteur identifie l'adjectif *provençal* mais du fait de sa prononciation, ne sait pas l'orthographier, d'où la notation phonétique de bas de page :

(7) L1 c'est quoi la danse folklorique
L2 la farandole et tout ça ce qui est la Provence on nous déguise en [ ][1] et puis
L1 en quoi
L2 en [ ][1] (corpus enfants dyslexiques, Zab,50,3)

---

[1] [πρ□ϖΕνσυν], probablement « Prouvençau » [πρυϖΕνσαυ] (cité dans le dictionnaire, au sens de « provençal »)

Mais il n'est pas toujours nécessaire de recourir à une notation phonétique. Dans le cas des morphèmes dont l'écriture est codifiée (régionalismes, argots, jeux, onomatopées, emprunts), on peut rétablir l'orthographe dans le texte, comme pour *mama mia* (ex. 8) :

(8) vous vous rappelez - quand le lendemain matin - elle a hurlé mama mia (corpus parodie enfants, Mor,I, 10,5)

2.2. Les notations phonétiques correspondent à des morphèmes non interprétables

Ces morphèmes que le transcripteur ne peut identifier comme des mots de la langue seront appelés « jargons ». Leur transcription ne peut pas se faire en orthographe standard et un espace vide entre crochets dans le corps du texte signale leur présence. Leur prononciation est restituée autant que possible en note de bas de page, entre crochets, en API. Dans les deux exemples qui suivent, le transcripteur n'a pas pu interpréter les morphèmes (9 et 10) :

(9) eh bien quand vous - c'est quand je vous [ ][1] - c'est quand je vous - je vous redonne votre fils - c'est après les vacances regarde le calendrier (corpus parodie enfants, Mor, II,6,5)

---

[1] [λεβε]

(10) L1 que fait ta maman
L2 elle [ ][1] et
L1 elle fait du ménage et puis elle fait autre chose (corpus enfants, Romans)

---

[1] [δεμελα)Z]

Ces notations phonétiques sont parfois nombreuses dans les transcriptions d'aphasiques : bon nombre de mots sont déformés[2]. L'étude des textes aphasiques[3] a montré que c'est le manque du mot qui fait défaut au locuteur. La recherche envahit le texte : elle peut aboutir (le mot attendu est alors produit par le locuteur ou par le thérapeute) ou ne pas aboutir et les corpus montrent alors le désarroi du patient qui n'arrive pas à dire ce mot[4].

---

[2] *cf.* bibliographie, études de Giorgini (1991) et Truong (1997) sur les déformations phonémiques des mots dans les corpus aphasiques.

[3] Cette étude s'est faite en collaboration avec Claude Loufrani à l'Université de Provence, responsable d'un recueil de textes sur l'aphasie.

[4] Ce désarroi se lit au travers des paroles du patient dans les corpus : *je ne sais pas le dire, je ne trouve pas le mot* ... (*cf.* Roubaud & Loufrani, 1993)

Pour un lecteur averti, les notations phonétiques de bas de page peuvent traduire cette recherche[5] (ex. 11) :

(11) j'ai [ ][1] j'ai [ ][2] [ ][3] [ ][2] [ ][4] j'ai remis le téléphone et je suis resté sur le téléphone sur le lit (corpus aphasique,Tru97, 74,9)

---

[1] [ρακλ↔με]
[2] [ρακλ↔]
[3] [ρεκοΣε]
[4] [ραρκλε]

Dans cet exemple, le lecteur averti peut faire l'hypothèse que les notations phonétiques en bas de page sont des réalisations approximatives[6] du morphème-cible *raccroché* (*j'ai raccroché le téléphone*) que le patient tente de produire par cinq essais successifs, dont deux essais sont identiques et correspondent au même numéro de note (note n°2). Cette recherche prend fin dès qu'un autre verbe *remis* (*j'ai remis le téléphone*) est produit par le patient.

Plusieurs questions se posent alors :

1. Quand utilise-t-on les notations phonétiques ?

   La transcription de deux corpus de locuteurs aphasiques (que nous identifierons par I et II[7]) dans un mémoire de DEA a montré qu'il fallait poser des limites. Le transcripteur avait abusé des notations phonétiques qui masquaient des phénomènes tout à fait normaux : amorces de mots ou déformations phonémiques.

2. Où les place-t-on ?

   Doit-on les rejeter en bas de page au risque de laisser apparaître un texte à trous ou doit-on les placer dans le corps du texte[8] ?

3. Comment traiter les recherches lexicales ?

---

5 Cette recherche se traduit aussi dans le débit du locuteur qui varie de 30 à 330 mots/mn, dans la durée et la fréquence des pauses (notées par : - pour une pause brève et -- pour une pause longue), dans la longueur des silences, dans les suites de syllabes inaudibles (notées par des X), dans le nombre de relances du thérapeute.

6 Pour la notion d'approximations lexicales, *cf.* Roubaud (1987, 1988).

7 Pour les deux corpus étudiés, le débit est en moyenne de 100 mots/mn pour le corpus I et de 110 mots/mn pour le corpus II (pour un locuteur ordinaire, il est de 200 mots/mn).

8 Il est d'ailleurs précisé dans les conventions de transcription du G.A.R.S. (version mars 2000) que pour les textes « particuliers » (jeunes enfants, pathologies, étrangers), des conventions spécifiques sont établies, selon les besoins, pour placer des notations phonétiques dans le texte.

Doit-on différencier dans le texte les morphèmes qu'on ne peut vraiment pas interpréter (ex. 9) de ceux qui apparaissent au moins comme des recherches du mot (ex. 11) ?
C'est au travers d'exemples, extraits de corpus de locuteurs aphasiques, que nous essaierons de répondre à ces questions.

3. Les notations phonétiques et leur place dans les corpus aphasiques

3.1. Les déformations phonémiques

Comme dans les corpus de locuteurs ordinaires, certaines notations phonétiques correspondent à des morphèmes interprétables. On a intérêt à les écrire en orthographe dans le corps du texte et à restituer leur prononciation en note de bas de page (ex. 12 à 18).

C'est le cas des confusions de phonèmes :

- entre une consonne sourde /v/ et une sonore /f/

(12) j'ai j'étais là-bas en qua- en euh au mois de mars quarante-trois et à la fin de à la à la fin[1] la fin de la (Tru98,II,55,3)

---

[1] prononcé [ϖE)]

- entre deux liquides /r/ et /l/ (ex. 13) ou une liquide /r/ et une constrictive /Z/ (ex. 14)

(13) Thérapeute : il y a quelque chose qui va pas et puis
Patient : ils allaient[1] il m'a envoyé au docteur D- David à Lyon (Tru98,I,18,14)

---

[1] prononcé [αρε]

(14) Thérapeute : la guerre était terminée vous vous êtes engagé en trente-neuf
Patient : je je me suis je me suis ren- je me suis engagé[1] en- engagé pour la la la (Tru98,II,53,15)

---

[1] prononcé [α)γαρε]

- entre deux voyelles /□)/ et /α/

(15) et puis je vous dit avant elle elle venait chez sa chez sa sœur chaque année chaque année[1] et puis après elle allait passer trois euh trois ou quinze ou trois semaines chez chez son chez son fils (Tru98,II,66,7)

---

[1] prononcé [σ□)καvε]

C'est aussi le cas des omissions de phonèmes, comme dans cet exemple où l'omission du /j/ par le locuteur peut laisser entrevoir la production d'un lapsus « marré » pour *marié* :

(16) et ensuite la guerre est finie je me suis marié[1] ma- marié (Tru98,II,52,6)

---

[1] prononcé [μαρε]

C'est encore le cas des ajouts de phonèmes. Dans l'exemple qui suit, le locuteur transforme un /t/ final en /sk/ :

(17) c'est un peu bête[1] (Nolot,I,52,7)

---

[1] prononcé [βEσκ]

Le morphème peut aussi être identifié grâce à la connaissance culturelle que se partagent le locuteur et le transcripteur. Ainsi dans l'exemple 18, le patient imite la voix du général de Gaulle et lui emprunte la formule *Vive la France* qui terminait ses discours, en déformant par deux fois le mot *France* :

(18) Thérapeute : vous avez fait euh le général De Gaulle l'autre jour là
Patient : vive la France[1] vive la France[2] nous les Français mais je peux pas c'est pas la même voix (Tru98,II,60,7)

---

[1] prononcé [κα)σ]
[2] prononcé [κα)]

Rétablir l'orthographe de ces mots dans le corps du texte, c'est alors les identifier comme des morphèmes de la langue. Il y a également un avantage à rassembler en bas de page les erreurs de prononciation : cela facilite une étude des déformations les plus fréquentes dans un corpus de locuteur aphasique.

3.2. Les jargons

Les mots « jargonnés » sont parfois nombreux :

(19) Thérapeute : et qu'est-ce que vous avez dit
Patient : je sais pas rien c'est pour le les les les les [ ][1] les XXX celle qui était [ ][2] à à Saint-Tropez (Tru98,II,56,3)
---
[1] [ζιϖα)]
[2] [τολοκε]

La notation phonétique de ces mots dans le corps du texte rétablirait une bonne lisibilité du texte. En effet, un texte jargonné risque d'être moins lisible si le corps du texte est parsemé de « vides » (notés [ ]), le lecteur devant sans cesse se reporter aux notes de bas de page au risque de perdre le sens du texte. Et l'écoute du texte aligné avec son signal sonore ne correspondra pas avec la transcription puisque aucune marque graphique ne correspondra à la suite de phonèmes produite. C'est pour cela que nous noterons dorénavant les notations phonétiques dans le corps du texte (ex. 19') :

(19') je sais pas rien c'est pour le les les les les [ζιϖα)] les XXX celle qui était [τολοκε] à à Saint-Tropez

Afin de faire le tri entre ce qui était un jargon et ce qui ne l'était pas, nous avons relu systématiquement des transcriptions pour les améliorer. C'est une entreprise longue, indispensable, car sous l'apparence de jargons, nous avons découvert des morphèmes identifiables. Cela a été le cas :

- des noms propres non identifiés par le transcripteur

Dans l'exemple qui suit, Coblence, prononcé en allemand Koblenz, a été transcrit phonétiquement, comme un jargon[9] :

(20) Thérapeute : vous m'avez dit bon euh votre votre épouse n'est pas française
Patient : elle est de de près de de [k] [κoβλE))σ] sur sur le sur le Rhin (Tru98,II,61,7)

alors qu'il aurait fallu transcrire :

(20') elle est de de près de K- Koblenz sur sur le sur le Rhin

---

9 Dans la fiche descriptive de ce corpus, le transcripteur aurait dû indiquer que certains des mots du locuteur étaient prononcés en allemand comme le nom de Coblence.

- des fautes contre la norme

L'auxiliaire *avoir* est souvent employé là où il faudrait *être* selon la norme et cette erreur a été transcrite comme un jargon :

(21) mais mais je me sens bien de suite euh une amie de ma sœur est [αϖ↔νψ] à onze à dix heures du soir (Tru98,II,78,14)

Il aurait fallu transcrire :

(21') mais mais je me sens bien de suite euh une amie de ma sœur est a venu à onze à dix heures du soir

- des amorces

Toute amorce de mot peut être transcrite dans le corps du texte, selon les conventions de transcription : signalée par un tiret collée au mot (par exemple : *m-* , amorce de *mais*). Tout locuteur, ordinaire ou aphasique, fait de nombreuses amorces et leur transcription dans les corpus participe à l'authenticité des textes. Mais une erreur serait de les noter phonétiquement, lorsqu'on peut faire *mieux*, comme dans ces exemples où le transcripteur n'a pas réussi à les identifier :

(22) Thérapeute : est-ce que vous pourrez me citer des noms de journaux - locaux
Patient : [κο] le Nice matin (Tru98,I,41,4)

Il aurait fallu rétablir l'amorce de «*comme* :

(22') co- le Nice matin

(23) Patient : et puis j'étais je je mangeais avec chez mes mes mes cousins et puis tout d'un coup ce bras il répondait pas
Thérapeute : quel bras
Patient : celui-ci il ne répondait pas et et [σ↔κιμορ] il tombait il tombait vous savez (Tru98,II,75,16)

On aurait pu rétablir l'amorce du syntagme verbal *ce qui m'aurait* :

(23') celui-ci il ne répondait pas et et ce qui m'aur- il tombait il tombait vous savez

Ces amorces peuvent subir des déformations, comme dans les exemples ci-dessous :

(24) mon père était pas [πι] prisonnier et et ma mère était là quoi (Tru98,II,54,12)

On pourrait rétablir l'amorce de *prisonnier*, pour laquelle le /r/ de la première syllabe a été omis :

(24') mon père était pas pi- prisonnier et et ma mère était là quoi

(25) mais et par moment et et et par moment je [βυ] je doutais (TRU98,II,84,14)

On pourrait rétablir l'amorce de *doutais*, pour laquelle la labiale /b/ a été remplacée par la dentale /d/ :

(25') mais et par moment et et et par moment je bou- je doutais

Un nouveau problème se pose alors (ex. 24 et 25) : doit-on écrire en orthographe standard ce que l'on entend (ex. 24' et 25') ou doit-on mettre cette prononciation en note de bas de page (ex. 24'' et 25'') et rétablir l'orthographe de l'amorce ?

(24'') mon père était pas pri-[1] prisonnier et et ma mère était là quoi

---

[1] prononcé [πι]

(25'') mais et par moment et et et par moment je dou-[1] je doutais

---

[1] prononcé [βυ]

Nous pensons que les amorces peuvent toutes être traitées au maximum de la même façon, qu'elles aient ou non subi une déformation phonémique, et c'est pour cette raison que nous préférons écrire en orthographe standard l'amorce de ce qui a été prononcé (ex. 24' et 25'). Nous réservons les notes de bas de page à la prononciation des morphèmes produits entièrement (ex. 12 à 18).

4. Les notations phonétiques et les recherches lexicales

Certaines notations phonétiques, même si le transcripteur ne peut les identifier comme des mots de la langue, sont des réalisations approximatives d'un morphème-cible. C'est le cas du verbe *raccrocher* de l'exemple 11 (les notations figurant désormais dans le corps du texte) :

(11) j'ai [ρακλ↔με] j'ai [ρακλ↔] [ρεκοΣε] [ρακλ↔] [ραρκλε] j'ai remis le téléphone et je suis resté sur le téléphone sur le lit

Même pour un transcripteur averti, cette recherche n'est pas toujours sûre à cent pour cent, témoin cet exemple :

(26) Thérapeute : ben racontez justement cette histoire là
Patient : oh il y a pas longtemps il y a il y a peut-être d- dou- douze ans de ça douze ans j'étais avec mon ami euh euh j'aimais bien aller sans [δα)σαταλε] le soir parce qu'il y avait un un où on on donnait tout en en mar- en en mangeant voyez-vous (Tru98,II,56,14)

Doit-on laisser la notation phonétique ou peut-on rétablir le syntagme *dans cette allée* ? Dans ce cas, il est possible que *sans* précédant le syntagme *dans cette allée* soit la préposition *dans*, prononcée [σα)], le locuteur répétant deux fois la préposition comme le ferait un locuteur ordinaire : *dans dans cette allée*. Mais n'est-on pas en train de trop interpréter le texte ?

