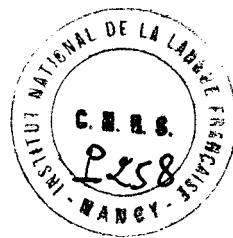


2 JUIL. 1997

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

RECHERCHES
SUR
LE FRANÇAIS
PARLÉ

N° 14 - 1997



1997

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

SOMMAIRE

Dominique WILLEMS, Histoire, linguistique et sources orales	11
René COPPIETERS, Quelques réflexions sur la question des données: corpus et intuitions	21
Gaston TUAILLON, Les calembours en toponymie dans le domaine francoprovençal	47
Mireille BILGER, Mylène BLASCO, Paul CAPPEAU, Berthille PALLAUD, Frédéric SABIO et Marie-Josée SAVELLI, Transcription de l'oral et interprétation; illustration de quelques difficultés	57
Claire BLANCHE-BENVENISTE, Transcription et technologies	87
Sonia BRANCA, Transcription et édition de manuscrits. Quelques problèmes autour de la "représentation" des textes	101
Paul CAPPEAU, Données erronées : quelles erreurs commettent les transpositeurs ?	115
Claire BLANCHE-BENVENISTE, A propos de <i>Qu'est-ce que c'est</i> et <i>C'est quoi..</i>	127
Marie-Noëlle ROUBAUD, Le sujet dans les énoncés pseudo-clivés	147
Christine ROUGET, <i>Espèce de, genre de, sorte de</i> à l'oral et à l'écrit	173

PARTICIPANTS

Mircille BILGER, Université de Perpignan
Claire BLANCHE-BENVENISTE, Université de Provence, Aix-Marseille 1
Mylène BLASCO, Université de Clermont-Ferrand
Sonia BRANCA, Université de Provence, Aix-Marseille 1
Paul CAPPEAU, Université de Poitiers
René COPPIETERS, department, Mason Hall, Pomona college, Claremont,
California, Etats-Unis
Berthille PALLAUD, CNRS, Université de Provence, Aix-Marseille 1
Christine ROUGET, Université de Vancouver, British Columbia
Marie-Noëlle ROUBAUD, Université de Provence, Aix-Marseille 1 et
IUFM
Frédéric SABIO, Université de Provence, Aix-Marseille 1
Marie-Josée SAVELLI, Université Stendhal, Grenoble III
Gaston TUAILLON, Université Stendhal, Grenoble III
Dominique WILLEMS, Université de Gand, Belgique

AVANT-PROPOS

Ce numéro de *Recherches sur le français parlé*¹ est construit autour du thème du recueil, de la transcription et de l'utilisation des données.

Dominique Willems réfléchit sur l'intégration des sources orales à d'autres disciplines que la linguistique (en particulier l'histoire) et montre que leur intégration permet une meilleure connaissance de l'évolution de la langue.

René Coppieters développe son expérience de linguiste générativiste face au problème des données attestées dans l'analyse linguistique.

Sonia Branca, à partir d'un travail d'édition de manuscrits de la période révolutionnaire, souligne à quel point il peut être utile de disposer de corpus transcrits au plus près.

Le groupe GEDO rappelle les problèmes liés à l'interprétation des données orales (les phénomènes de multi-transcription) et montre l'intérêt de noter certains faits prosodiques.

Paul Cappeau examine les erreurs les plus fréquentes qui sont commises par les transcrip-teurs débutants.

Gaston Tuaille, fort de son expérience de dialectologue, examine les relations entre transcription et interprétation(s) de certains toponymes dans le domaine franco-provençal.

Claire Blanche-Benveniste rend compte de l'état actuel des propositions concernant la transcription sous l'influence des nouvelles technologies. Dans un deuxième article, elle montre la nécessité de tenir compte des données orales pour compléter la description grammaticale des interrogatifs.

Marie-Noëlle Roubaud s'intéresse aux particularités du sujet dans les pseudo-clivées ; elle en dégage l'organisation et la spécificité.

Christine Rouget montre des différences significatives entre deux corpus oral et écrit à propos des formules *espèce de*, *genre de* et *sorte de*.

1 - Ce numéro 14 a été pour l'essentiel coordonné par les membres du Groupe d'Études sur les Données Orales. Contrairement à ce qui avait été annoncé dans le numéro précédent, le guide du transcrip-teur n'apparaîtra pas. L'article présenté ici par ce groupe n'en reprend que quelques réflexions.

HISTOIRE, LINGUISTIQUE ET SOURCES ORALES

0. En 1990, lors d'un colloque international sur le thème « Données orales et théorie linguistique » (Gand, mai 1990)¹, la réflexion s'était organisée autour de trois questions principales : (1) la linguistique de l'oral, une autre linguistique ? ; (2) les corpus oraux, pour quoi faire ? ; (3) la variation, où et comment la décrire ?

Cinq années plus tard nous aimerions reprendre les deux premières questions, en élargissant quelque peu la perspective.

Pour la première question, il nous paraît intéressant de comparer la linguistique, en ce qui concerne l'intégration et l'impact des sources orales, à d'autres disciplines où la même problématique se pose, et en particulier l'histoire². L'évolution, en 25 ans, d'une « histoire orale » en marge de l'histoire « traditionnelle » et utilisant ses propres canaux de diffusion (colloques, associations et revues), vers une histoire « sans adjectifs », qui intègre les sources orales parmi d'autres, est particulièrement révélatrice.

Pour la deuxième question, nous proposons de compléter les conclusions de 1990 à partir de l'analyse de travaux concrets utilisant des corpus oraux. Il s'agira de mesurer l'apport des données orales dans l'analyse de certaines structures syntaxiques particulières (listages,

¹ Cf. Travaux de Linguistique 21, 1990.

² Notre intérêt pour l'histoire orale découle d'une participation de plusieurs années au projet d'histoire orale du département d'histoire contemporaine de l'Université de Barcelone. Cf. en particulier M. Vilanova et D. Willems, Problems and methods in Oral History in Spain (1980), ainsi que *Lengua y poder en Cataluña en los años treinta* (1985).

structures à sujet inversé, pseudo-clivées), dans l'interprétation de la négation ou dans l'étude de langages spécifiques tel le discours aphasique³.

1. Si l'intérêt pour les sources orales en histoire est ancien, l'histoire orale en tant que mouvement et méthodologie spécifiques remonte aux années 1970⁴. Il est intéressant de se pencher sur les spécificités de cette approche, sur son évolution pendant ces 25 ans, sur l'impact qu'elle a exercé sur l'activité historique en général, et sur son influence en ce qui concerne l'analyse des sources écrites. De comparer finalement ce bilan avec la situation actuelle en linguistique.

En synthétisant considérablement⁵, les apports de l'HO se situent essentiellement sur quatre plans :

1.1. Sur un plan purement documentaire, les sources orales fournissent des *données complémentaires* par rapport aux sources écrites. Celles-ci sont souvent fragmentaires, même en ce qui concerne des événements historiques importants telle la deuxième guerre mondiale (par exemple, l'histoire de la résistance en France, ou celle de la bombe atomique au Japon). Sur un autre plan, l'histoire du continent africain par exemple nous est surtout connue à travers les documents écrits coloniaux, qui, quoique importants, ne représentent qu'un point de vue sur les événements.

A côté de leur caractère fragmentaire, il est intéressant de signaler que les archivistes nous mettent en garde contre un appauvrissement qualitatif du document officiel contemporain : l'utilisation de nouveaux moyens de communication (le télécopieur ou le téléphone par exemple) ainsi que

³ Nous nous sommes basés en particulier sur les thèses de doctorat suivantes : Callebaut B., *La négation en français contemporain. Une analyse pragmatique et discursive*, Gand, 1989 ; Cappeau P., *Le sujet postposé en français contemporain. Analyse syntaxique*, Aix-en-Provence, 1992 ; Loufrani Cl., *Analyse de discours de locuteurs classés comme aphasiques. « De l'art de tourner autour du mot »*, Aix-en-Provence, 1989 ; Ndiaye M., *L'analyse syntaxique des joncteurs*, Aix-en-Provence, 1989 ; Roubaud M.N., *Étude des constructions pseudo-clivées dans le cadre de l'approche pronominale*, Aix-en-Provence, 1994.

⁴ L'HO a fait sa première apparition au XIV^e Congrès des Sciences Historiques à San Francisco en 1975, sous forme d'une table ronde en marge du congrès officiel. Au congrès suivant à Bucarest en 1980, l'HO était officiellement représentée par une section propre.

⁵ Pour faire la présente synthèse, nous nous sommes largement basés sur les textes préparés en vue de la session spéciale consacrée à l'HO du XVIII^e Congrès des Sciences Historiques (Montréal, août 1995). Le thème proposé pour cette session, présidée par Ph. Joutard, était précisément l'évaluation de 25 années d'histoire orale. Les textes sont publiés en version espagnole dans *Historia y fuente oral*, 14, 1995.

l'exigence croissante de transparence publique se fait au détriment de la qualité des documents écrits⁶.

1.2. Sur le plan thématique, les sources orales permettent d'autre part d'*élargir le champ historique* en y incluant des domaines largement ignorés par les sources écrites traditionnelles : l'histoire des minorités (?) sociales et ethniques ou, plus globalement, de tous les groupes éloignés des centres de pouvoir, l'histoire de la vie quotidienne, des régions, des campagnes, etc. L'HO s'est ainsi en quelque sorte « spécialisée » en l'histoire des groupes par ailleurs invisibles (les femmes, les noirs, les analphabètes, etc.). C'est cette thématique particulière qui a contribué à faire de l'HO en quelque sorte une histoire « en marge », où la source orale - seule disponible - est considérée comme objet historique à part entière. Un domaine particulièrement fertile est celui des « histoires de vie », menant à une réflexion sur la mémoire historique et sur les représentations de l'histoire chez ceux qui l'ont vécue ou subie.

1.3. Mais c'est sans doute sur le plan méthodologique et plus largement épistémologique que se situe la contribution majeure de l'HO. Celle-ci s'est en effet, plus systématiquement peut-être que l'histoire traditionnelle, penchée sur des questions d'ordre méthodologique. C'est que, avec l'utilisation de la technique de l'entrevue pour la constitution des documents oraux, l'histoire, d'une science essentiellement interprétative, devient également une *science expérimentale*. L'historien perd sa position de pur interprète pour aider à créer lui-même sa source, dont, par le biais du mécanisme question/réponse, il est en quelque sorte co-responsable, sa subjectivité se mêlant à celle de son témoin. Il peut aussi dans une certaine mesure contrôler la source en multipliant les entrevues ou en les répétant dans le temps. Cette co-responsabilité de l'historien mène à un intérêt renouvelé pour le processus même de création d'une source historique : celle-ci n'est plus un savoir reçu, mais une connaissance à construire. On s'intéresse de plus en plus au contexte et aux conditions de production ; les erreurs, oublis et autres scories font partie du matériel historique⁷. Sur ce point, la terminologie en usage chez les historiens japonais est révélatrice : la langue parlée y reçoit le nom de langue « écoutée » et la transcription devient l'« écoute transcrite », mettant ainsi en lumière le rôle actif de l'historien dans le processus historique⁸.

⁶ Cf. J. P. Wallot, « Archivistique et histoire orale au Canada », in *Historia y fuente oral*, 14, 1995.

⁷ Cf. Ph. Joutard, « Le témoignage oral et la recherche historique : progrès ou déclin ? », *ibidem*, 1995.

⁸ Cf. Tadahide Hirokawa, « Les sources orales au Japon », *ibidem*, 1995.

La question la plus débattue est certainement celle de la *subjectivité* de la source orale : subjectivité considérée d'abord comme un handicap, conséquence inévitable d'une méthodologie spécifique ; subjectivité revendiquée par la suite comme objet historique à part entière ; subjectivité d'ailleurs largement partagée avec de nombreuses sources écrites.

Avec le problème de la subjectivité se pose celui, essentiel pour l'historien, de la *fiabilité* des documents oraux. Le substantif « sources » par opposition à « données », plus utilisé en linguistique, traduit bien cette préoccupation : le document comme point de départ d'une interprétation, comme preuve d'une hypothèse. Notons sur ce point l'attitude contradictoire qui consiste d'une part à accorder toute foi aux témoignages oraux dans le cadre des procès juridiques, ou aux témoignages écrits quels qu'ils soient, mais de se méfier d'autre part des sources orales en général. Une attitude qui n'est pas sans nous rappeler celle des grammairiens et des linguistes il n'y a pas si longtemps, assimilant trop facilement langue écrite à langue soignée, langue parlée à langue populaire et incorrecte... L'opposition individuel /v/ officiel ne coïncide pas avec l'opposition oral /v/ écrit : il y a des sources orales officielles (cf. la tradition orale) comme il y a des sources écrites individuelles (des lettres p. ex.). Il reste néanmoins que la source orale présente des caractéristiques particulières : elle est en effet artificiellement suscitée en vue d'une exploitation historique, alors que la source écrite est « spontanée » dans la mesure où sa raison d'être ne se situe pas dans son exploitation historique. Par ailleurs, la source écrite est souvent simultanée aux événements décrits, le témoignage oral étant, lui, le plus souvent postérieur. L'opposition mémoire/histoire prend ici tout son sens.

Les archivistes posent eux aussi le problème de la fiabilité des sources, mais sur le plan plus concret de la qualité des documents eux-mêmes : ils insistent sur la nécessité d'une transcription bien faite, d'une contextualisation précise du document et s'interrogent sur les possibilités de réutilisation. Ils posent finalement le problème de la conservation des documents oraux (stabilité physique, compatibilités techniques etc.). Le temps du dilettantisme est révolu. Finie aussi l'époque des premières archives orales (Université de Columbia), où jusqu'aux années 60, seule la transcription était considérée comme document historique, le support oral de l'enregistrement étant tout simplement détruit...

Les études d'HO affichent finalement un intérêt marqué pour *l'interdisciplinarité* : les histoires de vie concernent tant les

anthropologues et les sociologues que les historiens ; la source orale se prête par ailleurs à une extension vers l'image et les médias ; les rapports avec la littérature (e.a. les travaux sur la narrativité) sont aussi fréquemment à l'ordre du jour.

1.4. Ces aspects spécifiques de l'HO, surtout sur le plan thématique et méthodologique, ainsi que la méfiance des historiens classiques expliquent sans doute l'évolution en première instance vers une discipline autonome possédant ses canaux de diffusion propres. Ainsi, depuis 1977, les colloques internationaux d'HO se succèdent à intervalles réguliers. On assiste également à la création dans divers pays d'associations d'HO et de revues spécialisées⁹. A ce mouvement d'autonomie a succédé récemment un mouvement d'intégration : certaines revues disparaissent ou changent de perspective ; le terme même d'HO fait place à celui plus neutre d'histoire à base de sources orales. En bref, l'accent est mis sur la complémentarité des sources et l'unicité de la recherche historique.

Dans cette évolution globale, il est intéressant de constater des différences importantes d'un pays à l'autre : si en Angleterre ou en Italie l'HO s'est constituée un terrain propre, représentant aussi une idéologie marquée, en France, l'utilisation de sources orales n'a jamais donné lieu à une discipline séparée : ceci se traduit par une réticence à l'emploi du terme même d'HO, l'absence d'une association spécifique ou d'une revue propre. Les raisons en sont bien connues¹⁰ : citons en particulier l'aspect centralisateur de la culture française et sa méfiance par rapport à la diversité des parlers et l'oral en général et l'apparition d'autre part dès 1929, avec la revue *Annales*, d'une « nouvelle » histoire, orientée vers l'étude des mentalités et utilisant les sources orales parmi d'autres bien avant l'essor de l'HO.

2. En passant du champ historique au champ linguistique, il est intéressant de constater un certain nombre de parallélismes, ainsi que des différences significatives.

2.1. Sur le plan *thématique*, la linguistique de l'oral s'est attaquée à certains sujets et niveaux d'analyse privilégiés. Sur le plan discursif, il va de soi que l'analyse de la conversation ou du dialogue - genre éminemment oral - occupe une place importante. Certains centres de recherche en ont fait leur principal objet d'étude (cf. Genève ou Lyon).

⁹ Cf. e.a. Oral History (Angleterre), Journal of Oral history (Etats-Unis), Oral History Review (E.U.), Historia y fuente oral (Espagne), Fonti Orali (Italie).

¹⁰ Fait souvent souligné cf. Ph. Joutard, 1983 ; M. Trebitsch, 1992.

Les spécialistes de l'oral se sont également intéressés plus particulièrement aux phénomènes d'énonciation et de subjectivité, qui occupent à l'oral une place plus prépondérante qu'à l'écrit.

Les différences dans le mode de production, et plus particulièrement la multidimensionnalité de l'oral, ont mené à un intérêt renouvelé bien qu'encore modeste pour la prosodie¹¹ d'une part, les phénomènes gestuels de l'autre. En syntaxe, on s'est également intéressé aux phénomènes propres à la « production » de la langue parlée : répétitions, ellipses, hésitations, piétinements, dislocations, parenthèses e.a. (cf. les études du GARS à Aix-en-Provence). Ces phénomènes touchent surtout le niveau de la macrosyntaxe, où les différences entre oral et écrit sont les plus manifestes. Une place à part devrait être attribuée aux recherches portant sur la langue parlée des aphasiques, mettant précisément en lumière les problèmes que peut poser la production avec ses processus de sélection et de combinaison.

Parfois ces recherches typiquement « orales » ont jeté de nouvelles lumières sur certains phénomènes de l'écrit : E. Roulet a ainsi mis le doigt sur les aspects dialogiques des monologues écrits ; d'autres, ont étudié la « gestation » de l'écrit à travers l'analyse de brouillons.

Sur le plan de la syntaxe propositionnelle, on s'est en premier lieu intéressé aux structures qui différaient de façon sensible de celles de l'écrit (les formes spécifiques de l'interrogation, la disparition du *ne* négatif, l'emploi de relatives non standard etc.). On a constaté que ces différences étaient en réalité peu nombreuses et qu'elles permettaient souvent de mieux éclairer le système de la langue tout court, l'analyse de l'oral complétant en quelque sorte les zones d'ombre de l'écrit. C'est le cas par exemple du système des pronoms derrière l'impératif (*donne m'en /v/ chantes-en*) qui ne devient compréhensible qu'en prenant en considération aussi bien les données orales qu'écrites. Il en va de même pour les phénomènes de l'apposition et de la coordination dont l'étude gagne en profondeur en les situant dans le cadre plus large des « listes », dont l'oral présente une exploitation plus variée.

La plupart des études concernent des phénomènes qui existent tant à l'oral qu'à l'écrit. La prise en considération de données orales a mené :

a) à la révision de certains outils descriptifs, telle la notion de « phrase » ou de « subordination » par exemple. Toutefois, plutôt que

¹¹ Des progrès importants dans l'analyse prosodique ont par ailleurs été faits dans le domaine de la synthèse de la parole, dans une perspective toutefois différente.

de surimposer un outillage nouveau à la panoplie existante, on songe à réviser celle-ci afin d'y intégrer les données orales, à créer en quelque sorte un outillage passe-partout ;

b) au déplacement des limites syntaxiques et lexicales de certains phénomènes : ainsi, une étude récente sur le pseudo-passif révèle une utilisation moins contrainte de cette structure à l'oral qu'à l'écrit ; le contraire se fait jour dans le cas des structures à sujet inversé, dont l'oral se sert moins que l'écrit du moins dans le contexte d'une phrase simple ;

c) à la délimitation de zones où l'usage est flou ou hésitant, comme dans le cas de la variabilité du pronom relatif « lequel » ou du genre du mot « espèce » (*une espèce de fleur* / *un espèce de retour*) ;

d) à la possibilité finalement de prévoir dans quel sens pourrait aller l'évolution de la langue (cf. la fréquence à l'oral des structures relatives du type « le pays que je suis né » ou des questions indirectes du type « je me demande qu'est-ce qu'il fait » par exemple).

2.2. Sur le plan *méthodologique*, mis à part le fait que la prise en compte de l'oral nécessite des méthodes spécifiques pour l'obtention des données (enquêtes, enregistrements, manipulations postérieures), le fait le plus frappant est sans doute l'orientation essentiellement descriptive de la plupart des recherches sur la langue parlée.

Bien plus qu'une opposition entre linguistique de l'oral et linguistique de l'écrit, il s'agit d'une opposition entre une démarche inductive et descriptive, que les données de départ soient écrites ou orales, et une démarche essentiellement déductive et hypothétique, la vérification pouvant se faire là aussi sur des données d'observation écrites ou parlées, à côté des données d'intuition. Si depuis les années soixante, l'intérêt pour la description a été mitigé, on constate les dernières années un regain important de ce type de recherche en particulier sous la pression des projets de traitement informatique du langage naturel, où le manque de descriptions précises et d'inventaires exhaustifs se fait cruellement sentir. Notons par ailleurs que si la linguistique s'est longtemps détournée de l'étude des réalisations concrètes, pour s'intéresser au système plus abstrait, ceci ne s'est pas fait uniquement au détriment de la langue parlée : les réalisations proprement écrites n'ont pas non plus fait l'objet de descriptions précises (cf. les études bien éparpillées sur le paragraphe, la ponctuation, la variété des formes de l'écrit etc.).

Sur le plan méthodologique, l'analyse de l'oral présente toutefois des contraintes spécifiques : la nécessité de travailler sur un corpus de données d'observation et les difficultés que présente le recours à des manipulations. Là où l'étude du système abstrait peut se faire sur des données d'intuition obtenues par introspection, ceci n'est pas le cas pour la description de l'usage, parlé ou écrit. Comme le signale F. Gadet¹² « on ne peut saisir les conditions d'emploi d'une forme qu'en faisant appel à des considérations statistiques ou de fréquence qui ne jouent pas habituellement un rôle central dans le travail grammatical ». A cela s'ajoute une difficulté particulière pour l'oral : « Notre capacité réflexive sur la langue (en tant que locuteur et que linguiste), bâtie sur de l'écrit standard, s'applique difficilement à d'autres données ». Et compte tenu de la dissymétrie de statut entre langue standard et langue familière, il serait illusoire d'interroger un « locuteur natif ». Signalons pour finir que l'intonation, qui joue un rôle décisif à l'oral, est plus difficilement manipulable que les éléments segmentaux. Il convient donc d'être particulièrement prudent en expérimentant avec des données orales.

2.3. La nécessité de récolte des données par corpus pose également le problème de la conservation et de la réutilisation de ces données. Le problème a été clairement posé par les historiens et plus particulièrement par les archivistes en ce qui concerne les sources orales. Les linguistes se sont également posé la question et récemment des efforts d'harmonisation ont été entrepris dans le domaine de l'édition des textes oraux, en ce qui concerne les normes de transcription et l'accessibilité des corpus¹³. Sur un autre plan, il serait certainement utile de contextualiser d'avantage les exemples oraux : dans une étude récente, chaque exemple oral était accompagné d'une réelle « carte d'identité » explicitant tous les aspects pertinents de la situation d'énonciation.

3. L'utilisation des données orales en linguistique n'a pas donné lieu à une « autre » linguistique. Elle a par contre complété tant l'analyse du système, en y ajoutant des dimensions supplémentaires ou des données permettant de combler les « trous » de l'écrit, que celle de l'usage. Un aspect original concerne les péripéties de production de l'oral. Si l'objet est donc partiellement différent, les méthodes d'analyse rejoignent celle de la linguistique descriptive, bien que sporadiquement on puisse trouver des

¹² Cf. « Les outils grammaticaux au risque de l'oral non standard » in *Données orales et théorie linguistique*, Travaux de Linguistique, 21, 1990, p.14.

¹³ Cf. e.a. la banque de données orales connue sous le nom de Valibel (M. Francard).

données de langue parlée dans des approches de nature plus déductive¹⁴. En conclusion, en linguistique, il s'agit plus d'amplification du système que d'un système spécifique, d'intégration que d'indépendance.

En histoire, les changements paraissent plus profonds. L'utilisation de sources orales, si elle n'a pas vraiment donné lieu à une autre histoire, bouleverse tout de même profondément l'objet historique en reposant le problème de la vérité historique, de la subjectivité et du lien entre histoire et mémoire. Sur le plan méthodologique, la technique de l'entrevue pose le problème du rôle de l'historien dans la création de la source et ajoute au caractère essentiellement interprétatif de l'histoire une dimension expérimentale certaine.

Dominique WILLEMS

¹⁴ Cf. A. Zribi-Hertz, « L'oral, la syntaxe et l'astérisque : questions méthodologiques, avec et sans réponses », *Lynx*, 18, 1988.



Références

- Analyse grammaticale de corpus oraux*, 1988, Lynx, 18.
- BLANCHE-BENVENISTE Cl. et JEANJEAN C., 1987, *Le français parlé*, Paris, Didier.
- BLANCHE-BENVENISTE Cl. et alii, 1990, *Le français parlé : études grammaticales*, CNRS.
- « Données orales et théorie linguistique », 1990, *Travaux de Linguistique* 21 (éd. D. WILLEMS).
- FRANCARD M., 1989, VALIBEL. « La première banque de données sur les variétés orales du français en Belgique », in *Travaux de Linguistique* 18, pp.165-168.
- GADET F., 1990, « Les outils grammaticaux au risque de l'oral non standard », in *Travaux de Linguistique* 21.
- Historia y fuente oral*, 14, 1995 « Por una historia sin adjetivos »
- JOUTARD Ph., 1983, *Ces voix qui nous viennent du passé*, Paris, Hachette.
- Recherches sur le français parlé*, Université de Provence (13 volumes).
- TREBITSCH M., 1992, *Du mythe à l'historiographie*, in D. VOLDMAN, « La Bouche de la Vérité », *Cahiers de l'IHTP*, 21.
- VILANOVA M. et WILLEMS D., *Problems and methods in Oral History in Spain*, in *Rapports du XV^e Congrès International des Sciences Historiques*, Bucarest, 1980, pp. 546-554.
- WILLEMS D., « Lenguaje escrito y lenguaje oral », in *Historia y fuente oral*, 1, 1989, pp.97-107.

QUELQUES RÉFLEXIONS SUR LA QUESTION DES DONNÉES : CORPUS ET INTUITIONS

A priori, on pourrait croire que la question des données sur lesquelles baser l'étude du langage n'ait pas lieu de se poser. Toutes les données peuvent apporter leur contribution : intuitions et jugements de grammaticalité, collections de corpus oraux et écrits, expériences psychologiques, neurophysiologie, etc¹. En pratique, pourtant, il est loin d'être toujours clair comment harmoniser ces différents types de données. Les grandes orientations théoriques en linguistique se sont presque toujours caractérisées par le statut privilégié qu'elles attribuaient à telle ou telle catégorie de données et tout naturellement le sens qu'elles ont reconnu au terme « langage » s'en est trouvé profondément affecté. Je discuterai ici deux types de données, réflexion et jugements intuitifs d'une part, corpus de langage oral de l'autre, et leurs affinités avec des attitudes fondamentalement différentes envers le langage.

I. 1. Pour comprendre la place que les jugements d'intuitions en sont venus à occuper en linguistique, il faut sans doute rappeler un peu d'histoire. Au départ, il pourrait sembler plus naturel de prendre le langage parlé, et le corpus créé en collectant ces emplois, comme base et réalité linguistique la plus immédiate. C'était là, dans les années quarante et cinquante, la position de l'école de Bloomfield aux EU, ce qu'on a appelé le structuralisme américain, qui ne reconnaissait comme donnée valable que le comportement langagier, et en particulier le corpus oral, qu'elle

¹ C'est d'ailleurs ce qu'affirme N. Chomsky (1986), qui déclare aussi que la grammaire générative n'est pas une théorie, mais un topique de recherches. Dans ce travail, j'emploierai le terme GG pour faire référence à la tradition théorique et méthodologique initiée par Chomsky, même si celle-ci s'est effectivement de plus en plus diversifiée, et inclut à ce stade, au delà des théories successives développées par Chomsky lui-même, des théories aussi diverses que la grammaire relationnelle, Lexical-Functional Grammar, Head-Driven Phrase Structure Grammar, etc.

confondait d'ailleurs avec toute la réalité linguistique, et sur lequel elle tentait principalement d'appliquer une analyse phonologique.

Mais concevoir le réel n'est pas simple. Suivant une approche purement externe, le structuralisme américain ne voulait aborder le comportement que dans sa réalité physique, prétendant expliquer le langage humain sans faire référence au domaine mental, compris comme inefficace et d'ailleurs trop subjectif pour pouvoir être un objet de science. Seul existait vraiment le comportement tel qu'on peut l'observer de l'extérieur. La psychologie béhavioriste refusait tout rôle actif aux opérations mentales dans le comportement, lequel devait pouvoir se comprendre à travers un lien d'association presque direct entre input et output, stimulus et réponse, ainsi que le cri « Aïe ! » semble associé directement à une stimulation douloureuse. La théorie prônait donc la réductibilité du langage à des facteurs, stimulus et réponse, purement externes, que seule l'histoire de l'organisme pouvait rendre, par couches d'associations successives, plus compliqués. Le point de vue intérieur à l'esprit/cerveau, celui de l'esprit/cerveau qui donne forme et représentation mentale à la réalité extérieure, manipule cette représentation pour en déduire un comportement complexe et à la limite imprévisible, n'était donc ni possible ni surtout justifiable.

La révolution cognitive en psychologie et en linguistique a retourné notre compréhension du problème, et l'attitude chomskienne en linguistique doit se comprendre en réaction aux excès du béhaviorisme, excès qu'avaient d'ailleurs évités les écoles structuralistes en Europe. Depuis sa célèbre critique de *Verbal Behavior*, livre écrit par un des chefs de file de la psychologie béhavioriste, Skinner, Chomsky a fait remarquer dans une longue série d'ouvrages que le langage n'est pas seulement une réalité externe, mais aussi, et d'abord, une réalité interne et mentale. Le langage n'est pas transparent mais opaque à toute description naïve, c'est-à-dire à toute description qui n'attribuerait pas à l'organisme une forme de « connaissance préalable » du sujet abordé, et cela parce que l'être humain lui-même ne vient pas au monde comme une « page blanche », mais possède et développe une structure mentale propre, extrêmement complexe, dont l'origine ne peut être que son héritage biologique. C'est cette structure qui organise et donne forme à l'interaction de l'homme avec le monde. De même que les sens, le langage n'offre donc ni un simple moyen d'enregistrement, ni un reflet immédiat et transparent du monde extérieur. Cette remarque s'applique, en particulier, aux éléments mêmes du discours. Les catégories de sons à travers lesquelles nous interprétons la chaîne parlée ne se retrouvent pas comme telles, physiquement, dans le donné acoustique. Il faut déjà connaître une langue, et cela à tous les niveaux (phonologique, morphologique, lexical, syntaxique, etc.) pour pouvoir reconstruire les séries de phonèmes, morphèmes, etc. à travers lesquels un

locuteur analyse les énoncés produits « dans sa langue ». C'est pour cette raison précisément que la transcription d'un corpus oral représente, comme le font bien remarquer D.-C. Giovannoni et M.-J. Savelli 1990, un travail cognitif complexe : c'est un travail d'interprétation, de reconstruction de ce que les locuteurs ont pu dire, et non la production mécanique d'un fac-similé écrit d'un texte oral. De même, les catégories sémantiques en termes desquelles nous classifions notre univers ne sont pas de simples copies conformes de la structure du monde, et leur organisation reflète l'organisation et le mode d'appréhension de l'esprit humain tout autant que la nature. Le système linguistique peut donc se comprendre comme une forme de calcul engendrant des représentations mentales à toute une série de niveaux, d'un niveau phonétique concret à un niveau sémantique abstrait. Dans ses grandes lignes, et même dans les détails, ce calcul est régi par la structure biologique et innée de l'être humain².

C'est en fait cette pré-construction du langage qui rend possible son acquisition. C'est elle aussi qui sous-tend la diversité apparente des langues du monde. La grammaire générative (GG) dans ses dernières incarnations (théories du gouvernement et du liage, des principes et des paramètres, minimalisme), en portant son attention sur la question de la grammaire universelle (GU), tente précisément de mettre à jour ces formes *a priori* du langage, en particulier dans le domaine syntaxique.

Il y a donc, à la base de l'approche générative, affirmation du caractère irréductiblement mental du langage. On comprend alors la prépondérance accordée aux données de l'intuition, qui semblent bien sûr le mieux à même de nous donner des indices sur la structure interne du langage, tout en devenant elles-mêmes les phénomènes clefs que toute théorie linguistique se doit d'expliquer en priorité. En syntaxe et en sémantique il a d'ailleurs toujours été indispensable de tenir compte des données de la réflexion intuitive : jugements de grammaticalité, d'acceptabilité, d'opposition, de congruence avec un contexte donné, etc. La GG a simplement formulé un cadre théorique et méthodologique qui justifie et rend compte de cette situation.

On remarque très vite, cependant, qu'il y a parfois divergence entre les jugements intuitifs et la réalité langagière de ce qu'on dit. Un corpus de conversation familière transcrit sur papier peut même présenter aux yeux des locuteurs qui l'ont produit un côté scandaleux, tant il offre d'hésitations, de répétitions, de ruptures de construction, de phrases tronquées, sinon de langage « incorrect ». On pourrait prendre comme preuve de cette

² Ceci est vrai indépendamment de la réponse à donner à la question de la spécificité biologique du langage, c'est-à-dire de l'existence d'un module linguistique indépendant, isolé, et fonctionnant sur des principes différents du reste de la cognition humaine, question qui reste encore débattue.

in correction potentielle de l'oral le fait que les locuteurs se corrigent parfois eux-mêmes dans la parole. A la suite de Saussure, Chomsky va alors privilégier la langue, à laquelle il donne une interprétation psychologique, et qu'il appelle la compétence, sur la parole, la performance³. Au niveau des données, cela veut dire que la GG va privilégier la réalité des intuitions que nous avons sur notre langage, dans toute leur richesse et leur complexité, à la réalité de l'emploi du langage dans la parole. La GG à ses débuts rejoint ici indirectement la grammaire normative en comprenant la performance, le langage oral en particulier, comme potentiellement fautif, rempli d'erreurs d'inattention, de scories de toutes sortes. On explique cette situation en concevant la performance comme un reflet indirect du « vrai » langage tel que nous le connaissons et qui par définition ne peut se tromper, notre compétence. La performance reflète la complexité de la mise en œuvre concrète de cette connaissance dans tout l'ensemble des mécanismes et processus physiologiques et mentaux qui donnent naissance au comportement externe (compréhension qui renvoie à une vision modulaire du cerveau/esprit, vision qui s'est depuis largement vérifiée). Puisque les données immédiates de la performance orale spontanée sont ambiguës, les linguistes vont devoir travailler d'après leurs intuitions sur des exemples construits spécifiquement pour tester les hypothèses qu'ils auront avancées sur la compétence, notre connaissance mentale du langage. On insistera même sur le parallélisme de méthode entre recherche scientifique dans les sciences naturelles et construction de mini-expériences en linguistique. De cette manière la réunion d'un corpus oral (ou écrit) devient à la limite non-pertinente, en tout cas inadaptée à l'étude de la compétence, du fait qu'elle n'inclura que tout à fait accidentellement les exemples critiques nécessaires à la confirmation d'une hypothèse donnée, et qu'il faudra d'ailleurs toujours construire et vérifier le statut de tout l'éventail d'exemples que cette hypothèse prédit comme agrammaticaux.

L'enthousiasme déclenché par la grammaire générative est né du sentiment que l'on avait trouvé là un cadre théorique et méthodologique fécond et solide qui devait permettre à terme à la raison humaine de construire une compréhension théorique du langage comparable à notre compréhension théorique dans les domaines naturels.

³ On a souvent fait remarquer qu'il existe une différence essentielle entre Chomsky et Saussure. La langue, pour Saussure, est un phénomène social et c'est la société entière qui en est la garante et la reproductrice, alors que la compétence est une réalité purement psychologique. Chomsky subordonnerait donc le social au psychologique alors que Saussure comprenait le psychologique comme se conformant à, ou formé par, le social. Cette opposition apparente n'est en fait pas aussi claire qu'il n'y paraît. Sous le terme de langue, Saussure visait principalement le lexique, et n'avait pratiquement rien à dire sur la syntaxe. Pour Chomsky également, l'origine du lexique est sociale, et au départ la théorie de la compétence n'avait pas grand chose à en dire. C'est d'abord la syntaxe qui a fait l'objet de son étude.

La distinction plus récente que N. Chomsky 1986 établit entre *I-language* et *E-language* reprend sous une autre forme les mêmes notions. Seule la connaissance mentale interne au cerveau/esprit, ce qu'on appelait autrefois la grammaire, offre un objet approprié à la linguistique. Le langage externe, quel que soit son origine, et indépendamment de son statut grammatical, ne peut rendre compte de ce que l'être humain connaît quand il connaît une langue. Décrire la connaissance linguistique que possède un locuteur, c'est décrire la structure complexe du système de représentations qui organise son langage et informe ses intuitions linguistiques, non pas dresser une liste, d'ailleurs impossible à constituer, des phrases d'une langue et de leurs significations possibles.

I. 2. De ce paradigme théorique et méthodologique influent est sorti un grand nombre de découvertes importantes sur le langage, en particulier en syntaxe. Il en est aussi sorti un curieux paradoxe. Je vais brièvement l'exposer ici, sans chercher pour le moment à le résoudre ou à l'expliquer.

- La compréhension biologique et innée du langage conduit à affirmer la priorité du langage oral, « naturel », sur le langage écrit, qui, du point de vue de notre héritage biologique, ne peut évidemment qu'être dérivé. Mais en pratique, la GG s'est détournée des données du langage oral dans leur spécificité. Je n'en donnerai ici que deux exemples, mais ils sont fondamentaux. D'abord, c'est le fait que le point de départ de toute l'analyse grammaticale en GG a été pendant très longtemps, et reste d'ailleurs pour beaucoup maintenant encore, la phrase. Or comme l'on fait remarquer, entre autres, les travaux du GARS pour le français parlé, R. Macaulay () et P. Hopper 1988 pour l'anglais, le langage oral ne s'articule tout simplement pas autour de la notion de phrase. C'est ainsi qu'il n'y a pas de parallèle immédiat dans les travaux écrits dans le cadre de la GG à la distinction que l'équipe du GARS établit entre microsyntaxe et macrosyntaxe. Par ailleurs, il existe toute une série de constructions, du genre *il m'a donné tout ce que j'avais envie* en français, qui apparaissent régulièrement à l'oral, mais que les locuteurs qui les produisent refusent généralement comme agrammaticaux. La réflexion intuitive ne reconnaît donc pas toujours certaines régularités de langue qui n'apparaissent qu'à l'oral.

- La conception psychologique de la compétence, qui se comprend comme la construction du langage par et dans le cerveau/esprit d'un individu, affirme que la seule réalité immédiate est celle de la connaissance linguistique individuelle telle qu'elle se construit de manière indépendante chez chacun, d'une part, et celle de la grammaire universelle (GU), notre héritage biologique commun, qui guide et circonscrit le champ des possibilités du langage individuel de l'autre. Chomsky va d'ailleurs insister sur le caractère dérivé et même illusoire de la notion de langue nationale, telle que le français, l'anglais, qui correspond à une construction politique

et sociale, mais pas linguistique⁴. En pratique, cependant, la grammaire générative s'est de coutume restreinte à l'étude du langage dans sa forme standard, et s'est désintéressée de la question de la variabilité entre individus, alors là même qu'elle en posait la possibilité, sinon même la probabilité⁵.

D'une théorie a priori tournée vers l'oral et l'individuel, est donc sortie une approche concordant plutôt avec le langage standard et en particulier l'écrit. Je voudrais considérer ici la possibilité que le choix des données aient au moins en partie conditionné ce résultat.

II. 1. Le travail sur corpus s'était donc, pour un temps, retrouvé identifié à un point de vue théorique indéfendable. Pourtant, avec le triomphe de l'approche cognitive et de la compréhension interne de la connaissance linguistique, ce n'est évidemment pas tant l'approche par corpus comme telle qui était mise en cause, mais la conception que les chercheurs en avaient à une certaine époque. Ce n'est pas l'idée que le comportement soit une source de données utiles et valables qui s'est trouvée discréditée, mais l'idée de refuser toute détermination interne, c-à-d. mentale, au comportement. De fait, toute une branche de la linguistique s'est reconstituée depuis autour de l'analyse de la parole, utilisant cette fois une approche ouvertement mentaliste. C'est tout d'abord le travail de Dell Hymes, centré autour de la notion de compétence communicative, et ensuite le développement de l'analyse du langage parlé, ou analyse du discours, sous toutes ses formes (approche variationnelle, ethnographie de la communication, analyse de la conversation, sociolinguistique interactionnelle, etc.). L'analyse du discours va bientôt faire ressortir le côté systématique du langage oral et de l'interaction verbale, même si les règles découvertes diffèrent, au moins au niveau descriptif, de celles que

⁴ Voir par exemple N. Chomsky (1986).

⁵ Je ne suis bien sûr pas le seul à avoir remarqué cette anomalie. J. McCawley (1982), p.159 écrivait déjà : « I wish to dissociate myself from an assumption that is so popular among linguists that it is difficult to find anyone who disputes it, namely the assumption that people who talk the same have the same linguistic competence ». Il ajoute à la page suivante « In the rare cases where linguists have looked for interpersonal variation in language, they have generally found it ». Ce dont McCawley parle ici, c'est de la probabilité non seulement de l'existence de différences de grammaire entre les personnes qui parlent différemment, mais de différences de grammaire entre les personnes qui paraissent avoir le même langage. Les données de la performance orale seraient donc de ce fait doublement opaques et ambiguës. Le travail que j'ai fait sur les différences entre locuteurs natifs et non-natifs en français, R. Coppieters (1987), tendrait à supporter cette hypothèse. Dans les théories récentes, l'absence de variations importantes entre locuteurs d'une même langue est postulée par la thèse qu'une langue représente en fait le résultat d'un choix de position (setting) pour un nombre restreint de paramètres, choix qui déterminerait de manière nécessaire la structure de la langue en question.

l'on obtient par l'étude des intuitions, et si les chercheurs ont souvent tendance à faire appel à des notions de nature dite « fonctionnelle » plutôt que purement « formelle »⁶.

On remarque aussi que dans une large mesure les deux approches, GG et analyse du discours, se sont organisées non seulement autour de données différentes, mais aussi autour de niveaux d'analyse différents. De même que la syntaxe générative correspond au travail basé sur des jugements intuitifs passés sur des exemples construits, alors que l'analyse du discours correspond au travail sur corpus, de même syntaxe = phrase, alors que discours = énoncé, généralement défini comme un fragment de langage plus étendu, plus inclusif que la phrase.

Il n'en reste pas moins vrai que les deux approches ne conduisent pas nécessairement à une vision unifiée du langage, et que l'analyse du langage parlé continue à entretenir des rapports incertains, et souvent tendus, avec l'approche de la GG.

II. 2. Avant de considérer la question du rôle de la nature des données dans cette situation, je voudrais d'abord écarter un faux problème. Il est vrai que l'approche par corpus amène souvent à remarquer des exemples qui avaient échappé au chercheur travaillant uniquement sur des exemples construits. C'est là une fonction importante de la recherche par corpus, qui permet de compléter notre information sur le langage. Comme le disent M. Blasco et P. Cappeau 1992, p12, «... l'intuition seule ne peut suffire à faire le tour des faits de langue pertinents lorsqu'on envisage de travailler sur des problèmes descriptifs ». Il n'y a évidemment là rien de très révolutionnaire, et je ne crois pas qu'un générativiste refuserait a priori l'aide apportée, même quand il ne va pas lui-même la chercher. D'ailleurs, on trouve de plus en plus d'articles importants, écrits dans le cadre génératif au sens large, qui s'appuient en tout ou partie sur des exemples tirés de corpus, même si ces corpus sont d'habitude littéraires et non oraux. Je pense à des travaux comme A. Zribi-Hertz 1989 et C. L. Baker 1995 sur l'anaphore en anglais, et les études très fines de N. Ruwet 1994, 1995 sur les verbes de sentiments. De même, il ne faut évidemment pas confondre métonymiquement l'approche générative avec telle ou telle analyse précise offerte dans son cadre méthodologique. Dire que la GG analyse telle ou telle phrase de telle ou telle manière est un non-sens. Le terme GG recouvre d'abord une série de théories en évolution constante, il n'y a donc pas « une » théorie générative, ni « une » analyse générative ; ensuite, à l'intérieur même d'une des théories proposées dans le cadre conceptuel de la GG, toute analyse précise d'une phrase donnée reste toujours sujette à discussion et à révision : au niveau de la description, la question n'est pas

⁶ Voir D. Schiffrin (1994) pour une introduction détaillée aux travaux de langue anglaise sur l'analyse du discours.

celle de la seule analyse, mais de la meilleure analyse dans un cadre théorique donné.

D'ailleurs, à la base, les descriptions syntaxiques développées par le GARS dans le cadre de l'approche pronominale et celles offertes dans le champ de la GG se recoupent très largement. Sur bien des points, et plus particulièrement en microsyntaxe (place accordée au verbe, notions de rection, d'associés, de possibilité de remplacement par un pronom, etc.), les deux approches n'offrent même guère de différences significatives, à part l'insistance sur un certain niveau de formalisation de la description en GG. Les travaux du GARS ont incontestablement mis en lumière des faits nouveaux et significatifs, mais il n'y a aucune incompatibilité de principe entre l'approche générative et l'approche pronominale. Bien au contraire, on pourrait dire que jargon, rhétorique et tradition s'allient ici pour cacher un assez large accord de principe⁷.

II. 3. Il existe cependant une série de domaines où le point de vue du GARS et de l'analyse du langage parlé d'une part, et l'approche générative de l'autre, conduisent à remarquer et à mettre en valeur des phénomènes linguistiques profondément différents : je pense d'abord évidemment à tout ce qui reflète l'interaction linguistique, à tout ce qui est marqué par le côté actif et intentionnel du langage, si brillamment démontré par les philosophes anglais John Austin et Paul Grice, comme par exemple la construction coopérative de la référence (Cl. Blanche-Benveniste 1985) et du sens (H. Clark 1992) dans l'interaction orale, la confusion entre locuteurs (Cl. Loufrani 1985), etc., tous aspects que l'étude du langage parlé mène à découvrir dans la réalité intentionnelle du discours, mais qui restent complètement opaques à l'approche intuitive sur des exemples construits. Je pense par ailleurs à tout ce qui a trait au flot de l'information dans le discours, à la cohérence du discours, aspects qui ont aussi un impact au niveau syntaxique, comme c'est le cas pour la construction du paradigme verbal dans le discours (Cl. Blanche-Benveniste 1988), ou l'analyse des propositions relatives de B. Fox et S. Thompson 1990, dont le travail correspond de manière intéressante à certaines remarques de Cl. Blanche-Benveniste 1995 sur l'emploi de la construction *il y a... qui* en français parlé.

III. On pourrait objecter que les différences entre analyse du langage parlé et jugements intuitifs reflètent tout simplement le fait que le langage parlé est de la performance, du langage mis en œuvre. Les découvertes auxquelles j'ai fait allusion dans le cadre du langage parlé illustreraient ce

⁷ Il en est de même, bien sûr, pour les différentes composantes de la mouvance générative elle-même, dont les différences terminologiques cachent souvent les similarités.

qui arrive quand on passe de la simplicité, de l'unicité, de l'homogénéité de la compétence à la complexité, à l'hétérogénéité de la performance. Notons donc que les intuitions sont également, bien sûr, de la performance, et reflètent elles aussi la mise en œuvre d'une série de processus mentaux. Les intuitions sur des phrases construites de toute pièce pour les besoins de la cause dans la recherche d'exemples et de contre-exemples à une hypothèse donnée, sont dépendantes du contexte dans lequel elles sont produites, même si le contexte spécifique, dit « hors contexte », dans lequel on considère les intuitions est celui de la réflexion sur le langage, et non pas celui de l'action à travers le langage. L'approche intuitive hérite donc elle aussi du contexte où elle se constitue toute une série de caractéristiques.

Le fait de mentionner une forme linguistique quelconque (mot, phrase,...), nous fait toujours construire une représentation mentale de cet énoncé. La réflexion sur un mot quelconque, hors de tout contexte, tel « oiseau », donne naissance à une représentation mentale, celle d'un oiseau relativement précis, qu'on appelle le prototype de l'oiseau et qui ressemble plutôt à un moineau qu'à un pingouin, une autruche, une poule ou un vautour par exemple. De même, la considération d'une phrase hors contexte nous la fait comprendre comme un texte prototypique. C'est comme prototype du mini-discours clos s'offrant à la réflexion dans son unicité, qui se suffirait à lui-même et ne dépendrait que peu ou prou d'un acte d'échange avec autrui que la notion de phrase individuelle semble exister⁸. De plus, le travail du chercheur se fait par écrit, et le texte écrit est lui-même prototypique à la réflexion du locuteur lettré. Le fait que les jugements intuitifs se retrouvent plus proche de l'écrit que de l'oral n'a donc rien de surprenant, puisque les conditions d'emploi du langage sont les mêmes (auditoire indéfini, temps pour la réflexion et la correction, prime au langage normatif à travers l'effort de normalisation et d'effacement des différences entre locuteurs qui marque l'écrit, etc.).

