

Université de Neuchâtel
Faculté des Lettres et Sciences humaines

Mémoire de licence en linguistique française (filière français moderne)
Sous la direction de Marie-José Béguelin

**Traiter la diversité des formes interrogatives en français :
apports et limites de l'approche variationniste**

Géraldine Zumwald
Juin 2010

Table des matières

Avant-propos	4
1 Variation et variationnisme	6
1.1 Le concept de <i>variable sociolinguistique</i>	6
1.1.1 Naissance du concept de <i>variable sociolinguistique</i> en phonologie.....	6
1.1.2 Elargissement de la <i>variable</i> aux autres niveaux de la langue.....	7
1.2 Les débats sur l'extension de la variable phonologique à la syntaxe.....	9
1.2.1 La nature des variables/ typologie	9
1.2.2 La question du sens	12
1.2.2.1 <i>Equivalence sémantique</i>	14
1.2.2.2 <i>Equivalence fonctionnelle</i>	15
1.2.3 Les facteurs qui influencent la variation.....	17
1.2.3.1 <i>Facteurs sociaux, stylistiques et linguistiques</i>	17
1.2.3.2 <i>Facteurs pragmatiques</i>	19
1.3 Explications de la variation	21
1.3.1 Etendue de la variation : où et comment se manifeste-t-elle ?	21
1.3.2 Considérations terminologiques.....	24
2 Le domaine de la question	26
2.1 Généralités	26
2.1.1 La forme interrogative et l'acte de langage de question	26
2.1.2 Caractéristiques formelles.....	27
2.1.2.1 <i>Questions totales et partielles</i>	27
2.1.2.2 <i>Les structures interrogatives</i>	27
2.1.3 Valeur illocutoire	29
2.1.4 Détermination de l'objet d'étude	32
2.2 Etudes quantitatives	33
2.2.1 Présentation.....	33
2.2.2 Traitement et classement des données	33
2.2.2.1 <i>Questions totales</i>	34
2.2.2.2 <i>Questions partielles</i>	35
2.3 Les études variationnistes.....	41
2.3.1 Présentation.....	41
2.3.1.1 <i>Déroulement général</i>	42
2.3.1.2 <i>Le corpus de Coveney</i>	42

2.3.1.3 Le corpus d'Elsig	43
2.3.2 Etablissement des catégories de variables et de variantes	43
2.3.3 Contextes variables et contextes catégoriques	46
2.3.3.1 Critères de sélection des variantes	46
2.3.3.2 Questions totales catégoriques	49
2.3.3.3 Questions partielles catégoriques	52
2.3.4 Facteurs variables influençant la variation	53
2.3.4.1 Traitement informatique/statistique	53
2.3.4.2 Choix des facteurs	55
2.3.4.3 Attribution des fonctions communicatives	55
2.3.4.4 Facteurs linguistiques	59
2.3.4.5 Facteurs pragmatiques	61
2.3.4.6 Facteurs socio-stylistiques	63
2.3.4.7 Interactions entre facteurs	66
3 Synthèse critique.....	68
3.1 Difficultés méthodologiques liées aux études variationnistes (et quantitatives) sur les questions.....	68
3.1.1 Définition de <i>la question</i>	68
3.1.2 La dimension interprétative des études variatonnistes.....	69
3.1.3 Mesurer les causes de la variation	71
3.2 Objections au traitement variationniste des questions	73
3.2.1 Les particularités de chaque forme	73
3.2.2 La notion d'équivalence.....	76
3.3 Applications du cadre théorique variationniste aux questions	77
3.4 Limites des études variationnistes et suggestions alternatives	79
3.4.1 Les structures non traitées.....	79
3.4.2 La notion de « stratégie ».....	81
3.4.3 Les questions dans différents registres de langue	82
Conclusion	85
Bibliographie	87

Avant-propos

La diversité des structures syntaxiques disponibles pour former une question en français constitue une difficulté pour les apprenants de français langue étrangère. Les francophones, eux, font spontanément usage d'une grande variété de formes interrogatives dans leurs interactions quotidiennes. Pourtant, au moment de réfléchir sur l'emploi des différentes formes de la question, il n'est pas aisé, pour le locuteur francophone, d'expliquer les raisons de ses choix. En effet, que dire des différences d'usage ou de sens (s'il en est) entre les trois phrases suivantes ? :

Est-ce qu'il est parti ? Il est parti ? Est-il parti ?

Comment expliquer la valeur et l'emploi de la particule *est-ce que* ? Enfin, dans les exemples ci-dessous, pourquoi les trois premières phrases sont-elles acceptables alors que la quatrième ne l'est pas ? :

*Où finit le voyage ? Quand finit le voyage ? Comment finit le voyage ? *Pourquoi finit le voyage ?*

Même pour un locuteur natif, ces questions ne sont pas faciles à élucider. Nous savons qu'au moment de poser une question, les locuteurs du français ont le choix entre plusieurs structures syntaxiques susceptibles de combiner des éléments linguistiques identiques. Mais comment ces choix sont-ils effectués ? Dans quelle mesure sont-ils arbitraires ou au contraire motivés ou contraints ? Qu'est-ce qui influence l'utilisation de l'une ou l'autre forme interrogative ? Plutôt que de répondre exhaustivement à ces questions, ce travail a pour but de présenter différentes manières d'analyser les formes interrogatives et leur emploi en français. Plus particulièrement, nous chercherons à déterminer dans quelles mesures la méthodologie variationniste peut constituer une approche pertinente de la diversité des formes interrogatives.

Le domaine de la question a fait l'objet de nombreuses études en linguistique. Une certaine partie d'entre elles¹ s'intéressent aux différentes structures syntaxiques de l'interrogation et à ce qui peut influencer leur usage. Les facteurs susceptibles de privilégier l'utilisation de l'une ou l'autre forme de question sont nombreux ; en guise d'introduction et sans entrer dans les détails, nous pouvons citer :

- Les facteurs stylistiques (situation d'interaction et relation entre les locuteurs)
- Les facteurs sociaux (identité et caractéristiques sociales du locuteur)
- Les facteurs linguistiques (nature des éléments linguistiques de la phrase)
- Les facteurs sémantico-pragmatiques (sens et fonction de la question)
- Les facteurs géographiques (région de la francophonie)

Cette dernière catégorie de facteurs sera peu traitée dans ce travail, puisque nous nous concentrerons sur l'usage des structures interrogatives à un endroit donné et en un moment donné.

¹ Voir par exemple Renchon (1967) et Al (1975) pour des approches principalement syntaxiques, Borillo (1978) pour une approche syntaxique et fonctionnelle, Quillard (2001) pour une approche sociolinguistique et fonctionnelle.

L'étude du rôle des facteurs sociaux et stylistiques intéresse particulièrement les sociolinguistes, certains d'entre eux considérant les différentes formes de question comme un cas de variation syntaxique. Mais la méthode variationniste est-elle un bon moyen d'appréhender la diversité formelle des interrogations ? Ce n'est évidemment pas l'avis de tout le monde puisque la notion même de « variation syntaxique » reste controversée depuis son apparition dans les années 60². En effet, la *variable sociolinguistique*, principal outil du variationnisme, a été conçue pour la phonologie et son extension à d'autres niveaux de langue fait l'objet de contestations. Afin d'éclairer les enjeux de cette question litigieuse, ce travail débutera par une brève rétrospective partant de la naissance de la variable sociolinguistique (en phonologie) pour déboucher sur les débats relatifs au concept de *variable syntaxique*. Enfin, pour situer le variationnisme dans un contexte plus large, le point 1 se terminera par un survol des différentes théories de la variation dans la langue.

Ce n'est qu'après avoir exposé les enjeux et les méthodes du variationnisme que nous aborderons le thème de l'interrogation et du traitement de ses différentes formes. Une introduction générale à la problématique de *la question* (point 2.1) permettra d'aborder globalement ce vaste domaine de la langue et de fixer certains concepts utiles à notre travail. Enfin, nous arriverons au cœur de notre sujet avec l'examen de différentes études portant sur l'emploi des formes de l'interrogation en français.

Le point 2.2 sera consacré à trois études quantitatives³ menées dans les années 1960-1970, sur des corpus tirés de la langue orale. Ces recherches, portant sur l'usage réel de la langue, c'est-à-dire sur des données attestées, donnent un compte rendu quantitatif de l'utilisation des différentes formes de la question ; parallèlement, et dans une certaine mesure, elles s'intéressent aux raisons de l'emploi de ces différentes formes, en recourant principalement à des explications d'ordre linguistique. Cette volonté d'expliquer les fréquences et les conditions d'utilisation des formes interrogatives a récemment motivé Aidan Coveney (2002) et Martin Elsig (2009) à mener deux grandes recherches sur l'interrogation, à l'intérieur du cadre théorique variationniste. Bien que les moyens utilisés par les études quantitatives diffèrent passablement de ceux des études variationnistes, l'influence des premières sur les secondes est notable ; Coveney et Elsig eux-mêmes se posent en successeurs des auteurs d'études quantitatives traditionnelles. Malgré leur orientation sociolinguistique, nous verrons au point 2.3 que les études variationnistes sont loin de se limiter à l'examen du rôle des facteurs stylistiques et sociaux. Le compte rendu détaillé des études de Coveney et d'Elsig nous permettra d'apprécier l'application des méthodes variationnistes sur le cas des formes de la question en français.

Finalement, nous reviendrons au point 3 sur les fondements théoriques et les applications du concept de variable sociolinguistique au domaine de la question. Nous réfléchirons aux apports et aux limites des approches quantitatives et variationnistes dans leur tentative d'expliquer l'usage des différentes formes de question.

² Des critiques ont été formulées, par exemple, par Lavandera (1978) et Romaine (1984).

³ Pohl (1965), Söll (1983) et Behnstedt (1973)

1 Variation et variationnisme

1.1 Le concept de *variable sociolinguistique*

1.1.1 Naissance du concept de *variable sociolinguistique* en phonologie

Dans un ouvrage publié en 1972⁴, William Labov désigne sa thèse de maîtrise⁵ comme sa « première étude de l'action des structures sociales dans le changement linguistique » (1976, 48). Si le terme « changement linguistique », renvoie à une perspective diachronique, c'est surtout l'emploi synchronique de la langue que Labov analyse dans sa thèse menée sur l'île de Martha's Vineyard. En fait, il veut montrer que tout changement linguistique observable en diachronie tire son origine d'un phénomène de variation synchronique. Dans cette première étude, Labov limite ses recherches au domaine de la phonologie : il observe donc les différentes prononciations de certains phonèmes, sur la durée et en un moment donné. Si l'étude des liens entre variation synchronique et changement diachronique représente une partie importante du travail de Labov, ce sont principalement ses découvertes sur la structure de la variation synchronique qui nous intéresseront et dont nous rendrons compte dans ce chapitre.

En phonologie, il est établi qu'un phonème peut être prononcé de différentes manières : on parle alors de ses différentes *réalisations*. Un phonème dont la prononciation varie est considéré comme une *variable linguistique* et ses différentes réalisations en constituent les *variantes*. Avant Labov, comme l'explique Danièle Godard,

les structuralistes avaient mis à jour l'existence de deux types de variables, celles dont la réalisation est conditionnée par l'environnement linguistique, et celles qui sont dites « libres » » (1992, 51).

On pensait alors que l'utilisation des variantes libres n'était pas motivée et les études phonologiques n'y accordaient que peu d'importance. A Martha's Vineyard, Labov s'est intéressé aux phonèmes de l'anglais /ay/ et /aw/ qui connaissent chacun plusieurs réalisations différentes. On considérait généralement que l'alternance entre ces différentes réalisations relevait de l'arbitraire et qu'il s'agissait donc d'un cas de variation libre. Mais Labov a cherché à comprendre pourquoi la prononciation de ces deux phonèmes changeait d'un locuteur à l'autre, à l'intérieur d'une même communauté linguistique. Partant d'une démarche sociolinguistique, il s'est intéressé aux données sociales de l'île et de ses habitants. Il a donc comparé les productions linguistiques des locuteurs avec leurs caractéristiques sociales pour constater que les habitants de l'île les plus attachés à leur territoire et à leurs traditions étaient ceux dont la prononciation de /ay/ et /aw/ était le plus souvent centralisée.

Pour mettre au jour de tels phénomènes, Labov a appliqué des méthodes quantitatives sur des productions linguistiques orales attestées provenant de divers informateurs. Chaque informateur a été classé selon des critères sociaux et toutes les occurrences des variables étudiées ont été comptées. Labov précise qu'avant de mesurer l'influence des facteurs sociaux sur l'utilisation des variantes, il faut considérer l'environnement linguistique dans lequel

⁴ L'ouvrage original est *Sociolinguistic Patterns* (1972) mais nous nous référerons dans ce travail à la traduction française de 1976 (*Sociolinguistique*).

⁵ « Les motivations sociales d'un changement phonétique », thèse de maîtrise (MA thesis), publiée la première fois dans *Word*, en 1963.

apparaît la variable. En effet, la nature du phonème qui suit ou précède directement la variable étudiée, le fait qu'elle soit placée sous un accent ou encore le type de lexème dans lequel elle apparaît sont autant de facteurs qui peuvent influencer ou contraindre l'utilisation d'une variante. Par exemple, Labov a remarqué que certaines catégories de consonnes subséquentes aux phonèmes /ay/ et /aw/ favorisaient leur centralisation. Ce n'est qu'après avoir pris en compte l'influence de l'environnement linguistique que l'on peut comparer la fréquence d'utilisation des variantes avec les caractéristiques sociales de l'informateur. Enfin, en étendant la comparaison à l'ensemble des informateurs, des affinités peuvent apparaître (de manière plus ou moins marquée) entre certains groupes sociaux et l'utilisation de l'une ou l'autre variante.

Dans sa deuxième étude,⁶ Labov a continué ses recherches au niveau de la variation phonologique, mais il s'est basé cette fois-ci sur l'anglais parlé dans la ville de New York. Il a observé le fonctionnement de nombreuses variables phonologiques, comme celui du phonème /r/, dont il a analysé les différentes réalisations (dans ce cas, présence ou absence) en position postvocalique comme dans *car* ou *fourth*. Pour ce faire, il a interrogé anonymement les employés de trois grands magasins new-yorkais de différents standings. Il a constaté que la prononciation du /r/ postvocalique était unanimement valorisée par les membres de la communauté linguistique new-yorkaise (Labov parle alors de « variante de prestige »). Il a ensuite montré que la distribution de cette variante de prestige était corrélée à la classe socio-économique de la clientèle visée par le magasin d'où provenaient les données. Mais cette fois-ci, Labov ne s'est pas limité à mesurer l'influence des facteurs sociaux sur le choix des variantes. En effet, il a remarqué que l'emploi de l'une ou l'autre des variantes d'une variable ne dépendait pas seulement de son environnement linguistique et des caractéristiques sociales du locuteur mais aussi du contexte stylistique dans lequel elle apparaissait. Selon lui, « les divers styles de discours envisagés se rangent le long d'une dimension unique, l'attention portée à la parole (...) » (1976, 161). Afin d'obtenir deux styles différents dans les magasins new-yorkais, Labov a différencié les énoncés donnés comme première réponse à sa question des énoncés répétés après qu'il eut feint de ne pas avoir compris. Il obtint ainsi un deuxième énoncé identique, mais « emphatique et prononcé avec soin » (1976, 101). En comparant les deux énoncés, Labov a remarqué que la variante de prestige apparaissait plus facilement lors de la deuxième réponse à la question, lorsque *l'attention portée à la parole* était accrue.

Labov fut l'un des premiers à exposer des corrélations systématiques entre phénomènes de variation phonologique et facteurs sociaux ou stylistiques. Les variantes *libres* étaient considérées comme arbitraires et donc insignifiantes, jusqu'à ce que Labov dévoile le pouvoir conditionnant de certains facteurs sociaux et stylistiques. C'est alors que l'on commença à parler de variables *sociolinguistiques* en référence à l'influence des facteurs sociaux sur leur utilisation.

1.1.2 Elargissement de la *variable* aux autres niveaux de la langue

A la suite de ses deux premiers travaux consacrés aux liens entre facteurs socio-stylistiques et variation phonologique, Labov continua de s'intéresser aux rapports entre langue et société tout en élargissant son domaine de recherche. Comme le soulignent James et Lesley Milroy, un des points forts du travail de Labov fut de montrer que la variabilité dans la langue peut être structurée (2000). Or, la variabilité se manifeste aussi bien aux niveaux morphologique et syntaxique que phonologique et Labov a rapidement tenté d'étendre ses

⁶ *The Social Stratification of English in New York City*, thèse de doctorat publiée en 1966

méthodes d'analyse de la variation hors de la phonologie. En effet, il pensait pouvoir y déceler un conditionnement socio-stylistique comparable à celui qu'il avait établi en phonologie. « En congruence avec le postulat structuraliste d'homologie des niveaux », Labov a donc transposé le concept de variable sociolinguistique à d'autres niveaux de la langue (Gadet, 1992, 12).

C'est dans le quartier new-yorkais de Harlem que Labov continua ses recherches sur la structure de la variation linguistique. Il y choisit comme objet d'étude l'anglais vernaculaire noir américain dont il examina des phénomènes de variation tant au niveau phonologique que morphologique. En morphologie, il a par exemple étudié l'effacement de la copule au présent:

- (1) She is out of the game.
- (2) She out of the game. (Labov, 1972, 311)

Selon lui, ces deux exemples constituent deux variantes morphologiques dont la distribution peut être influencée aussi bien par des facteurs linguistiques que socio-stylistiques.

Dans la préface à l'ouvrage *Sociolinguistique* de Labov (1976), Pierre Encrevé souligne le contraste entre les deux premières études de Labov évoquées plus haut et celle de Harlem :

Le premier Labov ne travaillait guère la structure linguistique : il établissait les covariations, prouvait l'existence de variations régulières, mais n'en tirait pas de conséquences quant à la structure des systèmes linguistiques. En même temps que celui de « vernaculaire » pour désigner l'objet de l'enquête, le second Labov introduit un autre terme, désignant l'objet de l'analyse linguistique : « la grammaire ». (1976, 26)

On comprend donc que l'extension de la *variable sociolinguistique* s'inscrit dans un changement d'orientation général des travaux de Labov, désormais enclin à décrire et expliquer la structure de la langue, soit-elle phonologique ou syntaxique. Son intérêt pour la grammaire, combiné à son attrait pour la stratification sociale de la langue se matérialisent dans l'élaboration du concept de « règles variables ». Alors que la grammaire générative distingue les règles obligatoires et les règles optionnelles, Labov introduit le concept de règles variables pour tenter d'expliquer dans quelles mesures et pour quelles raisons (extra ou intralinguistiques) les règles optionnelles sont appliquées ou pas.⁷ Comme l'explique Françoise Gadet, on peut distinguer « plusieurs étapes dans le processus qui conduit progressivement Labov à la généralisation du traitement de la variation au moyen de règles variables » (1992, 12). Si l'article de Harlem constitue la première extension hors de la phonologie, il faut attendre 1972 pour qu'elle arrive au niveau de la syntaxe. C'est alors que Labov, toujours à Harlem, décide d'étendre l'analyse de la variation à des structures syntaxiques. Par exemple, il considère que les deux phrases négatives (3) et (4) sont les variantes d'une même variable, la première étant la variante vernaculaire et la seconde la variante standard :

- (3) He don't like nobody that went to no prep school.

⁷ Cedergren et Sankoff (1974) ont ensuite créé un programme qui permet de mesurer statistiquement le poids de chaque règle variable sur le choix des variantes. Il existe aujourd'hui plusieurs logiciels d'analyse en règles variables, mais, comme le signale Coveney (2002, 35), leur emploi n'implique pas nécessairement que l'on analyse l'application de règles présentes dans la grammaire générative. Selon lui, ces logiciels peuvent être utilisés comme de simples outils statistiques, pour mesurer tout type d'alternances dans la langue, et ceci que l'on adhère ou non au concept de règles variables.

(4) He doesn't like anybody that went to any prep school. (Labov, 1976, 322)

Le type d'approche de la variation mise en place par Labov a reçu plusieurs appellations différentes : *l'école labovienne*, *la théorie variationniste* ou *corrélationaliste*, *la sociolinguistique quantitative* sont autant de noms servant à désigner les travaux basés sur le concept de variable sociolinguistique⁸. Nous l'avons vu, la caractéristique majeure de ces travaux est d'établir des liens entre une variable (soit un point de variation de la langue) et les facteurs socio-stylistiques⁹ qui la structurent. Les données linguistiques sont toujours fournies par des énoncés attestés et les méthodes d'analyse sont de nature quantitative et statistique. Si Labov affirmait en 1972 que le but de ses travaux était d'étudier « les modes possibles d'interaction entre l'étude du langage et celle de la société » (1976, 175), on considère aujourd'hui que le variationnisme se dirige plus vers l'analyse de la structure linguistique que vers celle de la société. En témoignent les propos de Lavandera et Coveney, la première définissant le variationnisme comme « une approche consacrée à la théorie grammaticale » (1988, 4 notre traduction) et le second affirmant qu'il s'agit d'une approche orientée vers le système linguistique plutôt que vers le locuteur (2002).

1.2 Les débats sur l'extension de la variable phonologique à la syntaxe

D'une manière générale, la notion de *variable sociolinguistique* et les découvertes de Labov sur la structuration extralinguistique de la variation furent bien accueillies au niveau de la phonologie. En revanche, l'extension de la variable sociolinguistique à des niveaux autres que phonologiques fut reçue avec un enthousiasme mitigé. D'un côté, elle fut acceptée sans hésitations et utilisée de suite pour l'étude de cas de variation non phonologique ; c'est le cas par exemple de Gillian Sankoff, qui traita des phénomènes de variation syntaxique à travers les règles variables créées par Labov dans un article paru en 1973.¹⁰ Mais d'un autre côté, de nombreuses voix contestataires commencèrent à s'élever suite à un article de Beatriz Lavandera dénonçant l'inadéquation du concept de variable sociolinguistique pour l'analyse de la variation non phonologique. Un demi siècle plus tard, le débat sur la légitimité de l'utilisation de la méthode variationniste au-delà de la phonologie divise encore les linguistes. Le cœur du débat est constitué par la question de la possibilité d'équivalence sémantique entre deux structures (morpho)syntaxiques différentes. En effet, Labov a établi que, pour que deux segments linguistiques puissent être traités comme les variantes d'une variable, ils devaient vouloir « dire la même chose » (1976, 366). Si la question ne se pose pas au niveau de la phonologie (les phonèmes n'impliquant pas de sens), elle devient essentielle lorsqu'on élargit la variable sociolinguistique à d'autres niveaux de langue. Cependant, l'extension de la variable phonologique à la syntaxe soulève d'autres problèmes comme celui de définir les unités linguistiques susceptibles de constituer la variable ainsi que les facteurs qui entrent en compte dans la sélection des variantes.

1.2.1 La nature des variables/ typologie

En phonologie, la variable et ses variantes s'organisent toujours selon le même schéma : la variable coïncide avec un certain phonème et ses variantes correspondent aux différentes prononciations (ou à l'absence) de ce phonème. Si l'étude de la variation phonologique connaît une telle constance d'organisation, c'est que la phonologie est un domaine de la langue doté d'une grande cohérence interne. En effet, le système phonologique

⁸ Termes donnés dans *Sociolinguistics, the essential readings*, (2003), partie 5 : « Language and variation », introduction

⁹ ou autres facteurs (cf. point 1.2.3)

¹⁰ « Above and beyond phonology in variable rules »

d'une langue donnée est composé d'un nombre restreint de phonèmes définis oppositivement les uns par rapport aux autres par leurs différences fonctionnelles. La liste des variables phonologiques disponibles est donc limitée et clairement définie. Quant à leurs variantes, s'agissant des réalisations d'un phonème, elles sont certes plus nombreuses et moins nettement identifiables que les phonèmes eux-mêmes. Cependant, les variantes d'une variable phonologique trouvent une certaine unité dans leur nature commune (dans tous les cas, il s'agit de *sons*, non porteurs de sens).

Une des premières questions qui se pose lorsqu'on examine la possibilité de l'élargissement de la variable phonologique est celle des unités linguistiques qui vont constituer les variantes et les variables non-phonologiques. Si, en phonologie, les unités susceptibles de varier se limitent au nombre de phonèmes constitutifs du système phonologique de la langue étudiée, il n'en va pas de même pour les autres niveaux de la langue. Il suffit de passer de la phonologie à la morphologie pour voir apparaître toutes sortes de variables morphologiques potentielles, de différentes natures. En comparaison aux phonèmes, les morphèmes sont plus difficiles à définir oppositivement les uns par rapport aux autres. De plus, ils peuvent appartenir à diverses catégories (pronom, particule, terminaison, etc.) et remplir diverses fonctions grammaticales (sujet, complément, etc.). Ainsi, même si l'on laisse de côté les questions sémantiques pour se concentrer sur des considérations formelles, on constate que l'identification de ce qui peut constituer une variable et ses variantes est plus complexe en morphologie qu'en phonologie. Mais le degré de complexité s'accroît encore lorsque l'on passe au domaine de la syntaxe et que les variantes en alternance impliquent différentes structures syntaxiques.

Bien que l'on puisse établir, grossièrement, une échelle de complexité croissante entre les niveaux phonologique, morphologique et syntaxique, il n'est pas toujours possible de définir des frontières nettes entre ces trois domaines (en témoigne l'existence des termes *morphophonologie* et *morphosyntaxe*). On peut donc se demander s'il est vraiment pertinent de classer les différentes variables selon le type d'élément (phonème/morphème/structure syntaxique) qui varie. Pour répondre à cette question, nous nous baserons sur les réflexions que Suzanne Romaine et Françoise Gadet ont menées sur le sujet. Toutes deux proposent une classification des variables (ou phénomènes de variation)¹¹ qui tiennent compte de la porosité entre les différents niveaux de langue.

Dans un article intitulé « On the problem of syntactic variation », Romaine a établi une typologie des variables basée sur les éléments qui varient et les facteurs qui peuvent influencer ou contraindre leur utilisation (1984). Elle appelle *purement phonologiques* (« *pure* » *phonological*) les variables qui ne peuvent être conditionnées que par leur environnement phonologique, comme le /r/ postvocalique en anglais dont la réalisation dépend uniquement de la nature (consonne ou voyelle) du segment phonétique qui le suit. Selon Romaine, il ne faut pas confondre ce type de variables avec celles qui subissent l'influence de facteurs grammaticaux en plus de facteurs phonologiques. Elle qualifie cette deuxième catégorie de variables de *morphophonémique* et en donne comme exemple, toujours pour l'anglais, l'effacement des consonnes *t* et *d* en fin de mots ; la prononciation de ces consonnes finales dépend non seulement de la nature phonologique du segment subséquent, mais aussi de la nature monomorphémique ou bimorphémique du mot dans lequel elles apparaissent. La variable *morpho-syntaxique* ou *morpho-lexicale* constitue le troisième type de variable répertorié par Romaine. Comme la variable morphophonémique, elle peut être contrainte par des facteurs phonologiques et grammaticaux, mais elle implique des

¹¹ Romaine parle de types de *variables* ; Gadet de types de *variation*.

morphèmes qui comprennent plus d'un phonème. Il s'agit par exemple de la présence ou absence de *que* dans des phrases du type :

(5) Je pense (que) ça a été plutôt un snobisme. (G. Sankoff, 1973, 54)

Romaine pense qu'il est important de distinguer les variables morpho-syntaxiques, qui sont des cas de présence/absence, des variables *purement syntaxiques* (quatrième catégorie de variables, « *pure* » *syntactic*). Pour elle, seules les variables *purement syntaxiques* présentent une alternance de constructions syntaxiques, par exemple entre une structure active et une structure passive ou entre les auxiliaires *avoir* et *être* en français.

Romaine regrette que le troisième et le quatrième type de variable soient souvent confondus sous la dénomination de « variable syntaxique ». Pour elle, la distinction est d'autant plus importante qu'elle estime que les variables morpho-syntaxiques peuvent être conditionnées par des facteurs sociaux, au contraire des variables *purement syntaxiques*. Selon sa typologie, seules les trois premières catégories de variables peuvent être investies par des facteurs sociaux, et donc traitées comme des variables sociolinguistiques. En effet, Romaine pense que la variable *purement syntaxique* présente des questions d'équivalence sémantique trop différentes des trois autres types de variables pour que l'on puisse la traiter au travers d'un même outil théorique.

Dans « La variation, plus qu'une écume » (1997b), Gadet distingue trois grandes catégories de variation : phonologique, morphologique et syntaxique. Comme Romaine, elle remarque que certains cas de variation phonologique sont conditionnés par des facteurs phonologiques et grammaticaux, comme c'est le cas en français pour les phénomènes de liaison. Concernant la variation morphologique, Gadet différencie les cas de présence/absence des cas d'alternances. Elle considère que la présence ou absence d'un élément comme le *ne* de négation est un phénomène proche de la phonologie et donc sans incidence sur le sens d'un énoncé. Quant aux alternances, elles les classe en deux catégories : les *variantes locales* et les *variantes à effet sur le système*. Pour illustrer ces deux types d'alternance, elle donne l'exemple du pronom sujet *elles* : lorsque l'alternance porte sur la prononciation ([e] au lieu de [el]), il s'agit d'une variante locale qui n'a de conséquence qu'au niveau phonologique. Par contre, lorsque le pronom apparaît sous la forme neutralisée [il], il s'agit d'une variante à effet sur le système (ici changement de genre). Dans le domaine de la variation syntaxique, Gadet distingue les alternances qui impliquent toute une phrase (par exemple, les structures interrogatives) de celles qui ne touchent que des segments plus courts (par exemple, les relatives).