Le cas inverse apparaît dans l'exemple qui suit : le transcripteur a rétabli l'orthographe standard d'un morphème *à pied* qui semble inadapté dans ce texte :

(27) on on (n') est pas divorcé elle m'aime toujours même l'autre jour elle m'a habité euh elle m'a à pied ap- [πα] [παλ] elle m'a appelé euh non on se voit elle elle est venue chez chez ma soeur (Tru98,II,64,14)

Le locuteur aphasique semble rechercher le verbe *appeler* et le lecteur comprend difficilement la présence du lexique *à pied*, donné par le transcripteur. Il aurait fallu laisser la notation phonétique (ex. 27') :

(27') on on (n') est pas divorcé elle m'aime toujours même l'autre jour elle m'a habité euh elle m'a [απφε] ap- [πα] [pal] elle m'a appelé euh non on se voit elle elle est venue chez chez ma soeur

Cette recherche est facilement repérable, si le mot produit possède tous les phonèmes du mot attendu mais dans le désordre (ex. 28) :

(28) euh je je je vous en donne je vous en- dans le dans le dans le euh euh [ψναZ] non nuage (Tru98,II,71,11)

Cette inversion dans l'ordre des phonèmes n'est pas caractéristique des locuteurs aphasiques, témoin cet exemple où c'est le thérapeute qui inverse les deux dernières syllabes alors que le patient les répète correctement :

(29) Thérapeute : [αλιζε] et faites-le si vous pouvez
Patient : allez-y et allez-y et non alors pas pas très bien (Nolot,I,47,11)

Cette recherche est également repérable dans les différents essais, plus ou moins nombreux, produits par le patient en quête d'un mot (ex. 30 et 31) ou d'un syntagme (ex. 32) :

- recherche du mot *autorisation*

(30) je voulais faire la guerre mon père euh sans un mot il m'a écrit son [οριζ] son [σολιτασφ□)] ses autorisations que je pouvais y aller (Tru98,II,53,7)

- recherche du mot *contacts*

(31) Thérapeute : vous ne vivez plus ensemble
Patient : non non non non j'ai été la voir en Allemagne il y a deux ans je suis resté dix jours je m- je mangeais chez elle avec avec sa soeur et j'a- j'allais le soir pour euh pour dormir à l'hôt- à l'hô- à l'hôtel voyez-vous nous sommes n- nous aurons de de bons euh de bons [κ□)] [κ□)] [κ□)] [κ□)κατρ] [κ□)τρ] [κ□)κατρ] [κ□)κατρ] merde [κ□)] [κ□)κατ] [κ□)κατ] c'est bon comme ça non elle
Thérapeute : contacts (Tru98,II,65,8)

C'est l'orthophoniste qui clôt cette recherche en prononçant le mot *contacts* que le locuteur, par dix essais successifs, tente de prononcer. Pour ce dernier exemple, nous rétablirons l'orthographe standard des amorces *con-* du mot *contacts*, ce qui donne (ex. 31') :

(31') n- nous aurons de de bons euh de bons con- con- con- [κ□)κατρ] [κ□)τρ] [κ□)κατρ] [κ□)κατρ] merde con- [κ□)κατ] [κ□)κατ] c'est bon comme ça non elle

- recherche du syntagme *contremaître de fabrication*

(32) Thérapeute : oui j'avais oublié de vous poser cette question là euh votre métier
Patient : j'é- j'étais chez Nestlé comme comme [komEt] [k□)μEτρεσα))] de [)fregyrasj□)] de - non attendez de [k□)τρ] [k□)τρ™μEρ] contremaître de de de [)fa)] [fa)pretasj□))] fabrication chez Nestlé (TRU98,II,70,1)

Le patient cherche à la fois le mot *contremaître* (4 essais) et le mot *fabrication* (3 essais). La recherche porte en fait sur tout le syntagme : *contremaître de fabrication*. Comme pour l'exemple précédent, nous rétablirons en orthographe standard les amorces *contr-* de *contremaître* et l'amorce déformée *fan-* de *fabrication* (ex. 32') :

(32') j'é- j'étais chez Nestlé comme comme de [komEt] [k□)μEτρεσα))] de [)fregyrasj□)] de - non attendez de contr- [k□)τρ™μEρ] contremaître de de de fan- [fa)pretasj□))] fabrication chez Nestlé

Nous nous sommes demandée s'il fallait indiquer au lecteur la présence de ces recherches lexicales. On pourrait envisager de placer, dans le texte, un signe graphique, comme dans l'exemple qui suit :

j'é- j'étais chez Nestlé comme comme [komEt = contremaître] [)k□)μEτρεσα)) = contremaître] de [)fregyrasj□) = fabrication] de - non attendez de contr- [k□)τρ™μEρ = contremaître] contremaître de de de fan- [)fa)pretasj□)) = fabrication] fabrication chez Nestlé

Mais, de toute évidence, cela ne ferait que surcharger le texte et nuire à sa lisibilité. Distinguer les recherches du mot dans les notations phonétiques sera la tâche du linguiste, travaillant sur le corpus. Le nombre d'essais, l'apparition de phonèmes appartenant au morphème-cible et les amorces seront des indices précieux pour identifier ces recherches.

5. Conclusion

La tâche du transcripteur, face à un corpus de locuteur aphasique, est multiple.

Dans un premier temps, il doit noter phonétiquement, dans le corps du texte, ce qui est prononcé par le locuteur et qu'il n'interprète pas immédiatement comme un mot de la langue.

Puis vient une étape plus difficile : faire le tri de ces notations phonétiques. En orthographe standard pourront être restitués les phénomènes ordinaires comme les amorces, les déformations phonémiques (dont la prononciation sera notée en bas de page) ...

Dans le corps du texte ne figureront que les notations phonétiques correspondant à des morphèmes non interprétables. Mais il appartiendra au linguiste de repérer, parmi ces morphèmes, ceux qui sont les indices d'une recherche lexicale. Cela suppose que le transcripteur donne du sens à la parole du locuteur aphasique, qu'il soit capable de repérer les réalisations approximatives d'un morphème-cible sans interprétations excessives.

Ce n'est que si le transcripteur a fait un bon usage des notations phonétiques que le travail sur le corpus pourra alors se faire.

Références

Bernstein, B. (1975). *Langage et classes sociales. Codes socio-linguistiques et contrôle social.* Paris : Éditions de Minuit.

Blanche-Benveniste, C., Bilger M., Rouget C. & van Den Eynde K (1990). *Le français parlé. Etudes grammaticales.* Paris : CNRS.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français.* Paris-Gap : Ophrys.

Dorze, G. de (1985). *L'aphasie et le processus de lexicalisation.* Thèse de doctorat, Université de Montréal, Montréal.

Giorgini, M. (1991). *Étude d'un locuteur aphasique : corpus Y. L. : déformation phonémique et dénomination.* Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.

Loufrani, C. (1988). Un corpus de locuteur aphasique, originalité ou régularité ? *Recherches sur le français parlé, 8*, 59-78.

Loufrani, C. (1990). *Analyse du discours de locuteurs classé comme aphasiques : de l'art de tourner autour du mot.* Thèse de doctorat de linguistique française. Université de Provence, Aix-Marseille 1.

Loufrani, C. & Roubaud, M-N. (1990). La notion d'approximation : langage ordinaire ou pathologique. *Recherches sur le Français parlé*, 10, 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*, 195-210.

Nolot, V. (1991). Exploitation syntaxique et lexicale de corpus de locuteurs aphasiques. Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.

Roubaud, M.-N. (1987). *L'approximation lexicale.* Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille 1.

Roubaud, M.-N. (1988). *Les mystères de l'approximation lexicale.* Mémoire de DEA de linguistique française. Université de Provence, Aix-Marseille 1.

Roubaud, M-N. & 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*, 85-113.

Roubaud, M.-N. & Loufrani, C. (1999). Éclairage des corpus de type aphasique par les grilles. *Recherches sur le français parlé, 15*, 41-57.

Truong, K.-Y. (1997). *Déformations sémantiques et comptage des verbes chez deux locuteurs classés aphasiques.* Mémoire de maîtrise de linguistique française. Université de Provence, Aix-Marseille I.

Truong, K.-Y. (1998). *Classement et études sémantiques des verbes thématiques et de quelques verbes dits de grande fréquence présents chez deux locuteurs classés aphasiques.* Mémoire de D.E.A.de linguistique française. Université de Provence, Aix-Marseille I.

Propositions pour un étiquetage en « séquences fonctionnelles »

Mireille BILGER* & Estelle CAMPIONE**
*Université de Perpignan
**Université de Provence

Résumé

De nombreux chercheurs en Traitement Automatique des Langues ont essayé de mettre au point des stratégies d'analyse partielle (ou *shallow parsing*) des énoncés, faisant apparaître des segments intermédiaires entre le mot et la « phrase », qui sont le plus souvent des suites de mots très courtes, dont le statut linguistique n'est pas très clair et leur caractérisation est généralement faite dans la littérature à l'aide d'exemples plutôt que de définitions précises. Nous adoptons une approche opposée, et nous proposerons une stratégie d'analyse partielle non pas en segments minimaux, mais au contraire en « séquences » maximales qui représentent les constituants fonctionnels rencontrés dans les textes, sans entrer dans le détail de leur composition. En effet, nous faisons l'hypothèse qu'un tel niveau d'analyse, relativement aisé à réaliser manuellement, peut être utile pour l'étiquetage de grands corpus, à partir desquels des analyses plus fines pourront être menées.

1. Introduction

Reconnaissant la difficulté, à la fois théorique et pratique, d'une analyse totale des énoncés, de nombreux chercheurs en Traitement Automatique des Langues ont essayé de mettre au point des stratégies d'analyse partielle (ou *shallow parsing*), faisant apparaître des segments intermédiaires entre le mot et la « phrase ». Le plus souvent, ces segments sont des suites de mots très courtes appelées *chunks* (Abney, 1991) ou

Recherches sur le français parlé, 17, 117-136
©2002, *Université de Provence*

syntagmes non récursifs (Vergne, 1999). Nous empruntons à Vergne (1999) un exemple de découpage :

(La mesure) (de concentration) (de chlorophylle) (est utilisée) (pour estimer) (la biomasse)

Le statut linguistique de ces segments n'est pas très clair, et leur caractérisation est généralement faite dans la littérature à l'aide d'exemples plutôt que de définitions précises, même si des justifications prosodiques (Vergne, 1999) ou psycholinguistiques (Abeillé *et al.*, 2000) sont avancées.

Nous adopterons dans cet article une approche opposée, et nous proposerons une stratégie d'analyse partielle non pas en segments minimaux, mais au contraire en « séquences » maximales qui représentent les constituants fonctionnels rencontrés dans les textes, sans entrer dans le détail de leur composition. Nous faisons l'hypothèse qu'un tel niveau d'analyse, relativement aisé à réaliser manuellement, peut être utile pour l'étiquetage de grands corpus, à partir desquels des analyses plus fines pourront être menées. L'approche utilisée essaie de concilier à la fois une analyse en termes de dépendances et une analyse en termes de constituants, et s'apparente, par certains aspects, aux grammaires en chaînes (*strings grammars*) de Harris (1962) et Salkoff (1973).

2. Description du corpus

Pour tester notre démarche, nous avons appliqué ce découpage sur un corpus de taille réduite, totalisant près de 4000 mots, mais nous avons tenu à ce qu'il soit constitué de trois types de productions ; la répartition entre ces différents types est la suivante :

- un corpus d'extraits de textes écrits tirés de la presse, (*Ça m'intéresse, Sciences et Vie*) totalisant 527 mots ;
- un corpus constitué de textes « fabriqués » [1] destinés à être lus pour une recherche de parole en laboratoire, (*Euroml*), de 2682 mots ;
- un corpus d'oral spontané (extrait de l'enquête *Corpus de Référence de Français parlé),* de 760 mots.

[1] Il s'agit de la partie française d'un corpus de parole de laboratoire (Euroml) qui est constitué de passages de cinq phrases liées sémantiquement entre elles et lus par dix locuteurs qui avaient comme consigne de simuler une intonation naturelle.

3. Principes d'analyse

Nous n'avons pas cherché à recenser de manière exhaustive les différentes structures syntaxiques susceptibles d'être rencontrées dans n'importe quel texte. Nous nous sommes appuyées uniquement sur les données qu'offrait notre corpus, et nous nous sommes essentiellement préoccupées des constructions syntaxiques organisées autour d'un verbe.

Les constructions sans verbe rencontrées dans le corpus ont cependant reçu une étiquette qui en signale simplement le type, comme pour les trois exemples suivants :

(1)	Const. Adv.	(ici les pompiers)
(2)	Connecteur	(et)
	Const. Nom.	(moi qui sais pas résister)
(3)	Const.Adj.	(pareil là-bas)

Nous nous réservons le droit de rendre compte de leur constitution interne et des relations qu'elles peuvent entretenir avec d'autres unités syntaxiques au cours d'une analyse ultérieure.

En revanche, les constructions syntaxiques organisées autour d'un verbe reçoivent une analyse plus fine. Elles vont pouvoir être décomposées en diverses « séquences » [2] qui entretiennent avec cet élément plusieurs types de relation de dépendance. Le calcul de ces relations se fonde sur l'analyse syntaxique développée dans le cadre de l'Approche Pronominale (*cf.* Blanche-Benveniste *et al.*, 1984, 1990).

4. Les séquences

Ce que nous entendons par « séquence » est une unité syntagmatique qui occupe une grande fonction syntaxique, comme celle de Verbe, Sujet, Objet, *etc.* La différence qu'il y a entre ces « séquences » et les habituels « constituants » se situe dans le fait qu'on ne cherche pas à rendre compte des différents éléments qui les constituent. On délimite les séquences en prenant, en fait, « tout ce qui se rapporte à » l'élément fonctionnel (verbe, sujet, objet, *etc.*). On essaie ainsi, en premier lieu, de délimiter les grandes masses qui renvoient à des fonctions. Chaque « séquence » intègre donc des éléments de catégories diverses.

---

[2] Nous avions déjà introduit ce terme de « séquence » dans Bilger (1999).

C N R S ATILF NANCY 2

4.1. La séquence Verbe

Ce que nous désignons par *SéqV* correspond à ce que nous pouvons appeler une construction verbale « minimale », autrement dit, cela correspond à la partie verbale de la construction, dans son extension linéaire. Elle comporte le « bloc verbal » constitué de la forme verbale tensée, des pronoms clitiques « sujet » et « compléments », des particules de négation, ainsi que des participes, des infinitifs, des verbes modaux et de certains adverbes.

Les éléments insérés à l'intérieur de cette séquence ont pour particularité de dépendre de la forme verbale en question et de ne pas pouvoir développer une syntaxe interne. C'est cette double caractéristique qui nous fait intégrer dans cette zone « verbe » les pronoms clitiques « sujet » et « compléments », puisque par définition, ces formes clitiques ne sont pas « développables ». Dans certaines langues (*cf.* Creissel, 1991), cette séquence « Verbe » est d'ailleurs représentée comme une seule unité avec des éléments reliés graphiquement par un trait d'union ou sous forme d'affixes amalgamés sur une base.

Certains adverbes sont liés par des portées réciproques et présentent des dépendances qui leur sont propres, cependant, comme ils sont en définitive branchés sur la forme verbale en question, nous les incluons également dans la séquence « Verbe ».

Voici quelques exemples des différentes extensions que peut prendre la *SéqV* :

- Verbe seul :
  (accompagnait)
- sous forme de liste :
  (construit et dessine)
- pronoms clitiques sujet + objet + verbe :
  (ils les pollinisent)
- verbe + négation :
  (n'est pas)
- verbe + adverbe
  (sont généralement)
- auxiliaire + pp. :
  (ont attaqué)
- faire (laisser) + infinitif :
  (il la fait construire)

- verbe de perception + infinitif :
  (il voyait construire)
- modal + infinitif :
  (devrait s'incurver)
  (il devrait pouvoir les acheter)

La *SéqV* maximale pourrait donc avoir une extension du type :

il ne devrait pas pouvoir les avoir souvent fait polliniser et fertiliser

Les éléments susceptibles de composer la « séquence verbe » semblent pouvoir se calculer facilement. Cependant pour pouvoir la prévoir correctement, il faudrait pouvoir mettre au point la microgrammaire des auxiliaires, des causatifs, des modaux, (*cf.* par exemple, le travail de Chu, 1988). De même, il faudrait connaître la distribution des adverbes qui peuvent venir s'y insérer, et en définir les différentes classes, tant il est vrai qu'à l'intérieur de cette catégorie adverbiale, et notamment en ce qui concerne la négation, il existe un degré d'autonomie non négligeable.

4.2. Les séquences Sujet, Objet et Rection

Nous utilisons des étiquettes telles que *SéqS, SéqO* et *SéqR* pour signaler que la construction verbale se compose, cette fois, de constituants fonctionnels susceptibles de présenter une syntaxe interne. Ces constituants peuvent se réaliser sous une forme lexicale simple ou sous la forme d'une liste. Ils peuvent également comporter des :

- Relatives
- Que-Phrase
- Appositions
- Éléments prépositionnels
- *etc.*

et apparaître sous différents dispositifs, entre autres :

- Extraction en « c'est... que/qui »
- il y a ... qui
- pseudo-clivée

Quel que soit le développement syntagmatique de ces séquences, ces dernières entretiennent obligatoirement une relation de proportionnalité avec une proforme. Cette proforme permet de repérer quelle fonction syntaxique la séquence occupe auprès du verbe constructeur.

Ainsi, l'énoncé suivant :

Cinq voyages vers le Pôle sud et un séjour au Groenland n'ont pas guéri ce glaciologue de la passion qu'il nourrit pour les deux extrémités de la terre, dont il aime étudier les entrailles de glace

est proportionnel à « ils ne l'en ont pas guéri », ce qui permet de dégager quatre séquences du type :

SéqS	(Cinq voyages vers le Pôle sud et un séjour au Groenland)
SéqV	(n'ont pas guéri)
SéqO	(ce glaciologue)
SéqO	(de la passion qu'il nourrit pour les deux extrémités de la terre, dont il aime étudier les entrailles de glace)

Pour l'énoncé du type :

la semaine dernière, mon amie est allée chez le médecin, parce qu'elle avait une angine

le découpage proposé sera le suivant :

SéqR	(la semaine dernière)
SéqS	(mon amie)
SéqV	(est allée)
SéqO	(chez le médecin)
SéqR	(parce qu'elle avait une angine)

Les séquences « Rection » entretiennent une équivalence avec les proformes respectives « quand » et « pourquoi », l'ensemble des séquences *sujet* + *verbe*+ *objet* avec « elle y est allée »

Chercher ces relations de proportionnalité avec les proformes nous permet de rendre compte du schéma syntaxique de la construction mise en cause. Cela nous permet aussi de remettre à plus tard l'analyse de la syntaxe développée à l'intérieur des séquences.