Un phénomène similaire se produit au niveau de notre conceptualisation de la langue en général. Prenons le cas précis du français oral et écrit. Il y a des langues où la différence entre norme écrite et oral familier, par exemple, est telle que l'idée de confondre ces deux extrêmes sous un même système est immédiatement rejetée. C'est le cas de l'arabe littéraire et de l'arabe parlé, par exemple, ou de plus personne ne confondra les différentes formes géographiques que prend la langue parlée. De même, il y a des langues, comme l'anglais pour beaucoup de locuteurs, où les différences entre oral et écrit sont comparativement plus faibles. En français, au contraire, les différences entre système écrit normatif et oral spontané sont

⁸ Il faut remarquer que ce qui peut servir de phrase à l'écrit, c'est-à-dire ce qui est inclus graphiquement entre deux points, diffère d'une culture à l'autre. Ainsi en français on peut écrire « Il est parti comme ça. Sans dire un mot. » Cela ne pourrait pas s'écrire en anglais, où la notion de phrase écrite correspond à celle de « phrase complète ».

souvent importantes, même si, dans la culture, il est d'usage de concevoir et de parler *du* français comme s'il s'agissait d'un système, d'un langage, unique. C'est le langage écrit qui est alors régulièrement identifié au « vrai » français, au langage correct ; l'oral, chaque fois qu'on remarque sa différence de l'écrit, sera compris comme fautif, imparfait, relâché, etc. même là où il n'offre que peu de variations entre classes sociales. Je pense par exemple, en syntaxe, à la négation, à la dislocation, etc. Dans l'esprit du locuteur lettré, le prestige, et la force de l'image visuelle offerte par le langage écrit, s'impose, alors qu'il n'existe aucun système de représentation stable équivalent pour l'oral.

Il n'est pas inutile à ce stade de rappeler que le statut de monolingue est en réalité l'exception et non pas la règle dans le monde, et que le monolingue en fait n'est lui-même pas véritablement « mono » lingue, même s'il a tendance à se concevoir lui-même comme uniforme, régulier, et semblable à lui-même⁹.

Cette réaction uniformisante offre une bonne illustration des effets bien connus de la catégorisation humaine (voir par ex. G. Lakoff 1987), que l'on peut résumer de la façon suivante :

La langue est un catégorie ; les catégories sont conçues comme homogènes, c'est-à-dire que leurs éléments sont supposés partager les mêmes caractéristiques distinctives (même si en réalité il n'y a pas d'ensemble de caractéristiques qui réunissent tous les membres d'une catégorie linguistique et les distinguent des autres catégories voisines) ; toute catégorie est perçue en termes d'un de ses éléments, qu'on appelle son prototype ; la langue écrite est le prototype de la langue pour les locuteurs lettrés ; la catégorie de la langue est donc identifiée à la langue écrite.

⁹ A l'observation, il est effectivement évident que le français n'offre pas un mais plusieurs systèmes différents, selon des lignes de séparation géographiques et sociales aussi bien que suivant l'axe oral/écrit. Un exemple lexical en est les différentes formes adverbiales employées par les locuteurs de différentes régions pour ancrer temporellement l'énoncé, i.e. *maintenant/tout-à-l'heure/tout de suite/à c't'heure/...* ; un autre exemple, syntaxique, est offert par les variations dans le domaine des clitiques (clitique datif « éthique », par ex. « *alors vous voyez l'argent il me s'en va + de ma retraite* », dans le midi ; emploi de *y* pour certains objets directs dans le centre ; emploi de *en* régime d'un nom dans la langue écrite mais pas dans la langue orale (voir CI Blanc-Benveniste 1990), etc). Un exemple phonologique est celui du rapport entre forme masculine et féminine des adjectifs. Dans le langage écrit, le féminin est dérivé du masculin en ajoutant la lettre *e* (petit > petite, grand > grande). Dans le langage parlé, le masculin est dérivé du féminin en laissant tomber la dernière consonne (/p'tit/ > /p'ti/, /gra-d/ > //gra-/). En réalité, cependant, un locuteur, comme l'ont illustré les travaux du GARS, va souvent jouer sur les différences de systèmes pour renvoyer aux différents contextes auxquels ces systèmes sont associés. Pour ce qui est de l'oral et de l'écrit, il n'y a d'ailleurs pas deux systèmes différents, mais un domaine de langage structuré de manière complexe, en termes de paramètres qui varieront de manière contextuelle, d'après le genre littéraire et les circonstances de l'oral.

Si ce raisonnement est correct, il indique qu'il y aurait en fait un rapport assez direct entre intuitions et *literacy*. Une série d'interviews que j'ai conduites il y a quelques années dans la campagne française semble vérifier cette hypothèse. Dans de nombreux domaines, en effet, les locuteurs que je nommerai « oraux », dans la vie desquels le langage écrit (lecture et écriture) n'occupe pas de place ou une place insignifiante, n'ont souvent pas ou peu de jugements de grammaticalité ou d'intuitions précises sur des phrases individuelles, contrairement aux locuteurs « lettrés ».

Puisqu'il s'agit d'un phénomène qui n'a jusqu'ici fait l'objet d'aucune étude détaillée, j'en donnerai quelques exemples.

IV. Au cours du printemps 1989, alors que je m'intéressais à la possibilité de variations d'intuitions entre locuteurs d'origines géographiques différentes, j'ai interviewé un certain nombre de Français de diverses régions, en reprenant au départ des phrases que j'avais déjà utilisées dans un travail précédent sur les différences d'intuitions entre locuteurs natifs et non-natifs en français (voir Coppieters 1987).

Je fus très vite ramené à la raison : les locuteurs que j'avais choisis d'interviewer, des personnes qui avaient vécu à la campagne toute leur vie et n'avaient donc habituellement reçu qu'une instruction scolaire relativement réduite, qui étaient donc relativement moins directement affectés par la norme écrite, ne se laissaient tout simplement pas intégrer dans le cadre d'analyse que j'avais développé lors de mon précédent travail, orienté lui sur des personnes venant principalement de milieu universitaire, et donc comparativement fort lettrées. Ce n'était pas seulement une question de différences d'intuitions, c'était une question de nature d'intuitions.

Le premier exemple que je donnerai ici concerne un couple de retraités à Ploubalay, un village des Côtes d'Armor en pays Gallo. Lui (L2) a passé sa vie active comme pêcheur sur les bancs de Terre-Neuve, pendant qu'elle (L1) tenait la maison, le ménage, et s'occupait des enfants pendant ses longues absences. Ils savent bien sûr tous deux lire et écrire, mais la lecture et l'écriture n'ont guère de place dans leur vie. Au niveau syntaxique, leur langage se présente comme un français oral assez standard sauf pour quelques caractéristiques spécifiques probablement conservées au dialecte gallo, en particulier l'absence du mode subjonctif, remplacé partout par l'indicatif :

1. L2. *fallait bien qu'il part gagner la croûte*
2. L1. *nous on apprenait comme ça mais après il fallait que je leur apprends + alors ils me disaient faut que tu fais trois ponts et deux ponts*

Dans un premier temps, j'ai expliqué que je m'intéressais aux différences de parler dans les différentes régions de France, et que je leur présentais des phrases pour savoir si elles leur paraissaient naturelles ou non, en

insistant sur le fait que ce que je cherchais, c'était si cela se disait dans la région. Il y a bien sûr toujours le danger que les personnes interviewées perçoivent ce genre d'activité comme un test de leurs connaissances de la norme. Cependant, en insistant sur le caractère arbitraire de la norme, en faisant appel à leurs sentiments régionalistes et anti-parisiens, au droit de chacun de parler comme il l'entend sans qu'on lui impose comment, et enfin au fait qu'ils étaient dans la position de m'enseigner quelque chose de leur patrimoine culturel, j'ai d'habitude trouvé des collaborateurs pleins de bonne volonté, sinon même d'un certain enthousiasme.

Le genre d'interaction que l'on trouve sur l'enregistrement de la conversation que j'ai eue à Ploubalay illustre ma naïveté de « lettré » qui n'avait eu jusque alors aucun problème à faire comprendre ce qu'il recherchait. (L3 = moi).

3. L3 j'ai fait beaucoup de fagots + vous employez le mot fagot
 L2 oui c'est bon ça
 L3 j'ai fait beaucoup de fagots mais c'est des petits fagots + ou bien est-ce que vous dites + ils sont des petits fagots + qu'est-ce que vous préférez
 L2 ah + on peut dire petits ou des gros + la grosseur qu'on veut quoi
 L3 oui mais si je dis j'ai fait beaucoup de fagots mais c'est des petits fagots ou
 L2 des moyens des moyens fagots
 L3 ou des moyens fagots + mais on dit c'est des petits fagots
 L2 c'est des petits fagots
 L3 ou bien ils sont des petits fagots
 L1 ils sont des petits fagots
 L3 j'ai fait beaucoup de fagots mais ils sont des petits fagots
 L1 oui
 L2 on peut dire j'ai fait des gros aussi
 L3 il y a une différence pour vous entre + j'ai fait beaucoup de fagots mais c'est des petits fagots et + j'ai fait beaucoup de fagots mais ils sont des petits fagots
 L1 non ça dépend dans le temps on faisait des fagots dans les + dans les + fermes + ils sont petits qu'on disait
 L2 on était à celui qui en faisait le plus quoi + cent fagots c'est déjà beau alors si il faisait des petits + il avait plus que celui qui faisait des moyens quoi

Ce genre de malentendu se répète à travers tout l'enregistrement :

4. L3 j'ai fait cela pour lui gagner
 L2 ah ben peut-être + mais + on pense plutôt à nous [rire]

Parfois, dans la façon dont le locuteur répète la phrase, on peut trouver des indications sur ce qu'il dirait lui-même :

5. L3 j'ai fait cela pour moi gagner + pour moi gagner + non + oui
 L2 gagner au tiercé + ben je joue jamais + mais on pourrait dire ça + oui + j'ai joué pour gagner
 L3 je suis content que j'ai gagné
 L1 ah oui + on est très heureux de gagner

La discussion sur *voilà la chienne que j'ai envie d'un chiot* se termine par cette phrase qui, pour L1, résume tout : *oh + j'aime les bêtes moi*. De même, pour une phrase qui mettait en scène un politicien, L1 s'avoue vaincue et déclare : *la politique on s'intéresse pas du tout + je sais pas vous répondre hein*.

Dans la plupart des cas, ces deux locuteurs réagissaient évidemment au contenu sémantico-pragmatique des phrases que je leur proposais, et non à leur forme.

On pourrait alors croire qu'il n'y a ici qu'un long quiproquo. Pourtant, on trouve aussi quelques exemples où un jugement de nature formelle est exprimé sans hésitation et de manière parfaitement directe. C'est le cas en particulier d'un emploi de l'imparfait dit de rupture (que dans mon expérience, très peu de locuteurs oraux reconnaissent) :

6. L3 en 1885 Victor Hugo mourait + c'est différent de + en 1885 Victor Hugo est mort + pour vous est-ce que c'est différent
 L2 est mort moi je dirais
 L3 en 1885 Victor Hugo mourait + ou + en 1885 Victor Hugo est mort
 L2 est mort plutôt
 L1 oui

Ici, on voit bien que les deux locuteurs n'ont aucun mal à identifier ce qui leur paraît approprié.

Une autre stratégie que l'on rencontre chez les locuteurs oraux est celle où l'on reprend l'énoncé pour le reformuler en entier. C'est souvent le cas pour cette retraitée (L4) de Saint André de Roquelongue, un village des Corbières dans le Languedoc, ouvrière agricole toute sa vie, dont le langage est par ailleurs extrêmement complexe et organisé :

7. L3 j'ai acheté la bouteille que tu avais envie
 L4 on dit + tu m'as commandé une bouteille et je te l'apporte
8. L3 j'ai besoin d'un kilo de farine de seigle + et puis + vous revenez et vous dites + voilà la farine que tu avais besoin d'un kilo
 L4 non je dis je suis allée à l'épicerie + et je te porte la farine que tu m'as commandée + on ne dit pas je te porte la farine + voilà

Dans un certain nombre de cas, elle n'hésite cependant pas à trancher :

9. L3 j'ai fait beaucoup de fagots + mais ils étaient seulement des petits fagots
 ou + c'étaient seulement des petits fagots
 L4 c'étaient + pas ils étaient
10. L3 qui est Victor Hugo + il un écrivain du 19ème siècle ou + c'est un écrivain du 19ème siècle
 L4 non + c'est un écrivain du 19ème siècle

Comme pour les locuteurs précédents, elle refuse l'imparfait de rupture :

*11. L3 en 1845 + Napoléon est mort ou + en 1845 Napoléon mourait
L4 non + Napoléon est mort + est mort pas mourait + c'est pas du français*

et elle refuse également le pronom *en* régime d'un nom (voir Cl. Blanche-Benveniste 1990) :

*12. L3 tu vois cette vigne + j'en connais bien le propriétaire
L4 non je dirais + là de cette vigne + elle est de telle personne ou de tel
propriétaire + mais on ne dit pas j'en connais + j'en connais non non on le
dit pas ça*

D'autres locuteurs oraux enfin hésiteront beaucoup à se prononcer formellement sur l'acceptabilité d'une phrase. C'est le cas d'un autre locuteur (L5) de St André de Roquelongue, également ouvrier agricole toute sa vie, excellent conteur, et dont le fils a fait une brillante carrière dans l'administration d'une grande entreprise. Sa réponse à la question du choix entre *ils étaient/c'étaient seulement des petits fagots* est typique de ses réactions :

*13. L5 je ne vois pas moi c'est là que exactement moi je peux pas définir si la
chose qui était la meilleure
L3 une sonne mieux +
L5 je vous comprends c'est à dire que non c'est un peu différent + oui moi
je vois pas une chose qui tranche bien sur l'une ou sur l'autre + voyez*

Il existe donc toute une série de locuteurs oraux pour lesquels l'approche intuitive donne des résultats beaucoup moins tranchés qu'avec les locuteurs lettrés, et dont le langage est de ce fait relativement opaque à une approche purement intuitive. Au contraire des locuteurs lettrés, beaucoup de locuteurs oraux semblent prêts à accepter des exemples qu'ils n'emploieraient sans doute pas eux-mêmes. Par ailleurs, les locuteurs oraux hésitent souvent beaucoup plus à trancher entre les exemples. Là où il comprend la phrase, un locuteur oral aura souvent tendance à l'accepter. Les intuitions des locuteurs oraux offrent donc une vision moins définie, moins nette de la langue, vision qui correspond bien sûr mieux à l'image qu'un ensemble de corpus oraux assez diversifié donne de la réalité langagière.

Le développement de nos intuitions formelles semble donc accompagner l'apprentissage et l'usage de l'écriture, et par là d'une approche formelle de la langue. Quand on rapproche de ce phénomène le fait que les locuteurs lettrés ont généralement tendance à favoriser les types de phrases qu'ils écrivent, et à refuser les constructions qui ne se retrouvent qu'à l'oral, on

voit l'importance de l'influence du langage écrit sur la forme que prennent nos intuitions.

V. Le fait que les données intuitives viennent avec un bagage significatif ne veut pas dire que les données offertes par un corpus ne soient pas, elles aussi, problématiques. La difficulté fondamentale posée par le corpus reste bien sûr celle de son interprétation.

D'une part, tout corpus donné en termes linguistiques, qu'ils soient phonétiques ou lexicaux, représente déjà une interprétation de la réalité, ce n'est pas un reflet direct de la réalité à l'état « brut », et en particulier ce n'est pas la réalité telle que les locuteurs la comprenaient au fur et à mesure de leur interaction (à supposer que cette notion ait effectivement un sens). C'est toujours de la réalité telle qu'elle a été organisée et comprise par après, presque toujours par d'autres locuteurs. Il est important de se le rappeler. Il est donc impossible de construire des corpus comme le GARS sans avoir recours, au moins implicitement, aux intuitions des transcrip-teurs (l'analyse lexico-morphologique est elle-même un reflet de la compréhension sémantico-pragmatique, voir D.-C. Giovannoni et M.-J. Savelli 1990). Une excellente illustration de ce genre de question est offerte par les nombreuses multi-transcriptions que l'on trouve régulièrement dans les corpus oraux, et dont la question a été abordée avec beaucoup d'intelligence dans les travaux du GARS. L'on peut trouver bizarre que les transcrip-teurs puissent entendre des réalisations aussi différentes que /comment, quand même/ ou /tout, qui est/ pour une même section d'enregistrement. Mais c'est oublier que la transcription est précisément une interprétation, exactement comme l'est notre compréhension du langage d'autrui d'ailleurs, contexte dans laquelle les petits malentendus sont en fait assez fréquents quand on y prête attention.

On trouve des indices de ce même phénomène dans le cas de la personne qui écrit directement son texte, mais dont le découpage lexical et donc l'orthographe, n'est pas standard, ainsi le cas d'une jeune fille de ma connaissance qui orthographiait *quand même* sous la forme « commême ». Il n'y a pas seulement le fait que la personne qui écrit « commême » interprète la chaîne parlée (celle produite par autrui aussi bien que la sienne propre) différemment des personnes qui écrivent « quand même », il y a aussi d'autres mises en rapports lexicaux, avec *comme* et non avec *quand*, etc. On oublie souvent que même dans le cas de la personne qui utilise la graphie standard, cette graphie n'est en réalité pas transparente, que l'on peut justement très bien se cacher derrière elle.

D'autre part l'interprétation d'un corpus demeure elle-même toujours incomplète, et donc en attente d'interprétation. Que voulait dire tel ou tel locuteur ? Sur quelle compréhension de la situation ? A quoi réagissait-il ? Quel système linguistique se trouve ici mis en œuvre ? Dans une certaine

mesure, ce sont tous les problèmes de l'analyse littéraire qui se retrouvent ici. Dans un corpus, on n'est pas toujours sûr de ce que le locuteur voulait dire, ni de comment il comprenait les relations entre les différents éléments de son énoncé (voir par ex. M. Bilger 1989, p. 104-5). On peut essayer de reconstruire, d'interpréter le texte, mais à la limite, il pose le même genre de problèmes interprétatifs que tout autre texte, celui des intentions de l'auteur ou des auteurs, etc. Il y a donc un certain degré d'indécidabilité des données dans un corpus, indécidabilité qui correspond à l'ouverture du langage parlé à la construction coopérative, mais qui contraste avec l'impression de contrôle direct et total présentée par les jugements intuitifs. Dans le cas des exemples construits de toute pièce, les intuitions du linguiste, et aussi celles du lecteur de l'article où se trouvent mentionnés ces exemples, sont contemporaines à la lecture de l'exemple. Puisque le contexte est celui de la réflexion, ils ne présupposent pas de la même façon une reconstruction ou une remémoration du contexte dans lequel le texte a été produit, avec tous les problèmes que cela engendre. Puisque toute la chaîne créatrice s'élabore à l'intérieur de l'esprit du linguiste, il en découle un sentiment de contrôle sur les données mentales, qui gardent un côté incertain dans le cas du corpus.

Par ailleurs, le corpus reste sélectif, il ne peut rendre compte de toute la réalité : nombres de possibilités linguistiques ne se retrouvent jamais dans un corpus, tout simplement parce que le langage humain est trop vaste, et l'on n'est jamais sûr de la raison pour laquelle telle ou telle donnée est absente d'un certain (ensemble de) corpus : cela peut représenter un accident, une impossibilité dans le contexte, dans tout contexte, ou au contraire un choix.

Enfin, les particularités graphiques d'un corpus pourraient faire croire qu'on est proche de capturer l'oral dans toute sa spécificité, ce qui n'est évidemment pas le cas. De la situation réelle à la transcription, on perd tous les phénomènes paralinguistiques, bien sûr, tels que les gestes et le contexte physique, mais aussi le ton de la voix, son volume et son timbre, des éléments phonétiques et prosodiques comme la manière dont les sons sont prononcés, etc. tous aspects qui peuvent jouer un rôle déterminant dans notre compréhension d'un locuteur. On perd aussi et surtout l'intonation, qui elle fait certainement partie de la forme conventionnelle du langage oral, mais pose un problème particulier, dans la mesure où il n'existe pas de système adapté à la rendre facilement lisible à l'écrit.

Un problème précis qui se pose à quiconque veut construire un corpus, comme le GARS, est celui du choix d'une méthode de transcription et d'une graphie, choix qui n'est évidemment pas innocent et qui est conditionné par les buts visés. Dans la mesure où l'on est amené à suivre une orthographe standard pour simplifier la lecture du corpus, comme c'est

le cas pour des corpus que l'on veut étudier syntaxiquement, on implique évidemment un système plus uniforme et plus 'écrit' que ce n'est le cas.

Par ailleurs, il ne faut pas oublier que le corpus crée une nouvelle réalité, sélective. Implicitement, un ensemble de corpus tend lui aussi à prétendre qu'il n'y a qu'un système puisqu'un même système de représentation est employé. De même qu'en travaillant avec les intuitions, les linguistes ont tendance à identifier langage et domaine sur lesquels nous avons des intuitions, ce qui va trop loin, de même il est important de ne pas identifier réalité linguistique avec corpus. A terme, un phénomène de séduction se produit. Toute représentation fortement structurée a tendance à nous éblouir, toute distinction bien charpentée nous attire, et vise à s'imposer à s'identifier au réel. Il faut se rappeler que le corpus est un des outils qu'on se sert pour aider à se représenter cette réalité.

VI. Ces différences de perspective conduisent à une question extrêmement complexe, celle du problème de l'homogénéité de la langue et de la cohérence des données.

La GG conçoit le langage comme un objet défini, préexistant à l'analyse, et interroge l'état de langue atteint par l'adulte, qu'elle comprend comme un état stable et fixe, mis à part de mineures variations lexicales. Ayant choisi la réflexion comme approche au langage, elle définit le langage comme un objet de réflexion. Elle se détourne explicitement de la question de la mise en œuvre et de l'évolution du langage, qu'elle conçoit comme une question de performance. L'analyse du langage parlé choisit au contraire de saisir le langage comme action et intention. Elle mène à reconnaître les aspects moins stables et plus contextuels du langage, les variations (individuelles, géographiques et sociales, mais aussi contextuelles, de genre et de niveau de langue...), la façon dont les locuteurs essayent d'atteindre leur but, bref le rôle fonctionnel du langage, tout ce qui fait que le langage, jusque dans sa forme grammaticale, se crée et se recrée dans le discours (voir P. Hopper et S. Thompson 1980, 1984).

Théoriquement au moins, la GG résout le problème de la cohérence des données en faisant d'une part référence aux intuitions issues de la grammaire individuelle, l'unicité du cerveau/esprit de l'individu étant alors présumée garante de l'unicité du système, d'autre part en posant l'existence d'une grammaire universelle qui sous-tendrait l'ensemble des langages individuels, dont les langages individuels seraient en quelque sorte des incarnations possibles mais pouvant correspondre à des choix d'analyse, sinon même des choix paramétriques, distincts. La réalité de la pratique est cependant plus complexe, puisque les linguistes générativistes cherchent presque toujours à vérifier leurs jugements, à attribuer les indices de grammaticalité à la langue et non pas à l'individu, etc. Il reste en effet la question de la possibilité de la communication et de la construction d'un

espace intersubjectif, à laquelle l'approche centrée sur l'individu n'offre pas de réponse.

Au contraire, l'approche du GARS et de l'analyse du langage parlé tend à mettre en valeur la construction coopérative du langage dans l'interaction verbale, à insister sur le côté social du langage ainsi que sur le caractère non-fermé du système individuel.

La question de la cohérence des données n'en reste pas moins posée. Prenons l'exemple de citations tirées de corpus oraux de provenances assez diverses. Dans quelle mesure peut-on les considérer comme reflétant *un* système linguistique unique, ou au contraire *des* systèmes plus ou moins distincts ? C'est une question qui n'est en fait pas résolue, même si les travaux du GARS font parfois référence à un système unique, mais sans expliquer comment le définir, le comprendre, ou le justifier.

L'ensemble incluant toutes les variétés de langue parlées à l'intérieur d'une communauté sociale, régionale, ou nationale donnée représente à l'évidence un champ de connaissances linguistiques plus ou moins disparate¹⁰. C'est ainsi qu'on expliquera le fait que les différents membres d'une communauté linguistique donnée n'ont pas le même éventail de discours, et qu'ils ont aussi des intuitions opposées sur certaines séries de phrases. N. Dorian 1994 fait d'ailleurs remarquer que même dans une communauté linguistique restreinte et apparemment très homogène, on trouve des variations de langage purement individuelles. Vouloir faire entrer toutes les variations dans un même et unique système linguistique pose nécessairement la question de la réalité de la notion qui sous-tendrait et donnerait cohérence à ce système. Une approche descriptive que l'on pourrait dire géographique se devrait de cataloguer toutes les possibilités découvertes dans une région ou un pays donné (à supposer que l'on puisse vraiment faire cela), mais ne pourrait bien sûr garantir l'unicité linguistique de cet ensemble. La définition saussurienne de la langue comme phénomène social pourrait sembler offrir une telle base, mais il resterait à expliquer comment un phénomène intersubjectif comme le concept saussurien de la langue (dont la métaphore serait plutôt le dictionnaire commun) pourrait rendre compte d'un domaine à la structure interne aussi complexe que la syntaxe. Personne, à ma connaissance, ne l'a fait jusqu'ici.

Par ailleurs, la communauté linguistique est aussi bien évidemment le point de référence de l'individu, qui est exposé en pratique à un large éventail de discours et qui a généralement tendance à adapter, sinon même à moduler contextuellement son comportement sur celui d'autrui. Tout locuteur doit être capable de comprendre des formes de langage qui diffèrent du sien, et de conformer son discours à son auditoire.

¹⁰ Je ne vois pas que le résultat de l'étude du langage sur corpus doive ou puisse nécessairement aboutir à une grammaire unique, un cadre grammatical plus large uniformément applicable à toutes les formes d'une langue.

La question de l'homogénéité et de la cohérence des connaissances linguistiques des locuteurs reste donc posée, et les descriptions construites à partir de corpus donnent une image plus complexe de la réalité que celle donnée par les travaux génératifs. Les études sur corpus sont de nature à mettre en relief l'hétérogénéité contextuelle du langage, la nécessité de reconnaître, sinon d'intégrer, différents niveaux de systèmes : le niveau individuel, celui des individus en interaction directe et constante, le niveau du groupe géographique et/ou social relativement homogène, et celui de la communauté culturelle et linguistique au sens large, mais aussi les différents niveaux oraux et littéraires. Les connaissances linguistiques d'un locuteur ne peuvent se réduire à un seul de ces niveaux.

L'approche intuitive a donc tendance à nous faire percevoir la langue comme un phénomène homogène, alors que, suivant Mikhaïl Bakhtine, l'on peut aussi concevoir l'être humain, et en particulier le langage, comme polyphonique.

VII. Pour résumer, les intuitions tendent à correspondre à la vision suivante du langage :

- langage compris comme pré-créé, cadre linguistique donné, il s'agit uniquement de mettre en œuvre un connaissance préexistante.
- langage existe donc « hors contexte », c'est-à-dire dans le contexte de la réflexion
- langage hors du discours et donc non-intentionnel, inactif
- locuteur monolingue, unitaire,
- monologue
- les mots ont donc un sens a priori
- vision unitaire du langage, le langage est invariable, puisque le langage est supposé unitaire et précréé.

Au contraire, les données recueillies par corpus mènent à la vision suivante :

- langage non-unitaire, fragmenté, différencié d'après le contexte, l'interlocuteur, etc.
- langage en cours de création, et non pas complètement pré-créé.
- langage intentionnel, orienté vers l'action
- création de l'information au fur et à mesure.
- création de la référence dans le discours
- création du sens des mots dans le discours
- locuteur pluriel, dialogue et polyphonie au lieu de monologue

La question reste posée de savoir si, et jusqu'à quel point, l'on parle de la même chose. Le langage est-il un phénomène univoque ou non ? Rien n'est moins sûr. En tout cas, il serait tout à fait inapproprié de vouloir restreindre notre approche à l'une ou l'autre alternative, et il serait utopique à ce stade d'exiger l'harmonisation des résultats des différentes approches. Les énormes progrès faits dans les différents domaines ne pourraient pas avoir eu lieu en l'absence des différences de données et de méthodes, et les discours tranchants et doctrinaires sur ce qu'est le langage ne font qu'illustrer notre tendance à confondre l'horizon de nos préoccupations immédiates et les réponses que nous avons su leur apporter avec le seul horizon, et les seules préoccupations, possibles (ou, de manière plus normative, les seules préoccupations valables).

René COPPIETERS

Bibliographie

- BAKER C.L. 1995, « Contrast, discourse prominence, and intensification, with special reference to locally free reflexives in British English », *Language*, vol. 71, n° 1, 63-101.
- BAKHTINE Mikhail 1977, *Le marxisme et la philosophie du langage*, Paris, Éditions de minuit.
- BILGER Mireille 1989, « Les réalisations en *et tout (ça)* à l'oral », *Recherches sur le Français Parlé*, n° 9, 97-109.
- BLANCHE-BENVENISTE Claire 1985, « La dénomination dans le français parlé », *Recherches sur le Français Parlé*, n° 6, 109-130.
- BLANCHE-BENVENISTE Claire 1988, « La notion de contexte dans l'analyse syntaxique des productions orales : exemples des verbes actifs et passifs », *Recherches sur le Français Parlé*, n° 8, 39-57.
- BLANCHE-BENVENISTE Claire 1990, « Grammaire première et grammaire seconde : l'exemple de *EN* », *Recherches sur le Français Parlé*, n° 10, 51-74.
- BLANCHE-BENVENISTE Claire 1993, « Répétitions de lexique et glissement vers la gauche », *Recherches sur le Français Parlé*, n° 12, 9-34.
- BLANCHE-BENVENISTE Claire 1995, « Le semblable et le dissemblable en syntaxe », *Recherches sur le Français Parlé*, n° 13, 7-33.
- BLASCO Mylène et CAPPEAU Paul 1992, « Sujet postposé et double marquage », *Recherches sur le Français Parlé*, n° 11, 11-30.

- CHOMSKY Noam 1986, *Knowledge of Language : Its Nature, Origin, and Use*, New York, Praeger.
- CLARK Herbert H. 1992, *Arenas of Language Use*, Chicago, The University of Chicago Press.
- COPPIETERS René 1987, « Competence differences between native and non-native speakers », *Language*, vol. 63, n° 3, 544-573.
- DORIAN Nancy 1994, « Varieties of variation in a very small place : social homogeneity, prestige norms, and linguistic variation », *Language*, vol. 70, n° 4, 631-698.
- FOX Barbara et THOMPSON Sandra 1990, « A discourse explanation of the grammar of relative clauses », *Language*, vol. 66, n° 2, 297-316.
- GIOVANONNI Dominique-Catherine et SAVELLI Marie-Josée 1990, « Recueil des données en français contemporain », *Recherches sur le Français Parlé*, n° 10, 19-38.
- HOPPER Paul 1988, « Emergent grammar and the *a priori* grammar postulate », in D. TANNEN (éd.) *Linguistics in Context : Connecting Observations and Understanding*, Norwood, Ablex Press, 117-34.
- HOPPER Paul et THOMPSON Sandra 1980, « Transitivity in grammar and discourse », *Language*, vol. 55, n° 2, 251-299.
- HOPPER Paul et THOMPSON Sandra 1984, « The discourse basis for lexical categories in universal grammar », *Language*, vol. 60, n° 4, 703-752.
- LAKOFF George 1987, *Women, Fire and Dangerous Things : What Categories Reveal About the Mind*, Chicago, The University of Chicago Press.
- LOUFRANI Claude 1985, « Le locuteur collectif, typologie de configurations discursives », *Recherches sur le Français Parlé*, n° 6, 169-193.
- MCCAWLEY James D. 1982, *Thirty Million Theories of Grammar*, Chicago, The University of Chicago Press.
- RUWET Nicolas 1994, « Être ou ne pas être un verbe de sentiment », *Langue française*, 103, 45-55.
- RUWET Nicolas 1995, « Les verbes de sentiment peuvent-ils être agentifs ? », *Langue française*, 105, 28-39.
- SCHIFFRIN Deborah 1994, *Approaches to Discourse*, Oxford, Blackwell.
- ZRIBI-HERTZ Anne 1989, « Anaphor binding and narrative point of view : English reflexive pronouns in sentence and discourse », *Language*, vol. 65, n° 4, 695-727.

LES CALEMBOURS EN TOPONYMIE DANS LE DOMAINE FRANCOPROVENÇAL

Les noms de lieu viennent très souvent du vocabulaire patois. Un nom commun associé à une signification peut cesser d'être utilisé pour désigner tous les lieux conformes à sa signification, pour ne plus désigner qu'un lieu précis : le nom commun devient un nom propre. Ainsi le nom féminin *lanshi* [LA\$Qi] bien connu en Savoie désigne « un endroit herbu étroit, mais étroit dans le sens horizontal et long dans le sens de la pente ». Ce nom commun devient souvent un nom propre, au singulier ou au pluriel : *La Lanche de l'Envers*, *Les Lanches*. Il ne sert alors qu'à désigner un lieu précis.

En Vallée d'Aoste et en Savoie, les ingénieurs des cadastres sardes (celui de 1730-1750 et celui de 1830-1840) ont dû écrire ces mots en français ; en Savoie, après 1860, les topographes de la carte d'État-major (aujourd'hui carte de l'IGN : Institut Géographique National) et les géomètres des différents cadastres ont dû continuer à transcrire en français ces noms de lieux originellement patois. Ces hommes de l'art qui relevaient d'une façon remarquable les élévations, les surfaces, les méandres des ruisseaux ont fait ce qu'ils ont cru bon de faire avec les noms des lieux que les habitants leur prononçaient soit en patois pur, soit en patois quelque peu francisé.

Transférer en bon français des formes patoises de Champagne, de Lorraine ou de Picardie et de toutes les régions du domaine d'oïl n'était pas une entreprise trop ardue. Les transferts interdialectaux fonctionnaient très bien dans tout le domaine d'oïl, comme nous le prouve la transmission des textes médiévaux. La mise en français des toponymes des différentes régions d'oïl n'était que l'un de ces cas de transferts interdialectaux. Mais en domaine francoprovençal, les différences

linguistiques entre patois et français aggravait les problèmes, si bien qu'on a parfois commis des bévues. On s'est même laissé aller à quelque désinvolture inspirée par le mépris de la réalité locale, que ce soit la réalité géographique ou, beaucoup plus souvent, la réalité linguistique. Les fautes sont alors monumentales et pittoresques.

Le mécanisme qui explique ces erreurs est la traduction par calembour, syllabe par syllabe, sans aucun souci de la signification en patois. On conserve ainsi tous les sons du patois et on obtient un mot français qui n'a rien à voir pour le sens avec le toponyme originel. C'est exactement comme si l'on traduisait *De profundis* par « Des profs ont dix ».

Des calembours dus à des attractions paronymiques interviennent naturellement dans l'évolution des langues : la première syllabe du mot *sorcrote* (français du XVIII^e siècle) emprunté à l'alsacien *sûrkrût* (allemand *Sauerkraut*) est devenue *chou-*, parce qu'il y a du chou dans la *sorcrote* ou *choucroute*. Les patois savoyard de la région de Chambéry appellent [lar'mota] « le lézard gris » ; dans cette région, un patois a créé un calembour [l ara'mota] ce qui signifie « laboure-motte ». L'étymologie populaire peut s'appuyer sur un calembour et modifier un mot. Mais, pour ce qui est de l'écriture en français d'un mot patois qui sert de nom de lieu, il s'agit d'autre chose que d'étymologie populaire. Il s'agit d'une opération inspirée par le mépris de la langue locale qui aboutit à trouver un mot français ou latin (un exemple) correspondant aux sons du patois et non au sens du mot. Cela aboutit à des non-sens souvent cocasses.

Premier exemple :

Sur le cadastre de la commune de Macot, en Tarentaise, (on dit aujourd'hui *Macot-La Plagne*) il y a un nom de lieu que les patoisants prononcent [soflA fwA] : *Sofla - Foua*. Il s'agit d'une expression signifiant « Souffle-Feu ». Quel événement justifie cette appellation ? Un petit feu de berger est peut-être devenu un jour incendie ; ou bien une lanterne (même tempête) a été éteinte par le vent en pleine nuit. Peu importe : c'est un endroit très venté. Le géomètre du cadastre a conservé tous les sons des mots patois et a abouti, par calembour, à la traduction parfaitement stupide :

Sofla - Foua
↓
Sauf - La - Foi

Certes Macot se trouve entre *Longefoy* et *Sainte-Foy*, mais tout de même ! Tout de même ! Est-ce que je faisais le clown avec mon calembour sur *De profundis*. Je voulais simplement être pédagogue.

Qu'aurait-il fallu écrire ? Pour une expression patoise qui a un sens très clair et dont les deux mots ont un correspondant direct en français, on aurait pu proposer une vraie traduction *Souffle-Feu*. Si l'on veut conserver tous les sons du patois, pour que le toponyme soit prononcé de la même façon en français et en patois, on aurait pu écrire *Soffla-Foua* ou *Soffla-Foua*. Toutes ces solutions auraient été bonnes et cent fois préférables à la sottise officielle du cadastre. Pourra-t-on la corriger ? Ce n'est pas sûr, car l'écrit est revêtu d'un caractère sacré, dont se sert si souvent la paresse qui ne cherche qu'à recopier, même bêtement.

Deuxième exemple :

Le deuxième exemple témoigne justement de la longévité de la bêtise. Entre Genève et Bellegarde, le Rhône coupe un chaînon du Jura et coule entre deux falaises calcaires. Cet endroit devrait s'appeler le *Pas des Cluses* ou le *Défilé des Cluses* et le vieux fort français qui le domine pour surveiller deux frontières du royaume de France, la frontière avec Genève et celle de la Savoie-Piémont devrait porter le nom de *Fort des Cluses*. En terminologie locale, celle des montagnes calcaires des Préalpes et du Jura, le mot *cluse* désigne un passage plus ou moins large dominé de chaque côté par des rochers abrupts. C'est le cas en cette partie du cours du Rhône. Le fort s'appelle, depuis plusieurs siècles, le *Fort de l'Écluse* ; or il n'y a pas d'écluse et il n'y en a jamais eu. Comment en est-on arrivé là ? Tout simplement grâce à une traduction par calembour.

En francoprovençal, la contraction des propositions *à* et *de* avec les articles définis a été tardive ; elle est, aujourd'hui encore, loin d'être accomplie dans tout le domaine linguistique. Les cartes 1557 et 1559 de l'*Atlas linguistique du Jura et des Alpes du Nord* montrent que de nombreux patois bugistes et quelques patois savoyards disent encore *de lo* au masculin pluriel et *de lé* au féminin pluriel, ainsi que *a lo* et *a lé*. Un régionalisme du français parlé dans quelques villages savoyards a été souvent relevé par ceux qui se moquent des tournures incorrectes de certains provinciaux : c'est l'expression « *jouer à les cartes* ». Tout cela explique que la première désignation du défilé que le Rhône traverse entre des barres rocheuses a été, en patois : *lo pa de lé kluzé* et cela signifie « le pas » ou « le passage des cluses ». Les géomètres et topographes ont gardé toutes les syllabes de l'expression patoise et leur ont donné le seul sens qu'elles puissent avoir en français : « de

l'écluse ». L'absence de toute écluse sur le Rhône à cet endroit et à toute époque n'a pas empêché la sottise, tant est grande la tyrannie de ceux qui veulent servir le français à la lettre, syllabe par syllabe. Si vous dites [d'le klyze], cela ne peut pas signifier « des cluses », car dans ce cas il y aurait la contraction de la préposition et de l'article ; ce que vous prononcez ne peut, en français, signifier qu'une chose : « de l'écluse ». Ainsi a été faite cette écluse imaginaire. Elle ne sera jamais détruite, car le toponyme parle d'elle depuis des siècles, sans qu'elle ait jamais existé. Telle est la façon dont les serviteurs de la langue française ont traité la réalité locale et les langues régionales : ignorance et mépris jusqu'aux conséquences les plus absurdes.

Troisième exemple :

Tout près de ce Fort de l'Écluse, à plus de mille mètres au-dessus, on lit sur toutes les cartes le nom bizarre d'un sommet : *le Crédo*, 1624 mètres d'altitude. Les topographes ont risqué un calembour en latin, plutôt que de comprendre ce que les gens de l'endroit voulaient dire, en désignant cette haute-chaume. *Le Crédo* n'est pas un lieu où souffle l'esprit religieux ; c'est un large sommet arrondi couvert d'une prairie, alors que ses versants sont rocheux ou boisés. Ce sommet en prairie est appelé et décrit par les gens de l'endroit par l'expression *Crêt d'Au* ; le mot *Au* (cf. *S^t Jean d'Aulps* qu'on prononce *d'Au*) signifie « prairie » ; ses rapports avec le mot *Alpe* sont probables et d'ailleurs les scribes médiévaux écrivaient *Sanctus Johannes de Alpibus*. *Le Crêt d'Au* signifie simplement « le sommet en prairie » et n'a rien de religieux, si ce n'est par calembour.

Quatrième exemple :

Puisque nous parlons de hautes-chaumes, voici un autre cas d'un sommet herbu, dont le nom dauphinois a subi un calembour, joliet celui-là. Dans le massif de la Chartreuse, les sommets peuvent s'appeler *Som*, comme le *Grand-Som* qui domine de ses deux mille et quelques mètres le couvent de la Grande Chartreuse. Au sud-ouest du *Grand-Som*, un autre sommet porte le même nom précédé d'un autre adjectif : le *Charmant-Som*. Est-il vraiment charmant ? Ce qui est certain, c'est qu'il est couvert d'une vaste prairie, d'une « charme » en « som », d'une prairie au sommet. Le mot *charme* est une variante francoprovençale du mot qui est *chaume* en français ; il y a le même rapport entre *chaume* et *charme* qu'entre *baume* et *barme*, deux formes fréquentes en toponymie. Les formes avec *r* s'expliquent par un trait de la phonétique francoprovençale qui change souvent en *r* un *l* devant consonne labiale (*p, b, m, v, f*). Des

dérivés du mot *Alpe* peuvent avoir un *r* : *Arpette* désigne un alpage de basse altitude et *Arpon*, un alpage élevé : le *Mont-Charvin* au-dessus de St Jean de Maurienne est un mont chauve. Dans le cas de « Charme-en-Som », *charme* signifie « prairie » et donne une signification claire et exacte de ce sommet. Ce toponyme *Charme-en-Som* s'inscrit parfaitement dans la création des mots par les montagnards : les noms de lieux qu'ils forment sont descriptifs et exacts. Ils ne se soucient guère des effets esthétiques que peuvent avoir sur quelques imaginations les étendues d'herbe au sommet d'une montagne. Mais le *Charmant-Som* aura une longévité assurée : c'est joli et bon pour le tourisme. D'ailleurs *charmant* se trouve dans tous les dictionnaires qui ne connaissent pas une *charme* « prairie ».

Cinquième exemple :

Le mot *araignée* aussi figure au dictionnaire et c'est pour cela qu'un pré aux formes tout à fait régulières s'appelle *Araignée* sur le cadastre de la commune de St Maurice de Rotherens, dans la Savoie proche du Rhône. Ce pré est un terrain sableux et il y a moins de cinquante ans que le propriétaire exploitait là une petite sablonnière. Le patois du lieu a dû connaître le mot *arénier* pour dire « sablonnière ». Les gens ont appelé ce pré, *l'Arénier*. Comment a réagi le géomètre qui a rédigé le cadastre ? En français, les sons qu'on prononce pour dire *arénier* ne font penser qu'au nom de la bête. *Arénier* n'existe pas dans le dictionnaire, même si on peut y lire *arène* ; dans le dictionnaire, il n'y a que le mot *araignée* qui corresponde à ces sons. C'est ainsi que le mot *araignée* a donné son nom à la sablonnière. Encore un parfait calembour. Cet exemple nous permet de poser deux questions :

- Est-ce que le mot *arénier*, dérivé d'*arène*, aurait constitué une offense grave contre la perfection de la langue française ?
- Vaut-il mieux offenser la logique et le bon sens qu'un sentiment de purisme linguistique ?

En obéissant sagement au dictionnaire, le rédacteur du cadastre de St Maurice de Rotherens n'a pas craint d'offenser le bon sens.

Sixième exemple :

La même obéissance au dictionnaire a fait commettre bien d'autres calembours. Les Valdôtains et les Savoyards de Haute Tarentaise n'auraient jamais commis celui de ce sixième exemple, car ils savent ce que signifie l'adjectif *pers, perse*. C'est un adjectif qui signifie « bleu » ; il est commun aux patois de Haute Tarentaise et de la Vallée d'Aoste, de part et d'autre de la crête des Alpes.

Sur le territoire de la commune de Sainte-Foy (en Tarentaise), il y a un petit lac de montagne aux eaux très claires. Comme il est petit, les gens ne lui ont pas donné le nom de lac, mais l'ont appelé, non sans ironie, *gouille*, c'est-à-dire « flaque d'eau ». Ses eaux claires reflètent le ciel. Ce petit lac bleu s'appelle donc tout naturellement : *la Gouille Perse*. Ou plutôt il devrait porter ce nom ; mais un esprit sans doute bien intentionné a chassé les Perses de la carte d'État-major de la France et les a remplacés par un poisson, comme cela est naturel dans un lac. Ainsi il a obtenu *la Gouille Perche*. Pauvre lac bleu !

Septième exemple :

Les patois savoyards ont un mot pour dire « gazon » ou plus exactement « maigre gazon un peu trop sec » ; le mot désigne aussi un « pré aride ». C'est le mot *teppe*, en patois *tèpa* [tE :pa] ou [tEppa] ou [tÈpa]. Le toponyme *les Teppes* est fréquent en Savoie.

La voyelle [É] devant une consonne labiale, [p] ou [m] surtout, a tendance à s'ouvrir en [a], en français, comme dans les patois. Le [a] phonétique du mot *femme* (qui s'écrivait de façon plus sensée en ancien français : *fame*) est un è ouvert qui est devenu *a*, à cause de la consonne labiale qui suivait. Cela arrive aussi dans les toponymes savoyards : le mot *les Teppes* devient *les Tappes* et parfois aussi *l'Etape*, comme cela s'est réalisé aux Contamines-Montjoie. Un calembour de plus.

Huitième exemple :

Parfois le calembour phonétique est si présent et si ridicule qu'il faudrait faire un effort pour l'éviter. A quoi pense le lecteur de la carte IGN (feuille LANSLEBOURG) quand il lit le nom de deux petits alpages superposés : *la Fesse d'en bas*, *la Fesse d'en haut*, au nord de Lanslebourg ? En patois, le mot *fèssi* [fEsi] signifie « bande » : c'est le continuateur direct pour la forme et pour le sens du mot latin FASCIA. Dans les patois mauriennais, ce mot *fèssi* désigne la longue bande dont on emmaillottait les bébés. Un pré étroit et long, mais étroit dans le sens de la pente et long dans le sens horizontal a pris métaphoriquement le nom de « bande », c'est-à-dire en patois de *fèssi*. Ce nom commun a servi de base à des toponymes, qu'il aurait mieux valu écrire *Faisse*. Cette orthographe aurait eu l'avantage d'être étymologique, puisque le mot latin que *Faisse* continue commence aussi par FA- ; de plus et surtout cette orthographe aurait pu orienter l'esprit loin du calembour que l'orthographe adoptée suggère et impose.

Neuvième exemple :

Une commune savoyarde du canton de La Rochette porte le nom bizarre de *Bourget en Huile*. Tout le plateau environnant situé entre 700 et 1000 mètres d'altitude est désigné par le mot pluriel *Les Huiles*. Cette désignation surtout orale pour un ensemble de cinq communes est très courante dans les propos des habitants ; mais, comme cela est naturel pour les noms de petits pays, elle n'a pas été transcrite par la carte de l'IGN. Quoi qu'il en soit, à lui seul le nom *Bourget en Huile* présente une énigme suffisante, car il ne s'agit pas du produit tiré d'un oléagineux, mais d'un avatar du mot savoyard signifiant « aiguille ». Il faut dire pourtant que cette déviation de sens ne découle pas d'un calembour pur et simple, mais qu'elle est le résultat quasi nécessaire de difficultés graphiques. Un exemple semblable à Valloire permet de mieux suivre la métamorphose du mot.

Les Aiguilles d'Arves (3500 mètres) forment trois pointes très caractéristiques. Par l'aiguille méridionale, ce massif touche à la limite départementale entre Hautes Alpes et Savoie. L'ensemble des trois aiguilles sépare la commune de Valloire à l'est, des communes d'Albiez-Montrond et de St Jean d'Arves, à l'ouest. La présence de ces trois aiguilles est si forte dans le paysage que les alpages et les torrents portent le nom des trois montagnes, sous un nom français, *les Aiguilles*, qui en patois s'appellent *léz Ulyé* [yʔe]. Cette francisation est judicieuse, car le seul cas où l'on ait voulu conserver la prononciation du mot patois aboutit à la regrettable confusion avec le mot français *huile*.

Sur la commune de Valloire, la carte de l'IGN (feuille LA GRAVE) donne le nom de quelques alpages situés dans le vallon des Aiguilles, entre 1850 et 1950 mètres d'altitude : quelques chalets groupés sont désignés par *les Aiguilles*, tandis qu'un chalet isolé à 500 mètres au nord s'appelle *aux Huiles*. Le lecteur français ne songe pas au sens « aiguille » ; or ces *Huiles* sont bien le résultat d'un calembour associé aux insurmontables difficultés que rencontre celui qui veut écrire en français le mot qui, en patois de Valloire, signifie « aiguille ».