Gadet ne se prononce pas aussi clairement que Romaine sur les types de variation susceptibles d'être conditionnés par des facteurs sociaux. Cependant, elle affirme que les alternances *à effet sur le système* (en morphologie), de même que les alternances syntaxiques, sont fortement susceptibles de modifier le sens d'un énoncé, ce qui n'est pas compatible avec un traitement variationniste.

Même si les classes établies par Romaine et Gadet ne coïncident pas entièrement, elles présentent des points communs utiles pour distinguer certains types de variables ou de variation. Premièrement, on remarque qu'il existe deux manifestations différentes des phénomènes de variation, ces derniers apparaissant soit sous la forme de *présence/absence*, soit sous la forme d'*alternance(s)*. Ces deux formes de variation peuvent apparaître aussi bien au niveau phonologique que morphologique. Par contre, lorsque l'on passe au niveau de la syntaxe (ou de ce que Romaine appelle la *variable purement syntaxique*), on ne rencontre plus que des alternances. Romaine et Gadet s'entendent sur le fait que, lorsqu'on sort du domaine

de la phonologie, les phénomènes de variation qui se mesurent en terme de présence/absence sont moins susceptibles d'altérer le sens d'un énoncé que ceux qui impliquent des alternances. Un autre point qui ressort de la comparaison des observations de Romaine et Gadet est le manque d'uniformité de la variation à l'intérieur de chaque niveau de langue (phonologique, morphologique ou syntaxique). Ainsi, lorsque l'on parle de variable morphologique, il peut s'agir de la présence ou absence d'un morphème (par exemple le *ne* de négation), comparable à un phénomène de variation phonologique dans la mesure où il n'altère en rien l'organisation grammaticale ou sémantique de la phrase. Mais il peut tout aussi bien s'agir d'une alternance entre deux ou plusieurs morphèmes appartenant à différentes catégories grammaticales et susceptibles d'apporter des nuances sémantiques à la phrase ou au discours. Des divergences de nature tout aussi significatives peuvent être rencontrées parmi différents phénomènes de variation dite « syntaxique ».

Au vu de ce qui précède, il faudrait donc éviter de donner trop d'importance aux qualificatifs (morphologique/syntaxique) que donnent les chercheurs aux différents phénomènes de variation qu'ils observent. On retiendra plutôt la différence entre les phénomènes de présence/absence et ceux d'alternance, les premiers étant apparemment moins susceptibles d'altérer le sens d'un énoncé. Les alternances présentent également plus de complexité lorsqu'il s'agit de déterminer les limites formelles de la variable et de ses variantes. Ainsi, Gadet (1997a) constate que la grande diversité de structures relatives du français, surtout à l'oral, rend la classification formelle en variables et variantes très ardue : si la plupart des relatives se classent dans les catégories « standard », « résomptive », « réduite » ou encore « pléonastique », certaines, comme (6) ne correspondent à aucun type :

(6) J'ai touché la voiture où la personne était dedans c'est la personne qui va chanter maintenant. (Gadet, 1997a, 13)

Pour Gadet, le variationnisme n'est pas la méthode adéquate pour traiter des cas d'alternances aussi complexes que les relatives en français. D'ailleurs, dans les faits, elle remarque que la plupart des travaux variationnistes traitent de phénomènes proches de la morphologie.

1.2.2. La question du sens

Le concept de variable sociolinguistique implique que toutes les variantes concurrentes partagent le même sens. Cette condition est remplie d'office pour les variables phonologiques puisqu'un phonème n'est lié à aucune signification. Cela dit, toutes les unités linguistiques supérieures aux phonèmes sont porteuses de sens, si bien que la question de l'équivalence des variantes se pose inévitablement dès que l'on sort du domaine de la phonologie. Il s'agit certainement de la question la plus litigieuse des débats sur l'extension de la variable, car elle touche au problème controversé de la possibilité d'une équivalence de sens entre deux phrases ou deux énoncés différents. Comme le remarque Claire Blanche-Benveniste, les débats sur la variation syntaxique ressemblent parfois « aux anciens débats sur la synonymie », même si, en principe, la variation syntaxique ne fait pas intervenir des éléments lexicaux mais grammaticaux (1997a, 19). En effet, selon la définition de Godard, les variantes syntaxiques ne se différencient que par leur structure :

[...] ce sont deux séquences ayant le même sens et comportant les mêmes items lexicaux, mais non les mêmes suites ou les mêmes hiérarchies de catégories. (1992, 56)

Le problème qui se pose alors, est celui de l'incompatibilité apparente entre le traitement variationniste de la syntaxe et « l'hypothèse de compositionnalité », que Godard définit ainsi :

[...] le sens d'une expression complexe se construit par la combinaison du sens des items lexicaux selon le plan fourni par la syntaxe. (1992, 56)

Cette hypothèse, largement répandue, est décrite par le sémanticien John Lyons comme étant « absolument centrale en sémantique formelle moderne » (1995, 112, notre traduction). Mais doit-on vraiment en déduire, comme le fait Godard, que toute modification de structure syntaxique entraîne nécessairement une modification de *sens* ? Avant de considérer la possibilité d'une équivalence sémantique entre deux variantes syntaxiques, il faut s'interroger sur la portée même du mot *sens*. Comme l'explique Donald Winford, « Les sociolinguistes n'ont pas toujours précisé la base exacte sur laquelle devait reposer l'équivalence sémantique, ce qui a provoqué une certaine controverse » (1996, 184, notre traduction). En effet, la sémantique moderne distingue différents types de sens et la question de savoir lequel est approprié à l'étude de la variation syntaxique a fait l'objet de beaucoup de discussions. Avant d'exposer les positions de divers linguistes sur le sujet, nous introduirons brièvement les différents types de sens communément distingués par la sémantique moderne, tels qu'ils sont présentés par Lyons dans *Semantics* (1977, vol. 1 et 2) et *Linguistic Semantics : An introduction* (1995).

Pour commencer, Lyons mentionne une différence essentielle entre le sens *descriptif* et le sens *non descriptif*. Il ajoute que les termes *référentiel*, *cognitif*, ou *propositionnel* peuvent être utilisés dans un sens proche ou égal à celui de *descriptif*. Le sens descriptif est lié aux informations factuelles contenues dans une phrase ou un énoncé. Pour que deux phrases ou propositions puissent être considérées comme équivalentes au niveau de leur sens descriptif, il faut qu'elles partagent les mêmes *conditions de vérité*, c'est à dire les conditions nécessaires pour qu'elles soient considérées comme vraies ou fausses. Le sens non descriptif comprend selon Lyons, le sens expressif et le sens social, mais il précise que de nombreux auteurs groupent ces deux notions sous un seul terme. Il estime que le terme *interpersonnel* est le plus approprié pour désigner ce qui est commun au sens expressif et au sens social. Alors que le sens expressif renvoie aux caractéristiques du locuteur, le sens social est en rapport avec ses relations sociales (ces deux aspects sont bien sûr liés). Le sens non descriptif comprend également ce que certains ont appelé sens *émotif* ou *affectif*, c'est à dire la capacité d'un énoncé à « produire un certain effet sur l'auditeur » (Lyons, 1977:1, 175, notre traduction).

La distinction entre sens descriptif, expressif et social, n'est pas la seule pertinente lorsque l'on se situe au niveau des énoncés, dans un contexte d'énonciation particulier.¹² Depuis les travaux de John L. Austin, le sens d'un énoncé est aussi défini sur la base de sa force illocutoire et de ses effets perlocutoires. La force illocutoire d'un énoncé a trait à « son statut en tant que promesse, menace, demande, déclaration, exhortation, etc. » tandis que son effet perlocutoire agit sur « les croyances, attitudes ou comportement du destinataire » (Lyons, 1977:2, 731, notre traduction). Ces aspects sont liés au *sens pragmatique* d'un énoncé (on parle aussi de sa *fonction communicative*). Ils ne peuvent être définis qu'en considérant le contexte énonciatif dans lequel apparaît l'énoncé ainsi que les intentions des interlocuteurs.

¹² Selon la définition du Dictionnaire de linguistique Larousse (2001), « L'énonciation est l'acte individuel d'utilisation de la langue, alors que l'énoncé est le résultat de cet acte, c'est l'acte de création du sujet parlant devenu alors *ego* ou *sujet d'énonciation* ».

1.2.2.1 Equivalence sémantique

En 1978, Lavandera publie un article intitulé « Where does the sociolinguistic variable stop ? », dans lequel elle exprime ses doutes quant à l'utilisation d'un outil méthodologique unique (la variable sociolinguistique) pour traiter, d'une part, des cas de variation phonologique, où l'équivalence sémantique des variantes est certaine, et d'autre part, des cas de variation morphologique ou syntaxique pour lesquelles l'équivalence sémantique doit être prouvée. Dans *Sociolinguistique*, Labov soutient que pour que deux variantes veuillent dire la même chose, elles doivent partager la même « valeur dénotative ou assertorique » (*referential or truth value* dans le texte original) (1976, 366). Plus tard, dans sa réponse à Lavandera, il indique qu'il base l'équivalence sémantique des variantes sur leur sens représentationnel, tel qu'il est décrit par Bühler en 1934. Il précise encore que le sens représentationnel n'inclut pas le sens expressif ou affectif. John Lyons considère que la notion de sens représentationnel qui apparaît chez Bühler est très proche de ce qu'il appelle sens descriptif. On peut donc en déduire que Labov souhaite baser l'équivalence sémantique des variantes uniquement sur leur sens descriptif.

Peu convaincue par l'extension de la variable sociolinguistique, Lavandera regrette que les premiers travaux variationnistes menés en syntaxe, comme l'étude des passifs de Labov et Weiner¹³, ne traitent pas plus en détail la question du sens. Pour elle, la question de l'équivalence sémantique entre structures actives et structures passives sans agent ne peut être résolue que par l'affirmation vague que les variantes veulent « dire la même chose ». Elle estime qu'il est risqué de prétendre à une identité sémantique entre deux variantes non phonologiques, surtout si les études variationnistes ne se basent pas sur une théorie du sens bien définie.

Les bases du débat ayant été posées par Lavandera et Labov, de nombreuses réactions se firent entendre, tant pour soutenir que pour réfuter la possibilité de l'équivalence sémantique des variantes. Certains proposèrent également des manières d'affiner ou de modifier les conditions sur lesquelles doit se baser l'équivalence des variantes.

Dans sa réplique aux critiques de Lavandera, Labov explique que la définition du sens est beaucoup moins précise dans une perspective sociolinguistique que dans une perspective linguistique formelle (1978). Pour lui, les différences subtiles entre plusieurs sens représentationnels (ou descriptifs) sont pertinentes pour la linguistique formelle, mais pas pour la sociolinguistique : si le but de la linguistique formelle est d'explorer et d'étendre les nuances du sens représentationnel, celui de la sociolinguistique est de les restreindre. Le chercheur David Sankoff défend l'utilisation des méthodes variationniste pour l'analyse de la variation syntaxique et partage l'opinion de Labov sur l'équivalence sémantique des variantes (1988). Il admet que, au contraire des unités phonologiques et morphologiques, deux unités lexicales ou syntaxiques différentes présentent presque toujours des utilisations où leurs significations diffèrent. Si certains pensent que ces différences sont perceptibles d'une manière ou d'une autre à chaque fois que l'une ou l'autre forme est utilisée, Sankoff n'est pas de cet avis. Pour lui, les différences de sens référentiel entre deux unités syntaxiques peuvent être *neutralisées* dans le discours, c'est-à-dire dans un contexte d'énonciation particulier.

Dans un article paru en 1997, Blanche-Benveniste rapporte des discussions qui eurent lieu dans les années 1970-80 entre un groupe de sociolinguistes de Montréal et un autre du GARS d'Aix (dont elle faisait partie). Alors que les premiers étaient prêts à reconnaître

¹³ Il s'agit de « Constraints on the agentless passive », publié en 1983 dans *Journal of Linguistics* 19.

l'équivalence sémantique entre différentes alternances syntaxiques telles que les auxiliaires *avoir* et *être*, les seconds soutenaient « qu'à toute différence de forme correspondait une différence de sens » (Blanche-Benveniste, 1997a, 20). Ainsi, pour le groupe d'Aix, si différentes formes, au sens proches, étaient distribuées différemment selon différentes catégories sociales, on devait considérer que ces différents groupes de personnes « ne se disaient pas exactement « la même chose », et que la façon de se comprendre était toujours approximative » (1997a, 20). Le parti pris par Blanche-Benveniste est donc de considérer qu'une modification dans la forme va de pair avec une modification de contenu, « ne serait-ce que dans les façons subtilement différentes d'organiser l'information » (1997b, 56). Gadet mentionne également l'impossibilité d'une équivalence sémantique totale comme un obstacle à l'extension de la variable phonologique au niveau syntaxique. Dans le cas d'alternances syntaxiques, elle pense qu'il serait plus raisonnable de considérer que les différentes formes peuvent « dire des choses proches à propos du même référent » plutôt que « dire la même chose » (1997b, 11).

Comme Blanche-Benveniste et Gadet, de nombreux linguistes affirment que deux variantes syntaxiques ne peuvent pas « vouloir dire la même chose » au même titre que deux variantes phonologiques. Mais certains, convaincus de cela, acceptent néanmoins que l'on puisse baser l'équivalence sémantique des variantes sur leur sens descriptif, comme le préconise Labov. Ainsi, Godard, très critique face au concept de variable syntaxique, est prête à en accepter l'utilisation si l'équivalence sémantique entre les variantes ne concerne que le sens référentiel et non d'autres aspects sémantiques plus subtils. De même, Winford conçoit que le concept de variable syntaxique puisse être valable si l'on base l'équivalence sémantique des variables sur leur seul sens descriptif.¹⁴

On constate donc que l'affirmation de Labov selon laquelle les variantes d'une variable (socio)linguistique doivent « vouloir dire la même chose » est diversement interprétée et acceptée. Pour certains, l'équivalence sémantique des variantes n'est pas concevable hors de la phonologie, alors que pour d'autres, elle peut être envisagée si on la définit en termes d'équivalence de sens descriptif. En plus de ces considérations sémantiques, on peut se demander si, pour « vouloir dire la même chose », les variantes ne devraient pas renvoyer à une valeur pragmatique commune.

1.2.2.2 *Equivalence fonctionnelle*

Pour Lavandera, nous l'avons vu, il n'est pas souhaitable d'établir des variables syntaxiques sur la base de l'équivalence sémantique des variantes. Mais ce n'est pas pour autant qu'elle décide d'abandonner le concept de variable syntaxique. Elle suggère, sans développer son propos, d'éliminer le critère d'équivalence sémantique pour le remplacer par un critère d'équivalence fonctionnelle. L'idée d'une équivalence fonctionnelle des variantes est reprise et approfondie par Romaine dans sa réponse à Lavandera et Labov. Selon elle, le critère d'équivalence de sens descriptif des variantes n'est pas suffisant et doit impérativement être complété par un critère d'équivalence pragmatique (ou fonctionnelle). Elle s'étonne d'ailleurs que Labov, ayant toujours dit s'intéresser aux aspects sociaux du langage, base l'équivalence des variantes sur des critères de sens descriptif plutôt que sur des critères fonctionnels basés sur l'usage.

¹⁴ Winford souligne toutefois que le sens *référentiel* dont parle Labov ne correspond pas tout à fait au sens *descriptif* de Lyons. C'est, selon lui, sur le sens descriptif qu'il faut baser l'équivalence sémantique des variantes. Nous considérons néanmoins, pour les besoins de ce travail, que les sens *référentiel* et *descriptif* sont deux concepts identiques.

Après Lavandera et Romaine, d'autres linguistes ont abordé l'idée de définir l'équivalence des variantes sur la base de leur sens pragmatique. Winford considère la possibilité de définir l'équivalence des variantes uniquement sur l'identité de leur fonction dans le discours (1996). Il pense qu'une telle définition de la variable syntaxique est envisageable mais diffère passablement de la définition communément attribuée au concept de variable sociolinguistique. Il explique que lorsque l'on considère que deux énoncés ont le même sens pragmatique, il n'est pas question de leur sens descriptif mais de leur portée dans un contexte donné, c'est-à-dire de leur force illocutoire et peut-être de leur effet perlocutoire. Or, on sait qu'une même force illocutoire peut être transmise à travers des énoncés sans aucune équivalence formelle (unités lexicales et syntaxiques différentes). Winford mentionne l'intérêt de cette approche au niveau pragmatique, tout en soulignant qu'elle s'éloigne beaucoup d'une approche basée sur la variable sociolinguistique conçue par Labov ; si la constance de sens descriptif est abandonnée au profit du sens pragmatique, on analyse alors différentes stratégies discursives utilisées dans un même but communicatif. On dit des choses différentes pour communiquer une même intention et il n'est alors pas possible de corréliser ces différentes stratégies à des facteurs sociaux, comme c'est le cas pour certaines alternances au contenu descriptif commun.

Winford insiste avant tout sur la nécessité, pour toute étude sur la variation, de préciser quel type de sens sera considéré comme constant avant d'analyser les unités qui varient. C'est aussi l'avis de Sven Jacobson, qui est prêt à accepter diverses méthodes d'analyses de la variation syntaxique, pour autant que le sens qui relie les formes alternantes soit précisément défini (1989). Il pense que c'est au chercheur de déterminer le type de sens pertinent pour son analyse de la variation syntaxique. Après avoir choisi le sens qui restera constant, (descriptif, pragmatique, éventuellement stylistique ou affectif), le chercheur peut définir les contraintes variables qui agissent sur le choix de l'une ou l'autre forme syntaxique. Si Jacobson propose donc plusieurs pistes pour l'analyse de la variation, il faut préciser toutefois qu'il n'utilise pas le terme de *variable (socio)linguistique*. Il semble que la « variation » qu'il analyse au moyen d'outils quantitatifs et statistiques comprenne davantage de phénomènes que ceux qui répondent aux exigences de la variable linguistique de Labov.

Certains chercheurs recourant à l'approche variationniste labovienne aux niveaux morphologique ou syntaxique considèrent que l'équivalence fonctionnelle des variantes est aussi importante que leur équivalence sémantique. C'est le cas d'Aidan Coveney, dont nous analyserons, dans la deuxième partie de ce travail, une étude variationniste sur les formes interrogatives. Selon lui, les différentes variantes d'une variable syntaxique ne peuvent être considérées comme équivalentes que si elles présentent le même sens descriptif et la même fonction communicative. Sankoff pose également l'équivalence descriptive et fonctionnelle des variantes comme un critère essentiel pour l'identification des variantes d'une variable non phonologique (1988). Il explique que deux formes morphologiques ou syntaxiques présentent souvent des distinctions sémantiques ou fonctionnelles dans certains contextes, mais que ces distinctions peuvent parfaitement être *neutralisées* dans d'autres contextes. Le travail du chercheur variationniste consiste alors à identifier les contextes où le changement de forme correspond à un changement de sens ou de fonction et les contextes dans lesquels ces différences ne sont pas pertinentes. Les premiers, considérés comme *catégoriques* et non variables, devront être exclus de l'analyse variationniste. Sankoff et Coveney insistent tous deux sur la difficulté et l'importance de cette étape interprétative, cruciale dans toute recherche variationniste en syntaxe. Selon eux, pour pouvoir juger de l'équivalence entre deux variantes, il faut bien connaître la communauté linguistique étudiée ainsi que les situations d'interaction d'où proviennent les données linguistiques. Sankoff estime que la situation la plus avantageuse est celle où le chercheur fait lui-même partie de la communauté

en question. Cette proximité favoriserait la précision des jugements d'équivalence, tout comme la présence du chercheur pendant l'enregistrement des données, comme le recommande Coveney.

1.2.3. Les facteurs qui influencent la variation

Labov et les défenseurs de la variable syntaxique ont tendance à souligner les ressemblances entre le conditionnement de la variation phonologique et celui de la variation syntaxique. Ils constatent que le fonctionnement et la nature des différents facteurs influençant la variation sont similaires au niveau de la phonologie et aux autres niveaux de la langue, ce qui soutient le postulat selon lequel la variable sociolinguistique est un concept aussi valide en phonologie qu'en morphologie ou en syntaxe. Au contraire, ceux qui doutent que la notion de variable sociolinguistique puisse s'appliquer à la morphologie ou à la syntaxe s'attachent à montrer les divergences entre le conditionnement de la variation au niveau de la phonologie et aux autres niveaux de la langue. Les divergences les plus souvent mentionnées touchent à la possibilité d'un investissement social des variantes (morpho)syntaxiques et à l'influence possible de facteurs pragmatiques dès lors que l'on sort de la phonologie.

1.2.3.1 Facteurs sociaux, stylistiques et linguistiques

En phonologie, Labov a établi que l'usage des variantes d'une variable pouvait être influencé par trois types de facteurs: sociaux, stylistiques et linguistiques. Les facteurs sociaux et stylistiques agissent toujours de manière *variable*, c'est-à-dire qu'ils peuvent favoriser l'usage d'une variante, mais pas la contraindre. Lorsqu'on remarque qu'une variante est corrélée à un certain groupe social, cela ne veut pas dire que ce groupe fait toujours usage de cette variante mais qu'il l'utilise dans un pourcentage élevé. En revanche, les facteurs linguistiques peuvent agir tantôt de manière *catégorique*, tantôt de manière *variable*. Lorsque le contexte linguistique dans lequel apparaît la variante est catégorique, il contraint l'utilisation de l'une ou l'autre variante et ne laisse donc pas place à la variation. L'influence du contexte linguistique sur la variation peut être illustrée par les phénomènes de liaisons en français. En effet, le contexte linguistique est souvent déterminant dans le caractère obligatoire, variable ou interdit de la prononciation de certaines consonnes finales.

(7) Ils sont là.

Dans l'exemple (7), la présence de *là* et de sa consonne initiale empêche, en principe, la prononciation de la consonne finale de *sont*.

(8) Ils sont avec nous.

En (8) par contre, la présence d'une voyelle après *sont* permet deux prononciations différentes: [sõtavek] ou [sõavek]. Ainsi, les caractéristiques linguistiques de la phrase dans laquelle apparaît la variable peuvent contraindre l'usage de l'une ou l'autre variante et donc agir comme des facteurs catégoriques. Labov explique que les facteurs qui influencent la variation fonctionnent de la même manière en phonologie que dans les autres niveaux de la langue. Ainsi, en phonologie comme en syntaxe, les facteurs sociaux et stylistiques sont toujours variables alors que les facteurs linguistiques peuvent être catégoriques ou variables.

De même, la nature des facteurs sociaux et stylistiques reste inchangée lorsque l'on passe de la phonologie aux autres niveaux de la langue. Les facteurs sociaux pouvant influencer l'usage des variantes sont nombreux : en général, ils ont trait à la classe sociale ou économique des locuteurs, à leur âge, leur sexe ou encore leur profession. Nous l'avons vu, les différents contextes stylistiques se rangent, suivant Labov, selon le « degré d'attention

donnée à la parole ». Ces différents degrés sont graduels et il est parfois difficile de catégoriser le style du discours d'où proviennent les variantes étudiées. Selon les études, on peut distinguer deux ou plusieurs styles différents, mais il n'y a pas de méthode unique pour identifier les différents styles. Labov a créé, en 2001, un « arbre de décision » destiné à définir si, lors d'un entretien, le style est familier ou soigné (voir point 2.3.4.6). Pour son étude variationniste sur les interrogatives, Martin Elsig a choisi d'utiliser cet arbre de décision, tout en soulignant que l'exhaustivité d'un tel modèle est discutable (2009). Pour ce qui est des facteurs linguistiques, leur nature varie énormément d'une variable à l'autre, aussi bien au niveau phonologique que syntaxique. Nous avons déjà mentionné plusieurs facteurs linguistiques (phonologiques ou grammaticaux) qui pouvaient agir sur les variables phonologiques. Au niveau morpho-syntaxique, on peut citer comme exemples la nature ou l'ordre des différents éléments d'une phrase, la catégorie et la fonction grammaticale d'un mot, sa longueur, ou encore ses propriétés sémantiques ou phonologiques.

Malgré ces similitudes de nature et de fonctionnement, souvent évoquées par les partisans de la variable syntaxique, certains soutiennent que les facteurs conditionneurs de variation n'agissent pas de la même manière en phonologie et aux autres niveaux de la langue. C'est principalement l'action des facteurs sociaux qui est mise en cause. En effet, de nombreux chercheurs doutent que les variables non-phonologiques puissent être investies socialement au même titre que les variables phonologiques. Beaucoup approuvent Alain Berrendonner lorsqu'il déclare qu'au niveau de la syntaxe, « l'exploitation sociolinguistique des variantes est un phénomène partiel et secondaire ».¹⁵ Le peu de fréquence des variables syntaxiques est souvent considéré comme un frein à leur investissement social. Comme l'explique M. Rydén, les structures syntaxiques sont moins souvent répétées que les structures phonologiques et sont de ce fait moins susceptibles d'être conditionnées par des facteurs sociaux.¹⁶ Gadet constate également que les fréquences des occurrences sont toujours plus faibles au niveau de la syntaxe qu'au niveau de la phonologie. Selon elle, les marqueurs sociaux sont nombreux en phonologie, moindre en morphologie et rares en syntaxe (1992 et 1997b). La corrélation entre la fréquence d'occurrence des variables et leur degré de stratification sociale a été établie par Labov dès le début de ses recherches sur la variation. Pour lui, la variation syntaxique présente certes moins de possibilités pour l'investissement social que la variation phonologique, mais elle reste intéressante pour l'étude des effets de la stratification sociale sur la langue.

Comme nous l'avons vu précédemment, les premières études de Labov mettaient surtout en évidence le rôle des facteurs sociaux sur la variation, soit-elle phonologique ou (morpho)syntaxique. Plus tard, ses études ont légèrement changé d'orientation et il s'est intéressé autant à l'influence du social sur la langue qu'à sa structure même. Ainsi, dans l'étude des passifs qu'il a menée avec Weiner, Labov a remarqué que les alternances entre voies actives et passives étaient conditionnées principalement par des facteurs linguistiques, mais aucunement par des facteurs sociaux ou stylistiques. Comme le note Lavandera, Labov ne parle pas, dans l'étude des passifs, de variable *sociolinguistique* mais de variable *linguistique*. Or, elle constate que Labov utilisait ces deux termes indistinctement dans ses premières études sur la variation, alors que les variables étudiées étaient toujours conditionnées par des facteurs sociaux. Elle souhaite que la notion de *variable* (soit-elle qualifiée de *sociolinguistique* ou de *linguistique*) s'applique uniquement à des situations où la variation présente un conditionnement social, comme c'était le cas lors de son introduction par Labov.

¹⁵ Cité par Blanche-Benveniste (1997a, 28)

¹⁶ Propos rapportés par Jenny Cheshire (1996)

S'il est vrai, comme le mentionne Lavandera, que le concept de variable a été introduit par Labov pour analyser l'impact des facteurs sociaux sur la variation, sa portée s'est élargie à l'analyse de cas de variation sans conditionnement social. Peut-on alors encore parler de variable ? Pour Labov, il faut parler de variable *sociolinguistique* lorsqu'il y a corrélation avec des facteurs sociaux, et de variable *linguistique* lorsque ce n'est pas le cas. Plusieurs chercheurs se servent de cette distinction, mais ils n'utilisent pas toujours les mêmes termes. Par exemple, Coveney parle de *variable* seulement s'il y a corrélation avec des facteurs sociaux ; lorsque ce n'est pas le cas, il préfère parler d'*alternances*. Selon lui, la méthodologie variationniste peut tout aussi bien servir à l'analyse du conditionnement des variables que des alternances (2002). Pour Winford également, les cas de variation syntaxique se rangent en deux catégories, selon qu'ils peuvent ou non présenter un conditionnement social. Il ne propose cependant pas de terme spécifique pour distinguer ces différents types de variation. On remarque ainsi que la signification attribuée au mot *variable linguistique* diffère d'un auteur à l'autre ; pour certains, il est nécessairement lié à un investissement social des variantes alors que pour d'autres, il implique un conditionnement par des facteurs linguistiques ou extralinguistiques qui ne sont pas nécessairement sociaux. Il faut donc observer, dans toute recherche sur la variation linguistique, l'usage que l'auteur fait du concept de *variable*.

1.2.3.2 Facteurs pragmatiques

Lorsque l'on sort de la phonologie, les facteurs sociaux, stylistiques et linguistiques ne sont pas les seuls à pouvoir influencer l'usage des variantes. Il faut également compter avec l'influence possible de facteurs pragmatiques, liés par exemple à la fonction communicative des variantes ou à leur place dans la conversation (donc, à leur sens en contexte). Nous avons expliqué que l'influence des facteurs sociaux était plus faible au niveau de la variation syntaxique qu'au niveau phonologique et que cette différence pouvait s'expliquer par la moindre fréquence des variables syntaxiques. Or, on peut aussi considérer que le conditionnement des variables syntaxiques par des facteurs pragmatiques inhibe toute possibilité d'investissement social.