4.2.1. La séquence Sujet (*SéqS*)

Les données du corpus nous ont permis de relever des séquences « Sujet » d'extension différente ; cette séquence se réalise sous une forme simple, sans ajout[3], comme dans :

(chacun) prend parti
(la lumière) se déplace
(Basley) utilise et perfectionne

[3] Nous reprenons ici le terme utilisé, entre autres, par Salkoff (1973).

Elle peut également se réaliser sous la forme d'une liste :

(les petits animaux et les insectes) se sauvent

ou présenter une forme avec ajout(s). Ces ajouts entretiennent alors avec l'élément central plusieurs types de fonctions, entre autres, celles de « complément », de « modificateur » et « d'apposition ». Et, ils vont pouvoir se réaliser dans plusieurs catégories : « nom », introduit ou non par une préposition, « adjectif », « relative », *etc.* En voici quelques exemples relevés dans le corpus :

- ajout de type « complément »

  (les gens du pays)
  (l'accès par la traverse des Sables Jaunes)

- ajout de type « modificateur »
  - sous forme adjectivale

    (des astres très massifs)

  - sous forme de relative

    (l'ambition, qui forme le caractère des gouvernements populaires)
    (le nombre de postes qu'il propose)

- ajout de type « apposition »
  - sous forme nominale

    (Le Pr. Ben Basley, spécialiste des sciences de l'environnement à l'université du Colorado,)

  - sous forme adjectivale :

    (les athéniens, non contents d'avoir délivré leur patrie)
    (des insectes, attirés par le nectar)

  Chaque type d'ajout peut se réaliser sous la forme d'une liste :

  (le jeu de cache-cache du soleil et de la lune)
  (des connexions entre la couronne, le chromosphère et la surface du soleil)

ou bien semblent pouvoir se combiner :

(Amilcar, surnommé Barca, père d'Annibal, d'une famille chère au peuple, soutenu de beaucoup de crédit et d'un grand génie)

et s'imbriquer les uns aux autres de manière récursive :

(la lumière d'une étoile observée au ras du soleil, qui est suffisamment massif pour courber sensiblement l'espace à son voisinage)

La séquence maximale Sujet pourrait donc être du type :

(Le grand Paul de Dijon, dit « le chanceux », frère de Marie et de Marc, originaire d'un quartier cossu de la ville et que sa famille a délaissé)
(a gagné le concours)

la même séquence sous dispositif :

c'est le grand Paul de Dijon, dit « le chanceux ", frère de Marie et de Marc, originaire d'un quartier cossu de la ville et que sa famille a délaissé **qui a gagné** le concours

Les arrangements des différents ajouts sont complexes, et leurs combinatoires difficilement prévisibles.

Ces différents types d'ajouts que nous venons de mentionner sont ceux qui sont traditionnellement recensés pour rendre compte de la possible constitution du syntagme nominal ou « du nom expansé » (cf. Chonin & Delhay, 1998).

Les exemples cités ici n'éliminent pas bien entendu tous les types des SéqS que l'on peut rencontrer dans les textes. La *SéqS* peut se réaliser sous une autre forme que celle d'un syntagme nominal, par exemple, sous la forme d'une « que phrase » ou d'un infinitif :

(que Paul soit si souvent absent) est regrettable
(ne pas le lui avoir donné) était une erreur

4.2.2. La séquence Objet (*SéqO*)

Les compléments qui sont nécessaires au sens et à la construction du verbe, autrement dit les « éléments de valence complément », reçoivent une seule et même étiquette, celle de *SéqO*. Aucune autre information n'est signalée, ni sur la forme que peut prendre le complément en question, ni sur les traits syntactico-sémantiques auxquels il renvoie.

Les « séquences Objet » de catégorie nominale admettent à peu près les mêmes types d'extensions que les *SéqS*, ce qui pourrait donner comme séquence maximale :

(on a photographié)
(le grand Paul de Dijon, dit « le chanceux », frère de Marie et de Marc, vivant dans un quartier cossu de la ville et que sa famille a délaissé)

(on a pensé au)
(grand Paul de Dijon, dit « le chanceux »frère de Marie et de Marc, vivant dans un quartier cossu de la ville et que sa famille a délaissé)

Seul le dispositif aurait sa forme modifiée en « c'est... que » :

c'est le grand Paul de Dijon, dit « le chanceux », frère de Marie et de Marc, vivant dans un quartier cossu de la ville et que sa famille a délaissé qu'on a photographié

Mais bien d'autres types de syntagme pourront être étiquetés comme *SéqO*, entre autres, les « que-phrases », les « si-phrases », les infinitifs, les formes adjectivales, les adverbes. Voici quelques exemples tirés du corpus :

il préfèrerait (que ce soit pas une cuisine végétarienne)
il aimerait savoir (s'il y aura un wagon restaurant)
il n'ose pas (le bouger)
il est désolé (de ne pas avoir pu venir dîner)
il semblait (très proche)
il se comporte (très bien)

4.2.3. La séquence Rection (*SéqR*)

Les éléments qui sont construits par un verbe, mais qui ne le caractérisent pas, sont signalés par l'étiquette *SéqR*. Ceci dit, il n'est pas toujours aisé de statuer sur le côté définitoire ou non que certains éléments peuvent avoir auprès du verbe. Comme cela a été dit dans Blanche-Benveniste (1990), la distinction entre éléments de « valence » et de « rection » est une affaire de lexicographie beaucoup plus que de syntaxe. L'utilisation de dictionnaires de valences tels que celui élaboré dans le cadre du projet Proton[4] (van den Eynde & Mertens, 2001) pourrait donc servir de base à cette distinction.

Dans ce sens, nous n'utilisons pas nécessairement l'étiquette *SéqR*, pour les constituants (prépositionnels ou non), du type :

à Giverny
avec le pépiniériste Joseph Bory
depuis une dizaine d'années
de manière parfaitement correcte

cela dépendra essentiellement du lexique verbal mis en cause. Ainsi pour les verbes, tels que :

habiter (à Giverny)
cohabiter (avec le pépiniériste)
naître (depuis une dizaine d'années)
se comporter (de manière correcte)

ces éléments seront étiquetés comme *SéqO*, dans la mesure où ils semblent être essentiels au sens et à la construction de chacun de ces verbes, alors qu'avec une autre forme verbale, comme par exemple le verbe « peindre », l'ensemble pourra recevoir l'étiquette *SéqR* :

---

[4] Ce dictionnaire est consultable sur le Web à l'adresse : http://bach.arts.kuleuven.ac.be/PA/protonfr.html (mai 2002).

(peindre des paysages)
(à Giverny)
(avec le pépiniériste)
(depuis une dizaine d'années)
(de manière correcte)

Cette analyse différenciée essaie de s'appuyer au maximum sur des données statistiques, mais en leur absence, elle dépend souvent de la simple intuition que l'on peut avoir de la construction du verbe en question.

Les séquences introduites par un marqueur, tel que :

- parce que
- quand
- comme (comparatif)
- si (d'hypothèse)
- pendant que
- pour que
- *etc.*

si elles sont bien de type rectionnel[5], reçoivent également l'étiquette *SéqR*.

4.3. La séquence Associée (*SéqAss*)

Certains éléments gravitant autour du verbe ont l'apparence de compléments régis par cette forme, ils n'en partagent cependant aucune des propriétés, entre autres : refus de l'extraction en « c'est... que » et aucune équivalence avec une proforme. Nous avons dressé, (*cf.* Blanche-Benveniste et al., 1990), d'une part, une liste des introducteurs de « séquence associée » :

- comme (causal)
- car
- si (de prémisse)
- quant à
- afin que
- puisque
- bien que

[5] Avec certains verbes, les séquences introduites par « quand » ou « comme » sont des compléments valenciels ; elles reçoivent alors l'étiquette « SéqO » :
elle est née quand tu es parti
elle se comporte comme son père le voulait

d'autre part, une liste d'introducteurs qui peuvent, selon le contexte, être des marqueurs de Séquences « associées » ou « rectionnelles », par exemple :

- quand
- parce que[6]
- à mon avis
- selon

La position de la séquence en question, avant ou après la *SéqV*, peut être un critère déterminant, ainsi le « comme (causal) » à valeur d'associé est réalisé avant la *SéqV*. En revanche, le « comme (comparatif) » à valeur rectionnelle se trouve généralement après la *SéqV*.

Il en est de même pour les « séquences » introduites par « si ». Le « si d'hypothèse » rectionnel peut se trouver aussi bien avant qu'après la *séqV*, alors que le « si de prémisse » se trouve généralement avant cette séquence.

Établir ces deux listes, qui ne sont d'ailleurs pas exhaustives, n'épuise pas, bien entendu, tous les cas de *SéqAss* possibles. Bien d'autres formes pourront recevoir aussi cette étiquette :

SéqAss	(heureusement)
SéqV	(il y a pas eu)
SéqO	(de bagarre)

de même que les formes lexicales réalisées sous la forme d'une dislocation, comme dans :

SéqAss	(cette fille)
SéqV	(elle ne parlait pas)
SéqAss	(moi)
SéqV	(je la plains pas)

4.4. La séquence Recteur Faible (*SéqRf*)

On se réserve la possibilité de ne pas traiter automatiquement le premier verbe tensé comme la séquence du verbe constructeur. Si le premier verbe peut être interprété comme un recteur faible, on traite la séquence (y compris le « que ») comme telle. On peut établir une liste de verbes qui pourraient supporter ce statut :

---

[6] Les données du corpus de français parlé nous ont d'ailleurs amenées à vouloir distinguer de manière particulière les SéqAss introduites par « parce que » dans la mesure où elles ont un statut très particulier : elles produisent souvent un effet de « parenthèse » et peuvent avoir un développement relativement important.

- Germain précise que
- la Général a dit textuellement que
- les égyptiens pensaient que
- celle-ci stipule notamment que
- cela signifie que
- ainsi que l'affirme Einstein

on pourrait rajouter à cette liste :

- croire que
- trouver que
- être convaincu que
- *etc.*

Dans le cas où ces formes verbales ont le statut de « recteur faible », c'est-à-dire, dans le cas où on peut les traiter comme des « introducteurs », ces verbes semblent avoir pour caractéristique, entre autres, de ne plus vraiment exercer de sélection sur leur valence.

5. L'étiquette « Insertion »

L'analyse en terme de dépendance ou de non dépendance, par rapport notamment à la *SéqV,* et que nous avons appliquée jusqu'ici, ne permet pas de rendre compte de toutes les composantes rencontrées dans les textes. De fait, certains éléments semblent être « imperméables » à ce type d'analyse ou présentent un intérêt limité à être traités selon ces termes. Ces éléments ont pour particularité de pouvoir s'insérer entre diverses séquences ou à l'intérieur même des séquences, sans en modifier les relations. Nous signalons cette caractéristique par l'étiquette « Insertion » ; nous avons relever trois types d'insertion :

- Insertion Modalité (*InsertMod*)
- Insertion Phatique (*InsertPh*)
- Insertion Parenthèse (*InsertParenth*)

5.1 L'insertion Modalité (*InsertMod*)

Un certain nombre d'adverbes « modaux », comme « peut-être », « sûrement », « sans doute », *etc.*, ont la possibilité d'être suivis de la forme « que » et de précéder aussi bien les séquences réalisées avant la *SéqV* que d'introduire directement cette dernière :

InsertMod	(Peut-être que)
SéqR	(si je lui chante une chanson)
SéqV	(elle se calmera)

SéqR	(si je lui chante une chanson)
InsertMod	(peut-être que)
SéqV	(elle se calmera)

Nous signalons cette possibilité par l'étiquette « *InsertMod* ». Nous utilisons la même étiquette pour le tour interrogatif en « est-ce que » :

InsertMod	(Est-ce que)
SéqR	(si je lui chante une chanson)
SéqV	(elle se calmera)

SéqR	(si je lui chante une chanson)
InsertMod	(est-ce que)
SéqV	(elle se calmera)

5.2 Insertion Phatique (*InsertPh*)

Nous utilisons l'étiquette *InsertPh* pour toute une série de formes qui sont employées comme telles dans les productions et qui peuvent apparemment s'insérer entre n'importe quelle séquence. En voici quelques exemples :

InsertPh	(bon)
SéqAss	(le début de mon voyage)
SéqV	(ça a été)

SéqV	(je suis restée)
SéqO	(trois mois)
InsertPh	(hein)

SéqV	(passez-moi)
InsertPh	(s'il vous plaît)
SéqO	(les réclamations)

5.3. Insertion Parenthèse (*InsertParenth.*)

Nous signalons par cette étiquette, toute construction verbale qui va pouvoir s'insérer entre les différentes séquences déjà citées ou à l'intérieur même de ces séquences, et pour laquelle aucune autre étiquette ne peut être apposée, comme dans l'exemple suivant :

SéqV	(j'ai amené)

SéqO	(des affaires d'hiver des affaires d'été plus euh)
InsertParenth.	(à cette époque j'avais j'étais en maîtrise il me restait le mémoire à faire)
SéqO	(plus euh les livres tout ce qu'il me fallait pour faire mon mémoire)

6. Récapitulatif

Ces différentes séquences permettent de prévoir un enchaînement syntagmatique optimal du type :

SéqAss + SéqRf + InsertMod + SéqS + SéqV + SéqO + SéqR + InsertPh + InsertParenth

L'ordre d'apparition[7] de ces séquences ainsi que leurs points d'insertion sont relativement peu contraints, à l'exception de deux séquences : la séquence « Recteur faible » et la séquence *InsertMod* qui devront nécessairement apparaître avant la séquence « Verbe ».

Nous pouvons proposer un exemple fabriqué qui fait apparaître l'ensemble de ces séquences :

> Évidemment, les collègues ont pensé que si on acceptait d'accompagner l'enseignant, ben peut-être que les jeunes pourraient visiter, je sais pas si visiter est le bon mot, la ville de Paris, parce que c'est comme ça il faut un encadrement

et que l'on peut étiqueter de la manière suivante :

SéqAss	(Évidemment)
SéqRf	(les collègues ont pensé que)
SéqR	(si on acceptait d'accompagner l'enseignant)
InsertPh	(ben)
InsertMod	(peut-être que)
SéqS	(les jeunes)
SéqV	(pourraient visiter)
InsertParenh	(je sais pas si visiter est le bon mot)
SéqO	(la ville de Paris)
SéqAss.PCQ	(parce que c'est comme ça il faut un encadrement)

Il est à noter que nous n'avons pas trouvé une telle suite de séquences dans le corpus que nous avons étudié, ni dans les productions écrites, ni dans les productions orales. Voici quelques extraits étiquetés :

---

7 Dans un autre cadre, Morel (1997) a également travaillé sur la combinatoire des constituants qui peuvent se trouver en « préambule » de la séquence Verbe.

- Extrait tiré du corpus de presse écrite :

Avec son groupe de recherche du CIRES, le Pr. Ben Balsley, spécialiste des sciences de l'environnement à l'université du Colorado, à Boulder, mène chaque année des campagnes de relevés atmosphériques en divers endroits du globe. Plutôt que les habituels ballons-sondes, Basley utilise et perfectionne depuis une dizaine d'années un système capable d'emporter à plus de 5000 m d'altitude toutes sortes d'instruments de mesure, de maintenir ces capteurs à diverses hauteurs choisies et, surtout, de les récupérer à leur point de départ.

Seul un cerf-volant, ou un ballon captif pour les jours sans vent, peut dresser dans le ciel le câble sur lequel le CIRES fait circuler ses appareillages de mesure.

Science & vie, n°983, août 99, p. 57.

SéqR	(Avec son groupe de recherche du CIRES)
SéqS	(le Pr. Ben Basley, spécialiste des sciences de l'environnement à l'université du Colorado, à Boulder)
SéqV	(mène)
SéqR	(chaque année)
SéqO	(des campagnes de relevés atmosphériques en divers endroits du globe)
SéqAss	(Plutôt que les habituels ballons-sondes)
SéqS	(Basley)
SéqV	(utilise et perfectionne)
SéqR	(depuis une dizaine d'années)
SéqO	(un système capable d'emporter à plus de 5000 m d'altitude toutes sortes d'instruments de mesure, de maintenir ces capteurs à diverses hauteurs choisies et, surtout, de les récupérer à leur point de départ)
SéqS	(Seul un cerf-volant, ou un ballon captif pour les jours sans vent,)
SéqV	(peut dresser)
SéqO	(dans le ciel)
SéqO	(le câble sur lequel le CIRES fait circuler ses appareillages de mesure)

- Extrait tiré du corpus Euroml :

j'ai des problèmes avec mon adoucisseur d'eau : le niveau est toujours trop haut, ce qui fait que le trop plein coule sans arrêt. Vous pourriez pas m'envoyer un réparateur mardi matin ? C'est le seul jour de la semaine que j'ai de libre. Si vous êtes d'accord, soyez gentil de me confirmer le rendez-vous par écrit.