Ce mot *ulyi* se prononce [yʔi] et fait au pluriel *ulyé* : [yʔe]. Examinons toutes les difficultés qu'ont rencontrées ceux qui ont voulu écrire [lez ·yʔe] en français. Trois difficultés s'enchevêtrent :

1. l'inexistence en français du groupe final de mot [yʔ] ou [yj] ;
2. la complexité de la convention graphique *-ill* pour le son [ʔ] devenu en français [j] entre le XVIII^e siècle et la moitié du XIX^e siècle ; elle a deux lectures : *fil*, mais *ville* ;
3. les difficultés dues, avant le XVIII^e siècle, à la non distinction graphique entre la voyelle [y] et la consonne [v] : consonne et

voyelle s'écrivant à l'initiale par une lettre pointue, alors qu'à l'intérieur de mot, elles s'écrivaient indifféremment par une lettre arrondie. (Laissons le problème du tréma, qui n'intéresse pas l'initiale).

Ainsi le mot patois [n ·yʔi] « une aiguille » aurait dû s'écrire : *n' villi* ; le pluriel [lez ·yʔe] transposé en graphie française du XVII^e siècle aurait été *les villes*. Et comment éviter cette homographie avec l'un des mots les plus fréquents en toponymie ? Il fallait absolument éviter d'écrire *v* pour la voyelle [y] à l'initiale du mot : d'où le *h* et la graphie *les Huilles*. L'*h* initial est une astuce, nécessaire à une époque, pour faire comprendre que la lettre *u* représente une voyelle. L'*h* initial du français *huile* a la même fonction, à l'époque où l'on ne distinguait pas graphiquement *u* et *v*.

Les Huilles, soit. Un patoisant savoyard peut lire [lez ·yʔ]. Mais un Français ignorant le patois et ne connaissant aucun mot dans sa langue qui finisse par les sons [yʔ] ou [yʔ] ou [yj], comme on dirait aujourd'hui ? Les dictionnaires de rimes ne citent aucune série de mots en *-uille* valant [yj]. Et cela s'explique par la phonétique évolutive du français. En effet dans ce cas, devant [ʔ], la diphtongue [yi] serait une diphtongue de coalescence formée d'une voyelle [y] originelle et d'un élément [i] inaccentué et issu du contexte palatalisant, ou une diphtongue conditionnée. Quelle que soit sa nature, cette diphtongue a été d'abord accentuée sur la voyelle la plus ancienne [yi]. Les assonances des textes médiévaux l'attestent. Mais toutes les diphtongues [yi] ont fait basculer l'accent sur le [i]. Cette bascule de l'accent explique la prononciation de mots comme *truite*, *conduire* et aussi *huile* et *aiguille*. Un Français ne peut pas lire la graphie *huile*, sans mettre l'accent sur le [i], alors qu'un patoisant savoyard peut fort bien lire *huile* : [yʔ] ou [yj]. Dans ce cas, le calembour a été le résultat nécessaire de ces difficultés graphiques et de cette incompatibilité phonétique entre langue française et patois savoyard. Les topographes ne sont en rien responsables de ce calembour ; ils ne peuvent être responsables que du passage de [ʔ] à [l], c'est-à-dire de la graphie *huile* à *huile*. L'attraction du mot français a tout au plus favorisé cette simplification ; elle n'est intervenue que très faiblement et à la fin du processus.

Revenons à la commune *Bourget en Huile* : les mêmes graphies, la même incompatibilité et les mêmes lectures ont produit le même résultat. Mais pourquoi donc ce pays s'appelait-il les « aiguilles » ? Est-ce à cause des petites pointes, des toutes petites aiguilles de la chaîne des Hurtières qui ferme le pays à l'est ? Ce n'est pas certain, car vraiment elles ne marquent pas le paysage, comme le font les Aiguilles d'Arves. Mais ce qui est certain, c'est que le pays faisait partie d'un fief d'une famille noble

dont on connaît le nom dès le XI^e siècle : un document nous a transmis le nom d'un seigneur du lieu PETRUS DE ACU¹. « Pourquoi, me direz-vous, cette famille s'appelait-elle DE ACU, "de l'Aiguille" ? » Je vous répondrais volontiers qu'il ne faut pas s'engager dans le problème insoluble qui porte sur l'antériorité de la première poule ou du premier oeuf. Cette « aiguille » savoyarde qui devient nécessairement *huile* nous a suffisamment occupé l'esprit avec son calembour sans coupable.

Une remarque pourtant, qui est utile pour la bonne transcription des toponymes patois dans une langue nationale. Les Mauriennais, près des Aiguilles d'Arves, ont toujours, sauf une fois, traduit en français, ils ont adopté le mot français *aiguille* ; ainsi ils ont trouvé la solution la plus claire, la plus commode, l'échappatoire peut-être. Mais dans le cas de solution impossible, le mot français, si éloigné qu'il soit des sons que le patois prononce pour sa forme authentique et originelle, oui, le mot français est, dans ce cas, bien préférable au non-sens et à l'absurde.

Dixième exemple :

Le mot occitan de Provence *gour* signifie « vallée », « vallée profonde », « trou d'eau », « lit d'un torrent » ; ce mot est là depuis une époque antérieure à l'arrivée des Romains. En Savoie, cette ancienne famille lexicale a donné des féminins, avec le sens de « vallée », « vallée profonde », « lit d'un cours d'eau ». Le meilleur exemple est le nom du village perché au-dessus de la vallée de l'Isère qui est très profonde à cet endroit, entre Villaroger et Tignes : *La Gurra* (prononcez [gyra]). Le phonétisme du radical peut varier quelque peu, dans les différents patois de Savoie : [gur], [gyr], [gØr].

Au bord du lac d'Annecy, sur la commune de Veyrier du Lac, un petit cours d'eau qui dévale dans le lac, à partir du Mont Veyrier a creusé un lit assez bien marqué. Près de ce petit torrent, on observe un toponyme bizarre, belliqueux : *Les Guerres*. Il s'agit du mot patois, [gØra] et [gØre] au pluriel, qu'il aurait mieux valu écrire *Gæures*, même si rien de semblable n'existe dans le petit Larousse. Il est possible qu'on soit tombé dans le piège que pose l'orthographe complexe de *cueillir* ; dans ce cas, recourir à *Gæures* résout la difficulté.

*
* *

¹ Voir Gros, Adolphe, *Dictionnaire étymologique des noms de lieu de la Savoie*, Belley, 1935 ; p. 575, sous *Huiles* (à la lettre U).

On pourrait continuer longtemps, tant la matière est abondante. J'étudierai encore quelques cas, mais je voudrais les citer plus rapidement.

1. *Pierre à Voir* ou *Pierre à Voix*, sur le territoire de Sallanches, près du hameau des Vorziers, continue le patois *Pira Avoua* [pir avwa] = « pierre aiguë ». Le féminin *avoua* continue le latin ACUTA de façon régulière, avec insertion d'un [v] au milieu d'un groupe de voyelles. *Pierre Aiguë* aurait été facile à comprendre pour tous ; *Pierre Avoua* aurait continué la prononciation traditionnelle.

2. Les cols s'appellent parfois *Tracol* ou *Trécol*. À St Gervais, le patois a fermé la première voyelle de *Trécol*, ce qui donne la prononciation *Trico*. Pourquoi écrire *Tricot* ? Pour rendre obligatoire le calembour avec l'objet tricoté ? *Tricol* aurait été préférable.

3. Sur la commune de La Giétaz, une *Croix des Buts* est une *Croix des Byus*, c'est-à-dire des « bouleaux ».

4. A St Foy en Tarentaise, dans un pays où l'église paroissiale se situe à 1050 mètres, on a un *Plan de l'Eglise*, à 2000 mètres. Toujours ce *de lé* pour l'article contracté *des* ; quant à *-glise*, c'est une déformation pour le patois *gluiré*, forme locale de ce qui est dans d'autres parties de la Savoie : *Glières*. Il faudrait le *Plan des Gluires*.

5. Sur la feuille Ugine 3431 de la carte IGN, près de Gotty, une *Touvière* « lieu avec des rochers en tuf » est devenue une *Trouvière*.

6. A Macot, le toponyme *les Colosses* représente le patois [kolyo - orse] « couloir de l'ours ».

7. Sur le versant nord des arêtes de la Meije, entre 3000 et 3400 mètres, l'arête de la Meijette se termine au sud-est par une petite pointe qui devrait s'appeler « la pointe de l'extrémité ». Et c'est ainsi qu'elle s'appelait dans l'esprit de ceux qui lui ont donné le nom patois *lo Sérè dou Tsavon* = « le petit serre du bout ». *Tsavon* qui pourrait se franciser en *chavon* est de la même famille que le verbe *achever*. Ce qui aurait dû être le *Serret du Chavon* est devenu le *Serret du Savon*. Ce n'est pourtant pas le pays des montagnes érodées et arrondies comme un morceau de savon qui a déjà bien servi. Un rocher qui s'appelle *Savon* tout près des arêtes de la Meije ! Quel sens de la géographie et de l'alpinisme !

8. Et encore : le *col du Haut du Four*. Vous n'avez sans doute aucun mal à imaginer pourquoi un four à pain donne son nom à un col. Il s'agit du *col de l'Au Defour* « du pâturage (qui est) en dehors » ou « à l'extrémité ». Les mots issus de FORIS « en dehors » et de l'adjectif FORANUS « qui vient du dehors » servent parfois à désigner les extrémités du territoire agricole d'un hameau ou de la commune. Ainsi dans la vallée du Vénéon, sur la commune de St Christophe en Oisans, *Champhorent* (qu'on aurait mieux fait d'écrire *Champ Foran* ou *Chanforan*) est le nom de la dernière terre cultivée et de la dernière maison avant les alpages.

9. Parfois l'homophonie parfaite entre patois et français se réalise sur des mots qui ne permettent pas la correction. Sur la commune d'Aillon-le-Jeune, dans le massif des Bauges, deux trous dans les rochers portent les noms suivants : *le Trou de l'Agneau* et *la Grotte des Fées*. Le premier n'appelle pas de commentaire, mais le second repose sur un contre-sens. Il ne s'agit pas de « fées », car le patois dirait *foyé* [fo :jE]. S'il dit [fe], c'est qu'il veut dire « la grotte des brebis ». Le nom de cet animal est *fya* au singulier et *fé* au pluriel. D'ailleurs *le Trou de l'Agneau* et *la Grotte des Brebis* constituent un ensemble cohérent dans un monde agropastoral. Tout cela étant assuré, que faut-il écrire sur la carte ? On ne peut tout de même pas écrire *la Grotte des Fés* : tout le monde comprendrait « Fées » et se récrierait devant la faute d'orthographe. Le vrai sens *Grotte des Brebis* est très éloigné pour la prononciation. La solution est difficile.

*
* *

La frontière franco-italienne n'est pas une barrière infranchissable pour ce vice bizarre qu'est le calembour toponymique.

En suivant la route de la Vallée d'Aoste, entre Pré Saint Didier et La Salle, on passe par un lieu qui s'appelait *Mont-Bardon* construit sur le radical *Bard-* qu'on retrouve dans le *Fort de Bard*. Le *Mont-Bardon* a ses lettres de noblesse. Il figure dans une chanson de geste du XII^e siècle, sur une liste des lieux qu'il faut traverser pour aller de Paris à Rome. C'était au Moyen Age. Puis les hommes ont connu la poudre à canon ; ils créèrent l'artillerie et construisirent des forts. La France et l'Italie se chamaillèrent et une forteresse italienne fut construite à Mont-Bardon qui devint *Bombardon*.

Sur la crête des Alpes, entre les communes de Rhêmes Notre Dame et de Val d'Isère, se dresse une belle cime à 3600 mètres d'altitude, la

Tsantéleina. Les cartes françaises de l'IGN ont conservé les lettres qui permettent de retrouver la prononciation du patois [èA\$te·IEjna]. La francisation attendue *Chantèlène* aurait fait penser bêtement à « Chante-Hélène », ce qui aurait été un nouveau calembour. Le mot est difficile à expliquer : le suffixe féminin [-Ejna] s'ajoute à une partie [èA\$teI] qui fait penser à deux autres mots français :

- soit *Cantal* qui est formé sur un radical signifiant « montagne » ;
- soit *chantel / chateau* construit sur *chant* « côté » (*mettre un objet sur le chant*).

Les deux sens conviennent : « montagne » ou « bord d'un territoire ». L'IGN a très bien fait de conserver une allure patoise qui évite le calembour avec *chanter* et avec *sainte*. Hélas ! la carte italienne propose *Truc Sainte Hélène* pour une hauteur du Val de Rhêmes proche de la *Tsantéleina*. Il serait bon que les deux communautés de Savoie et de Vallée d'Aoste s'entendent sur une forme unique, qui essaierait de ne pas faire penser à une quelconque Hélène, chantusc ou sainte.

En un autre point de la frontière franco-italienne, un désaccord subsiste entre la carte française et la carte italienne : c'est entre Modane et Bardonnèche. Mais sur ce toponyme, les cartes ont tort, toutes les deux. Le même col situé à environ 2400 mètres d'altitude entre le col du Fréjus à l'est et le col de Vallée Etroite à l'ouest ne porte pas le même nom pour les Français et pour les Italiens. En fait il a bien le même nom patois, mais ce patois est interprété d'une façon en Italie et d'une autre en France. C'est le *col de la Roue* du côté français et *il colle della Rho* du côté italien. Rappelons que Modane est français depuis cent trente ans et que Bardonnèche a fait partie de la province du Dauphiné et du royaume de France, pendant trois siècles. Cette situation permet d'imaginer que la faute italienne *Rho* est peut-être due à des cartographes dépendant de la province française du Dauphiné et que la faute française *Roue* a été faite par des fonctionnaires savoyards dépendant du roi de Piémont ou de Sardaigne, selon l'époque. De toute façon, de part et d'autre, les fautes ont été faites par des gens qui ignoraient le patois occitan de Bardonnèche et le patois francoprovençal de Modane.

Ces patois possèdent aujourd'hui encore un adjectif qui signifie « en pente raide ». C'est l'adjectif *aro / arote* [aro] / [a-rot'] en occitan de la région de Bardonnèche (les formes sont de Salbertrand) et *arou/arouta* en patois de la région de Modane (les formes sont du Bourget). Ces mots viennent du latin RUPTUS. Les formes les plus anciennes qu'on ait trouvées dans les pièces d'archives commencent par *R* (Oulx, 1189 et

1223)² (Modane 1459). Dans cette région, les mots commençant par *R* ont tendance à se doter d'un *a*- prothétique. Une montagne (« alpage ») de forme ronde s'appelle aujourd'hui *Arrondaz* sur la commune de Modane, le verbe « raccommode » se dit *arminda*. Quand cet *a* prothétique est apparu dans les formes, tous les scribes en ont fait la voyelle de l'article *la* : *de la Rouz* (Modane, 1622) ; *de la Roz* (XVII^e siècle, Turin) ; *de la Rou* (carte Borgonio, XVII^e siècle). Personne n'ose écrire *la Roue*, car tous savaient à cette époque que le mot « rouc » se disait *roua* [rwa] à Modane comme à Bardonnèche. Mais il ne fallait pas écrire un féminin *la*. Aujourd'hui on a une forme bizarre sur la carte italienne : *la Rho* qui désigne un alpage de Bardonnèche et le col lui-même. La carte de France écrit *la Roue*, le calembour et l'erreur.

La solution est simple : il suffit d'obéir au patois de Bardonnèche et écrire la forme masculine *l'Arò* pour le col et pour l'alpage et du côté français, il suffit d'obéir au patois mauriennais et écrire *l'Arou*. Les deux formes sont masculines, signifient la même chose « en pente raide » et se ressemblent suffisamment pour être reconnues par les lecteurs des deux cartes. Après avoir commis des erreurs différentes de part et d'autre, on pourrait se mettre d'accord sur la vérité : elle rapproche.

*
* *

Toutes ces analyses critiques ne visent pas à corriger toutes les erreurs des cartes. Il y a des erreurs établies, institutionnalisées, intouchables. Qu'on n'y touche pas ! Que l'Écluse sans écluse, que le *Crédo* sans foi continuent à dire bien fort et bien haut que l'autorité centrale a toujours fait peu de cas de ce que pensent les petites gens et qu'elle s'est toujours moquée de leurs patois.

Même si elles n'ont aucune conséquence sur les cartes nationales, ces analyses critiques peuvent être utiles. Il suffirait que ceux qui aiment leur petite patrie fassent un relevé complet des bizarreries dont l'autorité centrale a affublé les noms de lieux de chez eux ; qu'un fichier soit ainsi constitué et qu'il soit conservé dans des bibliothèques du département et d'abord aux archives. Chaque fiche de ce sottisier devrait donner sur chaque erreur : sa localisation précise, la réalité lexicale en patois, l'explication de cette désignation vraie, l'analyse de la faute et une

² Ouvrage déjà cité d'Adolphe Gros : p. 491

proposition pour une éventuelle correction. Ce fichier pourrait ainsi être consultable par ceux qui veulent comprendre le sens des noms de lieux et non par ceux – et ils sont nombreux – à qui il est indifférent qu'une sablonnière soit appelée araignée.

G. TUAILLON³

³ Centre de dialectologie - Université Stendhal - Domaine universitaire - 1180 Av. Centrale - B.P. 25 Grenoble cedex 9.

TRANSCRIPTION DE L'ORAL ET INTERPRÉTATION. ILLUSTRATION DE QUELQUES DIFFICULTÉS

Les travaux du GARS consacrés à l'établissement de corpus de français parlé¹ ont montré à quel point la transcription est intimement liée à l'activité d'interprétation. Il est bien connu que le linguiste ne recueille pas les sources orales à la manière dont un magnétophone enregistrerait des séquences sonores, mais est constamment amené à faire certaines hypothèses sur ce qui a « vraiment » été dit par le locuteur dont on souhaite consigner les paroles.

Ce sont les difficultés d'interprétation inhérentes à l'activité d'écoute ainsi que les conséquences que celles-ci peuvent avoir sur la pratique de la transcription qui seront l'objet de cet article.

Tout d'abord, nous rappellerons que très souvent la compréhension de certaines séquences sonores n'est pas aussi univoque qu'on pourrait le penser, et peut amener le transcripateur à hésiter entre plusieurs interprétations concurrentes : un inventaire et une analyse des phénomènes les plus fréquents de « multi-écoute » seront proposés.

En second lieu, nous illustrerons la manière dont certains phénomènes prosodiques peuvent orienter le transcripateur vers telle ou telle interprétation, et nous essaierons de montrer que l'indication de ces phénomènes dans la transcription pourrait être une solution utile.

¹ L'ouvrage de référence reste sur ces questions celui de Blanche-Benveniste et Jeanjean (1987, *Le français parlé. Edition et transcription*, CNRS).

1- CONCURRENCES ENTRE TRANSCRIPTIONS

Plusieurs travaux du GARS², ont montré que les perceptions auditives concurrentes, ou encore les alternances orthographiques, pour une « seule » séquence sonore, mettent en cause plusieurs phénomènes ; les inventaires qu'on peut en faire révèlent que divers éléments de la langue sont concernés.

Ces phénomènes d'écoute et de transcription ont en commun qu'ils rendent indispensable la présence de multi-transcriptions dans le texte. Si, comme le soulignent ces auteurs (p.102), les équipes travaillant sur des corpus oraux ont le plus souvent décidé de conserver dans la transcription les « multi-écoutes », certaines d'entre elles, ne les mentionnent pas ou seulement partiellement (Mouchon et Fillol, 1980 ; Thibault et Vincent, 1990 ; Vion, 1990, par exemple). Et généralement, elles mettent l'accent exclusivement sur les « impuretés acoustiques » sources de ces multi-écoutes.

Pourtant, comme le développent largement Blanche-Benveniste et Jeanjean (1987, p 102 à 110), l'écoute semble être avant tout une activité interprétative qui provoque plus d'erreurs perceptives que ne le font les difficultés d'ordre purement perceptif. L'oreille n'est pas un traître, elle est surtout asservie à la recherche de signification. Ce fait a pour conséquence qu'il est vain de prétendre déterminer avec certitude quand la source auditive est seule responsable de l'hésitation perceptive du transcripateur : « *Phonéticiens et psychologues s'accordent à dire que la perception est un processus actif qui s'appuie sur la compréhension des énoncés* ». (Blanche-Benveniste et Jeanjean, 1987, p. 103).

Cette nécessité de multi-transcription provient de deux sortes de difficultés qu'il peut paraître arbitraire et abusif de distinguer parce qu'elles se conjuguent toujours. L'expérience de transcription montre en, effet, que l'activité d'interprétation de l'énoncé oral enregistré constitue l'essentiel de l'activité de transcription et remodèle, à l'occasion, les données auditives présentes ; mieux, l'activité perceptive est une activité interprétative « accrochée » à la perception des sonorités d'une façon telle que, parfois, si le contexte s'y prête, des sonorités ne sont entendues que si l'interprétation change (Giovannoni et Savelli, 1990).

Il reste qu'il nous paraît judicieux, pour l'exposé, de distinguer et classer les problèmes rencontrés et les différents processus à l'œuvre dans l'activité de transcription même si, en dernier ressort, ils sont toujours étroitement imbriqués.

2 Blanche-Benveniste et Jeanjean, 1987; Blanche-Benveniste, 1990 et 1993.

1.1. PROBLÈMES TOUCHANT À LA QUALITÉ SONORE.

Certaines difficultés lors de la transcription tiennent à des défauts techniques acoustiques que ce soit dû à la source orale, à l'enregistrement de cette source orale ou aux qualités auditives du transcrip-teur. Tout cela constitue des pièges pour la perception, pièges que les améliorations des techniques d'enregistrement contournent certainement mais ne peuvent éliminer totalement. Elles conduisent à obtenir des multi-transcriptions voire même des passages **inaudibles** (qui, par convention, sont notés **X** ou **XXX**, selon qu'ils semblent comporter une syllabe ou plus) :

il y a eu un grand un gros scandale de : de transfert de : d'ordures ménagères d'Allemagne en France /qu'il a, XXX/ des ordures ménagères qui étaient vendues (Ordures Ménagères, I, 104-106)

Cela peut être dû à plusieurs sortes de facteurs dont le point commun est qu'ils provoquent une hésitation perceptive qui se soldera bien sûr par une possibilité d'hésitation interprétative et donc de multi-transcription. Il reste qu'il ne s'agira pas d'une hésitation interprétative seule ; elle s'accompagne d'une alternance auditive. Les facteurs responsables de ces hésitations « auditives » sont multiples : bruits environnants lors de l'enregistrement, chevauchements de paroles, caractéristiques articulatoires des locuteurs (parlers régionaux, énoncés d'enfants, paramètres individuels, etc.), caractéristiques auditives du transcrip-teur. En général, les guides de transcription insistent sur ces obstacles potentiels de nature très technique et font des recommandations diverses pour les éviter ou les compenser.

1.2. PROBLÈMES DE SIGNIFICATION.

La deuxième source de difficultés tient à ce que la transcription d'un énoncé oral est toujours le résultat simultané d'une activité de signification arrimée sur une activité de perception. En effet, un énoncé même correctement perçu peut rester **incompréhensible** (donc ne pouvoir recevoir qu'une transcription phonétique) ou être transcrit de façon erronée. C'est souvent ce qui se passe avec les noms propres ou les termes, techniques ou non, qui peuvent être inconnus du transcrip-teur. Dans ce cas, la perception de la séquence sonore n'est pas en défaut, seule la construction de sens l'est ; l'élément est audible mais ne parvient pas à être interprété ; cela a été le cas pour la séquence suivante :

toutes ces espèces appartiennent à la même grande famille qui est elle-même divisée en deux familles disons les XXX et les XXX (Bon, 13,4-5)

Elle est restée sans proposition jusqu'au moment où un spécialiste de préhistoire a confirmé que ces séquences sonores notées par XXX, qui avaient pourtant été bien retranscrites phonétiquement par le transcripteur, correspondaient aux lexèmes suivants : « ornitischiens » et « saurischiens ».

Cet élément audible peut également donner lieu à une transcription erronée ; cela a été le cas pour :

*Monsieur Untel était venu monter Verenis (au lieu de : Bérénice) au Théâtre français (Métiers)
on utilise alors un lingelair (au lieu de : linge vert) (Métiers)*

On notera au passage que, du moins dans ces deux exemples, le transcripteur a toutefois identifié la catégorie sémantique à laquelle appartenait le terme sur lequel il a hésité : il a reconnu un nom propre dans le premier cas et un terme technique dans le deuxième cas³.

L'élément peut donc être audible mais ne pas parvenir à être interprété correctement. Ce qui est alors reconnu est sémantiquement probable mais ne correspond plus à certaines sonorités entendues ce dont le transcripteur ne semble pourtant pas s'apercevoir. Ainsi le sujet dans son activité interprétative modifie à son insu le signifiant. Le signal acoustique enregistré, lui, reste unique ; il est produit et reproduit, de façon invariable, grâce à l'appareil enregistreur tant que les conditions techniques de bon fonctionnement sont réunies⁴. Mais, l'oreille humaine n'est pas un magnétophone, et le traitement humain du signal acoustique n'est pas homologue à celui d'un enregistreur. Le fonctionnement de cet appareil n'en offre qu'une analogie. C'est la raison pour laquelle l'interprétation d'un énoncé oral peut avoir pour conséquence (et non pour cause : la transcription en faisant foi) une perception erronée.

Dans certains cas, comme celui de difficultés articulatoires (corpus d'aphasiques (Loufrani, 1990), d'enfants très jeunes ou déficients (Pasquale,

3 Ces observations corroborent celle qui a été faite par Blanche-Benveniste et Jeanjean (1987, p.106) lors de la transcription d'un enregistrement de bonne qualité: plusieurs solutions étaient proposées (*soumettre, souligner, signer*) également possibles par rapport au contexte et pourtant différentes phonétiquement "On comprend qu'on puisse parler "d'hallucinations auditives". Tout se passe comme si l'essentiel était qu'il y ait là un verbe à l'infinitif et que cela commence par "s".".

4 On sait, par ailleurs, que "l'écoute enregistrée diffère toujours de l'écoute directe" (Blanche-Benveniste, 1987); l'énoncé oral enregistré, même dans des conditions techniques "propres", n'est pas reproduit fidèlement: il y manque une hiérarchisation des sons qui peut rendre bruyant et inaudible un entretien que, pourtant, l'enquêteur avait entendu directement confortablement.

1989), les transcriptions orthographiques s'arrachent ainsi lentement et laborieusement de l'étape de transcription phonétique.

En dehors de ces cas, on constate des hésitations dues à des divergences d'écoute ou d'interprétation pour une même séquence sonore qui donnent lieu à des solutions concurrentes dans la transcription. Ces solutions multiples seront notées entre barres obliques : /x, y/, ou entre parenthèses, quand elles renvoient à certaines alternances d'orthographe grammaticale.

Dans un premier temps, nous proposerons un inventaire d'exemples types de multi-écoute qui se limitera à donner la liste des phénomènes d'ordre phonétique que nous avons pu observer ; dans un deuxième temps, nous rappellerons les contextes, mais aussi les secteurs grammaticaux, qui semblent être plus particulièrement touchés par ce phénomène de multi-écoute ; pour terminer, nous mentionnerons certains cas d'alternances orthographiques.

1.3. QUELQUES OBSERVATIONS D'ORDRE PHONÉTIQUE DU PHÉNOMÈNE DE LA MULTI-ÉCOUTE.

1.3.1. Certaines écoutes divergentes sont, semble-t-il, dues à la perception ou non d'une séquence sonore.

- Cela peut toucher **une voyelle** ; c'est ce que nous pouvons relever dans les exemples suivants :

et ça aide/â, Ø/ un certain nombre de manœuvres (94- Brel)
on a /eu, Ø/ une recrudescence énorme de brûlures (94- Brel.)
il y a des moments /où, Ø/ il vaut mieux être ho- ho- honnête (94- Brel)
et oui /et, Ø/ cette dame que vous venez d'entendre elle s'appelle Madame XXX (94- Brel.)
j'y étais allé /et, Ø/ je rentrais (Volattles, 1,8)

- dans d'autres cas, la divergence d'écoute porte sur la présence ou l'absence d'**une consonne** :

tu /entres, rentres/ dans la catégorie des découverts non autorisés (Lieutaud, 7,2)
vous les / recevez, recevrez/ les uns après les autres (Laubreton, 4,10)
donc tu /as, vas/ trouvé, trouver/ un défaut qu'il n'a jamais eu (Martin, 9,12)
Olivier parlant /Ø, l'/ espagnol (Anguenot, 2,6)
le Petit prince /s', Ø/ est lavé (Maret,7,9)
tu as des ce qu'on appelle des commandes en fait ce sont des freins - une commande dans chaque main qui /agissent, agit/ sur le bord de la chute (Dutertre, 5,4)
on a fait ça /en, dans/ quoi (94-Brul.)

c'est pour ça /que, Ø/ je + pour moi c'est + l'amitié ça fait partie aussi des événements quoi (Lic 95, corpus Amitié, 2, 24)

- ou encore, sur la perception ou non d'une séquence comportant plus d'un phonème :

*ils sont /là, Ø/ dans des cages (Baumettes, 21,16)
ils ils /vous, Ø/ vident tout (idem, 38,4)
en disant que /nous, Ø/ tout ce que nous demandons (idem,60,4)*

Certaines de ces divergences d'écoute mettent en jeu la perception ou non d'une répétition d'une même séquence :

*c'est /c'est, Ø/ une langue (94-Cux)
tous les détenus que j'ai rencontrés ont une /une, Ø/ dentition épouvantable (Baumettes, 10)
le racisme est /est, Ø/ fort (idem, 5,16)
le particulier euh sera obligé d'avoir plusieurs poubelles dans sa maison /au moins, Ø//au moins dans sa cuisine en tout cas (Ordures ménagères, 1, 176-178)*

1.3.2. Les écoutes concurrentes peuvent également être liées à un problème d'alternance auditive.

- Cette alternance auditive peut se situer entre deux voyelles, comme dans les exemples suivants :

*il y a /les, la/ sœur (s) de de des les sœurs de la Petite Sirène (Garrigues, P, 1, 18)
en fonction /du, des/ poids des Euclides (...)
je ne suis même pas certain que ça leur /a, ait/ fait plaisir (Delp. 21, 11)*

A noter que la confusion entre les voyelles /e/, /É/ et /ø/, /œ/ est fréquente, comme s'il y avait une neutralisation de timbre entre ces deux séries de voyelles⁵ :

*mais je / veux, vais/ répondre à Jean (Navale, 3é, 3)
je parle / des, de/ Chakras (Lachaise, 8, 6)
non ce sont /deux, des :/ choses complètement différentes (Ben, 2, 9)
on /commençait, commence/ le travail (Volatiles, 1,2)
j'étais /celle, seule/ qui lui /posait, posais/ des questions (Chati, 15, 13)
alors moi /j'ai, je/ /dit, dis/ c'est sûr (Kling, 6,14)
moi je /te, t'ai/ /dis, dit/ une chose (Poux, 15,11)*

de même que la confusion entre voyelle orale et nasale :

il faut que tu /attendes, entendes/ sinon je vais la repiquer (Baumettes, 25,4)

⁵ Cette neutralisation pourrait rappeler celle notée entre /œ/ et /o/ dans la prononciation parisienne.

je trouve que la le le /mode, monde/ que la France a choisi (Delp. 11,11)

- certaines multi-écoutes concernent **une alternance auditive entre deux consonnes** :

*je sais pas si tu as déjà vu /les, des/ parachutes actuels (Dutertre, 10, 15)
je /le, me/ résume (Baumettes, 67,2)
les ateliers donnaient sur /l'autre, notre/ rue (Feldman, 16,15)
alors mon grand-père lui il /s'était, s'est fait/ /engagé, engager/ dans les Forces
Françaises (Cati, 9,6)
la stratégie à /des fins, desseins/ militaire(s) (Antonini)
c'était l'époque où on envoyait de très grandes tournées de théâtre que les pays
étrangers étaient censés /de cueillir, recueillir/(Ben, 3, 12)
on /m', n'/ a pas précisé (Anguot, 1, 6)
les conditions de /d', l'/ hygiène (Volatiles, 1, 7)*

- ou **une alternance auditive entre deux séquences autres** ; ce type d'alternance peut sembler *a priori* plus surprenant, on en rencontre cependant de nombreux exemples dans les transcriptions :

*ça /vous, te/ dérange la pollution (94- Brel)
disons /je veux, faut/ pas mélanger les deux (Ben, 2,11)
eh bien dans l'État /il est, il y a/ effectivement des gens dont la fonction est d'être
médiateur (Ben, 13,8)
à l'heure actuelle vous /voyez, vivez/ ça de l'autre côté de l'Est (Delp. 17,8)
on y a amené le second /groupe, couple/ de français (Anguot, 14,15)
le décollage s'est très bien passé / parce que, puisque/ tu décolles pratiquement
sur place (Anguot, 13,12)⁶*

1. 4. LES SECTEURS TOUCHÉS PAR CES PHÉNOMÈNES DE MULTI-ÉCOUTE.

1. 4. 1. Le contexte de la **liaison** semble souvent provoquer des écoutes divergentes ; il peut être difficile, en effet, de percevoir les limites de mot dans ce type de contexte :

*à l'origine /c'était, c'est/ un chien de berger (...)
il /était, est/ intolérable qu'on puisse le condamner à mort (Peil, 4,13)
il faut poser le problème à /un autre, notre/ niveau (Brun, CFI)
on /est, était/ allé dans une cabine (Lic94-SA)*

⁶ A noter que dans une prononciation non normative ces deux séquences "parce que" et "puisque" sont moins différentes que ne le laisse croire l'orthographe : /pask/ et /pyisk/ ou même /pisk/.

Dans ces exemples, les multi-écoutes semblent provenir du fait que l'on peut entendre :

- soit une suite « fondue » de deux voyelles avec absence effective du phénomène de liaison :
 - *c'était un chien*
 - *il était intolérable*
 - *on était allé*
- soit une simple liaison (matérialisée par le signe =) :
 - *c'est = un chien*
 - *il est = intolérable*
 - *on est = allé*

Les limites de mots sont donc bien différentes : la forme verbale « être » a été perçue soit sous la forme « **était** », soit sous la forme « **est** ».

Dans l'exemple suivant : à /*un autre, notre/ niveau.*, le problème est identique, il a été perçu :

- soit une suite « fondue » de deux voyelles avec un phénomène de liaison :
 - *le problème à un = autre niveau*
- soit une absence de forme :
 - *le problème à Ø notre niveau*

Dans ce cas aussi, les limites de mots sont différentes.

1. 4. 2. Comme cela a déjà été mentionné (Blanche-Benveniste et Jeanjean, 1987), les éléments de la langue sont touchés de manière très différenciée par ces phénomènes de multi-écoute ; les divergences d'écoute portent régulièrement sur les **éléments clitiques** de la langue, et en particulier sur :

- les déterminants :

les, des/ portes de four ouvertes (LSL.94)
il y a les, la/ sœur(s) (Garrigues, P, 18)
toujours /la, les/ même(s) question(s) (Bon, 4, 14)
ah oui les, tes/ parents ils ils critiquent (Bon, 5, 9)

- les pronoms :

/ça, ce/ sera consacré au sourds (LSL.94)
alors je /vous, Ø/ rappelle la question (LSL.94)
il /nous, vous/ faut des années (Tri. 1, 2)

Elles portent également sur les **conjonctions** ; on note en particulier la confusion fréquente entre « parce que » et « puisque » :

oui /alors, parce que/ ce qui est curieux (LSL.94)
c'est une façon un petit peu de garder le lien quand même avec euh l'extérieur
/parce que, puisque/ ils ont quand même des difficultés au sein de leur(s)
famille(s) (Lic 94)

ou encore sur **la morphologie verbale** ; c'est le cas pour les désinences :

vous les / recevez, recevrez/ les uns après les autres (Laubreton, 4, 10)
nous /demandons, demandions/ (Baumettes, 60,7)
(.) une commande dans chaque main qui /agissent, agit/ sur le bord de la chute (Dutertre, 5,4)
oui tout d'abord je /voulais, voudrais/ te dire que le Cameroun (Lic 95, Cameroun, 31, 2)

ainsi que pour **l'auxiliaire** :

elle /a, Ø/ /réussi, réussit/ son concours (Bus, D1,3)

A ce sujet, il est intéressant de noter que certains contextes grammaticaux, conjugués à des phénomènes d'ordre phonologique, semblent favoriser ces écoutes divergentes sur la présence ou non de la forme auxiliée avec « avoir » :

alors moi /je, j'ai/ /dis, dit/ c'est sûr (Kling, 6,14)

une prononciation méridionale, dans laquelle l'opposition entre /e/ et /É/ en syllabe ouverte n'existe pas, amplifie par ailleurs ce type de phénomène :

ça /a, Ø/ /donné, donnait/ lieu à une théorie qui est celle de la dégénérescence (Caill. 5,1)
/c'était, ça a été/ très joli (Angunot, 1,6)
/c'était, ça a été/ prouvé (Baumettes, 61,8)
je crois vous l'avoir dit que /j'ai été, j'étais/ inspecteur de l'Education Nationale (Feldman, 1,1)

1. 4. 3. Les éléments lexicaux sont moins fréquemment touchés par le phénomène de la multi-écoute ; ils n'en sont pas, cependant, totalement exclus, comme le prouvent les exemples suivants :

on était un million cinq cent mille /fois, francs/ moins cher (Pozzana, 5, 10)
mais ça peut être aussi à usage local euh euh / par frottement, parfaitement/ (Brossard, 84, 14)
ils ont essayé de /voir, vouloir/(LSL.94)
l'opinion publique pourra se se /sentir visée, sensibiliser/ (Bollet)
pour ceux qui ont /eu, vu/ des sourds (LSL 94)

Les écoutes concurrentes pour une même séquence sonore sont plus fréquentes que ce que l'on pourrait croire ; les solutions multiples semblent être choisies en fonction de ce qui est prédictible ou plausible dans le contexte ; dans ce sens, les découpages en morphèmes se font en

conformité avec le sens local ou global des unités discursives, et ce, à l'intérieur d'un faisceau de probabilités.

1. 5. LES FORMES HOMONYMES⁷

Dans le travail de transcription, on rencontre en définitive peu d'hésitation entre les formes homonymes, tant il est vrai que, là aussi, généralement le contexte sémantique et/ou grammatical, de même que l'intonation, permet d'éliminer ce genre d'ambiguïté ; certaines hésitations subsistent cependant, et semblent toucher de manière différenciée certaines formes.

1. 5. 1. « là, la » : nous relevons des cas d'hésitation en ce qui concerne l'interprétation de la forme phonique /la/; dans les exemples suivants, celle-ci s'interprète soit comme un « là » locatif, soit comme le prédéterminant défini féminin qui aurait été répété :

*et regardez bien la Pointe du Raz + on voit + on voit /là, la/ la vierge (Barian, 14, 1)
donc + tu vois un peu /là, la/ déjà la notion de + patrie (Alger, 3, 13)*

Dans certains cas, le prédéterminant semble introduire un syntagme qui reste inachevé :

*alors j'ai le souvenir de ça des bras- des brasseries et /là, la/ euh le souvenir des voitures à chevaux (Dabat, 11, 10)
la mer / elle, 0 / remue vous l'entendez mais /là, la / mais le mais le temps le temps est calme vous auriez dû venir un jour de tempête (Barian, 6, 3)
et pour /là, la/ + les maladies enfin pour les les docteurs ça se passait comment (Barbara, 8, 16)*

Cette forme phonique /la/ peut également recevoir une triple interprétation ; celles que nous venons de signaler ainsi que celle du pronom complément postposé, comme dans :

regardez écoutez /- la, là, la / regardez regardez la vague (Barian, 12, 6)

1. 5. 2. « ces, ses » : il est parfois difficile de donner une seule interprétation à cette suite phonique ; même en contexte large, les cas d'hésitation ne sont pas rares :

⁷ L'inventaire, présenté ici, est constitué des seuls cas d'alternances orthographiques que nous avons pu relever au cours de la lecture d'environ une dizaine d'heures d'enregistrements retranscrits; dans ce sens, il ne peut prétendre à l'exhaustivité.

*il fallait qu'on mette en relief /ces, ses/ contrastes par l'image (Chaig., 9,4)
 on l'a filmé on l'a filmé /de, deux/ jour(s) et en fait il était en train de préparer
 /ces, ses/ images (Chaig., 26, 5)
 il y avait des gens qui parlaient pour la France + et qui allaient dans les colonies
 françaises + pour bah + développer /ces, ses/ colonies (Alger, 5,4)
 quand l'Europe et ses colonies + même en pro- profitant des richesses de /ces,
 ses/ colonies + euh sortent de la guerre (Alger, 13, 11)*

1. 5. 3. « eux, euh » : nous relevons souvent dans les transcriptions des cas d'alternance en ce qui concerne cette suite phonique qui peut être interprétée soit comme la forme pronominale soit comme la marque d'hésitation :

*Polonais qui /euh, eux/ tous les Polonais parlent russe (Pic, 25,11)
 comment dire euh des aides pour /eux, euh/ ++ pour l'école quoi (Aides, 5,9)
 certains élèves qui montrent leur(s) note(s) /eux, euh/ pour voir un peu /eux, euh/
 ce qu'ils ont fait (Aides, 10,1)
 c'était des des + jeunes gens qui sortaient à peine de l'adolescence + /eux, euh/
 ils ont pris conscience qu'ils avaient (Alger, 10, 5)*

1.5.4. des sons vocaliques /ø/ et /' / dans :

« deux, de » :
*on l'a filmé on l'a filmé /de, deux/ jour(s) et en fait il était en train de préparer
 (Chaig., 26, 5)
 le Christ a dû voir /de, deux/ vieux tombeaux (Jérusalem, 9, 12)*

« ceux, ce » :
*et ils cherchaient les Juifs et les communistes /tous ceux, tout ce/ qui étai(en)t
 communiste(s) et et Juif(s) (Grotta, 1, 10)*

1. 5..5. D'autres formes sont homonymes, mais nous avons rencontré peu de contextes qui permettent de proposer légitimement plusieurs interprétations ; c'est en particulier le cas des formes suivantes pour lesquelles nous n'avons trouvé qu'un seul exemple où la multi-interprétation est possible.

- « **c'est, s'est** » :
ah ben il a dit ça /c'est, s'est/ répété sur France Inter (Capp., 128, 27)

- « **où, ou** » :
*il y a même un petit village du côté de Saint-Gilles + à Saint-Gilles même
 d'ailleurs + /où, ou/ du côté d'Aigues Mortes où il y a encore les noms (Alger,
 7,7)*

- « **parties, partis** » :
*malheureusement la forêt est assassinée peu à peu d'abord par les gaullistes
 ensuite par les socialistes + c'est un exemple absolu de voir que ces deux /parties,
 partis/ de l'opinion font les mêmes trucs (casane, 9,2)*

- « **regardez, regarder** » :
quand vous rentrez chez vous vous /regardez, regarder/ dans la glace + vous regrettez + normal (Comm. 64, 9)
- « **à accepter, a accepté** » :
il est venu /à accepter, a accepté/ ce qu'on lui offrait (Lafuge, 95)

Certaines formes sont homonymes pour une partie des francophones. Que ce soit le cas pour le locuteur ou le transcritteur, ces prononciations vont alors pouvoir se prêter à des interprétations concurrentes ; c'est le cas, par exemple, des voyelles nasales dans « **hein, un** » :

c'était /hein, un/ des péplums ou des choses comme ça ou des fois des péplums des fois autre chose (Dabat, 23,14)

1. 6. LES DÉCOUPAGES EN MOTS

Dans les exemples précédents, l'hésitation de transcription porte sur des formes monosyllabiques (*ou / où ; c'est / s'est*). Quand ces phénomènes d'hésitation concernent plusieurs syllabes, se conjuguent alors des variations dans le découpage des mots graphiques ; c'est ce que nous avons dans les exemples suivants :

je fais des choix difficiles + mais je l'ai fait, les fais/ au regard de ma conscience (Capp. 43,20)
toute la tribu suivait donc l'émigration, les migrations/ de bisons (Past. 15,3)
je /connaissais, connais c'est/ c'est c'est les seules (Volatiles, 11,4)

Il en de même pour les formes suivantes:

- « **d'eux, de** » :
une personne au chômage qui a des enfants euh et elle s'occupe beaucoup plus de s- /d'eux, de/ ++ de ses enfants (Bonnet,15,12)
- « **qu'eux, que** » :
donc mon père + ma mère se rencontrent en Algérie alors /qu'eux, que/ normalement ils étaient (Alger, 12,16)
il restait plus /que, qu'eux/ les vieux docteurs (Barbara, 9,14)
si vous avez faim parce /qu'eux, que/ ils ont dit oui (Indiens, 1,16)

Des interprétations multiples (conjuguant homonymie et délimitation des mots) peuvent, là aussi, être réservées à une partie des francophones ; c'est semble-t-il le cas pour :

on /a mené, amenait) les enfants en seconde (Aides, 2,1)

Une même suite sonore va pouvoir être interprétée comme un morphème à part entière ou comme une simple amorce de morphème ; ce sont ces deux types d'interprétation que nous avons dans les exemples suivants :

- la suite sonore /s/ est interprétée soit comme la forme pronominale « se » soit comme l'amorce de « seront » :

quand vous serez arrivée à mon âge il y aura beaucoup de choses qui /se, se-/ seront passées ma petite dame (Barian, 11,16)

- la suite /pa/ est interprétée soit comme la particule négative « pas à Montpellier » soit comme une amorce probable de « Paris » :

ça a été prouvé à (pas, Pa-/ à Montpellier ça a été prouvé à Lyon /Barbara, 10,7)

- la suite /ã/ a été interprétée soit comme la préposition « en » soit comme l'amorce de la préposition « entre » qui aurait été produite par anticipation :

dans la forêt de Saint-Germain euh c'est /en, en -/ les les écuries où j'ai mes chevaux c'est entre Maison Lafitte et Saint-Germain /Casane, 8,14)

- Il en est de même pour /a/ et /E&/ :

*il fallait tout tout tout briquer à fond euh tous les soirs c'est à dire toutes les machines passer tous les tapis là, a-/ avec une sorte de produit /Corpus Volatiles l. 9)
il a vécu dans son petit euh bateau euh (hein, un, in -/ insubmersible /Barian, 3,13)*

1. 8. LES CAS DU NOMBRE ET DU GENRE⁸

Les phénomènes d'homonymie, qu'ils portent sur une ou plusieurs syllabes, sont des interprétations se traduisant par des solutions graphiques concurrentes (*a* ou *à*, par exemple). Les cas du nombre et de genre offrent également des exemples d'incertitudes exclusives (*leur* ou *leurs*, *son ami* ou *amie*). Ils permettent aussi une double interprétation, une interprétation « multiple », en quelque sorte ; c'est le cas lorsque ce phénomène touche les formes à valeur distributive :

certains élèves qui montrent leur(s) note(s) /eux, euh/ pour voir un peu (Aides, 9, 16)

L'usage de parenthèses est un artifice typographique déjà largement utilisé pour rendre compte de ces alternances orthographiques. Il est sans doute préférable, pour plus de simplicité, de s'en tenir à ce type de notation

⁸ Une étude de ces cas a été réalisée au GARS par Villaginès-Serra, 1995.

même si tous les phénomènes d'homonymie envisagés ici se traduisent par des alternances orthographiques notées jusque là par la convention suivante : /x, y/.