Romaine déclare que les variables syntaxiques (ce qu'elle appelle *pure syntactic variables*) ne peuvent être investies socialement à cause du rôle essentiel des facteurs pragmatiques, qui constituent selon elle les véritables conditionneurs de la variation syntaxique. Pour elle, toute alternance de structures syntaxiques engendre des changements au niveau du sens pragmatique. Ainsi, le choix de l'une ou l'autre structure dépendrait des nuances pragmatiques que le locuteur veut transmettre mais jamais de ses caractéristiques sociales. Romaine accepte cependant que des facteurs sociaux puissent influencer l'usage des variantes morphosyntaxiques ; ces dernières étant structurellement moins complexes que les variantes syntaxiques, elles pourraient alterner sans impliquer nécessairement de distinctions pragmatiques. C'est le même type de réflexion qui sous-tend les propos de Jenny Cheshire lorsqu'elle affirme qu'on ne peut traiter des cas de variation en termes de variable linguistique que lorsqu'ils se situent à l'extrémité morphologique du continuum entre morphologie et syntaxe. En effet, elle pense que le sens grammatical impliqué par les variables syntaxiques est trop complexe pour ne pas être lié à des distinctions pragmatiques.¹⁷

D'une manière générale, il est admis que les facteurs pragmatiques sont d'une grande importance dans le conditionnement des variables syntaxiques. Si certains linguistes pensent qu'ils excluent l'action de facteurs sociaux, d'autres estiment que ces deux types de facteurs

¹⁷ propos rapportés par Winford (1996)

peuvent agir conjointement et qu'il est possible de mesurer leurs actions respectives. C'est ce que fait Coveney lorsqu'il étudie la variabilité du *ne* de négation et des structures interrogatives en français. Selon lui, tous les facteurs potentiellement influents sur l'usage d'une variante peuvent et doivent être pris en compte dans l'analyse de la variation. Il faut observer, pour chacune des occurrences d'une variante, son contexte linguistique et stylistique, sa fonction communicative et les caractéristiques sociales du locuteur qui en fait usage. Toutes ses données doivent être codées et introduites dans un système d'analyses statistiques (par exemple le programme informatique VARBRUL). Coveney précise que l'utilisation d'un outil statistique tel que VARBRUL ne préjuge en rien de la nature sociolinguistique de la variable: on peut choisir d'étudier une alternance et décider ensuite s'il s'agit d'une variable sociolinguistique ou pas. Sankoff propose de traiter les différents facteurs impliqués dans la variation de manière similaire à Coveney. Il explique :

Lorsqu'on s'aperçoit qu'un choix a été fait au cours de la performance linguistique et qu'il a pu être influencé par des facteurs tels que la nature du contexte grammatical, la fonction discursive de l'énoncé, le sujet, le style, le contexte interactionnel ou les caractéristiques personnelles ou sociodémographiques du locuteur ou des autres participants, il est difficile de ne pas recourir aux notions et méthodes de l'inférence statistique, même si ces dernières ne doivent constituer qu'un outil heuristique pour tenter de saisir l'interaction des nombreux composants d'une situation complexe. (1988, 151, notre traduction)

On comprend donc que les méthodes statistiques d'analyse de la variation peuvent être utilisées à d'autres fins que celles de mesurer les corrélations entre la langue et le social. Qu'elles puissent révéler des liens entre caractéristiques sociales des locuteurs et usage des variantes est discutable, surtout lors de l'analyse d'alternances syntaxiques complexes. Dans son étude sur les interrogations, Coveney accorde peu d'importance au conditionnement des variantes par des facteurs sociaux alors que l'effet des facteurs pragmatiques et linguistiques est largement traité. Il admet d'ailleurs que les interférences entre les nombreux facteurs pragmatiques et les quelques facteurs sociaux réduisent la fiabilité des résultats obtenus quant au conditionnement social des variantes. Il affirme en revanche que le rôle des facteurs sociaux est plus marqué et plus sûr avec la variable, beaucoup moins complexe, du *ne* de négation (présence/absence).

Comme les facteurs linguistiques, les facteurs pragmatiques peuvent être tantôt catégoriques, tantôt variables. Pour chaque occurrence d'une variante, l'analyste doit décider si une ou d'autres variantes peuvent apparaître dans le même contexte sans que cela n'altère le sens descriptif ou la fonction de l'énoncé. Une certaine fonction communicative peut contraindre ou exclure l'usage d'une variante : les facteurs pragmatiques sont alors catégoriques. Dans d'autres cas, la fonction communicative agit de manière variable, c'est-à-dire qu'elle influence l'usage d'une forme sans la contraindre. Nous le verrons au point 2.3, le travail de Coveney sur les structures de l'interrogation montre combien l'attribution de fonctions communicatives précises à chacune des occurrences des variantes est complexe et fastidieuse. Il admet d'ailleurs que l'établissement des fonctions communicatives est « un exercice particulièrement problématique au vu des nombreux désaccords sur comment et dans quelle mesure les aspects pragmatiques du sens des énoncés peuvent être représentés » (2002, 123, notre traduction). Il pense que beaucoup d'oppositions à l'usage de la variable hors de la phonologie ont l'apparence d'objections théoriques mais relèvent en fait de problèmes méthodologiques. Si Coveney prend le risque d'aborder ces questions, d'autres jugent que leur complexité ainsi que l'importance de la dimension interprétative incombant au chercheur variationniste mettent en doute la fiabilité de ce type de recherches.

1.3 Explications de la variation

1.3.1 Etendue de la variation : où et comment se manifeste-t-elle ?

Nous avons expliqué aux points précédents que le courant variationniste était né en phonologie, dans le but de déterminer les facteurs sociaux et stylistiques qui pouvaient structurer la variation linguistique. Certaines des variantes « libres » établies par les structuralistes ont été réanalysées en variantes « sociolinguistiques » dès lors que leur schéma d'usage a pu être corrélé avec des facteurs socio-stylistiques. Lorsque les méthodes variationnistes ont commencé à s'appliquer à la morphologie ou à la syntaxe, le concept de « variantes » a dû être réexaminé et précisé : les variantes (socio)linguistiques ont généralement été définies comme différentes unités formelles présentant une identité de sens descriptif.¹⁸ L'extension des méthodes variationnistes hors de la phonologie a aussi posé la question de l'investissement pragmatique des variantes : on a remarqué qu'en plus (ou au lieu) des facteurs sociaux et stylistiques, des facteurs pragmatiques étaient susceptibles de peser sur le choix des variantes non phonologiques.

Les phénomènes de variation présents dans la langue sont nombreux et variés : ils concernent tout ce qui est hétérogène, en opposition à ce qui est stable et homogène. Pour aborder la thématique de la variation linguistique, on doit s'interroger, d'une part, sur ce qui varie et, d'autre part, sur les facteurs qui causent et expliquent cette variation. Le variationnisme ne prend en compte qu'une partie des éléments de la langue qui varient, et ceci, selon un certain angle d'analyse. Les données linguistiques analysées par les méthodes variationnistes ne comprennent que des unités qui partagent au moins le même sens descriptif. Les explications données quant à la structuration de ces unités variables concernent principalement les caractéristiques sociales des locuteurs ou le contexte stylistique de l'interaction (mais certaines études variationnistes tentent aussi de lier l'organisation de la variation à des aspects pragmatiques de la communication). Or, la question de savoir *ce qui varie* dans une langue et *en fonction de quoi* peut être traitée selon différentes perspectives.

Tous les niveaux de la langue peuvent présenter des phénomènes variables: au point 1.2, nous avons abordé des cas de variation provenant des niveaux phonologique, morphologique et syntaxique. Le domaine lexical, peu traité par les méthodes variationnistes, comporte également de nombreuses possibilités de variation. Selon Berrendonner, le « matériau variationnel » de la langue se résume à « quelques schèmes, ou types de variation (deux ou trois), qui rendent compte de la nature de toutes les variantes possibles », et ceci indépendamment du niveau de la langue où il se situe (1988, 46). Premièrement, il y a variation lorsque l'appartenance d'une certaine unité à une certaine classe linguistique n'est pas bien définie: c'est le cas des noms de la langue française dont le genre est flottant, comme *hymne* ou *météorite*. Le deuxième « schème variationnel » décrit par Berrendonner concerne les cas d'indécisions liés à la forme d'une structure (ordre des mots) ou d'une unité morphologique. Ces indécisions peuvent venir du fait que le système linguistique présente une carence ou, au contraire, du fait qu'il offre plusieurs formes rivales. Ainsi, les diverses structures interrogatives du français sont autant de formes concurrentes présentes dans le système. Quant à la carence, elle apparaît lorsqu'« il n'y a pas d'opération de linéarisation

¹⁸ Cette définition ne fait toutefois pas l'unanimité, certains considérant que les véritables variantes ne peuvent être que phonologiques, et d'autres préconisant, en plus ou à la place de l'identité de sens descriptif, une identité de sens pragmatique.

prévue dans le système, ou bien il manque une forme dans le lexique » (Berrendonner, 1988, 46).

La structuration du matériau variationnel de la langue peut être influencée par des facteurs linguistiques internes ou par des facteurs extralinguistiques. En abordant les méthodes variationnistes, nous avons montré comment des facteurs linguistiques (propriétés phonologiques d'un énoncé, nature de ses constituants, etc.) pouvaient peser sur l'usage des unités variantes. Quant aux facteurs extralinguistiques agissant sur la variation, ils sont souvent classés en quatre catégories, selon une terminologie en *dia-*:

- La variation *diatopique* concerne tous les phénomènes de variation en relation avec l'espace : le français varie par exemple suivant qu'il est parlé en Belgique ou au Québec, au nord ou au sud de la France.
- La variation *diachronique*¹⁹ est liée au passage du temps et aux changements de la langue au fil des époques. Elle peut être observée en analysant des états de langue provenant d'époques révolues ou en comparant, à une époque donnée, les parlers de personnes présentant une grande différence d'âge.
- La variation *diastratique*, résulte des différences sociales entre les locuteurs (classe socio-professionnelle, âge, sexe, etc.). Selon Gadet, « tous les traits sociaux qui divisent une société et peuvent être porteurs d'oppositions et de conflits sont susceptibles de constituer des axes de variation » (1997a, 5).
- La variation *diaphasique* est observable chez un seul et même locuteur. Elle est souvent qualifiée de *stylistique* et dépend principalement du contexte interactionnel dans lequel se trouve le locuteur. Beaucoup d'éléments peuvent entrer en compte dans la caractérisation d'un contexte interactionnel. On peut citer, à titre d'exemples : les personnes en présence et le type de relation qu'elles entretiennent, le milieu (professionnel, familial, etc.), ou encore la nature des propos échangés (solennels, humoristiques, techniques, etc.).

A ces quatre axes de variation, Gadet (2003) ajoute celui de *diamésique*, qui concerne la variation due aux différents médiums (oral et écrit) de la langue. Elle précise que le français ne connaît pas de formes réservées à l'un ou l'autre médium mais que l'on peut observer des tendances caractéristiques du mode oral ou écrit.

La sociolinguistique s'intéresse principalement à la variation diastratique (ou sociale) et diaphasique (ou stylistique), plutôt qu'aux phénomènes qui relèvent de la diachronie ou de la diatopie. Nous l'avons vu, certaines études variationnistes souhaitent également intégrer le rôle des facteurs pragmatiques dans la variation. L'importance des facteurs pragmatiques dans la structuration de la variation est bien mise en évidence par Berrendonner (1987 et 1988). Il explique qu'au travers de la variation, la langue offre différentes possibilités d'expression qui peuvent être exploitées par les locuteurs comme autant de stratégies discursives différentes (Berrendonner parle alors de stratégies « lectales »).

¹⁹ La terminologie en *dia-* place le changement diachronique aux côtés de phénomènes de variation synchroniques : comme nous l'a fait remarquer Alain Berrendonner lors de la soutenance de ce mémoire, cette présentation est discutable. On ne peut pas, selon lui, considérer le changement diachronique comme un phénomène de variation.

Pour illustrer quelques-unes de ces stratégies, nous nous baserons sur le mot *amour*, donné par Berrendonner comme exemple d'un nom dont le genre n'est pas défini en français. Le locuteur confronté au choix du genre d'un tel mot peut opter pour une stratégie de *sélection-généralisation* ou pour une *spécification*. Dans le premier cas, il déterminera un genre et s'y tiendra : cette stratégie consiste donc à éliminer la variation en sélectionnant une variante au détriment des autres. S'il recourt à la *spécification*, le locuteur déterminera des conditions d'emploi spécifiques pour chaque variante. Cette stratégie coïncide avec les dictats de la norme dans le cas du mot *amour* : les grammaires préconisent de réserver le genre masculin au singulier et le genre féminin au pluriel. Mais le locuteur peut également décider de ne pas choisir, par exemple en employant une forme neutralisée (« l'amour » ou « les amours », qui ne donnent pas d'indication de genre) ou simplement en renonçant à l'utilisation de l'unité variable.

Selon Berrendonner, les différentes stratégies lectales mises en œuvre dans le choix d'une unité linguistique variante servent divers objectifs :

Remédier aux conflits opératifs, prévenir les ambiguïtés et autres difficultés de décodage, favoriser la reconnaissance perceptive ou la mémorisation des structures, assurer le rendement maximal des moyens mis en œuvre, telles sont les fonctions assurées par la variation. (1987, 51)

La manière dont les locuteurs font usage de la variation linguistique peut donc être liée à différentes stratégies discursives. Toutefois, l'influence des prescriptions normatives sur l'organisation de la variation dans le discours n'est pas négligeable. En effet, un grand nombre des variations présentes dans la langue sont condamnées ou réglées par les grammaires normatives, ces dernières recourant fréquemment aux techniques de *sélection-généralisation* ou de *spécification* pour différencier les emplois dits « corrects » de ceux dits « incorrects ». Berrendonner parle de « conflits » entre, d'une part, la commodité et l'efficacité communicative qu'offrent les possibilités de variation de la langue et, d'autre part, la censure dont elles font l'objet au niveau normatif. Suivant les situations, le locuteur préférera s'exprimer conformément aux exigences de la norme (et en montrer ainsi sa maîtrise) ou user pleinement des possibilités communicatives offertes par la variation linguistique.

Les études de la variation peuvent également s'interroger sur les relations entre la nature des unités linguistiques qui varient et le type de facteurs extralinguistiques impliqués dans la variation. Il n'existe évidemment pas de relations univoques entre telles unités variables et tels facteurs extralinguistiques, mais certaines affinités peuvent être observées. Par exemple, le niveau phonologique présente de nombreuses variations sur le plan diatopique : on reconnaît généralement un locuteur francophone d'une autre région que la sienne par sa prononciation, c'est-à-dire son accent, plutôt que par son usage de la syntaxe ou du lexique. En revanche, la phonologie change peu lorsqu'un locuteur se trouve dans différentes situations interactionnelles, la variation diaphasique du français étant plus marquée par des changements lexicaux et morphosyntaxiques que phonologiques. En outre, nous avons abordé au point 2 la question des différences fondamentales entre le niveau phonologique (n'impliquant pas de sens) et les niveaux supérieurs de la langue, porteurs de sens. Si Gadet admet que l'influence de facteurs sociaux peut intervenir à tous les niveaux de la langue, elle réfute pourtant l'utilisation de la variable sociolinguistique hors de la phonologie :

Une problématique sociolinguistique offre un cadre satisfaisant pour traiter de la variation en phonologie, mais est à la fois insuffisante et inadéquate pour saisir le jeu syntaxique. (Gadet, 1997a, 141)

Le peu d'intérêt des méthodes variationnistes pour la variation lexicale peut s'expliquer par le fait que le lexique, plus que toute autre unité linguistique, véhicule du sens. L'exigence d'identité de sens descriptif est donc difficile à remplir dès lors qu'interviennent différentes unités lexicales.

1.3.2 Considérations terminologiques

Le terme « variation » peut recouvrir de nombreux phénomènes : il peut référer autant à des propriétés linguistiques internes de la langue qu'à des facteurs extralinguistiques. On l'utilise pour désigner des phénomènes relevant de la synchronie comme de la diachronie (bien que dans ce cas, on préfère souvent le terme « changement » à celui de « variation »). Pour Gadet (1997b), la notion de variation était déjà mal définie avant son utilisation par la sociolinguistique, cette dernière n'ayant fait qu'amplifier la confusion qui régnait autour de ce concept.

Dans sa présentation du numéro 115 de *Langue française*, Gadet explique que le mot « variation » a été utilisé dès les débuts de la linguistique moderne sans que son statut conceptuel n'ait été défini. Avec l'apparition de l'expression « variation libre », la « variation » a commencé à désigner plus précisément le cas de différentes unités formelles (généralement des phonèmes et des morphèmes) pouvant être utilisées alternativement pour transmettre un même contenu. Le concept a ensuite été introduit en sociolinguistique pour désigner autant les phénomènes linguistiques qu'extralinguistiques (on parle de la « variation » des voyelles aussi bien que de la « variation » sociale). Aujourd'hui, le statut conceptuel du mot « variation » reste peu clair,²⁰ si bien que Gadet recommande d'être attentif à la diversité des usages auxquels il peut servir.

Les définitions du mot « variation » trouvées dans des ouvrages de référence linguistique²¹ témoignent du caractère flou de la notion. Dans tous les ouvrages que nous avons consultés, la variation est expliquée comme le résultat de facteurs extralinguistiques : caractéristiques sociales des locuteurs, diversité de situations interactionnelles et d'espaces sont donnés comme causes de variation²², sans que celle-ci ne soit vraiment définie. Pour le *Dictionnaire des sciences du langage*, elle est la « manifestation de la variabilité des langues naturelles », alors que le dictionnaire *Linguistique et sciences du langage*, la décrit comme un « phénomène » qui fait qu'une langue « n'est pas la même » en fonction des facteurs extralinguistiques ci-mentionnés. Pour *La grammaire d'aujourd'hui*, la variation « s'observe dans l'espace ou selon les couches socioculturelles ». Enfin, dans *Termes et concepts de l'analyse du discours*, elle est définie comme un « phénomène central des langues, qui signale leur nature sociale ». Le caractère vague de ces définitions ne surprend guère lorsque l'on songe à toutes les facettes de la variation qu'elles doivent englober. Hormis ces quelques formulations générales, on constate que les définitions de la variation tournent principalement autour des facteurs extralinguistiques qui la causent, alors qu'une description du fonctionnement linguistique de la concurrence entre unités structurales manque. Les ouvrages *Termes et concepts* et *Dictionnaire des sciences du langage* donnent bien une liste des unités structurales susceptibles de varier, mais ils n'expliquent pas comment elles peuvent se substituer. De plus, il n'est fait aucune mention des facteurs pragmatiques pouvant motiver l'usage du matériau variationnel.

²⁰ Ainsi, Gadet présente-t-elle le numéro 115 de *Langue française* comme une tentative d'éclaircissement de la notion.

²¹ Il s'agit de : *Linguistique et sciences du langage* (2007), *Dictionnaire des sciences du langage* (2004), *Termes et concepts pour l'analyse du discours* (2001), *La grammaire d'aujourd'hui* (1986)

²² Le changement diachronique est souvent mentionné, mais différencié de la variation et traité à part.

Nous avons également examiné les entrées « variantes » et « variables » de nos quatre ouvrages de référence, ces deux mots étant utilisés par les études variationnistes selon une acception bien précise. Seul le dictionnaire de *Linguistique et sciences du langage* comprend une entrée pour le mot « variable », mais il est décrit dans une optique structuraliste comme une « quantité susceptible de prendre différentes valeurs », l'exemple donné étant la catégorie SN (syntagme nominal), dont les différentes valeurs peuvent être *je*, *la table*, *Hélène*, etc. Le *Dictionnaire des sciences du langage* et celui de *Linguistique et sciences du langage* ont une entrée pour le mot « variante », mais il n'est pas décrit selon une perspective de mise en relation avec des facteurs extralinguistiques, comme c'est le cas des variantes faisant l'objet d'études variationnistes. Les deux dictionnaires décrivent les variantes comme des phonèmes ou des morphèmes pouvant apparaître en distribution complémentaire ou de manière libre. La seule mention pouvant rappeler la possible exploitation sociale des variantes est le terme « variantes *stylistique* », donné comme synonyme de variante libre. Dans le *Dictionnaire des sciences du langage*, seuls les phonèmes et les morphèmes sont considérés comme des variantes. L'entrée « variante » de la *Grammaire d'aujourd'hui* renvoie aux entrées « allomorphe », « allophone », « morphème », « orthographe » et « phonétique », ce qui montre que les variantes sont là aussi associées à des unités phonologiques et morphologiques. Seul le *Dictionnaire des sciences du langage* fait allusion à d'autres unités linguistiques, lorsqu'il indique (nous soulignons) :

On appelle *variantes* les diverses réalisations effectives d'une unité fonctionnelle, notamment dans les domaines phonologique et morphologique. (300)

Ces différents éléments montrent que la définition usuelle du terme « variante » ne correspond pas à la définition qu'en donne la sociolinguistique et les études variationnistes. En effet, la corrélation des variantes à des facteurs extralinguistiques n'apparaît pas dans ces définitions comme une de leurs caractéristiques centrales : les variantes sont simplement distinguées en libres ou contraintes. Par ailleurs, l'existence de variantes syntaxiques n'est pas prise en compte, ces dernières n'étant pas ancrées dans les théories grammaticales en termes de variantes libres ou contraintes.

En bref, la variation linguistique est difficilement réductible à un concept unique, sa définition oscillant entre un ensemble de causes et d'effets intra et extralinguistiques. Ainsi, les différentes approches théoriques de la langue n'étudieront pas obligatoirement les mêmes facettes de la variation : le dialectologue se focalisera sur les phénomènes de variation dus aux changements d'espaces, le diachronicien s'intéressera aux changements dans le temps tandis que le grammairien étudiera les principes linguistiques ou normatifs réglant l'utilisation du matériau variationnel. Mais, si l'influence des facteurs diatopiques et diachroniques est facilement isolable de l'influence d'autres facteurs extralinguistiques, ce n'est pas le cas des facteurs diastratiques, diaphasiques et pragmatiques. En effet, il peut s'avérer extrêmement difficile de déterminer, dans un segment de discours, le poids respectif de chacun de ces facteurs sur la variation. Nous devons en tenir compte dans notre analyse des méthodes adéquates à l'analyse de la diversité des formes interrogatives en français. Quant aux termes « variable » et « variante », leur portée varie passablement selon qu'on les emploie pour des descriptions grammaticales ou sociolinguistiques. Dans ce travail, nous tenterons de réserver les termes « variables » et « variationnisme » au contexte du paradigme labovien de l'analyse de la variation. En revanche, nous utiliserons les mots « variation » et « variabilité » dans un sens plus général, sans forcément les rattacher à une théorie particulière.

2 Le domaine de la question

2.1 Généralités

Le point 2.1 nous servira à donner un aperçu des principales caractéristiques de l'interrogation en français, tandis que des considérations plus particulières seront faites dans les points 2.2 et 2.3. La question sera abordée ici dans une perspective globale, c'est-à-dire aussi bien dans ses aspects formels que fonctionnels. Pour illustrer les différents types de questions que nous présenterons, nous recourrons soit à des exemples littéraires²³, soit à des exemples tirés d'études sur la question.

2.1.1 La forme interrogative et l'acte de langage de question

Lorsqu'on parle de *question*, on peut penser aux structures grammaticales qui leur servent de support ou à la fonction communicative liée à l'acte d'interroger. Certaines études sur la question s'orientent plus vers ses caractéristiques formelles alors que d'autres s'intéressent prioritairement à ses fonctions pragmatiques. Souvent, ces deux aspects sont traités conjointement puisque forme et fonction sont étroitement liés. Il faut préciser toutefois que les liens qui unissent formes interrogatives et actes de langage de question ne sont pas univoques. En tant qu'acte de langage, la question se place aux côtés de l'assertion et de l'ordre pour former les trois « actes primitifs »²⁴ du langage :

La question serait ainsi l'un de ces trois « archi-actes » - actes basiques, et vraisemblablement universels, dont tous les autres actes de langage ne seraient que des formes spécifiques. (Kerbrat-Orecchioni, 1991a, 5)

A ces trois actes fondamentaux correspondent trois types de phrases : affirmatives, interrogatives et impératives. Mais il n'est pas rare qu'un acte de langage soit accompli par une structure formelle qui ne lui est traditionnellement pas associée : on parle alors d'actes de langage *indirects*. Ainsi, les phrases interrogatives peuvent remplir d'autres fonctions que celles d'interroger ; elles servent par exemple à formuler des requêtes, comme en (9) :

(9) Tu peux me résumer l'affaire du crédit industriel, maintenant ? (Pennac, 68)

De même, l'acte de question peut être réalisé en l'absence de toute structure interrogative. Par exemple, une phrase affirmative comme (10) peut parfaitement fonctionner comme une question :

(10) Je veux savoir ton nom. (Koltès, 25)

Les termes « question » et « interrogation » sont donc ambivalents : ils peuvent référer autant à un acte de langage qu'à une structure formelle, ces deux aspects étant souvent (mais pas toujours) liés. Dans le domaine de la question, Catherine Kerbrat-Orecchioni recommande de « bien distinguer dans la terminologie le signifiant et le signifié pragmatique (le support de la valeur illocutoire et cette valeur elle-même) » (1991a, 12). Elle propose de désigner le signifiant de la question par le syntagme « structure (ou phrase) interrogative », tout en

²³ Les exemples littéraires proviennent de trois œuvres différentes : *La chute* d'Albert Camus (roman), *La fée carabine* de Daniel Pennac (roman) et *Roberto Zucco* de Bernard-Marie Koltès (pièce de théâtre).

²⁴ Terme de A.M. Diller, cité par Kerbrat-Orecchioni (1991a, 5)

soulignant qu' « il est souvent dans ce domaine bien difficile de démêler signifiant et signifié » (1991a, 13).

2.1.2 Caractéristiques formelles

Les structures interrogatives sont nombreuses en français, si bien que l'interrogation serait « l'une des formes de la langue française qui présente les plus nombreuses variations » (*La grammaire d'aujourd'hui*, 347). Nous n'introduisons ici que quelques-unes des distinctions qui peuvent être établies entre les nombreuses structures interrogatives (les points 2.2 et 2.3 nous permettront de traiter la diversité formelle des questions plus en détail). Les structures interrogatives indirectes ne seront pas traitées dans ce travail, leurs caractéristiques formelles et fonctionnelles présentant des particularités spécifiques au discours indirect. En effet, l'interrogation indirecte est « contenue dans une phrase qui peut être énonciative, injonctive ou interrogative ». De plus, elle ne partage « ni l'intonation, ni la ponctuation (point d'interrogation) de l'interrogation directe » (*Le Bon Usage*, 484-485).

2.1.2.1 Questions totales et partielles

On peut classer les questions directes en deux grandes catégories, suivant que l'interrogation porte sur l'intégralité de la phrase ou sur un de ses composants. Dans le premier cas, il s'agit de questions *totales* (voir (11) et (12)), alors que dans le deuxième cas, on parle de questions *partielles* (voir (13) et (14)).

(11) Est-ce que je vous intimide ? (Koltès, 56)

(12) Vous estimez qu'il a une tête de tueur ? (Camus, 44)

(13) D'après toi, qu'est-ce qu'ils cherchaient ? (Pennac, 136)

(14) Comment as-tu quitté les rails, Roberto ? (Koltès, 17)

Les questions totales sont susceptibles de recevoir une réponse en oui/non, ce qui n'est en principe pas possible pour les questions partielles. Ces dernières comportent toujours un mot interrogatif (*qui, quand, où, pourquoi*, etc.), qui indique sur quel composant de la phrase porte la question. Enfin, les questions dites « alternatives » ou « disjonctives », comme en (15), ne comprennent pas de mot interrogatif mais n'appellent pas non plus de réponse en oui/non. Elles forment ainsi une troisième catégorie de questions :

(15) Alors, Ponthard, cette déposition, vous la signez, ou je vous tue ? (Pennac, 275)

2.1.2.2 Les structures interrogatives

Les structures interrogatives sont variées en français : les questions partielles, autant que les questions totales, peuvent correspondre à plusieurs schémas syntaxiques différents. La plupart des grammaires en distinguent trois :

- l'inversion du verbe et du sujet : exemple (14)
- l'utilisation de la particule *est-ce que* : exemples (11) et (13)
- l'ordre sujet-verbe (SV) : exemple (12)

Cette dernière catégorie de phrase interrogative est parfois qualifiée d' « intonative » : il est vrai que dans le cas d'une structure SV, et en l'absence de mot interrogatif, c'est souvent l'intonation qui permet de distinguer les questions des assertions. Toutefois, l'étude de Louise Fontaney (1991) sur l'intonation des questions montre que ce n'est pas toujours le cas. En effet, Fontaney observe que l'intonation montante n'apparaît pas systématiquement pour les

questions totales à structure SV. Si la plupart de ces questions sont bien assorties d'une intonation ascendante (à divers degrés), une partie d'entre elles présentent des schémas prosodiques identiques à l'assertion. Ainsi, l'exemple (16), fourni par Fontaney, fonctionne comme une question même s'il présente une structure syntaxique et une intonation caractéristiques de l'assertion :

(16) Ouais d'toute manière tu...t'as rien contre les homosexuels (Fontaney, 1991)

Si de telles structures peuvent être interprétées comme des questions, c'est grâce au « savoir relatif des participants » : dans l'exemple (16), c'est « l'interlocuteur qui détient le savoir », de sorte que l'énoncé peut difficilement être interprété comme une assertion (Fontaney, 1991, 129). En effet, il est rare d'asserter sur des opinions ou des sentiments appartenant à autrui. Pour expliquer ce phénomène, Labov distingue les « faits-A », les « faits-B » et les « faits-AB », posant A et B comme deux interlocuteurs d'un dialogue. Les faits-A ne sont connus que de A, les faits-B que de B alors que les faits-AB sont connus des deux interlocuteurs. Labov suggère que lorsqu'un énoncé est formulé sous forme assertive²⁵ mais qu'il porte sur un fait ne concernant que l'interlocuteur, il est alors interprété comme une question.²⁶ On constate ainsi que l'intonation n'est pas l'unique marque distinctive des questions sans inversion, *est-ce que* ou mot interrogatif. Dès lors, nous préférons renoncer au terme de « question intonative » et garder celui de « structure SV ».