SéqV	(j'ai des problèmes)
SéqO	(avec mon adoucisseur d'eau)
SéqS	(le niveau)

SéqV	(est toujours)
SéqO	(trop haut)
SéqRf	(ce qui fait que)
SéqS	(le trop plein)
SéqV	(coule)
SéqO	(sans arrêt)
SéqV	(vous pourriez pas m'envoyer)
SéqO	(un réparateur)
SéqO	(mardi matin)
SéqV	(c'est)
SéqO	(le seul jour de la semaine que j'ai de libre)
SéqAss	(si vous êtes d'accord)
SéqV	(soyez gentil de me confirmer)
SéqO	(le rendez-vous)
SéqO	(par écrit)

- Extrait tiré du corpus de français parlé :

mais au Portugal par contre je suis partie l'hiver - et je ne savais pas pour combien de temps je partais si c'était l'hiver ou plus longtemps ce qui fait que j'ai emmené des affaires d'hiver des affaires euh d'été -plus euh à cette époque j'avais j'étais en maîtrise il me restait le mémoire à faire - plus euh donc euh les livres tout ce qu'il me fallait pour faire mon mémoire là-bas -

Connecteur	(mais)
SéqR	(au Portugal par contre)
SéqV	(je suis partie)
SéqO	(l'hiver)
Connecteur	(et)
SéqV	(je ne savais pas)
SéqO	(pour combien de temps je partais si c'était l'hiver ou plus longtemps)
SéqRf	(ce qui fait que)
SéqV	(j'ai amené)
SéqO	(des affaires d'hiver des affaires euh d'été plus euh)
InsertParenth	(à cette époque j'avais j'étais en maîtrise il me restait le mémoire à faire)
SéqO	(plus euh donc euh les livres tout ce qu'il me fallait pour faire mon mémoire là-bas)
InsertParenth	(mais je m'en rappelle pas là
SéqO	(mais pas beaucoup)
InsertPh	(enfin)
SéqV	(je suis partie donc euh)
InsertPh	(je sais pas)
SéqO	(le vingt-cinq décembre ou le vingt-six un truc comme ça)
Connecteur	(et alors)
Const.Nom	(là dans ma 4L)
SéqV	(j'avais

SéqO	(une 4L fourgonnette)
Connecteur	(et)
SéqV	(je pensais)
InsertPh	(enfin je sais pas)
SéqV	(je pensais)
Connecteur	(mais)
SéqV	(j'avais prévu de dormir)
SéqO	(dans ma 4L)
InsertPh	(enfin bon)
SéqV	(j'avais emmené)
SéqO	(de quoi camper)
InsertPh	(euh enfin comme j'ai dit bon)
séqR	(pour le l'hiver le printemps l'été tout)

7. Distribution

Le tableau 1 donne les fréquences absolues et en pourcentages des différents types de séquences dans les trois corpus. On constate des différences de distributions, qui sont visualisées dans la figure 1. Par exemple, les *SéqS* sont beaucoup plus fréquentes à l'écrit que dans les autres corpus. Afin de juger de la significativité des différences observées, nous avons appliqué un test de χ^2 aux effectifs de chacune des séquences. La colonne p du tableau 1 donne la probabilité d'erreur en affirmant la présence d'un effet significatif : on considère généralement qu'il n'y a pas d'effet significatif au delà d'une valeur de p égale à 0,01 ou 0,05. Les lignes du tableau ont été classées par significativité décroissante.

On voit que les séquences qui discriminent le plus les trois corpus sont :

- *InsertPh* (beaucoup plus représentées à l'oral)
- *SéqS* (beaucoup plus représentées à l'écrit)
- *Connecteur* (beaucoup plus représentées à l'oral)
- *SéqO* (plus représentées dans Eurom)

On observe aussi des différences significatives, mais moins prononcées pour les séquences :

- *SéqR* (plus représentées à l'écrit)
- *SéqAss* (plus représentées à l'écrit)

Par contre la proportion de *SéqV* et *SéqRf* ne diffère pas de façon significative entre les trois corpus.

On notera que nous n'avons pas appliqué un test de χ^2 aux séquences ayant un petit effectif, sur lesquelles nous ne pourrions tirer de conclusions statistiques fiables.

	Ecrit		**Eurom**		**Fref**		**Signif.**	
TypeSéq	*Freq*	*%*	*Freq*	*%*	*Freq*	*%*	*p*	
InsertPh	0	0,0	7	1,0	21	13,7	0,000	oui
SéqS	18	17,6	49	6,8	2	1,3	0,000	oui
Connecteur	5	4,9	39	5,4	22	14,4	0,000	oui
SéqO	25	24,5	252	35,0	36	23,5	0,005	oui
SéqR	17	16,7	64	8,9	10	6,5	0,018	oui
SéqAss	9	8,8	27	3,8	4	2,6	0,032	oui
SéqV	24	23,5	226	31,4	45	29,4	0,258	non
SéqRf	3	2,9	27	3,8	3	2,0	0,519	non
Const.Nom	1	1,0	13	1,8	1	0,7	//	//
InsertMod	0	0,0	10	1,4	1	0,7	//	//
SéqAss.PCQ	0	0,0	0	0,0	4	2,6	//	//
InsertParenth	0	0,0	0	0,0	2	1,3	//	//
Autres	0	0,0	5	0,7	2	1,4	//	//
Total	**102**	**100**	**719**	**100**	**153**	**100**		

Tableau 1. Fréquence des différents types de séquences.

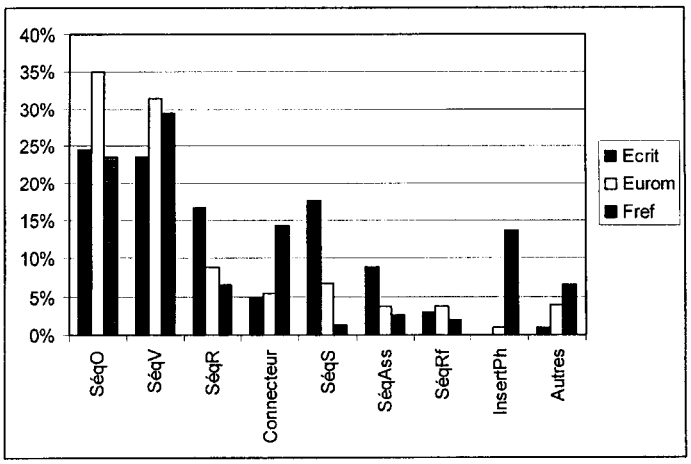

Figure 1. Distribution des séquences en fonction du corpus.

8. Conclusion

Ces propositions d'étiquetage de textes suivis peuvent facilement se prêter à un certain nombre de critiques ; on peut leur reprocher de s'appuyer sur une analyse à la fois trop « compliquée » ou trop « simpliste » ; on peut s'interroger sur l'utilité même de cet étiquetage. Ces propositions doivent avant tout être interprétées comme une première tentative qui nécessite un approfondissement.

Ceci dit, il nous semble qu'en ce qui concerne l'analyse, cet étiquetage en présuppose le moins possible et peut facilement être reproductible. La notion de « séquence fonctionnelle » permet de faire une analyse en grandes masses sans entrer immédiatement dans les détails. Cet aspect est loin d'être négligeable, tant on sait à quel point l'entrée immédiate dans les détails peut être compliquée et propice à des analyses divergentes. Cette procédure minimale de description présente l'avantage de rendre comparables des productions qui pourraient apparaître très différentes les unes des autres de par la simple constitution interne de leurs séquences, de même elle permet de rendre comparables des corpus de langues différentes et d'en faciliter ainsi l'étude contrastive.

Références

Abney, S. (1991). Parsing by chunk. In R. Berwick, S. Abney & C. Tenny (Eds), *Principles-Based Parsing* (pp.257-278). Dordrecht : Kluwer Academic Publishers.

Blanche-Benveniste, C., Bilger, M., Rouget, C. & van den Eynde, K. (1990). *Le français parlé. Études grammaticales*. Paris : Éditions du CNRS.

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J., & van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris : Selaf.

Bilger, M. (1999). Coordination : analyses syntaxiques et annotations. *Recherches sur le français parlé, 15*, 255-272.

Choi-Jonin, I. & Delhay, C. (1998). *Introduction à la méthodologie en linguistique*. Strasbourg : Presses Universitaires de Strasbourg.

Chu, X-Q. (1989). Examen de la structure de l'enchaînement des verbes modaux. *Recherches sur le français parlé, 9*, 75-86.

Creissels, D (1991). *Description des langues négro-africaines et théorie syntaxique*. Grenoble: Ellug.

Morel, M.-A. (1997). Stratégies intonatives et syntaxe du discours dans l'oral spontané. *Traitement Automatique des langues(T.A.L.), 38-1*, 83-100.

Salkoff, M. (1973). *Une grammaire en chaîne du français. Analyse distributionnelle*. Paris : Dunod.

Vergne, J. (1999). *Étude et modélisation de la syntaxe des langues à l'aide de l'ordinateur. Analyse syntaxique non combinatoire.* Mémoire d'habilitation à diriger des recherches. Université de Caen, Caen.

Abeillé, A., Clément, L. & Kinyon, A. (2000). Buildling a treebank for French. In *Proceedins of the Second International Conference on Language Resources and Evaluation (LREC 2000),* Athènes.

Harris, Z. (1962). *String analysis of sentence structure.* The Hague : Mouton.

van den Eynde, K. & Mertens, P. (2001). *La syntaxe du verbe, l'approche pronominale et le lexique de valence PROTON.* Preprint n°174, Department of Linguistics, Leuven : K.U. Leuven.
[en ligne : http://bach.arts.kuleuven.ac.be/pmertens/papers/proton.pdf].

Comment présenter les « types de phrase » dans une grammaire de référence du français d'après corpus ?

José DEULOFEU
Université de Provence

Résumé

L'objet de cet article est de contribuer à établir des cadres pour un projet de grammaire du français à partir de données authentiques sur corpus électroniques. Il y est discuté de l'opportunité d'un chapitre consacré aux « types de phrase ». Cette notion, communément admise dans les grammaires de références est critiquée. Il est montrée qu'elle rassemble des faits disparates et hétérogènes. Cette inadéquation est liée à l'utilisation d'un modèle syntaxique à une seule composante. Il est proposé d'utiliser un cadre qui oppose macro et microsyntaxe de façon à répartir la matière entre types de constructions grammaticales et types de structures macrosyntaxiques.

1. Introduction

L'éditeur Longman (Biber, 1999) vient de sortir une grammaire de référence de l'anglais qui fait à juste titre l'admiration de beaucoup d'entre nous, linguistes descriptivistes du français. Il nous prend à rêver de disposer d'un pareil instrument pour le français.

Mais si nous voulons construire une grammaire de référence du français fondée sur l'examen de corpus authentiques d'après le modèle britannique, nous allons rencontrer un problème parmi d'autres : quelle terminologie adopter pour en permettre une consultation aisée par un public de spécialistes comme par un public de personnes cultivées ? Quelles que soient les difficultés de détail, il me semble facile d'arriver à un consensus

Recherches sur le français parlé, 17, 137-165.
©2002, *Université de Provence*

pour une terminologie grammaticale de base qui allie cohérence linguistique et efficacité pédagogique dans des domaines tels que le syntagme nominal, la complémentation verbale, sans doute l'ensemble des structures que l'on peut rassembler sous l'appellation de « constructions grammaticales ». Je voudrais montrer que ce résultat sera beaucoup plus difficile à obtenir dans la présentation de ce que l'on appelle dans les grammaires les « types de phrases ». Pourtant cette présentation tient une place capitale dans ces ouvrages qu'ils soient d'inspiration descriptive ou normative. C'est en effet dans ce chapitre que sont présentées les structures syntaxiques majeures de la langue, au-delà desquelles commence l'analyse du discours. On connaît d'autre part l'importance de l'organisation des constituants majeurs de la phrase pour caractériser typologiquement les langues (Greenberg, Comrie). Dans la perspective plus particulière d'établir une grammaire de référence du français, construite à partir de corpus authentiques, qui puisse être consultée parallèlement à des grammaires comparables d'autres langues européennes, c'est à ce niveau que semblent pouvoir être repérées les ressemblances et les différences les plus typiques entre les langues. Tous les enseignants français d'allemand et de latin ont eu à souligner en terme de position du verbe dans les grands types de phrase (principales et subordonnées) les différences typologiques importantes qui existent entre la langue source et les langues cibles. Parmi les langues romanes, les phrases « à inversion du sujet » obéissent à des contraintes beaucoup plus strictes en français que dans les autres langues. Le français apparaît, en compensation comme beaucoup plus riche en tournures prédicatives du type : *il y a un chien qui aboie* ; *il y un truc, c'est qu'il est parti.*

Les présentations pédagogiques s'appuient donc crucialement sur la notion de type de phrase. Cependant lorsqu'on regarde à la fois les faits que recouvre la notion dans chaque grammaire et la terminologie employée, on ne peut qu'être frappé par le fait que les divers types de phrases distingués par ces grammaires sont en décalage avec une présentation qui s'appuierait sur les résultats communément admis de la linguistique descriptive dans ce domaine. J'illustrerai ces difficultés en examinant le traitement de la notion dans deux grammaires largement utilisées dans l'enseignement supérieur français. J'essaierai donc d'expliquer pourquoi la notion de type de phrase est sans doute celle où les grammaires de référence ont le plus de difficulté à proposer une terminologie descriptivement acceptable. Je proposerai deux pistes d'explication.

La première est la conséquence d'une confusion d'objectifs, au moins dans le cas de la France : les grammaires de référence, y compris celles

destinées à l'enseignement supérieur, se donnent l'impossible tâche de concilier présentation pédagogique et vulgarisation linguistique. Le terme phrase recouvre en fait deux objets distincts. D'un côté, une notion « pratique », au sens de Berrendonner (1989) et Beguelin (2000) développé plus loin, à savoir la forme « standard » que les grammaires normatives recommandent de donner aux énoncés écrits : la phrase commence par une majuscule et finit par un point et se présente sous la forme Sujet-verbe-complément. Et d'un autre côté, l'unité maximale de l'analyse syntaxique, qui peut prendre des formes beaucoup plus variées que l'unité pratique. Cette confusion entre unités pratiques et unités descriptives aboutit à normaliser les formes d'énoncés que l'on prend en compte et à marginaliser tout ce qui n'est pas phrase verbale. Du point de vue de l'analyse, une autre conséquence est que la description se fera sans discussion à partir de la phrase « pratique », rebaptisée « phrase canonique ». Celle-ci sera analysée de façon détaillée et spécifique, alors que les autres formes d'énoncés ne seront pas décrites pour elles-mêmes, mais à travers le miroir déformant ou le lit de Procuste des catégories établies à partir du modèle canonique.

La deuxième explication demeure interne à l'analyse linguistique. La notion de type de phrase recouvre de ce point de vue deux ordres distincts d'unités descriptives. Elle désigne d'abord des sous-types de constructions grammaticales. Ainsi, ce que l'on appelle, par exemple, « phrase passive » est un sous-type de construction verbale, qui par ailleurs reste un énoncé de type SVO. Au contraire, la « phrase à détachement », et, une grande partie des « phrases sans verbe », ne constituent pas des sous types de constructions grammaticales, mais des sous types d'énoncés paratactiques « binaires », non ramenables au schéma SVO et dont on peut montrer que l'analyse relève d'un autre niveau que celui de la syntaxe des catégories grammaticales.

La conclusion serait qu'une terminologie visant à la cohérence descriptive devrait renoncer à la notion de « phrase » source de toutes les confusions. Cet abandon pose problème, car les locuteurs construisent leurs représentations naïves de la syntaxe autour de la notion de phrase issue de la pratique scolaire et j'ai montré ailleurs (Deulofeu, 1995) qu'il est difficile d'imposer une nouvelle terminologie, si on ne parvient pas en même temps à modifier les représentations que les usagers se font de l'objet. Je proposerai donc une voie pour modifier ces représentations.

2. Les « types de phrase » dans la Grammaire Méthodique de Riegel *et al.* (1994)

Prenons l'exemple d'une grammaire qui est en passe de devenir la grammaire de référence pour les premiers cycles d'enseignement supérieur en France : Riegel *et al.* (1994). Dans la plupart des domaines de la description syntaxique (complémentation verbale, structure du groupe nominal et adjectival...), elle présente un état de la question qui tient compte des dernières recherches et en tire des synthèses intéressantes, bons compromis entre vulgarisation scientifique et présentation à visée pédagogique. Voyons alors comment est organisé le chapitre 11 explicitement intitulé : « types de phrases ». Les grandes rubriques en sont :

- Les types énonciatifs : assertion, interrogation, exclamation, injonction
- Le type logique : la phrase négative
- Les types fondés sur le réarrangement communicatif : phrase passive, phrase impersonnelle, phrases à procédés d'emphase (dislocation, extraction)
- Les phrases « atypiques » : phrases nominales, phrases à présentatifs, incises et incidentes.