Dans le cas du nombre et du genre, on relève deux sortes d'exemples qu'il nous semble utile de distinguer :

- dans un certain nombre de cas l'alternance orthographique signale une difficulté d'interprétation. La solution graphique alternative est en concordance avec une hésitation « référentielle » (qu'elle soit exclusive ou multiple). C'est la cas dans les exemples suivants :

tout ce qu'on pouvait avoir comme + biens(s) comme amitié(s) affection (Alger, 3,3)
si tu allais à l'épicerie si tu étais connu(e) si tu avais de l'argent + et ben /on, il(s)/ sortai(en)t un kilo de sucre (Barbara, 1,16)
quand on descendait on (n') était pas trop sûr(es) de les retrouver (Barbara, 7, 14)
on a toujours été très bien soigné(es) (Barbara, 11,10)

L 1- vous l'avez transformé(e) /cet, cette/

L 2- non

L 1- ce livre non (Casane, 17,6)

les fumeurs avaient pris l'habitude de garder leur(s) mégot(s) pendant la guerre tu vois (Barbara, 5,1)

On peut utiliser la même convention pour la troisième personne de l'imparfait pour lesquelles il n'existe pas de marque phonique distinguant le singulier du pluriel ; ces cas d'ambiguïtés sont nombreux et les exemples présentés ici appartiennent tous à cette catégorie d'indécidable « référentielle » :

vous auriez compris mieux compris la vie de mon mari dans le temps + puis quand je voyais où il(s) dormai(en)t (Barian, 6,7)
c'est des super-hommes qui qui qui qui sont attirés par la mer parce qu'il y a des années il(s) pratiquai(en)t la grande pêche c'est-à-dire qu'il(s) partai(en)t quinze jours en mer d'Irlande et puis au bout de quelques années je n'avais pas d'enfant donc je lui ai dit écoute (Barian, 10,8)
mais lui il était heureux + il partait avec ses collègues avec ses avec ses avec son son équipage il(s) partai(en)t euh quinze jours en mer d'Irlande (Barian, 12,9)
quand vous l'avez eu tout-à-l'heure mais il(s) vous aurai(en)t rien dit eux ne disent rien (Barian, 13,3)
alors j'étais voué à la rue à peu près parce qu'il(s) n'avai(en)t pas le temps de s'occuper de moi (Dabat, 12,9)
ils ont conti- continué quand même après même que qu'il y avait du tabac + il(s) continuai(en)t quand même après et c'est pour ça que ta maman avait ramassé un mégot dans Prismic + tiens pour papa hein (Barbara, 5,4)

Pour d'autres cas, on peut être tenté de donner deux orthographes qui pourtant ne semblent pas forcément associées à une hésitation interprétative. Or, on constate que précisément dans ces cas, l'usage orthographique n'est pas stable. C'est le cas pour les syntagmes en *de* dans *confiture de pomme(s), de groseille(s), etc* :

non il n'y avait pas d'homme(s) (Bonnet, 15,12)
faire euh la pêche au(x) thon(s) (Barian, 12,10)
à l'époque où l'Alsace était tantôt française tantôt allemande elle n'arrêtait pas de
changer de camp(s) (Alger, 10,2)
et pour les cartes de rationnement(s) (Barbara, 3,10)
elle savait pas faire d'addition(s) ni des de multiplications (Aides 8,9)

En conclusion, l'ensemble de ces multi-transcriptions sont guidées (et en fait surtout limitées) par le recours à la cohérence textuelle. La vraisemblance attendue sur les propos tenus limite les solutions de multi-transcriptions et donc les calembours souvent possibles ou significations farfelues. Les savoirs et expériences du transcripteur jouent un grand rôle dans le travail d'interprétation du corpus oral (Kerbrat-Orechionni, 1980, Kleiber, 1995). Toutefois, l'inconvénient d'un recours systématique à la cohérence textuelle est d'éliminer, parfois abusivement des solutions inattendues c'est-à-dire improbables au regard du contexte et qui pourtant correspondent à la signification proposée par l'énonciateur. Il est vrai que parfois le locuteur, conscient des risques de mauvaise interprétation de son énoncé, précise la signification qu'il entend transmettre :

les journalistes dont les revendications étiqes sans h ne sont même plus entendues rêvent d'un autre monde (Fr. Culture, commentateur, 27/11/96)
on ne sait pas là-dedans qui c'est qui sait. Ça peut s'écrire de deux façons: qui c'est qui sait, qui sait qui c'est (Lacan, séminaire d'un autre à l'Autre 1968-1969, p.375)

Le corpus oral donne ainsi lieu à un texte multiple malgré le caractère unique de sa source. La fréquence des malentendus, contresens et quiproquos a conduit les linguistes à cesser de considérer ces phénomènes comme des « bavures » (Fuchs et Le Goffic, 1975, Kerbrat-Orechionni, 1980). En ce qui concerne la transcription, comme le soulignent Claire Blanche-Benveniste et Colette Jeanjean (1987, p. 115) son objectivité ne peut être totalement garantie ; « *Le texte authentique est un mythe rassurant, il n'est pas vérifiable* »⁹. La rigueur de la démarche impose

⁹ "Si j'entends aussi bien "a" que "b" dans une séquence de français parlé, je ne saurai donc jamais ce que le locuteur a "dit en vrai". Il semble bien que la réponse soit "non", même si on améliore les appareils d'enregistrements, et l'oreille de celui qui écoute, même si l'on met les gestes et les mimiques et même si l'on questionne le locuteur sur

cette pluralité de transcriptions pour un énoncé oral¹⁰ et ce travail scrupuleux de transcription révèle qu'il y a des zones du texte où surgissent des problèmes grammaticaux persistants.

2- REGROUPEMENTS PROSODIQUES ET « FLOTTEMENT SYNTAXIQUE »¹¹

Les chercheurs en syntaxe qui fondent leurs descriptions sur le recueil préalable de données orales se trouvent tôt ou tard confrontés à la nécessité d'intégrer à leurs transcriptions certaines indications prosodiques indispensables. Par « indispensables », on entendra exclusivement ici les marques prosodiques qui conditionnent très directement l'interprétation syntaxique, et dont l'absence dans la transcription peut rendre l'analyse indécidable, ou tout au moins hasardeuse. Pour donner un exemple simple, la séquence :

je j'ai énormément de mal non pas à me réveiller je me réveille immédiatement je me lève et je me dis il faut que je me prépare (Rhumato., 222)

peut être - en l'absence d'indications prosodiques - interprétée de deux manières : soit *immédiatement* est rattaché, comme élément temporel postposé, à la construction précédente (*je me réveille immédiatement*) ; soit il est rattaché comme complément antéposé à la construction qui suit (*immédiatement je me lève*). De même, dans l'exemple suivant, *évidemment* peut être analysé comme associé¹² à la construction verbale qui précède ou à celle qui suit :

ce qu'il a voulu dire. Le texte authentique fuit, du moins par certains bouts." (Blanche-Benveniste et Jeanjean, 1987, p.112)..

10 Dans l'extrait de corpus Volatiles (l. 16), on pourrait ajouter une multi-écoute: "avant parce que (*tout, tous, tou-*) toutes les personnes. En fait, il en a certainement plusieurs encore qu'une autre re-lecture permettrait de révéler. "Où s'arrêter quand on est un transcritteur?": c'est une question que Claire Blanche-Benveniste et Colette Jeanjean posent (1987, p.150) mais à laquelle elles ne répondent pas.

11 Partie rédigée par F. Sabio.

12 Dans le cadre théorique de l'approche pronominal, on désigne comme *associés* les éléments qui ont l'apparence de compléments sans en posséder les propriétés : *évidemment, d'après moi...* A l'opposé, on emploiera plus loin le terme d'"éléments régis" pour désigner les séquences (sujets ou compléments) qui sont sous la dépendance syntaxique d'un verbe. Voir Blanche-Benveniste, et al. (1990).

alors euh cette mission consiste pour la Poste à aller chercher ou à distribuer le courrier au siège des entreprises évidemment ce ce service est effectué est effectué moyennant paiement d'une redevance annuelle (Alfonsi, 1, 16, corp. C97).

Les ambiguïtés réelles sont évidemment moins fréquentes qu'on pourrait le penser dans la mesure où assez souvent le lexique, ainsi que certains indices tirés du contexte, indiquent laquelle des interprétations syntaxiques concurrentes est la plus probable ; dans l'exemple qui suit, il est raisonnable de penser que l'adverbe est à rattacher à la construction de droite :

le feu arrive au vert tranquillement je démarrai (Sard, 1, 10, corp. C116)

Cependant, à la seule lecture du texte, les cas indécidables restent nombreux, comme on le verra plus loin, et rendent indispensable un minimum d'indications prosodiques dans les transcriptions. Une solution envisageable reviendrait à utiliser dans la transcription certains signes de ponctuation (point, virgule...) à des fins de désambiguïtation syntaxique ; nous resterons fidèles aux recommandations données par Cl. Blanche-Benveniste et C. Jeanjean (1987), de ne pas avoir recours à ce procédé de notation, qui est typique de la langue écrite orthographiée et qui ne peut en aucun cas être assimilé à un système de notation de la prosodie.

Une autre possibilité consisterait à faire figurer dans les corpus de langue parlée une transcription prosodique disposée en regard de la transcription du texte : plusieurs systèmes de codage intonatif ont été proposés pour le français¹³. Cependant leur sophistication-même les rend difficilement utilisables, non pas tant - comme on le pense souvent - parce qu'ils nécessitent l'utilisation d'outils techniques, tels que le détecteur de mélodie, souvent indisponibles pour les non-phonéticiens, mais surtout parce que l'adoption d'un système de transcription prosodique est toujours subordonnée à certaines options théoriques et méthodologiques qui constituent une spécialité à part entière¹⁴.

La solution proposée ici poursuit un objectif beaucoup plus modeste mais sans doute plus réaliste : elle consiste d'abord à faire un relevé des difficultés d'interprétation syntaxiques qui sont directement imputables à l'organisation prosodique, après quoi il devra être possible de proposer non pas un système de codage prosodique au vrai sens du terme, mais plutôt un système assez simple d'indications à même de constituer une « aide à l'interprétation syntaxique » des textes oraux, et qui pourra - si on le juge utile - prendre place parmi les conventions de transcription. Pour revenir

¹³ Voir les propositions du groupe PACOMUST (Astesano et al.), de l'Université de Provence, ou encore le système qu'a proposé P. Mertens.

¹⁴ Ajoutons qu'il n'existe à ce jour aucune méthode standardisée de notation intonative qui serait comparable à l'API pour la transcription segmentale.

aux exemples donnés plus haut, il nous semble que le fait d'indiquer que les syntagmes *immédiatement* ou *évidemment* sont perçus par le transcripteur comme « branchés à gauche » ou « branchés à droite » pourrait suffire à faciliter l'analyse syntaxique, sans exiger aucune formalisation prosodique compliquée.

C'est de ce relevé qu'il sera question ici : nous avons parcouru plusieurs corpus de français parlé, et avons systématiquement recueilli les énoncés pour lesquels l'absence d'indications prosodiques peut constituer une gêne pour l'interprétation syntaxique : il nous semble en effet qu'un repérage préalable des secteurs de la grammaire qui apparaissent les plus sensibles aux faits d'organisation prosodique pourra être utile à la réflexion ultérieure sur la consignation des données orales. Evidemment, il est impossible de prétendre que, pour chacun des énoncés dont il va être question, les faits prosodiques seraient *à tous les coups* capables d'opérer la désambiguïsation espérée : qu'ils en soient seulement susceptibles est suffisant pour notre propos.

Insistons sur le fait que nous nous placerons exclusivement dans la perspective de la perception prosodique globale : pour reprendre l'exemple déjà donné, c'est le fait que l'adverbe *immédiatement* puisse être perçu comme lié à ce qui précède ou à ce qui suit qui représente pour nous un intérêt, dans la mesure où cela pourra avoir des conséquences directes sur l'analyse syntaxique ultérieure. Par contre, la nature précise des corrélats phonétiques qu'utilisent les auditeurs pour percevoir la présence ou l'absence d'un phénomène de rattachement prosodique (présence de pauses, faits temporels relatifs à l'allongement, indices intonatifs) ne sera pas du tout évoquée ici.

3. 1. PROSODIE ET REGROUPEMENT MICRO-SYNTAXIQUE : LES SEGMENTS FLOTTANTS

La difficulté d'analyse la plus fréquente dans nos relevés apparaît lorsque, dans une séquence de deux constructions verbales, un syntagme peut être affilié avec autant de vraisemblance à la construction gauche qu'à la construction droite. A. Berrendonner et M.-J. Reichler-Béguelin (1989, 108) parlent de segmentation floue et de clauses flottantes pour décrire ce phénomène. Comme l'ont noté Cl. Blanche-Benveniste et C. Jeanjean (1986, 141), les conséquences de la difficulté de segmenter ne sont peut-être pas très grandes pour l'information que donne le texte, mais elles sont importantes pour le découpage syntaxique.

3.1.1. Segments flottants à interprétation syntagmatique

Ce premier type constitue un cas de figure assez simple puisque l'affiliation du segment flottant à droite ou à gauche n'occasionne aucun changement de statut syntaxique : dans les exemples donnés ci-après, chacune des séquences en italiques peut être affiliée à la construction précédente ou suivante, mais son statut syntaxique, lui, reste inchangé quel que soit le mode de rattachement considéré. Il ressort que le plus souvent, les éléments flottants de ce type semblent être des associés (Cf. note 1). Moins attendu peut-être est le fait que certains éléments régis par le verbe (les « véritables » compléments) peuvent flotter entre deux constructions : c'est le cas notamment des temporels et des locatifs. Voici quelques exemples (les éléments flottants sont notés en gras) :

segments flottants associés

qu'est-ce que vous avez à dire aux parents parfois vous avez à leur dire qu'ils ne vous comprennent pas euh d'après ce que j'ai entendu vous avez toutes les raisons d'être heureux mais... (cité par Berrendonner et Reichler-Béguelin, 1989, 109)

*vous allez à votre poissonnier **naturellement** + vous prenez un peu de la de du poisson + de fond (Douss, 9, 13, corp. C121)*

*c'est l'horreur **le fonctionna- le fonctionnariat en médecine** c'est l'horreur (Rhumato., 224)*

*pendant une minute on s'est écouté respirer **Léo et moi** on s'est dit l'ami est parti (Lefebvre, 6, 5)*

*et c'était fantastique **Paris** c'était vraiment la la ville quoi (Lefebvre, 13, 9)*

segments « flottants » régis

valeur temporelle :

*j'ai porté plainte j'ai pris un avocat et j'ai eu gain de cause + **six mois après** euh : l'affaire était réglée (Sard, 9, 13, corp. C116)*

*ça y est il coulait **le mois d'après** il était au chômage (Baral ; 21, 8)*

*je j'ai énormément de mal non pas à me réveiller je me réveille **immédiatement** je me lève et je me dis il faut que je me prépare (Rhumato., 222)*

*lui je l'attends je l'attends depuis **des années** lui je lui téléphone chez lui et il répond plus maintenant il est plus là (Lefebvre, 1, 3)*

valeur locative :

*parce qu'il y a beaucoup d'ironie **dans ce film** il y a il y a de l'ironie (cité par Berrendonner et Reichler-Béguelin, 1989, 109)*

*si nous avons un très bon taux de : synthèse protéique + **dans le lait** nous aurons des : possibilités de transformation(s) fromagère(s) meilleure(s) (Detem, 9, 3, corp. C121)*

autres :

je pensais pas que c'était le plus important pour les malades c'était fondamental bien sûr mais c'était pas suffisant (Rhumato., 221)
le collège international de philosophie est une instance qui est euh basée à Paris + comme beaucoup de euh de de centres de de + de recherche ++ donc elle est basée c'est une instance qui est basée au 1 rue Descartes à Paris (Solerieu, C83, 1, 1)

3.1.2. Segments flottants à interprétation syntagmatique ou paradigmaticque

Dans ce cas, le segment flottant peut être interprété de deux manières :

- s'il est interprété comme branché à gauche, il devra être analysé - d'un point de vue paradigmaticque - comme la seconde instanciation « en liste » d'une place syntaxique déjà actualisée¹⁵;

- s'il est interprété comme branché à droite, il sera alors analysé - d'un point de vue syntagmaticque - comme occupant une place syntaxique « à lui tout seul ».

Quatre cas de figure apparaissent très régulièrement dans nos relevés :

1) Interprétation en « associé » ou « régi »

Soit cet exemple :

alors X j'ai expliqué à à ce monsieur à ce colonel je lui disais mon camarade est dans l'infanterie (Lefebvre, 20, 11) [paradigm. : 2e instance Objet prép. / syntagm. : Associé]

Deux organisations syntaxiques sont envisageables, selon que *à ce colonel* est perçu comme branché sur ce qui précède ou sur ce qui suit :

- si *à ce colonel* est branché sur la construction *j'ai expliqué à ce monsieur*, il devra être analysé comme la seconde instanciation, « en liste », de l'objet prépositionnel de *expliqué*. La présentation en grille syntaxique¹⁶ permet de visualiser la nature paradigmaticque du lien entre *à ce monsieur* et *à ce colonel* :

¹⁵ Ainsi, dans une séquence comme *j'ai expliqué à ce monsieur à ce colonel*, on dira que l'élément *à ce colonel* est en liste avec *à ce monsieur* : il possède la même fonction syntaxique, et vient apporter une précision lexicale à la dénomination initiale, sémantiquement plus vague. On sait que le développement de listes constitue un procédé fondamental dans l'activité de dénomination (Cf. Blanche-Benveniste 1985).

¹⁶ Le GARS utilise cette représentation en grilles syntaxiques afin de visualiser le déroulement syntagmaticque et paradigmaticque des textes, oraux comme écrits. Pour une présentation récente, voir Blanche-Benveniste (1995)

j'ai expliqué à
 à ce monsieur
 à ce colonel
 S+V Objet

- si par contre à ce colonel est perçu comme branché sur la construction de droite, il sera analysé comme un associé antéposé à la construction en dire (aucune liste paradigmatic n'apparaît alors) :

à ce colonel je lui disais mon camarade est dans l'infanterie

On voit ainsi que, contrairement aux éléments flottants à interprétation uniquement syntagmatique, traités *supra*, le rattachement des éléments du type à ce colonel à droite ou à gauche induit une différence de statut grammatical : complément dans un cas, associé dans l'autre.

Voici d'autres exemples du même type :

*lorsque ça a mijoté une vingtaine de minutes je prends la grande poêle la vraie poêle à paella il faut que ce soit une poêle euh d'un diamètre assez grand (Dupuy, C83, 8, 12) [paradigm. : 2e instance Objet non prép. / syntagm. : Associé]
 tout ordinateur a besoin pour travailler de programmes qui sont stockés dans la mémoire centrale et le cœur de l'ordinateur c'est l'unité centrale de traitement (Deglie, C83, 1, 12) [paradigm. : 2e instance Locatif / syntagm. : Associé]*

Nous avons en outre relevé quelques exemples dans lesquels l'interprétation comme réactionnel induit un fait de non-connexité : dans les énoncés suivants, on peut soit interpréter l'élément flottant comme un élément associé branché à droite, soit comme la seconde instanciation de la position sujet, avec en outre un phénomène d'« éloignement linéaire » :

*mon frère a toujours eu des dons mon frère qui a trois ans de plus que moi il a toujours eu des dons [branch. à gauche : 2e instanc. Sujet (non connexe) / branch. à droite : Associé]
 il y a la lumière aussi qui est très très importante à respecter le : la lumière solaire avec les les ultraviolets en général on essaye d'en avoir le maximum possible pour chauffer l'eau (Chagui, C83, 4, 1) [branch. à gauche : 2e instanc. Sujet (non connexe) / branch. à droite : Associé] 17*

2) Le cas des proformes

Ici, la double interprétation - en liste ou pas - surgit au sein d'une seule et même construction verbale dans laquelle le sujet est réalisé par une proforme clitique :

17 Cet énoncé est plus complexe que le précédent puisque le sujet "la lumière" se trouve singularisé au sein d'un dispositif en "il y a ... qui".

il faudra que tu fasses attention parce que ça ça t'indique qu'il y a différentes différents aspects de cette pièce (Sanch, 20, 15, corp. C116) [paradigm. : 1e instanc. Sujet / syntagm. : Associé]

ça ça me frappe beaucoup (Rhumato., 220) [paradigm. : 1e instanc. Sujet / syntagm. : Associé]

vous: vous les faites revenir (Douss, 9, 6, corp. C121) [paradigm. : 1e instanc. Sujet / syntagm. : Associé]

et vous et vous vous restez toujours chez vous (Lefebvre, 11, 4)

Les proformes en gras peuvent être interprétées soit comme un phénomène d'hésitation qui se traduit syntaxiquement par un « piétinement » paradigmatique sur la place sujet, soit comme une organisation en *associé antéposé* + *sujet*. Cette double interprétation ne peut, évidemment, survenir que lorsque la proforme non clitique - interprétable comme associé - et la proforme clitique sujet sont morphologiquement identiques : c'est le cas pour « *elle elle* », « *nous nous* », « *vous vous* », « *elles elles* » et « *ça ça* ». L'interprétation en liste est le plus souvent corrélée à un très fort allongement sur le premier pronom, indice d'hésitation.

3) Les séquences adjectivales

Les constructions nominales qui comportent plusieurs rections adjectivales peuvent poser un problème d'analyse syntaxique très spécifique. Soit une séquence comme :

une femme énorme dégueulasse (Lefebvre, 7, 8)

Il peut être difficile de savoir si l'élément régi le plus à droite du constructeur nominal (*dégueulasse*) est construit par l'ensemble *une femme énorme*, ou par le constructeur nominal seulement (*une femme*).

Dans la première interprétation - totalement syntagmatique - on aurait un parenthésage du type :

((*une femme énorme*) (*dégueulasse*))¹⁸

Dans la seconde, la représentation en grilles ferait clairement apparaître un phénomène de liste paradigmatique :

<i>(une femme)</i>	<i>(énorme)</i>
	<i>(dégueulasse)</i>

Ainsi, bien que les interprétations concurrentes n'induisent pas ici une opposition de type régi/associé, comme précédemment, on retrouve

¹⁸ Blanche-Benveniste (1995) parle d'organisation "à encerclement".

néanmoins l'hésitation entre interprétation syntagmatique et paradigmaticque.

Relèvent du même fonctionnement les exemples suivants :

une maladie particulière à soigner (corpus rhumato) [syntagm. : régi par « une maladie particulière » / paradigm. : régi par « une maladie »]
 (...) *et en particulier avec notre société actuelle occidentale (corp. Rhumato)*
 [syntagm. : régi par « notre société actuelle » / paradigm. : régi par « notre société »]
c'est des pièces encastrées emboîtées (Sanch, 20, 1, corp. C116) [syntagm. : régi par « des pièces encastrées » / paradigm. : régi par « des pièces »]

Dans les exemples suivants, la seconde rection nominale est constituée par une relative :

j'ai un tas d'auteurs que j'adore que je relis tout le temps (corp. rhumato)
 [syntagm. : régi par « un tas d'auteurs que j'adore » / paradigm. : régi par « un tas d'auteurs »]
des personnes âgées qui ont une maladie chronique (Ghildi, 33, 6) [syntagm. : régi par « des personnes âgées » / paradigm. : régi par « des personnes »]
des gens qui viennent qui ont l'œil qui coule (Ghildi, 41, 1) [syntagm. : régi par « des gens qui viennent » / paradigm. : régi par « des gens »]¹⁹

4) Rection nominale et rection verbale

Dans les séquences suivantes, le segment flottant peut être analysé soit comme une rection verbale, soit comme une rection nominale :

je démarrais mais + les panneaux des panneaux qui étaient placés euh: très bas sur le trottoir me cachaient la visibilité de la route + qui venait en face la route de Lavéra (Sard, 1, 10 corp. C116) [syntagm. : régi par « venir (en face) » / paradigm. : régi par « la visibilité »]
des choses qu'on pense qu'on dirait évidemment beaucoup plus mal (corp. rhumato) [syntagm. : régi par « penser » / paradigm. : régi par « des choses »]²⁰
c'était sur l'autoroute vers Aubagne (corp. Accident de voiture, 1, 14) [syntagm. : régi par « être » / paradigm. : régi par « l'autoroute »]

Par exemple, le dernier de ces énoncés pourrait recevoir une analyse dans laquelle l'enchaînement grammatical s'opère de manière exclusivement syntagmatique :

c'était sur l'autoroute vers Aubagne

¹⁹ Les deux derniers exemples sont donnés par Blanche-Benveniste (1995).

²⁰ Pour cet exemple, l'écoute de l'enregistrement rend évident que c'est l'interprétation paradigmaticque qui est la bonne.

Dans cette interprétation, la séquence *sur l'autoroute vers Aubagne* serait conçue comme régie par le verbe « être », et signifierait *sur l'autoroute qui va vers Aubagne*.

Mais on pourrait concevoir une seconde analyse qui ferait de *sur l'autoroute* un syntagme homologue à *vers Aubagne* :

*c'était sur l'autoroute
vers Aubagne*

Ici, *vers Aubagne* constitue une reprise en liste de la rection locative du verbe *être*, dont le sens (*à hauteur d'Aubagne*) est clairement différent de l'interprétation précédente. Les exemples donnés plus haut connaissent un fonctionnement identique.

D'autres exemples se prêtent à une double interprétation en rection nominale et/ou rection verbale, sans toutefois induire aucun phénomène de rupture paradigmatique. C'est le type même de structure qui a le plus souvent retenu l'attention des chercheurs en prosodie :

il a peint la jeune fille en noir (cité par M.-C. Hazaël-Massieux, 1983, 105)
les Dupont ont vendu leur château en Espagne (cité par A. Di Cristo, in Rossi et al., 1980, 283)
Daniel a téléphoné à la bouchère de Rousset (cité par A. Di Cristo, in Rossi et al., 1980, 286)
ce travail donc a été réalisé dans le cadre de la valorisation + du: pastoralisme dans les régions méditerranéennes (Detem, 1, 5, corp. C121)
mon père avait toujours une un /type, X/ qui lui prêtait une maison en Italie (Lefebvre, 16, 11)

La différence de parenthésage serait du type :

[(il a peint) (la jeune fille) (en noir)] opposé à [(il a peint) (la jeune fille en noir), [(il a téléphoné) (à la bouchère) (de Rousset)] opposé à [(il a téléphoné)(à la bouchère de Rousset)].

3. 2. PROSODIE ET REGROUPEMENT MACRO-SYNTAXIQUE

Il ne s'agit plus ici de choisir entre le caractère régi ou non d'une unité, ni d'identifier l'élément recteur d'une construction, mais d'évaluer la configuration macro-syntaxique globale d'un énoncé, qu'on décrit en termes de « dispositifs de la rection ». Nous nous contenterons ici de citer quelques phénomènes bien connus.

- Certaines séquences en *c'est... que/i* peuvent s'analyser soit comme des constructions relatives, soit comme des phénomènes d'extraction :

c'est une boîte que tu tournes (Sanch, 14, 2, corp. C116)
c'est une révision que je fais (Pialat, 46, 7, corp. C96)
c'était toujours les mêmes trucs qui revenaient (Anguenot, 5, 16, corp. C94)

- Une autre difficulté d'analyse provient du fait que les éléments régis par un verbe peuvent dans bien des cas être placés non pas après mais avant le verbe. On parle en ce cas de dispositif à binarisation :

plus personne on connaît (Canadell, 11, 27)
Un grand troupeau de pleurnicheurs ils formaient (écrit, Céline)

Or, quand les énoncés apparaissent non pas isolément mais au sein d'une séquence textuelle, il est parfois délicat d'apprécier le caractère antéposé de certains compléments. Soit l'exemple :

il a bricolé jusque jusqu'à l'exode quoi ouais jusqu'au mois de juin ça a duré du mois de septembre au mois de juin (Baill, 5, 12, corp. C47)

Deux propositions d'analyse sont possibles, que nous disposons sous forme de grilles syntaxiques :

<i>il a bricolé</i>	<i>jusqu'à l'exode</i>
<i>jusqu'au mois de juin</i>	
<i>ça a duré</i>	<i>du mois de septembre au mois de juin</i>

C'est la solution la plus simple puisque chacune des deux constructions verbales connaît la même organisation linéaire : sujet + verbe + objet temporel. Par contre, dans la grille suivante,

	il	a bricolé	jusqu'à l'exode
jusqu'au moi			
	s de juin	ça	a duré
du mois de septembre au mois de juin			

la construction *jusqu'au mois de juin ça a duré* s'analyserait comme un dispositif à binarisation. En outre, *du mois de septembre au mois de juin* forme une liste avec le syntagme précédent *jusqu'au mois de juin*. A noter que cette dernière analyse suppose que *duré* s'achève sur une intonation basse « de fin de phrase », et que *du mois de septembre au mois de juin* constitue un groupe intonatif autonome²¹.

- Le troisième mode d'organisation macro-syntaxique qui intéresse notre propos est celui que l'on désigne sous le terme de « projections »²² : les

21 Dans l'écriture, ce texte pourrait être ponctué ainsi : "Il a bricolé jusque jusqu'à l'exode. Jusqu'au mois de juin ça a duré. Du mois de septembre au mois de juin."
 22 Voir Bilger (1983) et Blanche-Benveniste et al. (1990), pp. 151 sq.

projections sont des séquences macro-syntaxiques en *Préfixe + Noyau* dont les termes connaissent une affiliation paradigmatique avec ceux de la construction verbale précédente. Soit l'exemple :

nous avons euh : + au point de vue de la : + diversité de ces protéines + un une plus grande diversité donc au point de vue de la qualité du lait une meilleure qualité à la transformation fromagère (Detem, 5, 3, corp.C121)

Dans les deux interprétations proposées, cette construction verbale met en jeu certains développements paradigmatiques :

1) Dans la première analyse, c'est la place objet qui se développe en liste (« nous avons ça (et) ça ») :

nous avons (...)	une plus grande diversité donc au point de vue de la qualité du lait
	une meilleure qualité à la transformation fromagère

2) Dans la seconde interprétation, il y a une projection paradigmatique de deux places, qu'on peut gloser par *nous avons à ce point de vue ceci, donc, à ce point de vue cela* (le « donc » servant de joncteur aux deux unités) :

nous avons	au pt de vue de la diversité de ces protéines	une plus grande diversité
donc	au pt de vue de la qualité du lait	une meilleure qualité (...)

3. 3. LES RUPTURES DE CONSTRUCTIONS

Il arrive assez souvent dans les productions orales spontanées que certaines constructions syntaxiques se trouvent brutalement interrompues par les locuteurs. Parfois, les seules marques segmentales ne permettent pas d'affirmer avec certitude qu'on a affaire à une rupture de construction : certains faits prosodiques, tels qu'un fort allongement à valeur d'hésitation, constituent dans ces cas-là des marques capitales de disjonction syntaxique. En voici trois exemples :

j'ai beaucoup de copines et de copains mais plus de copines euh qui euh qui me téléphonent pour me parler me parler d'elles pendant euh un certain temps et je suis toujours très très frappée que euh bon elles m'aiment bien (corpus rhumato, 246, 7)

La lecture du texte ne laisse supposer aucune rupture de construction ; pourtant il est évident lorsqu'on écoute cet extrait que la construction initiée par *je suis toujours très très frappée que* est interrompue, et que la construction *elles m'aiment bien* n'a aucun lien de dépendance avec ce qui précède. Ainsi la locutrice ne veut pas dire du tout qu'elle est frappée que ses copines l'aiment bien. Comme elle l'explique un peu plus loin, ce qui la

frappe c'est de voir à quel point les gens ne sont pas dans l'échange ça c'est quelque chose de très frappant.

Voici un second exemple mettant en jeu une construction pseudo-clivée :

alors ce que je fais pour moi ben c'est ce qui est le plus important maintenant c'est de d'essayer de faire de la peinture (corpus Rhumato)

On perçoit une rupture de construction après le premier *c'est*. La grille correspondante à cette interprétation serait :

alors ce que je fais pour moi ben	c'est
ce qui est le plus important maintenant	c'est de d'essayer de faire de la peinture

Ce dernier exemple a été produit par un homme politique à la télé :

quelqu'un de mon âge qui vous dirait comme ça à brûle-pourpoint je vais être candidat parce que c'est ridicule

En fait, il n'est pas question pour le locuteur d'expliquer qu'il veut être candidat *parce que c'est ridicule*. A l'audition de cet extrait, les indices prosodiques font clairement apparaître que la construction relative se trouve brutalement interrompue après *parce que*, laissant la séquence causative « en suspens ».

3. 4. REMARQUE FINALE

A la lecture des corpus de français parlé, il apparaît que les cas d'ambiguïté structurelle imputables à l'absence d'indications prosodiques dans la transcription se ramènent en fait à un nombre restreint de phénomènes linguistiques, tels que la possibilité pour un syntagme

- d'être branché à droite ou à gauche,
- d'être situé à deux niveaux différents de hiérarchie syntaxique,
- de se trouver réalisé sous certains dispositifs macro-syntaxiques (comme l'extraction),
- de faire partie d'une construction interrompue.

Nous avons entrepris de consigner²³, dans certaines de nos transcriptions, ces phénomènes de rattachement prosodique en nous limitant aux cas pour lesquels l'impossibilité d'interpréter la structure syntaxique est bien réelle, et à condition, bien évidemment que l'écoute de la bande sonore fournisse des indices phonétiques clairs, ce qui est loin d'être toujours le cas.

On pourrait indiquer les faits de rattachement prosodique et les effets syntaxiques qu'ils induisent de la manière suivante :

« rattaché à la construction verbale suivante / précédente »
 « branché à gauche/droite : interprétation syntagmatique /
 paradigmatique »
 « régi par le verbe X / le nom Y »
 « interruption de construction »

Le groupe GEDO se compose des membres suivants :

Mireille BILGER	Mylène BLASCO
Paul CAPPEAU	Berthille PALLAUD
Frédéric SABIO	Marie-Josée SAVELLI

Bibliographie

- ASTESANO, C., BERTRAND, R., BROUSSEAU, M., CHAFCOULOFF, M., DI CRISTO, A., GHIO, A., HIRST, D., LAPIERRE, S., NICOLAS, P., ROMEAS, P., SABIO, F., VINCENT, M. 1995. « Le projet PACOMUST, un corpus de parole continue multistyle. Objectifs et choix méthodologiques », *TIPA*, 16, 9-38.
- BERRENDONNER, A., REICHLER-BÉGUELIN, M.-J. 1989, « Décalages : les niveaux de l'analyse linguistique », *Langue Française*, 81, 99-125.
- BERRENDONNER, Alain, 1981, *Éléments de pragmatique linguistique*, Paris, Les éditions de minuit
- BILGER, Mireille, 1983, *Étude distributionnelle de la coordination par ET*, Thèse de troisième cycle, Université de Provence.

²³ Dans nos transcriptions, nous faisons figurer l'information prosodique sous forme de notes de bas de pages.

- BLANCHE-BENVENISTE, C. 1985, « La dénomination dans le français parlé : une interprétation pour les 'répétitions' et les 'hésitations' », *Recherches sur le français parlé*, 6, 110-130.
- BLANCHE-BENVENISTE C. & JEANJEAN C., 1987, *Le français parlé. Transcription et édition*. Didier Érudition, Paris.
- BLANCHE-BENVENISTE, C., BILGER, M., ROUGET, C., EYNDE, K. v.d., MERTENS, P. 1990, *Le français parlé. Études grammaticales*, Paris, Éditions du C.N.R.S.
- BLANCHE-BENVENISTE C., 1993, « Faire des phrases », *Le français aujourd'hui*, 101, *Norme(s) et pratiques de l'oral*, 5-15.
- BLANCHE-BENVENISTE C., 1993, Une description linguistique du français parlé, *Le Gré des Langues.*, 8-23.
- BLANCHE-BENVENISTE, C. 1995, « Le semblable et le dissemblable en syntaxe », *Recherches sur le français parlé*, 13, 7-32.
- CERVONI Jean, 1987, *L'énonciation*. PUF, Paris
- COSNIER J. ET KERBRAT-ORECCHIONI Catherine (Eds.), 1987, *Décrire la conversation*. Presses Universitaires de Lyon
- de SAUSSURE F., 1972, *Cours de linguistique générale.*, Paris, Payot.
- DEVOS R., 1989, *A plus d'un titre.*, Paris, Ed. Olivier Orban.
- DOLTO F., 1981, *Au jeu du désir*. Le Seuil.
- ENCREVÉ, Pierre, 1988, *La liaison avec et sans enchaînement. Phonologie tridimensionnelle et usages du français*, Paris, Seuil.
- FUCHS C. & LE G OFFIC P., 1975, *Initiation aux problèmes des linguistiques contemporaines*. Hachette université, Paris.
- GADET Françoise, 1989, *Le français ordinaire*, Paris, A. Colin.
- GALISSON R. & COSTE D., 1976, *Dictionnaire de didactique des langues.*, Paris, Hachette.
- GIOVANNONI Dominique & SAVELLI M. J., 1990, « Transcrire, traduire, orthographier le français parlé », *Recherches sur le français parlé*, 10, 19-37.
- HAZAËL-MASSIEUX Marie-Christine. 1983, « Le rôle de l'intonation dans la définition et la structuration de l'unité de discours », *BSLP*, LXXVIII, 99-160.
- KERBRAT-ORECCHIONI C. 1980, *L'énonciation de la subjectivité dans le langage* A. Colin, Paris.
- KERBRAT-ORECCHIONI C., 1996, « Texte et contexte », *Scolia*, 6, 39-60.
- LACAN, Jacques, 1966, *Les Ecrits*, Paris, Seuil

- LACAN, Jacques, 1968, *D'un autre à l'Autre*, Texte photocopé d'une transcription.
- LOUFRANI, C., 1990, *Analyse de discours de locuteurs classés comme aphasiques. De l'art de tourner autour du mot*. Thèse de doctorat, Université de Provence, Département de Linguistique Française, Aix.
- MERTENS, P. 1987, *L'intonation du français. De la description linguistique à la reconnaissance automatique*, Thèse de Doctorat, Université K.U.L., Leuven.
- MOUCHON, J., & FILLOL F., 1980, *Pour enseigner l'oral*. Paris, Cédic.
- PARISOT, H., 1962, *Lewis Carroll*. Paris, Ed. P. Seghers.
- PASQUALE, A. 1989, *Étude diachronique du langage d'un enfant inintelligible de 7;4 à 8;5. Mémoire de maîtrise 1988-1989*. Université d'Aix-Marseille I, Département de Linguistique française.
- RAINGEARD M. & LORCHEIDER U, 1977, Edition de corpus de français parlé, *Recherches sur le Français Parlé, 1*, 14-29.
- ROSSI, M., DI CRISTO, A., HIRST, D., MARTIN, P.I., NISHINUMA, Y. 1981, *L'intonation de l'acoustique à la sémantique*, Paris, Klincksieck.
- THIBAUT P. & VINCENT D., 1990, *Un corpus de Français parlé. Montréal 84 : Historique, méthodes et perspectives de recherche*, *Recherches Sociolinguistiques/1*, Bibliothèque nationale du Québec.
- VILAGINÉS SERRA, E., 1995, *Étude sur les indécisions à l'oral du nombre et du genre*, Université de Provence, Département de linguistique française, mémoire de maîtrise.
- VION R., 1992, *La communication verbale. Analyse des interactions*, Hachette, Paris.

TRANSCRIPTIONS ET TECHNOLOGIES

Beaucoup de débats engagés sur l'utilisation des corpus de langue parlée ont évolué récemment, sous l'influence des nouvelles technologies diffusées par les ressources informatiques. Je voudrais ici rappeler quelques questions importantes issues directement de ces avancées techniques.

Lorsque le terme de *corpus* apparaît en France, vers 1923 (d'après A. Rey¹), il n'est pas tourné vers les technologies modernes. Il vient du droit ; *corpus juris* désigne depuis fort longtemps le recueil des lois du droit romain. Lorsque, à partir de 1956, Gougenheim, Rivenc et Sauvageot enregistrent des conversations pour élaborer *Le Français fondamental*, nous dirions qu'ils collectaient des *corpus*. Mais eux ne se servaient pas de ce mot. Ils n'utilisaient que le terme d'*enquête*.

Vers les années 1970, le mot *corpus*, doté de sa nouvelle acception importée des États-unis, s'installe dans la nomenclature des linguistes français. Dans cette nouvelle acception, il a une allure technique. « Le mot est pédant », écrivait R. L. Wagner en 1973, « autant que celui de *campus* »².

A l'époque actuelle, la diffusion et l'utilisation des corpus de langue parlée ont pris en Europe une grande extension. A quoi servent-ils ? A des fins d'études universitaires sans applications extérieures au domaine : grammaire, recherche lexicale, pragmatique étude des interactions. Mais également à la recherche appliquée, par exemple dans le dialogue homme-machine, les modèles de langage pour entraîner les systèmes d'analyse automatisés, les bases de données lexicales pour des dictionnaires, etc. Les

1. A. Rey, 1992, *Dictionnaire Historique de la Langue française*. Paris: éditions Le Robert.

2. R.L. Wagner, 1973 *La Grammaire française, volume 2. La grammaire moderne. Voies d'approche. Attitude des grammairiens*. Paris: SEDES, p. 100. Le pluriel latin, *corpora*, adopté par d'autres langues, ne s'implante pas en français; on dira "des corpus".

traitements des *corpus* se sont diversifiés et sont devenus, dans certaines entreprises, hautement technologiques.

« The area of natural language processing is particularly sensitive to technological advance » (G. Leech, EAGLES 1996)

Je propose de rendre compte, dans les grandes lignes, des informations que diffuse à ce sujet un groupe de chercheurs européens réunis dans un organisme nommé EAGLES (Expert Advisory Group in Language Engineering Standards). Ce groupe s'intéresse depuis quelques années à la standardisation de ce qu'il est convenu d'appeler *la linguistique sur corpus*. J'utiliserai particulièrement les publications de Joachim Llisterri (Universidad Autonoma de Barcelona) pour la transcription³, de Geoffrey Leech (University of Lancaster) pour les annotations grammaticales⁴, et de John Sinclair pour la typologie des textes⁵. Je ferai allusion aux conventions établies par NERC (Network of European Reference Corpora) et TEI (Text Encoding Initiative).

1. LES CONVENTIONS DE TRANSCRIPTION

Le travail sur corpus a amené les linguistes à avoir un respect des données au même titre que dans d'autres domaines de recherche⁶. Lorsqu'il s'agit de représenter par écrit des données orales, ce respect des données exige qu'on fasse des choix (on ne peut pas être fidèle à tous les phénomènes en même temps) et qu'on développe des conventions commodes. Dans la perspective actuelle d'échanges de travail en Europe, ces conventions doivent être largement partagées.

La démarche habituelle consiste à partir, en premier lieu, d'une transcription orthographique (complétée éventuellement par divers systèmes d'annotations). En effet, la plupart des chercheurs qui travaillent actuellement sur les langues parlées tiennent à disposer de grands corpus

³. Llisterri, J. (1994a) Spoken Texts. Draft-Work in Progress, EAGLES Document EAG-CS/IR T7.1, October 1994.

⁴. Leech G. (1994) Morphosyntactic Annotation . Draft- Work in Progress. EAGLES Document EAG-GCS/IR-T3.1.

⁵. Sinclair J. (1994), Corpus Typology. Draft- Work in Progress. EAGLES Document EAG-CSG/IR-T1.1. J'utiliserai aussi les informations glanées au Colloque de EAGLES qui s'est tenu à Madrid du 6 au 8 janvier 1996.

⁶. G. Leech, 08/01/96, p. 2).

pour en faire éventuellement des études quantitatives. Ils font rarement, en ce cas, des transcriptions phonétiques. C'est en partie pour une question de temps de travail, bien que certaines transcriptions phonétiques puissent aujourd'hui être faites semi-automatiquement. Mais c'est surtout affaire de spécialisation.

On se sert de l'Alphabet Phonétique International pour les études consacrées spécifiquement au signal sonore (*Speech Research*). Mais il est très rare de rencontrer des transcriptions de langue parlée (*Spoken Language*) faites systématiquement en API. Cette différence entre *Speech Research*, portant sur l'aspect phonique du langage, et *Spoken Language Research*, (difficile à rendre en français), portant sur l'étude grammaticale, lexicale ou discursive de productions orales, a représenté pendant longtemps une frontière majeure entre deux disciplines. Les études grammaticales et lexicales cherchent à disposer d'échantillons de langage « spontané », stocké en grandes quantités. Les travaux d'ordre phonique portent généralement sur le langage « de laboratoire », représenté par des collections beaucoup plus restreintes. Les avancées technologiques récentes vont sans doute les rapprocher, de l'avis de J. Llisterrí, parce que le langage « de laboratoire » pourra absorber davantage de discours dit « spontané ».

« Even general purpose corpora of impromptu, unrehearsed, unscripted, non elicited informal conversations now seem to arouse some interest in speech research as they can be used as test-beds for speech recognition systems » (Teubert, 1993 :4)⁷.

D'autre part, depuis qu'il est possible de faire coïncider la représentation orthographique et une représentation automatique du signal sonore, les techniques sont davantage mises en commun. On peut en effet, au prix d'une certaine dose d'intervention manuelle, transcrire en faisant correspondre une ligne d'écriture orthographique, une ligne de transcription phonétique et une ligne de représentation prosodique. Mais, pour l'instant, comme la mise en place des différents paramètres n'est pas simple, le travail sur la langue parlée n'utilise en général la transcription phonétique que pour un petit nombre d'annotations rajoutées à la transcription orthographique.

Les mots sont donc transcrits comme des unités lexicales graphiques, dont on n'étudie pas le détail des réalisations. Les conventions NERC adoptées en 1992 précisaient que, dans ces transcriptions, on devait suivre les standards orthographiques admis conventionnellement ; on n'utilisait les contractions de mots que dans la mesure où elles figuraient dans un dictionnaire de référence. Les frontières de phrases étaient marquées par un

⁷ W. Teubert (1993) *Phonetic/phonemic and prosodic annotation*. Final Report NERC-WP1-8&E&, Manneheim: IDS (cité par Llisterrí 1994-a:5).

point et une majuscule. On utilisait des guillemets pour les citations et le discours rapporté ; on ne faisait aucun usage de la virgule.

En plus de la transcription orthographique, on a souvent proposé de noter les divers « événements » qui accompagnent la prise de parole⁸ comme les allongements, l'accentuation ou divers autres phénomènes phoniques. On a proposé des conventions particulières pour les formes non standard, pour les abréviations, les acronymes, les mots épelés, les coupures dans les énoncés, avec ou sans pause, les amorces, les chevauchements, et quantité d'autres « événements ». On a proposé de noter les unités intonatives, les hauteurs, les intensités, le débit de parole et les pauses. Certains transcripseurs représentent les qualités de la voix, le chantonement, la voix criée, les passages lus, les passages chantés, les morceaux incompréhensibles ou simplement devinés, etc. Certains ont envisagé de noter systématiquement les éléments non verbaux : gestes, mimiques, regards, attitudes, qualité des rires, voire des toux, bruits divers. Enfin, il a souvent été demandé d'adjoindre à la transcription un descriptif des difficultés qu'on a rencontrées pour la faire. En 1992, J. P. French a publié des propositions visant à unifier ces pratiques⁹.

J. Sinclair¹⁰ signalait des abus : les néophytes, lorsqu'ils se lancent dans l'exploitation des corpus de langue parlée, ont tendance à insister sur les aspects les plus bizarres ou les plus anecdotiques. Le résultat est souvent un intérêt assez disproportionné accordé aux éléments comme les onomatopées, les claquements de langue, les raclements de gorge, les rires, toutes choses qu'on aura du mal à intégrer dans une description linguistique. Ces abus avaient peu de conséquences fâcheuses quand ils portaient sur des études limitées. Mais, à partir du moment où l'on établissait de grands corpus, subventionnés par des fonds de recherche, ces abus avaient des conséquences financières très voyantes.

Il y a actuellement de grands débats sur l'intérêt que présentent ces annotations. Certains y voient une procédure d'enrichissement et d'éclaircissement des textes. Mais G. Leech, assez pessimiste, disait, au colloque EAGLES de janvier 1996, qu'on aurait pu aussi bien parler, dans bien des cas, de « pollution » ou « corruption » des textes !

Ces interpolations peuvent être effectivement gênantes, et les chercheurs de EAGLES recommandent maintenant d'avoir toujours à disposition une version du *texte nu*, dépourvue d'annotations. Il se peut que cela crée des

⁸. J. Llisterra recommande particulièrement la publication de J.A. Edwards et M.D. Lampert (eds.), *Talking data: Transcription and coding in discourse research*. Hillsdale, NJ: Laurence Erlbaum Ass.

⁹. French, J.P. (1992) *Transcription proposals: multi-level system*. NERC-WP 4 50. Birmingham: University of Birmingham.

¹⁰. Colloque de Lisbonne, 3/10/95.

difficultés, et que certains textes semblent, en raison du manque d'annotations, peu intelligibles par endroits. Mais cela semble préférable à l'excès inverse, qui consiste à charger les textes, d'emblée, d'une grande masse d'informations. Il arrive, actuellement, que certains chercheurs dépensent une grande énergie à surcharger leurs corpus d'annotations diverses et que d'autres chercheurs en dépensent autant pour effacer cette surcharge.

Un certain nombre des ambitions de fidélité aux phénomènes sonores vont tomber en raison des avancées technologiques, par exemple la possibilité d'avoir, sur un CD-rom, des transcriptions couplées avec les enregistrements. On pourra envisager, par exemple, de vérifier les phénomènes prosodiques d'un passage transcrit orthographiquement, ou les liaisons, sans dépendre d'un système de notation étranger à la recherche en cours.

Au fur et à mesure que progressent ces technologies, le travail spécifique du transcripateur se précise. J. Sinclair demandait que les chercheurs s'entendent sur une sorte de *niveau zéro* de la transcription, faite orthographiquement avec le minimum d'indications supplémentaires¹¹. On ménage ainsi, dit-il la facilité de lecture. En fait, les conventions NERC de 1992 proposaient de distinguer quatre niveaux :

- Le premier donnerait une représentation orthographique avec un minimum de ponctuation (en 1996, Sinclair suggère qu'on omette la ponctuation). A ce niveau, aucune marque des interactions, ni même des changements de locuteurs.
- Le niveau deux introduirait les tours de parole des locuteurs et certains éléments non verbaux.
- Le niveau trois aurait des indications d'intonation et d'interactions. Les frontières d'unités tonales et syllabiques seraient indiquées. Seuls des phonéticiens pourraient le faire, et sur de bonnes qualités d'enregistrement.
- Le niveau quatre additionnerait les indications acoustiques et phonétiques, les schémas mélodiques, un tracé du fondamental (avec d'excellents enregistrements).