Ainsi les interrogations totales à structure SV sont-elles interprétées comme des questions grâce à leur intonation et/ou à leur contenu, ce dernier devant être mis en relation avec le savoir respectif des interlocuteurs. Les structures interrogatives SV ne font donc intervenir aucune marque syntaxique caractéristique de l'interrogation. En revanche, les deux autres structures interrogatives susmentionnées présentent des marques syntaxiques spécifiques ou étroitement liées à l'acte de question. Ainsi, la particule interrogative *est-ce que* ne s'utilise en français que pour former des questions. *Le Bon Usage* la décrit comme un « introducteur de l'interrogation » : *est-ce que* apparaît en effet en début de phrase, indiquant d'emblée une valeur interrogative. Aussi qualifiées d'« interrogations périphrastiques », les questions construites avec *est-ce que* conservent (normalement) l'ordre canonique SV.

La troisième structure de l'interrogation répertoriée par toutes les grammaires consiste en une inversion de l'ordre canonique sujet-verbe. Dans les structures interrogatives, les phénomènes d'inversion se classent en deux catégories, suivant la nature du sujet. Si le sujet est un pronom personnel ou *ce*, on parle d'inversion « simple » et le sujet suit immédiatement le verbe :

(17) Puis-je, monsieur, vous proposer mes services, sans risquer d'être importun ? (Camus, 7)

Si le sujet n'est pas un pronom personnel ni *ce*, on doit recourir à l'inversion « complexe » : le sujet se place alors devant le verbe, mais il est repris par un pronom personnel, ce dernier suivant immédiatement le verbe pour créer une inversion :

(18) Les mensonges ne mettent-ils pas finalement sur la voie de la vérité ? (Camus, 127)

L'inversion (simple ou complexe) du verbe et du sujet est souvent associée à l'interrogation, mais elle peut également apparaître dans des énoncés non interrogatifs :

²⁵ C'est-à-dire une structure SV et un schéma prosodique caractéristique de l'assertion.

²⁶ Voir le compte-rendu de la théorie de Labov dans Kerbrat-Orecchioni (1991b, 93).

Elle est également déclenchée par divers adverbiaux antéposés (*Peut-être a-t-elle décidé de s'en aller, Sans doute ton frère a-t-il raison*), et elles apparaissent aussi dans certaines subordonnées à valeur de supposition (*Si courageux soient-ils, Aussi précieuse la devinâton...*). (Béguelin-Corminbœuf, 2005, 69)

Une quatrième structure interrogative, caractérisée par la présence du morphème interrogatif *-ti* (dont la graphie peut varier), est utilisée par certains locuteurs du français et mentionnée par quelques grammairistes. Cette tournure est attestée dès le XVIII^{ème} siècle, mais son usage se limite aujourd'hui à quelques régions de la francophonie, principalement le Québec et la Normandie. *-ti* s'est formé sur l'inversion du pronom personnel de troisième personne (*est-il, va-t-il, mangent-ils*), mais son emploi s'est généralisé à toutes les personnes. Il apparaît directement après un groupe verbal qui garde en principe l'ordre sujet-verbe, comme le montrent les exemples (19) et (20), tirés du *Bon Usage* :

(19) ça va-t-il, mon gars ? (H.Bazin, *Vipère au poing* <Goosse-Grevisse>)

(20) Vous entrez-t'y, Docteur ? (Céline, *Voyage au bout de la nuit* <Goosse-Grevisse>)

Le dernier type de structures interrogatives dont nous rendons compte est celui des « phrases avec appendice » (le terme est d'Andrée Borillo). Elles ne constituent pas une catégorie de structure interrogative au même titre que celles décrites plus haut puisqu'elles peuvent présenter un ordre inversé du verbe et du sujet, un ordre sujet-verbe ou *est-ce que*. Leur particularité réside dans l'appendice qui leur est joint. Borillo (1978) distingue deux types d'appendices, le premier étant constitué par une construction verbale de type *tu crois, je me demande*, etc., et le second par une particule interrogative de type *hein, n'est-ce pas*, etc. Dans le premier cas, Borillo parle de « structures interrogatives postposées » et dans le deuxième cas, de « questions-reprises ». Dans sa thèse de 1978, Borillo propose une étude détaillée des « phrases avec appendice », rendant compte aussi bien de leurs propriétés structurelles que fonctionnelles.

2.1.3 Valeur illocutoire

Le point 2.1.2 nous a permis d'aborder la diversité formelle des structures interrogatives. Au niveau pragmatique, les questions présentent également une grande diversité car elles peuvent remplir des fonctions communicatives très variées. Par exemple, les questions qui représentent de véritables demandes d'information fonctionnent différemment des questions dites « rhétoriques ». Dans le cas des demandes d'information, le locuteur cherche à obtenir une information qui lui manque et ne préjuge aucunement de la réponse qui lui sera fournie. Borillo explique que les questions totales qui représentent de vraies demandes d'information peuvent être paraphrasées par une interrogative disjonctive sans que cela n'altère leur signification: ainsi, si une question simple du type « il est déjà parti ? » signifie la même chose que son alternative disjonctive « il est déjà parti, ou pas ? », il s'agit alors d'une véritable demande d'information, les réponses *oui* et *non* étant équiprobables (1978, 370-71). En revanche, une question comme (21), tirée des *Fables* de La Fontaine, n'appelle pas identiquement une réponse positive ou négative :

(21) Crois-tu qu'après un tel outrage

Je me doive fier à toi ? (La Fontaine, *Fables* <Bonhomme, 2005, 206>)

En fait, la question implique d'elle-même une réponse négative, de sorte que l'interlocuteur n'est pas tenu de répondre. En effet, l'une des caractéristiques des questions rhétoriques est de ne pas appeler de réponse :

Elle est le cas où le locuteur, fort de son jugement sur la proposition qu'il émet, pourrait très bien ne pas la présenter sur le mode interrogatif mais carrément l'asserter [...]. (Borillo, 1981, 6)

Entre les véritables demandes d'information et les questions rhétoriques, on trouve toute une gamme de questions diversement orientées (c'est-à-dire, qui présument plus ou moins fortement de la réponse demandée). Beaucoup de chercheurs se sont demandé s'il était possible de déterminer une force illocutoire commune à toutes les questions. Mais les avis divergent et dépendent pour beaucoup de la représentation que les auteurs se font de *la question*, c'est à dire de sa définition même. Comme le rapportent Berrendonner (2005) et Kerbrat-Orecchioni (1991a), les théoriciens des actes de langage ont souvent tenté de décrire l'acte de question sur la base de deux conditions, que nous nommerons, comme Berrendonner, α et β . La condition α concerne l'incertitude du locuteur et stipule que l'on ne pose une question que lorsqu'il nous manque une information. Quant à la condition β , elle a trait à l'effet des questions sur l'interlocuteur et implique que toute question appelle une réponse.

Berrendonner et Kerbrat-Orecchioni estiment que les conditions α et β ne sont pas adéquates pour distinguer la question des autres actes de langage. Selon Kerbrat-Orecchioni, la condition α est trop restreinte, étant donné qu'un locuteur peut poser une question sur un fait qu'il connaît déjà (pour une question d'examen par exemple). Quant à la condition β , elle la trouve trop large, tout acte de langage pouvant susciter une réaction comparable à une réponse. Berrendonner constate également que β n'est pas spécifique à l'acte de questionner ; il ajoute que les questions n'appellent pas forcément de réponse (il parle dans ce cas de questions « monologiques »).

Si Kerbrat-Orecchioni et Berrendonner refusent tous deux de définir la question sur la base de α et β , ils ne s'accordent pas pour autant sur la définition à adopter : pour la première, le domaine de la question a une portée plus restreinte que pour le second. Kerbrat-Orecchioni définit la question comme un

énoncé qui se présente comme ayant pour finalité principale d'obtenir de L2 un apport d'information. (1991a, 14)

Elle précise que cette définition rejoint celles de Ducrot, Borillo, Goffman, et de nombreux autres linguistes. Pour les partisans de cette définition, les questions rhétoriques ne font pas partie du domaine de la question étant donné qu'elles ne suscitent pas d'apport d'information. Elles sont parfois appelées « vraies fausses questions », en opposition avec les « fausses fausses questions ». Ces dernières sont représentées par les questions d'examens ou d'interviews, pour lesquelles le locuteur connaît déjà la réponse mais souhaite, pour une raison ou une autre, que l'interlocuteur la lui fournisse. Les « fausses fausses questions » font bien partie du domaine de la question tel que le conçoit Kerbrat-Orecchioni. Elles ne représentent toutefois pas le *prototype* de l'acte de question, ce dernier étant constitué par les questions « à réponses ignorées » (1991, 17).

Berrendonner fait partie de ceux qui souhaitent pouvoir englober toutes les variétés de questions (rhétoriques y compris) à l'intérieur d'un même modèle théorique. Pour ce faire, il renonce à l'approche illocutoire–interactionniste, considérant qu'elle ne permet pas d'appréhender la question de manière généralisante. Il préfère

décrire la question non pas comme une opération « sociale » qui serait exercée directement sur le destinataire en vue d'en obtenir une réaction, mais comme une opération « cognitive » accomplie sur le savoir partagé commun aux deux interlocuteurs. (Berrendonner, 2005, 10)

Pour expliquer le fonctionnement de la question, Berrendonner pose la conversation comme un « jeu à trois », faisant intervenir, en plus de deux locuteurs (L1 et L2), « un ensemble de représentations publiquement partagées, ou « mémoire discursive » (M) » (2005, 2). Les énonciations des locuteurs intégreraient M à travers deux opérations simultanées : un acte de langage primaire et

un « geste intonatif » (et éventuellement d'autres gestes ou mimiques), qui exprime(nt) une posture psychologique ou épistémique adoptée à l'égard de cette assertion. (2005, 10)

Selon cette conception, une question serait une « assertion non assumée » (Berrendonner, 2005, 13), puisqu'elle introduit dans M une assertion primaire qui est réfutée, à différents degrés, par la posture modale secondaire qui l'accompagne. A partir de là, les diverses combinaisons possibles entre l'opération primaire et secondaire engendrent différents types de questions. Dans le cas d'une véritable demande d'information, L1 fait entrer une assertion dans M tout en indiquant qu'il ne l'assume aucunement. L'état de M étant alors inconsistant, L2 devra approuver ou réfuter la validité de l'assertion pour laisser M dans un état acceptable. Quant à la question rhétorique, elle consiste à réfuter primordialement des éléments qui sont bien ancrés dans M, puis, secondairement, à ne pas en assumer la réfutation : la réfutation secondaire n'est donc pas validée et les éléments déjà implantés dans M y demeurent.

Comme Berrendonner, Marc Bonhomme (2005) définit *la question* de manière élargie et souhaite pouvoir traiter les questions rhétoriques et les demandes d'information à l'intérieur d'un même modèle. Il propose pour cela une définition globale de la question (sa « valeur illocutoire matricielle ») qui peut s'adapter à toutes les variétés de question grâce à son caractère modulaire. A partir d'une valeur illocutoire commune à toutes les questions, soit

l'expression adressée d'un doute sur une assertion ou un fait. (Bonhomme, 2005, 195)

Bonhomme propose des extensions basées sur des paramètres véridictaires et interactionnels. Tandis que le paramètre véridictaire indique le degré de sincérité du locuteur (le doute peut être sincère, mixte ou feint), le paramètre interactionnel est défini par la nature de la réponse, cette dernière pouvant être hétérolocutée ou autolocutée.²⁷ Selon la nature des paramètres véridictaires et interactionnels, différents types de question peuvent être distingués. Ainsi, pour une demande d'information, le doute est sincère et la réponse hétérolocutée alors que pour la question rhétorique, le doute est feint et la réponse autolocutée. Bien que Kerbrat-Orecchioni ait une vision de la question plus restreinte que Bonhomme, elle considère également que les questions partagent toutes une même valeur illocutoire (selon sa définition donnée plus haut). C'est selon elle au niveau perlocutoire que se situent les effets plus particuliers (et extrêmement variables) des questions.

Nous avons déjà mentionné comme autant de types de questions à valeur et fonctionnement pragmatiques divers, les questions rhétoriques, les questions d'examen et les demandes d'information. Nous avons également établi l'existence de questions plus ou moins orientées, situées entre les deux extrêmes constitués par la demande d'information, dont les chances de recevoir une réponse positive ou négative sont égales, et par la question rhétorique qui contient en elle-même une réponse implicite. Les questions orientées sont nombreuses et

²⁷ La réponse autolocutée peut être formulée explicitement ou signifiée implicitement.

de différentes natures, mais elles contiennent toutes des éléments qui permettent de déterminer plus ou moins clairement l'orientation de la réponse attendue. Les indicateurs de l'orientation de la réponse sont souvent de nature intonative : comme le remarque Fontaney (1991), les schémas prosodiques des questions sont différents lorsque le locuteur « croit savoir » et lorsqu'il exprime une véritable demande d'information. Mais l'orientation d'une question peut aussi être suggérée par des modalisateurs (« probablement », « je suppose », etc.) ou des éléments situationnels.²⁸

Pour clore cet aperçu des différentes possibilités de fonctionnement pragmatique des questions, nous mentionnerons encore les questions préliminaires, qui servent à introduire un acte souvent menaçant (*je peux vous poser une question indiscrete ?*²⁹), les questions notificatives (lorsque « le locuteur, en demandant à son interlocuteur s'il connaît une information, vise en fait à la lui communiquer »),³⁰ et les questions délibératives, que l'on se pose à soi-même. Les différents effets perlocutoires créés par les questions sont nombreux et il n'est pas possible d'en établir ici une liste exhaustive. L'inventaire des différents types de questions pourrait se poursuivre longuement, non seulement sur la base de leurs caractéristiques formelles et pragmatiques, mais aussi sur d'autres critères, comme celui de « la *relation* que la question entretient avec *son cotexte antérieur immédiat* » (Kerbrat-Orecchioni, 1991a, 21). C'est sur ce critère que Kerbrat-Orecchioni fait reposer la spécificité des « questions-répliques », qui reprennent tout ou partie de l'énoncé qui les précède directement, généralement pour en clarifier l'un ou l'autre aspect, comme en (22) :

(22) - Le vieux Risson n'est pas rentré, cette nuit.
(Merde. Ça, je n'aime pas)
- *Comment ça, pas rentré ?* (Pennac, 258)

Les questions-répliques ont souvent une fonction métacommunicative (lorsqu'elles interrogent sur le code de communication) mais elles peuvent aussi fonctionner comme des questions rhétoriques. Leurs particularités formelles et fonctionnelles, dues pour la plupart à leur dépendance à un énoncé conjoint, en font un objet d'étude particulier, rarement pris en compte dans les études générales sur la question.³¹

2.1.4 Détermination de l'objet d'étude

La base de nos recherches est constituée par l'existence de différentes structures interrogatives en français, notre but étant de découvrir comment il est possible d'en expliquer l'usage. *La question* sera abordée autant du côté de ses propriétés formelles que de son fonctionnement pragmatique et interactionnel : en effet, pour expliquer pourquoi une forme est préférée à une autre dans un contexte donné, il faut étudier l'influence de nombreux facteurs, ceux-ci pouvant être de nature linguistique aussi bien qu'extralinguistique. Ainsi, nous n'excluons *a priori* aucun type de question ni aucune structure interrogative, notre intérêt résidant dans l'explication des choix des locuteurs en matière de question, là où plusieurs structures syntaxiques sont disponibles.

²⁸ Les éléments (visuels, auditifs, etc.) du contexte d'interaction donnent souvent lieu à des inférences, qui demandent à être confirmées par l'interlocuteur. Voir à ce propos l'article de Véronique Traverso (1991). Sur les divers « orientateurs » de la question, voir l'article de Kerbrat-Orecchioni (1991b, 100-102).

²⁹ Exemple de Kerbrat-Orecchioni (1991a, 23)

³⁰ Borillo (1978, 616)

³¹ Pour plus de détails et une typologie des questions-répliques, voir Borillo (1978)

2.2 Etudes quantitatives

2.2.1 Présentation

A partir des années 1960, grâce à la simplification des techniques d'enregistrement, la linguistique commença à travailler de plus en plus sur la base de grands corpus tirés de la langue orale. Plusieurs études menées dans les années 60-70 s'intéressent à l'emploi des diverses formes de la question dans le langage oral: basées sur un large éventail de questions, elles quantifient les occurrences des divers types de structures interrogatives pour en déterminer les fréquences d'utilisation. Purement quantitatives, ces études ne traitent pas la question comme une *variable* au sens labovien du terme. En effet, elles ne présupposent pas d'équivalence sémantique entre les différentes formes de questions, ni de différenciations sociales dans leur utilisation; la question y est plutôt traitée comme un objet grammatical pouvant correspondre à différentes structures.

Nous aborderons les approches quantitatives de l'interrogation et les différents problèmes méthodologiques qu'elles doivent affronter à travers trois études publiées entre 1965 et 1973 : il s'agit de Pohl, « Observations sur les formes d'interrogation dans la langue parlée et dans la langue écrite non littéraire » (1965), Söll « L'interrogation directe dans un corpus de langage enfantin » (1983, mais l'article original en allemand date de 1971) et Behnstedt *Viens-tu ? Est-ce que tu viens ? Tu viens ?* (1973). L'étude de Behnstedt est sans aucun doute la plus vaste, fournie et détaillée des trois études. Pour des raisons de compréhension, nous nous sommes aidée du compte rendu de l'étude de Behnstedt donné dans Al (1975) en français, tout en recourant fréquemment au texte original en allemand. Behnstedt a tiré ses données de plusieurs sources différentes: un de ses corpus a été établi alors qu'il voyageait comme passager d'un poids lourd, l'autre provient d'entretiens menés avec des personnes de classe moyenne et un troisième est constitué de diverses émissions radiophoniques. Le corpus de Pohl provient de conversations observées entre son père et sa mère, tous deux belges et âgés respectivement de 80 et 70 ans. Il a également analysé des questions tirées de la correspondance de ses parents, mais nous n'avons pas pris en compte cette partie de l'étude. Quant à Söll, il a observé l'usage des questions fait par des enfants de 9 ans, lors de conversations libres et de jeux.

2.2.2 Traitement et classement des données

Une étape cruciale pour les études quantitatives est l'établissement de toutes les « catégories » devant être quantifiées. Dans le cas des études quantitatives sur la question, ces catégories sont formées par les différents types de structures syntaxiques de l'interrogation. Une fois que les catégories sont définies, il faut procéder au comptage de toutes les occurrences de chaque catégorie établie (on obtient alors des fréquences *observées*). Finalement, on peut calculer de manière statistique le pourcentage d'utilisation de chaque forme (on obtient cette fois-ci des fréquences *relatives*). La grande diversité des structures interrogatives, dont nous n'avons eu qu'un aperçu au point 2.1, permet diverses approches du sujet. L'analyse des études de Pohl, Söll et Behnstedt nous permettra d'appréhender des différences méthodologiques, tant dans l'établissement des catégories de base que dans les moyens mis en œuvre pour les compter.

Avant d'établir les différentes structures syntaxiques de l'interrogation, il faut déterminer d'une manière plus générale les types de questions incluses ou exclues de l'analyse quantitative. Les trois études ne prennent en compte que les questions directes qui comprennent au moins un verbe conjugué. Pohl explique qu'il a renoncé à inclure dans son

analyse « certaines phrases d'apparence elliptique dont le caractère interrogatif est parfois discutable » (503) : il s'agit de la catégorie des *questions-répliques* dont nous avons rendu compte au point 2.1. Söll ne signale aucun type de questions écartées de l'analyse. Il indique tout de même que certaines questions totales fonctionnant comme demande de confirmation (notamment celles recourant aux particules *hein*, *non*, *oui* et *quoi*) n'auraient pas pu apparaître à l'intérieur d'une structure syntaxique présentant une inversion ou *est-ce que*. Il a décidé d'inclure ces questions dans la catégorie SV (sujet-verbe), tout en signalant la présence et le nombre, ce qui permet à qui veut de les y soustraire. Al ne donne aucune indication quant à d'éventuelles questions exclues de l'étude de Behnstedt ; il semble que ce dernier n'en fasse pas mention, mais notre connaissance partielle du texte complet en allemand ne permet pas de l'affirmer.

2.2.2.1 Questions totales

Les catégories de structures interrogatives sont plus nombreuses et complexes au niveau des questions partielles qu'au niveau des questions totales. Pour ces dernières, les études quantitatives se basent essentiellement sur les trois grandes catégories de structures interrogatives communément distinguées par les grammaires, soit : l'inversion, l'utilisation d'*est-ce que* et l'ordre sujet-verbe. Le tableau 1 permet de comparer les différentes structures de l'interrogation totale telles qu'elles ont été établies par Pohl, Söll et Behnstedt. Pour favoriser les comparaisons, nous avons renommé toutes les catégories à partir d'abréviations dont la signification est donnée plus bas.

Tableau 1 - Catégories de structures syntaxiques pour l'interrogation totale

Pohl	Söll	Behnstedt
SV	SV	SV
ESV : 1. <i>est-ce que</i> 2. <i>n'est-ce pas que</i> 3. <i>hein que</i> 4. <i>-ti</i>	ESV	ESV
VS : 1. VScl (inversion simple) 2. SsnVScl (inversion complexe)	VS : 1. VScl 2. SsnVScl	VS : 1. VScl 2. SsnVScl
		SV <i>-ti</i>

S = sujet / Scl = sujet pronom clitique / Ssn = sujet syntagme nominal / V = verbe / E = *est-ce que* /

La principale difficulté classificatoire des structures de l'interrogation totale a trait au statut des questions faisant apparaître des particules autres qu'*est-ce que*, telles que *hein*, *non*, *-ti* etc. Les auteurs des trois études ont différentes manières de classer de telles structures. Pohl distingue, à l'intérieur de la catégorie ESV, les questions qui recourent aux particules *n'est-ce pas que*, *hein que*, et *-ti*. Behnstedt choisit lui de créer une catégorie spéciale pour les questions qui présentent la particule *-ti*. Quant à Söll, il n'a récolté aucune question avec *-ti* dans son corpus. Comme Behnstedt, il réserve la catégorie ESV aux questions exprimées au moyen d'*est-ce que*. Il considère que les interrogatives recourant aux particules *hein*, *non*, *oui* et *quoi* sont des demandes de confirmation et il les groupe dans la catégorie SV. Il indique en

outre que ces questions ne pourraient être exprimées par une autre structure interrogative sans perdre leur valeur sémantique de demande de confirmation.

2.2.2.2 Questions partielles

Au niveau des questions partielles, les différentes structures interrogatives sont nombreuses et syntaxiquement complexes. Elles posent plusieurs problèmes au moment de les classer, mais également au moment d'en établir les fréquences d'utilisation. Pohl, Söll et Behnstedt recourent à diverses méthodes pour faire face à cette complexité. Nous prendrons comme point de départ la liste des diverses structures syntaxiques possibles, telles qu'elles ont été établies par les trois auteurs.

Tableau 2 - Catégories de structures syntaxiques pour l'interrogation partielle

Pohl	Söll	Behnstedt
KSV	KSV	KSV
KESV	KESV : 1. K <i>est-ce que</i> SV 2. K <i>c'est que</i> SV 3. K <i>que</i> SV	KESV : 1. K <i>est-ce que</i> SV 2. K <i>c'est que</i> SV 3. K <i>que c'est que</i> SV
SVK	SVK	SVK
K(S)VS 1. KVS 2. KSsnVScl	KVS ³²	K(S)VS 1. KVS 2. KSsnVScl
		K <i>que</i> SV

K = mot interrogatif / S = sujet / Scl = sujet pronom clitique / Ssn = sujet syntagme nominal / V = verbe / E = *est-ce que* /

Les divergences que l'on peut observer dans ces trois classements proviennent du fait que les auteurs ne s'entendent pas sur le traitement des tours dits parfois « concurrents » d'*est-ce que*. Les exemples suivants, tirés de l'étude de Behnstedt (40-45), font apparaître des segments proches d'*est-ce que* dans la mesure où ils reprennent certains ou la totalité de ses constituants (à savoir *ce*, *est*, et *que*) :

- (23) Comment qu'ils disent déjà ?
- (24) Où c'est qu'i' crèche?
- (25) Qu'est-ce t'en penses ?
- (26) Quand c'est qu'c'est qu't'es v'nu la dernière fois ?
- (27) Qu'est-ce que c'est que ça ?

Le statut syntaxique de ces différents tours ainsi que leur éventuelle affiliation à la périphrase *est-ce que* sont discutés. L'étendue de ce travail ne nous permettant pas de traiter cette question en détail, nous nous contenterons de présenter une classification de ces différents tours, établie par Hector Renchon (1967, 157-173).

³² Söll n'a rencontré aucune inversion complexe dans les interrogations partielles de son corpus, ce qui explique qu'il n'ait créé aucune catégorie KSsnVScl. Pohl a rencontré des inversions complexes dans la partie écrite de son corpus, et Behnstedt dans les entretiens radiophoniques.

Adoptant une formulation très normative, Renchon qualifie les différents tours présentés à travers les exemples (23) à (27) de « pathologies de l'interrogation périphrastique » ; il en distingue trois grandes catégories. Premièrement, la « désinversion », qui correspond à l'exemple (24), ne fait que rétablir l'ordre sujet-verbe de la particule inversée *est-ce que*. Deuxièmement, la « désagrégation », due selon Renchon à la « paresse articulatoire » (160), est illustrée par les exemples (23) et (25). Renchon mentionne que la désagrégation peut aussi toucher la première syllabe d'*est-ce que* : il en donne plusieurs exemples, dont :

(28) Mais comment c'que moi j'ai jamais remarqué ça ? (Virgile, *Dialogue de la semaine, Pourquoi pas ?* <Renchon, 1967, 161>)

Les exemples (26) et (27) sont, selon Renchon, des formes « surpériphrastiques » qui proviennent de la « réduplication », d'*est-ce que*. Les formes surpériphrastiques peuvent se combiner avec un phénomène de désinversion (26) ou de désagrégation (29):

(29) Pourquoi qu'c'est qu'ils m'attendent ? R. Benjamin, *Gaspard* <Renchon, 1967, 171>

La classification de Renchon peut être discutée sur certains points. S'il est assez clair que *ce que* et *est-ce* sont des réductions phonologiques d'*est-ce que*, il est plus difficile de garantir que *c'est que* et *que* sont bel et bien dérivés d'*est-ce que*. Le statut de *que* est particulièrement problématique: Renchon lui-même signale qu'il peut également être analysé comme une « simple cheville syntaxique »³³ non spécifique aux énoncés interrogatifs et indépendante d'*est-ce que*.

Pohl est le seul à ne pas distinguer les différents tours concurrents d'*est-ce que*. Pour lui, *ce que*, *est-ce*, et *que* sont des « variantes morphologiques qui relèvent moins de la syntaxe que de la phonétique ou de la « parlure » » (504) ; il les réunit donc tous à l'intérieur de la catégorie KESV. En revanche, Behnstedt et Söll établissent des catégories distinctes pour chacun des tours périphrastiques rencontrés dans leur corpus. Même si Söll n'est pas convaincu que les tours *c'est que* et *que* tirent leur origine d'*est-ce que*, il décide néanmoins de les traiter comme des tours périphrastiques. Pour Behnstedt, *c'est que* est bien une désinversion d'*est-ce que* mais *que* ne provient pas de la réduction d'*est-ce que*. Il ne considère donc pas les questions en KqueSV comme des tours périphrastiques.

Passée cette première difficulté d'ordre classificatoire, l'analyse de l'utilisation des différentes formes de l'interrogation partielle pose toute une série de problèmes au moment de définir les fréquences d'utilisation de chaque forme. Ces problèmes sont dus aux nombreuses contraintes syntaxiques qui pèsent sur la construction d'une question partielle. En effet, l'acceptabilité des structures de l'interrogation partielle varie selon la nature de ses constituants. Par exemple, la nature du mot interrogatif détermine dans une certaine mesure les agencements syntaxiques possibles de l'interrogation partielle. Comme l'explique Pohl (507), « chaque mot interrogatif est lié à des habitudes syntaxiques particulières », de sorte que certaines structures syntaxiques ne sont possibles qu'avec certains mots syntaxiques et vice-versa. Cela est dû au fait que les mots interrogatifs correspondent à différentes catégories grammaticales³⁴ et remplissent différentes fonctions dans la phrase. Ainsi, lorsque *qui* ou

³³ Terme de G. et R. Le Bidois, cités par Renchon (1967, 164). Huirauds et Lerchs préfèrent le terme de « conjonction minimum » (cités par Behnstedt, 1973, 37).