2.1. Une approche non descriptive

Le principe de classement des types de phrase surprend par son caractère onomasiologique, alors que partout ailleurs dans le chapitre de syntaxe on part des formes pour aller au sens. De plus la présence de la rubrique *ad hoc* des « phrases atypiques » laisse entendre que le classement n'est pas en parfaite adéquation avec les faits. Les sous-rubriques, quant à elles, montrent bien que les types de phrase dans cette grammaire mêlent sans discussion explicite :

- des types de constructions. Certains types de phrases sont caractérisés par leur structure grammaticale : type de catégorie assurant la relation prédicative (nom ou verbe), sous-types de constructions à prédicat verbal (présence d'un modifieur du groupe verbal : la négation, phénomènes de diathèse : phrase passive et impersonnelle).
- types d'énoncés. Les critères de classification ne s'appuient plus sur les catégories grammaticales, mais sur d'autres principes

d'organisation : les procédés d'emphase, les modalités marquées par la prosodie.

Faute d'expliciter la distinction, les rubriques manquent de cohérence descriptive. Ainsi, la grande division des types de phrase fondés sur le réarrangement communicatif rassemble des types de constructions et des types d'énoncés. Les phrases passives et impersonnelles sont en effet des sous-types de constructions à prédicat verbal, puisque leurs différences reposent sur des opérations internes au système de la construction verbale (diathèse) ou de réalisation particulière du sujet (impersonnel) : ($_{CV}$ *mon frère a été vu*). Au contraire, les procédés d'emphase supposent des opérations relevant du niveau de l'énoncé et extérieures à la construction verbale : c'est celle-ci dans son ensemble qui est mise en relation avec le terme mis en emphase (mon frère -($_{CV}$ *je le vois - il a été vu*). D'un autre côté, cette approche classe dans deux rubriques différentes des types d'énoncés identiques. La structure d'énoncé thème-propos est un arrangement communicatif que l'on retrouve aussi bien dans *mon frère je le vois*, considérée comme une phrase à réarrangement communicatif typique et *Paris, bof!* ou *la cathédrale, tout droit*, classés comme « phrases atypiques binaires ». Descriptivement, on constate que dans les deux cas, les séquences sont des énoncés du même type (thème-propos) et ne diffèrent qu'au niveau des constructions grammaticales (présence ou absence de constructeur verbal dans le propos).

Si les critères ne sont pas hiérarchisés, ils ne sont pas non plus explorés systématiquement. Ainsi, la typologie des constructions formant un énoncé ne prend en compte que les cas où la relation prédicative est assurée par un verbe conjugué. De sorte que les constructions à « nexus prédicatifs non verbaux » ne forment pas une sous-rubrique des types de phrases sujet-prédicat, mais sont dispersées dans des sous-rubriques de chapitres différents : *et André qui ne revient pas* apparaît dans le sous-type des « phrases exclamatives » (p. 406), *voilà l'homme qui est tout surpris, il y a un verre qui est cassé, il y a un verre (de) cassé*, dans celui des phrases « atypiques à présentatifs » (p. 456). L'unité et la spécificité des constructions avec relative prédicat est pourtant bien affirmée (p. 485). D'autres types à prédicats non verbaux ne semblent même pas mentionnés (*et encore un livre par terre, deux hommes renversés par une voiture...*).

Les divers points de vue sous lesquels on peut considérer une unité syntaxique majeure et qui constituent des dimensions indépendantes : forme de la prédication, limites des relations de rection, divisions informatives explicitement marquées, expression de la modalité de phrase ne sont pas systématiquement explorés.

2.2. Statut trop privilégié de la phrase canonique verbale

De fait, le décalage entre cette présentation et une description des énoncés tient à un parti pris majeur : les types de phrase considérés comme typiques sont ceux que l'on peut ramener au modèle de la phrase assertive verbale choisi comme « canonique » :

> « c'est par rapport à ce schéma de référence que seront décrites toutes les phrases observables. » (*op. cit.*, p. 109).

Le cas de phrases observables qui ne seraient pas ramenables moyennant règles syntaxiques au modèle de la phrase canonique est cependant prévu par les auteurs :

> « Si néanmoins il n'y a pas de relation formulable en termes syntaxiques entre le modèle canonique et une phrase effectivement reconnue comme telle, cette phrase demeurera inanalysable à partir de ce modèle. L'existence de telles phrases irréductibles peut amener à remettre en cause l'adéquation du modèle descriptif ou du moins son unicité » (*ibid.*).

De telles phrases, par exemple les constructions binaires comme : *la cathédrale, tout droit* ; *magnifique, ce livre ; Victor Hugo hélas!* sont effectivement signalées dans la grammaire, mais seulement dans le chapitre intitulé « phrases atypiques », où il est dit que :

> « cette construction binaire est très souvent irréductible, en tant que telle à la phrase canonique » (*op. cit.*, p.459).

Une description est d'ailleurs esquissée :

> « Ce type de construction binaire réalisé de manière parataxique l'association d'un argument et de son prédicat sur laquelle se greffe celle d'un thème et d'un propos » (*ibid.* p. 460).

Mais comme on le voit, cette description ne précise que les catégories interprétatives (prédicat-argument, thème-propos). Elle n'explicite pas la relation syntaxique qui se cache sous la formule « de manière paratactique », de sorte que l'on ne peut en tirer aucun principe d'analyse ni aucune extension terminologique qui permettrait de caractériser la structure syntaxique des énoncés binaires. On voit que les auteurs n'appliquent pas leurs propres principes, puisque contrairement à ce qui était affirmé, on ne propose pas de nouveaux moyens d'analyse pour décrire la réalité linguistique dans sa complexité, mais c'est cette réalité que l'on accommode au modèle : est atypique ce qui n'est pas naturellement décrit par le modèle choisi au départ.

Or, il est clair que si l'on étend le corpus des énoncés observés, le nombre des énoncés atypiques organisés « de manière paratactique » va

s'accroître considérablement (*cf.* Deulofeu, 1977). Ainsi, on ne signale pas comme construction binaire des énoncés à élément nominal antéposé, bien attestés à l'oral et impossibles à ramener à une dislocation de la phrase canonique :

ce métier on se déplace tous les jours
il y a des régions on appelle pas ça de la même manière
la coupe du monde c'est la France qui a gagné
on ne détaille pas non plus l'analyse d'énoncés tels que :
ce garçon je veux plus lui parler

où la relation de coréférence n'est en fait pas obligatoire et où le statut de complément régi par le verbe de *ce garçon* est en fait impossible à prouver, ce qui met en difficulté l'analyse par « détachement ». N'est pas non plus soulevée la question de l'analyse en « constructions binaires » d'énoncés en apparence canoniques :

à mon avis il n'est pas rentré
quant à lui je ne l'aime pas

mais où le constituant initial, au vu de ses propriétés syntaxiques et malgré la présence de la préposition, n'est pas régi par le verbe.

L'entreprise qui consiste à décrire l'ensemble des énoncés attestés à partir du modèle de la phrase canonique n'est donc pas aisée, pour ne pas dire impossible. Pour justifier leur position, les auteurs renvoient à des débats datant de la première étape de la grammaire générative transformationnelle. Or la référence à la « kernel sentence » est aujourd'hui abandonnée par les promoteurs de ce modèle eux-mêmes. Ainsi dans le modèle « minimaliste » dernier avatar de la GGT, les nexus prédicatifs non standard et certaines formes d'emphase sont données comme des types « basiques » et non comme des sous-types que l'on pourrait dériver de la « kernel sentence » (*cf.* Rebuschi, 1999) et les modèles de « small clauses ». De plus, la grammaire générative transformationnelle ne peut pas servir de référence unique pour une présentation syntaxique, même si la démarche transformationnelle a quelques avantages pédagogiques. Outre ses développements internes incompatibles avec une vision unifiée du modèle de phrase de base sujet-prédicat verbal tensé, la GGT est concurrencée par d'autres modèles syntaxiques, formels (HPHG, grammaire des constructions de Fillmore, diverses variantes de lexique grammaire) ou fonctionnels (divers avatars du fonctionnalisme de Lambrecht à Givon) dont un caractère commun est d'admettre que chaque langue peut avoir plusieurs types de phrases « de base » distincts et irréductibles. Une présentation pédagogique de haut niveau ne peut éluder

ces débats et ne saurait privilégier, sans discussion, un de ces modèles en particulier.

Le sérieux scientifique de l'ensemble de cette grammaire ne pouvant être mis en doute, les choix du chapitre 11 ont sans doute une source extérieure à la démarche linguistique. La marginalisation des énoncés « atypiques » et le privilège accordé à la phrase minimale sont sans doute liés à des considérations d'ordre sociolinguistique non explicitées.

2.3. Quand la pratique définit la théorie

Beaucoup des énoncés qui ne sont pas ramenables à la phrase canonique sont évidemment non normatifs et ne seraient jamais écrits, ce qui pourrait justifier leur marginalisation. Mais cette liberté prise à l'égard des données n'est pas compatible avec une attitude descriptive et revient a reconnaître implicitement que les objectifs descriptifs sont subordonnées à d'autres objectifs. Quels pourraient-ils être ? Cette grammaire n'a pas, en fait, explicité le problème fondamental de la différence de nature entre outils d'analyse « pratiques » et outils d'analyse descriptifs, au sens de Berrendonner (1989) repris dans Deulofeu (1995) :

> « la linguistique s'efforce d'être une science dont l'objet est le langage appréhendé à travers ses diverses manifestations écrites, orales, standard ou non standard. Les grammaires en revanche découlent d'une démarche praxéologique : elles se sont construites comme des rationalisations de l'activité visant à construire une version écrite, normalisée et fonctionnellement réduite (le langage réduit à sa fonction de représentation de la réalité) de telle ou telle langue [...] L'activité pratique des locuteurs, en particulier des professionnels de la langue : écrivains, grammairiens, livre des unités issues de catégorisations pratiques : mot, phrases. Ces deux systèmes d'unités sont loin de se recouvrir, comme l'ont montré un certain nombre de recherches récentes » *(op. cit.*, p.32).

Sur la base de cette distinction entre unités pratiques et unités descriptives, c'est en fait la notion même de phrase qui, pour certains linguistes, n'a pas de statut d'unité descriptive, mais seulement d'unité pratique (*cf.* Deulofeu, 1955, *loc. cit.*) :

> « La difficulté est bien relevée par Halliday (1986 :193) : « the clause complex will be the only grammatical unit which we shall recognize above the clause. Hence there will be no need to bring in the term « sentence » as a distinct grammatical category. We can use it simply to refer to the orthographic unit that is contained between two stops. » »

Ceci s'interprète aisément dans les termes de la distinction proposée : le « clause complex » est l'unité descriptive justifiée dans l'analyse du linguiste, la « sentence » est l'unité pratique établie empiriquement après des siècles d'efforts pour créer une langue anglaise écrite. Dans le corps de son ouvrage, l'auteur montre que, sur bien des points, l'unité descriptive et l'unité pratique en concurrence ne recouvrent pas en fait la même réalité. Berrendonner (1990) et Blanche-Benveniste (1990) ont proposé une réflexion élaborée sur le thème. Blanche-Benveniste aborde même, à travers les notions de « compétence grammaticale première » et « compétence grammaticale seconde », la question capitale de l'intériorisation du travail des grammairiens par les locuteurs. Ainsi, les locuteurs qui ont suivi un long entraînement scolaire ou professionnel en viennent à utiliser spontanément des tournures ou des constructions qui supposent au départ un enseignement explicite. Cet enseignement, qui se donne comme enseignement de « la » langue, est fondé sur deux principes implicites : standardisation et enrichissement. Des tournures spontanées sont bannies parce qu'elles ne paraissent pas conformes à l'objectif visé : créer une langue écrite efficace ; des tournures absentes de la « compétence première » des locuteurs sont introduites toujours avec le même souci pratique. L'enseignement de la grammaire dans nos pays n'a pas pour objectif de donner une connaissance de la langue, mais bien d'en assurer une pratique, essentiellement écrite, efficace.

Les « types de phrase » de notre grammaire pourraient donc en partie recouvrir non pas des unités descriptives, mais un ensemble de notions pratiques permettant de décrire non pas tous les énoncés possibles, mais bien ceux qui sont privilégiés par une approche pratique de la langue : ceux dont la maîtrise est « rentable » pour l'expression écrite normative. Si donc les notions de phrase et de types de phrase sont considérées comme des notions de grammaire « pratique » et non des concepts de linguistique descriptive, et si l'on fait l'hypothèse que la grammaire de Riegel a fait le choix d'une orientation pratique dans ce chapitre, on retrouve une cohérence dans la présentation. Ce chapitre affirme la primauté du modèle de la phrase comme unité « pratique » et les auteurs y rassemblent sous l'appellation « type de phrases » les énoncés qui peuvent poser des problèmes à une pédagogie de l'écrit :

- Structures s'écartant de la structure de la phrase simple assertive avec des problèmes de norme (interrogatives, impersonnel, passif). La situation est particulièrement claire pour les phrases interrogatives. Les phrases rassemblées ici ne présentent aucune cohérence formelle (elles mettent en jeu des « procédés » très divers : prosodie, marquage

morphologique (est-ce que), construction particulière (qu- phrases, inversion du sujet clitique). Le seul intérêt du regroupement est qu'il permet d'indiquer comment employer ces formes en fonction de la situation d'énonciation (style familier, standard, recherché) et de préciser des normes d'utilisation que des étrangers ou des francophones peu familiers avec la langue écrite peuvent ignorer (impossibilité d'emploi de l'inversion du clitique en interrogative indirecte).

- Énoncés qui du point de vue de la grammaire de l'écrit apparaissent comme des outils stylistiques à utiliser avec discernement et uniquement dans certains types de discours : procédés de mise en relief, énoncés à procédés d'emphase, phrases nominales, dont l'emploi est systématiquement donné comme marqué (titres, exclamations).
- Apprentissage des formes normatives et enseignement des valeurs stylistiques des formes sont deux objectifs importants pour les grammaires pédagogiques : ces deux principes semblent bien avoir aussi inspiré la rédaction du chapitre 11. On peut regretter que le choix n'ait pas été explicité.

Il faut sans doute voir ici l'effet d'une des contraintes particulières qui pèsent sur l'enseignement de la linguistique en France : il doit, par tradition, prioritairement préparer les étudiants à être des enseignants de langue française de l'enseignement secondaire avant même d'être un enseignement scientifique désintéressé. D'où la nécessité, pour ces grammaires de référence, d'être des outils de formation pédagogique autant que synthèses scientifiques. Malgré les efforts des auteurs, les débats théoriques sont donc toujours biaisés par des impératifs pédagogiques, ce qui est particulièrement le cas pour tout ce qui tourne autour de la notion de phrase, fondement de la pratique pédagogique normative de la langue écrite. Ces contraintes semblent peser spécifiquement sur les auteurs français, elles n'ont pas guidé les auteurs de grammaire de référence anglo-saxons, comme en témoigne, par exemple, l'absence de concessions de ce type dans la grammaire de Quirk *et al.* (1985).

3. Les « types de phrases » dans la Grammaire de la phrase française de Le Goffic (1993)

Considérons à présent la « grammaire de la phrase française » de Le Goffic, ouvrage également souvent recommandé dans les premiers cycles universitaires français et qui offre, sur bien des points, des analyses séduisantes et novatrices. Cette grammaire ne consacre pas de chapitre à la notion de type de phrase, mais elle y fait explicitement référence, en se donnant cet objectif dans l'introduction :

> « Tous les types de phrase, simple et complexe, verbale et non verbale sont analysés, avec de très nombreux exemples : non seulement Paul donne un livre à Marie, mais aussi Ce film, moi, j'ai adoré !. Ca me plairait de manger vietnamien, pour changer ! ou la douce chose que d'aimer ! Oral ou écrit, familier ou littéraire, le français est un ; les exemples illustrent surtout la langue des échanges quotidiens, mais aussi le langage littéraire (notamment Proust) ou la presse. » (*op. cit.*, p. 5).

Je vais cependant montrer que, malgré cet objectif descriptif affirmé et son souci de clarifier la nature des critères d'analyse utilisés, cette grammaire, dans son analyse des formes de la phrase, ne parvient pas non plus à faire coïncider totalement ses rubriques avec celles qu'une démarche descriptive supposerait. Cette difficulté me paraît là encore liée à une volonté de compromis entre une attitude cherchant à conserver une référence constante à l'unité « pratique » phrase et une attitude descriptive, qui admet l'existence en français d' énoncés impossibles à ramener au modèle « canonique » de la phrase.

3.1. Le problème du corpus de référence

Le premier parti pris concerne les données, le second, la méthode. Du point de vue des données, l'auteur, malgré les déclarations liminaires en vient à définir, parmi les énoncés possibles en français, un sous corpus à partir duquel il va établir les catégories d'analyse. Ce sous corpus est celui des énoncés qui se laissent ramener au modèle de « *la* phrase typique, de référence (qui) est la phrase assertive (conclusive) à l'indicatif... *Marie chante...* » (*op. cit.*, p. 8). Pour exclure de ce noyau dur, d'autres énoncés qui ne suivent pas ce modèle, on invoque la distinction langue / parole :

> « cette définition idéalisée est d'autant plus nécessaire que, dans leur discours effectif, les locuteurs sont loin de s'exprimer toujours par une succession de séquences normées correspondant au moule canonique de la phrase : ils produisent non pas des phrases, mais des énoncés (souvent

incomplets ou mal délimités). Mais la description de ces énoncés suppose référence au modèle de la phrase canonique » (*ibid*).