Il est intéressant de remarquer que, à travers toutes ces discussions et ces mises au point, jamais ne s'est posée, dans le groupe EAGLES, la question des multi-transcriptions telle que l'envisage le GARS, c'est-à-dire la possibilité de proposer plusieurs versions graphiques d'un même passage, comme « il l'apprend / il la prend » ou « elle a marché/ elle marchait ». Dans les études du GARS, l'établissement même du texte est envisagé comme un problème difficile, qui met en jeu autre chose que l'automatisme

¹¹. J. Sinclair dans les recommandations faites au Colloque de Madrid en janvier 1996.

d'une transposition terme à terme, parce qu'une part d'interprétation, phonique ou sémantique, y intervient toujours. C'est pourquoi la transcription y est considérée comme une entreprise qui engage la théorie, et que les chercheurs ne peuvent pas abandonner à des « secrétaires ». Rien de semblable dans le groupe EAGLES, qui n'a pas, pour l'instant, été préoccupé par ce problème.

2. LES ANNOTATIONS

Elles commencent avec les marques de diverses sortes que l'on porte sur le texte transcrit, afin de pouvoir les exploiter ensuite pour une analyse suivie. Tout dépend des buts poursuivis. Les chercheurs intéressés par les interactions notent les événements qui ont accompagné la prise de parole, et particulièrement les attitudes des interlocuteurs, quand il y en a plusieurs : rires, regards, bruits, gestuelles, etc. On a développé, d'autre part, diverses façons de noter l'information grammaticale, en mêlant ce travail à celui de la simple transcription. On porte sur chaque mot un étiquetage morphologique (*tagging*), qui indique essentiellement à quelle partie du discours il appartient (POS, *Part of Speech Tagging*). On indique aussi, à un autre niveau, les relations syntaxiques (*parsing*).

On peut également porter des annotations sémantiques, surtout lorsqu'on prévoit de faire une application à des dictionnaires, et des annotations discursives, comme par exemple certains réseaux d'anaphore.

Il s'agit donc d'indiquer par des signes conventionnels (« interpolations » à l'intérieur du texte), des catégories d'analyse. Dans un grand nombre de cas, ces catégorisations ont été seulement des indications morphosyntaxiques sur les parties du discours. Mais, comme le reconnaît G. Leech en 1996, l'analyse n'est pas encore très avancée :

« Morphosyntactic annotation (which has been so far carried out extensively on English, but not on other languages) is at a relatively primitive stage of development » (1994 :4).

Dans certains cas, une partie du travail d'étiquetage peut être fait automatiquement, quitte à corriger ensuite à la main les assez nombreuses erreurs. Mais il faut bien se rendre compte du travail que représente la correction de l'étiquetage pour un corpus de 100 millions de mots. L'expérience des lexicographes, qui ont l'habitude d'indexer les éléments grammaticaux de leurs inventaires, peut servir, mais elle ne suffit pas à donner toutes les indications qu'on jugerait nécessaires pour faire une

analyse syntaxique. Le principe des annotations est le suivant : on accompagne la transcription d'un étiquetage qui forme une sorte de « préanalyse ». On envisage d'appliquer cette préanalyse à un grand nombre de textes, formant des corpus importants, de l'ordre d'un million de mots. En totalisant les étiquetages sur de grandes dimensions, on pourrait obtenir des statistiques de régularités et d'irrégularités grammaticales, qui échappent à la description faite sur de petites dimensions.

Le grand problème préliminaire est de savoir quelle limite minimale on se fixe pour le nombre de ces annotations, et quelle limite maximale, pour autant qu'on puisse en envisager une. On peut imaginer de fournir, comme l'équipe du GARS, des transcriptions qui ne sont accompagnées d'aucune codification grammaticale. C'est l'attitude minimaliste qu'on adopte si l'on pense que la grammaire des productions orales est encore peu connue, qu'il y a des données à dégager et des concepts nouveaux à mettre au point. En ce cas, il serait téméraire de vouloir tout étiqueter, dans la mesure où les analyses traditionnellement acceptées risquent parfois d'être mises en défaut. Ce serait le cas pour certaines parties du discours comme « préposition, conjonction, adverbe », et pour quantité de fonctions syntaxiques, en particulier les notions majeures de coordination et subordination. Dans cette perspective, le plus urgent serait d'assurer en premier lieu les principes de l'analyse et d'envisager seulement par la suite de faire un étiquetage des textes.

Les corpus de langue parlée apportent en effet, pour des entreprises comme celles du GARS, non seulement des données nouvelles, mais un ensemble de délimitations nouvelles des données que l'on avait déjà. Il n'est donc pas toujours facile d'analyser d'emblée la nouveauté.

« Ce sont souvent les exemples difficiles à classer, ou qui posaient des problèmes au départ, qui permettent par la suite de faire de nouvelles hypothèses et d'avancer dans l'analyse » (Monique Gibier, 1987, Mémoire sur l'accord des participes passés en français parlé, Université de Provence).

On peut au contraire décider que le matériel d'analyse dont disposent les linguistes actuellement est largement suffisant, et qu'on peut adopter une analyse en parties du discours (dotée de quelques modernisations) et une description, au moins « basique » des relations syntaxiques. Cet appareil descriptif, même imparfait, permettrait déjà de passer à une étape statistique rentable.

Le groupe EAGLES, en 1994, propose d'indiquer au moins les principales parties du discours, comme Nom, Verbe, Conjonction :

1 N [nom]	2. V [verbe]	3 AJ [adjectif]
4 PD [pronom, déterminant]	5. AT [article]	6. AV [adverbe]
7. AP [adposition]	8 C [conjonction]	9. NU [numéral]
10. I [interjection]	11 U [unique, non assigné]	12. R [résiduel]
13. PU [ponctuation]		

Une partie de ces désignations sont classiques et admises par tout le monde. D'autres permettent de signaler certaines zones délicates, comme par exemple AP, qui désigne les « adpositions », « prépositions », « circumpositions » ou « postpositions ». Le classement par U. sert à désigner les éléments qui, comme la particule de négation, sont en quelque sorte « uniques en leur genre », et pour lesquels il serait inutile de chercher une assignation à une catégorie :

1. Les marqueurs d'infinitif (anglais *to*)
2. la particule négative (anglais *not, n't*)
3. le marqueur existenciel (anglais *there*, danois *der*)
4. la seconde particule de négation (français *pas*)
5. les éléments anticipateurs (néerlandais *er*)
6. le marqueur de voie médio-passive (portugais *se*)
7. les particules préverbales (grec)

R. s'applique aux éléments qu'on ne peut pas classer dans des parties du discours : mots étrangers, formules.

On peut indiquer également certains de leurs attributs : Genre, Nombre et Cas pour le Nom ; éventuellement l'appartenance à des classes sémantiques comme « noms temporels », « adverbes de manière », etc.

G. Leech énumère ainsi des ajouts optionnels, représentant certaines caractéristiques mises au point dans différents courants de la linguistique contemporaine :

- pour le Nom : comptable ou massif
- pour l'Adverbe :
 - classe générale ou adverbe de degré (*très, si, tant*)
 - type interrogatif, relatif ou exclamatif (ex : *comment*)
- pour la conjonction :
 - simple
 - corrélative
 - initiale
 - non-initiale

Ces caractéristiques reçoivent un codage conventionnel, par exemple,

dans la présentation de G. Leech (1994) :

- nom commun, féminin, pluriel, comptable :
N122010
- verbe 3^{ème} pers, sing, conjugué, indicatif, passé, actif, principal,
non réflexif :
V3011141101200
- adjectif « général », au comparatif :
AJ2000000

Il est difficile, même avec tous ces recours, d'atteindre un point ultime de l'analyse, qui serait une sorte de perfection idéale. Leech donne l'exemple du verbe anglais, pour lequel on devrait prévoir une formule maximale assez importante, qui ne couvre pourtant pas le cas de l'infinitif :

V [[-301/ 002] 111/ 000121/ 000130] 0200001

Voici un exemple de cette annotation morphosyntaxique appliquée à une série de pronoms en italien :

PQNS1	Pron. pers. comm. sing. 1	io
PD141001001		
PQNS2	Pron. pers. comm. sing. 2	tu
PD241001001		
PQMS3	Pron. pers. masc. sing. 3	egli
PD311001001		
PQFS3	Pron. pers. femm. sing. 3	ella
PD321001001		
PQNP1	Pron. pers. comm. plur. 1	noi
PD142001001		
PQNS2	Pron. pers. comm. plur. 2	voi
PD242001001		
PQNP3	Pron. pers. comm. plur. 3	loro
PD342001001		

On peut arriver ainsi à des notations plus ou moins raffinées des informations grammaticales. On voit bien que, ici, ce type de codification reproduit, à peu de choses près, les classifications les plus traditionnelles de l'analyse « grammaticale ».

3. EXPLOITATIONS

Ces techniques d'analyse se sont répandues dans quantité de domaines sans que les linguistes s'en aperçoivent, et de façon que les exploitations qu'on en fait échappent à leur contrôle. Dans certains secteurs d'activité, le travail sur la langue parlée implique même d'emblée qu'on fait un étiquetage de tous les mots, ce qui est supposé lancer sur le texte un filet d'analyse dont on pourra de toute façon tirer quelque profit. Un grand nombre de chercheurs d'autres disciplines ont commencé à exploiter les annotations comme des techniques fermement assurées, ce qui peut susciter des problèmes quand il s'agit de textes peu conformes aux normes usuelles. S'ajoute à cela un inconvénient né de la répartition des tâches. Comme il est assez fastidieux de conduire ce travail d'étiquetage morphosyntaxique sur de grandes étendues de textes, les chercheurs délèguent parfois cette tâche, présentée comme élémentaire, à des personnes peu spécialisées. On risque d'aboutir, en ce cas, à une pratique des annotations peu cohérente et peu utile.

Je prendrai un exemple de résultat négatif dans la pratique de certains spécialistes médicaux de troubles du langage (logopédistes de Liège, par exemple¹²), à partir de corpus transcrits et annotés selon des procédés qu'ils croient « scientifiques » et qui, souvent, peuvent nous horrifier, même si le monde médical en fait, somme toute, un usage modéré.

Voici un exemple de l'étiquetage utilisé par des spécialistes des troubles du langage, orthophonistes ou logopédistes (M. F. Granier 1994, *Approches quantitatives des déficits en morpho-syntaxe de patients aphasiques*, Liège, p. 23). Il s'agit de l'enregistrement d'une personne hospitalisée à la suite d'un accident. On lui demande de raconter l'accident. Le texte de la transcription est présenté en deux formats. D'une part la version orthographique, coupée en séquences qui correspondent à des sortes de paragraphes, et en unités qui ressemblent aux « propositions » de la grammaire scolaire. D'autre part un étiquetage morphosyntaxique, sur lequel sont construits des calculs statistiques. Il s'agit, dans ce genre de recherche, de compter les « anomalies », aussi bien les énoncés inachevés que les réalisations non normatives ou les « agrammaticalités ». On compte les anomalies à partir de la version annotée.

A partir du moment où on dispose de la version annotée, on peut se

12. Marie-France Granier, 1994, *Approches quantitatives des déficits en morpho-syntaxe de patients aphasiques*, Université de Liège).

permettre, semble-t-il, d'abandonner la transcription orthographique de départ. Elle est considérée comme une étape préliminaire du travail, à laquelle on peut se référer pour vérification, mais qui n'est plus essentielle. La version « étiquetée » devient la base réelle de l'observation et de la réflexion. Comme elle procède par étiquetage de petits segments, elle ne rend pas compte (du moins pas dans cet état de la technique utilisée) des relations entre les grands constituants de l'énoncé.

Séquence n° 12

C'est à ce moment-là que
proS être prep detN conj \emptyset

qu'on a pris le parti de m'emmener dans un autre hôpital
conj pros aux v deefN gen cliD v prep indef adjq N

et c'est là-bas qu'on m'a trépané très vite etc
coor proS être adv conj proS cliD aux V adv adv (expression)

Séquence n° 14 :

Pour moi c'est comme si j'avais dormi quoi,
prep pro proS être loc conj proS aux V (intj)

sauf que quand je me suis réveillé
loc conj conj proS cli (refl) aux V

je ne savais plus parler quoi
proS aux nég V_{inf} (intj)

Les annotations cumulent, sur les segments graphiques les plus petits, les indications morphologiques et syntaxiques. Dans *c'est*, (séquence 12), l'auteur analyse comme un pronom (pro) qui est sujet (S). La notation par signe vide \emptyset , (*que* = conj. \emptyset), dans :

c'est à ce moment-là que qu'on a pris le parti de m'emmener

est faite pour signaler un inachèvement. L'analyse est la suivante : le locuteur a commencé une subordonnée en *que*, qu'il n'a pas continuée. Cette conjonction qui joue à vide, puisqu'elle n'est pas immédiatement suivie de sa subordonnée, sera comptée, parmi d'autres choses, comme un des indices du « manque de langage » du patient aphasique. Tous les signes « vides » seront comptés, et disponibles pour faire des comparaisons avec d'autres patients.

On voit vite à quel émiettement mène cet étiquetage. Dans la séquence

12, *c'est à ce moment-là que qu'on a pris le parti de m'emmenner* on a un luxe de catégorisation morphologique, mais aucun lien n'a pu être marqué entre *c'est* et *que*. La tournure clivée que forme *c'est... que...*, autour de *à ce moment-là*,

c'est à ce moment-là qu'on a pris le parti de m'emmenner

ne peut pas être identifiée. La partie analytique ne rend pas compte du rôle de *c'est... que*, et de la localisation du complément temporel à *ce moment-là*. Elle traite *c'est...* comme un simple verbe, puisque c'est la donnée morphologique immédiate, sans pouvoir envisager son rôle de support du dispositif clivé.

Il est bien évident que, pour un linguiste minutieux, ce simple travail d'étiquetage doit être complété par une analyse qui tienne compte des grandes relations syntagmatiques. C'est particulièrement délicat quand il s'agit des discours d'aphasiques, dans lesquels les grandes relations sont souvent masquées par des phénomènes superficiels d'hésitations et de bribes inachevées (par exemple *c'est à ce moment-là que qu'on a pris le parti...*). Mais on comprend que, pour des chercheurs qui ne sont pas des linguistes, et qui veulent rendre compte d'un grand nombre de productions, cette approche, même insatisfaisante, leur paraisse être déjà un « bon début ».

Les dirigeants de EAGLES signalent un autre inconvénient, d'un tout autre ordre. Étiqueter un texte en donnant des indications sur les classes de mots est, pour l'essentiel, un travail fait à la main. Pour des textes de grande ampleur, cela revient très cher, au point que G. Leech demandait récemment qu'on ne finance plus ce genre de travail quand il porte sur des textes longs. Les résultats, dit-il, sont souvent assez médiocres, parce qu'on le fait faire par des personnes peu spécialisées, et qu'on paie très mal. Il vaudrait mieux, estime-t-il, qu'on consacre les fonds de recherches à améliorer les possibilités d'analyse automatique des corpus.

4. CONCLUSION

Les avancées technologiques récentes sont en train de changer considérablement le travail linguistique sur les données des langues parlées, aussi bien pour les objectifs que pour les méthodes. Les réflexions faites actuellement au niveau européen sur les transcriptions confortent les choix de transcription « pauvre » qu'avait proposés l'équipe du GARS. Les discussions sur les annotations grammaticales, étalées en pleine lumière,

partagées par des chercheurs de formations diverses, et soumises à des questions de rentabilité financière, ont tout à coup pris une autre allure. La recherche des « bonnes solutions » d'analyse morphosyntaxique en sera sans doute profondément modifiée.

Claire BLANCHE-BENVENISTE

TRANSCRIPTION ET ÉDITION DE MANUSCRITS

Quelques problèmes autour de la « représentation » des textes

La tradition linguistique considère qu'il est impossible de travailler sur la langue sans la médiation d'une représentation graphique¹, c'est-à-dire sans soumettre nécessairement les énoncés à un processus qui les transforme en un objet de savoir, qui les démonte en quelque sorte pour les reconstruire au terme de la méthode de représentation adoptée. Le mot « transcription » désigne à la fois la méthode et son résultat, ou plutôt définit le résultat (le texte transcrit) une fois qu'il a été formé par l'opération de transcription.

Toute transcription soulève bien sûr un certain nombre de problèmes d'interprétation. Mais au-delà, et si naturelle qu'elle paraisse, elle suppose une réflexion sur la construction de l'objet que l'on observe, elle engage une théorie des données, plus exactement une réflexion sur son rôle dans l'élaboration de ce que le chercheur va considérer comme un fait, construire comme tel. Il ne s'agit donc pas d'une question périphérique mais d'un enjeu qui oblige les linguistes à cerner leur objet de description.

Nous voudrions revenir sur cette question à partir d'un travail d'édition de manuscrits de la période révolutionnaire². Les textes ont été rédigés par des « semi-lettrés » provençaux, certainement bilingues, que les événements politiques ont amenés à jouer le rôle de secrétaire de séance dans des assemblées politiques, à présenter leur défense devant des tribunaux révolutionnaires ou à s'adresser pour des motifs variés aux autorités administratives. Nous avons proposé une transcription

¹ J'entends par là aussi bien les versions orthographiques que les transcriptions phonétiques ou les courbes par lesquelles les phonéticiens représentent l'intonation.

² S. Branca-Rosoff et N. Schneider (1994).

diplomatique³ des documents, respectant autant que faire se pouvait les aspects matériels, et fidèle aux écarts d'orthographe et de ponctuation. En soi, le respect « des fautes » n'est ni utile, ni inutile. Il prend sens par rapport aux analyses qu'il permet ou non de renouveler. Nous essayerons de situer les questions que notre travail a permis de se poser. Nous avons montré ailleurs que ces sources éclairaient la façon dont s'était opérée la « francisation » des populations du Sud, et qu'elles permettaient de souligner le rôle de l'écrit public dans cette transformation. Il s'agit seulement de rappeler dans ces brèves notes que le travail préliminaire des « éditeurs » a eu des effets sur les interprétations ultérieures.

1- TRANSCRIRE, C'EST PROPOSER UNE RECONSTRUCTION.

La majorité des transcrip-teurs, qu'ils s'intéressent aux corpus écrits ou aux corpus oraux, rappellent que transcrire « ce que l'on voit ou ce que l'on entend » suppose en fait tout un travail d'interprétation. Cette opération n'est d'ailleurs pas fondamentalement différente de la lecture ordinaire que l'on met en œuvre lorsqu'on reçoit une lettre et que l'on cherche à comprendre le message qu'un correspondant vous a adressé. Par exemple, pour déchiffrer des unités graphiques dont le tracé est peu visible, le lecteur fait appel à de multiples processus d'interprétation plus ou moins inconscients.

Certains graphèmes entraînent plus que d'autres des confusions, ainsi les « u », les « n », les « m », lettres à jambages qui n'ont pas de hampe. Les lettres finales posent aussi problème car elles se confondent avec des boucles ornementales fréquentes dans les écritures de l'époque. N. Schneider avait par exemple lu *prendre* avec un « r » final. Seul l'appui sur le contexte a conduit à éliminer ce « r » :

Je vous prie faire prendre
et non
Je vous prie faire prendrer

Le « r » supposé n'apparaissant plus que comme un motif décoratif. L'identification de la lettre met donc en jeu le niveau du mot ou de l'énoncé global.

³ «Transcription diplomatique» vient de diplôme. Pour les diplômes, les chartes, les paléographes établissent des transcriptions fidèles permettant d'évaluer l'âge et le degré d'authenticité des documents.

La différence de tradition graphique et la méconnaissance du contexte historique accroissent les difficultés. Il faut s'accoutumer aux formes visuelles d'époques antérieures. La différence déjà délicate à repérer en imprimé entre « s » longue et « f » est parfois indécidable et on ne sait plus si on doit lire :

ils sont bien
ou
ils font bien.

Cela amène à multiplier les lectures et à pratiquer un va-et-vient, parfois complexe, de la lettre au mot, à l'énoncé et au contexte.

Les écarts orthographiques, le décalage historique et la mauvaise connaissance du contexte compliquent encore l'interprétation. Nous avons longtemps lu :

*Je suis Eté Comme un Lezard me jéter dans les bras de ce grand General
Carteaux
(je suis été comme un lézard me jeter dans les bras de ce grand général
Carteaux)*

Nous nous étonnions de rencontrer une tournure que nous ne connaissions ni en français ni en provençal. Le contexte qui évoque la conduite irréprochable d'un soldat nous a alors conduit à « lire » un « c » (d'où *César*, graphié *Cezard*) :

*Je suis Eté Comme un Cezard me jéter dans les bras de ce grand General
Carteaux
(je suis été comme un César me jeter dans les bras de ce grand général
Carteaux)*

Nous avons conservé pourtant les deux solutions d'une part parce qu'elles permettent de réfléchir au travail du transcripateur confronté à l'orthographe déviante et aux lettres délicates à distinguer pour lesquelles il faudrait à terme envisager une typologie ; d'autre part parce que si l'image du lézard nous paraît incongrue c'est sur la base d'un savoir qui demeure incertain. La double transcription est une prudence qui s'impose lorsqu'on ignore beaucoup du contexte des documents. Nous rejoignons sur ce point la pratique d'un certain nombre d'éditions critiques actuelles et la problématique défendue par les chercheurs du GARS.

Toute transcription est donc un acte d'interprétation global qui intègre à la fois l'analyse optique des éléments graphiques et l'analyse du texte et de la situation de communication. La désambiguïsation restant parfois

partielle, il est prudent de conserver une gamme d'interprétation. Enfin, il s'agit d'un acte de lecture dont la fidélité n'est jamais totalement garantie.

2- LE PASSAGE AU TEXTE DACTYLOGRAPHIÉ : DU VISIBLE AU LISIBLE

L'édition installe nécessairement une distance entre le document matériel d'origine et ce qui en est restitué où se perdent plus ou moins de ses caractéristiques. Régler cette distance, choisir ce que l'on va faire disparaître ne va pas de soi. Si on perd toujours l'aspect et l'odeur du papier, tout le reste, le choix des caractères, de la mise en page, du format, de l'orthographe etc., dépend de l'équilibre entre la facilité de lecture et la fidélité au document que chaque chercheur établit en fonction de ses buts.

De l'ordre du continu à l'ordre du segmentable

On pourrait présenter des fac-similés mais la quasi totalité des éditeurs dactylographie les textes car l'écriture manuscrite ralentit énormément la lecture. Pourtant ce choix n'est pas simple car en renonçant à la calligraphie on peut faire disparaître des témoignages intéressants. Le tracé d'une écriture, dans sa singularité incomparable, se perd. La dactylographie efface tout ce qui renvoie au geste graphique individuel, à l'orientation particulière de l'écriture, horizontale, ou verticale, aux tracés des lettres anguleux, ronds, irréguliers... toutes choses dont la graphologie nous a pourtant appris l'importance.

Or, la restitution de l'identité visuelle de l'écriture présente aussi une dimension collective intéressante. Alors qu'aujourd'hui les scribes en marge de l'alphabétisation peinent aussi bien sur le tracé des caractères que sur l'orthographe et la gestion des énoncés, les « semi-lettrés » du dix-huitième siècle marseillais tracent des caractères impeccables, jouent des pleins, des déliés et des boucles ornementales. Ceci nous rappelle les méthodes d'apprentissage dissociant la maîtrise de la graphie et l'apprentissage des aspects textuels.

Plus préoccupant encore, la frappe efface des phénomènes qui ont un effet sémantique, par exemple, la force de l'appui sur le papier ou le changement d'échelle pour des mots importants. Ces dernières caractéristiques visuelles sont iconiques, motivées - par opposition à l'arbitraire du graphème - et elles présentent des variations continues. Au contraire, le code graphique des lettres d'imprimerie est composé de segments discrets et tout ce qui est lié à l'expression et qui est de l'ordre du

continuum ne « passe pas » dans ce système de notation qui découpe un continuum en unités séparées.

En sacrifiant ces aspects visuels, le travail sur manuscrits se rapproche à nouveau de la transcription proposée par le GARS. Les chercheurs d'Aix-en-Provence, qui ont consacré de nombreux travaux à la transcription, ont expliqué comment ils avaient élaboré un code adapté à la nature de la recherche syntaxique, conservant les ratés, lapsus, amorces de l'oral mais renonçant à rendre compte du volume de la voix et des particularités phonétiques. Ils n'intègrent pas à leurs transcriptions d'indications phonétiques. Faute de procédés efficaces et simples, les courbes intonatives ne sont pas mentionnées à l'exception de quelques fragments, confiés à des spécialistes. Dans la transposition graphique, les inflexions de la voix, le rythme de la parole, les intonations se perdent, ainsi que presque tout ce qui a trait à la prosodie.

Trois expériences de transcription

Au-delà des choix concernant l'intonation, nous revenons rapidement sur l'expérience du GARS et sur les options des médiévistes et des généticiens concernant les manuscrits. D'abord parce que nous devons beaucoup à ces recherches mais aussi pour situer par rapport à leur expérience notre choix de ne pas « normaliser » nos manuscrits en les réorthographiant et de ne pas les ponctuer en fonction d'une interprétation qui lève le maximum d'ambiguïtés.

Cette décision nous sépare évidemment de la plupart des historiens des temps modernes et de nombreux éditeurs de textes littéraires⁴ qui poussent la recherche de lisibilité à son terme et présentent des manuscrits dont l'orthographe et la ponctuation, voire la disposition en paragraphes ont été retouchées. Nous avons au contraire tenté de respecter l'orthographe et la ponctuation originales, y compris les abréviations et les majuscules.

Les philologues ont mené toute une réflexion sur le travail d'établissement des textes dont B. Cerquiglini a bien retracé l'évolution dans *L'Éloge de la variante*⁵. Au XIX^e siècle, les médiévistes qui déchiffraient les manuscrits et qui les remettaient en circulation cherchaient à reconstituer un texte original défiguré par les erreurs des copistes. Ils

⁴ La position est tout autre chez les spécialistes du Moyen Âge qui visent un public de spécialistes et peuvent donc proposer des éditions savantes. Le confort de lecture du public contemporain l'emporte à partir du XVI^e siècle. A. Tournon (Conférence au DEA de linguistique française de S. Branca-Rosoff, 1990) a montré que les éditeurs de Montaigne n'hésitaient pas à présenter le texte selon des normes de ponctuation modernes, sacrifiant le mouvement de l'argumentation à nos unités habituelles, des phrases commençant par une majuscule et terminées par un point.

⁵ ainsi qu'à S. Branca-Rosoff (1983), dans *RSFP* n° 5.

voulaient retrouver la version de « l'auteur ». À cette nostalgie pour l'origine s'est substitué l'intérêt pour les variations, chacune attribuable à un énonciateur autonome (le copiste), devenu partie prenante dans la circulation du texte. Les parties variantes ne sont pas des erreurs mais des zones de flottement intéressantes en soi. Cl. Blanche-Benveniste est revenue à plusieurs reprises sur la parenté des pratiques d'édition entre transpositeurs de l'oral et spécialistes des manuscrits⁶.

Le courant de la critique génétique s'intéresse aux manuscrits de travail des écrivains qu'ils transcrivent minutieusement. Autour de L. Hay (1982, 1990) d'A. Grésillon et de J.-L. Lebrave (1983, 1988, 1990), des chercheurs défendent l'intérêt du strict respect du manuscrit pour l'étude de la production littéraire⁷. Lorsqu'on dispose des reformulations, des modulations parfois infimes auxquelles se livrent certains écrivains comme Proust, on peut entrer dans la « fabrique »⁸ de leur texte. La critique s'oriente vers une typologie des opérations de réécriture propres à chaque écrivain (ajouts, suppression, substitutions, déplacements...). Il est intéressant de souligner l'intrication entre le choix d'une codification et l'objet de la recherche. Par exemple, puisqu'il s'agit de travailler sur l'écriture comme processus, comme activité, il faut pouvoir noter différemment ce qui a été immédiatement corrigé au fil de la plume, l'élément biffé étant suivi d'une adjonction (J.-L. Lebrave emploie le \$) et ce qui a été remplacé après coup et qui vient s'inscrire entre deux lignes (le & note la substitution après coup). Dans cet exemple de Proust, on observe un premier choix, *s'installer*, une tentative de reformulation dans le cours même de l'écriture, se placer, suivie d'un retour au point de départ. Proust au moment de la relecture, alors qu'il occupe la position du lecteur, a préféré revenir à *s'installer* :

*Cette ligne blanche au dessous de laquelle court (s'installer/se placer
\$/s'installer &) la fenêtre (J.-L. Lebrave 1990 : 147)*

L'édition est un compromis entre la nécessaire transparence de la transcription et la recherche de fidélité notamment lorsqu'il s'agit de noter les étapes de la construction des énoncés.

⁶ Comme elle l'a récemment rappelé (1994, 59), c'est d'ailleurs à J. Monfrin, directeur de l'École de Chartres, que Cl. Blanche-Benveniste avait demandé une préface au livre rédigé avec C. Jeanjean, *Le français parlé. Édition et transcription*.

⁷ à L'Institut des Textes et Manuscrits Littéraires (ITEM).

⁸ Les éditions critiques des généticiens correspondent à tout un courant littéraire dont l'objet paraît être l'écriture en acte, la littérature comme production plutôt que comme produit. F. Ponge (1970) a ainsi édité l'ensemble des brouillons qui l'ont conduit au poème *Le pré*.

Les transcrip-teurs du GARS posent aussi le problème des possibilités de la transcription lorsqu'ils justifient leur choix de l'orthographe et leur refus de la ponctuation. L'orthographe implique que l'on renonce à capter les variantes de prononciation pour ne pas bricoler au coup par coup un système qui est fait pour transcrire des morphèmes et non des variantes phonétiques⁹. Le choix est d'abord dicté par un souci de lisibilité. Mais il permet tout autant de mettre en valeur la proximité syntaxique de l'oral et de l'écrit. Enfin, les opérations de transposition orthographique rapprochent le français du petit enfant qui prononce mal certaines consonnes, le français d'un Hongrois qui aurait gardé ses intonations et celui du Parisien lettré. Le système de représentation adopté contribue fortement à produire un effet de système linguistique unique.

Cependant, comme les faits dont les chercheurs du GARS cherchent à rendre compte sont essentiellement de nature syntaxique, ils n'utilisent pas la ponctuation pour éviter de projeter sur la transcription une pré-analyse syntaxique intuitive. Ils renoncent par là même à supprimer à ce niveau les ambiguïtés des séquences que l'intonation peut parfois démarquer à l'oral et qui à l'écrit - du moins dans notre tradition graphique - seraient isolées par l'emploi des signes de séparation, parenthèses guillemets et tirets¹⁰.

Quand cette tradition graphique n'est pas respectée, le problème se pose par exemple en cas de rupture de sources énonciatives. Voici un exemple où les limites du discours rapporté sont indécidables à la lecture. Les guillemets permettraient d'isoler deux des interprétations possibles où varie ce qui est à attribuer au locuteur et ce qui est à attribuer aux jeunes (deuxième source énonciative mise en scène) :

alors qu'est-ce que c'est maintenant être de Pourrières parce que les jeunes ils disent bien je suis de Pourrières c'est rien (Corpus Britta Langhans)

alors qu'est-ce que c'est maintenant être de Pourrières parce que les jeunes ils disent : « bien je suis de Pourrières » c'est rien !

alors qu'est-ce que c'est maintenant être de Pourrières parce que les jeunes ils disent bien : « je suis de Pourrières ; c'est rien ! »

⁹ Toutes sortes de problèmes de limites entre phonologie et morphologie se posent. Restituer un segment "manquant", ce qui est légitime au niveau phonologique, risque d'être déjà une prise de position sur le fonctionnement morphologique. Transcrire /ja/ par *il y a*, c'est décider que le sujet est exprimé et ne s'entend pas pour des raisons de variation phonétique puisqu'il n'y a pas d'opposition phonique pertinente dans cette place (cf la discussion menée par D.C. Giovannoni et M.-J. Savelli dans *RSFP*, 1990).

¹⁰ cf P. Mertens (1990, 175) pour les changements de registre intonatif signalant les incises et F. Sabio ici même.

Ce qu'il faut retenir c'est donc que toute transposition graphique préconstruit un modèle de corpus qui rend intelligible certains faits mais qui en estompe d'autres et qui entretient toujours une relation complexe avec les faits dont elle cherche à rendre compte.

Le témoignage de l'orthographe : des effets de proche et de lointain

Pour notre part, nous souhaitons rassembler des documents montrant le travail d'appropriation de l'écrit auquel s'étaient livrés des Provençaux peu-lettrés à l'occasion des événements révolutionnaires. Aussi, nous rendons compte des hésitations du texte (ratures, répétitions) et des passages à la ligne ainsi que des changements de page. Nous n'effaçons pas les écarts orthographiques et nous respectons la ponctuation d'origine.

Voici un exemple de texte transcrit au plus près, en respectant notamment l'absence totale de ponctuation. À la date du 8-16 prairial an 2, (14 juin 1794), les démolisseurs de bâtiments de la Ville de Marseille demandent à pouvoir commencer leurs travaux¹¹ :

Aux Citoyens administrateur
Du distric
Citoÿen les àquereur de là demollition du raipaire
Sectionnaire de lauratoire vous esposent quil ont fait
l'aquisition le 9 floreal et qu'il non peut Commencer leurs
demollition àtendut qui y a deux Chanbrée
de vollontaire do loser Sur le dit local qui les enpaice de
commencer leur demollition et Comme les dits àquereur
ont un terme ficcé pour effetuer leur demollition il vous
prie deprendre enconsideration leurs esposets
pour que vous donné les ordre naisaisaire pour faire placer
les dits vollontaire dens dautre Chambre pour que nous
puision Commenser à demollir et remplir nos engagement
enver la nation vous observent que Cet la segonde petition
qui vous prente a marseille le
14 praireal 2me anée repbque
Charles Caillot pour [-]

11 CODE DE TRANSCRIPTION

le signe	[]	encadre les séquences raturées.
le signe	{x}	indique une rature illisible.
le signe	[-]	indique un mot ou une séquence illisibles
	[être]	indique une interprétation douteuse, ici <i>être</i>
	[etre/cten]	signale deux interprétations possibles, ici <i>etre</i> ou bien <i>eten</i>
le signe	< >	indique les ajouts sur l'interligne supérieur.
le signe	<>	signale un ajout dans la marge gauche
le signe	<">	signale un ajout dans la marge droite.
le signe	//	signale un changement de page

Des documents publiés tels quels permettent de poser, niveau par niveau, la question de leur appartenance à la tradition écrite française. Ils permettent de situer les hésitations des lettrés - contemporains des événements ou non - quant à leur statut.

On peut aborder le problème par une remarque de F. Brunot. Tentant un bilan des progrès du français après la Révolution, F. Brunot cite un extrait des archives de Saône-et-Loire qu'il commente ainsi :

Un paysan facétieux, répondant à un enquêteur de l'Empire lui disait : *Depeu la Révolutiun, je commençon de franciller esé bein* (HLF IX première partie).

La proximité des parlers de la Haute-Loire et du français rend la réplique difficile à interpréter. Le paysan est-il facétieux ou bien estime-t-il de bonne foi qu'il parle français, en interprétant - s'il en est conscient - ses écarts en termes de registre ? Faut-il placer une coupure là où le paysan lui-même ne différencie peut-être pas et à partir de quel seuil de compétence va-t-on estimer qu'il s'agit bien de français ? Lorsque la distance linguistique entre deux variétés est faible, on voit bien que certains (le paysan) peuvent y voir deux niveaux d'un unique système linguistique, tandis que d'autres (ici, l'enquêteur de l'Empire que semble appuyer F. Brunot) peuvent les constituer en langues distinctes.

Le néologisme transparent, *franciller*, n'aurait sans doute pas suffi à produire un tel effet d'extériorité, ni même le vieux stéréotype paysan *je commençons* dont on peut se demander s'il s'agit d'une première ou d'une quatrième personne. On remarque le rôle joué par la graphie incorrecte qui accentue l'effet non normatif. Le procédé n'a aucune justification dans :

je commençon

et il transforme ce qui aurait pu être interprété comme un « accent » en production étrange dans :

depeu = depuis
esé = assez
bein = bien

De fait, les morphèmes coutumiers du français sont méconnaissables. On voit à quel point les choix de graphies pèsent dans l'évaluation (français/non français) des documents. À cause du risque de dévalorisation des productions, on peut être tenté de corriger car le poids de l'orthographe est si fort, en France notamment¹², que les lecteurs « ordinaires » risquent

¹² Toute écriture est un élément essentiel dans la constitution de cet imaginaire collectif qu'est une langue. Le corse en notant, les sons /dj/ par ghj a adopté une

de ne plus pouvoir les lire comme des textes. Ils ne les recevront pas comme disant quelque chose, et s'arrêteront à la déviance de la forme. On peut chercher aussi à marquer l'appartenance des textes à la communauté linguistique française, à « montrer » qu'ils appartiennent au français national.

Mais en corrigeant les documents, on efface toute trace du travail d'appropriation de ce français sans pouvoir « récupérer » l'information. En les présentant tels quels, avec leur orthographe et leur ponctuation déviantes, on donne à voir les étapes de ce long processus. C'est l'approche qui a été préférée.

Cette approche permet également de travailler de façon intéressante sur la variation graphique et sur la syntaxe.

Une sorte de coulée graphique

Je rappellerai seulement l'intérêt d'un travail systématique sur l'orthographe et ses difficultés, renvoyant à un article écrit en collaboration avec C. Guirand et N. Schneider sur les mêmes documents (1989). Malgré la faible maîtrise de l'orthographe dont témoignent certains scripteurs, les textes permettent de s'interroger sur la variation graphique et de repérer les écarts qui s'expliquent seulement par la mauvaise maîtrise du code et ceux qui s'expliquent par des variantes phonétiques - dans ce cas, ils sont relativement fréquents et recourent des remarques des dictionnaires normatifs du XVIII^e et du XIX^e siècles. Les documents permettent aussi de s'interroger sur l'évolution de la norme graphique : les usagers sont pris entre les variantes de l'ancien système (lettres muettes, faible emploi des accents, forte présence du y...), et une incorporation plus ou moins rapide des nouvelles normes (accents, recul du y...).

orthographe très éloignée de l'italien ggio ; l'emploi d'un groupe graphique qui n'est pas utilisé en italien cherche à « individualiser » le corse (le problème du corse est bien la conquête d'un sentiment d'autonomie linguistique par rapport à l'italien, en particulier au toscan dont il est très proche). On peut adopter des graphies utilisées dans des parlers voisins pour mettre la langue sous un patronage culturel de prestige : Les débats actuels que rapporte T. Magot (1989) sur la standardisation de la graphie du gallo sont un bon exemple du caractère factice de l'écriture : les codificateurs du gallo ont minoré certaines variantes :

«Les consonnes finales sont généralement muettes, mais seront écrites car prononcées dans certains parlers (29)».

ils ont choisi d'inscrire la langue dans la famille romane en adoptant les graphies conventionnellement utilisées dans ces langues :

«Le gallo est une langue romane. L'utilisation d'un système graphique autre que celui utilisé dans toutes les langues romanes ne donnerait pas au gallo plus de spécificité. C'est pourquoi on écrira "causer" (= parler) et non "kawzer" (29)».

Des documents publiés tels quels permettent de poser, niveau par niveau, la question de leur appartenance à la tradition écrite française. Ils permettent de situer les hésitations des lettrés - contemporains des événements ou non - quant à leur statut.

On peut aborder le problème par une remarque de F. Brunot. Tentant un bilan des progrès du français après la Révolution, F. Brunot cite un extrait des archives de Saône-et-Loire qu'il commente ainsi :

Un paysan facétieux, répondant à un enquêteur de l'Empire lui disait : *Depeu la Révolutiun, je commençon de franciller esé bein* (HLF IX première partie).

La proximité des parlers de la Haute-Loire et du français rend la réplique difficile à interpréter. Le paysan est-il facétieux ou bien estime-t-il de bonne foi qu'il parle français, en interprétant - s'il en est conscient - ses écarts en termes de registre ? Faut-il placer une coupure là où le paysan lui-même ne différencierait peut-être pas et à partir de quel seuil de compétence va-t-on estimer qu'il s'agit bien de français ? Lorsque la distance linguistique entre deux variétés est faible, on voit bien que certains (le paysan) peuvent y voir deux niveaux d'un unique système linguistique, tandis que d'autres (ici, l'enquêteur de l'Empire que semble appuyer F. Brunot) peuvent les constituer en langues distinctes.

Le néologisme transparent, *franciller*, n'aurait sans doute pas suffi à produire un tel effet d'extériorité, ni même le vieux stéréotype paysan *je commençons* dont on peut se demander s'il s'agit d'une première ou d'une quatrième personne. On remarque le rôle joué par la graphie incorrecte qui accentue l'effet non normatif. Le procédé n'a aucune justification dans :

je commençon

et il transforme ce qui aurait pu être interprété comme un « accent » en production étrange dans :

depeu = depuis
esé = assez
bein = bien

De fait, les morphèmes coutumiers du français sont méconnaissables. On voit à quel point les choix de graphies pèsent dans l'évaluation (français/non français) des documents. À cause du risque de dévalorisation des productions, on peut être tenté de corriger car le poids de l'orthographe est si fort, en France notamment¹², que les lecteurs « ordinaires » risquent

¹² Toute écriture est un élément essentiel dans la constitution de cet imaginaire collectif qu'est une langue. Le corse en notant, les sons /dj/ par ghj a adopté une

de ne plus pouvoir les lire comme des textes. Ils ne les recevront pas comme disant quelque chose, et s'arrêteront à la déviance de la forme. On peut chercher aussi à marquer l'appartenance des textes à la communauté linguistique française, à « montrer » qu'ils appartiennent au français national.

Mais en corrigeant les documents, on efface toute trace du travail d'appropriation de ce français sans pouvoir « récupérer » l'information. En les présentant tels quels, avec leur orthographe et leur ponctuation déviantes, on donne à voir les étapes de ce long processus. C'est l'approche qui a été préférée.

Cette approche permet également de travailler de façon intéressante sur la variation graphique et sur la syntaxe.

Une sorte de coulée graphique

Je rappellerai seulement l'intérêt d'un travail systématique sur l'orthographe et ses difficultés, renvoyant à un article écrit en collaboration avec C. Guirand et N. Schneider sur les mêmes documents (1989). Malgré la faible maîtrise de l'orthographe dont témoignent certains scripteurs, les textes permettent de s'interroger sur la variation graphique et de repérer les écarts qui s'expliquent seulement par la mauvaise maîtrise du code et ceux qui s'expliquent par des variantes phonétiques - dans ce cas, ils sont relativement fréquents et recourent des remarques des dictionnaires normatifs du XVIII^e et du XIX^e siècles. Les documents permettent aussi de s'interroger sur l'évolution de la norme graphique : les usagers sont pris entre les variantes de l'ancien système (lettres muettes, faible emploi des accents, forte présence du y...), et une incorporation plus ou moins rapide des nouvelles normes (accents, recul du y...).

orthographe très éloignée de l'italien ggio ; l'emploi d'un groupe graphique qui n'est pas utilisé en italien cherche à « individualiser » le corse (le problème du corse est bien la conquête d'un sentiment d'autonomie linguistique par rapport à l'italien, en particulier au toscan dont il est très proche). On peut adopter des graphies utilisées dans des parlers voisins pour mettre la langue sous un patronage culturel de prestige : Les débats actuels que rapporte T. Magot (1989) sur la standardisation de la graphie du gallo sont un bon exemple du caractère factice de l'écriture : les codificateurs du gallo ont minoré certaines variantes :

«Les consonnes finales sont généralement muettes, mais seront écrites car prononcées dans certains parlers (29)».

ils ont choisi d'inscrire la langue dans la famille romane en adoptant les graphies conventionnellement utilisées dans ces langues :

«Le gallo est une langue romane. L'utilisation d'un système graphique autre que celui utilisé dans toutes les langues romanes ne donnerait pas au gallo plus de spécificité. C'est pourquoi on écrira "causer" (= parler) et non "kawzer" (29)».

La transcription au plus près est également intéressante pour la réflexion sur les unités d'écriture. On s'en rend compte en procédant pas à pas à un travail de normalisation. Une interprétation orthographique en morphèmes, à la façon du GARS, lève un bon nombre de difficultés et permet d'apprécier la compétence syntaxique et lexicale des scribes et leur bonne connaissance des formulaires :

Aux Citoyens administrateurs Du district,

Citoyens, les acquéreurs de la démolition du repère Sectionnaire de l'oratoire vous exposent qu'ils [en] ont fait l'acquisition le 9 floréal et qu'ils n'ont pu Commencer leur démolition attendu qu'il y a deux Chambrées de volontaires de logées Sur le dit local qui les empêchent de commencer leur démolition et Comme les dits acquéreurs ont un terme fixé pour effectuer leur démolition, ils vous prient de prendre en considération leur exposé pour que vous donniez les ordres nécessaires pour faire placer les dits volontaires dans d'autre Chambres pour que nous puissions Commencer à démolir et remplir nos engagements envers la nation (ils) vous observent que C'est la seconde pétition qu'ils vous présentent.

Mais si l'on n'établit pas de ponctuation, le texte reste difficile d'accès. L'absence de points et de virgules, l'emploi fréquent des C majuscules comme initiales de mots et non comme bornes syntaxiques donnent l'impression d'un continuum graphique ininterrompu. Le corps de la lettre enchaîne même par dessus les changements énonciatifs - classiques dans ces textes administratifs¹³. Les paragraphes existent à peine : les scribes se sont appropriés seulement les deux bornes de l'adresse et de la signature, les moments les plus ritualisés de l'ouverture et de la clôture de la pétition.

Ces indices négatifs recourent d'autres indices de continuité qui montrent que les scribes n'écrivent pas avec les unités modèles de l'écrit, la phrase et le paragraphe. En premier lieu, le texte progresse par reprises et donne l'impression de piétiner autour de l'information centrale. Dans le premier groupe (a), le même lexique (verbe recteur signifiant un processus et valence lexicale) apparaît au début précédé du modal puis en fin d'énoncé dans une relative. L'élément négatif est une fois portée par la modalité négative *ne*, et une fois par le sens lexical du verbe *empêcher de* :

a - ils n'ont pu **commencer leur démolition** attendu qu'il y a deux chambrées de volontaires de logées sur le dit local qui les empêchent de **commencer leur démolition**

¹³ Au début, les signataires se désignent de façon juridique à la troisième personne, puis la lettre bascule quand ils se font plus pressants vers la fin (pour que nous puissions commencer à démolir).

Dans le deuxième groupe, on trouve un changement de verbe *effectuer*, puis un retour au modal *commencer*, mais avec un changement de classe de mots *démolition*, passant à *démolir*.

b - Et comme les dits acquéreurs ont un terme fixé pour **effectuer leur démolition** [] pour que nous puissions **commencer à démolir** et remplir nos engagements envers la nation

La deuxième caractéristique est que à partir du noyau *ils vous prient*, le texte procède par accrochage successifs de propositions de plus en plus éloignées du point de départ, sans coupure linguistique, sans dégroupement. L'absence de hiérarchisation des relations est d'autant plus frappante que c'est le même marqueur de relation *pour* qui est employé.

Et comme les dits acquéreurs ont un terme fixé **pour** effectuer leur démolition, ils vous prient *de prendre en considération* leur exposé **pour** que vous donniez les ordres nécessaires **pour** faire placer les dits volontaires dans d'autre chambres **pour** que nous puissions commencer à démolir et remplir nos engagements envers la nation

Le premier *pour* dépend du substantif *terme*. Dans la construction suivante, *pour (que)* introduit des groupes situés à trois niveaux différents de la hiérarchie syntaxique. De plus, il rattache *que vous donniez des ordres à prendre en considération leur exposé*, alors que la liaison causale paraîtrait moins obscure s'il s'agissait de *prier*. Tout se passe comme si ce qui importait était l'effet d'enchaînement plus que la signification propre de *pour que*. Le même *pour* introduit les deux derniers segments ce qui semble les mettre sur un même plan. Ce n'est donc pas le manque de marqueurs qui caractérise ce type de textes à la fois non-standard et non-familiers - mais leur accumulation sans hiérarchisation.

L'absence de ponctuation est solidaire d'un modèle syntaxique non ramenable à la phrase. Inversement, la tradition grammaticale lettrée a établi parallèlement l'emploi des découpages visuels, le raccourcissement de la traîne des subordonnées et l'interdiction d'utiliser le même marqueur de relations dans des emplois syntaxiques différents. Chaque segment doit avoir un maximum d'autonomie et se situer à un niveau bien repérable de la structure hiérarchique. L'emploi de l'alinéa, le découpage en « phrases » et la répartition des marqueurs de relation selon les fonctions sont trois façons de ne pas confondre les niveaux d'enchâssement. Ces conventions qui facilitent la lecture, sont bien décrites au XVIIIe. Selon Condillac, l'alinéa qui permet de découper une pensée joue un rôle important :

[Racine cherche à développer une pensée *le théâtre doit beaucoup à Corneille*, il ne peut la développer] qu'autant qu'il en aperçoit toutes les parties. [...] Ainsi, sa pensée se décompose en trois parties qu'il distingue en les séparant par trois alinéas.

Vous voyez par là que dans le discours écrit, les alinéas contribuent à distinguer, d'une manière plus sensible, les différentes parties d'une pensée. Il marque où chacune finit, où chacune commence, et, par cet artifice, elles se démêlent beaucoup mieux (Condillac, 1775, rééd 1947, 448-449).