³⁴ La *Grammaire méthodique du français* et *La grammaire d'aujourd'hui* classifient les différents mots interrogatifs comme suit : PRONOMS : *qui, que, quoi* DETERMINANTS : *quel(le)(s), combien* de ADVERBES : *comment, où, pourquoi, quand, combien*. Elles mentionnent également l'existence de formes composées (*lequel*).

qu'est-ce qui fonctionnent comme sujet, ils ne peuvent apparaître qu'en tête de phrase et ils excluent la possibilité d'inversion du verbe et du sujet. Il est donc impossible de voir apparaître *qui* et *qu'est-ce qui* dans une structure d'ordre KSV, par ailleurs très fréquente dans des phrases du type *Pourquoi tu pars ?* ou *Comment il fait ?* Les mots interrogatifs *que* et *quoi* engendrent également des contraintes syntaxiques spécifiques lorsqu'ils correspondent à un objet direct ou à un attribut. Quant aux adverbes interrogatifs (*où*, *comment*, *quand*, *pourquoi* et *combien*), ils ont l'avantage de partager, grossièrement, les mêmes habitudes syntaxiques. A noter tout de même que *pourquoi* est moins propice à apparaître dans des structures KVS à sujet nominal que les autres adverbes interrogatifs :

(30) Où sont passées mes cigarettes ? (Al, 1975, 58)

(31) *Pourquoi est partie sa femme ? (Al, 1975, 89)

Sur la base d'exemples comme (30) et (31), Al (1975) déclare que l'inversion nominale (KVS_{sn}) est incompatible avec la présence de *pourquoi*. Cette configuration n'apparaît effectivement pas dans les corpus de Pohl, Söll et Behnstedt, mais elle n'est toutefois pas « impossible » : des exemples attestés en sont donnés notamment dans Renchon (1967, 44-45). Ce dernier précise que l'inversion du sujet nominal était d'usage avec *pourquoi* jusqu'au XVII^e siècle.

Le type de sujet des questions partielles constitue un autre élément susceptible d'engendrer des contraintes syntaxiques particulières. En effet, l'acceptabilité des structures interrogatives partielles peut varier en fonction de la nature nominale ou pronominal du sujet.³⁵ L'inversion complexe (KS_{sn}VS_{cl}) ne peut avoir lieu que si l'interrogation comporte un sujet nominal (le premier sujet est nominal et le second pronominal). En revanche, l'inversion simple (KVS) peut, parfois mais pas toujours, avoir lieu avec un sujet nominal, comme c'est le cas de (30). L'acceptabilité discutable de (31) est due, nous l'avons vu, à une restriction particulière entraînée par l'adverbe interrogatif *pourquoi*. Mais il existe d'autres cas de figure où l'inversion simple du verbe et du sujet n'est possible qu'avec un sujet pronominal : par exemple lorsque le verbe est suivi d'un complément ou d'un attribut. Nous avons construit les exemples (32) et (33) pour montrer comment la présence d'un complément (ici *les œufs*) peut rendre l'inversion simple du sujet nominal improbable³⁶ :

(32) Comment mange-t-il les œufs ?

(33)*Comment mange Paul les œufs ?

Ce ne sont là que quelques exemples des nombreuses règles et tendances qui gouvernent la structuration des questions partielles. La nature du mot interrogatif et du sujet ainsi que la combinaison de ces deux éléments sont importants pour déterminer les types de structures possibles. Mais la nature de l'attribut, du complément ou même du verbe est également susceptible d'engendrer des restrictions syntaxiques particulières (voir *Le Bon Usage*, rubrique « phrase interrogative », pour une description détaillée du fonctionnement grammatical de l'interrogation partielle).

³⁵ Suivant *Le Bon Usage*, nous considérons comme « sujets pronominaux », les pronoms personnels ainsi que *ce* et *on*, et nous incluons dans le terme « sujet nominal » le pronom *ça/cela*.

³⁶ Toutefois, selon la longueur du sujet ou du complément, les structures KVS_{sn} peuvent apparaître en présence d'un complément ou d'un attribut verbal (par exemple dans: *Comment peuvent peser des choses qui n'existent pas ?* Ionesco <Goosse-Grevisse> (2007, 497) ou dans *Comment ne serait pas malheureux un enfant sans père ?* Gide <Renchon> (1967, 75)).

Le fait que les contraintes syntaxiques de l'interrogation partielle varient en fonction des constituants de la phrase pose problème lorsqu'il s'agit d'établir les pourcentages d'utilisation de chaque structure interrogative. En effet, il n'est pas toujours possible de garantir l'interchangeabilité des structures interrogatives de la question partielle si les phrases que l'on confronte contiennent des mots interrogatifs différents, des sujets de différentes natures, ou encore d'autres éléments qui entraînent des restrictions syntaxiques particulières. Le risque de comparer des phrases dont les structures syntaxiques ne sont pas interchangeables est d'obtenir des résultats peu représentatifs des fréquences d'utilisations. Prenons le cas du pronom *que/quoi* : au contraire des autres mots interrogatifs, *que/quoi* ne peut apparaître avant une structure SV que s'il est suivi d'*est-ce que*.³⁷ En conséquence, la fréquence d'utilisation de la structure KESV est nettement plus élevée avec *que/quoi* qu'avec les autres mots interrogatifs. Des résultats qui ne montreraient pas l'importance du mot interrogatif *que/quoi* dans le pourcentage global d'utilisation de la structure KESV seraient peu représentatifs de l'utilisation effective de cette structure. Pour une représentation statistique pertinente de l'utilisation des structures interrogatives partielles, il faut donc favoriser les possibilités de commutation entre structures syntaxiques. La seule manière d'y arriver est de trier les différentes interrogations rencontrées. C'est ce que font Pohl, Söll et Behnstedt au moment de présenter les fréquences d'utilisation des structures de l'interrogation partielle.

L'étude de Behnstedt, sensiblement plus large et détaillée que les deux autres, expose ses résultats à travers un grand nombre de tableaux. La plupart de ces tableaux présentent, horizontalement, les 6 structures syntaxiques de l'interrogation partielle préalablement établies (voir tableau 2) et, verticalement, différentes catégories de phrases interrogatives. Ces catégories sont établies en fonction de la nature de certains des constituants de la phrase. Behnstedt présente ainsi plusieurs facettes des liens pouvant exister entre les constituants de la phrase interrogative et sa structure syntaxique. Par exemple, il présente dans un tableau les pourcentages d'occurrence des différentes structures interrogatives lorsque le sujet est formé du pronom *ça* ou *cela*. Dans un autre tableau, il divise toutes les interrogations selon la personne grammaticale du sujet, pour définir le nombre d'occurrence des structures avec chaque personne. Le tableau que l'on peut qualifier de plus « généralisant » est celui qui classe les interrogations partielles en fonction du mot interrogatif qu'elles contiennent ; c'est aussi celui dont la présentation s'approche le plus des comptes-rendus de Pohl et Söll. En effet, les études de Pohl et Söll, d'étendue plus modeste que celle de Behnstedt, se contentent de donner un aperçu global de l'utilisation des différentes formes interrogatives. Même s'ils se veulent le plus englobants possible, ces comptes-rendus doivent présenter un certain tri des données, pour les raisons de pertinence statistique dont nous avons parlé. Les phrases interrogatives sont donc séparées en différentes classes, selon des critères qui concernent tous le type de mot interrogatif et/ou sa fonction dans la phrase. Les différentes classes ainsi obtenues varient légèrement d'un auteur à l'autre : nous les avons reproduites dans le tableau 3 :

³⁷ *que/quoi tu fais ?* vs. *qu'est-ce que tu fais ?*

Tableau 3 - Classes d'interrogations partielles établies pour le calcul des fréquences d'utilisation des différentes structures syntaxiques.

Pohl	Söll	Behnstedt
<i>où</i>	sujet : 1) personnes 2) choses	<i>où</i>
<i>quand</i>	attribut : 1) personnes 2) choses	<i>quand</i>
<i>comment</i>	complément d'objet : 1) personnes 2) choses	<i>comment</i>
<i>pourquoi</i>	complément prépositionnel	<i>pourquoi</i>
<i>combien</i>	questions modales (<i>comment, où, pourquoi, quand</i>)	<i>combien</i>
<i>qui, quoi</i> ou <i>quel</i> , précédé d'une préposition	<i>combien</i>	<i>quoi</i> complément prépositionnel
<i>quel</i> (attribut)	<i>quel/lequel</i>	<i>quel</i> + verbe impersonnel
<i>quel</i> (épithète, non précédé d'une préposition)		<i>qui</i> prédicatif
<i>qui</i> (sujet)		<i>qui</i> complément prépositionnel
<i>qui, que, quoi</i> , objet direct ou attribut		

Ces trois classifications présentent plusieurs convergences, en particulier celles de Pohl et Söll. Nous nous contenterons de lister les points communs entre ces deux études, puisqu'elles partagent grossièrement la même étendue et la même vocation et que celle de Behnstedt contient, par ailleurs, de nombreux autres tableaux et classifications. Les constantes observables dans les classifications de Pohl et de Söll sont donc :

- L'isolement de chaque mot interrogatif qui n'est pas un adverbe
- L'isolement de *qui* sujet - *qui* précédé d'une préposition – *qui* objet direct ou attribut
- La distinction entre *que/quoi* précédés d'une préposition ou non

Les divergences que l'on rencontre dans les classifications de Pohl et Söll peuvent s'expliquer de diverses manières. Premièrement, on peut supposer que la nature des données détermine dans une certaine mesure la façon dont les auteurs les classent: l'établissement d'une catégorie donnée peut être jugée nécessaire si elle est largement représentée dans le corpus, mais elle peut être omise dans le cas contraire. Deuxièmement, il faut rappeler que le but des divisions présentées dans le tableau 3 est de classer les interrogations selon leur potentiel de transformation syntaxique, pour favoriser l'interchangeabilité des structures à l'intérieur d'une même catégorie. Or, l'acceptabilité grammaticale de l'interrogation partielle est extrêmement complexe et dépend d'un grand nombre de facteurs qu'il serait difficile de représenter à l'intérieur d'un même tableau. Conçues pour donner un compte-rendu général

de l'usage des structures interrogatives partielles, les études de Pohl et Söll doivent répondre à une double exigence, de pertinence statistique d'une part, de clarté et de concision de l'autre.

On remarque que dans ces conditions, Pohl et Söll ont jugé nécessaire de trier les différentes interrogations selon le type et/ou la fonction de leur mot interrogatif, mais pas selon la nature de leur sujet. Le fait que Pohl et Söll n'aient pas jugé pertinent de distinguer les sujets nominaux des sujets pronominaux dans la présentation finale de leurs résultats peut surprendre. Nous avons vu en effet que la nature du sujet pouvait modifier les exigences syntaxiques de l'interrogation partielle, au même titre que le type de mot interrogatif. Le choix de Pohl et Söll de ne pas distinguer les sujets nominaux des sujets pronominaux pourrait venir de la fréquence élevée des phénomènes de dislocation du sujet. Behnstedt et Söll constatent en effet que lorsqu'une question (totale ou partielle) présente un sujet nominal, il est très souvent disloqué, c'est-à-dire doublé d'un pronom sujet.³⁸ Comme le montrent les exemples suivants, tirés du corpus de Behnstedt (Al, 1975), la dislocation peut avoir lieu par anticipation (34) ou par reprise (35) :

(34) Comment est-ce qu'elles peuvent savoir, les autorités ?

(35) Votre femme, par exemple, elle approuve ?

Söll mentionne que sur les 15 interrogations totales à sujets nominaux de son corpus, 12 présentent un phénomène de dislocation (il ne donne malheureusement pas de chiffres pour les interrogations partielles). Quant à Behnstedt, il signale que la dislocation touche la « très grande majorité des questions à sujet nominal » (Al, 1975, 59). Il semble donc que les questions dépourvues de sujet pronominal soient rares (du moins dans la langue non soutenue). Dans ce contexte, on comprend que Pohl et Söll n'aient pas jugé nécessaire de classer les phrases interrogatives selon la nature de leur sujet.

L'examen des études quantitatives de Pohl, Söll et Behnstedt nous a permis d'approcher de plus près la diversité structurale des questions ainsi que les contraintes grammaticales qui pèsent sur leur formation, en particulier pour les questions partielles. Nous avons ainsi appréhendé la complexité grammaticale des questions et constaté que leur classification en vue d'une quantification présentait toute une série de problèmes, au niveau grammatical déjà. L'analyse des études de Pohl, Söll et Behnstedt nous a servi de préambule à l'analyse des études variationnistes qui seront présentées en 2.3. En effet, les études quantitatives sont chronologiquement antérieures aux études variationnistes et elles ne présupposent aucune différenciation sociale dans l'utilisation des différentes structures interrogatives. Elles se concentrent donc principalement sur les contraintes grammaticales qui gouvernent l'usage des structures interrogatives³⁹. Les études variationnistes, elles, partent du principe qu'il existe des explications sociales à l'usage variable des structures interrogatives. Cependant, elles doivent prendre en compte l'aspect grammatical de cette variation, tout autant que les études quantitatives traditionnelles. Les auteurs des études variationnistes dont nous rendrons compte se sont d'ailleurs inspirés de la méthodologie et des résultats de plusieurs études quantitatives, dont celles que nous venons de décrire.

³⁸ Ce phénomène n'est d'ailleurs pas spécifique aux énoncés interrogatifs.

³⁹ Behnstedt s'intéresse également à l'aspect social de l'utilisation des structures interrogatives, mais il ne se réclame pas du courant variationniste. Nous nous sommes concentrée sur ses analyses de la question au niveau grammatical.

2.3 Les études variationnistes

Nous retournons à présent au thème de la variable syntaxique, abordée au point 1, pour analyser la façon dont elle est utilisée dans le domaine de la diversité des questions en français. Le point 2.3 sera entièrement consacré au compte rendu de deux études variationnistes qui tentent d'expliquer la variation des structures interrogatives en fonction de divers facteurs linguistiques et extra-linguistiques. Ces deux études sont *Variability in spoken French : a sociolinguistic study of interrogation and negation* d'Aidan Coveney (2002, publié la première fois en 1996) ainsi que *Grammatical variation across space and time : the French interrogative system* de Martin Elsig (2009). Toutes les citations tirées de ces deux ouvrages sont des traductions que nous avons effectuées de l'anglais au français. Les exemples tirés des corpus de Coveney et d'Elsig seront reproduits tels qu'ils apparaissent dans leurs études.

2.3.1 Présentation

Les études variationnistes de Coveney et d'Elsig s'approchent des études quantitatives présentées au point 2.1 dans la mesure où elles répertorient toutes les structures interrogatives des questions totales et des questions partielles pour en déterminer les pourcentages d'utilisation. Cependant, les études de Coveney et d'Elsig se différencient des études quantitatives traditionnelles par le fait qu'elles s'inscrivent dans le cadre théorique variationniste élaboré par Labov. Or, nous l'avons vu, le concept de variable sociolinguistique est controversé en syntaxe, et son utilisation peut prendre différentes formes. Si le variationnisme offre un cadre théorique précis et bien défini pour l'analyse de la variation phonologique, il n'en va pas de même au niveau syntaxique, où il incombe au chercheur de régler les questions théoriques soulevées au point 1.

Nous le verrons, Coveney et Elsig n'abordent pas la question du sens et de l'équivalence des variantes de la même façon. Ils ont également chacun leur manière d'appliquer le concept de variable sociolinguistique à l'étude des questions. Tous deux partent pourtant d'un objectif commun, celui de comprendre et d'expliquer l'usage des structures interrogatives tel qu'il est fait au quotidien par les locuteurs du français. La base des études de Coveney et d'Elsig est constituée par l'existence d'un point de variation dans le système linguistique du français. Ainsi Coveney décrit-il son étude comme étant « orientée sur le système » plutôt que sur le locuteur (Coveney, 2002, 91). Il explique qu'une étude orientée sur le locuteur prendrait comme point de départ une fonction (par exemple la « demande d'information ») pour en analyser les différents moyens d'expression. Les approches variationnistes, elles, se basent sur la présence de plusieurs formes concurrentes mises à disposition par le système (ici les structures interrogatives) pour analyser les fréquences et les raisons de leur utilisation.

Avec ses multiples structures formelles, l'interrogation constitue un terrain d'analyse privilégié pour ceux qui cherchent à appliquer le concept de *variable sociolinguistique* au niveau syntaxique. En effet, on considère communément que les principales différences entre les formes de question en concurrence sont plutôt d'ordre stylistique ou contextuel que sémantique. Les structures interrogatives sont donc, à première vue du moins, de bonnes candidates pour une analyse variationniste. L'examen des études de Coveney et d'Elsig nous permettra d'apprécier les méthodes utilisées par les chercheurs variationnistes pour expliquer la variation de l'interrogation. Plutôt que sur les résultats des études, nous nous focaliserons principalement sur les différentes étapes et choix méthodologiques des deux auteurs. Nous observerons ainsi comment ils abordent les problèmes théoriques liés à l'extension de la

variable sociolinguistique à la syntaxe et comment ils répondent aux questions du débat présenté au point 1.2.

2.3.1.1 Déroulement général

Les grandes lignes méthodologiques communes aux deux études se présentent comme suit: dans un premier temps, les deux auteurs définissent les contextes qu'ils considèrent comme variables et les variantes qui peuvent alterner à l'intérieur de chaque contexte variable. Ensuite, ils distinguent parmi toutes les structures interrogatives de leur corpus celles qui peuvent être prises en compte comme variantes et celles qui doivent être exclues de l'analyse. Finalement, ils analysent quantitativement et statistiquement les pourcentages d'occurrence des variantes, en corrélation avec plusieurs facteurs, linguistiques et extralinguistiques.

A côté de ces similitudes méthodologiques, les études de Coveney et d'Elsig présentent de nombreuses divergences, à commencer par le système informatique qu'ils utilisent pour mesurer l'influence des différents facteurs sur la variation des questions. En effet, Coveney recourt à un simple logiciel d'analyse de données, tandis qu'Elsig utilise un logiciel d'analyse en règles variables (voir point 1.1.2). En plus de chercher à expliquer les facteurs linguistiques et extralinguistiques qui influencent la variation, Elsig s'intéresse aux règles grammaticales qui gouvernent la formation des questions. Ainsi, une partie importante de son étude (dont nous ne rendrons pas compte) est dédiée à la grammaire des phrases interrogatives, abordée selon une perspective générative : après avoir analysé les facteurs pouvant influencer l'usage des variantes, Elsig s'intéresse aux propriétés transformationnelles des questions et tente de tirer des implications théoriques des résultats de son analyse en règles variables. Nous verrons au point 2.3.4 que les conséquences pratiques qui découlent du recours à l'analyse en règles variables différencient les deux études sur plusieurs points. La seconde divergence méthodologique majeure entre les deux études concerne le type de facteurs dont Coveney et Elsig choisissent de mesurer l'effet sur la variation des questions. Alors que tous deux s'intéressent à l'influence de facteurs linguistiques, stylistiques et sociaux, seul Coveney considère l'influence de facteurs pragmatiques. De ce fait, l'étude de Coveney traite longuement des divers aspects fonctionnels des questions, tandis qu'Elsig ne s'y attarde guère.

2.3.1.2 Le corpus de Coveney

L'étude de Coveney est basée sur un corpus qu'il a recueilli lui-même, dans le département de la Somme, en Picardie. Pendant plusieurs années, Coveney a participé à des camps de vacances dans la Somme en tant qu'animateur. Au cours des années 80, il a recruté plusieurs personnes travaillant dans divers camps de vacances de France pour mener des entretiens enregistrés. Coveney avait préalablement logé quelques jours dans chaque camp d'où il a recruté ses informateurs, de sorte que ces derniers lui étaient déjà familiers. Les entretiens furent réalisés sous la forme d'une conversation informelle entre l'informateur et Coveney lui-même ; les sujets abordés pendant la conversation concernaient principalement les camps de vacances, une réalité bien connue de Coveney et de ses informateurs. Au total, Coveney a interviewé 30 personnes âgées de 17 à 37 ans (à l'exception de trois femmes quinquagénaires), toutes originaires du département de la Somme. Cette région abrite encore des locuteurs du dialecte picard, mais Coveney explique que la majorité de ses informateurs ne le parlent pas, bien qu'ils soient souvent capables de le comprendre. Selon Coveney, les locuteurs du dialecte picard sont rares et souvent âgés. Il n'a rencontré aucune production

picarde pendant la durée de son enquête mais il précise qu'on ne peut exclure une légère influence de ce dialecte sur l'usage des interrogations de ses informateurs.

2.3.1.3 *Le corpus d'Elsig*

Les données d'Elsig ne concernent que le français québécois ; elles se répartissent sur trois corpus, tirés de trois périodes différentes. En effet, en plus de vouloir établir la grammaire de l'interrogation du français populaire, Elsig souhaite en tracer l'évolution, du XV^{ème} siècle à nos jours. Le premier de ses corpus est un mélange de textes littéraires écrits entre le XV^{ème} et le XVII^{ème} siècle. Il s'agit principalement de pièces de théâtre, dont le style s'approche, selon Elsig, de la langue orale de l'époque. Le deuxième corpus se compose de contes, légendes et histoires personnelles enregistrés au cours du XIX^{ème} siècle, lors de rencontres informelles. Enfin, le corpus du XX^{ème} siècle (années 80) est constitué d'enregistrements réalisés sous la forme d'entretiens (il s'agit d'une partie du corpus connu sous le nom de *Ottawa-Hull French corpus*). Les 48 informateurs, d'âge variable, proviennent tous de la ville de Gatineau, au Québec. Elsig précise que les entretiens ont été réalisés selon des méthodes qui favorisent l'émergence d'un langage non surveillé : ainsi qualifie-t-il le français des deux corpus enregistrés d'« hautement vernaculaire » (Elsig, 2009, 39). Etant donné que nous ne souhaitons pas aborder la variation des questions de manière diachronique dans ce travail, nous ne nous intéresserons qu'à la partie de l'étude d'Elsig qui concerne le français du XX^{ème} siècle. Ainsi, lorsque nous mentionnons « le corpus » d'Elsig, il s'agit toujours et exclusivement de celui du XX^{ème} siècle.

2.3.2 Etablissement des catégories de variables et de variantes

Avant toute tentative de quantification et d'explication de la variation, les auteurs d'études variationnistes doivent définir les structures qui varient, comme c'est le cas pour les études quantitatives (cf. point 2.1). Selon la terminologie introduite par Labov et présentée au point 1, les différentes structures interrogatives en concurrence sont des *variantes* alors que le contexte linguistique dans lequel elles peuvent apparaître est une *variable* (on peut aussi, comme Elsig, parler de *contexte variable*). Coveney et Elsig travaillent sur la base de deux variables : les questions totales et les questions partielles. Dans les deux cas, ils ne prennent en compte que les questions directes. Comme eux, nous recourons parfois au mot *tokens* pour désigner les occurrences de structures interrogatives relevées dans leurs corpus.

A l'intérieur de la variable « question totale », Elsig distingue quatre variantes différentes, alors que Coveney n'en a que trois. Sans surprise, les trois variantes communes aux deux études sont l'inversion (VS), *est-ce que* (ESV) et l'ordre sujet-verbe (SV). La variante VS ne comprend que des inversions simples (avec pronom sujet clitique), car Coveney et Elsig n'ont relevé aucune inversion complexe dans leurs corpus. La quatrième variante rencontrée par Elsig mais absente du corpus de Coveney est celle qui fait apparaître la particule interrogative *-tu* (SV-*tu*)⁴⁰. L'utilisation de la structure interrogative en *-tu*, si elle a presque disparu en France, reste bien implantée au Québec.

⁴⁰ Nous utilisons ici la même graphie qu'Elsig : *tu*.

Tableau 4 - Les variantes de la question totale

Coveney	Elsig	
SV	SV	il part ?
ESV	ESV	est-ce qu'il part ?
VScl	VScl	part-il ?
	SV-tu	il part-tu ?

S = sujet/ Scl = pronom sujet clitique / V = verbe / E = *est-ce que*

Nous remarquons que ni Coveney ni Elsig ne crée de variante particulière pour les questions qui font apparaître des particules telles que *hein*, *non*, *n'est-ce pas*, etc.⁴¹ Comme nous le verrons plus bas, ces questions ont souvent été exclues de leurs analyses.

Au niveau des questions partielles, seul Coveney a décidé de compter les différentes variantes en vue d'en expliquer l'utilisation par les locuteurs. Il distingue six variantes différentes, représentées dans le tableau 5.

Tableau 5 - Les variantes de la question partielle chez Coveney

SVK	il part où ?
KSV	où il part ?
KESV	où est-ce qu'il part ?
KVScl	où part-il ?
KVSsn	où part Jean ?
K=S V	qui part ?

S = sujet / Scl = pronom sujet clitique / Ssn = sujet syntagme nominal / V = verbe / E = *est-ce que* / K = mot interrogatif /

On constate que Coveney distingue clairement les cas d'inversion du sujet pronominal (KVScl) et les cas d'inversion du sujet nominal (KVSsn). Il ne s'agit pas ici d'une distinction entre inversions « simples » et inversions « complexes », ces dernières étant absentes du corpus de Coveney. Si Coveney sépare les cas d'inversions simples selon la nature du sujet, c'est moins pour des questions d'interchangeabilité syntaxique (dont nous avons parlé au point 2.1) que pour des questions d'équivalence socio-stylistique. En effet, il a constaté que les variantes KVScl et KVSsn étaient liées à différentes valeurs socio-stylistiques. Puisqu'un des objectifs de l'étude de Coveney est de saisir les différences sociales et stylistiques dans l'emploi des structures interrogatives, il se doit de séparer, au départ, des structures qui présentent des divergences à ce niveau, même si elles sont syntaxiquement très proches. C'est la même raison qui a conduit Coveney à créer une variante spécifique pour les cas où S=K (la fonction sujet peut être remplie par *qui*, *lequel*, *combien de* + nom, etc.). En effet, Coveney explique qu'au niveau grammatical, K=S V pourrait être assimilé à une structure KSV où la position du sujet serait vide. Mais il constate que la variante K=S V est communément considérée comme standard (voir exemple (36)), au contraire de KSV, souvent perçue comme une structure non-standard (exemple (37)).

⁴¹ Par exemple: *ça se trouve dans le Vieux-Hull, hein ?* (Elsig, 2009, 44) / *c'est moi qui qu'ai la clé non ?* (Coveney, 2002, 117)

(36) Qui y a été ?

(37) Comment elles s'appellent ? (Coveney, 2002, 93-94)

Il faut préciser encore que Coveney n'inclut pas les structures en *qu'est-ce qui* + V dans la catégorie K=S V. Même si les questions en *qu'est-ce qui* + verbe comportent aussi un mot interrogatif qui fait office de sujet, Coveney les intègre à la variante KESV, parce qu'il considère qu'elles partagent la même valeur socio-stylistique que les autres structures KESV.

La variante KESV ne comprend que des occurrences d'*est-ce que* ou d'*est-ce*. Toutes les autres structures dites « rivaes » ou « dérivées » d'*est-ce que* sont absentes du corpus de Coveney. Cette absence est surprenante, si on la compare avec la richesse et le grand nombre d'occurrences des structures rivaes d'*est-ce que* dans le corpus d'Elsig (voir tableau 6 ci-dessous) ou avec la présence systématique de ce genre de structures dans les études quantitatives que nous avons analysées en 2.1. Coveney ne s'attarde pas sur le fait que son corpus présente si peu de variété en matière de structures dérivées d'*est-ce que*: il se contente de mentionner l'existence de plusieurs particules concurrentes d'*est-ce que* dans un point qui traite des structures n'apparaissant pas dans son corpus. Quant aux occurrences d'*est-ce* dans les questions partielles de son corpus, Coveney les considère comme des réductions phonologiques d'*est-ce que*. Estimant que l'usage d'*est-ce* à la place d'*est-ce que* n'entraîne aucune modification sémantique, socio-stylistique ou fonctionnelle, il décide d'inclure les interrogations qui contiennent cette particule dans la variante KESV.

On remarque que pour Coveney, la portée socio-stylistique des différentes structures interrogatives est un critère décisif dans l'établissement des variantes. Ainsi, si les structures KSV / K=S V d'une part, KVScl / KVSsn d'autre part sont considérées comme des variantes distinctes, c'est parce qu'elles sont connues pour ne pas partager les mêmes valeurs socio-stylistiques. Avant d'établir les variantes dont il allait étudier l'usage, Coveney a consulté de nombreuses grammaires ainsi que des enquêtes menées sur l'opinion de locuteurs natifs au sujet des différentes formes de l'interrogation. Il a ainsi pu déterminer, de manière générale, les valeurs sociales et stylistiques communément attribuées aux différentes structures interrogatives.

La question de l'interchangeabilité syntaxique des variantes n'entre pas en compte dans l'établissement des catégories de variantes. Ce n'est qu'au moment de coder et de quantifier les occurrences des variantes que cette question est prise en compte. Ainsi, dans la suite de l'analyse, les phrases au potentiel de transformation syntaxique réduit, comme celles qui comprennent un pronom interrogatif *que/quoi*, seront traitées différemment des phrases qui peuvent s'adapter à toutes les structures interrogatives variantes.