Ce parti pris de sélection des données est un premier indice que les plans des unités pratiques et des unités descriptives ne sont pas distingués clairement. Si « phrase canonique » peut relever du plan descriptif, « séquence normée », « énoncés incomplets » fait référence au plan des unités pratiques. Ces termes suggèrent que les énoncés dignes d'être décrits doivent observer certaines normes fonctionnelles. En gros et encore une fois, celles qui sont à la base de la constitution du style écrit standard. Cette grammaire va donc nous montrer essentiellement comment sont organisés les énoncés « sans fautes ». Elle accepte une tâche normative qu'elle pense pouvoir conjuguer sans dommage avec la tâche descriptive. De plus, l'utilisation du couple théorique langue/discours, pour opposer phrase canonique à énoncé « parfois incomplet » n'est pas sans problème.

Si l'opposition est comprise comme celle que l'on pourrait établir entre les structures disponibles dans la compétence des locuteurs et le mode de production de ces structures dans le discours effectif, tout linguiste est prêt à admettre qu'un énoncé tel que :

la la cette élè- non étudiante là est a chanté une comme ça voilà quoi

ne soit pas pris comme modèle d'une structure en langue (on ne va pas poser des groupes nominaux à deux articles !) mais soit ramené pour sa structure syntaxique au modèle canonique DET N Verbe DET N.

Mais le danger est que l'on considère également comme énoncés « incomplètement ou mal réalisés », et donc à analyser par rapport à la phrase canonique, des énoncés qui constituent des types d'énoncés en langue, différents du type sujet prédicat verbal, mais qu'il faut considérer d'un point de vue descriptif comme parfaitement « complets » : *le lait qui bout, un homme à la mer, merveilleux ce livre, à demain !* Il serait abusif, d'un point de vue descriptif, de ne pas les analyser en tant que tels et de les ramener à des énoncés canoniques qui en sont des paraphrases approximatives : *il y a le lait qui bout, un homme est tombé à la mer, ce livre est merveilleux, je vous dis à demain* (?). C'est une difficulté que perçoit l'auteur puisqu'il consacre un chapitre distinct de sa grammaire à la phrase sans verbe (Chapitre 17).

3.2. L'impossible analyse des phrases non canoniques

3.2.1. Quelles relations syntaxiques au-delà de la rection grammaticale ?

Cependant le parti pris méthodologique prend ici le relais pour interdire une analyse cohérente de ces énoncés sans verbe. Ce parti pris méthodologique consiste en effet à établir les catégories d'analyse sur les seules phrases à prédicat verbal et à essayer d'en étendre le champ aux phrases sans verbe. D'un point de vue descriptif, il n'y a pas de raison qu'on y parvienne et c'est d'ailleurs à un tel constat qu'aboutit l'auteur en fin de chapitre 17 :

> « en l'absence de relation (et d'accord) sujet-verbe, il ne va pas de soi que l'on continue de parler de sujet ». « le GN en position détachée fait penser plutôt à un thème postposé qu'à un sujet au sens strict. » [...] On parlera donc de sujet avec prudence, sachant qu'il n'est pas toujours aisé de déterminer le rôle de tous les éléments en présence dans une phrase sans verbe ». (*loc cit*, p. 510).

C'est seulement les hypothèses méthodologiques de l'auteur qui créent le problème. La contrainte d'établir les termes d'analyse à partir d'un seul modèle de phrase canonique, le refus de considérer que le type d'énoncé est une notion syntaxique de base et qu'il faut prévoir plusieurs types d'énoncés distincts et diversifier d'autant les outils d'analyse sont à l'origine de la difficulté, comme ils l'étaient de celles de Riegel. Les citations précédentes montrent bien que les principes d'organisation des énoncés avec verbes ne sont pas transposables à certains énoncés sans verbes, mais l'admettre serait mettre en cause l'axiome de la priorité donnée à l'analyse des phrases avec prédicat verbal et établir une rupture entre la « phrase » comme unité pratique et les unités descriptives permettant d'analyser les énoncés, rupture que Le Goffic, pas plus que Riegel, n'est prêt à envisager.

Le Goffic distingue un certain nombre de plans selon lesquels on peut analyser la phrase. Plan prédicatif (division Sujet-Prédicat), plan des fonctions syntaxiques (inventaire des « compléments »), plan sémantique, plan thématique (division en unités thème-rhème). Chaque plan comporte un certain nombre de termes d'analyse, mais la relation entre les unités des différents plans n'est posée que pour la phrase type où elles se recoupent et recouvrent le domaine sans « résidu » (*op. cit.*, schéma p. 10) :

	Le chat	mange	la souris
Str. fondamentale :	Sujet	Prédicat	
Constituants :	GN	GV	GN
Fonctions :	Sujet	Verbe	Compl. Dir. (Objet)
Rôles thématiques :	Thème	Rhème	
Rôles sémantiques :	Agent	Action	Patient
Modalité de phrase :	Assertion		

Quand on prend des exemples d'énoncés plus complexes, la relation entre le plan des formes (prédication, fonctions) et celui des composantes interprétatives (rôles thématiques ou sémantiques) va poser problème. Au début de l'ouvrage, il semble que sur le plan des formes (plan des fonctions syntaxiques) on ne dispose que des termes suivants : sujet, prédicat, compléments essentiels du verbe complément accessoires.

Les termes tels que thème-rhème, qui interviennent au plan de la structure thématique, définissent le rôle discursif de constituants exerçant indépendamment une fonction syntaxique (sujet, compléments). La structure thématique serait donc simplement une structure interprétative, qui ne permettrait pas de définir des constituants syntaxiques :

La « structuration thématique » est donc bien distincte de la structuration logico-grammaticale ou fonctionnelle. Souvent difficile à appréhender, elle intéresse essentiellement certains aspects de l'interprétation, au plan sémantique, et surtout au niveau du texte (p. 15). Mais l'auteur ajoute :

> « A ce compte elle pourrait apparaître marginale pour la grammaire. Mais on n'oubliera pas que c'est la dynamique communicative qui... apparaît en particulier comme responsable de plusieurs types importants et systématiques) de structures syntaxiques non canoniques : énoncés disloqués (avec prolepse, phrases impersonnelles il y a un livre sur la table, phrases incomplètes (énoncé réduit à sa partie rhématique) » (*op. cit.*, p. 15).

Ceci veut dire en clair que pour les énoncés canoniques, il s'agit d'une structure interprétative, alors que pour les structures non-canoniques, qu'elles soient verbales ou non, le plan thématique « serait responsable des formes syntaxiques » de ces énoncés. Cette formulation quelque peu contournée revient en fait à admettre que les fonctions définies dans la phrase canonique ne permettent pas d'analyser syntaxiquement les énoncés non canoniques. Pour y parvenir il faudrait donc définir de nouvelles fonctions syntaxiques en dehors de la complémentation verbale. Cela est évident pour les phrases sans verbe, mais même pour les phrases verbales, il apparaît que la classe des compléments classés comme non-essentiels extra-prédicatifs (*à mon avis, franchement, enfin...*) n'a en fait pas de statut

syntaxique dans le modèle. Ces compléments sont définis de façon purement négative en terme fonction syntaxique ou grammaticale : ils ne sont en aucune manière régis par le verbe ni comme catégorie (non-essentiels), ni comme prédicat (extra prédicatifs) et rien n'est dit sur le type de relation qui les intègre à la phrase :

> « De plus certains éléments syntaxiquement accessoires (tels que des adverbes de phrase : enfin) peuvent faire partie de la phrase sans appartenir en propre ni au sujet ni au prédicat » (*op. cit.*, p.8).

On voit, en revanche, qu'ils pourraient être définis de façon positive en termes de relations du « plan thématique » (ce seraient des constituants thèmes d'une structure d'énoncé réduite à la division fonctionnelle thème-rhème dont le rhème serait la construction verbale).

Mais pour résoudre ainsi le problème de l'analyse syntaxique de l'ensemble de ces éléments, il faudrait que le « plan thématique » permette de définir des unités syntaxiques et ne soit plus une simple composante interprétative. Il faudrait faire la distinction, indispensable d'un point de vue descriptif, entre unités syntaxiques à part entière donnant un statut syntaxique à des constituants non construits par le verbe (éléments détachés, adverbes de phrase comme *enfin*, *etc.*) et notions interprétatives précisant la valeur informative de constituants syntaxiques indépendamment définis. C'est à ces notions qu'on se réfère lorsqu'on caractérise tantôt comme thème, tantôt comme rhème, selon le contexte discursif, le constituant syntaxique sujet de beaucoup de gens sont venus (*cf. op. cit.*, p.15). Pour échapper à la confusion, il faudrait d'abord cesser de désigner par le même terme de « thème » des objets linguistiques aussi différents. Les relations syntaxiques distinctes des fonctions complément ou sujet qui permettraient d'y parvenir sont d'ailleurs esquissées par l'auteur dans le chapitre 3 intitulé : « les relations : linéarité et structure ». On y définit une « syntaxe des positions » autonome par rapport à la syntaxe des fonctions. Mais cette syntaxe est ramenée à la notion confuse de dynamique communicative, démarquée des positions du Cercle de Prague. Les notions comme début ou fin de phrase (p. 57) ont un statut qui oscille entre celui de constituants d'une structure syntaxique distincte de la syntaxe des fonctions et celui simples procédés de hiérarchisation de l'information contenue dans les constituants majeurs de la phrase verbale. Cette grammaire, pas plus que celle de Riegel, ne distingue, à côté des constructions grammaticales, des types d'énoncés.

3.2.2. Absence de prise en compte de la diversité des « nexus » prédicatifs

Le même statut ambigu entre composante interprétative et principe de structuration de la forme des énoncés affecte la division dite logico-grammaticale entre sujet et prédicat. Ce sujet logico-grammatical recouvre tantôt le sujet grammatical (phrase verbale) tantôt la notion de thème (phrases non verbale). Le problème tient à ce que, là encore, le modèle de référence pour l'analyse est la phrase canonique verbale. En effet, poser une relation syntaxique de prédication, distincte à la fois de la relation verbe-complément et de la relation thème-rhème, est utile lorsqu'on admet qu'il existe des phrases dont le prédicat n'est pas un verbe et qui ne sont pas pour autant de simples structures binaires rhème-thème. Ainsi à côté d'énoncés comme *magnifique, ce livre,* où la prédication n'est pas assurée par une catégorie grammaticale, mais par l'organisation de type « paratactique », on trouve des énoncés organisés autour de « nexus » prédicatifs non verbaux, notions que certaines analyses utilisent depuis Jespersen (voir notamment l'utilisation que fait de cette notion Ericksson, 1993 dans sa présentation de « la phrase française ») : *un carreau de cassé, le lait qui bout, un homme à la mer.* Cette notion de nexus trouve d'ailleurs son correspondant en grammaire générative sous le nom de « small clause », structurées comme des prédications de phrases ordinaires, à l'absence près de marques temporelles. La prise en compte d'une telle notion permet de définir des types de phrases en fonction de la nature de leur prédicat. Elle suppose que l'on accepte une dissymétrie dans les fonctions : alors que l'on va distinguer complément de verbe et complément de nom, ces notions étant définies par rapport à des systèmes de valences propres à chaque catégorie grammaticale, on parlera de sujet par rapport à une structure de prédication avec plusieurs catégories « prédicatives » possibles : relatives, groupes prépositionnels, participes...

Si l'on confère un statut structurant et non interprétatif au plan d'analyse en prédication ou nexus, la distinction entre compléments accessoires intra et extra prédicatifs devient inutile. Les compléments accessoires intraprédicatifs, sont des compléments régis par le prédicat, qu'il soit verbal ou non, tandis que les extra-prédicatifs trouvent leur statut syntaxique dans des relations de type paratactique vues plus haut. Il est d'ailleurs suggéré dans la grammaire elle même que les non essentiels extra-prédicatifs constituent des énonciations autonomes et sont affectés de poids informatifs spécifiques.

Il apparaît donc qu'un modèle descriptif des types d'énoncés en français doive, d'une part, distinguer entre syntaxe de la rection grammaticale (complémentation et prédication) et syntaxe des

« regroupements paratactiques » élémentaires interprétés comme thème-rhème dont la nature proprement syntaxique est à clarifier ; et, d'autre part, en syntaxe grammaticale, admettre des types distincts de nexus prédicatifs. Ceci permettrait une analyse non *ad hoc* et plus raffinée des phrases sans verbes (avec ou sans nexus verbal, avec telle ou telle structure « paratactique ») et de traiter avec les mêmes concepts, les compléments non essentiels extra-prédicatifs des phrases verbales au statut grammatical difficile à cerner. Mais il est clair qu'une telle démarche aboutirait à souligner la différence entre phrase au sens pratique et type d'énoncé au sens descriptif, ce qui apparaît comme une conséquence indésirable pour une grammaire qui cherche à allier des objectifs pédagogiques et des objectifs scientifiques.

Ce double parcours nous aura permis de clarifier les rapports entre approche « pratique » et approche descriptive des types de phrase. Si nous voulons articuler les deux points de vue, nous avons vu qu'il fallait commencer par bien les distinguer. On pourrait réserver la notion de type de phrase à la constitution d'une typologie des énoncés retenus par une approche pratique de la langue et dont les praticiens de la langue écrite recommandent l'usage dans telle ou telle situation d'énonciation particulière. Cette notion a donc plutôt sa place dans des ouvrages de techniques d'expression que dans des grammaires de référence et il est clair que dans cette perspective, le modèle de la phrase canonique verbale assertive reste central et les principes de la typologie peuvent reposer sur les fonctions en discours des divers types de phrases. Si l'on veut au contraire proposer des outils pour décrire les pratiques réelles des locuteurs, à l'écrit comme à l'oral spontané, il faudrait parler en termes de types d'énoncés et de types de constructions et non de types de phrases pour bien distinguer les unités descriptives des unités pratiques.

Nous avons vu que cette entreprise suppose que l'on s'appuie sur une double série d'unités syntaxiques : les unités organisées par les principes de rection et les unités organisées par les principes « paratactiques » en œuvre dans les phrases nominales et verbales non canoniques. L'enjeu est de montrer que l'on peut encore parler d'organisation syntaxique pour ces derniers principes et pas seulement des effets d'un « plan thématique » au statut incertain. Or, il existe effectivement des modèles syntaxiques qui proposent cette double approche de la syntaxe des énoncés et qui ont développé en conséquence une double terminologie. Si nous voulons que nos grammaires de référence soient descriptivement adéquates, nous devons donc rapidement passer en revue les concepts analytiques

développés par ces approches et voir ce que l'on peut en retenir pour une terminologie de référence des types d'énoncés.

3.3. Quelles unités descriptives recouvre la notion de type syntaxique de phrase ?

Vers les années 90, les nécessités de la description du français, notamment des énoncés d'oral spontané, ont amené un certain nombre de chercheurs à concevoir l'analyse syntaxique selon un modèle modulaire, avec une composante de syntaxe rectionnelle et une composante faite d'unités d'analyse distinctes des unités morphosyntaxiques, explicitement nommée macrosyntaxe (Blanche-Benveniste, 1990 ; Berrendonner, 1991) ou simplement distinguée en référence au corpus (grammaire de l'oral, intégration discursive des constituants, par opposition à leur intégration grammaticale chez Morel, 1992, 1998). Les unités dans les approches « macrosyntaxiques » sont définies selon des principes de variation concomitante de signifiant et de signifié dégagés depuis longtemps par des chercheurs travaillant à étendre l'adéquation empirique du cadre structuraliste classique (Bally , Perrot). Il n'y a pas à proprement parler de filiation historique entre les approches macrosyntaxiques et les démarches pionnières. Leur développement est la simple conséquence d'une démarche commune reposant sur le respect de la tradition structuraliste et de l'application conséquente de cette démarche à des corpus « problématiques » (langues non indo-européennes pour Perrot, variétés d'oral spontané pour les autres).

En tout état de cause, ces approches macrosyntaxiques permettent pour l'instant de donner dans beaucoup de cas une solution conforme aux principes fondateurs de la linguistique aux problèmes de description qui défient les syntacticiens. Elles ont malheureusement créé chacune leur terminologie, ce qui brouille la perception que l'on peut avoir d'une syntaxe au-delà de la rection. Il n'est donc pas inutile de présenter cette composante telle qu'elle se dégage des travaux de Perrot, les premiers dans ce domaine, où le lien avec la démarche de la linguistique structurale est bien établi.