Condillac examine ensuite la phrase. Les règles qu'il donne dans la *Grammaire* et plus encore dans *L'Art d'écrire* concerne l'importance du découpage en unités séparées¹⁴ et l'interdiction d'utiliser des phrases qui ne s'organisent pas nettement autour de la principale. Le privilège donné à la principale interdit en particulier *de construire une suite de propositions subordonnées les unes aux autres* comme dans :

Le Corrége étoit si rempli de ce qu'il entendait dire de Raphaël, qu'il s'étoit imaginé qu'il falloit que l'artisan qui faisoit une si grande fortune dans le monde, fût d'un mérite bien supérieur [...]

Ce n'est pas parce que les *que* sont répétés que nous sommes choqués de ces constructions [... c'est] parce que la même conjonction sert à marquer des subordinations différentes [...] Le fil des idées échappe, quand on subordonne trois ou quatre propositions successivement les unes aux autres (Condillac, 1775, rééd 1947, *De l'Art d'écrire* I. VII, 531).

Pour résumer ce qu'on peut attendre d'un travail d'édition de données-écrites-non-normées, on peut donc répondre que de telles données permettent de mieux situer par contraste les enjeux des techniques de normalisation du discours écrit, notamment pour ce qui est du niveau macro-syntaxique. La description patiente des usages en matière d'unités de textes permettra d'envisager une histoire qui ne soit pas seulement celle des discours normatifs des grammairiens ou des pratiques des élites. Pour cela, il nous faut disposer de corpus transcrits au plus près.

Sonia BRANCA

Bibliographie

ARABYAN Marc, 1994, *Le paragraphe narratif*, Paris, l'Harmattan.

¹⁴ B. Lamy dans sa rhétorique y insistait déjà : «C'est le fin de l'art de sçavoir couper le sens à propos» (cité dans J.-P. Seguin, 1993, 208).

- BLANCHE-BENVENISTE Claire, ET JEANJEAN Colette, 1986, *Le français parlé Édition et transcription*, Paris, Selaf.
- BLANCHE-BENVENISTE Claire, 1994, « Aspectos preliminares a una reflexion sobre la interpretacion », *Substratum*, vol II, 4.
- BRANCA-ROSOFF Sonia, 1983, « Les débats sur la variation au milieu du XIX^e siècle : les débuts de l'école de philologie romane », *RSFP*, n°5, Aix-en-Provence, Publications de l'Université de Provence.
- BRANCA-ROSOFF Sonia, GUIRAND Christine et SCHNEIDER Nathalie, 1989, « Les Archives révolutionnaires de Marseille ; état d'une recherche sur l'orthographe des textes », *RSFP* n°9, Aix-en-Provence, Publications de l'Université de Provence.
- BRANCA-ROSOFF Sonia et SCHNEIDER Nathalie, 1994, *L'Écriture des citoyens. Une analyse linguistique de l'écriture des peu-lettrés pendant la période révolutionnaire*, Publications de l'INALF, collection Saint-Cloud, Paris, Klincksieck.
- BRUNOT Ferdinand, 1939, *Histoire de la langue française, t. IX*, Paris, A. Colin (dans l'édition 1968).
- CONDILLAC Étienne Bonnot (abbé de) 1775, *Traité de l'Art d'écrire, Cours d'Instruction du Prince de Parme*, éd. G. Le Roy, Œuvres philosophiques de Condillac, 1947, Paris, PUF.
- FARGE Arlette, 1989, *Le Goût de l'archive*, La librairie du XX^e siècle, Paris, Le Seuil.
- GRÉSILLON Almuth, 1983, « Encore du temps perdu, déjà le texte de la Recherche », *Langages* 69.
- GRÉSILLON Almuth, mars 1988, « Les manuscrits littéraires : le texte dans tous ses états », *Pratiques* n° 57.
- GRÉSILLON Almuth, LEBRAVE Jean-Louis, VIOLLET Catherine, 1990, *Proust à la lettre. Les intermittences de l'écriture*, Dulerot éd. Tisson Charente.
- HAY Louis, 1982, *La genèse du texte*, éd. 1987, Paris CNRS.
- HAY Louis ed., 1979, *La critique génétique : origines et perspectives. Essais de critique génétique*, Paris, Flammarion.
- LEBRAVE Jean-Louis, 1983, « Lecture et analyse des brouillons », *Langage* 69.
- MAGOT Thierry, 1989, « Écrire le gallo », dans (vol. collectif), *La variation dans la langue en France du XVI^e au XIX^e siècle*, Paris, Éditions du CNRS.
- MERTENS Pet, 1990, « Intonation » dans Cl. Blanche-Benveniste, *Le français parlé. Études grammaticales*, Paris, Éditions du CNRS.
- PONGE Francis, 1970, *La fabrique du pré*, Paris, Skira.

SEGUIN Jean-Pierre, 1993, *L'invention de la phrase au XVIII^e siècle*, Paris, Louvain, Bibliothèque de l'Information grammaticale.

SEIJIDO Magali, *Histoire des remarques de F. Brunot sur le français parlé de 1789 à 1815 à partir du t X de l'ILF*, Maîtrise de lettres modernes sous la direction de Cl. Blanche-Benveniste.

DONNÉES ERRONÉES : QUELLES ERREURS COMMETTENT LES TRANSCRIPTEURS ?

Parmi les nombreuses difficultés que soulève la collecte de données orales (au travers de corpus longs), je voudrais ici faire état de distorsions assez récurrentes auxquelles se livrent les transcripateurs débutants. Le problème a déjà été évoqué dans Blanche-Benveniste et Jeanjean (1987), je tenterai d'en fournir ici une présentation plus détaillée. Je me suis servi d'une vingtaine de corpus de 15 minutes transcrits à l'université de Poitiers¹. J'ai relevé de façon systématique les séquences proposées par les étudiants et les versions finales qu'après écoutes j'ai retenues². Les écarts relevés reposent sur trois procédés : l'oubli, l'ajout ou la modification d'une séquence. A travers des exemples pris dans divers domaines, je montrerai que le hasard a peu à voir avec ces erreurs - qui peuvent être plus ou moins lourdes de conséquences pour les descriptions envisagées.

1. MODES DE PRODUCTION

Les erreurs de transcription peuvent tenir à des causes diverses. La méconnaissance des modes de production de l'oral conduit fréquemment le transcripateur à occulter certains phénomènes tels que les bribes et les

1 Deux personnes travaillent conjointement, ce qui permet d'obtenir « d'emblée » des transcriptions de bonne qualité, déjà soumises à la confrontation de plusieurs écoutes. La recherche des variantes entreprise pour cet article n'en est que plus pertinente. La mention « le transcripateur » désignera donc, le plus souvent, un couple d'individus.

2 Pour ce travail, seuls ont été retenus les exemples qui ne comportent pas de multi version.

amorces. Le contexte phonétique est souvent propice à une mauvaise perception des bribes surtout lorsqu'elles portent sur des éléments monosyllabiques. Les cas d'oubli sont assez répandus. Ainsi, le débutant transcrit une occurrence là où en fait deux items sont prononcés³ :

il : écrit -> il *il* écrit (St Gelais)
 donc sous l'occupation -> donc *sous le* sous l'occupation (Notre-Dame 2)
 la même personne de la société -> la même personne *de* de la société (Vayres)
 et euh : c'est le roi -> et euh *le* c'est le roi (Aliénor)

La difficulté ne tient pas seulement au nombre réduit de syllabes en cause. Des bribes de plus grande importance peuvent aussi être oubliées :

la chute du revêtement -> la chute *du revête* - du revêtement (Notre-Dame 1)

L'erreur peut venir d'une modification des bribes. Dans cet exemple, les éléments répétés sont mal recensés⁴ :

et les les tranches -> et *des* les tranches (Notre-Dame 1)

Le nombre d'éléments répétés est souvent mal identifié :

des ang- des anguilles -> des ang- des ang- des anguilles (Anguilles)

La transcription proposée initialement tend toujours à restituer un schéma plus habituel, plus proche des formats de l'écrit. On a ici une seule tendance qui va vers la « simplification » des données. Ainsi, on a relevé un seul cas où un phénomène de bribe avait été fautivement inséré dans la transcription :

margarine *de* de : bas de gamme -> margarine de : bas de gamme (Couscous)

³ Dans les exemples on indiquera toujours en premier la version proposée par le transcripateur débutant puis, après une flèche, la version finale retenue. Le titre du corpus est fourni entre parenthèses. Les séquences divergentes seront signalées en italiques.

⁴ Les déterminants sont la source d'erreurs multiples que l'on ne traitera pas ici :
on avait des grandes rigoles -> *on avait de grandes rigoles* (Notre-Dame 1)
devant le petit côté -> *devant un petit côté* (Notre-Dame 2)
la migration des papillons -> *la migration de papillons* (Futuroscope-3)
toucher les subventions -> *toucher des subventions* (Vayres)

Les enjeux pour l'analyse semblent ici bien plus importants.

Les divers écarts relevés dans cette partie présentent finalement une importance réduite. L'oubli d'une bribe ou d'une amorce a un retentissement assez mineur sur l'organisation morphosyntaxique de la langue. Les phénomènes qui relèvent de l'approximation lexicale sont assez souvent malmenés lors de la transcription. A plusieurs reprises ont été omises des séquences telles que élément démarcatif (*là, enfin*) introductif (*alors*) ou verbe recteur faible (*je trouve que*) :

c'est par exemple euh notre remplaçant -> c'est par exemple euh notre remplaçant *là* (École)
 un peu trop chauffée à l'avance -> un peu trop chauffée à l'avance *là* (Truite)
 pas mal de de bouillon de de liquide -> pas mal de de bouillon *enfin* de de liquide (Paëlla)
 ce mur ici -> *alors* ce mur ici (Notre-Dame 2)
 on prend on prend -> on prend *je trouve qu'*on prend (Paëlla)

Dans ce cas, on rencontre à nouveau des oublis mais aucun d'ajout. Les écarts entre la version primitive et la version finale ne se font donc pas de façon anarchique. Sur les trois procédés envisagés (oubli, ajout, changement), le premier est quasi exclusivement représenté. Cette singulière représentation laisse penser que les erreurs se plient à certaines règles et que leur nature dépend du secteur auquel elles se rattachent. L'observation d'autres types d'erreurs permettra amplement de le vérifier.

2. UNITÉS LEXICALES

Les erreurs recensées ne relèvent pas uniquement des phénomènes de production de l'oral. On rencontre aussi des fautes portant sur des éléments lexicaux. On réserve pour la partie suivante celles qui ont un retentissement syntaxique.

Les transcriptions peuvent présenter, en quelques endroits, un certain flottement lexical à cause duquel certaines séquences lexicales sont imparfaitement restituées. On trouve des exemples d'oublis :

et les portails étaient -> et les portails *latéraux* étaient (Nabu)
 il faut une crème -> il faut *faire* une crème (Tarte)

Les exemples avec changement des unités lexicales sont les plus fréquents. Des variations plus ou moins plausibles sont en quelque sorte « explorées » dans le contexte. Ces cas de « reconstruction » peuvent être favorisés par une proximité phonétique des séquences en cause :



vous *avez* Adam et Eve -> vous *voyez* Adam et Eve (Nabu)
 dans une cérémonie *bien peu* faste -> dans une cérémonie *pleine de* faste
 (St Gelais)
 les *fondations* ont disparu -> les *fortifications* ont disparu (Notre-Dame 2)
 plein de *choses* -> plein de *jeux* (École)

Les contextes avec des éléments symétriques peuvent faciliter ce relâchement de l'attention et provoquer une permutation des éléments :

quelques *années* après allait naître une^e seconde fille donc les deux époux
 semblent^o réconciliés mais euh quelques *temps* après + -> quelques *temps*
 [...] quelques *années* (Aliénor)

La permutation est d'ailleurs souvent source de transcription erronée :

nous irons *plus près tout à l'heure* -> nous irons *tout à l'heure de plus près*
 (Notre-Dame 2)

Des écarts plus importants se rencontrent aussi et les variations constatées ne manquent pas alors d'être, après coup, surprenantes. On est dans un phénomène de « mirage auditif », qui ne tient pas seulement à l'homophonie ou à l'homonymie :

de *faire* une visite XXX -> d'*assurer* une visite de mariage (Vayres)
 à quoi ça ressemble là le *bidule* -> à quoi ça ressemble là le *milieu* (Nabu)
 voilà oui chacun *mange* ce qu'il veut -> voilà oui chacun *rajoute* ce qu'il
 veut (Paëlla)
 vous incorporez + au *feu* -> vous incorporez au *fouet* (La Périgourdine)

Un certain nombre d'erreurs proviennent aussi d'une méconnaissance culturelle. Lorsqu'elle touche le locuteur, on court le risque d'une rectification involontaire par le transcripteur. Inversement, si elle tient au transcripteur, celui-ci construit généralement une séquence à partir d'éléments phoniques perçus. Seuls ces derniers cas nous intéresseront ici :

les très riches *œuvres* du Duc de Berry -> les très riches *heures* du Duc de
 Berry (Notre-Dame 2)
 enlever les *sucres* qui /Ø, y/ sont collés -> enlever les *sucs* qui /Ø, y/ sont
 collés (Estouffade)

Des problèmes orthographiques ou de découpage des unités peuvent s'ajouter. Ils traduisent le poids de l'écrit et permettent de mesurer le jeu complexe entre forme et sens auquel est soumis le transcripteur. Identifiant une forme ou des éléments sonores, il en livre une version écrite « connue » mais inadaptée, de fait, au contexte :

vous avez eu la chance de ne pas avoir *violé* le Duc -> vous avez eu la chance de ne pas avoir *Viollet-le-Duc* (Notre-Dame 2)
 le bordelais Bordeaux est *talonneur* entre le vin /des, de/ Graves et le Lillet
 -> le bordelais Bordeaux est = *à l'honneur* entre le vin /des, de/ Graves et le Lillet (Truite)

Ces erreurs en disent long sur les diverses ambiguïtés potentielles qui ne sont pas nécessairement levées dans les échanges de la vie courante (Fuchs : 1996). Les erreurs de transcription qui touchent le lexique relèvent pour l'essentiel d'une modification des éléments produits. L'oubli, fortement représenté lorsque les difficultés provenaient des modes de production, occupe ici une place marginale. Le travail du transcripateur opère ici au niveau du syntagme. A partir d'un moule phonique, il reconstitue une séquence qui puisse s'insérer sur le contexte. C'est pourquoi il propose des « équivalents » acceptables à partir de certains indices formels ou sémantiques.

3. PHÉNOMÈNES MORPHOSYNTAXIQUES

Dans cette partie seront évoquées quelques cas d'erreurs qui peuvent avoir un retentissement dans la description morphosyntaxique du français. Nous avons, pour l'essentiel, retenu des phénomènes dont la fréquence est peu élevée à l'oral ce qui rend plus grave toute erreur sur leur non repérage lors de la transcription. On comprend mieux ainsi l'importance non seulement du travail sur corpus mais tout l'aspect délicat et minutieux qui doit s'attacher à l'établissement des données. C'est souvent, comme on le verra par la suite, une meilleure connaissance des phénomènes, grâce à leur description, qui permet de renforcer la vigilance.

Les erreurs de transcription relevées sont souvent les conséquences visibles des représentations ou des préjugés concernant l'oral. Nous traiterons des points suivants : la négation, le sujet et le complément clitiques, les relations syntaxiques.

3.1. LA NÉGATION

La plupart des locuteurs français utilisent avec parcimonie la forme « ne » à l'oral. Mais le poids de la norme écrite reste important. C'est peut-être pourquoi dans les erreurs des transcripateurs on note plus d'ajouts que d'oublis. Un seul cas d'omission erronée a été relevé :

on trouve pas -> on *ne* trouve pas (Futuroscope -2)

Les autres cas relevés portent sur le rétablissement fautif de la négation complète, plus conforme aux habitudes de l'écrit :

on *ne* va pas discuter -> on va pas discuter (St Gelais)
 ça *n'*a pas l'air -> ça a pas l'air (St Pierre)
 il *n'*y a pas eu -> il y a pas eu (St Pierre)
 la porte *n'*est pas d'origine -> la porte */était, est peut-être/* pas d'origine (Notre-Dame 2)

Un autre phénomène intervient avec le « ne » de négation qui peut être confondu avec un autre élément clitique :

on *ne* sait pas du tout -> on *le* sait pas du tout (St Pierre)
 vous mettez pas du sel -> *ne* mettez pas du sel (Truite)

Ces fluctuations mettent en évidence à la fois la fragilité des données et l'extrême prudence qui doit accompagner les travaux de comptage, lorsqu'ils portent sur des phénomènes peu représentés. La réalisation de la négation pouvait facilement être mal évalué, si la transcription initiale avait été conservée en l'état.

3.2. LA RÉALISATION DU PRONOM SUJET

Pour des phénomènes rares à l'oral, les conséquences de telles erreurs de transcription peuvent devenir véritablement déterminantes et fausser en grande partie la pertinence des analyses. De tels problèmes sont favorisés par les éléments en cause - souvent des clitiques - difficilement perçus dans certains contextes. Mais la méconnaissance de certains pans de la grammaire du français favorise ce « relâchement » du transcritteur débutant. L'écart est d'autant plus difficile à identifier qu'aucune description détaillée ne peut servir de garde-fou. Le sujet offre deux contextes pertinents pour cette démonstration.

3.2.1. Verbes impersonnels

Dans le cas de « falloir », la présence ou l'absence du clitique « il » est bien souvent sujette à variation. Les erreurs de transcription reflètent cet usage flottant :

la aussi faudra -> là aussi *il* faudra (Nabu)

Pour cet exemple, le contexte phonétique prête à confusion. Reste que si détecter la réalisation du clitique pose quelquefois problème, c'est essentiellement dans des circonstances où il peut ne pas apparaître. Les transpositeurs proposent rarement des formes verbales (autres qu'un impersonnel) sans sujet. Le cas ci-dessous, seul de son espèce, apparaît au milieu de constructions inachevées, ce qui a pu dérouter le transpositeur. La séquence proposée « tout le monde s'appelle madame Morel » si elle frise l'absurde (dans le corpus du moins...) constitue une suite grammaticalement correcte :

il y en a une tout le monde^o s'appelle madame Morel -> il y en a une tout le monde^o *elle* s'appelle madame Morel (École)

Les deux versions peuvent être disposées de façon à mieux faire ressortir leur différence syntaxique :

tout le monde ^o	s'appelle	madame Morel	
tout le monde ^o			<construction inachevée>
<i>elle</i>	s'appelle	madame Morel	

3.2.2. Constructions enchaînées

Si l'on prend le cas des constructions enchaînées (juxtaposées ou coordonnées), le sujet « commun » n'est pas forcément présent devant chaque verbe à l'écrit (Blanche-Benveniste : 1995, Cappeau : 1996) :

Il fouillait, perquisitionnait, saisissait, recoupait, auditionnait. (Le Monde. 22-06-96)

Cette conformation est peu répandue à l'oral. Sur l'ensemble des corpus examinés pour ce travail, seuls 5 exemples ont pu être trouvés. On voit donc combien il importe de disposer d'une transcription fiable, d'autant plus fragile qu'elle ne repose bien souvent que sur la perception d'une forme clitique parfois réduite au phonème /i/. Les habitudes acquises au travers de l'écrit viennent ici perturber l'attention des transpositeurs et leur faire négliger cette écoute fine :

ils vont aller à Paris et vont aller être consacrés -> ils vont aller à Paris et *ils* vont aller être consacrés (Aliénor)

A nouveau, les transpositeurs « trébuchent » de façon sélective sur des réalisations clitics. La forme « il » n'est pas perçue aux endroits où sa non apparition est envisageable. On sait que ces écarts sont généralement

sélectifs : les formes valorisées sont plus fréquemment attribuées aux locuteurs qui ont l'air d'avoir du prestige.

Le sujet se prêterait à d'autres observations en particulier sur la concurrence des formes *l'on / on* et *ça / ce*. On retrouverait alors certaines des observations formulées pour la négation : la forme valorisée donne lieu à plus de perceptions erronées et les transcripteurs en proposent généralement un nombre trop élevé.

Par son caractère quasi obligatoire en français, les variations sur le sujet se limitent finalement à sa forme et à son « oubli » dans quelques contextes particuliers. Les erreurs relevées sont donc une nouvelle fois guidées, influencées par la structure de la langue.

3.3. LA RÉALISATION DES COMPLÉMENTS

Si pour le sujet les « oublis » étaient étroitement canalisés, pour le complément une plus grande disparité de situations existe : il suffit que la solution soit conforme à la grammaire. On trouve ainsi des exemples d'oubli d'un complément de forme clitique (monosyllabique et non accentué). La séquence incomplètement transcrite ne présente toutefois pas d'anomalie ou de bizarrerie :

on vient à la nécessité de la lumière -> on *en* vient à la nécessité de la lumière (Futuroscope -3)
 autrefois on pêchait -> autrefois on *en* pêchait (Anguilles)
 au moins il ne pleut plus -> au moins il n'y pleut plus (St Gelais)
 et mettre tremper ensuite euh dans un mélange -> et *le* mettre tremper ensuite euh dans un mélange (Anguilles)

L'ajout d'un complément, potentiellement disponible pour le verbe en cause, est aussi représenté :

écoutez ce que je vais *lui* dire -> écoutez ce que je vais dire (Vayres)
j'apprends aussi des *trucs* -> *ils nous apprennent* aussi des *choses* (Vayres)
 comment on *les* pêche -> comment on pêche (Anguilles)

Pour le complément, les erreurs présentent une variété plus grande qui correspond aussi à des schémas de construction plus nombreux. La même « liberté » se rencontrera pour les relations syntaxiques.

3. 4. LES RELATIONS SYNTAXIQUES

Les exemples recueillis sont trop peu nombreux pour permettre un examen approfondi de diverses relations syntaxiques. Mais il m'a semblé intéressant de montrer que les indications soulignées à maintes reprises dans ce travail étaient pleinement corroborées. Quelle que soit l'erreur commise, la solution proposée par l'étudiant est « grammaticalement correcte ». Les ajouts et les oublis qui, en général, changent la relation syntaxique dépendent étroitement du contexte dans lequel ils apparaissent. Ainsi, le transcripteur peut traiter comme complément ce qui apparaît comme une construction isolée :

ça dépend *de* chacun -> ça dépend *c'est* chacun (Paëlla)

Il peut aussi modifier le type de lien entre deux constructions (rection faible vs coordination) :

j'allais vous dire *que* vous ajoutez les petites noix -> j'allais vous dire *et* vous ajoutez les petites noix (Truite)

ou entre deux séquences (épithète vs relative) :

des gens disponibles -> des gens *qui sont* disponibles (Futuroscope -1)

L'indication d'un lien de rection (même faible) au travers de la présence d'une conjonction peut être proposée :

je pense *que* maintenant il est temps -> je pense maintenant il est temps (L'enjeu)

C'est encore la forme de la conjonction qui peut être atteinte :

c'est-à-dire *bien* que on met à chaque personne -> c'est-à-dire que on met à chaque personne (Estouffade)
pour qu'elles roussissent un peu ouais *pour* qu'elles brunissent on rajoute -> pour qu'elles roussissent un peu ouais qu'elles brunissent on rajoute (Fricassée)

C'est enfin l'élément introducteur qui peut varier (verbe vs conjonction) :

un peu de moutarde *il faut* quand même que la sauce soit un peu épaisse -> un peu de moutarde *pour* quand même que la sauce soit un peu épaisse (Paëlla)
avec un fouet *parce qu'il faut* que ça fasse un ruban -> avec un fouet *jusqu'à ce que* ça fasse un ruban (La Périgourdine)

Dans chacun de ces cas, le transcripateur retrouve ses repères. Sa version plus ou moins éloignée du résultat final, peut provenir d'oublis, d'ajouts, de changements, elle n'est jamais éloignée des structures du français. La « monstruosité » n'a pas sa place dans les erreurs que commettent les étudiants.

4. CONCLUSION

Le relevé des corrections apportées aux travaux de transcripateurs novices a permis de vérifier à quel point le locuteur est impliqué dans le travail de transcription et de quelle façon il tend à reconstruire les données. Il occulte souvent ce qu'il ignore ou maîtrise imparfaitement (les modes de production de l'oral), substitue le lexique dont il dispose, et adapte la syntaxe à certaines de ses habitudes. Par là, ce travail de correction permet de découvrir certaines zones originales de l'oral (celle où la syntaxe diverge de l'écrit par sa fréquence du moins). Les données réelles attestées ne le deviennent véritablement qu'après de multiples vérifications qui nous affranchissent de ce que nous sommes toujours prêts à entendre.

Paul CAPPEAU

Références

- BLANCHE-BENVENISTE, Claire et JEANJEAN Colette, (1987), *Le français parlé. Transcription et édition*, Paris, Didier Érudition.
- BLANCHE-BENVENISTE Claire, (1995), « Répéter ou ne pas répéter », Mélanges offerts à David Gaatone, Tel-Aviv.
- CAPPEAU Paul (1996), « Quand le sujet n'apparaît pas. Petite étude sur quelques cas », Strasbourg, *Scolia*.
- FUCHS Catherine (1996), *Les ambiguïtés du français*, Paris, Ophrys.

**A PROPOS DE
*QU'EST-CE QUE C'EST ET C'EST QUOI***

Les études sur le français parlé ont-elles fait progresser la connaissance grammaticale de la langue ? Il y a quelques années, nous apportions à cette question une réponse un peu pessimiste (Blanche-Benveniste et Jeanjean 1987 :86-90). Il nous semblait que la grammaire du français parlé avait finalement été assez peu exploitée, sinon pour venir compléter des analyses déjà établies par ailleurs. Près de 10 ans plus tard, il n'est pas sûr que les choses aient vraiment changé.

Pour beaucoup d'observateurs, tout se passe encore comme si la grammaire, en gros, avait été suffisamment décrite. Le français parlé n'apporterait rien de fondamental dans ce domaine, mais plutôt des connaissances sur des secteurs annexes : précisions sur le jeu des normes et des usages, conjectures sur les évolutions prévisibles, activités et interactions des locuteurs, ou modèles d'échanges sociolinguistiques. Pas question d'apporter du nouveau sur les subordinations ou sur les interrogations, sinon pour confirmer que les Français, dans leurs usages parlés quotidiens, auraient tendance à s'écarter des normes. En 1992, M. Voghera fait une constatation semblable pour l'italien : beaucoup de travaux sur l'italien parlé, mais très peu sur la syntaxe.

« I lavori esplicitamente dedicati alla sintasi sono pochissimi se confrontati col numero complessivo degli studi sul parlato » (p. 71).

Il reste pourtant de grands secteurs de la syntaxe française pour lesquels nous ignorons ce que sont les usages parlés les plus courants. C'est me semble-t-il le cas, en particulier, pour certains domaines que les grammairiens n'ont pas aimé traiter, comme certaines formes de questions.

Un observateur minutieux de la langue parlée, comme l'était Ph. Martinon en 1927, remarquait que nous étions souvent piégés par des

« indicibles » de langage. Il relevait, avec un certain agacement, qu'il nous arrivait de ne pouvoir légitimer aucun des usages mentionnés par les grammairiens, lorsque l'un était vraiment trop archaïque (« un grand nombre de personnes reculent ») et l'autre trop familier. Il ne reste plus alors qu'à « s'abstenir » ou à changer de tournure. Citons par exemple Ph. Martinon parlant des pronoms complément à l'impératif (p. 299) :

« A l'impératif sans négation, on a régulièrement *mettez l'y, renvoyez-les-y* ; et de même, avec *me* et *te*, pour *moi* et *toi*, *mets-t'y, prépare-t'y*. Toutefois, un grand nombre de personnes reculent devant ce *m'y* ou *t'y* ; quelques unes préfèrent *menez-y-moi* ; en réalité ni l'un ni l'autre n'est fort usité, et on aime mieux généralement s'abstenir ».

L'observation des usages dans les corpus contemporains semble donner raison à Martinon : on ne rencontre ni l'une ni l'autre des solutions qu'il critiquait. Mais on aimerait savoir comment et par quels moyens les usagers se débrouillent devant cet « indicible », et quelles autres tournures sont produites. Il y a peu de recherches sur ces points.

Je voudrais m'intéresser à certaines formes d'interrogation dont la complexité a souvent provoqué l'agacement des grammairiens. Ce sont celles que l'on trouve dans les questions sur l'attribut. En voici cinq formes :

- qu'est-ce ?
- qu'est-ce que cela ?
- qu'est-ce que c'est ?
- qu'est-ce que c'est que cela ?
- c'est quoi ?

Martinon estimait que, d'une façon générale, les tournures en *est-ce que*, même si elles étaient « absolument correctes », faisaient cependant « assez mauvaise figure dans la langue écrite » (p.131). Il fallait pourtant bien se résoudre à les utiliser dans ce genre de questions, puisque les autres formules étaient devenues impossibles et qu'on ne pouvait tout de même pas renoncer à poser ce type de questions :

« La langue la plus correcte elle-même recule aujourd'hui devant *qu'est-ce* ou *qu'était-ce*, qui deviennent *qu'est-ce que c'est ? Qu'est-ce que c'était ?* [...] Pour désigner un objet particulier, on ne peut plus dire *qu'est ceci ?* ou *qu'est cela ?* » (p. 132).

Il semblait voir là une sorte de « trahison de la langue », et déplorait que la « langue parlée » (au sens de langue « vulgaire »), en profite pour aller encore plus loin :

« Il faut donc dire au moins *qu'est-ce que cela ?* Mais la langue parlée ne s'en contente pas, et dit *qu'est-ce que c'est que ça*, où il y a trois *que*, deux *ce* et deux *est*, et qui fait rire assez justement les étrangers » (p.132).

Il me paraît intéressant d'aller voir de près ce que font réellement les Français, actuellement, avec ce genre de questions. Les corpus de français parlé nous en fournissent quantité d'exemples : c'est une des formes de questions les plus fréquemment attestées. Il n'est pas toujours aisé de les transcrire. Les éléments *que*, *c'est*, *ce*, ont souvent des formes raccourcies, difficiles à repérer. On peut confondre, à l'écoute, certaines des séquences, comme par exemple :

kEskTMsE
 kEsk^sE
 kEssE
 kTMsE

C'est donc une recherche qu'on ne peut faire qu'avec un corpus soigneusement établi, où les transpositeurs ont noté leurs incertitudes lorsqu'ils en ont eu.

Ces interrogations sont des domaines de la syntaxe du français contemporain difficiles à décrire à cause d'une grande variété de réalisations, que les grammairiens ont souvent voulu traiter uniquement en termes de niveaux de langue. De nombreuses grammaires présentent successivement l'interrogation dans la langue soutenue, dans la langue familière, dans la langue vulgaire, etc. La forme réputée la plus correcte est généralement traitée en même temps comme la forme originelle, dont les autres, moins correctes, et placées à des niveaux plus bas, seraient des dérivées.

Pour l'interrogation, c'est la forme avec postposition du sujet qui est généralement montrée en premier, *viendrez-vous ? Où va-t-il ?* Les autres, qui ont un sujet antéposé au verbe, *Il va où ?*, ou qui utilisent *est-ce que*, *Où est-ce qu'il va ?* sont présentées comme le résultat d'un affaiblissement de la bonne forme de départ : il semble qu'on perde la postposition du sujet et que *est-ce que* soit un palliatif. De cette manière, on a toujours l'impression que la forme normativement la plus correcte a également une prééminence logique et historique. Les formes moins correctes semblent apporter du désordre.

Ces classements sont souvent là faute de mieux, pour donner au moins un recensement des formes disponibles. A y regarder de près, en s'appuyant sur des exemples attestés dans des corpus, on s'aperçoit vite que les niveaux invoqués ne correspondent ni à des niveaux sociaux des locuteurs,

ni à des genres bien déterminés (interactions, narrations, récits, etc.). De plus, cette répartition en niveaux est la plupart du temps peu utile pour établir une filiation syntaxique entre les différentes formes.

Les emplois d'attribut, comme, par exemple, *Qu'est-ce ? Qu'est-ce c'est ?* ou *C'est quoi ?* forment apparemment une petite sous-partie d'un sous-chapitre de l'interrogation. Mais on y trouve en concentré des problèmes typiques de délimitations des données, d'interprétations de formes concurrentes, et d'explications à chercher pour les apparents « trous » dans le système. D'autre part, les attestations sont faciles à collecter. On en rencontre par exemple dans les demandes de définitions, dans les explications ou dans les jeux. Enfin, les emplois d'attributs sont grammaticalement assez spécifiques ; ils sont encore plus complexes que ceux du sujet, dont les grammairiens ont beaucoup parlé, et les « trous » du système y semblent encore plus flagrants. Les emplois de compléments et de sujets me serviront de points de comparaison par rapport à ceux de l'attribut.

Je rappellerai les difficultés de l'analyse morphosyntaxique ; je montrerai ce qu'apporte l'observation des corpus de français parlé, et enfin je proposerai une tentative d'interprétation pour la concurrence entre les deux formes les plus fréquentes.

1. DESCRIPTION MORPHOSYNTAXIQUE

1.1. APPARENTES FAIBLESSES DU SYSTÈME

L'interrogation sur l'inanimé ne dispose pas en français d'une morphosyntaxe simple. Les grammairiens l'ont souvent déploré en citant le cas particulièrement frappant du sujet : les formes *que* et *quoi*, réparties en distribution presque complémentaire, qui seraient des formes « simples », ne peuvent pas, sauf cas d'exception, servir pour le sujet. Le relais serait pris par la forme « renforcée », *qu'est-ce qui* :

- **que bouge ?* - (?) *quoi bouge ?*
- *qu'est-ce qui bouge ?*

Le Goffic (1994 :117) voit là une nécessité de renforcer l'identification de la chose questionnée. Un degré supplémentaire d'identification serait donné par l'adjonction de *c'est* :

- *qu'est-ce que c'est qui bouge ?*

de sorte qu'on en arriverait, selon lui (§159) à faire jouer trois fois de suite un mécanisme d'identification

La forme *quoi*, utilisable dans une tournure avec extraction, fournit une autre solution pour ce type de sujet :

- c'est quoi qui bouge

Cela fait donc au moins trois tournures, dont aucune n'est « directe », et qui sont souvent présentées comme des formes de compensation. P. Le Goffic invite à s'étonner de cette difficulté.

« Pourquoi cette impossibilité d'interroger au premier degré sur un inanimé sujet - impossibilité caractéristique du français, et dont seule l'accoutumance nous empêche de ressentir le caractère a priori surprenant ? » (1994 : 116).

Il y voit des raisons sémantiques fondamentales : interroger sur l'inanimé ne serait pas du tout la même opération que celle qui consiste à interroger sur l'animé. Pour l'animé, il est posé une question sur l'identité « qui est-ce ? », à laquelle on peut répondre par un nom propre. Pour l'inanimé, conçu nécessairement avec une « moindre individualisation », il est posé une question sur les propriétés. Le Goffic semble donc assez satisfait de constater qu'il y a, pour l'inanimé, des « trous » et des formes concurrentes. Cette irrégularité dans les formes serait un signe révélateur des irrégularités dans les significations mises en cause. Il me paraît intéressant de chercher des arguments pour vérifier cette hypothèse.

Un tableau des formes s'impose, pour chacune des fonctions majeures, sujet, complément et attribut, avec une comparaison permettant d'évaluer les lacunes de l'inanimé par rapport à l'animé. Un tel tableau est malaisé, en premier lieu à cause du foisonnement des données. J'ai pourtant écarté certains types d'interrogations « vulgaires » que citent les grammairiens, mais que je n'ai jamais rencontrées, comme cet exemple tenu pour « franchement populaire » :

- *quoi c'est donc qu'on me veut* (G. Duhamel, Grevisse § 390b)

Je n'ai retenu que les formes pour lesquelles j'avais plus d'une dizaine d'attestations, soit par oral soit par écrit.

Une seconde difficulté tient à l'analyse elle-même. Dans les tableaux qui suivent, j'ai traité le *est-ce que* interrogatif de *A quoi est-ce que tu penses ?* comme si c'était une extraction, *c'est à quoi que tu penses ?* avec une postposition du sujet *ce*, (que les grammairiens nomment « inversion »). En accord avec Damourette et Pichon, Riegel et alii et Le Goffic, je ne pense pas que cette analyse soit juste. Mais j'ai conservé cette terminologie qui

permet d'avoir une certaine symétrie dans les tableaux, et qui n'engage ici aucun problème central. Je n'avais aucun terme commode pour désigner les périphrases qui semblent impliquer une double focalisation, comme dans :

- à quoi est-ce que c'est que tu penses ?

Je les ai nommées « doubles », en m'inspirant de la terminologie de Le Goffic. J'ai appelé « directe » la place des pronoms interrogatifs placés là où se trouverait un syntagme nominal de même fonction, comme c'est le cas pour [*à quoi*] et [*à une difficulté des termes*], dans l'exemple oral suivant :

- c'est dû à quoi - c'est dû à une difficulté des termes
(Alf 25,6, La Poste)

Lorsque l'interrogatif est placé en tête de la construction, comme dans *à quoi tu penses*, j'ai désigné ce phénomène de *fronting* (Cf. Chisholm, 1984) par « place en tête ». J'ai conservé le terme d'« attribut » pour éviter d'avoir à justifier ici d'autres appellations (même si elles me semblent préférables, comme par exemple, à la façon de M. Gross, « objet de être »).

Au total, sept tournures ont été retenues. Même si cela n'est pas exhaustif, c'est déjà une base utile pour la comparaison entreprise ici. J'examinerai successivement les tournures utilisées pour interroger sur le complément prépositionnel, le complément prépositionnel, les sujets (animés et inanimés) et enfin les attributs, animés et inanimés.

1.2. LE COMPLÉMENT PRÉPOSITIONNEL

Il est commode de commencer par le complément prépositionnel car c'est là qu'on trouve la plus grande régularité. L'inanimé *quoi* et l'animé *qui* fonctionnent très régulièrement, sans aucune lacune. Il y a une seule forme, *quoi*, ou *qui*, et elle est disponible pour tous les emplois.

	inanimé	animé
1. directe	tu penses à quoi	tu penses à qui
2. place en tête	à quoi tu penses	à qui tu penses
3. + inversion	à quoi penses-tu	à qui penses-tu
4. extraction	c'est à quoi que tu penses	c'est à qui que tu penses
5. place en tête	à quoi c'est que tu penses	à qui c'est que tu penses
6. + inversion	à quoi est-ce que tu penses	à qui est-ce que tu penses
7. double	à quoi est-ce que c'est que...	à qui est-ce que c'est que...

1.3. COMPLÈMENT NON PRÉPOSITIONNEL QUE/QUOI

La distribution réciproque de ces deux formes a souvent été étudiée. En simplifiant les choses, on pourrait dire que *que* a uniquement sa place avant le verbe et *quoi* uniquement après :

- *que* me conseillez-vous (oral, Alf 30,9)
- mais lui il faisait *quoi* (oral, Puget D, 9,7)

Mais trois précisions s'imposent, selon que le verbe a un sujet ou non et selon la place de ce sujet.

La forme *quoi* ne peut pas se mettre avant un verbe doté d'un sujet, à quelque place qu'il soit, ce qui, dans le tableau suivant, exclut les tournures N°2 et N°5 :

- **quoi* tu vois, **quoi* vois-tu
- **quoi* c'est que tu vois, **quoi* est-ce que tu vois

Mais *quoi* se place librement après ou avant le verbe dépourvu de sujet, c'est-à-dire l'infinitif et le participe présent :

- faire *quoi* ? *quoi* faire ? - en faisant *quoi* ? en *quoi* faisant ?

La forme *que* se place uniquement devant un verbe, mais jamais devant un groupe sujet+verbe, ce qui, dans le tableau suivant, exclut les tournures N°2 et 5 :

- **que* tu vois ? - **que* c'est que tu vois ?

En revanche, *que* se place devant un infinitif, aussi bien que *quoi* :

- *que* faire ?

Dans le tableau suivant, je ne mentionnerai que les tournures à verbe conjugué, où se constate une distribution complémentaire de *que* / *quoi* :

	inanimé	animé
1. directe	tu vois <i>quoi</i>	tu vois <i>qui</i>
2. place en tête		<i>qui</i> tu vois
3. + inversion	<i>que</i> vois-tu	<i>qui</i> vois-tu
4. extraction	c'est <i>quoi</i> que tu vois	c'est <i>qui</i> que tu vois
5. place en tête		<i>qui</i> c'est que tu vois
6. + inversion	qu'est-ce que tu vois	<i>qui</i> est-ce que tu vois
7. double	qu'est-ce que c'est que tu vois	<i>qui</i> est-ce que c'est que tu vois

1.4. LES SUJETS

Ici, deux lacunes prévisibles. Les sujets, animés comme inanimés, ne peuvent pas avoir une place spéciale en tête (tournure N°2), puisque telle est leur position ordinaire, et ils ne peuvent évidemment pas provoquer une postposition du sujet (tournure N°3). L'inanimé a deux lacunes supplémentaires, qui tiennent aux mêmes contraintes sur *que/ quoi*, qu'on a vues précédemment : tournure N°1, là où on pourrait imaginer **Quoi bouge ?* et tournure N°5, *Quoi c'est qui bouge ?*. Cette dernière forme est parfois citée. Grevisse en donne seulement un exemple, produit par E. Ajar, qu'il estime « franchement populaire et tenu pour incorrect par tous les grammairiens » :

- « *Quoi c'est donc qu'on me veut ?* » (Grevisse p. 654)

je n'en ai jamais rencontré dans les corpus. Je retiendrai trois formes de questions sur les sujets inanimés, dont aucune n'est « directe » :

	inanimé	animé
1. directe		qui bouge
2. place en tête		
3. + inversion		
4. extraction	c'est quoi qui bouge	c'est qui qui bouge
5. place en tête		qui c'est qui bouge
6. + inversion	qu'est-ce qui bouge	qui est-ce qui bouge
7. double	qu'est-ce que c'est qui bouge	qui est-ce que c'est qui bouge

1.5. LES ATTRIBUTS

Le *qui* animé a l'ensemble des sept tournures, que le sujet soit un pronom *ce*, ou *tu* :

	animé	
1. directe	c'est qui	tu es qui
2. place en tête	qui c'est	qui tu es
3. + inversion	qui est-ce	qui es-tu
4. extraction	c'est qui que c'est (?)	c'est qui que tu es (?)
5. place en tête	qui c'est que c'est	qui c'est que tu es
6. + inversion	qui est-ce que c'est	qui est-ce que tu es
7. double	qui est-ce que c'est que c'est	qui est-ce que c'est que tu es

Pour l'inanimé, en revanche, on retrouve ici les lacunes dues aux contraintes sur *que/ quoi*, qui empêchent la production des tournures N°2

(**quoi c'est ?* / **que c'est ?*) et N° 5 (**quoi c'est que c'est ?* **que c'est que c'est ?*). Mais, de plus, quand le sujet est un pronom clitique, *ce* ou *tu*, la tournure N°4 (*c'est quoi que c'est*) ne semble pas très bonne ; je n'en ai jamais rencontré d'exemples. La tournure N°3 elle-même (*qu'est-ce ?*) est tenue pour suspecte. Le Goffic la juge « faible et quelque peu livresque » (p. 104), de sorte que seules les tournures N°1, 6 et 7, *c'est quoi ?*, *qu'est-ce que c'est ?* *qu'est-ce que c'est que c'est ?*, paraissent vraiment vivaces. Encore faut-il préciser que la tournure N°1, *c'est quoi*, a mauvaise presse. La plupart des grammairiens l'assignent à la « langue parlée familière » (Grevisse : 652, Riegel : 398) et même « relâchée » (Le Goffic : 110).

inanimé : sujets pronoms clitiques		
1. directe	c'est quoi	tu es quoi
2. place en tête		
3. + inversion	qu'est-ce	qu'es-tu
4. extraction	c'est quoi que c'est (?)	c'est quoi que tu es (?)
5. place en tête		
6. + inversion	qu'est-ce que c'est	qu'est-ce que tu es
7. double	qu'est-ce que c'est que c'est	qu'est-ce que c'est que tu es

Mais la situation est encore plus compliquée pour l'attribut inanimé lorsque l'on a comme sujet un pronom non clitique, *cela*, ou un syntagme nominal de type *un N*, *le N* (je prendrai pour exemple *un zouave*, issu d'une attestation du corpus).

Dans cette situation, aucune des sept tournures prévues jusqu'ici pour les autres fonctionnements n'est possible. A l'impossibilité (ou la difficulté) des tournures N°2, **qu'un zouave est ?*, 4, *c'est quoi qu'un zouave est ?* et 5, **que c'est qu'un zouave est ?*, déjà constatée, s'ajoute une grande gêne à produire la tournure N°1 (**cela est quoi ? un zouave est quoi ?*) et la grande rareté de la tournure N°3 (*qu'est cela ? qu'est un zouave ?*). Martinon(1927: 245) s'en méfiait :

« On ne le dit pas et on ne l'écrit guère » .

Grevisse trouve que cela fait partie d'une « langue assez apprêtée », et que c'est senti comme « peu clair » (1987, §388b et p. 1109).

Ajoutons qu'un problème particulier se pose pour les tournures N°6 et 7 : les formes telles qu'on pourrait les attendre d'après les modèles précédents sont bizarres, et jamais attestées :

- Qu'est-ce qu'un zouave est ?
- Qu'est-ce que c'est qu'un zouave est ?

A la place de ces formes attendues, on en trouve d'autres, d'usage très fréquent et très naturel, qui ressemblent aux précédentes, si ce n'est qu'elles n'ont pas le verbe *être* :

- Qu'est-ce qu'un zouave ?
- Qu'est-ce que c'est qu'un zouave ?

L'absence du verbe *être* change considérablement le problème posé au départ, qui consistait à s'intéresser aux formes de l'attribut. Par souci d'unifier le domaine, la plupart des grammairiens (par exemple Grevisse, Le Goffic) ont proposé de voir ici une ellipse du verbe *être* :

- Qu'est-ce que cela [est] ?
- Qu'est-ce qu'un zouave [est] ?
- Qu'est-ce que c'est que cela [est] ?
- Qu'est-ce que c'est qu'un zouave[est] ?

Sans me prononcer sur cette analyse, avec ou sans ellipse, je traiterai ces tournures comme des sortes de dérivées des constructions attendues.

Au total, le tableau des tournures est donc très lacunaire. Si on n'y intégrait pas les deux tournures suspectes N° 1 et 3, ni les dérivées des tournures N° 6 et 7, il serait même carrément vide :

attribut inanimé : sujets nominaux		
1. direct	cela est quoi (?)	un zouave est quoi (?)
2. place en tête		
3. + inversion	qu'est cela (?)	qu' est un zouave (?)
4. extraction		
5. place en tête		
6. + inversion		
dérivée	qu' est-ce que cela [est]	qu' est-ce qu'un zouave [est]
7. double		
dérivée	qu'est-ce que c'est que cela [est]	qu' est-ce que c'est qu'un zouave [est]

L'observation des usages montre que, à l'exception des « dérivées », toutes les tournures mentionnées dans ce tableau sont en fait exclues du français de conversation. Les tournures N° 1 et 3 sont même écartées de tous les usages, y compris les plus soutenus ; elles ne subsistent que comme archaïsmes. Il ne reste donc que les dérivées des tournures N°6 et 7, sans verbe *être* :

- *qu'est-ce qu'un colissimo (Alf 26, 10)*
qu'est-ce qu'un colissimo [est]
- *qu'est-ce que c'est que la vente pure (Ch 4, 11)*
qu'est-ce que c'est que la vente pure [est]

Pour questionner sur l'attribut inanimé en français contemporain, on a donc en fait deux solutions. Soit se passer du verbe *être*, soit l'utiliser, mais en ce cas dans des tournures « à détachement », très différentes des formules vues jusqu'ici. Les deux plus fréquentes dans les usages de français parlé contemporain sont de type *C'est quoi, le N ?* et *Qu'est-ce que c'est, un N ?*

- c'est quoi, un colissimo ?
- qu'est-ce que c'est, un colissimo ?

En voici des exemples :

- *c'est quoi le hayon élévateur (Alf 15,2)*
- *alors qu'est-ce que c'est un wagon-lit (Law 25,12)*

Les tenants de l'explication par ellipse rétabliront ici deux éléments : le verbe *être* et le *que*, qui permettent de retrouver une « syntaxe liée » :

- c'est quoi [que] le hayon élévateur [est]
- qu'est-ce que c'est [qu'] un wagon-lit [est]

Il sera intéressant de voir si les deux tournures attestées, en *quoi* et en *qu'est-ce*, sont équivalentes ou si elles correspondent à des visées sémantiquement différentes.

2. LES USAGES DANS LES CORPUS

Quelles sont les formules utilisées en français parlé contemporain ? On s'appuiera ici sur un petit échantillonnage de corpus de français parlé : personnes adultes de plusieurs âges, parlant soit dans des conversations entre égaux, soit dans des situations publiques où se marque l'autorité : discours, représentants de l'administration face aux administrés et clients, professeurs face aux élèves. L'échantillon comporte aussi des exemples d'enfants dans des conversations, dans des jeux, et dans des situations contraintes.

L'enquête a porté seulement sur l'attribut inanimé et sur quelques formes compléments utiles pour la comparaison. Les résultats sont assez clairs. On peut distinguer trois types d'usages. Certaines tournures, qui n'apparaissent que dans les exemples littéraires, sont totalement absentes du français parlé. Un deuxième type montre des tournures utilisées à l'oral dans des situations assez surveillées de parole publique. Le troisième type comporte

celles qui servent couramment dans les conversations ou les échanges entre pairs.