Elsig a renoncé à traiter les différentes structures de la variable « question partielle » en vue de déterminer les facteurs qui influencent leur usage. Il explique n'avoir pas obtenu assez d'occurrences de chaque structure pour mener une analyse en règles variables de la question partielle. Il a néanmoins rencontré trois structures dont l'usage est répandu dans son corpus : KESV, KSV et SVK. L'usage de l'inversion (du pronom clitique ou du groupe nominal) est extrêmement rare dans le corpus d'Elsig, qui la considère comme un phénomène « marginal, presque négligeable » (2009, 148). En fait, l'usage de la structure KESV domine nettement dans le corpus d'Elsig, puisqu'elle apparaît pour $\frac{3}{4}$ des questions partielles. A cause de la prévalence marquée de cette structure sur les autres, Elsig n'a pas jugé opportun de chercher à expliquer l'usage des variantes de la question partielle selon des facteurs

linguistiques et extralinguistiques.⁴² Il faut préciser qu'Elsig a compté comme des occurrences de la structure KESV les questions qui faisaient apparaître, entre le mot interrogatif et le verbe, des particules du type *c'est que*, *que c'est que*, *que*, etc., qu'il considère comme dérivées d'*est-ce que*. Au total, Elsig ne dénombre pas moins de 12 particules composées d'un ou de plusieurs des trois morphèmes *est*, *ce*, et *que*. Le tableau 6 contient la liste de toutes les particules rencontrées par Elsig dans les structures KESV, (elles apparaissent donc toutes après le mot interrogatif, quel qu'il soit). Nous avons classé ces particules par ordre de fréquence, le nombre entre parenthèses indiquant le total d'occurrences dans le corpus. Lorsque des exemples du corpus d'Elsig étaient disponibles, nous les avons ajoutés dans la colonne de droite.

Tableau 6 - Les dérivés d'*est-ce que* dans le corpus d'Elsig

<i>ce que</i>	(158)	pourquoi <i>ce que</i> la police, elle a pas sorti (...) ?
<i>est-ce que</i>	(151)	qu' <i>est ce qu'</i> elle va faire avec six-cents piasses par mois (...) ?
<i>est-ce</i>	(125)	
<i>c'est que</i>	(117)	
<i>ce</i>	(75)	hey tabarnouche, d'où <i>ce</i> tu viens toi (...) ?
<i>que</i>	(72)	où <i>qu'</i> il est?
<i>c'est</i>	(37)	
<i>ce que c'est que</i>	(16)	
<i>est-ce que c'est</i>	(15)	
<i>ce que c'est</i>	(5)	
<i>c'est que c'est que</i>	(6)	il dit qui <i>c'est que c'est qui</i> frappe sur la maison comme ça (...) ?
<i>ce que que</i>	(1)	

2.3.3 Contextes variables et contextes catégoriques

Seule une partie de toutes les questions totales et partielles des corpus de Coveney et d'Elsig sont prises en compte comme variantes relevant de l'une ou l'autre variable. En effet, une étape fondamentale de toute étude variationniste est la distinction entre les tokens qui peuvent être considérés comme des variantes et ceux qui sont jugés catégoriques. Ces derniers sont exclus de l'analyse statistique et seuls les tokens variant sont codés en vue de déterminer les différents facteurs qui influencent leur usage.

2.3.3.1 Critères de sélection des variantes

Avant de différencier les contextes variables et les contextes catégoriques, Coveney et Elsig déterminent d'une manière plus générale les questions susceptibles d'intégrer leurs analyses : il doit d'agir de questions directes qui comprennent un verbe, et dont la structure est clairement définie. Ainsi, les questions incomplètes, dont il est difficile d'établir le

⁴² Au niveau des questions totales, Coveney a affaire à un phénomène similaire puisque l'usage de SV surpasse de loin l'usage des autres variantes. Pourtant, à aucun moment il ne mentionne cette importante prévalence comme un obstacle à l'explication de la variation selon des facteurs intra et extra linguistiques.

commencement et/ou la fin, sont d'emblées écartées de l'analyse : il s'agit souvent de faux départs (lorsque le locuteur ne termine pas une phrase où la répète en la modifiant) ou d'anacoluthes (lorsque le locuteur perd le fil syntaxique de sa phrase). Une fois ce premier tri effectué, il est possible de déterminer la nature (variable ou catégorique) de tous les tokens restants.

Les études de Coveney et d'Elsig comportent une différence méthodologique de taille dans la distinction entre questions variables et questions catégoriques. Nous le verrons plus bas, cette différence est due notamment à une exigence spéciale de l'analyse en règles variables utilisée par Elsig pour traiter ses données. Pour chaque phrase interrogative rencontrée dans son corpus, Elsig analyse si toutes les variantes peuvent lui être substituées sans que cela n'entraîne d'énoncé grammaticalement inacceptable et sans que cela ne change le sens de l'énoncé. Il suffit qu'une seule des structures variantes ne puisse être utilisée pour que le token soit considéré comme catégorique, et donc exclu de l'analyse. Pour Coveney en revanche, il n'est pas nécessaire que toutes les structures variantes puissent être substituées à une certaine phrase interrogative pour que celle-ci soit prise en compte dans l'analyse. Si une seule structure lui est substituable, Coveney prend en compte le token et le classe comme *semi-catégorique*, comme c'est le cas de la question en (38) :

(38) ...l'enfant quand il sort de la colonie bon on a eu beau faire / euh le maximum / bon il en a retenu quelque chose./ il en aura un souvenir bien sûr./ *mais est-ce qu'on l'aura vraiment aidé pour qu'il reparte euh sur d'autres bases ?* euh -- pt -- c'est [BRUIT] c'est pas évident du tout./... (Coveney, 2002, 200)

Coveney considère que l'inversion pourrait se substituer à la structure ESV de la question en (38), alors que l'usage d'une structure en SV entraînerait une question non équivalente. Dans l'étude d'Elsig, où le concept de token semi-catégorique n'existe pas, la question en (38) aurait été classée comme catégorique, et donc exclue de l'analyse.

Dans toute étude variationniste, les deux principaux paramètres qui font que deux (ou plusieurs) tokens peuvent être considérés comme variantes d'une variable sont leur acceptabilité grammaticale et leur équivalence (sémantique et/ou fonctionnelle). Les études quantitatives traditionnelles ne procèdent généralement pas à l'exclusion des questions dont la structure syntaxique ne peut pas être remplacée par une autre pour des raisons d'équivalence ou d'acceptabilité. Les trois études dont nous avons rendu compte au point 2.2 se souciaient, dans une certaine mesure, de l'interchangeabilité grammaticale des différentes structures, mais elles n'abordaient que peu (ou pas) les questions d'équivalence au niveau sémantique ou pragmatique. Coveney souligne l'importance de l'exclusion des questions catégoriques et l'incidence de cette étape sur les résultats des fréquences relatives. Pour mesurer l'influence de l'exclusion des questions catégoriques, il a commencé par calculer les fréquences relatives de ses variantes en incluant les questions catégoriques, comme le font les études quantitatives traditionnelles. Lorsqu'il a ensuite fait le même calcul en excluant les questions catégoriques, la fréquence d'utilisation de la variante SV est passée de 79.4% à 70%, en faveur de la variante ESV dont la fréquence est montée de 20.6% à 29.4% (ceci dans le contexte variable de la question totale).

Ainsi, dans les études de Coveney et d'Elsig, deux (ou plusieurs) structures interrogatives ne peuvent pas être prises en compte comme variantes si elles ne sont pas grammaticalement interchangeables (le potentiel de transformation syntaxique des différentes structures interrogatives a été abordé au point 2.2). De même, deux structures interrogatives ne peuvent pas être considérées comme variantes d'une variable si elles ne confèrent pas le même sens à l'énoncé dans lequel elles apparaissent. En effet, nous avons expliqué au point 1

que toutes les variantes analysées dans un cadre théorique variationniste devaient « vouloir dire la même chose ». Nous avons également montré les diverses interprétations qui pouvaient découler de cette exigence théorique, ainsi que les objections auxquelles elle a pu donner lieu, certaines personnes considérant que l'équivalence de sens ne pouvait être garantie qu'au niveau des variables phonologiques. Les deux variables analysées par Coveney et Elsig sont des cas d'alternances entre plusieurs structures syntaxiques ; selon les typologies de variables données au point 1.2, elles font partie de la catégorie de variables la plus hautement exposée aux problèmes d'équivalence de sens. Coveney et Elsig doivent aborder ces questions d'équivalence au moment de distinguer les questions qu'ils vont considérer comme variantes d'une variable de celles qu'ils vont exclure de l'analyse. Il s'agit d'une étape interprétative délicate que Coveney et Elsig abordent chacun à sa manière. S'ils considèrent tous deux que l'équivalence des variantes concerne aussi bien leur sens descriptif que fonctionnel, l'application de cette exigence théorique diffère d'une étude à l'autre.

Coveney traite longuement et en détail du sujet de l'équivalence sémantique et fonctionnelle des variantes. Il consacre également une partie de son étude à réfuter les arguments de ceux qui considèrent la question de l'équivalence comme une barrière théorique à l'application de la notion de variable sociolinguistique en syntaxe.

Coveney ne parle de *variable sociolinguistique* que lorsqu'il est démontré que l'usage des variantes est corrélé à certaines caractéristiques sociales des locuteurs. Si ce n'est pas le cas, il parle d'*alternances* (*alternations* en anglais). Il considère que la procédure variationniste peut servir à l'étude des deux phénomènes, selon des modalités bien distinctes. En effet, pour Coveney, la variable sociolinguistique est un *concept* étroitement lié à une théorie sociolinguistique de la langue, dans la mesure où elle aide à « délimiter et différencier des communautés linguistiques » (Coveney, 2002, 248). En revanche, lorsque la méthodologie variationniste est utilisée pour l'étude d'une alternance, elle devient, selon Coveney, un simple *outil* analytique qui ne présuppose pas d'orientation théorique particulière. Coveney définit la variable (sociolinguistique) grammaticale⁴³ sur la base de sept critères que nous résumons ci-dessous (Coveney, 2002, 52-53) ; dans le cas des alternances, il considère que seuls les quatre premiers critères doivent être remplis.

1. Les variantes d'une variable (ou d'une alternance) doivent pouvoir se substituer sans que cela ne change le contenu propositionnel de l'énoncé dans lequel elles apparaissent (Coveney précise qu'on peut aussi parler de sens descriptif, conceptuel, cognitif ou dénotatif). En revanche, Coveney n'exige pas d'équivalence de sens expressif, tel qu'il est décrit par Lyons (voir point 1.2.2) : les variantes et les énoncés dans lesquels elles apparaissent peuvent donc différer par leur degré d'affectivité, d'emphasis ou de politesse.
2. Les variantes d'une variable doivent pouvoir se substituer sans que cela ne change la fonction communicative de l'énoncé dans lequel elles apparaissent (la notion de « fonction communicative » sera abordée plus bas). Selon Coveney, ce critère est inévitablement lié au jugement intuitif du chercheur. En effet, pour chaque occurrence d'une variante, le chercheur, aidé ou non, doit déterminer si l'usage des autres variantes est pragmatiquement équivalent. Si une variante ne confère pas le même sens pragmatique, elle doit être exclue de l'analyse et le contexte où elle apparaît doit être classé comme *semi-variable*. Pour Coveney, la meilleure

⁴³ Coveney préfère parler de variation « grammaticale » plutôt que de variation « syntaxique », en raison des restrictions parfois liées au terme « syntaxique ». Il explique en effet que le terme « syntaxe » peut être utilisé pour ne désigner que les relations existantes entre les différents éléments de la phrase. Comme il souhaite que la variable puisse traiter de phénomènes qui vont au delà de la phrase comme les pronoms, il choisit de la qualifier de « grammaticale » (Coveney, 2002, 29).

façon d'opérer cette sélection serait de recourir aux jugements d'un grand nombre de locuteurs natifs, idéalement les participants à l'enregistrement du corpus. Pour des raisons pratiques, il a procédé lui-même aux jugements d'équivalence, tout en demandant de l'aide à deux locuteurs natifs pour les énoncés qui lui posaient problème.

3. Les variantes d'une variable doivent pouvoir se substituer de manière à ce que l'énoncé dans lequel elles apparaissent conserve les mêmes segments lexicaux. Les variables qui se présentent sous la forme de *présence/absence*, doivent être *susceptibles* de contenir les mêmes segments lexicaux. Cette condition est présentée comme une conséquence de l'équivalence de sens descriptif exigée par le critère 1.
4. Les variantes doivent être formellement liées. La portée exacte de ce critère dépend de chaque variable et n'est pas applicable aux cas de *présence/absence*. Il se traduit souvent par l'appartenance des variantes à une même catégorie syntaxique (pronoms, phrases subordonnées, etc.).
5. Les variantes doivent être différenciées socialement à l'intérieur d'une communauté linguistique donnée (une différenciation stylistique n'est pas obligatoire, même si, dans la pratique, elle coexiste souvent avec la différenciation sociale).
6. Les variantes doivent apparaître dans le style vernaculaire d'au moins quelques locuteurs (on exclut ainsi des formes syntaxiques dont l'élaboration est plus caractéristique du langage écrit qu'oral).
7. L'alternance étudiée est plus susceptible de présenter une différenciation sociale si elle est spécifique à une langue que si elle est universelle. Par exemple, la présence de différentes structures interrogatives est un phénomène plutôt spécifique au français alors que l'alternance entre structures actives et structures passives est plutôt universelle.

Au début de son étude, Coveney ne se prononce pas sur la nature de la variation des structures interrogatives de la question totale et de la question partielle : il dit devoir s'appuyer sur ses résultats avant de pouvoir décider si les questions totales et partielles constituent deux variables ou pas. En fin de compte, il considère qu'elles répondent bien aux critères 1 à 7 et qu'on peut donc les considérer comme des variables sociolinguistiques.

Elsig, quant à lui, ne s'attarde pas sur la notion d'équivalence. Nous savons que sur les 1460 questions totales de son corpus, 828 ont été jugées catégoriques et donc exclues de l'analyse, tandis que seules 632 questions totales ont été retenues. Etant donné qu'Elsig ne travaille que sur la variable « question totale », il n'a pas affaire aux questions d'interchangeabilité grammaticale, spécifiques aux questions partielles. Cela signifie qu'il a écarté 739 questions pour des raisons d'équivalence. En observant les questions qu'il a exclues de l'analyse (voir plus bas) nous en apprendrons un peu plus sur sa manière de traiter l'équivalence des variantes.

2.3.3.2 Questions totales catégoriques

Alors qu'Elsig a gardé 632 questions totales sur les 1460 qu'il avait au départ, Coveney en a gardé 109 sur un total de 180. Etant donné qu'Elsig a renoncé à coder les structures interrogatives des questions partielles, nous ne pouvons comparer les structures interrogatives exclues par les deux auteurs qu'au niveau des questions totales. Nous aborderons plus bas les questions partielles exclues de l'analyse de Coveney.

Premièrement, les deux auteurs s'accordent sur l'exclusion de (pratiquement) toutes les questions qui présentent un *tag* interrogatif⁴⁴. Les questions *tags*, telles que les définissent Coveney et Elsig, ne se composent que d'un verbe, conjugué au présent avec un pronom sujet de deuxième personne, comme l'illustre la partie marquée en (39) :

- (39) - Ça a faite bien de quoi aux paroissiens, *tu sais* ?
- Ouais, je comprends (Elsig, 2009, 45)

Les verbes utilisés dans les questions *tags* sont souvent *voir*, *savoir* et *comprendre*. Selon Elsig, la fonction de ses questions « n'est pas de demander quelque chose à l'interlocuteur, mais plutôt d'entretenir la conversation » et d'obtenir une courte réaction de l'interlocuteur (Elsig, 2009, 45). Coveney explique que la fréquence d'utilisation des questions *tags*, très élevée chez certains locuteurs, pourrait biaiser les résultats. En effet, de telles questions apparaissent souvent sous la forme SV : les compter dans l'analyse augmenterait passablement la prégnance de la variante SV. Elsig exclut également toutes les questions qui se terminent par une particule interrogative du type *hein*, *non*, *hé*. Il explique qu'il est difficile de déterminer si l'interrogation porte sur l'intégralité de la phrase dans laquelle ces particules apparaissent ou seulement sur la particule. Coveney rend compte de la même difficulté, mais il inclut néanmoins les structures à particule *hein*, *non*, etc. dans l'analyse, lorsque leur orientation est clairement interrogative. C'est le cas du segment marqué en (40), qui présente une intonation montante sur le *vous* et que Coveney considère comme une question variable. Selon lui, la particule *non* ? signale ici « l'attente d'une réponse négative, peut-être en raison de l'expression faciale de A » (Coveney, 2002, 117).

- (40) B : ... et on a mis en système un carnet de chèques – les chéquiers *tu les as vus* ? non ?
A : ah non – pas encore (Coveney, 2002, 117)

Coveney et Elsig excluent également toutes les « questions échos » de leurs corpus. Les questions échos font partie de la catégorie que Borillo appelle « questions-répliques » (voir point 2.1.3) : elles reprennent mot pour mot (avec des modifications déictiques s'il le faut) une information ou une question qui précède. Coveney et Elsig n'ont pas compté les occurrences des questions échos parce qu'ils ont remarqué que leur structure était entièrement conditionnée par la structure de l'énoncé qu'elles reprennent, comme le montre l'exemple (41) :

- (41) Oui je le sais. *Tu le sais toi* ? Moi je m'en souviens pas (Elsig, 2009, 44)

Coveney a observé un phénomène de conditionnement similaire au niveau des questions alternatives. Ainsi, la deuxième partie de l'interrogation alternative (séparée de la première par la conjonction *ou*) présente toujours une structure identique à la première. Dès lors, Coveney renonce à compter comme structure variante la deuxième partie des questions alternatives.

Elsig a rencontré dans son corpus 51 questions formulées à l'aide du segment *c'est-tu*, comme c'est le cas en (42) :

- (42) Il y a – c'est-tu mieux de vivre en concubinage, que de- de se marier puis se séparer s- un an après ? (Elsig, 2009, 44)

⁴⁴ Ces questions correspondent aux « structures interrogatives postposées » que Borillo (1978) classe dans les « phrases avec appendice » (voir point 2.1.2.2).

Il a renoncé à introduire ces questions dans l'analyse car il considère que la formule *c'est-tu* est lexicalisée et qu'elle ne constitue donc pas une utilisation productive de la variante SV-*tu*. En outre, Elsig a écarté toutes les questions négatives car il a remarqué qu'elles apparaissent presque systématiquement avec la variante SV : sur les 88 questions négatives de son corpus, 80 présentent une structure SV et 8 une structure SV-*tu*. Elsig considère donc que les questions négatives constituent un « contexte pratiquement invariable » (2009, 43). Pour Coveney, nous le verrons plus bas, seules les questions négatives orientées négativement (c'est-à-dire, qui favorisent une réponse en *non*) sont considérées comme étant (presque) catégoriquement liées à la variante SV.

Le reste des questions totales exclues par Coveney et Elsig présentent toutes des structures SV. Il s'agit de questions auxquelles on ne peut substituer aucune autre variante sans que cela n'altère leur sens pragmatique. Elsig donne peu de renseignements sur la nature des fonctions communicatives de ces questions jugées catégoriques : il se contente d'expliquer que leur statut interrogatif n'est pas clair, et de donner l'exemple d'une question orientée affirmative (43) qui appelle une réponse positive. Elsig considère que la substitution de la question en (43) par une autre structure interrogative entraînerait un énoncé pragmatiquement inapproprié :

(43) Monsieur Côté, *je pense c'était ça le directeur ?* (Elsig, 2009, 44)

Coveney se montre plus exhaustif en distinguant trois types de fonctions communicatives qui contraignent l'usage de la variante SV. Ainsi, toutes les questions qui correspondent à l'une des trois configurations énumérées plus bas ont été classées comme catégoriques chez Coveney:

1. La question est une demande d'action qui présente le sujet *tu* ou *vous* accompagné d'un verbe non-modal.

(44) ...y a une petite fille qui me dit '*...tu me mets de de la couleur sur les noëils ?*'... (Coveney, 2002, 178)

2. La question est une suggestion et le verbe n'est pas modal.

(45) ...vous passez par ici ? (question accompagnée d'un geste montrant une porte au lieu d'une autre) (Coveney, 2002, 180)

3. La question sert à souligner une information qui vient d'être donnée, et elle n'appelle pas de réponse de l'interlocuteur ; elle correspond à ce que Coveney appelle *post-announcement* (voir aussi point 2.3.4.3).

(46) ...i te donnent pas ton permis parce que t as pas donné d'argent à la boîte. // *t vois ce que j veux dire ?*/ alors c'est pour ça //... (Coveney, 2002, 180)

Enfin, Coveney signale deux types de questions orientées qui contraignent presque systématiquement l'usage de la variante SV. Il s'agit des questions affirmatives orientées positivement (c'est-à-dire qui appellent une réponse en *oui*, comme c'est le cas en (43)) et des questions négatives orientées négativement, comme en (47).

(47) A : est-ce que vous pourriez me décrire un petit peu une journée moyenne – pour vous – pendant une colonie de vacances ?/

B : c'est-à-dire le prix de de le revient d'une journée ? / non * *c'est pas ça que vous voulez * dire ?*/

A : * euh – fin * - non – ce que vous faites pendant la journée - (Coveney, 2002, 182)

On voit que les questions qui apparaissent en (43) et (47) ne pourraient être substitués par les variantes ESV ou VS sans que cela ne change leur valeur pragmatique. Coveney a toutefois rencontré quelques rare cas où des questions orientées de cette manière pouvaient faire apparaître d'autres variantes et donc être incluses dans l'analyse.

2.3.3.3 Questions partielles catégoriques

Le pourcentage de questions partielles exclues est bien moindre que celui des questions totales : sur 111 questions partielles, Coveney n'en a exclu que 11. Comme pour les questions totales, Coveney a exclu toutes les questions échos au niveau des questions partielles. Le reste des questions partielles exclues ont été jugées catégoriques sur la base de critères linguistiques. Deux types de structures interrogatives linguistiquement catégoriques sont apparues dans le corpus de Coveney. Le premier concerne les questions qui présentent une structure KESV et dont $S = K = que$.

(48) Qu'est-ce qui te fait peur ? (Coveney, 2002, 186)

Le second type de structure interrogative linguistiquement invariable ne se présente qu'une fois dans le corpus :

(49) A : moi la dernière colo qu j'ai visité c'était une colo maternelle /... le personnel d service occupait tout l troisième étage du château./

B : ah oui alors là vraiment *c'était une colonie pour quoi alors ça ?* (Coveney, 2002, 186)

Selon Coveney, la question en (49) ne peut être substituée par une structure autre que SVK en raison de la confusion qui pourrait s'installer entre *pourquoi* et *pour quoi*. Il ajoute que plusieurs autres configurations syntactico-lexicales pourraient engendrer des questions catégoriques mais qu'elles n'apparaissent pas dans son corpus.

Si Coveney a répertorié plusieurs fonctions communicatives contraignant l'usage de certaines variantes au niveau des questions totales, il n'a pas observé ce phénomène au niveau des questions partielles. Ces dernières posent toutefois un autre type de problème, lié à l'identification de la nature directe ou indirecte de l'interrogation. En effet, Coveney signale qu'il est parfois difficile de distinguer si la structure SVK représente une interrogation directe ou indirecte. C'est le cas lorsque le lien syntaxique entre une structure SVK et le segment linguistique qui la précède n'est pas clair, comme en (50) et (51) :

(50) ...et ça c'est vrai que c'est un problème d'éducation euh/ c'est euh *comment on fait passer des tas de choses – auprès des gamins* / y a pas que le football à faire passer. /... (Coveney, 2002, 115)

(51) ...on avait discuté avec les ados parce que (...certains) Rosemary 'mais / euh a *sur quoi ça repose ?* le fait qu vous faites – fin si un très gros repas l matin midi presque rien puis l soir (bon)/... (Coveney, 2002, 116)

La structure SVK qui apparaît en (50) a été interprétée comme une interrogation indirecte, et donc exclue de l'analyse. En revanche, la structure SVK de (51) a été comptée comme une interrogation directe sur la base de son intonation (montante sur le mot *repose*) et du sens de l'énoncé (le locuteur rapporte une question posée par *les ados*).

La syntaxe floue de certains énoncés peut également poser problème au moment d'identifier les structures SVK des questions partielles. En effet, il est parfois difficile de savoir si le mot interrogatif est syntaxiquement lié au sujet et au verbe qui le précèdent ou s'il

en est indépendant. Ici encore, les caractéristiques prosodiques de l'énoncé aident à en éclaircir le statut syntaxique. Ainsi, le segment marqué en (52) n'a-t-il pas été interprété comme une occurrence de la variante SVK à cause de la pause (remplie par *euh*) qui sépare le verbe du mot interrogatif. Le contexte marqué en (53), n'a pas non plus été considéré comme variant, en raison d'une intonation descendante sur *enfants* (et marquant ainsi une coupure syntaxique) :

(52) ...à l'Ecole des Beaux Arts oui. / *j'en ai fait euh combien ?* quatre euh quatre cinq ans... (Coveney, 2002, 116)

(53) ...c'est une question de moyens / la preuve c'est que les gens – à faible revenu *ils ont beaucoup d'enfants. pourquoi ?* parce qu'ils sont aidés et assistés. / (Coveney, 2002, 116)

2.3.4 Facteurs variables influençant la variation

Après avoir exclu les questions de leurs corpus considérées comme catégoriques, Coveney et Elsig observent toutes les questions restantes pour tenter d'en expliquer la variation. Pour ce faire, ils considèrent l'influence de plusieurs types de facteurs : linguistiques, socio-stylistiques et, pour Coveney uniquement, pragmatiques. Pour mesurer l'influence et le poids de ces facteurs sur l'utilisation des variantes, Coveney et Elsig doivent procéder au codage de toutes les questions non catégoriques rencontrées dans leur corpus. Cela signifie que chaque token retenu doit être introduit dans le système informatique selon un système de codes. Les codes contiennent toutes les informations pertinentes sur le token, de manière à ce qu'on puisse analyser les facteurs variables susceptibles d'agir sur le choix de sa structure syntaxique.

A ce stade, les questions retenues par Coveney n'ont pas les mêmes propriétés transformationnelles que celles retenues par Elsig. En effet, nous l'avons vu, Elsig n'a conservé que les questions auxquelles on pouvait remplacer toutes les structures interrogatives variantes sans que cela n'altère leur acceptabilité ou leur équivalence. En revanche, Coveney a gardé toutes les questions qui pouvaient être remplacées par au moins une autre structure variante tout en conservant leur acceptabilité et leur équivalence. De ce fait, au moment de définir les facteurs qui influencent la variation, Elsig n'a affaire qu'à des interrogations dont le potentiel de variabilité est total alors que Coveney doit aussi traiter des interrogations semi-variables, c'est-à-dire dont une ou plusieurs alternatives sont bloquées.

2.3.4.1 Traitement informatique/statistique

Pour le traitement informatique de ses données, Coveney utilise un simple logiciel de traitement de données (à savoir, « Context Editor »). Il se contente de coder toutes les questions retenues et de les introduire dans deux fichiers (un pour les questions totales, l'autre pour les questions partielles). Chaque question se voit donc transformée en une ligne de codes qui comprend : la fonction communicative de la question, la structure variante utilisée ainsi que celles qui auraient pu l'être, la forme verbale utilisée (sujet+verbe+temps) et finalement, la personne à qui était adressée la question. Pour les questions partielles, le mot interrogatif apparaît également dans le code. Tous les codes étant alignés, le logiciel permet d'isoler et d'observer les questions qui présentent des caractéristiques communes. On peut ainsi visualiser toutes les questions qui présentent le même mot interrogatif, la même fonction communicative, la même forme verbale, etc. La manipulation des codes permet de détecter des affinités entre l'apparition d'une certaine variante et la présence d'une certaine caractéristique dans la question. Lorsque Coveney suspecte une affinité (c'est-à-dire l'effet probable d'un facteur sur une variante), il recourt à un calcul statistique pour en mesurer le

poids. Ainsi, l'étude de Coveney comprend plusieurs tableaux qui mesurent l'effet de différents facteurs sur la fréquence d'occurrence des variantes.

Etant donné que Coveney a affaire à des questions variables et à d'autres semi-catégoriques, il distingue deux types de facteurs (qu'il nomme « contraintes ») : les contraintes catégoriques et les contraintes variables. Les contraintes catégoriques excluent l'usage de l'une ou l'autre variante alors que les contraintes variables ne font qu'influencer l'usage de certaines variantes. Le fait que Coveney traite conjointement ces deux types de questions entraîne une particularité au niveau des comptes rendus quantitatifs. En effet, au moment de définir les fréquences observées des variantes, Coveney se base sur la formule suivante (par ailleurs tout à fait standard) :

$$\frac{\text{total des occurrences de la variante}}{\text{total des occurrences de la variante} + \text{total des occurrences potentielles de la variante}}$$

Or, les tokens semi-variables acceptent certaines variantes mais en rejettent d'autres, de sorte que le total des occurrences potentielles diffère souvent d'une variante à l'autre. Une autre conséquence de cette manière de faire est que le total des fréquences relatives n'arrive souvent pas à 100%. Ce phénomène complique légèrement la comparaison entre variantes, comme le montre le tableau suivant, donné dans Coveney (2002, 205) :

Tableau 7 - Fréquences relatives et observées des variantes de la question partielle chez Coveney

SVK	KSV	KESV	KVScI	KVSsn	K=S V
18/65	29/58	51/105	6/104	3/3	4/4
27.7%	50%	48.6%	5.8%		

Sur la ligne du milieu, les chiffres situés à gauche des barres obliques correspondent au total des occurrences de la variante alors que les chiffres de droite sont l'addition du total des occurrences et du total des occurrences potentielles de la variante. Sur la ligne du bas, on trouve les fréquences relatives, dont le total atteint 132%. Grâce à ce procédé, on peut observer que, malgré leurs faibles fréquences relatives, les variantes KSNP et K=S V ont été utilisées à chaque fois qu'elles pouvaient l'être (en d'autres termes, elles « réalisent toutes leurs occurrences potentielles » (Coveney, 2002, 205)).