4. Syntaxe et macrosyntaxe

4.1. Différence entre perspective fonctionnelle et morphosyntaxe globale

Perrot a formulé à diverses reprises les principes à partir desquels on peut évaluer les approches qui prétendent adopter dans l'analyse une perspective « informative ». Je les reprends ici sous une formulation récente (Perrot, 1997). C'est en se fondant sur le respect de ces principes que l'on pourra distinguer celles qui renouvellent l'analyse syntaxique en dégageant une composante macrosyntaxique de celles qui se contentent, dans la ligne de ce que nous avons vu dans les deux grammaires examinées, d'enrichir d'une perspective d'organisation de l'information les composantes interprétatives de la description. Perrot introduit la nécessité d'une analyse « morphosyntaxique globale des langues » à partir de ce rappel des axiomes fondamentaux de la démarche linguistique (les soulignements sont de mon fait) :

> « Le noyau dur, dans une science, est précisément ce qui résulte des caractères les plus fondamentaux de l'objet à étudier. Dans le cas des langues, on est fondé à voir leur trait le plus fondamental dans la propriété qui était au centre de l'enseignement de Ferdinand de Saussure lorsqu'il faisait de la linguistique une partie de la sémiologie et caractérisait la langue comme un système de signes associant un signifiant et un signifié. Il résulte de là que *l'identification des unités de la langue doit être fondée sur les variations concomitantes du signifiant et du signifié*, à tous les niveaux, et qu'un système linguistique s'organise précisément dans sa spécificité, sur la base de cette concomitance de variations. Ce qui ne s'inscrit pas dans le champ ainsi défini intéresse certes le linguiste, mais se situe hors du noyau dur et en revanche tout ce qui s'inscrit dans ce champ appartient au noyau dur.
>
> Ce dernier point est essentiel, et c'est précisément au nom de ce principe que l'analyse des phrases de la langue doit intégrer l'ensemble des structures syntagmatiques qui reposent sur une variation concomitante du signifiant et du signifié : non pas seulement les structures syntaxiques au sens traditionnel, sur lesquelles repose l'organisation de l'énoncé à partir des constituants liés les uns aux autres par des fonctions syntaxiques, mais aussi les structures qui organisent l'information contenue dans les phrases en tant que porteuses de messages. Tel est le sens de l'analyse morphosyntaxique globale dont les principes sont repris ci-après et qui vise à montrer qu'un ordre de valeurs traditionnellement traité par la stylistique appartient en fait au noyau dur de la linguistique. » (*op. cit.*, p. 34).

La morphosyntaxe globale articule donc une composante dont les unités sont établies selon les principes d'analyse des faits de rection avec

une composante comportant les structures « qui organisent l'information contenue dans les phrases en tant que porteuses de messages ». Le fait que les unités de ces structures soient établies dans le respect du principe de « concomitance » permet d'opposer cette approche à celles qui interprètent l'héritage de Bally en termes purement conceptuels :

> « L'autre principe de portée générale dont le respect conditionne le contrôle des faits analysés résulte de la nécessité de distinguer les points de vue selon lesquels les données linguistiques sont abordées. Les considérations qui précèdent ont déjà montré le danger d'une confusion entre les références dont les signes sont porteurs et les fonctions qu'ils assument dans la hiérarchie syntagmatique. Le point de vue « sémantico-référentiel » est l'un des trois points de vue que distingue C. Hagège, en reprenant sous d'autres dénominatifs ce que l'école de Prague considérait comme trois niveaux de l'analyse. Les deux autres points de vue correspondent l'un à l'organisation morphosyntaxique, l'autre à l'organisation du message (point de vue énonciatif hiérarchique). [...] Le niveau grammatical (Prague) ou point de vue morphosyntaxique (Hagège) correspond à la structuration de l'énoncé, le niveau du message (utterance dans l'expression en anglais de la thèse de Danes) ou point de vue énonciatif hiérarchique (Hagège) correspond à la structuration informative. *Dans les deux cas, les unités hiérarchisées en cause sont à considérer à la fois dans leur expression morphologique et dans leur contenu référentiel : le troisième point de vue (sémantique ou sémantico-référentiel) n'est donc pas à considérer comme spécifique par rapport aux deux autres : il ne représente que la saisie d'une des deux faces - le signifié - que possèdent par nature les unités engagées dans la combinatoire syntagmatique des phrases, l'autre face correspondant à leur signifiant.* » (*op. cit.* pp. 53-54).

C'est à partir de cette position théorique que j'ai critiqué l'organisation des modèles sous-jacents aux présentations grammaticales examinées plus haut, qui étaient donc des modèles construits sur la base d'une morphosyntaxe à un seul niveau étendue par une composante interprétative. J'ai montré en détail les conséquences du non respect de ce principe sur la cohérence interne de leur présentation. Perrot dégage nettement l'autonomie des unités macrosyntaxiques (ici le support) par rapport au référent et leur statut de signe en établissant le parallèle avec les unités microsyntaxiques :

> « Dans les structures auxquelles appartiennent les unités significatives, se définissent les fonctions, de façon spécifique dans chaque langue ; ces fonctions, exprimées par des moyens morphologiques très divers, sont des valeurs abstraites qui dans la réalisation concrète des phrases se sémantisent pour la saisie de la réalité extralinguistique ; le principe vaut aussi bien pour le message que pour l'énoncé. Dans l'organisation syntaxique de l'énoncé, la fonction subjectale reçoit un contenu sémantique variable d'une langue à

l'autre : elle recouvre des modes de participation au procès très divers dans une langue comme le français où le système verbal comporte une voix passive opposée à la vois active ; la fonction sujet ne se laisse alors définir qu'à un haut niveau d'abstraction, exprimé par la notion de partenariat privilégié avec le verbe dans la relation de prédication, ce partenariat correspondant à un rôle d'agent ou de patient selon la voix du verbe et se manifestant - c'est en définitive le seul critère solide - par le phénomène d'accord.

De la même manière la fonction support, par exemple, peut être remplie par des constituants de nature très variée : un syntagme nominal ou une subordonnée en fonction de circonstant, un adverbe de phrase, etc. : Pierre, je l'ai déjà rencontré, » dans un mois, il sera libéré » ; « si je peux, j'irai aux états Unis » ; « naturellement, il n'est pas satisfait ». La fonction informative est la même à travers des sémantisations diverses qui, dans la structure de l'énoncé, peuvent correspondre à des réalités morphosyntaxiques très diverses » (*op. cit.* pp. 53-54).

« la fonction grammaticale de support se sémantise comme personnage impliqué dans l'information à poser ou comme circonstance locale, temporelle, etc., ou comme hypothèse, etc. [...] les mêmes observations vaudraient pour le report, dont la fonction comme constituant postrhématique correspond à une stratégie du locuteur et n'implique aucunement que, d'un point de vue objectif, l'élément placé en rappel soit négligeable. Souvent il ne l'est pas, notamment dans les interrogations : le domaine évoqué par le mot interrogatif (qui est rhématique), mot par nature purement catégoriel, est fréquemment précisé par un élément en rappel : ta mère, elle a quoi, comme machine ? » (*op. cit.*, 1994 : 18).

4.2. Les unités macrosyntaxiques

A partir de ce cadre général d'analyse, et surtout de ce principe clair d'établissement des unités, Perrot a proposé un ensemble de concepts analytiques pour organiser le deuxième niveau morphosyntaxique :

« ...les deux structurations, celle de l'énoncé et celle du message, ont en commun de comporter un élément nucléaire nécessaire et suffisant pour constituer l'unité énoncé ou l'unité message : le prédicat dans le cas de l'énoncé, le rhème dans le cas du message ; l'énoncé peut donc se réduire à un prédicat et le message au rhème. [...] la structure syntagmatique qui s'organise par contraste autour du noyau rhématique affecte bien le contenu informatif du message, ou plus exactement la stratégie à laquelle le locuteur a recours pour présenter ce contenu : le noyau, qui a pour fonction de poser l'information, peut être précédé d'un segment qui a pour fonction d'introduire un élément servant de support, en quelque sorte, à cette information, et peut être suivi d'un élément qui, l'information étant posée, y apporte un complément dont le contenu, d'une manière ou d'une autre assure

> mieux l'interprétation du message. Plutôt que les termes de thème et de rhème, rendus dangereux par l'usage qu'on en a déjà fait dans la tradition, on a proposé de parler de support et d'apport, et le troisième terme nécessaire est alors celui de report pour le constituant post-rhématique. La réalité linguistique de cette structuration du message est validée par la prise en compte du signifié et du signifiant conjointement ; le signifiant fait intervenir essentiellement la prosodie, qui souvent constitue à elle seule le signifiant, tandis que la rupture syntaxique que constitue l'éjection en français produit un marquage fort du contraste entre support et apport. » (*op. cit.*, pp. 40-41).

Pour la discussion approfondie des critères d'établissement de ces unités, je renvoie aux articles de référence (notamment 1994 : 20-24, 1997 : 11-15).

4.2.1. Relations syntagmatiques et relations paradigmatiques

Le projet de Perrot consiste donc bien à résoudre les problèmes de description syntaxique par une extension des concepts et des méthodes linguistiques structuralistes utilisés pour les unités significatives au domaine des unités « informatives ». Perrot attire aussi notre attention sur d'autres possibilités qu'offre le cadre structuraliste d'étendre les limites de la description. Il s'agit d'exploiter au mieux la notion de relation paradigmatique entre unités :

> « D'autre part, la structuration de l'information comporte, outre ce contraste entre constituants, le marquage particulier d'un constituant, marquage qui est interprété comme une mise en valeur de ce constituant, mais dont la fonction doit être définie d'une manière plus précise à l'aide de notions concernant la charge informative du message. Les termes souvent utilisés de focalisation ou d'emphase s'appliquent au contenu informationnel dont le rhème ou apport est porteur, lorsque le constituant qui véhicule ce contenu est posé comme objet d'une sélection qui exclut les autres termes appartenant au même paradigme. Le français standard réalise ce marquage en recourant à c'est ...qui/que [...] Ce type de marquage de l'apport, appelé ici emphase, constitue dans l'organisation de l'information un phénomène paradigmatique en face du phénomène syntagmatique qui établit le contraste entre l'apport le support et le report. » (*op. cit.*, 1997 : 41).

Ceci veut dire que certaines constructions micro ou macrosyntaxiques seront caractérisées par la mise en relation d'une unité avec les autres unités du paradigme syntaxique auquel elle appartient. Certains marqueurs sont donc la face signifiante de ces opérations sur l'axe paradigmatique. L'emphase ou la focalisation n'est pas seulement une opération sémantique, c'est une unité linguistique comportant un signifiant et un signifié. Le modèle de Perrot permet de décrire ces énoncés non canoniques comme des

types morphosyntaxiques fondamentaux et non comme des déformations d'une hypothétique structures canonique sous l'effet d'une composante interprétative.

Une source de malentendus pourrait subsister, qui porte sur les relations entre les termes énoncé, phrase et message. On a vu plus haut que Perrot a renoncé aux termes thème et rhème, en raison de leur ambiguïté. Il a cependant fait une concession, à mon avis malheureuse, à la tradition en retenant le terme de phrase, pour désigner la structure morphosyntaxique globale « maximale ». Comme je l'ai suggéré, il y a nécessairement confusion à la lecture entre la phrase de Perrot, entendu au sens descriptif et l'usage galvaudé de phrase par tout un tas d'approches qui confondent sous ce vocable unité pratique et unité descriptive. Pour continuer le travail de clarification terminologique je propose les distinctions suivantes. Le terme neutre désignant la réalisation des structures morphosyntaxiques globales devrait rester celui d'énoncé. Un énoncé est la résultante de deux structurations : celle de la morphosyntaxe de rection et celle du message. Je propose d'utiliser le couple neutre microsyntaxe - macrosyntaxe pour désigner respectivement ces deux structurations. On parlera donc de structure microsyntaxique de l'énoncé pour référer à son organisation par les relations de rection (détermination, actance, prédication) et de structure macrosyntaxique de l'énoncé pour désigner son organisation par les relations qui le constituent comme message. Macrosyntaxe et organisation en message seront donc largement synonymes. La différence étant celle du point de vue dont on saisit les unités linguistiques : signifiant, pour macrosyntaxe, signifié, pour message.

4.2.2. Éléments pour une typologie syntaxique descriptive des énoncés

Une telle typologie suppose donc que l'on se place des deux points de vue : celui de l'énoncé comme construction (relations microsyntaxiques) et celui de l'énoncé comme message (relations macrosyntaxiques). Il faut définir d'un côté les types de messages possibles et de l'autre les constructions « grammaticales majeures » qui peuvent entrer dans chaque type de message. Une typologie syntaxique des énoncés suppose que l'on fournisse la combinatoire de ces deux modules. Je ne propose ici que les principes d'une telle typologie. Certains classements sont encore objets de débats, mais si le principe d'une syntaxe à deux modules est adopté, on peut faire l'hypothèse que l'on aboutira à des classifications cohérentes des types d'énoncés.

a. Types macrosyntaxiques

Un message comporte nécessairement un noyau « apport », qui n'est pas obligatoirement une construction verbale :

hélas !
tant mieux
quel désastre
Marie Chante

mais il peut prendre des formes plus complexes dont les plus fréquentes sont support apport ou apport report. Des constituants incises peuvent figurer entre ces unités ou à l'intérieur d'elles. La façon la plus élémentaire de réaliser ces unités est d'utiliser des éléments non linguistiques : geste déictique pour symboliser le support + mimique appréciative comme apport. mais bien évidemment des constructions grammaticales les plus variées peuvent figurer en apport (caractères droits) et en support (italiques). Ces constructions peuvent être microsyntaxiquement indépendantes :

ces souliers hélas !
ces souliers un travail d'artiste
ces souliers j'écrase les pieds de tout le monde
à mon avis il est malade
enfin nous y sommes arrivés

ou reliées par une relation de rection :

de ces souliers je ne parlerai pas

b. Types microsyntaxiques (constructions grammaticales) :

- Types de nexus

Parmi les constructions qui peuvent former le noyau d'un message, on doit au moins distinguer les constructions prédicatives des constructions nominales. Signalons en passant que les seuls énoncés qui auraient un titre à être appelés « phrases nominales » sont les énoncés constitués d'une construction nominale noyau de message :

le beau livre
la prise de Constantinople par les Turcs

Pour s'en tenir ensuite aux constructions prédicatives, on rappelle qu'il faut distinguer plusieurs types de prédicats ou de nexus, chacun d'entre eux peut constituer le noyau d'un message :

Marie chante
Marie qui chante

les livres sous la table, la vaisselle éparpillée, l'armoire éventrée, on a tout eu quoi

Des extensions de cette notion de prédication peuvent rendre compte des structures suivantes qui forment noyau de message, mais où on a de bonnes raisons d'analyser le présentatif comme auxiliaire de prédication et non comme prédicat principal :

il y a Marie qui chante
il y en a à qui je pourrai jamais parler
(la corrida) c'est le matador qui joue avec le taureau

Les formes de prédication constituent donc une dimension essentielle d'une typologie syntaxique des constructions pouvant faire noyau d'énoncé Bien évidemment la prédication à prédicat verbal conjugué doit faire l'objet d'un développement détaillé : c'est là que les phénomènes d'actance et de diathèse sont à l'origine de sous-types d'énoncés importants qu'il faut caractériser. Outre les types de prédications d'autres constructions sont à prendre en compte.

- Types à opérations sur l'axe paradigmatique

Les énoncés à prédicat verbal tensé peuvent être enrichis de constructions particulières qui amènent à des réalisations de messages spécifiques. On peut citer notamment les constructions qui consistent à isoler à gauche de la construction verbale un élément régi par le verbe et à le marquer d'une opération paradigmatique particulière (diverses formes d'extractions et de pseudo-clivées) :

à qui tu parles
à Paul je parle
c'est à Paul que je parle

Toutes ces constructions peuvent constituer un message indépendant comme les prédications, mais comportent une opération paradigmatique particulière.

- Autres types de constructions

On a pu aussi proposer d'identifier une construction avec inversion du sujet clitique, qui a des réalisations contraintes comme élément de message (forme des noyaux à modalités contraintes, qui ne peuvent être des assertions constatives) :

est-il bête ? / ! /

Une construction « présentative » avec sujet postposé et adverbe locatif antéposé, qui impose une modalité affirmative :

au fond étaient assises quelques personnes
au fond n'étaient pas assises quelques personnes
au fond quelques personnes n'étaient pas assises

On vient de lister le minimum de constructions distinctes que l'on doit poser pour établir une typologie des énoncés. Il va sans dire que ce minimum devra être complété si l'on veut couvrir l'ensemble des formes d'énoncés attestées.

4.2.3. Articulation de cette typologie avec d'autres domaines de la description

Cette typologie fondée sur le signifiant peut être aisément articulée avec d'autres, qui prendraient notamment en compte les possibilités, pour tel ou type d'énoncé, de recevoir telle ou telle modalité d'énoncé représentant tel ou tel acte de langage. Certains pensent que ces deux typologies peuvent se recouper directement :

> « we say that a sentence type exists when a certain communicative function is conventionally associated with a particular grammar structure » (Lambrecht et Michaelis, 1993 : 235).

Mais la relation est en fait indirecte (*cf.* Prince, 1998), ce qui justifie une double typologie, comme celle qui oppose fonctions syntaxiques et rôles sémantiques dans la complémentation verbale. Ainsi certains types syntaxiques sont libres de recevoir l'ensemble des modalités d'énoncé y compris la marque d'évaluation subjective que représente l'exclamation. C'est évidemment le cas des prédications verbales sans inversion de clitique, mais d'autres se combinent préférentiellement avec certaines modalités, comme par exemple les constructions à inversion du clitique sujet. Il faut cependant souligner que la liberté de combinatoire est souvent beaucoup plus grande qu'on ne croit. Ainsi les constructions généralement qualifiées d'interrogatives peuvent-elles recevoir d'autres modalités. La construction avec inversion du clitique sujet peut jouer le rôle d'un support non modalisé :

a-t-il soif on lui sert immédiatement à boire

ou même d'un apport assertif, à valeur exclamative :

est-il intelligent !