2.1. FORMES NON ATTESTÉES DANS L'ÉCHANTILLONNAGE DE FRANÇAIS PARLÉ

Les locuteurs n'utilisent pas le type *qu'est cela ?*, ni avec un pronom non clitique ni avec un syntagme nominal. Ils confirment donc l'avis des grammairiens pour qui il s'agit d'une tournure rare, sentie comme peu claire (Grevisse, §388b), qui paraît actuellement réservée aux usages écrits archaïsants :

- qu'est ce bruit ? - qu'est un héros ?
- qu'est la vieille langue ? (Littré)
- Qu'était cela ? De l'amour (Maupassant)
- Qu'est le plaisir ? (Courteline)
- Qu'est une foi sans les oeuvres ? (D. Rops)
- Que sera, ensuite, son avenir ? (De Gaulle)

Les locuteurs n'utilisent pas non plus le type *qu'est-ce ?*, dont les grammairiens parlent assez peu, même pas avec un élément détaché :

- Mais un baiser, à tout prendre, qu'est-ce ? (E. Rostand)

Il semble que cette tournure a vieilli. Dans une situation où l'on questionne sur un événement, par exemple un bruit soudain, les grammairiens laissent entendre qu'on pourrait dire :

- qu'est-ce ?

Mais cela sent le théâtre, et les locuteurs utilisent massivement l'autre tournure :

- qu'est-ce que c'est ?

2.2. LES EMPLOIS ATTESTÉS DANS DES SITUATIONS DE LANGAGE PUBLIC.

Les tournures dites « à ellipse du verbe *être* », comme *qu'est-ce que cela ?*, que je m'attendais à rencontrer plus souvent, semblent se trouver uniquement dans des situations de parole publique. Une employée de la poste l'utilise lorsqu'elle raconte comment elle parle aux clients de la poste

et lorsqu'elle donne des définitions relatives à son métier, avec des formules qui font partie des habitudes administratives :

- qu'est-ce qu'un colissimo ben un colissimo ce sont des paquets qu'on appelle à délais garantis (Alf 26, 10)

Un professeur l'utilise pour questionner ses élèves :

- qu'est-ce qu'un devin ?
- qu'est-ce qu'une boutique ?
(De Gaulmyn, in Kerbrat-Orecchioni : 268, 271)

Un orateur politique l'utilise par oral et par écrit :

- Qu'est-ce que le socialisme français aujourd'hui ?
(J. D. Bredin, cité par Ch. Plantin, in Kerbrat-Orecchioni : 76)

On trouvait régulièrement cette interrogation dans les textes poétiques du XIX^e siècle, où elle semblait fonctionner comme la forme la plus banale, sans marque stylistique spéciale. Hugo la choisissait souvent pour les fréquentes questions qu'il plaçait dans ses textes :

- Qu'est-ce, hélas! que le génie / Et la beauté ? (Cont XIX)
- Qu'est-ce donc que cet homme qui songe ? (467)
- Et d'abord, qu'est-ce que la justice ? (444)

On peut la considérer comme la seule tournure de prestige encore très vivace.

2.3. TOURNURES DE L'USAGE COURANT

2.3.1. Qu'est-ce que c'est que cela ?

C'est une question sans élément détaché, qui semble avoir, de ce fait, un caractère un peu soutenu, sans être pour autant littéraire. On la rencontre en français parlé dans les entrevues, dans les jeux et dans les activités scolaires.

- L1. alors dernière question qu'est-ce que c'est qu'un zouave
- L2. un zouave c'est quelqu'un - c'est quand on traite quelqu'un de zouave
(Law 30,5 : adulte questionnant enfants de 10 ans)
- qu'est-ce que c'est que la vente pure (Ch 4, 11 : dialogue à l'ANPE)

Il suffirait d'enlever le *que* qui assure, devant le syntagme nominal, la cohésion de l'ensemble, pour obtenir une tournure à détachement, représentée ici dans la ponctuation par une virgule :

- *qu'est-ce que c'est, la vente pure ?*

Cette tournure à détachement n'a pas du tout le même caractère soutenu que la précédente et ne s'emploie pas dans les mêmes circonstances.

2.3.2. Qu'est-ce que c'est

Il arrive que le référent ne soit pas donné directement dans la tournure, parce qu'il a été indiqué dans le contexte précédent. C'est une des formes de questions les plus fréquentes dans la conversation. Elle semble, comme le dit Le Goffic (p. 104), une des moins marquées stylistiquement. On la trouve chez tous les locuteurs.

- L1. *il y a des QCM c'est vrai*
- L2. *qu'est-ce que c'est*
- L1. *des Questionnaires Choix Multiples*
(Ch. 45, 2 : dialogue à l'ANPE)
- L1. *et en face - il y a - une cabane - qu'est-ce que c'est*
- L2. *qu'est-ce que - quelle cabane - qu'est-ce que tu racontes*
(Puget C,22,6 : enfants de 11 ans. narration)

2.3.3. Qu'est-ce que c'est, le N ?

C'est la même tournure que la précédente, accompagnée du référent de la question construit en détachement. C'est l'une des deux tournures que les locuteurs utilisent pour mentionner un sujet nominal, dans les situations peu contraintes. Les exemples sont très nombreux :

- L1. *qu'est-ce que c'est [...] un moulin*
- L2. *un moulin c'est comme une maison*
(Law 13, 2 : enfants de 10 ans, dans un jeu)
- L1. *alors qu'est-ce que c'est un wagon-lit*
- L2. *un wag quoi*
- L1. *un wagon-lit*
- L2. *un wagon-lit c'est un c'est comme un wagon de train*
(Law 25, 12) : adulte questionnant un enfant)
- L1. *qu'est-ce que c'est le yoga*
- L2. *le yoga c'est quand tu fais de la gymnastique comme ça*
(Law 28, 14 : enfants de 10 ans, dans un jeu)

2. 3. 4. *C'est quoi ?*

J'ai rappelé que la plupart des grammairiens jugent cette tournure très familière, à la limite de la vulgarité, semble-t-il. Elle se trouve pourtant chez tous les locuteurs. C'est la façon banale de questionner, dans des situations senties comme peu formelles, sur un terme technique qui vient d'être énoncé :

- L1. *et votre formation initiale c'est un CAP de vente*
- L2. *non*
- L1. *c'était quoi*
- L2. *c'était euh - je me rappelle pas comment ça s'appelait*
- L1. *bon enfin - c'était dans quoi - dans la cuisine dans la métallurgie dans quoi*

(Ch 3, 5 : employée de l'ANPE et un consultant)

2. 3. 5. *Le N, c'est quoi ? C'est quoi, le N ?*

C'est la même tournure que la précédente, avec le référent nominal construit en détachement. Elle a, de ce fait, une réputation plus mauvaise encore. On la rencontre pourtant fréquemment dans la langue de conversation, chez tous les locuteurs, en particulier dans les demandes de définition :

- L1. *niveau quatre c'est quoi*
- L2. *c'est les BAC mais en code informatique c'est code trois*
(Ch 54,5 : interviewer avec la dame de l'ANPE)
- L1. *ces batteries sont mises sur le hayon élévateur du camion*
- L2. *et c'est quoi le hayon élévateur*
(Alf 15,2 : amie parlant à l'employée de poste) :

Elle est courante chez les enfants, et dans les dialogues entre adultes et enfants :

- L1. *alors c'est quoi le ketchup*
- L2. *le ketchup c'est un truc rouge là qu'on met dans les spaghettis*
(Law 10,10 : enfants de 10 ans, dans un jeu)
- *c'est quoi comme marque votre chien*
(Cl 89,2 : enfant 8 ans parlant à un adulte)

Le Goffic (1994 : 110) juge très familier ce *c'est quoi*, au même titre que toutes les questions où le référent interrogé est laissé après le verbe, comme : *Vous êtes qui, vous ?* Pourtant cette forme de question offre plusieurs avantages, qui expliquent peut-être son succès. Elle peut se trouver dans une subordonnée comme *dire que...*:

- et les paquets de messagerie alors tu dis que c'est que c'est quoi
(Alf 18,14 : dialogue amie/ employée de la poste)

Elle permet de placer l'interrogatif dans la même position que les éléments de réponse :

- L1. et alors ce qu'il avait - si ça n'était pas la rougeole c'était quoi
- L2 c'était une angine
(CIW2,6 : dialogue entre adultes)

Enfin elle permet d'avoir la même forme d'interrogation pour l'attribut et pour différentes sortes de compléments, prépositionnels ou non : *je marque quoi, c'est dû à quoi, c'est comment* :

- je marque *quoi*
(Ch 79,12 : employée ANPE et consultant)
- ça veut dire *quoi* faire partie d'Antenne2
(Reumaux 46,5 : étudiante/ ingénieur)
- alors à ton avis cette cette difficulté euh euh à comprendre cela c'est dû à *quoi* c'est dû à c'est dû à une difficulté des termes
(Alf 25,6 : dialogue amie/ employée poste)
- ça sert à *quoi* - ça sert simplement au niveau du montage
(Reumaux 107,3 : ingénieur)
- ils sont rémunérés *par qui* alors
(Reumaux 46,10 : étudiante/ingénieur)
- ça s'écrit *comment* déjà
(Ch 102,12 : employée ANPE)
- L1. c'est *comment* un orque alors - c'est gris
- L2. un orque oui c'est gris - c'est noir et blanc - c'est comme ça
(Puget D1,2 : enfants de 10 ans)

Une question comme *c'est quoi le ketchup ?*, appuyée par autant de similitudes, semble bien installée dans les usages.

3. RÉPARTITION ENTRE LES DEUX FORMES À DÉTACHEMENT : *QU'EST-CE QUE C'EST, UN N ?* ET *C'EST QUOI, LE N ?*

Les attestations des corpus montrent donc que ce sont les deux tournures concurrentes dans le français parlé des conversations. Lorsque les

grammairiens les mentionnent, c'est souvent, comme on l'a vu, pour classer la question en *quoi* à un niveau inférieur à celle en *qu'est-ce que c'est*. Pourtant, à y regarder de près, ce n'est pas une affaire de niveau de langue, puisqu'on trouve toutes les tournures usuelles chez les mêmes locuteurs et dans les mêmes situations. Par exemple, dans un jeu sur les définitions de mots croisés, un adulte (études Bac + 5), questionne un enfant en utilisant à quelques secondes d'intervalles, pour le même référent, deux tournures qui ont la réputation d'appartenir à des niveaux différents :

- L1. alors dernière question *qu'est-ce que c'est qu'un zouave*
 - L2. un zouave c'est quelqu'un - c'est quand on traite quelqu'un de zouave alors il est pas content il est nerveux alors il casse tout
 - L3. et il dit il dit espèce de zouave pourquoi tu m'as fait mal
 - L2. il dit il dit espèce de zouave toi-même
 - L1. alors *c'est quoi un zouave*
- (L1 = adulte ; L2 et L3 = enfants de 10 ans ; Law30,5)

Je propose une hypothèse : la différence entre les deux ne serait pas une affaire de niveau, de relâchement dans l'expression ou de désir d'expressivité. Ce serait une différence d'ordre syntactico-sémantique.

Les exemples des corpus montrent une tendance nette à utiliser l'une de ces deux tournures avec des syntagmes nominaux à valeur générique, souvent accompagnés d'un article indéfini, et l'autre avec des nominaux à valeur plus spécifique, souvent accompagnés d'un déterminant défini. On trouve préférentiellement *qu'est-ce que c'est* ? avec un article indéfini et *c'est quoi* ? avec un article défini :

- *qu'est-ce que c'est un moulin, un wagon-lit, le yoga.....*
- *c'est quoi le hayon, le keichup, votre formation initiale....*

Les questions en *qu'est-ce que c'est* sont souvent équivalentes à *qu'est-ce que cela veut dire ? qu'est-ce que vous voulez dire ?* : Elles semblent agir comme des demandes de définitions générales :

- *qu'est-ce que vous voulez dire par flux*
- (Ch 60,14 : interview à l'ANPE)

Au contraire, les questions en *c'est quoi* visent souvent à élucider un terme qui vient d'être produit, de sorte qu'elles ont souvent une valeur nettement déictique. Il s'agit presque toujours d'un élément dont la détermination est spécifiée dans le contexte : *c'est quoi ce hayon que vous venez de mentionner ? C'est quoi ce ketchup en question ?* On en rencontre de nombreux exemples dans les dialogues où un locuteur lance un terme, que le deuxième reprend pour poser la question :

- L1. donc tu as la zone après - Europe - qui est en *PAL* sauf les Français
- L2. sauf la France - mais *PAL c'est quoi*
(Reumaux 75,14, ingénieur et étudiante)

- L1. j'aimerais que tu me parles un peu du *référentiel* de l'une de tes classes
- L2. alors *le référentiel c'est quoi déjà*
(93-2, Metall1,12)

On rencontre aussi cette tournure lorsque la question et la réponse sont formulées par le même locuteur :

- tu envoies ton montage - le banc de montage va prendre ces *trois trames-là* ce qui est - euh pour te donner une idée - *c'est quoi* - ça représente quoi - euh le huitième de la seconde - - un huitième de seconde
(Reumaux 97,14, ingénieur)

- maintenant *la radio le soir c'est quoi* bof tel chanteur qui va vous interpréter telles conneries
(Reumaux 66,16 : ingénieur)

La répartition entre les deux tournures rappelle la distinction entre forte et faible référentialité que T. Givón (1984 : 219-40) jugeait très importante pour la typologie des formes interrogatives, en particulier dans les interrogations portant sur les prédicats. Dans la langue Ute qu'il prenait pour exemple (Sud-Californie), les questions portant sur le prédicat sont nettement réparties selon qu'elles présentent un élément générique ou sur un élément spécifique (*type/token*), ce qu'on pourrait traduire approximativement, en français parlé, par :

- qu'est-ce c'est, comme sorte de chose ?
- c'est quoi, cette chose ?

Il serait intéressant de vérifier, avec une plus ample collecte de faits, la possibilité d'établir cette distinction sémantique entre ces deux tournures interrogatives, généralement peu décrites pour le français.

4. BILAN

Cette analyse est orientée vers une justification grammaticale et non stylistique des formes interrogatives. Il s'agit d'essayer de montrer que, derrière l'apparent désordre des formes, compliquées à établir et sans cesse

soumises à des « renforcements », il y aurait une organisation grammaticale. Je rappelle ici les principaux arguments.

Questionner sur l'attribut de l'inanimé ne se fait pas avec les mêmes tournures que pour les autres questionnements. Les complexités de la morphosyntaxe de *que / quoi* semblent donner raison à l'hypothèse sémantique de Le Goffic : l'inanimé est toujours présenté comme moins individualisable que l'animé. C'est sans doute pourquoi les formes semblent s'user et devenir désuètes aussi facilement. Il n'y a pas la même « force » dans l'inanimé *que* que dans l'animé *qui*. C'est peut-être une des raisons qui font que la tournure en *qu'est-ce ?* est incomparablement moins vivante que *qui est-ce ?*

L'interrogation sur l'attribut est plus délicate que toutes les autres, et c'est peut-être pourquoi le *que* attribut est moins vivace que le *que* complément : *Que voulait ce zouave ?* est un type plus fréquent que *qu'est un zouave ?* Il est probable (je n'ai pas pu le mesurer) que le *que* attribut est plus souvent « renforcé » que l'autre : *qu'est-ce que c'est, un zouave ?*

Pour les attributs de l'inanimé, les formes les plus vivantes sont indéniablement celles qui ont un détachement. C'est sans doute à relier à la tendance relevée, en français comme dans d'autres langues parlées (cf. Chisholm) à formuler toutes sortes d'interrogations dans des formules à détachement :

- *la formation là c'est une formation initiale hein*
(Ch 46,9 : consultation à l'ANPE)
- *c'est mercredi hein la formation collective*
(id. 101,12)

Enfin, j'ai proposé de voir, entre les deux formules à détachement portant sur l'attribut inanimé, deux valeurs syntactico-sémantiques différentes, d'après le degré de référentialité du sujet.

Ces hypothèses sont délicates à vérifier par de larges enquêtes quantitatives, dans la mesure où tous les locuteurs français semblent avoir des systèmes un peu mêlés, où l'on trouve à la fois des régulations grammaticales internes à la langue parlée et des influences fortes de formes apprises pour et par la langue écrite. Ceci est sans doute vrai, à des degrés divers, pour toute entreprise de description de la langue. On pourrait y voir des raisons de renoncer à décrire ces situations linguistiques un peu bâtarde. On peut y voir, au contraire des raisons de lancer des hypothèses qui permettent d'interpréter ces données hétérogènes. La description grammaticale de la langue parlée lance des défis redoutables. Les questions sur l'attribut en sont un exemple parmi quantité d'autres.

Claire BLANCHE-BENVENISTE

Bibliographie

- BLANCHE-BENVENISTE, Claire et JEANJEAN, Colette, 1987, *Le français parlé. Édition et transcription*. Paris : Didier-Érudition.
- BORILLO, Andrée, 1978, *Structure et valeur énonciative de l'interrogation totale en français*. Thèse de doctorat d'État, Université de Provence.
- CHISHOLM, William S. Jr. (ed.), 1984, *Interrogativity. A colloquium on the Grammar, Typology and Pragmatics of questions in seven diverse languages*. Amsterdam / Benjamins.
- GADET, Françoise, 1992, *Le Français populaire*, Paris : P. U. F. (Coll. « Que sais-je ? » n° 1172).
- GIVÓN, Talmy, 1984, « Utc », in CHISHOLM : 215-43.
- GRESILLON, Almuth, 1981, « Interrogation et interlocution », *DRLAV* 25 : 61-75.
- GREVISSE, Maurice, 1987, *Le bon usage. Grammaire française*. Treizième édition revue et corrigée par A. GOOSSE. Louvain-la-Neuve : Duculot.
- KERBRAT-ORECCHIONI, Catherine (dir.), 1991, *La Question*. Lyon : Presses Universitaires de Lyon (Collection « Linguistique et sémiologie »).
- LE GOFFIC, Pierre, 1994, *Grammaire de la phrase française*. Paris : Hachette.
- MARTINON, Henri, 1927. *Comment on parle en français. La langue parlée correcte comparée avec la langue littéraire et la langue familière*. Paris : Larousse.
- POHL, Jacques, 1965, « Observations sur les formes de l'interrogation dans la langue parlée et dans la langue écrite non littéraire », *Actes du Xème Congrès International de Linguistique et de Philologie Romanes*. Paris : Klincksieck
- RIEGEL, Martin, J.C. PELLAT et R. RIOUL, 1994, *Grammaire méthodique du français*. Paris : Presses Universitaires de France.
- VOGHERA, Miriam, 1992, *Sintassi e Intonazione nell'italiano parlato*. Bologna : Il Mulino.(Coll. Ricerca).

LE SUJET DANS LES ÉNONCÉS PSEUDO-CLIVÉS

Nous nous intéresserons aux énoncés du type :

ce qui compte c'est le voyage (oral, conversation, IA 85)
ce qui est très bon c'est le bœuf à l'ananas (oral, conversation, GARS I, ELOI)
Ce qui m'a plu, c'est que j'ai participé. (écrit, Aurélien, 7 ans)
ce qui compte c'est de ne pas se lancer comme des fous (oral, conversation, GG)
ce qui m'énerve c'est quand il crie (oral, conversation, GG)

nommés « pseudo-cleft sentences », depuis les travaux de E. Bach et S. Peters en 1968 (non publiés) suivis par ceux d'A. Akmajian en 1970¹. Nous garderons cette terminologie et parlerons « d'énoncés pseudo-clivés ».

1. LE SUJET DANS LES ÉNONCÉS PSEUDO-CLIVÉS

1.1. PROFORME ET LEXIQUE

Nous proposons de dire que dans un énoncé pseudo-clivé, le sujet est réalisé deux fois : une première fois, en tête de la construction, sous la forme d'une proforme, de valeur grammaticale, et une seconde fois, en fin d'énoncé, sous une forme lexicale réalisée après le verbe *c'est*

<i>ce qui</i>	<i>me</i>	<i>plaît</i>	<i>c'est</i>	<i>la linguistique</i>
<i>proforme</i>	<i>verbe constructeur</i>	<i>verbe d'équivalence</i>	<i>lexique</i>	

¹ Cf. Roubaud Marie-Noëlle, 1994, « Les constructions pseudo-clivées dans le cadre de l'Approche pronominal », thèse de 3^e cycle, pour une histoire de la terminologie.

La tournure pseudo-clivée, donne un statut particulier à la valence sujet de *plaire*. Nous la mettrons en parallèle avec d'autres dispositifs² tels le dispositif direct (ex. a) dans lequel le verbe ne réalise qu'une fois sa valence sujet et le dispositif d'extraction (ex. b) dans lequel le sujet est encadré par *c'est ... qui* et nous parlerons alors de « dispositif pseudo-clivé » :

- (a) dispositif direct : *la linguistique me plaît*
 (b) dispositif d'extraction : *c'est la linguistique qui me plaît*
 (c) dispositif pseudo-clivé : *ce qui me plaît c'est la linguistique*

Pour établir que *ce qui* est équivalent à *la linguistique*, il est indispensable de se placer dans un système de proportionnalité. Les exemples du corpus montrent qu'il faut envisager la relation syntaxique existant entre le verbe et le lexique pour déterminer le type de proforme de l'énoncé pseudo-clivé : la relation que le verbe entretient avec sa valence détermine la diversité des proformes (*ce qui, ce que, ce dont ...*)

valence + verbe constructeur	-----	proforme +	verbe constructeur
le voyage	compte	-----	ce qui
le bœuf à l'ananas	est très bon	-----	ce qui
les applaudissements	j'aime	-----	ce que
être légalisés	nous voulons	-----	ce que
d'aller en avion	j'aurais toujours rêvé	-----	ce dont j'aurais toujours rêvé
que la Provence va mal	je suis sûr	-----	ce dont je suis sûr

Nous proposons une analyse où le lexique n'est pas que l'expression lexicale spécifiée d'une proforme produite en tête d'énoncé³. La relation entre le lexique et la proforme est plus complexe, elle est avant tout syntaxique. Ce rapport⁴ entre la construction verbale contenant le lexique et celle contenant la proforme est constant : c'est celui de la relation syntaxique existant entre le verbe et sa valence.

² Nous empruntons ce terme de « dispositif » à Blanche-Benveniste Claire et alii, 1990, p.62.

³ Dans d'autres modèles, ces faits seront traités de manière anaphorique.

⁴ Les auteurs de *l'Approche pronominale* ont parlé de « rapport de proportionnalité » (1984), terminologie que l'on retrouve dans Blanche-Benveniste Claire et alii (1990): *Le français parlé*.

1.2. LE CAS DU SUJET

Ce que nous mettons en parallèle, dans le cas du sujet, c'est la relation syntaxique qui demeure constante entre le verbe et sa valence lexicale sujet et le verbe et sa proforme sujet. Nous constatons grâce aux exemples attestés du corpus, que l'énoncé pseudo-clivé n'est pas réductible à ses éléments composants. Une preuve en est donnée par les exemples où le sujet lexical est réalisé en *que* ou en *quand* :

- (a) *ce qui m'a plu c'est que j'ai participé*
 (b) *ce qui m'énerve c'est quand il crie*

Nous ne pouvons pas réduire ces énoncés pseudo-clivés à de simples énoncés (a') et (b'), ces derniers paraissant difficiles à réaliser :

- (a') ? *que j'ai participé m'a plu*
 (b') ? *quand il crie m'énerve*

C'est pour cette raison que nous rechercherons pas comme l'ont fait les générativistes ou les transformationnalistes la phrase-source qui serait à l'origine de l'énoncé pseudo-clivé (Cf. A. Akmajian :1970, F. Higgins :1973, M-L. Moreau :1970-71, J. Pinkham et J. Hankamer :1975, N. Chomsky :1977, E. Prince :1978, R. Geluykens :1988, E. Verheugd-Daatzelaar :1990). Nous considérons l'énoncé pseudo-clivé comme une construction syntaxique à part entière, dans laquelle le verbe entretient des relations particulières avec ses éléments valenciels.

C'est aussi pour des raisons syntaxiques que nous refusons d'appliquer à ces énoncés pseudo-clivés les termes de « procédé de mise en relief » (Wagner et Pinchon :1962, Corbeil :1968, Hanse :1983, Gardes-Tamine :1988), de « structure à présentatif » (Wartburg et Zumthor : 1958, Gadet :1989), de « gallicisme » (Prigent :1956, Léard :1992), de « pléonasmе » (Nyrop :1925) ou encore de phrase « à effets idiomatiques » (Laeven :1987). Ces concepts bloquent d'emblée l'analyse syntaxique car ils n'envisagent pas la relation syntaxique existant entre le verbe constructeur et sa valence.

D'ailleurs si nous recourons à un autre système de relation qui est celui des questions/ réponses, nous retrouvons cette proportionnalité qui s'établit entre la construction verbale avec pronom interrogatif non spécifié et celle avec lexique spécifié apparaissant dans la réponse :

QUESTION
qu'est-ce qui me plaît ?

RÉPONSE
*c'est la linguistique (qui me
 plaît⁵)
 c'est quand tu viens
 c'est qu'il a réussi*

Nous retrouvons aussi cette équivalence entre un pronom sujet peu spécifié et du lexique dans le cas de la dislocation de la valence sujet :

*ça me plaît, la linguistique
 ça me plaît, quand tu viens
 ça me plaît, qu'il a réussi*

Le sujet de *plaire* est doublement marqué⁶ dans cette construction verbale, une fois par le pronom *ça* et une seconde fois par du lexique.

Il nous a ainsi été possible de mettre en relation un élément peu spécifié (*ce qui, qu'est-ce qui, ça*) et du lexique sur une même place de construction syntaxique et dans trois systèmes de proportionnalité différents (le pseudo-clivage, la question/réponse et le double marquage). C'est la raison pour laquelle nous parlerons de « sujet lexical », quand nous aurons à considérer l'élément lexical réalisé dans la deuxième partie du dispositif pseudo-clivé.

1.3. TYPES DE PROFORME SUJET

Le corpus d'exemples attestés, constitué de 645 exemples de français parlé⁷ (pour 70 %) et de français écrit, montre que la proforme sujet se réalise de quatre façons différentes. Le tableau suivant présente, à l'horizontale, les différentes proformes sujets du corpus et, à la verticale, les formes prises par le lexique dans chacun des cas. Les chiffres renvoient au nombre d'exemples du corpus et les lettres aux exemples figurant sous le tableau :

⁵ A propos des questions/réponses, voir Blanche-Benveniste Claire, 1992, « A propos des énoncés sans verbe: les énoncés-réponses », p.66.

⁶ Pour l'analyse des dislocations, nous renvoyons à la thèse de 3^e cycle de Blasco Mylène, 1993, « Les énoncés à redoublement et dislocation en français contemporain. Analyse en double marquage ».

⁷ Ces exemples sont extraits du corpus global d'énoncés pseudo-clivés du volume II de ma thèse (1994). La proforme sujet représente 46% des énoncés globaux, le dispositif servant dans les autres cas à disposer la valence objet du verbe constructeur. Les énoncés oraux ont été relevés, pour la majorité, dans les corpus de français parlé du G.A.R.S. (Groupe Aixois de Recherches en Syntaxe) et dans celui d'Orléans. D'autres appartiennent à ma collection personnelle notée (Col.MN).

Proforme Lexique	ce qui	celui qui	le plus qui	le moins qui	TOTAUX
Syntagme Nominal	235 (a)	20 (i)	3 (l)	0	258 (40 %)
Pronom (prép.) Infinitif	26 (b)	4 (j)	0	0	30 (45 %)
« Que- phrase » ⁸	70 (c)	0	1 (m)	1 (o)	72
« Quand- phrase »	259 (d)	0	1 (n)	0	260 (40,3 %)
« Où- phrase »	13 (e)	0	0	0	13
« Si- phrase »	0	1 (k)	0	0	1
« Comment -phrase »	6 (f)	0	0	0	6 (55%)
« Quel- phrase »	4 (g)	0	0	0	4
TOTAUX	1 (h)	0	0	0	1
	614 (95 %)	25	5	1	645

Exemples en *ce qui* :

(a) *ce qui est très bon c'est le bœuf à l'ananas* (oral, conversation, GARS I, ELOI)

(b) *Ce qui travaille chez moi, c'est ça !* (écrit, roman, Col. MN)

(c) *ce qui compte c'est de ne pas se lancer comme des fous* (oral, conversation, GG)

(d) *Ce qui m'a plu, c'est que j'ai participé.* (écrit, Aurélien, 7 ans)

(e) *ce qui m'énerve c'est quand il crie* (oral, conversation, GG)

(f) *ce qui m'embête c'est si mon oculiste veut que je me pique moi-même* (oral, conversation, GG)

(g) *Ce qui compte c'est comment évoluent dans le temps les valeurs relatives au même institut.* (écrit, journal, GG)

(h) *ce qui me gêne par rapport à ça c'est quel est l'objectif qu'on poursuit* (oral, parole publique, Col. MN)

Exemples en *celui (celle, ceux, celles) qui* :

⁸ Cette terminologie de « Que-phrase », « Quand-phrase », « Où-phrase » ... facilite le repérage du type de subordonnée produite après *c'est* : l'élément introducteur *que, quand, où* ... figurant en premier. Ce type d'étiquettes a l'avantage de laisser de côté l'aspect sémantique que recouvrent les notions de « temporelles, conditionnelles » ... et qui ne correspondent pas toujours à la subordonnée produite (Roubaud, 1994:257).

- (i) *Celui qui les a tués c'est le tigre.* (écrit, enfant, 7 ans)
 (j) *celui qui fait le plus gros du travail c'est moi* (oral, conversation, LIC 91-6)
 (k) *celui qui est à côté là-bas c'est où J. se sert* (oral, conversation, CAPPE)

Exemples en le plus qui :

- (l) *le plus qui m'intéresse c'est le prix* (oral, conversation, CAPPE)
 (m) *le plus qui me fait râler c'est de voir que...* (oral, conversation, CAPPE)
 (n) *le plus qui m'embête c'est qu'il va sur la table de la télé et il y a les les fils tu sais derrière et j'ai peur qu'il m'attrape ça* (oral, conversation, LC 86)

Exemples en le moins qui :

- (o) *le moins qui me plaît c'est de faire la cuisine aux enfants* (oral, conversation, Col. MN)

La proforme sujet la plus représentée dans le corpus est *ce qui* (95 % des cas). C'est avec cette proforme qu'ont été relevés, dans le corpus, quelques exemples dans lesquels la « Que-phrase » sujet construit elle-même un dispositif pseudo-clivé :

- ce qui est fondamental c'est que ce qui compte c'est la relation qui se qui s'établit entre l'enfant et le thérapeute* (oral, conversation, LB 88)
ce qui est frappant c'est que ce qui se passe c'est une révolution bourgeoise (oral, conversation, Col. CBB)

Le dispositif pseudo-clivé est récursif, c'est-à-dire qu'un dispositif pseudo-clivé peut en construire un autre et ainsi de suite, alors que le dispositif d'extraction ne l'est pas, on ne peut pas extraire deux fois un sujet :

- c'est le chef qui habite à côté de chez moi qui l'a composé*
 * *c'est (c'est le chef qui habite à côté de chez moi) qui l'a composé*

Nous n'avons relevé que 25 exemples en *celui qui* réalisés principalement pour 20 d'entre eux avec du lexique nominal, 4 avec un pronom et 1 avec un locatif. Cette restriction sur la forme de lexique produit après *c'est* dans le cas d'une proforme en *celui qui* s'explique par l'analyse qui découle de l'étude des pronoms *ce* et *celui* dans le dispositif pseudo-clivé⁹. Considérons ces deux exemples, dont le premier est attesté :

- (1) *ce qui m'énerve c'est la prof* (oral, conversation, LIC 79-II)
 (2) *celle qui m'énerve c'est la prof*

⁹ Cf. Roubaud Marie-Noëlle, 1994, « Les constructions pseudo-clivées dans le cadre de l'Approche pronominale », thèse de 3^e cycle, pp.92-106.

Les pronoms *ce* et *celui* indiquent chacun une interprétation de la valence sujet de *énervé*. En (1), nous saisissons un ensemble de choses qui gravite autour du personnage et qui se résume à un *ça* globalisateur :

ce qui m'énerve c'est la prof
ça (la prof avec ses cours, ses manies, ses devoirs...)

En (2), nous saisissons le personnage qui ne peut s'identifier que par le pronom *celle-là*. Un contraste sur le lexique accentue cette saisie individualisante du sujet lexical :

celle qui m'énerve c'est la prof
celle-là (et pas une autre)

Cette saisie individualisante du sujet lexical oblige à un repérage « terme à terme » entre le pronom *celui* et le sujet lexical, si bien que l'accord en genre et en nombre entre ces deux éléments se fait obligatoirement :

ceux qui faisaient des mots croisés c'était des gens pas bêtes (oral, conversation, Orléans)
celles qui font le plus de bénéfice c'est les compagnies internationales (oral, conversation, Barnes)

Cette saisie du lexique est indépendante du trait (+ humain) ou (- humain) du sujet lexical. Avec du lexique de type (- humain), nous retrouvons les deux interprétations :

ce qui sonne c'est ce téléphone (ça)
celui qui sonne c'est ce téléphone (celui-là)

C'est la proforme qui impose son trait au lexique : le pronom *ce* étant chargé du trait (+ globalisant) et *celui* du trait (+ individualisant). C'est elle qui va catégoriser les éléments lexicaux qui lui seront proportionnels. C'est ainsi qu'une proforme en *celui qui* sera mise en équivalence avec quelque chose d'individualisé, c'est-à-dire avec du lexique nominal ou pronominal. Le seul exemple du corpus avec lexique verbal est un locatif :

celui qui est à côté là-bas c'est où J. se sert (oral, conversation, CAPPE)

Ce locatif représente en fait quelque chose d'individualisé (un magasin, dans le contexte), il n'échappe pas au trait (+ individualisant) du pronom *celui*. Par contre, il n'aurait pas été possible d'avoir après *c'est* du lexique verbal proportionnel à une proforme en *celui* :

* *celui qui m'a plu c'est que j'ai participé*
* *celui qui est beau c'est de se nourrir*

Les chiffres du tableau indiquent que dans 55 % des cas environ, le lexique produit après *c'est* est de type verbal, ce qui explique en partie le faible pourcentage d'énoncés en *celui qui* et l'écrasante majorité des énoncés en *ce qui* pouvant construire tout type de lexique, nominal ou verbal.

Notre corpus laisse également apparaître six exemples oraux en *le plus qui/le moins qui*, qu'aucun grammairien ne mentionne :

- le plus qui m'intéresse c'est le prix (oral, conversation, CAPPE)*
- le plus qui m'embête c'est qu'il va sur la table de la télé et il y a les les fils tu sais derrière j'ai peur qu'il m'attrape ça (oral, conversation, LC 86)*
- le plus qui a eu de chance dans tout ça ça a été le roi du Bon Marché (oral, conversation, LC 90-VI)*
- le plus qui a fait du mal dans Marseille c'était la milice et la Gestapo (oral, conversation, LC 90-VI)*
- le plus qui me fait râler c'est de voir que... (oral, conversation, CAPPE)*
- le moins qui me plaît c'est de faire la cuisine aux enfants (oral, conversation, Col. MN)*

Cette proforme en *le plus qui* ou *le moins qui* laisse passer tout type de lexique (nominal, verbal) sans le catégoriser, à la différence des proformes *ce qui* et *celui qui*. De tels exemples ne sont pas connus de tous les locuteurs français. Nous avons pu parler à ce propos de régionalisme syntaxique¹⁰.

Nous allons, dans un second temps, dégager les particularités du sujet lexical dans ce dispositif de réaction verbale. Pour cela, une étude distributionnelle est nécessaire car elle va faire apparaître que le sujet lexical, dans la construction pseudo-clivée, se réalise dans d'autres catégories grammaticales et varie dans sa composition. Dans un dernier temps, nous étudierons le comportement de ce sujet, tant du point de vue morphologique, syntaxique que sémantique.

¹⁰ Pour plus de détails, consulter: Roubaud M.N., « Les énoncés pseudo-clivés en *le plus / le moins* », à paraître.

2. ÉTUDE DISTRIBUTIONNELLE DU SUJET DANS LES ÉNONCÉS PSEUDO-CLIVÉS

2.1. CATÉGORIES GRAMMATICALES DE SUJETS LEXICAUX

Dans un énoncé pseudo-clivé, l'ordre des éléments est régulièrement *verbe + sujet lexical* :

proforme sujet + verbe + objet + c'est + sujet lexical
ce qui plaît à Marie c'est la linguistique

Cette position privilégiée du sujet lexical, en fin d'énoncé, change en partie la répartition des catégories grammaticales de sujets en français. Dans un dispositif direct ou extrait, les sujets infinitifs ou en *que* sont peu fréquents. Dans son étude sur les sujets, Laurence Honoré (1992-1993) a relevé, dans son corpus de 67 articles de presse, 4120 sujets dont 19 seulement étaient de type verbal :

Et que sa prophétie se réalise l'avantage bien.
 (écrit, journal, L. Honoré, « L'Événement du Jeudi », 17/12/92)
Que les journalistes ne soient pas des saints n'a peut-être aucune importance.
 (écrit, journal, L. Honoré, « L'Événement du Jeudi », 17/12/92)

Lors de son étude de 8 corpus de français parlé, elle n'a recensé, sur 1883 sujets, aucun sujet de type verbal. Tous étaient du type :

on s'éclatait on se défoulaît (oral, conversation, L. Honoré, corpus « Jeux », 5.9)
cette feuille me retourne (oral, conversation, L. Honoré, corpus « Sécurité Sociale », 1.8)

Dans le dispositif pseudo-clivé, le sujet lexical est autant verbal que nominal, comme l'indiquent les pourcentages dans le tableau précédent. Mais ce sont surtout les sujets nominaux et les « Que-phrases » sujets qui sont les plus représentés, suivis par les infinitifs (11 %) :

ce qui compte c'est pas la ville d'Orléans c'est le gouvernement (oral, conversation, Orléans)
ce qui compte à Tours c'est qu'il y a une vie de nuit des étudiants (oral, conversation, Orléans)
ce qui compte c'est de ne pas se lancer comme des fous (oral, conversation, GG)

C'est dans ce dispositif que nous relevons des infinitifs sujets en *pour*, qu'aucun grammairien, même Olof Eriksson (1993), dans son inventaire des constituants syntaxiques de la phrase française, ne relève comme sujet possible d'un verbe :

ce qui est embêtant c'est pour se placer (oral, conversation, Orléans)
ce qui a été le plus dur c'était pour trouver le pont de Saint-Cloud (oral, conversation, S-C)

Quant aux sujets nominaux, ils sont plus nombreux dans le corpus d'énoncés pseudo-clivés que les sujets pronominaux, dans une proportion de 8 syntagmes nominaux pour un seul syntagme pronominal :

ce qui prime eh bien ce sont c'est les tests psychotechniques (oral, conversation, LC 90-III)
ce qui travaille chez moi, c'est ça ! (écrit, conversation, Col. MN)

Avec un pourcentage de représentation dans le corpus de 40 % d'exemples, la catégorie nominale de sujets lexicaux l'emporte largement sur la catégorie pronominale. Le dispositif pseudo-clivé est le lieu privilégié des sujets nominaux, écrit et oral confondus.

2.2. COMPOSITION DES SUJETS LEXICAUX

Le fait que le sujet lexical soit produit en fin d'énoncé influe non seulement sur la catégorie grammaticale de sujet mais aussi sur la composition de ce syntagme. L'étude des exemples n'a révélé que 20 % de syntagmes nominaux se réduisant au déterminant et au nom :

ce qui est étrange aussi c'est cette mer (oral, conversation, LIC 89-A)

Le corpus présente en effet de longs sujets nominaux ou pronominaux, surtout à l'oral, avec de fréquentes énumérations (ex. 1 à 3), des contrastes (ex.4 à 7), ce qui augmente bien évidemment la longueur du sujet lexical :

1. *ce qui marche le mieux dans Fantasia c'est Mickey avec l'Apprenti Sorcier c'est Beethoven avec la Symphonie pastorale* (oral, conversation, LIC 92-2)
2. *ce qui est intéressant c'est justement ceux qui n'y arrivent pas ceux qui souffrent euh ceux qui ont des problèmes* (oral, conversation, LB 88)
3. *ce qui vous intéresse dans cette affaire Mesdames Messieurs c'est moins ce qu'on a pu ce qu'on a pu vous raconter sur le passé de Leduc mais davantage ce qu'il a pu réellement faire le soir du drame* (oral, parole publique, corpus d'avocats, TSCHA)

2. ÉTUDE DISTRIBUTIONNELLE DU SUJET DANS LES ÉNONCÉS PSEUDO-CLIVÉS

2.1. CATÉGORIES GRAMMATICALES DE SUJETS LEXICAUX

Dans un énoncé pseudo-clivé, l'ordre des éléments est régulièrement *verbe + sujet lexical* :

proforme sujet + verbe + objet + c'est + sujet lexical
ce qui plaît à Marie c'est la linguistique

Cette position privilégiée du sujet lexical, en fin d'énoncé, change en partie la répartition des catégories grammaticales de sujets en français. Dans un dispositif direct ou extrait, les sujets infinitifs ou en *que* sont peu fréquents. Dans son étude sur les sujets, Laurence Honoré (1992-1993) a relevé, dans son corpus de 67 articles de presse, 4120 sujets dont 19 seulement étaient de type verbal :

Et que sa prophétie se réalise l'avantage bien.
 (écrit, journal, L. Honoré, « L'Événement du Jeudi », 17/12/92)
Que les journalistes ne soient pas des saints n'a peut-être aucune importance.
 (écrit, journal, L. Honoré, « L'Événement du Jeudi », 17/12/92)

Lors de son étude de 8 corpus de français parlé, elle n'a recensé, sur 1883 sujets, aucun sujet de type verbal. Tous étaient du type :

on s'éclatait on se défoulait (oral, conversation, L. Honoré, corpus « Jeux », 5.9)
cette feuille me retourne (oral, conversation, L. Honoré, corpus « Sécurité Sociale », 1.8)

Dans le dispositif pseudo-clivé, le sujet lexical est autant verbal que nominal, comme l'indiquent les pourcentages dans le tableau précédent. Mais ce sont surtout les sujets nominaux et les « Que-phrases » sujets qui sont les plus représentés, suivis par les infinitifs (11 %) :

ce qui compte c'est pas la ville d'Orléans c'est le gouvernement (oral, conversation, Orléans)
ce qui compte à Tours c'est qu'il y a une vie de nuit des étudiants (oral, conversation, Orléans)
ce qui compte c'est de ne pas se lancer comme des fous (oral, conversation, GG)

C'est dans ce dispositif que nous relevons des infinitifs sujets en *pour*, qu'aucun grammairien, même Olof Eriksson (1993), dans son inventaire des constituants syntaxiques de la phrase française, ne relève comme sujet possible d'un verbe :

ce qui est embêtant c'est pour se placer (oral, conversation, Orléans)
ce qui a été le plus dur c'était pour trouver le pont de Saint-Cloud (oral, conversation, S-C)

Quant aux sujets nominaux, ils sont plus nombreux dans le corpus d'énoncés pseudo-clivés que les sujets pronominaux, dans une proportion de 8 syntagmes nominaux pour un seul syntagme pronominal :

ce qui prime eh bien ce sont c'est les tests psychotechniques (oral, conversation, LC 90-III)
ce qui travaille chez moi, c'est ça ! (écrit, conversation, Col. MN)

Avec un pourcentage de représentation dans le corpus de 40 % d'exemples, la catégorie nominale de sujets lexicaux l'emporte largement sur la catégorie pronominale. Le dispositif pseudo-clivé est le lieu privilégié des sujets nominaux, écrit et oral confondus.

2.2. COMPOSITION DES SUJETS LEXICAUX

Le fait que le sujet lexical soit produit en fin d'énoncé influe non seulement sur la catégorie grammaticale de sujet mais aussi sur la composition de ce syntagme. L'étude des exemples n'a révélé que 20 % de syntagmes nominaux se réduisant au déterminant et au nom :

ce qui est étrange aussi c'est cette mer (oral, conversation, LIC 89-A)

Le corpus présente en effet de longs sujets nominaux ou pronominaux, surtout à l'oral, avec de fréquentes énumérations (ex. 1 à 3), des contrastes (ex.4 à 7), ce qui augmente bien évidemment la longueur du sujet lexical :

1. *ce qui marche le mieux dans Fantasia c'est Mickey avec l'Apprenti Sorcier c'est Beethoven avec la Symphonie pastorale* (oral, conversation, LIC 92-2)
2. *ce qui est intéressant c'est justement ceux qui n'y arrivent pas ceux qui souffrent euh ceux qui ont des problèmes* (oral, conversation, LB 88)
3. *ce qui vous intéresse dans cette affaire Mesdames Messieurs c'est moins ce qu'on a pu ce qu'on a pu vous raconter sur le passé de Leduc mais davantage ce qu'il a pu réellement faire le soir du drame* (oral, parole publique, corpus d'avocats, TSCHA)

4. *Ce qui l'anime aujourd'hui, ce n'est plus leur zèle réformateur mais une vision élargie des diverses variétés de la pratique historique créative. (écrit, journal, Col. MN)*

5. *ce qui compte c'est pas la ville d'Orléans c'est le gouvernement (oral, conversation, Orléans)*

6. *Ce qui effraie le plus dans les partis, ce n'est pas ce qu'ils disent, c'est ce qu'ils négligent ou refusent de dire. (écrit, roman, Col. MN)*

7. *euh ce qui le rend sélectif ce n'est pas le chlorure de sodium en tant que chlorure de sodium mais la quantité de chlorure de sodium (oral, conversation, HOPWO)*

Il serait difficile de réaliser de tels sujets lexicaux dans un dispositif direct à cause de la longueur du sujet à gauche du verbe :

2' ? *ceux qui n'y arrivent pas ceux qui souffrent euh ceux qui ont des problèmes est (sont) intéressant (s)*¹¹

5'. ? *pas la ville d'Orléans mais le gouvernement compte*

Mais c'est surtout avec un sujet lexical de type « Que-phrase » que nous avons relevé les plus longs sujets lexicaux. Outre des énumérations, nous avons rencontré, à l'intérieur de la « Que-phrase », des constructions verbales en relation de dépendance (ex.8-9), des paroles rapportées en série (ex.10) :

8. *ce qui va le différencier de Breuil l'abbé Breuil c'est que l'abbé Breuil isole tous les objets tandis que Leroy-Gourand va lui au contraire penser qu'il faut essayer de faire des fouilles horizontales (oral, conversation, LB 86)*

9. *ce qui est important à comprendre c'est que si tu n'effectues pas une instruction particulière tu ne pourras pas sauvegarder tes données traitées (oral, conversation, LC 90-IV)*

10. *ce qui m'a paru bizarre c'est que quand je lui ai dit je vous mets à l'ordre quel ordre Monsieur il m'a dit non non non laissez laissez j'ai l'habitude je le ferais moi-même (oral, conversation, LC 92-8)*

La production d'une « Que-phrase » dans la deuxième partie du dispositif permet au locuteur de développer lexicalement son sujet, de produire plusieurs constructions verbales en série comme dans cet exemple :

12. *ce qui m'avait aussi le plus intrigué c'était que quelques jours plus tard repassant par la plaine d'Assidouillette j'avais remarqué que euh les nomades s'étaient arrêtés et m'étant approché je me suis aperçu qu'ils labouraient le champ (oral, conversation, LIC 89-D)*

¹¹ Il est tout à fait possible de réaliser cet exemple si on allonge l'énoncé à droite du verbe: *ceux qui n'y arrivent pas ceux qui souffrent euh ceux qui ont des problèmes sont intéressants à étudier*

Il est difficile d'envisager de tels sujets dans un dispositif direct, à gauche du verbe :

- 9'. ? *que si tu n'effectues pas une instruction particulière tu ne pourras pas sauvegarder tes données traitées est important à comprendre*
 10'. ? *que quand je lui ai dit je vous mets à l'ordre quel ordre Monsieur il m'a dit non non non laissez laissez j'ai l'habitude je le ferais moi-même m'a paru bizarre*

Par contre, ces « Que-phrases » sujets sont possibles avec un autre type de construction syntaxique qui rejette le sujet lexical en fin d'énoncé, à droite du verbe, comme le fait le dispositif pseudo-clivé. Il s'agit de la construction impersonnelle:

- 9". *il est important de comprendre que si tu n'effectues pas une instruction particulière tu ne pourras pas sauvegarder tes données traitées*
 10". *ça m'a paru bizarre que quand je lui ai dit je vous mets à l'ordre quel ordre Monsieur il m'a dit non non non laissez laissez j'ai l'habitude je le ferais moi-même*

La presse utilise d'ailleurs volontiers la construction impersonnelle, qui lui permet de réaliser de longs sujets lexicaux :

- Il est urgent que les pouvoirs publics respectent le vote des personnels en reconnaissant sans plus tarder la FSU, en lui faisant, dans tous les organismes consultatifs et dans les négociations, la place que les personnels lui reconnaissent par leur vote. (écrit, presse, Col. MN)*

Dans une construction pseudo-clivée, la construction syntaxique figure dans sa totalité dès la première partie du dispositif (exemple : *ce qui est important à comprendre*). Tous les éléments grammaticaux étant en place, le locuteur peut, dans la deuxième partie du dispositif (après *c'est*), développer le sujet lexical en produisant des énumérations, des commentaires, en recherchant son lexique, en rapportant les paroles de quelqu'un... Il en est de même avec les constructions impersonnelles. La distribution des éléments lexicaux sur l'axe syntagmatique joue un rôle primordial dans la réalisation des catégories grammaticales de sujet et dans la composition de cet élément.