Comme mentionné plus haut, Elsig recourt à l'analyse en règles variables pour mesurer l'influence des facteurs qui agissent sur la variation. Le système d'analyse en règles variables, dont nous avons donné un aperçu au point 1.2.2, est un outil d'analyse statistique développé originellement par Cedergren et Sankoff (1974). Elsig considère l'analyse en règles variables comme une « méthode analytique standardisée », du moins « dans le cadre de la sociolinguistique et de la théorie de la variation » (Elsig, 2009, 33). Il utilise la dernière version en date du logiciel VARBRUL, à savoir, la version nommée « GoldVarb ». L'utilisation de ce programme requiert que toutes les variantes soient disponibles pour chaque token traité.

Chez Elsig, les facteurs analysés comme potentiels conditionneurs de variation sont divisés en *groupes de facteurs*, chacun composé de deux ou plusieurs *facteurs individuels*. Par exemple, le groupe de facteurs nommé *identité du sujet* contient trois facteurs individuels : *tu*,

vous, et autres. A l'intérieur d'un groupe de facteurs, tous les facteurs individuels doivent être clairement définis et s'exclure mutuellement (c'est une des raisons pour lesquelles Elsig renonce à mesurer l'effet des facteurs pragmatiques). Dans un premier temps, les groupes de facteurs sont confrontés un par un aux données, pour déterminer s'ils exercent une influence notable sur le choix des variantes. De plus, cette étape permet de détecter d'éventuelles interactions entre facteurs (voir point 2.3.4.7). Une fois cette première étape passée, tous les facteurs retenus sont codés et introduits dans le logiciel d'analyse en règles variables. Ce dernier teste ensuite l'influence conjointe de tous les groupes de facteurs et indique lesquels ont une influence significative sur le choix des variantes. Il calcule également le poids de chaque facteur individuel à l'intérieur d'un groupe donné. Finalement, les résultats de l'analyse sont présentés dans deux grands tableaux (un pour les facteurs linguistiques et stylistiques, l'autre pour les facteurs sociaux) et la signification des chiffres qui y sont représentés sont expliqués par Elsig.

Les résultats de Coveney sont donnés de manière plus éparse que ceux d'Elsig ; on ne trouve pas chez Coveney de tableau qui représente l'ensemble des pourcentages d'utilisation des variantes en fonction de chaque facteur. Si l'étude de Coveney perd ainsi une certaine lisibilité par rapport à celle de Elsig, elle semble offrir plus de flexibilité au chercheur. Nous le verrons au point 2.3.4, les résultats de Coveney sont présentés à travers un grand nombre d'observations, certaines n'étant données qu'à titre d'hypothèses.

2.3.4.2 Choix des facteurs

Pour définir les facteurs susceptibles d'influencer l'usage des variantes de la question, Coveney et Elsig se sont tous deux basés sur les résultats d'études menées précédemment (souvent des études quantitatives) ainsi que sur le contenu de plusieurs grammaires. Ils ont également pris en compte les phénomènes observés en manipulant leurs propres données.

Elsig ne rend compte que de trois types de facteurs : linguistiques, sociaux et stylistiques. Il explique qu'il n'a pas jugé opportun de mesurer l'influence éventuelle de facteurs pragmatiques, cela pour deux raisons. Premièrement, il estime qu'il est extrêmement difficile, voire impossible, de définir certaines caractéristiques pragmatiques des questions telles que leur degré d'orientation, c'est-à-dire l'attente du locuteur d'une réponse positive ou négative. Deuxièmement, il explique que les différents traits pragmatiques des questions ne sont pas toujours mutuellement exclusifs, de sorte qu'on ne peut pas les introduire dans un logiciel d'analyse en règles variables comme celui qu'il utilise. Puisqu'il ne connaît pas de programme fiable qui permette d'analyser l'influence de facteurs qui ne s'excluent pas mutuellement, il renonce à rendre compte du conditionnement qui pourrait provenir des caractéristiques pragmatiques des questions.

Coveney considère quant à lui que la fonction communicative joue un rôle important dans le choix des variantes. Sa manière de traiter les données, moins standardisée et plus flexible que celle d'Elsig, lui permet de mesurer l'influence de tels facteurs. Pour mesurer l'impact des facteurs pragmatiques sur le choix des variantes de l'interrogation, Coveney doit préalablement attribuer une fonction communicative à toutes les questions de son corpus.

2.3.4.3 Attribution des fonctions communicatives

Coveney explique que l'attribution des fonctions communicatives est une étape « problématique » car la représentation du sens pragmatique d'un énoncé fait l'objet de nombreux désaccords. C'est donc sur la base de plusieurs sources différentes que Coveney

crée son propre dispositif pour l'identification des fonctions communicatives des questions : il explique s'être inspiré de trois approches différentes :

- Les catégories de forces illocutoires développées par Austin (1962) et Searle (1969)
- Les catégories de l'analyse conversationnelle élaborées par plusieurs chercheurs et résumées dans Levinson (1983)
- Une analyse communicative de l'interrogation établie pour l'anglais par Leech (1983)

Le dispositif classificatoire de Coveney se compose de trois « groupes de caractéristiques », composés chacun de plusieurs traits distinctifs, ou caractéristiques (*features* en anglais). Ces trois groupes concernent chacun un aspect différent de la pragmatique des questions.

Le premier groupe concerne la relation entre le locuteur, l'énoncé et l'interlocuteur. Etant donné qu'il ne comprend qu'un nombre réduit de traits distinctifs, nous en donnons la description complète et en reproduisons le tableau donné par Coveney (2002, 124). **A**, **B** et **C** sont les trois traits distinctifs qui constituent ce premier groupe et auxquels Coveney attribue une valeur + ou -. En fonction de la combinaison entre les valeurs de ces trois traits, une caractéristique pragmatique est attribuée à la question.

Tableau 8 - Premier groupe de caractéristiques pragmatiques de l'étude de Coveney

A : Le locuteur de l'énoncé est également l' « auteur » des mots de l'énoncé

B : L'auteur est/était aussi le destinataire

C : L'énoncé est une répétition identique (ou presque) de l'énoncé récent d'un autre locuteur

A	B	C	
+	+	-	question auto-adressée
-	-	-	question citée
-	+	-	question citée auto-adressée
-	-	+	question écho
+	-	-	question « ordinaire »

A l'image de ce premier groupe, chaque groupe de caractéristiques comprend plusieurs traits distinctifs auxquels on attribue une certaine valeur. La combinaison de ces valeurs détermine une certaine caractéristique pragmatique de l'énoncé. Finalement, la fonction communicative de chaque question se voit définie sur la base d'une ou de plusieurs de ces caractéristiques pragmatiques.

Le deuxième groupe de caractéristiques concerne tout ce qui touche aux connaissances et croyances du locuteur et, dans une moindre mesure, de l'interlocuteur. C'est à l'intérieur de ce groupe que Coveney détermine le degré d'orientation des questions. Pour établir les divers traits distinctifs de ce groupe, Coveney différencie p (le contenu propositionnel des questions totales) et q (la variable manquante des questions partielles, représentée par le mot interrogatif). Les traits distinctifs de ce groupe, au nombre de 9, comprennent les éléments suivants : « il est sûr que l'auteur ne sait pas si p ou q est vrai », « l'auteur suppose (ou pense que le destinataire suppose) que p est vrai » ou « l'auteur croit que p ou q est particulièrement important ». Les différentes combinaisons entre les 9 traits de ce groupe engendrent 11 caractéristiques pragmatiques différentes. En voici quelques exemples : « question affirmative qui attend un *oui* », « affirmation provisoire », « affirmation négative emphatique qui attend

un *non* »⁴⁵. Les affirmations emphatiques correspondent, chez Coveney, aux questions rhétoriques dont nous avons parlé au point 2.1.

Enfin, le troisième groupe sert à définir les caractéristiques du contenu propositionnel de l'énoncé. Il comprend 9 traits différents, du type: « p est présenté comme un fait objectif », « p = tu veux dire X », « p = je devrais faire X ». Pour ce groupe de caractéristiques, les valeurs attribuées aux différents traits peuvent être +, - ou /, ce dernier symbole signifiant que la valeur peut aussi bien être positive que négative. Pour les demandes d'information par exemple, on ne peut attribuer de valeur (ici +) qu'au premier des trois traits donnés plus haut. En effet, les traits « p = tu veux dire X » et « p = je devrais faire X » peuvent être tantôt présents, tantôt absents d'une demande d'information. Les combinaisons entre les 9 traits du groupe font ressortir 16 caractéristiques pragmatiques : en plus de la demande d'information déjà mentionnée, on trouve de nombreuses autres sortes de demandes, de même que des offres, des vérifications, etc.

Une caractéristique pragmatique issue d'un certain groupe peut parfaitement se combiner avec une caractéristique d'un autre groupe pour former la définition pragmatique (ou fonction communicative) complète de la question. Au total, Coveney ne dénombre pas moins de 55 fonctions communicatives différentes. En regardant les codes des fonctions communicatives attribuées aux questions du corpus, on remarque que certaines fonctions sont définies par un seul critère alors que d'autres accumulent plusieurs caractéristiques : *a* et *b* sont deux codes de fonctions communicatives rencontrés dans la liste de codes des questions totales de Coveney :

a. [clar] = demande de clarification

b. [e clar neg non] = question écho, demande de clarification, question orientée négative, qui appelle une réponse en non.

Parmi les nombreuses fonctions communicatives répertoriées par Coveney, on remarque que la grande majorité des questions se voient définies en tant que demande, offre, suggestion, vérification ou assertion. Les demandes sont divisées en de nombreux types (demandes d'opinion, de rappel, d'action, etc.), tandis qu'il n'existe qu'une sorte d'offre et une sorte de vérification (il s'agit en fait d'une vérification des connaissances de l'autre). Les suggestions, qui ne sont que des suggestions d'action, se divisent en deux groupes, selon qu'elles incluent ou qu'elles excluent le locuteur. Quant aux assertions, elles peuvent être emphatiques ou « tentées » (*tentative* en anglais). Les assertions emphatiques correspondent à une affirmation de polarité inverse à celle de leur structure, selon la définition souvent donnée des questions rhétoriques. Coveney appelle « assertions tentées » les questions qui agissent comme des assertions de doutes (et que l'on peut paraphraser par des assertions qui commencent par *peut-être que* ou *je ne sais pas si*). L'exemple (54) présente une assertion tentée qui pourrait être paraphrasée, selon Coveney, par *je ne sais pas si c'est bien*. Il explique que lorsqu'elles ne sont pas orientées, les assertions tentées se présentent toujours sous la forme d'une question alternative, comme c'est le cas en (54) :

(54) ...culturellement on a eu une influence là-dessus. / mais *est-ce que c'est bien ou mal ?* j sais pas. De toute façon ils étaient obligés / ... (Coveney, 2002, 140)

⁴⁵ Mais puisque les affirmations emphatiques n'attendent pas de réponse, il faut comprendre « assertion emphatique négative **qui implique une affirmation correspondante positive** » (Coveney, 2002, 158).

D'autres caractéristiques peuvent ensuite se greffer (ou pas) sur l'une des cinq fonctions de bases (la demande, l'offre, la suggestion, la vérification ou l'assertion). Ces caractéristiques concernent, par exemple, l'orientation de la question, son destinataire, le fait qu'elle soit une citation ou un écho, etc.

Il existe en outre trois types de questions qui ne correspondent à aucune des cinq fonctions de base données plus haut. Les deux premières sont ce que Coveney appelle *pre-announcement* et *post-announcement*. Dans le cas d'un *pre-announcement*, le locuteur demande à son interlocuteur s'il est au courant d'un fait tout en sachant qu'il ne l'est vraisemblablement pas. Le locuteur n'attend donc pas de réponse, de sorte que l'on ne peut interpréter ces questions comme des vérifications des connaissances de l'autre. En fait, le *pre-announcement* (55) sert à mettre en relief l'information qui va être délivrée (il semble correspondre à la « question notificative » de Borillo, présentée au point 2.1.3). Quant au *post-announcement* (56), il est utilisé pour mettre en relief une information qui a déjà été délivrée.

(55) ...il a passé son permis / il a très bien conduit // et puis l bonhomme euh / il était sûr de l'avoir son permis hein / le bonhomme lui a dit 'vous (n) l'avez pas' / (alors il comprenait pas) / tu sais pourquoi il (n l) a pas eu ? parce qu'il avait oublié d mettre la ceinture. // il avait oublié d mettre la ceinture...(Coveney, 2002, 148)

(56) ...sa mère était économe dans la colonie où on était avec Christine à c moment-là / j la connais elle avait neuf ans. / maintenant elle a vingt-quatre ans. / tu vois un peu c qui s'est passé ? elle est prof de gym et tout maintenant. (Coveney, 2002, 150)

Le troisième type de questions qui ne correspond à aucune des cinq fonctions de base est ce que Coveney appelle les *sub-topic-introducing questions*. Ces questions ressemblent à un *pre-announcement* dans la mesure où le locuteur pose une question à laquelle il répond immédiatement. Cependant, la *sub-topic-introducing question* ne sert pas à délivrer une information que le locuteur juge particulièrement intéressante, comme c'est le cas du *pre-announcement*. Son but est plutôt d'aider le locuteur à développer son sujet comme il l'entend: en simulant une question qui lui serait posée, le locuteur organise son propos, comme en (57).

(57) ...hein parce que / l'économat ça (me) prenait quand même pas tout mon temps. / alors ça consiste en quoi ? eh ben – ça consiste euh à animer / à animer un groupe d'animateurs c'est le cas de le dire. / (comme) je vous l'expliquais tout à l'heure euh i faut quelquefois les pousser...(Coveney, 2002, 147)

Le système mis en place par Coveney pour attribuer une fonction communicative précise à toutes les questions de son corpus lui a permis de venir à bout de cette tâche sans trop de difficulté. Il explique que le fait d'avoir pris part aux entretiens en tant que participant à la conversation l'a aidé à résoudre plusieurs ambiguïtés. Finalement, seules une dizaine de questions n'ont pas pu être définies précisément sur le plan pragmatique. Selon lui, la catégorie de questions à laquelle il est le plus difficile d'attribuer une fonction communicative est celle des questions citées (lorsque le locuteur n'est pas l'auteur original de la question). Il explique que deux problèmes se posent avec les questions citées: d'une part, il est parfois difficile de savoir si la question est véritablement citée ou pas et d'autre part, le contexte des questions citées est souvent moins fourni que pour les autres questions. Ainsi, le manque de contexte en (58) empêche de savoir si le destinataire de la question citée est le locuteur lui-même ou une autre personne :

(58) si ç avait été un accident j'aurais dit 'mon dieu / *qu'est-ce qui s'est passé ?* ' euh – c'est vrai à la suite euh des fois / d'une malveillance de quelqu'un... (Coveney, 2002, 153)

Puisque la question en (58) n'est pas réelle mais hypothétique, Coveney pense que le locuteur ne s'est pas soucié de lui imaginer un destinataire précis : il considère que la fonction communicative de cette question est indéterminée. L'exemple (59) présente une autre question à laquelle Coveney n'a pas pu attribuer de fonction communicative claire :

(59) (...)...y avait des shorts euh des / des maillots d corps des chaussettes dans tous les coins / on les retrouvait plus / euh certaines étaient mouillées f euh *qu'est-ce qu'on fait ?* on s rhabille sur place et on retourne jouer ailleurs euh / euh – bon là y a une espèce de flottement (...) (Coveney, 2002, 154)

Coveney n'a pas réussi à déterminer si cette question était donnée comme une citation (par exemple d'un enfant qui aurait posé la question à l'animateur) ou comme une *sub-topic-introducing question*.

2.3.4.4 Facteurs linguistiques

Coveney discute d'un grand nombre de facteurs linguistiques qui pourraient influencer l'apparition des variantes de la question totale et de la question partielle. La plupart de ces facteurs ont déjà été abordés au point 2.2, de sorte que nous ne mentionnerons ici que les facteurs que Coveney a dû prendre en compte en traitant son corpus. Les facteurs analysés par Coveney sont divisés d'une part selon la variable sur laquelle ils agissent (question totale ou questions partielle) et d'autre part selon leur nature, catégorique ou variable.

Au niveau des questions totales, Coveney n'a répertorié qu'un seul facteur linguistique catégorique : lorsque le sujet est *je* et que le verbe est au présent, l'inversion (VScl) est très souvent exclue (à l'exception de certains verbes communs du type *pouvoir, être, avoir*, etc.). Ce facteur a le même effet au niveau des questions partielles, puisqu'il empêche la variante KVScl. Coveney a pris en compte deux autres contraintes catégoriques qui agissent sur les variantes des questions partielles: la première concerne la variante KSV, incompatible avec le mot interrogatif *que/quoi*. La seconde, « plus difficile à spécifier » (Coveney, 2002, 193), n'affecte qu'une question du corpus :

(60)...où est l'inconvénient ?... (Coveney, 2002, 193)

Coveney remarque que cette question pourrait difficilement être exprimée par les variantes KSV et KESV. Il pense que ce phénomène est dû à la longueur du sujet, passablement plus importante que celle du verbe ; selon lui, un sujet nominal accompagné d'un verbe monosyllabique pourrait exclure l'usage des variantes KSV et KESV.

Coveney n'a répertorié aucun facteur linguistique variable qui aurait une influence notable sur le choix des variantes de la question totale. Il faut préciser que parmi les trois variantes de la question totale établies par Coveney, seules deux ont été utilisées par les locuteurs de son corpus : ESV et SV. C'est donc dans le choix entre ESV et SV que Coveney dit n'avoir repéré aucune contrainte linguistique variable de taille. En revanche, il a remarqué que certaines contraintes linguistiques variables pouvaient exercer une influence sur les variantes de la question partielle. La plus évidente de ces contraintes, et la plus étendue dans le corpus de Coveney, concerne l'influence du mot interrogatif sur le choix des variantes KESV, SVK et KSV. L'effet de *que/quoi, comment, combien* et *où* est notable sur les variantes de Coveney. Ce dernier regrette cependant que la faible occurrence des mots interrogatifs *quand, quel, pourquoi* et *qui* ne permette pas d'obtenir des résultats statistiques

pertinents à leur sujet. De même, le peu d'occurrence des variantes KVScl, KVSsn et K=S V fait qu'il est difficile d'établir des contraintes variables qui agiraient sur leur apparition. Enfin, Coveney a remarqué que la longueur du groupe sujet-verbe-complément (SVC), en combinaison avec celle du mot interrogatif, pouvait avoir une influence sur l'usage de la variante SVK. Selon ses calculs, la variante SVK a peu de probabilité d'apparaître lorsque le groupe SVC a sensiblement plus de syllabes que le mot interrogatif.

Avant de procéder à l'analyse statistique en règles variables, Elsig a distingué huit groupes de facteurs linguistiques susceptibles d'influencer le choix des variantes de la variable *question totale*. Selon la procédure donnée plus haut, il a testé individuellement l'influence de chaque groupe de facteurs pour déterminer s'il était pertinent de les introduire dans le programme d'analyse en règles variables. Elsig a renoncé à introduire deux des huit groupes de facteurs parce qu'ils ne se sont pas révélés influents dans la première analyse (il s'agissait des groupes *temps et mode du verbe* ainsi que *classe grammaticale du verbe*). Il a également renoncé à coder le facteur *identité du verbe lexical*, pour des raisons pratiques cette fois-ci : si chaque verbe lexical apparaissant dans les questions devait constituer un facteur individuel, ces derniers seraient extrêmement nombreux et difficiles à introduire dans un programme d'analyse statistique.

Elsig a donc gardé cinq groupes de facteurs linguistiques, qu'il a codés et introduits dans le logiciel GoldVarb. Le nombre de facteurs individuels à l'intérieur de chaque groupe de facteurs varie de deux à cinq : au total, l'analyse comprend 16 facteurs linguistiques individuels. Les cinq colonnes du tableau 9 présentent les cinq groupes de facteurs linguistiques ainsi que tous les facteurs individuels introduits dans l'analyse en règles variables d'Elsig (Elsig, 2009, 93).

Tableau 9 - Facteurs linguistiques de l'analyse en règles variables d'Elsig

Identité du sujet	Fréquence du verbe	Verbes cognitifs	Longueur du verbe	Traitement parallèle
• vous • tu • autres	• une seule occurrence • rare • fréquence moyenne • fréquent • très fréquent	• verbe cognitif • autres	• monosyllabique • dissyllabique • polysyllabique	• pas de token qui précède • token précédent identique • token précédent différent

L'influence des verbes cognitifs sur le choix des variantes a été détectée par Elsig avant l'analyse en règles variables, alors qu'il testait l'influence de l'identité du verbe sur les variantes. L'analyse en règles variables a confirmé que cette catégorie de verbes favorisait l'usage de la variante VScl. Le groupe de facteurs nommé « traitement parallèle » (*parallel processing* en anglais), sert à mesurer si la présence d'une question dans les 20 lignes qui précèdent peut influencer le choix de la variante (selon une hypothèse qui veut que l'on répète facilement la même structure plusieurs fois dans une séquence de parole). Ce facteur s'est révélé actif sur la variante ESV, cette dernière étant favorisée lorsqu'elle est précédée d'une

autre question ESV. La fréquence du verbe agit principalement sur la variante SV-*tu*, particulièrement sollicitée lorsque la fréquence du verbe est basse. En revanche, la variante SV-*tu* se raréfie lorsque le verbe a plus de deux syllabes : en effet, les résultats d'Elsig montrent que plus le verbe est long, plus VScl et ESV sont utilisés, au détriment de SV-*tu*. Finalement, le facteur qui exerce la plus grande influence sur le choix des variantes est la nature du sujet. Elsig observe que la variante VScl n'apparaît qu'avec des sujets de deuxième personne (à l'exception d'une occurrence avec le sujet *ils*). Mais l'influence de la nature du sujet est également notable au niveau de la variante SV-*tu*, puisque qu'elle n'apparaît jamais lorsque le sujet est *vous*.

Selon les résultats d'Elsig, la variante SV est la seule dont l'usage n'est pas influencé par les cinq groupes de facteurs linguistiques analysés. Elle représente pourtant environ un tiers de toutes les occurrences de questions. Elsig pense que ce phénomène pourrait être dû au fait que la variante SV ne présente aucun « mécanisme syntaxique pour exprimer l'interrogation » (2009, 97). Cette dernière observation fait partie des nombreuses réflexions d'Elsig sur les liens entre la syntaxe de l'interrogation (considérée dans une perspective historique et générative) et l'usage qui en est fait. Le chapitre 5 de son livre (« Interprétation et discussion des résultats ») est principalement consacré à la recherche de tels liens. Nous n'en rendrons pas compte ici car ces considérations ne concernent pas l'analyse variationniste à proprement parler.

2.3.4.5 Facteurs pragmatiques

Comme pour les facteurs linguistiques, Coveney distingue, parmi les facteurs pragmatiques qui influencent la variation des questions, ceux qui exercent des contraintes catégoriques et ceux qui exercent des contraintes variables.

Coveney constate qu'au moins deux fonctions communicatives peuvent exclure l'une ou l'autre des variantes de la question totale. Premièrement, il remarque que les assertions tentées (exemple 54 ci-dessus) ne peuvent être exprimées par la variante SV. Il suppose, en outre, que les assertions emphatiques et les questions *sub-topic-introducing* excluent également l'usage de la variante SV mais le peu d'occurrence de ces fonctions communicatives ne lui permet pas de l'affirmer. La deuxième contrainte catégorique exercée sur les questions totales est liée aux questions appelées *pre-announcements* (exemple 55), pour lesquelles Coveney constate que la variante ESV est indisponible.

Les fonctions communicatives peuvent également exercer des contraintes catégoriques sur les variantes de la question partielle, mais leur rôle est quelque peu difficile à définir. Coveney observe que la variante SVK est souvent indisponible, pour des raisons qui ne sont pas linguistiques mais qui ne sont pas clairement liées à une fonction communicative particulière. La seule fonction communicative qui exclut assez clairement l'usage de SVK est l'assertion emphatique : pour l'exemple (61), comme pour toutes les assertions emphatiques de son corpus, Coveney estime que l'usage de SVK serait « particulièrement inapproprié » (2002, 202) :

(61) ..alors c'est à la fois un petit handicap mais *qu'est-ce tu veux ?* ça fait partie des règles de vie... (Coveney, 2002, 202)

Pour les cas, relativement nombreux, où l'on n'a pas affaire à une assertion emphatique mais où l'usage de SVK semble néanmoins inapproprié, Coveney recourt à une explication dont nous rendrons compte plus bas, car elle concerne principalement les facteurs variables.

Les facteurs variables agissant au niveau des questions totales peuvent être résumés de la manière suivante : toutes les questions qui n'attendent pas de réponse de l'interlocuteur favorisent la variante ESV, tandis que dans les cas où la question appelle une réponse de l'interlocuteur, la variante SV est largement sollicitée. Les « questions qui n'appellent pas de réponses de l'interlocuteur » forment ainsi un groupe de fonctions communicatives qui comprend toutes les questions citées ainsi que les questions auto-adressées.

Au niveau des questions partielles, Coveney fait plusieurs hypothèses sur des liens éventuels entre certaines fonctions communicatives et certaines variantes. Il s'agit d'observations isolées, relevant plus de l'analyse au cas par cas que de l'analyse quantitative. Coveney a toutefois repéré des phénomènes significatifs dans l'usage de la variante SVK. Le premier n'est qu'une supposition puisque Coveney ne dispose pas d'un nombre suffisant de preuves pour l'affirmer. Coveney suppose donc que la variante SVK a plus de chance d'être choisie que les autres lorsque le locuteur attend une réponse (d'autrui ou de lui-même) : il s'agit là encore d'une influence qui concerne tout un groupe de fonctions communicatives. Mais cette contrainte variable n'est pas la seule à agir sur l'usage de la variante SVK. En fait, Coveney considère que l'information contenue dans le groupe sujet-verbe (et complément(s) éventuel(s)) influence également l'utilisation de SVK : selon lui, cette variante serait favorisée lorsque l'information du groupe verbal a peu d'importance (parce qu'elle a déjà été donnée ou parce qu'elle est supposée connue). Ainsi, le corpus de Coveney comprend beaucoup d'exemples de structures SVK utilisées dans des situations telles que (62), lorsque le locuteur reprend une information qui vient d'être donnée, ou telles que (63), lorsque l'information contenue dans le groupe verbal est clairement partagée par les interlocuteurs :

- (62) X : Bertrand est-ce que j peux avoir des agrafes et d la colle ?/
 B : est-ce que tu peux avoir des agrafes et d la colle ?/
 X : s'il te plaît/
 B : j crois que là y a une agrafeuse. / (combien...) pas beaucoup y en a encore. /
 X : merci/
 B : *t en veux combien d'agrafes ? /...* (Coveney, 2002, 224)

- (63) mais vous êtes en France pour combien de temps ?/ (Coveney, 2002, 224)

Ce facteur, que Coveney qualifie de discursif et pragmatique, se différencie des autres facteurs donnés jusqu'à présent dans la mesure où il n'est lié à aucune des fonctions communicatives que Coveney a attribuées aux questions. Si Coveney considère que ce facteur n'est pas contraignant (il agit de manière variable), il estime qu'il pourrait être à l'origine de plusieurs des cas (mentionnés plus haut) où la variante SVK est exclue sans que l'on puisse clairement définir pourquoi. Ainsi, si l'information contenue dans le groupe verbal est tout à fait nouvelle, l'utilisation de SVK n'est souvent pas appropriée. La question de l'exemple (64) constitue un des nombreux cas dans lequel Coveney doute de la disponibilité de la structure SVK et de son équivalence pragmatique avec les autres variantes :

- (64) ...on téléphone on écrit en disant ' mais *quand est-ce que vous venez nous inspecter ?* on veut vous voir euh' / alors qu'auparavant [RIRES] / on voulait pas... (Coveney, 2002, 225)

Coveney estime qu'il est peu satisfaisant voire impossible de remplacer la question donnée en (64) par la variante SVK (*vous venez nous inspecter quand*). Selon lui, ce phénomène provient probablement du fait que l'information « vous venez nous inspecter » n'est pas considérée comme acquise par l'interlocuteur.

2.3.4.6 Facteurs socio-stylistiques

Coveney et Elsig considèrent tous deux que les caractéristiques sociales des locuteurs ainsi que divers éléments liés au contexte stylistique de l'énonciation peuvent jouer un rôle dans le choix des différentes variantes de la question totale (et partielle pour Coveney). Coveney, mesure l'influence des facteurs sociaux et stylistiques après avoir fait l'inventaire des facteurs linguistiques et pragmatiques qui agissent sur la variation des questions. Quant à Elsig, il mesure l'influence des facteurs stylistiques dans le même tableau que celui qui rend compte de l'influence des facteurs linguistiques. Il procède ensuite à une seconde analyse en règles variables, dans un autre tableau, pour mesurer le poids des facteurs sociaux.

