

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 modeler 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.