3. ÉTUDE SYNTAXIQUE DU SUJET DANS LES ÉNONCÉS PSEUDO-CLIVÉS

3.1. LE SUJET ET LES ACCORDS

Le dispositif pseudo-clivé offre deux possibilités de saisie sémantique du sujet selon le type de proforme employée (*ce qui* ou *celui qui*) si bien que les accords dans la construction verbale seront différents.

3.1.1. Les accords en nombre

La proforme en *celui qui* oblige à une saisie individualisante du sujet lexical, à un repérage terme à terme entre le pronom *celui* et le sujet lexical :

ceux qui me plaisent c'est les journaux-là (eux-là)

La variation en genre et en nombre du pronom *celui* va de pair avec celle du sujet lexical et l'accord morphologique entre le sujet lexical et le verbe est ainsi respecté :

Celui qui l'écoute c'est son oncle Edmundo. (écrit, enfant, Col. MN)
celle qui a eu cette idée c'est la femme d'un collaborateur de Koch (oral, conversation, HOPWO)
ceux qui peuvent perdre quelque chose ce sont les Occidentaux et certainement pas les Soviétiques dans cette démarche (oral, T.V., LB 86)

Par contre, la proforme *ce qui* oblige à une interprétation globalisante du sujet lexical, équivalent à un *ça*, qui peut être marqué par du lexique singulier ou pluriel :

ce qui me plaît c'est le cinéma (oral, conversation, Col. Gadet)
ce qui me tient c'est mes parents ici (oral, conversation, Orléans)

L'accord morphologique se fait entre le verbe et la proforme *ce qui* mais pas avec le sujet lexical. C'est aussi ce qui se passe dans le cas du double marquage du sujet :

le cinéma, ça me plaît
mes parents, ça me tient

Il en est de même pour le cas de l'adjectif attribut du sujet. C'est la proforme, occupant la place syntaxique de sujet, qui détermine l'accord morphologique de l'adjectif attribut. Avec une proforme en *celui qui*, le repérage se faisant terme à terme, les accords morphologiques ont lieu tout au long de la construction verbale (entre la proforme, l'adjectif attribut et le sujet lexical) :

celle qui est dangereuse c'est la petite délinquance

Mais pris dans un dispositif pseudo-clivé avec une proforme en *ce qui* globalisante, l'adjectif attribut reste au masculin singulier, quels que soient le genre et le nombre du sujet lexical :

ce qui est dangereux c'est la petite délinquance (oral, conversation, RAVOU)
ce qui est très beau c'est la reprise des animaux (oral, conversation, Orléans)
ce qui est malheureux c'est les retraites des petits vieux (oral, conversation, Orléans)

C'est également ce qui se passe dans les dislocations, quand le lexique est repris au moyen du pronom globalisateur « c' » :

la petite délinquance c'est dangereux
la reprise des animaux c'est beau
les retraites des petits vieux c'est malheureux

3.1.2. Les accords en mode

Notre corpus atteste de couples d'exemples en *ce qui* pour lesquels le mode du verbe dans la « Que-phrase » sujet est aussi bien au subjonctif qu'à l'indicatif :

Ce qui m'inquiète un peu, c'est que Tintin n'ait pas encore réparé. (écrit, B.D., HERGE)
ce qui m'inquiète beaucoup c'est qu'il sait piloter un avion (oral, conversation, Col. MN)

Ce qui est important c'est que l'on ne soit pas mort. (écrit, roman, A. Christie)
ce qui est important pour moi c'est que ça permet à la personne qui est avec toi de s'exprimer (oral, conversation, HAMMA)

Des nuances sémantiques, dues à la valeur de l'indicatif ou du subjonctif en français, s'attachent à cette différence d'emploi :

ce qui m'ennuie c'est qu'on ne peut pas sortir (certitude : on ne peut pas sortir)
ce qui m'ennuie c'est qu'on ne puisse pas sortir (probabilité)

ce qui m'intéresse c'est qu'on apprend l'arabe (certitude : on apprend l'arabe)
ce qui m'intéresse c'est qu'on apprenne l'arabe (probabilité)

Dans un dispositif direct ou d'extraction (en *c'est ... qui*), l'indicatif est souvent difficile :

c'est important que l'on ne soit pas mort ? c'est important que l'on n'est pas mort

mais :

ce qui est important c'est que l'on ne soit pas mort
ce qui est important c'est que l'on n'est pas mort

cela me gêne que vous fassiez tout le travail ? cela me gêne que vous faites tout le travail

ce qui me gêne c'est que vous fassiez tout le travail
ce qui me gêne c'est que vous faites tout le travail

Le fait que le mode indicatif soit possible dans la « Que-phrase » du dispositif pseudo-clivé peut résoudre certains des problèmes posés par les formes du subjonctif, notamment quand il s'agit d'exprimer un fait passé. Nous savons combien il est difficile pour une partie des locuteurs d'employer l'imparfait du subjonctif (remplacé d'ailleurs le plus souvent par un présent du subjonctif) si bien que la possibilité de remplacer ce mode par l'indicatif permet au locuteur de ne pas se soucier des règles de concordance des modes en français :

c'était horrible qu'il revînt sur les lieux du crime
ce qui était horrible c'était qu'il revenait sur les lieux du crime

La construction pseudo-clivée est moins contraignante à ce niveau là que la construction impersonnelle ou les autres dispositifs.

3.2. DE NOUVEAUX SUJETS SYNTAXIQUES

3.2.1. Des nexus nominaux

Dans ce type de constructions apparaissent, ce que Jespersen (1937, traduction française de 1971 :204), appelle : des « nexus ». Considérons ces exemples :

1. *ce qui manque c'est des boîtes pour aller danser (oral, conversation, Orléans)*
2. *ce qui est commode c'est le facteur deux fois par jour (oral, conversation, Col. CBB)*

3. *ce qui est affreux c'est le fasciste qui refuse de répondre* (oral, conversation, LB 86)

La relation qui unit les deux éléments du sujet lexical (le nom et le syntagme qui le suit) n'est pas une simple relation de juxtaposition ou de « jonction » (pour reprendre la terminologie de Jespersen) que l'on trouve dans :

4. *ce qui manque c'est des boîtes de nuit*

Dans les exemples 1 à 3, il y a mise en relation des deux éléments constitutifs du sujet. La valence sujet n'est plus proportionnelle à un simple « ça » globalisateur (comme dans l'exemple 4) mais elle est constituée de la relation unissant les deux éléments constitutifs du sujet, que l'on pourrait paraphraser verbalement de cette façon :

1'. *ce qui manque c'est des boîtes dans lesquelles nous pourrions danser*
 2'. *ce qui est commode c'est le facteur qui passe deux fois par jour*
 3'. *ce qui est affreux c'est le fasciste quand il refuse de répondre*

La saisie du sujet lexical ne se fait plus en fonction du syntagme nominal pris comme seule unité (comme dans l'exemple 4) mais en fonction de la relation qui existe entre le nom et le syntagme qui le suit. C'est ce que Jespersen appelle « nexus » (là où les auteurs de *L'Approche pronominale* parle d'un « ça relationnel »). Derrière la proforme sujet en *ce qui* se cachent des relations de nexus qui sont elles-mêmes des relations syntaxiques. Le dispositif pseudo-clivé offre au verbe la possibilité de construire non plus de simples sujets nominaux (comme dans l'exemple 4) mais des nexus nominaux.

3.2.2. De nouvelles formes de sujets

La construction pseudo-clivée fait apparaître à droite du verbe, après *c'est*, des sujets lexicaux qui seraient difficilement réalisables dans un dispositif direct :

ce qui compte c'est ce que vous écrivez (oral, conversation, Orléans)
 (d. direct) *ce que vous écrivez compte*

ce qui intéresse Foucault c'est comment pouvoir et savoir sont liés (oral, conversation, GG)
 (d. direct) *comment pouvoir et savoir sont liés intéresse Foucault*

Il en est de même avec des sujets lexicaux de type « Que-phrase » qui ne peuvent être construits directement par le verbe. Ils ont besoin pour apparaître d'être construits par un nom comme *le fait* (le syntagme en *le fait*

que ... étant équivalent à un ça), des problèmes de mode seront à résoudre également :

ce qui ne va pas c'est que la table est ronde (oral, conversation, GG)
(d. direct) ? *que la table est/soit ronde ne va pas*
le fait que la table soit ronde ne va pas

ce qui les caractérise c'est qu'ils ont quatre pattes (oral, conversation, LIC 89-B)
(d. direct) ? *qu'ils ont/ qu'ils aient quatre pattes les caractérise*
le fait qu'ils aient quatre pattes les caractérise

ce qui rendait les choses particulièrement difficiles c'est que la variation est double (oral, parole publique, P. Encrevé)
(d. direct) ? *que la variation est/soit double rendait les choses particulièrement difficiles*
le fait que la variation soit double rendait les choses particulièrement difficiles

Des subordonnées en *quand, si, quel* deviennent également sujet lexical dans une construction pseudo-clivée comme en témoignent les exemples du corpus, impossibles à réaliser dans un dispositif direct :

ce qui m'énerve c'est quand il crie (oral, conversation, GG)
* *quand il crie m'énerve*

ce qui m'embête c'est si mon oculiste veut que je me pique moi-même (oral, conversation, GG)
* *si mon oculiste veut que je me pique moi-même m'embête*

ce qui me gêne par rapport à ça c'est quel est l'objectif qu'on poursuit (oral, parole publique, Col. MN)
* *quel est l'objectif qu'on poursuit me gêne*

Le remplissage lexical de la place syntaxique de sujet est moins contraignant dans la construction pseudo-clivée que dans les autres dispositifs de rection verbale. C'est ainsi que nous pouvons parler d'une extension du pouvoir de construction du verbe.

Cette liberté de construction du sujet lexical touche aussi la nature des éléments lexicaux qui entrent dans une énumération. Notre corpus témoigne de la combinaison de sujets lexicaux de différente nature grammaticale. Les exemples 1 à 3 fournissent, sur la place du sujet lexical, une série formée d'un syntagme nominal et d'un syntagme infinitif, l'exemple 4, d'un syntagme nominal et d'une « Que-phrase » et l'exemple 5, d'un syntagme pronominal et d'un syntagme infinitif :

1. *ce qui était important c'était la forme* avoir la perspective (oral, conversation, LA 84)
2. *ce qui m'amuse le plus ce n'est pas des mots extraordinaires* plutôt que de trouver une formulation astucieuse d'un mot (oral, conversation, Orléans)
3. *ce qui caractérise le chrétien c'est cette veille* veiller (oral, conversation, LC 90-IV)
4. *Ce qui caractérise toutes ces valences, ce n'est pas exactement l'interrogation*, c'est qu'elles laissent la modalité du verbe « en suspens » ... (écrit, ouvrage linguistique, Col. MN)
5. *ce qui intéresse les autres c'est ce qu'ont vécu les gens* c'est pas d'avoir un cours sur le scoutisme (oral, conversation, LIC 89-A)

Il serait bien difficile d'envisager de telles énumérations dans un autre dispositif : ? la forme avoir la perspective étaient important ? ce qu'ont vécu les gens et pas d'avoir un cours sur le scoutisme intéresse les autres

3.2.3. Un pouvoir de construction plus grand pour le verbe

Le verbe de l'énoncé pseudo-clivé a un pouvoir de construction supérieur à celui qu'il a dans un autre dispositif. Dans l'exemple suivant, la construction impersonnelle n'admet pas un sujet de type « Que-phrase » :

- il est bien français de ne pas respecter les limites de vitesse* (ex. cité par J. Deulofeu, 1988 :122)
 * *il est bien français qu'on ne respecte pas les limites de vitesse*

Dans une construction pseudo-clivée, l'une ou l'autre forme de sujet lexical est possible :

- ce qui est bien français c'est de ne pas respecter les limites de vitesse*
ce qui est bien français c'est qu'on ne respecte pas les limites de vitesse

Nous retrouvons cette extension du pouvoir de construction du verbe avec les énoncés en « être + adjectif ». Dans les exemples qui suivent, l'adjectif devient attribut d'un sujet lexical qu'il ne pourrait pas réaliser dans un dispositif direct. C'est ainsi que *cher* ou *nouveau* (qui peuvent être attributs d'un sujet infinitif) sont mis en relation avec une « Que-phrase » sujet (ex.1-2) et que les adjectifs *premier*, *hostile* ou *impitoyable* sont reliés à un sujet lexical de type verbal (ex.3-5) :

1. *ce qui est cher là-dedans c'est que c'est moi qui ai tout pris et pas les autres* (oral, conversation, LIC 79-II)
2. *ce qui est nouveau c'est que moi et quelques trois ou quatre confrères nous avons récupéré cette prise d'empreinte* (oral, conversation, LC 90-VII)
3. *Ce qui est premier, c'est de construire l'unité.* (écrit, ouvrage linguistique, Col. MN)
4. *ce qui me serait hostile c'est que tu lui parles de mes défauts* (exemple fabriqué)
5. *ce qui serait impitoyable c'est que tu lui coupes les vivres* (exemple fabriqué)

Il faudrait, pour ces trois derniers exemples, revenir à un syntagme nominal en « *le fait de/que* » pour rétablir les acceptabilités dans le dispositif direct :

le fait de construire l'unité est premier
le fait que tu lui parles de mes défauts me serait hostile
le fait que tu lui coupes les vivres serait impitoyable

3.3. SYNTAXE ET SÉMANTISME

3.3.1. Cas des adjectifs

La construction syntaxique influe sur le sémantisme de certains adjectifs construits dans la première partie du dispositif. En effet, la liste de fréquence des adjectifs du corpus pseudo-clivé diffère de celle du *Français fondamental* (Gougenheim et alii :1964). Comme fréquences les plus fortes figurent ici des adjectifs évaluatifs comme *important, intéressant, sûr, dommage* ainsi que l'adverbe *bien* (Roubaud, 1994 :148). La structure pseudo-clivée favorise l'entrée d'adjectifs ayant cette valeur sémantique ou les fait entrer avec ce sens. C'est ainsi que certains adjectifs perdent leur caractère de propriété physique en entrant dans la construction pseudo-clivée et deviennent des adjectifs évaluatifs :

ce qui est lourd c'est l'administration (oral, conversation, LIC 90)
ce qui était dur aussi c'est les lèvres qui étaient toutes coupées par l'oxygène
(oral, conversation, LIC 92-8)
ce qui était net c'était une marginalisation de Serge July (oral, parole publique, Col. MN)
ce qui est beau c'est que elle répond après une deuxième lettre le dix-huit (oral, conversation, LC 86)

D'autres perdent leur trait (+ personnel), habituellement attaché à une personne, si bien que leur lecture ne peut plus être qu'appréciative :

ce qui est idiot c'est que j'aime les chansons là nulles (oral, conversation, BOAVI)
Ce qui semblait bien mort par exemple, c'était la route. (écrit, roman, Col. Sandfeld)

Les adjectifs qui possèdent une double lecture : qualifiante et relationnelle, que Martin Riegel (1985 :118) appelle « des adjectifs doubles », entrent dans les énoncés pseudo-clivés avec leur valeur qualifiante. La syntaxe annule la possibilité d'une lecture relationnelle. C'est ainsi que dans cet exemple :

ce qui est très giscardien c'est la politique que vous menez

l'adjectif n'est pas mis en relation avec le personnage de Giscard mais il est pris au sens appréciatif de « habile » ou « maladroit ».

Un autre aspect de l'effet de la syntaxe sur le sens de certains éléments lexicaux est que certaines subordonnées, proportionnelles à la proforme sujet, ont des effets de sens particulier.

3.3.2. Cas des « si-phrases »

Considérons ces énoncés pseudo-clivés dont le sujet lexical est une « si-phrasé » :

ce qui m'embête c'est si mon oculiste veut que je me pique moi-même (oral, conversation, GG)
ce qui serait grave c'est si il y a un accident mortel (oral, conversation, RAILO)
ce qui aurait été dramatique c'est si ça aurait été un coup d'état (oral, conversation, Col. MN)

Sandfeld (1936, IV :339) parle d'une « condition » exprimée par *si*. Wagner (1939 :537) et G. Guillaume (1965 :50) préfèrent le terme de « supposition ». A. Delaveau (1990), dans sa thèse sur la conjonction *si*, montre que dans ce type de conditionnelles, la principale est jugée « vraie » par rapport à la conditionnelle :

« cela m'embête » est vraie dans le cas où « mon oculiste veut que je me pique moi-même »

L'effet obtenu est un effet de supposition dans le temps induite par la conjonction *si*. Avec d'autres éléments introducteurs comme *quand* ou *que*, cet effet est annulé pour devenir répétitif (avec *quand*) ou causatif (avec *que*) :

ce qui m'embête c'est quand mon oculiste veut que je me pique moi-même
(cela m'embête chaque fois que mon oculiste veut que je me pique moi-même)
ce qui m'embête c'est que mon oculiste veut que je me pique moi-même
(que mon oculiste veut que je me pique moi-même est la cause de mon ennui)

Ces conditionnelles en *si* n'apparaissent qu'avec des verbes ou adjectifs évaluatifs, comme dans les exemples précédents. Avec un verbe comme *donner raison*, cet effet d'hypothèse dans le temps est annulé :

ce qui me donne raison c'est si mon oculiste veut que je me pique moi-même

La construction pseudo-clivée autorise la réalisation de sujets lexicaux de type supposition dans le temps, difficilement réalisables dans une autre construction, comme le fait remarquer A. Valli (1981, note 15) avec cet exemple :

ce qui serait grave c'est s'il y a un accident mortel ? il serait grave s'il y a un accident mortel ? s'il y a un accident mortel serait grave

3.3.3. Cas des « quand-phrases »

C'est aussi avec ce type de verbes ou adjectifs évaluatifs qu'apparaissent des subordonnées en *quand* qui ne sont pas des valences temporelles ordinaires. Elles ont été identifiées dans le dispositif direct par Colette Jeanjean (1984). Nous les retrouvons dans le dispositif pseudo-clivé :

ce qui m'énerve c'est quand il crie (oral, conversation, GG)
ce qui l'énervait le plus c'est quand elle parlait toujours de ses voyages (oral, conversation, GG)
ce qui me fait mal c'est quand on appuie dessus (oral, conversation, Col. CBB)
ce qui m'a fait sourire c'est quand on a parlé de valse (oral, conversation, Col. MN)

Ces subordonnées n'ont pas le statut des autres subordonnées temporelles en quand, on ne peut pas les extraire : ? c'est quand il crie qui m'énerve ? c'est quand on appuie dessus qui me fait mal ? c'est quand on a parlé de valse qui m'a fait sourire

X. Chu (1987 :259) a parlé à propos de ces subordonnées « d'un *quand* conçu de manière substantivante » :

ce qui m'énerve c'est quand il correspond à toi criant tu m'énerves

L'énoncé pseudo-clivé produit n'est donc pas équivalent à celui-ci :

ça m'énerve quand il crie
c'est quand il crie que ça m'énerve

Du fait de sa valeur substantivante, ces temporelles n'acceptent pas de modal (Cf. Chu, op. cité), les énoncés suivants paraissant bizarres :

ce qui m'énerve c'est quand il peut crier
ce qui l'énervait le plus c'est quand elle pouvait parler de ses voyages
ce qui me fait mal c'est quand on peut appuyer dessus
ce qui m'a fait sourire c'est quand on a pu parler de valse

Ces subordonnées en *quand* sont du type de celles identifiées par Colette Jeanjean, dans son article « *toi quand tu souris* ». Elles sont liées syntaxiquement à un paradigme *toi* dans la même valence, que l'on peut reconstruire :

ce qui m'énerve c'est lui quand il crie
ce qui l'énervait le plus c'est elle quand elle parlait toujours de ses voyages
ce qui me fait mal c'est toi quand tu appuies dessus
ce qui m'a fait sourire c'est lui quand il a parlé de valse

Il est possible de ne pas réaliser la forme longue de la structure « *toi/quand*-phrase » des exemples ci-dessus. Le locuteur peut produire l'une ou l'autre des formes courtes :

ce qui m'énerve c'est lui
ce qui m'énerve c'est quand il crie

ce qui me fait mal c'est toi
ce qui me fait mal c'est quand tu appuies dessus

Mais la saisie que l'on a du lexique sujet n'est pas la même. La forme courte avec pronom présente un aspect quelconque du personnage alors que celle avec la subordonnée en *quand* en montre un aspect particulier. Un contraste sur la conjonctive accentue la saisie particulière que l'on a du personnage, lorsque seule la subordonnée en *quand* est réalisée :

ce qui m'énerve c'est quand il crie et pas quand il pleure
ce qui me fait mal c'est quand tu appuies dessus et pas quand tu défais le pansement

La structure pseudo-clivée convient parfaitement aux verbes et adjectifs évaluatifs si bien que seront réalisés à l'intérieur de la construction, des sujets lexicaux aux effets de sens particuliers comme ces subordonnées en *si* et en *quand* dont nous venons de parler.

CONCLUSION

Cette étude sur le sujet dans les énoncés pseudo-clivés a montré qu'un des intérêts majeurs de ce dispositif de rection verbale réside dans l'enrichissement du sujet lexical tant du point de vue de sa composition que de sa réalisation dans des catégories grammaticales diverses. La place privilégiée du sujet lexical en fin d'énoncé et le grand nombre de sujets de

type *ça* offrent au verbe la possibilité de construire de nouvelles formes de sujets. Des effets de sens particulier de certains sujets lexicaux sont attachés au type de verbe ou d'adjectif entrant dans la construction.

Nous venons de voir qu'il est difficile, pour certains exemples, de produire le sujet lexical dans un autre dispositif. Il n'est pas possible de réduire la construction pseudo-clivée à une phrase simple car ce serait détruire les équilibres syntaxiques. Nous ne pouvons pas, dans ce cas, adhérer à l'explication d'Eriksson (1993 :94) qui, pour justifier la présence de certains infinitifs sujets dans les énoncés qui nous intéressent, parle en termes de « distorsions dans l'analyse ». Il est dommage que beaucoup de grammairiens ignorent cette construction qui passe encore pour être une phrase emphatique ou un procédé de mise en relief. L'analyse du sujet dans les énoncés pseudo-clivés montre que la structure pseudo-clivée est une construction syntaxique majeure du français.

Marie-Noëlle ROUBAUD

Bibliographie

- AKMAJIAN A., 1970, « On deriving cleft sentences from pseudo-cleft sentences », *Linguistic Inquiry*, n°1, 149-168.
- BLANCHE-BENVENISTE Claire, DEULOFEU José, STEFANINI Jean et VAN DEN EYNDE Karel, 1984, *Pronom et syntaxe. L'Approche pronominale et son application au français*, Paris, Didier Érudition, INALF.
- BLANCHE-BENVENISTE, BILGER Mireille, ROUGET Christine et VAN DEN EYNDE Karel 1990, *Le Français parlé : études grammaticales*, Paris, C.N.R.S.
- BLANCHE-BENVENISTE Claire, 1992, « Sur un type de nom 'évaluatif' portant sur des séquences verbales », *Review of applied linguistics*, n°97-98, K.U. Leuven, 1-25.
- BLANCHE-BENVENISTE Claire, 1992, « A propos des énoncés sans verbe : les énoncés-réponses », *Recherches sur le français parlé*, n°11, 57-85.
- BLASCO Mylène, 1993, « Les énoncés à redoublement et dislocation en français contemporain. Analyse en double marquage », thèse de doctorat de linguistique française, Université d'Aix-Marseille I.
- CHOMSKY Noam, 1977, « On WH-movement », P.W.Culicover, T. Wasow and A. Akmajian (eds), *Formal Syntax*, New-York, Academic Press, 71-132.
- CHU Xiao-quan, 1987, « Étude sur les verbes modaux en français contemporain », thèse de doctorat de linguistique française, Université de Provence.

- CORBEIL Jean-Claude, 1968, *Les structures syntaxiques du français moderne. Les éléments fonctionnels dans la phrase*, Paris, Klincksieck.
- DELAVEAU Annie, 1990, « La conjonction 'si' dans ses emplois interrogatifs et conditionnels », thèse de doctorat de linguistique française, Université de Paris VII.
- DEULOFEU José, 1988, « Les couplages de constructions verbales en français parlé : effet de cohésion discursive ou syntaxe de l'énoncé », *Recherches sur le français parlé*, n°9, 111-141.
- ERIKSSON Olof, 1993, *La phrase française. Essai d'un inventaire de ses constituants syntaxiques*, Suède, Acta Universitatis Gothoburgensis, Romanica Gothoburgensia XLII.
- GARDES-TAMINE Joëlle, 1988, *La grammaire*, Volume II, Paris, A. Colin.
- GELUYKENS Ronald, 1988, « Five types of clefting in English discourse », *Linguistics*, n°26, 823-841.
- GOUGENHEIM G., RIVENC P., MICHEA R. et SAUVAGEOT A., 1964, *L'élaboration du Français Fondamental (1° degré)*, Paris, Didier.
- GUILLAUME Gustave, 1965, *Temps et verbe*, Paris, Champion.
- HANSE Joseph, 1983, *Nouveau dictionnaire des difficultés du français moderne*, Paris, Duculot.
- HIGGINS F. Roger, 1973, « The pseudo-cleft construction in English », thèse M.I.T.
- HONORE Laurence, 1992-1993, « Quels sujets ? Relevé des sujets dans 8 corpus de français parlé et 67 articles de presse. Comparaisons, ressemblances et différences. », mémoire de maîtrise, Université de Provence, département de linguistique française.
- HONORE Laurence, 1993-1994, « Autant que faire se peut. L'infinitif sujet en français contemporain, description des usages. », mémoire de D.E.A., Université de Provence, département de linguistique française.
- JEANJEAN Colette, 1981, « L'organisation des formes sujets en français de conversation. Étude quantitative et grammaticale de deux corpus », *Recherches sur le français parlé*, n°3, 99-134.
- JEANJEAN Colette, 1984, « 'Toi quand tu souris'. Analyse sémantique et syntaxique d'une structure de français parlé peu étudiée », *Recherches sur le français parlé*, n°6, 131-165.
- JESPERSEN Otto, 1937, *Analytic Syntax*, réimpression de 1969, New-York, : Holt, Rinehart and Wintson et traduction française de 1971 par Anne-Marie Léonard, Paris, Ed. de Minuit.
- LEARD Jean-Marcel, 1992, *Les gallicismes. Étude syntaxique et sémantique*, Paris-Louvain La Neuve, Duculot.

- MOREAU Marie-Louise, 1970-71, « Grammaire Transformationnelle du français. Trois aspects de la syntaxe de 'c'est' », thèse de doctorat de linguistique française, Université de Liège.
- PINKHAM Jessie et HANKAMER Jorge, 1975, « Deep and Shallow Clefts », *Paper from the Eleventh Regional Meeting of the Chicago Linguistic Society*, 429-450.
- PRINCE Ellen F., 1978, « A comparison of WH-clefts and IT-clefts in discourse », *Language*, n°54, 883-906.
- RIEGEL Martin, 1985, *L'adjectif attribut*, Paris, P.U.F.
- ROUBAUD Marie-Noëlle, 1994, « Les énoncés pseudo-clivés dans le cadre de 'l'Approche pronominale' », thèse de doctorat de linguistique française, Université d'Aix-Marseille I.
- ROUGET Marie-Christine et SALZE Laurence, 1985, « C'EST ... QUI, C'EST...QUE : le jeu des quatre familles », *Recherches sur le français parlé*, n°7, 117-139.
- SANDFELD Kr., 1936, *Syntaxe du français contemporain. I. Les propositions subordonnées*, Paris, Droz.
- SCAPPINI Sophie-Anne, 1987-1988, « Étude des extractions en C'EST...QU- : différenciation entre une relative et une extraction en C'EST...QU- », mémoire de maîtrise, Université de Provence, département de linguistique française.
- TOURATIER Christian, 1987, « Définition typologique du sujet », *Travaux 5. Typologie des langues*, CLAIK, pp.59-79.
- VERHEUGD-DAATZELAAR Els, 1990, *Subject arguments ans predicate nominals. A study of French copular sentences with two Nps*, Amsterdam-Atlanta, Rodopi B.V.
- VALLI André, 1981, « Note sur les configurations dites 'pseudo-clivées' en français », *Recherches sur le français parlé*, n°3, 194-211.
- WAGNER Robert-Léon, *Phrases hypothétiques commençant par si dans la langue française des origines à la fin du XVI^e siècle*, Paris, Droz.
- WAGNER Robert-Léon et PINCHON Jacqueline, 1962, *Grammaire du français classique et moderne*, Paris, Hachette.

ESPÈCE DE, GENRE DE, SORTE DE À L'ORAL ET À L'ÉCRIT¹

Cette analyse de [det espèce / genre / sorte de N] ([det N1 de N2]) a été tout d'abord menée à partir d'un corpus de 203 exemples réels, tirés des transcriptions de français parlé du Groupe Aixois de Recherche en Syntaxe. Les résultats ont été ensuite comparés aux données de l'écrit littéraire (1108 exemples extraits de 44 romans et essais contemporains de la base de données ARTFL de Chicago²).

Quand on s'intéresse à la syntaxe de *espèce de*, la nécessité d'une étude en contexte, à partir d'exemples réels, se fait particulièrement sentir. C'est que les structures en *espèce de* sont par essence ambiguës. A partir d'un exemple isolé (et inventé) comme :

dans cette maison, il y avait trois espèces de chats

il est impossible de décider s'il s'agit d'exactement trois animaux vaguement félins (interprétation approximative) ou d'au moins trois chats d'espèces différentes (interprétation sous-catégorisatrice).

Pour comprendre le mécanisme de ces structures, j'ai observé de près leurs propriétés distributionnelles, les phénomènes d'accord en genre et en nombre à l'intérieur et à l'extérieur du bloc [det N1 de N2] et enfin les déterminants qui apparaissent devant le premier substantif. Ces deux derniers points sont des secteurs où il y a une assez grande latitude en français ; il est donc particulièrement important de disposer d'exemples

¹ Cet article est une version remaniée d'une communication donnée en mai 1994 à l'université Queen's, Kingston, Ontario (Colloque "Oralité").

² Il s'agit de la base de données qui a servi à l'élaboration du Trésor de la Langue Française à Nancy, augmentée de quelques textes.

oraux, réputés « naturels » et « spontanés », en tout cas non revus par des correcteurs professionnels³.

1- PROPRIÉTÉS DISTRIBUTIONNELLES

Cette analyse des syntagmes nominaux (SN), inspirée de l'approche pronominale, a été développée dans Rouget 1989. Il y a deux façons de « disloquer » un SN, la lexicalisation à retardement et le double-marquage, et elles révèlent des propriétés utiles pour l'analyse des SN.

Le double-marquage correspond à une double réalisation d'un SN, c'est-à-dire à une réalisation lexicale ou semi-lexicale complète d'un SN couplée à un pronom clitique coréférent (un deuxième SN) ; Cette construction est possible quelle que soit l'interprétation d'*espèce / genre / sorte* :

*1- ils prennent leur pâte + leur sorte de pâte et ils la mettent dans les scourtins
(oral - Puger-Ville C 20)
1a - ils la prennent, leur sorte de pâte*

*2- on a beaucoup parlé de mouvoir + pour qualifier ce genre d'établissement(s)
(oral - Maison de retraite 3)
2a - pour le qualifier, ce genre d'établissement(s)*

La lexicalisation à retardement ressemble au double-marquage car il s'agit aussi d'un redoublement du même référent à l'intérieur d'une construction verbale. Cependant, ce sont deux structures très différentes du point de vue syntaxique. La lexicalisation à retardement correspond à des constructions nominales où on a un SN donné sans lexique (sous forme de pronom non-clitique contenant la détermination du SN) suivi d'une forme incomplète [de N], sans détermination, où apparaît le lexique, comme « retardé » :

*3a - c'est une agitation une danse de Saint-Guy
3b - c'en est une, de danse de Saint-Guy*

La construction avec lexique « retardé » est difficile pour les SN avec approximatifs : elle a tendance à les transformer en sous-catégorisateurs.

*3- c'est en gros une agitation sur place une espèce de danse de Saint-Guy
(oral - Microbiologie 82)*

³ Dans la partie "propriétés distributionnelles", ces exemples réels ne seront pas simplement observés mais ils seront l'objet de manipulations et de jugements d'acceptabilité.

3c - ? *c'en est une, d'espèce de danse de Saint-Guy*

4- *c'est une sorte de maquette qu'on achète (oral - Train 4)*

4a - ?? *c'en est une, de sorte de maquette qu'on achète*

4b - ? *c'en est une qu'on achète, de sorte de maquette*

4b me semble meilleur que 4a car l'incidence de la relative déterminative tire le lexème *sorte* vers l'interprétation sous-catégorisatrice : *cette sorte, la sorte qu'on achète en magasin* par opposition à *la sorte qu'on fabrique soi-même*.

La lexicalisation à retardement est parfaitement compatible avec tous les exemples de sous-catégorisateurs :

5- *il y a deux sortes de cartes il y en a une qui s'appelle la carte locale...*

(oral - Carte bleue 4)

5a - *il y en a deux, de sortes de cartes*

6- *dans le genre de photo que je fais ce qui : ce que je recherche c'est...*

(oral - Photo 22)

6a - *dans celui que je fais, de genre de photo*

Cette différence de comportement montre que l'approximatif, contrairement au sous-catégorisateur, n'est pas un N1 « tête » ordinaire⁴ mais plutôt un élément qui s'intègre au syntagme déterminant du N2 « tête » : le double-marquage est possible car il opère à l'extérieur des frontières du syntagme (en redoublant un SN complet par un autre SN pronom), tandis que la lexicalisation à retardement, qui intervient à l'intérieur des syntagmes (en séparant la détermination de la tête du SN), est exclue.

Les constructions sous-catégorisatrices avec déterminant indéfini ou quantifieur et en position de complément direct peuvent se pronominaliser par *en*, tout comme les constructions quantifiées (*j'ai bu un verre de vin / j'en ai bu un verre, de vin*) :

7- *puis j'ai fait plusieurs sortes de circuits (oral - Train 1)*

7a - *j'en ai fait plusieurs sortes, de circuits*

8- *il y a + un autre genre de carte(s) qui s'appelle la carte Express*

(oral - Carte bleue 6)

8a - *il y en a un autre genre, de carte(s)*

Cette possibilité n'est pas offerte aux exemples avec approximatifs dans les mêmes conditions, ce qui confirme que ces approximatifs ne sont pas des têtes nominales :

3d - **c'en est une espèce, de danse de Saint-Guy*

⁴ Pour une théorie du syntagme nominal, voir Corbett et alii (1993).

**elles en font une espèce, de danse de Saint-Guy*
*4b - *c'en est une sorte, de maquette*
**j'en fais une sorte, de maquette*

2- L'ACCORD.

Les phénomènes d'accord peuvent se manifester soit à l'intérieur du segment [det N1 de N2], soit à l'extérieur, sur un verbe ou un adjectif. Ce dernier type d'accord (accord externe) a déjà été amplement décrit par les grammairiens (à partir de données écrites) et les données orales n'apportent rien de nouveau. Les phénomènes d'accord interne, en revanche, ont été moins étudiés (probablement parce qu'ils sont quasi-inexistants à l'écrit) et c'est un point où l'oral a quelque chose à nous apprendre.

2.1. L'ACCORD EXTERNE.

L'accord d'éléments extérieurs au syntagme avec soit N1, soit N2, donne des indications précieuses pour l'interprétation d'*espèce / genre / sorte*. Logiquement, l'accord avec la « tête » N2 privilégie l'interprétation approximative :

9- nous habitons une sorte de grand domaine à l'intérieur duquel on voyait...
(oral - Souvenirs 4)
*9a - *nous habitons une sorte de grand domaine à l'intérieur de laquelle...*

et l'accord avec la « tête » N1 l'interprétation sous-catégorisatrice :

6- dans le genre de photo que je fais ce que je recherche c'est...
6b - dans le genre de photo que j'ai fait
≠ 6c - dans le genre de photo que j'ai faite (approximatif)

Malheureusement, la plupart de ces accords ne se manifestent pas à l'oral⁵ (ex. 6) et on est obligé de considérer les données écrites si on veut avoir un corpus d'exemples assez important. Quand on le fait, on s'aperçoit que les faits ne sont pas aussi clairs : il semble y avoir une certaine latitude dans les habitudes de rédaction. La même structure en *c'est*, employée par le même auteur, présente une fois l'accord avec N1 et l'autre avec N2 :

⁵ La probabilité d'avoir à la fois une différence de genre ou de nombre entre les deux substantifs et un syntagme verbal ou adjectival où cette différence soit audible est en effet plutôt faible.

10- *Ce n'est pas le genre de réponse qu'il aurait fait l'année dernière.*
(écrit - Beauvoir, *Mandarins*, p. 258)

11- *Ce n'était pas exactement le genre de vacances qu'il avait rêvées.*
(écrit - Beauvoir, *Mandarins*, p. 89)

Faut-il en déduire que *genre* est sous-catégorisateur en 10 et approximatif en 11 ? Même problème pour ces exemples avec relatives branchées sur *une autre sorte*, qui est pourtant une forme clairement sous-catégorisatrice⁶ :

12- *Une autre sorte d'héritiers, qui lira Hegel plus sérieusement, choisira...*
(écrit - Camus, *Homme Révolté*, p. 182)

13- *Mais il donnait lui-même naissance à une autre sorte de nihilistes qui, faisant de l'ennui un principe d'action, identifieront leur suicide avec...*
(écrit - Camus, *Homme Révolté*, p. 182)

Enfin, on trouve quelques exemples isolés qui ne révèlent rien d'autre que les habitudes orthographiques des rédacteurs. Ainsi, on rencontre l'expression *toute(s) sorte(s) de* la plupart du temps au pluriel (43 formes contre 3), et l'accord externe avec le N2 dans l'exemple suivant n'est donc pas si révélateur :

14- *J'entends des plongeurs et des bruits de rames et des rires et toute sorte de gens qui poussent des cris.* (écrit - Claudel, *Poésies Diverses*, p. 831)

2.2. L'ACCORD INTERNE.

Les accords qu'il est intéressant d'observer à l'intérieur du syntagme sont de deux types : l'accord de genre entre le premier déterminant et N1 ou N2 et l'accord de nombre entre [det N1] et N2. Devant l'approximatif *espèce*, on peut observer une variation de genre du déterminant quand le N2 est masculin : *une espèce de* ou *un espèce de* (accord qualifié de *sylléptique* par la grammaire traditionnelle⁷) :

15- *c'était une espèce de tract recto verso* (oral - *Syndicat* 8)

16- *et c'est + c'est une + un espèce de tissu artificiel + qu'on met dans les toitures*
(oral - *Aborigènes* 26)

L'accord sylléptique est très fréquent à l'oral : sur 24 exemples de *espèce de* avec N2 masculin, on a 15 fois *un espèce*. Cependant, ce phénomène ne

⁶ Voir infra, "Les déterminants".

⁷ Voir notamment Grevisse (1986), §422.

semble pas s'étendre de façon significative aux autres approximatifs, ce qui indiquerait que c'est la structure phonique de *espèce* (son initiale vocalique) qui le rend possible. Je n'ai rencontré que deux exemples avec *genre*, dont un où le locuteur se corrige :

17- *c'était une genre de grosse maison à pignons (oral - CR 1993)*

18- *ils avaient un casque + assez plat et une genre et : une espèce de résille dessus tu vois (oral - Louissette 89)*

Au niveau syntaxique, la syllepse s'explique par la nature accessoire, comme greffée après coup, de l'approximatif ; l'accord passe littéralement par-dessus. Elle serait impossible dans l'interprétation sous-catégorisatrice d'*espèce* :

19- *il n'aime qu'une espèce de chat(s), les chats siamois (sous-catég.)*
 ≠ 19a - *il n'aime qu'un espèce de chat, son vieux matou pelé (approximatif)*

Ce phénomène, même s'il est discuté dans Grevisse (1986 : 693), n'est pas du tout représenté dans mon corpus écrit, si ce n'est un exemple où on lit *cet espèce* :

20- *Le troupeau hideux des marchands de biens et de stocks*
Et cet espèce de beau garçon qui se perd dans son époque (écrit - Aragon, p. 242)

L'accord de nombre entre N1 et N2 est aussi révélateur. Comme le suggère déjà l'exemple 19, un changement de nombre est possible entre le sous-catégorisateur et le sous-catégorisé. Il apparaît qu'un tel changement est moins fréquent dans l'interprétation approximative, où N1 et N2 sont le plus souvent tous les deux au pluriel (conformément à la structure syntaxique de l'approximatif)⁸. On peut comparer 21, qui s'interprète spontanément comme approximatif à 21a, qui glisse vers le sous-catégorisateur :

21- *ils avaient tous les mêmes genres de tics c'était affreux (oral - CR 1994)*
 21a - *ils avaient tous le même genre de tics nerveux*

Les données écrites ne vont pas aussi clairement dans ce sens, surtout en ce qui concerne *genre* et *sorte*, les deux lexèmes où on peut voir le plus de différences entre l'oral et l'écrit. *Genre* n'apparaît jamais au pluriel dans les exemples écrits ; *sorte* est au pluriel dans seulement 12 % des cas, alors qu'à l'oral, on le trouve plus souvent au pluriel qu'au singulier (36 formes contre 24). Ceci a pour conséquence qu'à l'écrit, N1 singulier et N2 pluriel se rencontrent souvent dans des structures approximatives :

⁸ Bien sûr, deux N de même nombre n'interdisent pas l'interprétation sous-catégorisatrice.

22- ...serait obtenu de l'extérieur, par une sorte d'actes énumérables qui portent le signe d'une beauté retouchée. (écrit - Bachelard, *Poétique de l'Espace*, p. 106)

23- ...après une sorte d'arrachés violents on tombait tout à coup sur un quatuor à cordes. (écrit - Schaeffer, p. 44)

A l'oral, on aurait plutôt :

22a - par des sortes d'actes énumérables
23a - après des sortes d'arrachés violents

3- LES DÉTERMINANTS.

Voici, présentés en tableaux, les déterminants utilisés devant *espèce / genre / sorte* dans chacune des interprétations et dans les deux corpus.

3.1. CORPUS ORAL.

	approximatif	sous-catégorisateur	indécidable	total
espèce(s)	un(e) 56 des 6 Ø 1 63		cel(te) 6 6	69
genre(s)	un(e) 12 des 2 (tout) le +m. 2 le(s) même(s) 2 les 1 partitif 1 Ø 1 21	le (+modifieur) 3 le même 2 quels 1 un autre 1 ce -là 1 8	ce 44 ? 45	74
sorte(s)	une (+m) 22 des 4 possessif 1 27	toutes 19 numéral 7 plusieurs 4 différentes 1 beaucoup de 1 32	cette 1 1	60
total	111	40	52	203

Les deux exemples avec déterminant \emptyset ne sont pas des exemples d'insultes (qui n'ont pas été retenus dans mon corpus de base) mais des structures avec la préposition *en* :

24- *c'était en genre(s) de bouteille(s) + (Viandox 2)*

L'exemple avec déterminant « ? » est un cas de multi-écoute inutilisable. Les modifieurs (m.) indiqués sont des relatives ou des adjectifs qui portent sur N1 ou N2.

Les déterminants démonstratifs introduisent des énoncés qui sont presque toujours ambigus (colonne « indécidable »)⁹, sauf un cas qui a été classé comme sous-catégorisateur. C'est qu'il est désambiguïsé par la particule *-là* qui apparaît après N1 :

25- *on (n') a pas ce genre-là de cours en France (oral - Erasmus Salamanque 28)*

25a aurait plutôt une interprétation d'approximatif :

25a - *on (n') a pas ce genre de cours-là en France*

Ce n'est pas un critère assez fiable pour en faire un test syntaxique, il faut que la particule soit réalisée dans l'exemple attesté. Comment choisir en effet entre *ce genre-là de choses* et *ce genre de choses-là* ? La différence de sens, s'il y en a une, est complètement neutralisée quand *ce genre* est suivi du lexème *chose* (et c'est la majorité des cas). Pour être systématique, j'ai tout mis dans la colonne « indécidable ».

Les déterminants qui figurent dans les énoncés non ambigus sont assez prévisibles : on a une majorité d'indéfinis dans la colonne des approximatifs et des déterminants « identifiants » dans la colonne des sous-catégorisateurs (quantifieurs, *un autre, le même, quel*).

⁹ Pour une analyse détaillée de cette structure, voir Rouget (à paraître).

3.2. CORPUS ÉCRIT.

	approximatif	sous-catégorisateur	indécidable	total
espèce(s)	une 159 cet(te) 29 la +m. 27 aucune 6 des 5 ces 2 228	une +m. 7 diverses 2 quelle 2 toute autre 1 numéral 1 toutes 1 14	cette 10 toute 8 18	260
genre	le +m. 3 3	le +m. 24 ce 6 un (+m.) 7 quel 4 41	ce 42 le +m. 5 47	91
sorte(s)	une 595 des 7 une certaine 3 ces 1 606	toute(s) 48 la +m. 20 numéral 17 ces 9 quelle(s) 7 aucune 6 une autre 5 plusieurs 1 différentes 1 114	cette 37 37	757
total	837	169	102	1108

Première remarque : il y a plus de déterminants différents dans chaque emploi à l'écrit qu'à l'oral (mais n'oublions pas qu'il y a aussi plus d'exemples écrits au départ). Ils ont en gros les mêmes caractéristiques : indéfinis du côté des approximatifs et « identifiants » du côté des sous-catégorisateurs.

Deuxième remarque : les déterminants démonstratifs, qui introduisent des énoncés indécidables à l'oral, sont parfois désambiguïsés à l'écrit. On rencontre également la forme *le genre + modifieur* répartie dans les trois colonnes. C'est que j'ai tenu compte des indications apportées par les phénomènes d'accord externe dans ce classement ; ainsi, l'exemple

10- *Ce n'est pas le genre de réponse qu'il aurait fait l'année dernière.*

a été automatiquement mis du côté des sous-catégorisateurs, alors que l'exemple

11- *Ce n'était pas exactement le genre de vacances qu'il avait rêvées.*

a été classé comme approximatif. Bien sûr, on peut se poser des questions sur la validité de ce critère, d'autant plus que les auteurs (ou correcteurs) semblent faire un peu n'importe quoi, comme on l'a vu plus haut. Mais c'est un critère formel dont j'ai tenu compte pour le classement des formes écrites.

CONCLUSION.

Cette petite comparaison de *espèce / genre / sorte* à l'oral et à l'écrit révèle des différences significatives entre les deux corpus. On peut voir en premier lieu que les formes sont rarement spécialisées dans le même emploi : *genre* est plutôt approximatif à l'oral alors qu'il est presque toujours sous-catégorisateur à l'écrit ; c'est l'inverse pour *sorte*, plus souvent sous-catégorisateur qu'approximatif à l'oral mais approximatif vedette de l'écrit. On peut voir également qu'à l'écrit, malgré une plus grande variété apparente, il existe des déterminants de prédilection qui rassemblent jusqu'à 90 % des occurrences de chaque type (*une espèce de, une sorte de, ce genre de, toutes sortes de*), ce qui rend finalement l'oral plus varié et plus riche. L'analyse syntaxique montre enfin que même si les phénomènes d'accord externes ne sont pas utilisables à l'oral (car ils sont pour la plupart inaudibles), les phénomènes d'accord internes, eux, y sont plus représentés qu'à l'écrit. De plus, ces derniers semblent constituer des indices plus sûrs pour l'analyse syntaxique car ils sont nettement plus systématiques. Tout ceci confirme l'idée qu'il faut utiliser les données de l'oral, et pas seulement les données écrites, si on veut faire une description détaillée d'un point de syntaxe aussi subtil que la frontière entre approximatifs et sous-catégorisateurs.

Christine ROUGET

Références

BLANCHE-BENVENISTE, Claire et alii., 1990, *Le français parlé - Études grammaticales*, Paris : Éditions du C.N.R.S.

- CORBETT GREVILLE, G., FRASER NORMAN M. ET MCGLASHAN SCOTT, 1993, *Heads in grammatical theory*, Cambridge : Cambridge University Press.
- CORBETT GREVILLE, G. , 1983, *Hierarchies, targets and controllers : agreement patterns in Slavic*, London : Croom Helm.
- GREVISSE, Maurice , 1986, *Le bon usage*, 12e édition, Gembloux : Duculot.
- ROUGET, Christine , 1989, *Application de l'approche pronominale à la syntaxe du nom, les formes N1 de N2*, Université de Provence, thèse de doctorat non publiée.
- ROUGET, Christine (à paraître) « espèce de, genre de, sorte de : approximatifs ou sous-catégorisateurs ? », *Actes du 7e colloque de Psychomécanique du Langage*.

Imprimé sur les Presses



Dépôt légal - 1^{er} Trimestre 1997
ISBN - N° 2-85399-400-7