De même la construction « clivée » avec un pronom *qu-*, ne constitue pas nécessairement des noyaux de message à valeur d'interrogation partielle, elle peut en effet prendre une valeur assertive, notamment dans la construction dite « relative de liaison » :

à qui tu n'aimes pas parler ?
J'ai rencontré Pierre. A qui tu n'aimes pas parler, je crois.

Ces exemples montrent bien la nécessité de proposer une typologie des énoncés sur base syntaxique, puisque, aussi bien, aucune construction n'est associée de façon univoque avec une modalité de phrase.

5. Quelles présentations possibles des types de phrase dans les grammaires de référence ?

5.1. Une présentation sur bases descriptive est-elle possible ?

Au bout du compte, une typologie syntaxique à visée descriptive des énoncés possibles en français ne semble pas imposer un surcroît important de terminologie, par rapport à l'approche « pratique ». Trois ou quatre termes permettent de décrire la structure de l'énoncé comme message (apport, support, report, incise, dans le cadre de Perrot ou leurs équivalents dans les autres approches « macrosyntaxiques ») ; quelques autres viennent à bout des constructions grammaticales : prédication verbale et non verbale, éléments appartenant à la valence verbale, éléments régis par le prédicat, constructions clivées, constructions à inversion du clitique sujet....Dans ce domaine, seule la prise en compte des nexus à prédicats non verbaux et le rejet de certains compléments circonstanciels traditionnels hors de la rection verbale entraîne un changement dans les habitudes scolaires d'analyse. Le gain de ces aménagements est évidemment que l'on est libéré de notions *ad hoc* devenues inutiles : phrases non verbales, détachements, emphase, phrases atypiques, compléments extra-prédicatifs et bien d'autres (*cf.* Deulofeu, 1995 : 28-32). Rien ne semblerait donc s'opposer à l'adoption d'une terminologie de ce type.

L'obstacle le plus difficile à surmonter est en fait de l'ordre des représentations que l'on se fait de la langue. Il tient à la difficulté qu'a le public lettré que constitue la communauté universitaire à repenser la syntaxe en dehors du modèle de la phrase scolaire. Des années d'apprentissage nous ont fait intérioriser le modèle et le métalangage qui va avec. Des habitudes aussi ancrées ne sont pas faciles à modifier.

5.2. Une stratégie pour une adaptation « pratique » réaliste du modèle descriptif

Si l'on veut créer chez le public cultivé des dispositions pour une attitude descriptive objective, il peut paraître indispensable d'abandonner la notion de type de phrase pour celles de type de messages et de types de constructions. C'est seulement alors qu'il pourra appréhender le français écrit et oral comme un tout cohérent. La condition nécessaire pour y parvenir est de lui faire remettre en cause sa représentation de la langue française fondée sur le primat de la phrase simple écrite. Une stratégie possible pour y parvenir semble être de confronter les futurs utilisateurs de grammaires, le plus tôt possible, aux réalités de la langue spontanée telle qu'elle apparaît à l'observation. C'est donc pendant le cours de sa formation supérieure que ce public doit être confronté à la complexité et à la spécificité de la langue spontanée, que le modèle scolaire antérieur occulte, et cela sans doute à juste titre. Dans notre expérience au Département de Linguistique Française à Aix, nous avons constaté que les étudiants se sentent démunis au moment d'analyser du français parlé authentique avec les outils de la grammaire pour la phrase française écrite. On peut profiter de ce léger désarroi pour suggérer des outils nouveaux. Ainsi, l'enregistrement et la transcription de français oral spontané jouent ce rôle de révélateur et fournissent l'occasion d'une remise en cause des représentations « scolaires de la langue ». Pour ce qui est des notions permettant d'analyser les types d'énoncés difficiles, nous avons facilité le passage d'une syntaxe à l'autre par une étape de transition qui utilisait deux termes par le biais desquels le lien s'établissait entre micro et macrosyntaxe : la notion d'élément associé à une construction noyau d'énoncé, qui met de l'ordre dans les « circonstants » et les « dislocations ». Et la notion de dispositif d'une construction pour couvrir le domaine des nexus non canoniques et des clivées. Il serait intéressant de tester les possibilités pour ce métalangage de transition d'être utilisé à plus grande échelle. Des années d'aller et retours entre pratique et théorie des énoncés d'une langue semblent nécessaires pour réaliser une rupture qui ne peut être que salutaire aux pédagogues comme aux enseignants de linguistique.

Références

Bally, C. (1944). *Linguistique générale et Linguistique française*. Berne : A. Verlag.

Berrendonner, A., Reichler-Beguelin, M.-J. (1989). Décalages : les niveaux de l'analyse linguistique, *Langue Française, 81*, 99-125.

Berrendonner, A. (1991). Pour une macro-syntaxe. In D. Willems (éd.), *Données orales et théories linguistiques* (pp.25-31). Paris –Louvain : Duculot.

Blanche-Benveniste, C., Bilger, M., Rouget, C., & Van den Eynde, K. (1990). *Le français parlé. Études grammaticales*, Paris : Éditions du CNRS.

Blanche-Benveniste, C. (1990). Grammaire première et grammaire seconde : l'exemple de en, *Recherches sur le français parlé, 10*, 52-73.

Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris-Gap : Ophrys.

Danon-Boileau, L., Morel, M.-A. (1995). L'oral : invariance et variations, *Intellectica, 1*.

Deulofeu, J. (1977). Les constructions binaires en syntaxe, *Recherches sur le français parlé, 1*.

Deulofeu, J (1995). Le dilemme de la modernisation des terminologies linguistiques, *Travaux de Linguistique, 31*.

Deulofeu, J. (1999). Questions de méthode dans l'étude du morphème que en français contemporain, *Recherches sur le français parlé, 15*, 163-198.

Fillmore, C.-J., Kay, P. (1992). *Construction grammar. A course book*. Berkeley : University of California.

Firbas (1966). On defining the theme in fonctional sentence analysis, *Travaux Linguistiques du Cercle de Prague, 2*, 239-256.

Givon, T. (1984). *Syntax : a functional typological introduction*. Amsterdam : J. Benjamins.

Hagege, C. (1982). *La structure des langues*. Paris : Presses Universitaires de France.

Halliday, M. (1985). *An introduction to functional grammar*. London : Arnold Edition.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge : Cambridge University Press.

Le Goffic, P. (1993) *Grammaire de la phrase française*. Paris : Hachette.

Michaelis, I., Lambrecht, K. (1996). Toward a construction based theory of language function : the case of nominal extraposition, *Language, 72/2*, 215-247.

Morel, M.-A., Danon-Boileau, L. (1998). *Grammaire de l'intonation*. Paris-Gap : Ophrys.

Morel, M.-A. (1992). Intonation et thématisation, *L'information Grammaticale, 54*, 26-34.

Perrot, J.(1967). Observations sur la structure de l'énoncé en hongrois, *B.S.L., 62/1*, 122-133.

Perrot, J. (1994). Éléments pour une typologie des structures informatives, *Mémoires de la Société de Linguistique de Paris, n.s.t. II : La phrase : énonciation et information, 13*

Perrot, J. (1997). Visée communicative. In J. Feuillet (ed.), *Actance et valence dans les langues de l'Europe*. Berlin : EUROTYP.

Prince, H. (1998). On the limits of syntax, with reference to left dislocation and topicalisation. In P. Culicover & L. Mcnally (eds.), *The limits of syntax., Syntax and Semantics 29*. New-York : Academic Press.

Résumés de thèses

Étiquetage semi-automatique de la prosodie dans les corpus oraux : algorithmes et méthodologie

Estelle Campione

Thèse dirigée par Jean Véronis, soutenue le 08 décembre 2001

L'objectif de cette thèse est de mettre en évidence la faisabilité de la transcription prosodique de grands corpus, à l'aide d'outils semi-automatiques destinés à (1) réduire le coût humain de la transcription, (2) lui apporter une base objective. En effet, alors que l'on dispose désormais pour l'écrit de corpus de dizaines, voire de centaines de millions de mots, les corpus oraux transcrits avec des indications prosodiques sont rares et dispersés. La prosodie semble pourtant indispensable à l'interprétation des transcriptions d'oral, et surtout à l'étude des relations qu'elle entretient avec les phénomènes syntaxiques et pragmatiques.

Nous proposons un codage prosodique « large » qui délimite seulement les unités majeures, dans une perspective de lisibilité et d'exploitation de grands corpus pour l'analyse des phénomènes syntaxiques et pragmatiques. Notre travail est sous-tendu par la volonté constante de limiter les présupposés théoriques, de façon à produire des corpus annotés les plus neutres possibles vis-à-vis des théories particulières, et nous avons fait le choix de transcrire seulement les événements qui correspondent à une réalité objective et qui sont par conséquent repérables par l'instrumentation tels que les proéminences mélodiques, les pauses silencieuses, les « marques du travail de formulation » (*euh* et allongements syllabiques), *etc.* L'exemple ci-dessous montre un fragment de corpus transcrit avec notre système. La notation est volontairement assez simple : les énoncés sont découpés en segments qui apparaissent chacun sur une ligne séparée, précédée de la position temporelle en secondes à partir du début de l'enregistrement ; les montées et descentes intonatives importantes sont notées par les flèches (↗ et ↘), les allongements syllabiques d'hésitation sont marqués par deux-points (:) ; les tirets (- et --) marquent deux niveaux de pause.

0.0	on on dit il y a une anecdote qui dit pourquoi: - Corton-Charlemagne ↘
4.5	-
5.1	ben parce que le: Charlemagne euh - paraît-il ↗ ne buvait que des rouges ↗
9.0	-
9.5	parce qu'il voulait pas ↗
10.4	-
10.9	il il se tachait ↗
12.0	-
12.6	sa: sa: - il ne ne buvait que* des Blancs pardon ↗
15.9	-
16.3	euh parce qu'il ne voulait pas ↗ se tacher ↗
17.9	-
18.2	sa: sa barbe ↗
19.7	-
20.2	et donc euh - c'est pour ça qu'un jour ^ les habitants de: - d'Aloxe-Corton: - ont dit ↗ mais alors →

A l'aide d'un corpus de travail constitué de près de six heures de parole (lue et spontanée) dans cinq langues (anglais, allemand, espagnol, français et italien), que nous analysons avec des outils informatiques et statistiques, nous proposons dans un premier temps une analyse rigoureuse et systématique des différents indices et de leur interaction. Nous étudions ainsi près de 60000 mouvements mélodiques et 6000 pauses, et l'influence de facteurs tels que la langue, le sexe, le débit ou le type de parole. Si nos résultats confirment le plus souvent, en les complétant, les études précédemment publiées, ils apportent parfois un point de vue nouveau, voire contredisent certaines données tenues pour acquises.

Dans un deuxième temps, nous proposons un algorithme qui prend en compte l'ensemble des contraintes et interactions pour produire en sortie un codage prosodique des corpus. Nous avons découplé de façon soigneuse d'une part la phase d'extraction d'indices, automatique ou manuelle, et d'autre part l'interprétation de ces indices à l'aide de règles et contraintes. Cette approche modulaire permet une meilleure compréhension du rôle des différents indices et un travail de mise au point plus aisé. Elle permet aussi une certaine indépendance de la langue : les modules de bas niveau sont applicables à diverses langues et ont été testés sur les cinq langues de notre corpus, et seul le système de règles d'interprétation est spécifique à chaque langue (dans le cadre de ce travail, nous proposons un tel système seulement pour le français).

Étude critique d'un remarqueur de la fin du XVIIe siècle : Nicolas Andry de Boisregard

Magali Seijido

Thèse dirigée par Claire Blanche-Benveniste, soutenue le 08 décembre 2001

Cette thèse est une étude critique d'un auteur peu connu - Nicolas Andry de Boisregard - qui a publié deux recueils de remarques complémentaires : les Réflexions sur l'usage présent de la Langue Françoise ou Remarques nouvelles et critiques touchant la politesse du langage édité en 1689 et la Suite des Réflexions Critiques sur l'usage présent de la Langue Françoise édité en 1693.

Le genre des Remarques et des Observations sur la langue du XVIIe siècle a été pendant longtemps peu étudié, pourtant il constitue un contrepoint intéressant aux grammaires traditionnelles : il a tendance à être axé sur des données plutôt que sur des théories. Les « remarqueurs » traitent des subtilités et des idiotismes de l'usage, de faits de langue dans le détail plutôt que dans l'ensemble ; ils portent un intérêt particulier aux usages douteux, à la variation, à l'instabilité et au changement de la langue. Leurs témoignages sont donc très précieux pour l'histoire du français et de sa standardisation. Cependant, pour un lecteur moderne, l'accès à ce contenu est parfois rendu difficile par la forme même de ces ouvrages : ils se présentent généralement comme des recueils de courtes remarques détachées. Il n'y a pas de lien logique entre les articles, et sans index, il est peu aisé d'y rechercher un phénomène précis ou de les exploiter pour l'étude d'un problème grammatical.

Le premier objectif de cette thèse est de proposer des outils qui permettraient de faciliter la consultation des Réflexions (1689) et de la Suite des réflexions (1693) et qui pourraient constituer un matériau pour l'histoire de la grammaire, pour l'étude de la variation et de la grammaticalisation. J'ai réalisé un classement moderne des remarques en cinq rubriques : prononciation et orthographe, grammaire, lexique, rhétorique et règles de bienséance, qui aboutit à une nouvelle présentation : j'ai « scanné » les remarques des deux recueils et je les ai reproduites en suivant cette réorganisation.

Je propose une présentation très générale des remarques grammaticales et une étude plus détaillée du verbe et des problèmes de cohésion syntaxique. En grammaire, la majorité des remarques aborde des questions liées à la morphosyntaxe des catégories. Toutes les catégories sont représentées : déterminant, pronom, substantif, adjectif, adverbe, préposition, connecteur. Andry fait de nombreuses observations sur le verbe et le groupe verbal ; il s'intéresse principalement aux valences mais il fait également des remarques sur la morphologie, l'accord, les modes et les temps.
Bien que ce classement ait demandé un travail de catégorisation, il n'a pas pour objectif d'aboutir à une grammaire organisée. Les « remarqueurs » et les grammairiens n'ont pas la même démarche. Andry, en refusant de faire une grammaire, en refusant l'ordre fondé sur les parties du discours, ne propose pas de théorie générale sur l'ensemble de la langue. Deux sortes d'erreurs d'interprétation consisteraient à prêter aux « remarqueurs » des théories qui n'ont pas lieu d'être ou à les qualifier de descriptivistes comme l'a fait Chomsky à propos de Vaugelas. Le rôle d'Andry est de participer concrètement à la standardisation du français en observant et en commentant principalement les usages qui posent problèmes et la variation. A titre d'exemple, si l'on prend le traitement qu'il fait des valences verbales, il ne s'intéresse pas aux principes généraux : la majorité de ses remarques portent sur des verbes qui ont plusieurs constructions concurrentes en usage. Il adopte alors plusieurs positions. La position la moins fréquente consiste à sélectionner une construction et à condamner l'autre. Dans d'autres cas, il accepte la « variation » et considère les deux constructions comme strictement équivalentes. Les constructions concurrentes peuvent également différer par une nuance de style mais le plus souvent il les distingue par le sens : elles peuvent correspondre à une simple opposition de traits (+ /- animé, + /- fort, + /- humain) ou à deux sens différents.
Dans cette thèse, je propose également des hypothèses sur l'évolution de la définition du bon usage au XVIIème siècle. J'interviens contre la simplification qui associe, sous le règne de Louis XIV, le bon usage à l'usage de la cour. Un lecteur aussi attentif que R. A. Lodge note que « au cours des cinquante-cinq ans du règne absolu de Louis XIV, la concentration du pouvoir dans la personne du roi assura le maintien du rôle de la cour comme garant du bon usage. » Or, Andry observe peu les usages en fonction de ce que nous appellerions « la stratification sociale ». La cour ne sert de modèle que pour les règles de prononciation. Pour discuter de la norme grammaticale, il se réfère principalement à des productions écrites : plus de cent ouvrages. Plusieurs types de productions sont représentés : le

domaine technique, le domaine poétique et le domaine de la littérature en prose. Un phénomène est tout à fait remarquable : il accorde une place centrale à la traduction ; d'une manière générale, elle tend à remplacer la cour en matière de bon usage.

Composition et mise en page
Estelle CAMPIONE
Estelle.Campione@up.univ-aix.fr

Maquette de couverture
PUP - Valérie JULIA

ACHEVÉ D'IMPRIMER
SUR LES PRESSES NUMÉRIQUES DE
DICOLORGROUPE
À AHUY (FRANCE)

Dépôt légal 2e trimestre 2002
ISBN 2-85399-506-2
ISSN 0224-0041