Coveney s'intéresse à l'influence de trois types de facteurs sociaux sur les variantes des questions totales et partielles : l'âge, la classe socio-professionnelle et le sexe. Il divise les locuteurs en deux classes d'âge : les 17-22 ans et les 24-37ans. Coveney n'a interrogé que trois personnes âgées de plus de 37 ans : il s'agit de trois femmes entre 50 et 60 ans. Ces femmes forment un groupe à elles seules et Coveney mentionne leurs usages, tout en précisant que leur nombre réduit ne permet pas d'établir des comparaisons ou des généralités. Pour définir la classe sociale de ses informateurs, Coveney s'est basé sur leur profession (ou celle des parents, pour les étudiants). Il a ensuite attribué chaque catégorie socio-professionnelle à l'une des trois classes suivantes : populaire, moyenne ou supérieure. Coveney exemplifie chaque classe selon les professions suivantes: agriculteurs et ouvriers pour la classe populaire, instituteurs et commerçants pour la classe moyenne, professeurs et professions libérales pour la classe supérieure.

Avec un total de 30 informateurs, Coveney ne compte que peu de sujets à l'intérieur de chaque catégorie sociale établie selon les trois distinctions données plus haut. Certaines personnes n'ayant pas produit beaucoup de questions, le nombre de tokens est très faible pour certaines catégories de locuteurs. Pour parvenir néanmoins à comparer les utilisations des variantes parmi les différentes catégories de locuteurs, Coveney a compté tous les tokens à l'intérieur d'un groupe et il en a établi les fréquences relatives. Il explique que cette manière de faire s'éloigne de la méthodologie conventionnelle, qui consiste à calculer les fréquences relatives d'utilisation pour chaque locuteur et à faire ensuite la moyenne de toutes ces fréquences individuelles pour obtenir les fréquences relatives d'un groupe. Coveney a dû renoncer au calcul des fréquences individuelles parce qu'un certain nombre de ses informateurs n'ont pas produit assez de tokens. Dès lors, ses résultats sont biaisés par le fait que les locuteurs qui ont produit beaucoup de tokens sont plus représentés que ceux qui en ont produit peu.

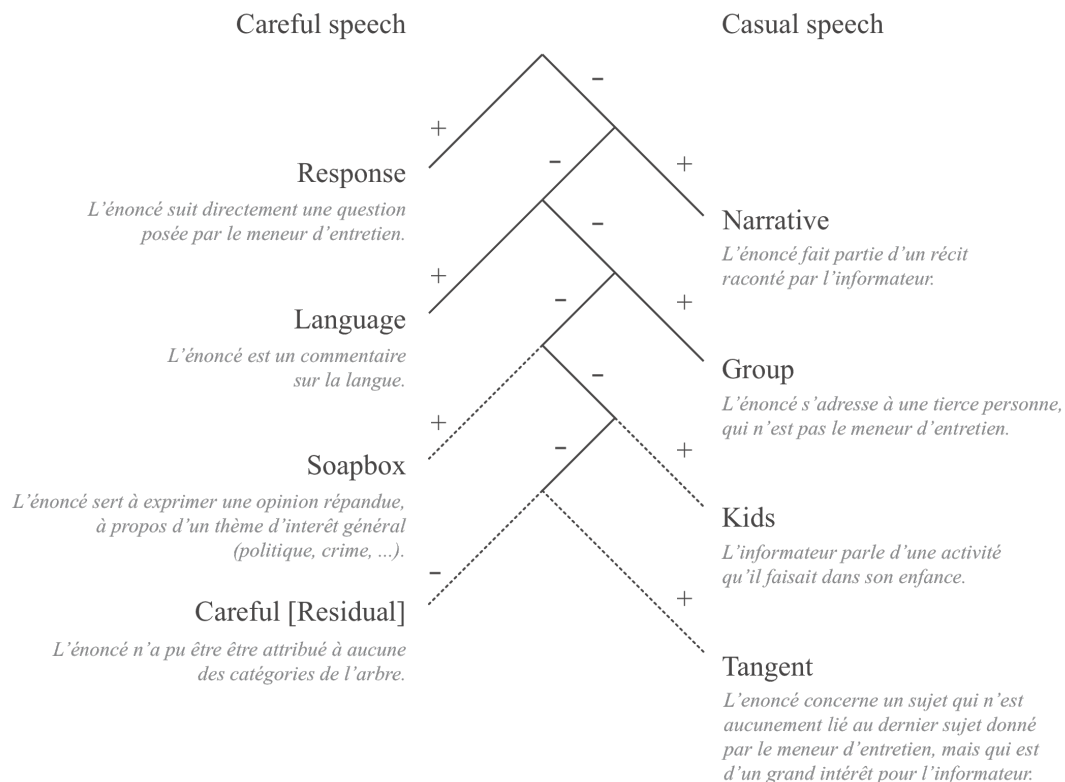
En dépit de cet inconvénient méthodologique, Coveney analyse les fréquences d'utilisation relative des variantes pour chaque groupe de locuteurs. Les résultats de la classe populaire ne sont pas pris en compte au niveau des questions totales en raison d'un nombre insuffisant de tokens (en tout, sept questions totales ont été prononcées par les informateurs de la classe populaire). Pour les questions partielles, les fréquences d'utilisation des variantes K=S V et KVS_{sn} ont dû être omises, là encore, à cause de leur nombre très réduit.

Au niveau des questions totales, Coveney observe que la variante ESV est favorisée par le groupe de locuteurs les plus âgés ainsi que par la classe supérieure. Il remarque que cela corrobore les observations des grammairiens qui considèrent souvent que ESV a une valeur socio-stylistique plus élevée que VS. Coveney indique toutefois que ces résultats doivent être interprétés avec prudence, en raison du nombre réduit d'informateurs (30 locuteurs divisés en plusieurs catégories) et de tokens (109 questions totales et 111 questions partielles). Ceci est d'autant plus vrai au niveau des questions partielles, où le total des questions est réparti entre

quatre variantes. Les résultats montrent de légères différences d'usage entre les différentes catégories de locuteurs (la plus marquée est l'usage de KESV favorisé par les femmes) mais Coveney constate que dans l'ensemble, les résultats sont moins clairs qu'au niveau des questions totales.

Etant donné que les informateurs de Coveney n'ont participé qu'à un seul entretien dont les modalités étaient similaires du début à la fin, il est difficile de mesurer l'influence de facteurs stylistiques ou contextuels dans leur manière d'employer les structures interrogatives. Coveney a toutefois procédé à une analyse que l'on peut qualifier de stylistique puisqu'elle examine l'influence du destinataire de la question sur le choix des variantes de la question. En fait, Coveney a voulu mesurer si les questions qui lui étaient posées (alors qu'il menait l'entretien) se présentaient différemment que les questions posées à d'autres personnes (pendant les entretiens, les informateurs étaient parfois interrompus et leurs échanges avec d'autres personnes du camp étaient enregistrés). Il n'a pris en compte pour cette analyse que les questions qui appellent une réponse de l'interlocuteur. Les résultats ne sont pas très représentatifs à cause du peu de questions adressées à de tierces personnes, mais ils suggèrent que l'utilisation des variantes diffère peu lorsque les questions sont posées à Coveney ou à d'autres personnes.

Elsig explique que l'exigence, requise par le système d'analyse en règles variables qu'il utilise, de ne mesurer que des facteurs clairement définis et mutuellement exclusifs est plus facile à respecter au niveau des facteurs linguistiques que des facteurs stylistiques. Pour parer à la difficulté que l'on rencontre à attribuer un style à chaque token de manière claire et exclusive, Elsig s'est basé sur « l'arbre de décision » de Labov que nous avons mentionné au point 1.2, et dont nous reproduisons ici le schéma donné par Elsig (2009, 65). Les commentaires explicatifs (en italiques) que nous donnons sous chaque branche de l'arbre ne sont pas présents dans le modèle original.



Suivant la nature de l'énoncé dans lequel la variante apparaît, cette dernière sera associée à un style soutenu (à gauche) ou à un style familier (à droite). L'arbre se lit de haut en bas et la première décision à prendre concerne la caractéristique « réponse » : si la variante apparaît dans la première phrase qui répond à une question du chercheur, elle est automatiquement associée au style soutenu. Si ce n'est pas le cas, il faut passer à la deuxième décision qui concerne le caractère narratif de l'énoncé : si l'énoncé fait partie d'un récit du locuteur, il sera placé dans la catégorie de style familier. Si l'énoncé n'est ni une réponse, ni un récit, on passe à la troisième décision, et ainsi de suite jusqu'au bas de l'arbre. Le groupe de facteurs « style » d'Elsig ne comporte que deux facteurs individuels, qui correspondent aux deux côtés de l'arbre : style soutenu et style familier.

Le second facteur stylistique pris en compte par Elsig concerne la place de la question dans l'entretien entre l'informateur et le chercheur. Ce groupe de facteurs comprend trois facteurs individuels, selon que la question se trouve en début, au milieu ou en fin d'entretien. Partant de l'hypothèse que plus l'entretien avance, plus l'informateur se sent à l'aise, et donc libre de parler de manière plus détendue, Elsig pensait que l'usage des formes colloquiales pourrait être plus important vers la fin que vers le début de l'entretien. Finalement, l'analyse a infirmé cette hypothèse : sans qu'Elsig ne puisse expliquer pourquoi, il a observé que la variante SV (liée à un style soutenu en français québécois) était plus fréquente à la fin des entretiens, au détriment des variantes SV-*tu* et VScl, généralement liées à un style détendu.

Dans son deuxième tableau d'analyse en règles variables, Elsig mesure l'effet de quatre facteurs sociaux. Il en avait prévu cinq au départ mais il a renoncé à introduire dans l'analyse le facteur « éducation », en raison de son interaction avec les facteurs « profession » et « âge »⁴⁶. Le tableau 10 reproduit les quatre groupes de facteurs sociaux retenus par Elsig avec leurs facteurs individuels respectifs :

Tableau 10 - Les quatre groupes de facteurs sociaux chez Elsig

quartier de domicile	sexe	âge	profession
Mont-Bleu (classe moyenne-supérieure)	homme	15-34	non qualifiés et sans emploi
Vieux-Hull (classe ouvrière)	femme	≥ 35	vente et services
			spécialistes et cadres

Le groupe de facteurs qui s'est révélé le plus influent sur le choix des variantes de la question totale est l'âge. Elsig observe que la variante SV-*tu* est privilégiée par les jeunes locuteurs, qui tendent à éviter VScl. Il estime que cette tendance pourrait indiquer que la variante SV-*tu* est en train de subir un changement et de s'imposer comme variante *par défaut*.

Elsig observe en outre que la variante ESV fonctionne clairement comme une variante de prestige puisqu'elle apparaît le plus fréquemment chez les locuteurs hautement qualifiés,

⁴⁶ L'interaction entre le facteur éducation et le facteur âge résulte du fait que les jeunes informateurs sont systématiquement plus éduqués que leurs aînés : ainsi, aucun informateur âgé de moins de 34 ans ne correspond à la catégorie « éducation primaire » tandis qu'aucun informateur entre 45 et 64 ans n'entre dans la catégorie « éducation supérieure ».

dans des contextes stylistiques soutenus. Elle est également plus fréquente parmi les locuteurs du quartier de Mont-Bleu que parmi ceux qui habitent à Vieux-Hull.

2.3.4.7 Interactions entre facteurs

Dans leurs recherches de corrélations entre les différents facteurs intra et extra-linguistiques et l'utilisation des variantes de l'interrogation, Coveney et Elsig sont conscients de la possibilité d'interaction entre les effets des différents facteurs. Ces interactions peuvent avoir lieu entre facteurs de même nature (linguistiques, sociaux, etc.), mais aussi entre des facteurs de natures différentes.

Avant de procéder à ses analyses en règles variables, Elsig a découvert plusieurs interactions, parmi certains facteurs linguistiques d'une part, et parmi certains facteurs sociaux, d'autre part. Or, il est essentiel que les groupes de facteurs introduits dans le programme d'analyse en règles variables n'interagissent pas entre eux, car le programme n'est pas fait pour traiter les interactions entre facteurs. Chaque groupe de facteurs doit donc avoir une influence indépendante des autres. Pour détecter les facteurs qui interagissent, Elsig compare la première partie des analyses (où chaque groupe de facteur est considéré individuellement) et une première analyse non définitive en règles variables. Lorsque le poids d'un groupe de facteurs change passablement entre ces deux étapes, Elsig mesure son interaction avec d'autres groupes de facteurs. C'est ainsi qu'il a constaté que le facteur « longueur du verbe » interagissait avec le facteur « identité du sujet » au niveau de la variante VScl. Ainsi, la longueur du verbe a deux effets opposés sur l'apparition de la variante VScl, suivant la nature du sujet. Lorsque le sujet est *vous*, la fréquence de VScl augmente proportionnellement au nombre de syllabes du verbe alors qu'avec le sujet *tu*, l'usage de VScl est favorisé par les sujets courts. Pour remédier à ce problème, Elsig a dû renoncer à mesurer le poids du facteur *sujet* sur l'apparition de la variante VScl. Au niveau des facteurs sociaux, Elsig n'a pas pu introduire le groupe de facteurs « éducation » dans l'analyse en règles variables à cause d'interactions avec les groupes de facteurs « âge » et « profession ». L'analyse finale en règles variables ne doit donc contenir que des facteurs dont les effets ne peuvent interagir entre eux.

Coveney estime que les interactions entre les facteurs liés à l'éducation et ceux liés à la profession sont bien connues ; il décide donc de ne prendre en compte que les seconds. Il explique en outre que l'effet des différents facteurs (linguistiques, pragmatiques, sociaux) sur le choix des variantes de l'interrogation est toujours simultané ; en effet, au moment de poser une question, le choix du locuteur peut être influencé par plusieurs facteurs de différentes natures, dont il n'est pas toujours évident de mesurer les poids respectifs. Cette difficulté peut être illustrée par la question que se pose Coveney au moment d'éclaircir le lien entre l'usage des structures ESV et KESV, les questions citées et l'âge des locuteurs. En effet, Coveney remarque que les questions citées tendent à se présenter sous la forme ESV ou KESV ; il remarque aussi que les questions citées, fréquentes dans le groupe des 24-37 ans, sont presque absentes dans le groupe 17-22 ans. Or, les locuteurs de 24-37 ans utilisent passablement plus de structures KESV et considérablement plus de structures ESV que les plus jeunes. D'où la question suivante énoncée par Coveney :

Les questions citées tendent-elles à avoir des structures ESV et KESV parce qu'elles sont citées, ou parce qu'elles sont produites principalement par des locuteurs qui utilisent de toute façon plus ESV et KESV » ? (2002, 236)

Coveney pense qu'il existe une motivation communicative dans l'usage des structures ESV et KESV (i.e. signaler qu'aucune réponse n'est attendue) et estime dès lors que le caractère cité

de la question influence l'usage de ESV et KESV. Mais il ne s'interroge pas sur l'effet que peut avoir le grand nombre de questions citées sur les fréquences d'apparition des structures ESV et KESV chez les 24-37 ans. Il nous semblerait pourtant intéressant de chercher à savoir dans quelle mesure la préférence du groupe des 24-37 ans pour les structures ESV et KESV peut s'expliquer en fonction du nombre de questions citées qu'ils utilisent.

Les questions qui viennent d'être soulevées montrent qu'il n'est pas toujours évident d'attribuer l'apparition de l'une ou l'autre structure interrogative à une cause précise. En établissant des liens directs entre l'usage de certaines structures et de certains éléments extra ou intra-linguistiques dans l'énoncé, on risque d'aboutir à une image faussée des schémas d'usage des questions. Une observation et une confrontation minutieuse des différents effets observés (c'est-à-dire, de l'influence de chaque facteur) semble donc primordiale pour éviter de telles dérives. Enfin, il faut compter avec l'influence de facteurs difficilement perceptibles et/ou pondérables : Coveney estime que plusieurs facteurs ont probablement agi sur la manière dont ses informateurs ont utilisé les variantes de l'interrogation, sans qu'il n'ait pu en mesurer l'influence. Ainsi, la personnalité de l'informateur, sa relation avec le chercheur de même que sa perception de la situation de communication et de ses normes sont autant de facteurs dont Coveney dit ne pas avoir pu rendre compte, en dépit de leur rôle probable sur le choix de ses informateurs en matière de formes interrogatives.

3 Synthèse critique

Cette dernière partie nous servira à peser les avantages et les inconvénients des approches variationnistes de l'interrogation, à la lumière des considérations générales sur la variation et la variable sociolinguistique données au point 1. Nous nous baserons principalement sur les deux études variationnistes présentées au point 2.3, mais les études quantitatives du point 2.2 seront également mentionnées là où des comparaisons peuvent être établies.

3.1 Difficultés méthodologiques liées aux études variationnistes (et quantitatives) sur les questions

3.1.1 Définition de *la question*

Nous avons constaté au point 2.1 que *la question* était un objet linguistique difficile à décrire et à circonscrire. Au niveau formel, elle présente de multiples structures, dont plusieurs ne servent pas uniquement à interroger. Au niveau de ses fonctions, la question est également très vaste et les chercheurs ne s'accordent pas tous sur la nature des paramètres sémantiques et pragmatiques qui entrent en compte dans sa définition. Plusieurs tentatives de définitions sémantico-pragmatiques ont été évoquées au point 2.1.3: la première, basée sur la double condition d'ignorance du locuteur et d'appel de réponse, met l'accent sur les véritables demandes d'information, mais se révèle insatisfaisante pour traiter les nombreux cas où le locuteur n'ignore pas (ou pas complètement) la réponse. Deux autres types d'approches ont été décrites : celles qui cherchent à englober toutes les questions au moyen d'une définition de type modulaire et celles qui restreignent le domaine de la question en excluant, par exemple, les questions rhétoriques.

Aucune des études quantitatives et variationnistes que nous avons présentées ne donne de définition de la question. Du fait que ces études s'intéressent avant tout aux structures syntaxiques de l'interrogation, on pourrait penser que la définition fonctionnelle des questions ne les concerne guère, du moins au niveau des questions partielles, qui présentent toujours, dans leur structure, une marque lexico-syntaxique de l'interrogation.⁴⁷ Ceci dit, les questions totales se présentent très souvent sous la forme sujet-verbe, qui peut servir aussi bien à former des assertions que des questions. Sur quels critères se basent donc les études quantitatives et variationnistes pour déterminer quels énoncés sont pris en compte comme questions ? Ce problème se pose en vérité aussi bien au niveau des questions partielles que des questions totales. En effet, il nous semble important de savoir si les questions partielles rhétoriques sont prises en compte comme n'importe quelle autre question ou si elles sont écartées de l'étude en raison de leur valeur assertive. Il serait utile, de ce point de vue, que les auteurs d'études quantitatives ou variationnistes déterminent clairement les limites et les caractéristiques (formelles et fonctionnelles) des objets qu'ils appellent « questions » ou « interrogations ».

Pour Coveney et Elsig, le critère qui fait foi dans l'identification des interrogations considérées comme variantes est la possibilité de substitution par les autres structures variantes. Ce critère implique que les structures en alternance soient syntaxiquement interchangeables et que leur substitution n'entraîne aucune modification du statut sémantico-

⁴⁷ Même si, nous l'avons vu, il est parfois difficile de distinguer, au niveau des structures de l'interrogation partielle, les questions directes des questions indirectes (voir point 2.3.3.3).

pragmatique de la question. Les nombreuses catégories de fonctions communicatives de Coveney indiquent qu'il a une vision large du domaine pragmatique des questions: en effet, seules les définitions les plus englobantes (cf. Berrendonner et Bonhomme, point 2.1.3) incluent dans le domaine de la question les énoncés que Coveney appelle « assertions tentées » ou « assertions emphatiques ». Il est en revanche plus difficile de définir si Elsig aborde le domaine de la question dans son sens le plus large, ou d'une manière plus restrictive. En fait, deux passages de son étude se rapportent à ce sujet : lorsqu'il présente la méthodologie de son étude, il explique que l'on doit éliminer les variantes de l'interrogation qui apparaissent « dans un contexte qui n'est pas interrogatif » (2009, 35). Plus tard, lorsqu'il présente les questions du corpus du XIX^{ème} siècle, il dit que la majorité des tokens variants sont des demandes d'information, mais que « certains tokens remplissent d'autres fonctions communicatives telles que demandes de permission ou *pre-announcements* » (2009, 50). Il ajoute que ces fonctions sont maintenues quelle que soit la variante utilisée. Cela signifie qu'Elsig considère les demandes de permission et les *pre-announcements* comme faisant partie du contexte interrogatif. Mais il n'explique pas quelles sont pour lui les limites de ce « contexte interrogatif » ; aucune mention n'est faite, par exemple, des questions rhétoriques, et l'on en sait finalement peu sur la diversité fonctionnelle des questions qu'il analyse.

Le problème de la définition et de la portée pragmatique des questions est esquivé davantage encore au niveau des études quantitatives traditionnelles. En effet, les études quantitatives analysées en 2.2 ne procèdent pas au tri entre questions catégoriques et questions variables, de sorte que la substitution des diverses structures interrogatives n'est pas un critère de sélection des questions. Il n'existe en réalité aucun critère de sélection défini, ou clairement exposé, qui serve à déterminer quels énoncés sont considérés comme des questions, et donc pris en compte dans l'analyse quantitative. On ignore comment sont traités les nombreux énoncés dont la valeur, interrogative ou assertive, est difficilement définissable (par exemple, les demandes de confirmation, fortement orientées vers le *oui* et indissociables de la structure SV). Le critère de sélection des études variationnistes, basé sur la possibilité de substitution des variantes, n'est certes pas toujours bien défini, mais il garantit une certaine uniformité dans les questions sélectionnées. On peut toutefois s'interroger sur la validité de ce critère, du fait qu'il dépend des jugements intuitifs du chercheur sur l'acceptabilité et l'équivalence d'énoncés créés artificiellement par substitution des variantes.

3.1.2 La dimension interprétative des études variationnistes

Au point 2.3, nous avons pu constater que Coveney et Elsig étaient amenés à émettre des jugements sur chaque token de leurs corpus. Les études variationnistes sur la question comportent donc une dimension interprétative de taille, sur laquelle repose une partie importante des résultats. Le jugement du chercheur est sollicité au moment de définir quelles questions peuvent être prises en compte comme variantes. Pour chaque token, le chercheur doit décider si l'utilisation d'autres variantes produit une question linguistiquement acceptable et fonctionnellement équivalente à la question énoncée originalement.⁴⁸

Au point 1, nous avons vu que le jugement du chercheur est toujours sollicité lorsque la variable (socio)linguistique est utilisée hors de la phonologie et qu'elle implique des unités porteuses de sens. Selon Sankoff, si le chercheur ne fait pas lui-même partie de la communauté linguistique qu'il analyse, il doit en avoir une connaissance suffisamment

⁴⁸ Si le chercheur décide de s'intéresser aux fonctions communicatives des questions et à leur influence sur la variation, il doit également juger de la valeur pragmatique des questions pour leur attribuer une fonction communicative.

approfondie pour pouvoir « déduire les intentions des locuteurs » (1988, 154, notre traduction). Dans le cas des questions, qui font alterner plusieurs structures syntaxiques et qui correspondent à la catégorie de variable la plus complexe, cette tâche peut être particulièrement délicate. Pour venir à bout de certains doutes interprétatifs, Coveney a demandé à deux locuteurs natifs de juger de l'équivalence de plusieurs structures interrogatives (toutes négatives). Il en est ressorti que les jugements des deux locuteurs natifs ne concordaient que dans la moitié des cas qui leur avaient été soumis. On comprend donc que certaines interprétations sont extrêmement difficiles, voire impossibles à résoudre de manière satisfaisante et indiscutable. Les exemples (65) et (66) ci-dessous, tirés de l'étude de Coveney, nous serviront à illustrer les difficultés qui peuvent apparaître au moment de juger l'équivalence des variantes.

Coveney considère que la question en (65) est un contexte semi-catégorique, en raison de l'indisponibilité de la variante SV. Il lui semble que l'usage de *ils sont bien arrivés ?* au lieu de *est-ce qu'ils sont bien arrivés ?* est incompatible avec la fonction communicative de la question, qu'il décrit comme étant citée et orientée positivement. Pourtant, sur la base du contexte (certes peu fourni) donné en (65), nous pensons que la structure SV pourrait servir à transmettre une question citée qui appelle une réponse en *oui* :

(65)...ils ont dû aller à l'usine et dire 'est-ce qu'ils sont bien arrivés ?' et tout ça euh / ... (Coveney, 2002, 201)

Dans l'exemple (66), Coveney considère de nouveau que la question (décrite comme une question rhétorique) correspond à un contexte semi-catégorique, où la variante SVK (*l'inconvénient est où ?*) est exclue :

(66)...y a certains professeurs qui euh / qui acceptent / euh que / qu les enfants à l'éc à l'école euh les tutoient euh / fin / j dis où est l'inconvénient si ? mais moi je n'accepte pas qu les enfants m tutoient. / ... (Coveney, 2002, 202)

L'emploi de SVK nous paraît effectivement inapproprié pour la question en (66), de même que l'emploi de toute autre variante que celle utilisée originalement. Nous aurions donc, *a priori* et de manière purement intuitive, classé la question en (66) comme catégorique. Notre but n'est pas ici de contredire les jugements de Coveney ni d'en proposer des améliorations. Coveney est évidemment mieux placé que nous pour juger, puisqu'il a participé lui-même aux entretiens et qu'il a accès à la totalité du contexte dans lequel apparaissent les questions, ainsi qu'à leurs caractéristiques prosodiques. Nous souhaitons simplement souligner que des difficultés (ou controverses) interprétatives peuvent facilement apparaître dans une étude variationniste non phonologique, *a fortiori* lorsque la variable est un objet aussi complexe que l'interrogation.

Certaines divergences interprétatives peuvent être observées entre l'étude de Coveney et celle d'Elsig : elles apparaissent clairement au niveau de la description des questions jugées catégoriques, c'est-à-dire dont la structure ne peut pas varier (voir point 2.3.3.2). Ainsi, Elsig classe d'emblée toutes les questions qui contiennent la particule *hein* comme catégoriques alors que Coveney considère que certaines d'entre elles acceptent un changement de structure. Elsig se montre également plus radical que Coveney dans son traitement des questions négatives, puisqu'il décide de toutes les exclure de son analyse. Coveney, lui, distingue entre questions négatives catégoriques et questions négatives variables.

Les incertitudes interprétatives semblent inévitablement liées à l'utilisation de la méthodologie variationniste dans le domaine des questions. Cela signifie que les chiffres

résultant de ces études sont toujours approximatifs et qu'ils devraient être considérés comme des indicateurs de certaines tendances plutôt que comme des valeurs exactes.

3.1.3 Mesurer les causes de la variation

Nous avons évoqué en 2.3.4.7 les phénomènes d'interactions qui peuvent apparaître lorsque l'on cherche à mesurer les influences de plusieurs facteurs sur la variation. Outre ces problèmes d'interactions, l'observation des divers facteurs susceptibles de conditionner la variation des formes interrogatives peut poser plusieurs difficultés d'ordre classificatoire. Avant de calculer statistiquement l'influence de tel ou tel facteur sur l'apparition des différentes formes de question, les chercheurs variationnistes doivent réaliser un travail important de catégorisation et d'identification. Ainsi, chaque facteur dont le chercheur suspecte une influence sur l'usage des questions doit être clairement défini et délimité, de sorte qu'il puisse ensuite être repéré dans tous les tokens où il entre en compte. Quelle que soit la nature des facteurs dont on veut mesurer l'influence, la question de savoir ce qu'on mesure et comment est primordiale et parfois problématique.

La mesure de l'influence des facteurs linguistiques est certainement celle qui pose le moins de difficultés. En effet, le système linguistique, composé de nombreuses unités discrètes, se prête plutôt bien à des découpages en différentes catégories. Ainsi, des facteurs tels que la nature du sujet, la longueur ou le temps du verbe semblent relativement bien définis et facilement identifiables dans un énoncé. Cependant, certains aspects de la phrase, notamment sémantiques, sont moins nettement mesurables. C'est le cas, par exemple, des verbes cognitifs, dont Elsig considère qu'ils influent sur le choix des structures de la question totale. Dans l'étude d'Elsig, les verbes cognitifs sont définis comme « dénotant des états ou des activités cognitives » (avec pour exemples *penser*, *trouver* et *comprendre*) (Elsig, 2009, 62). Sur la base de cette définition, il est difficile de connaître la portée précise de la classe des verbes dits cognitifs : on ne sait pas, par exemple, si elle inclut des verbes tels que *voir*, *regarder* et *entendre*⁴⁹. La classe des verbes cognitifs étant floue, sa portée peut facilement varier d'une étude à l'autre. On voit donc que certaines caractéristiques formelles de l'énoncé peuvent poser des problèmes classificatoires, même si ce n'est pas le cas de la plupart des facteurs linguistiques observés par Coveney et Elsig.

Les paramètres qui entrent en compte dans la caractérisation sociale des locuteurs peuvent être plus ou moins factuels et circonscrits, selon les cas. Si l'on observe le tableau 10 (point 2.3.4.6), qui représente les différents groupes de facteurs sociaux d'Elsig, on constate que seul le groupe de facteurs « profession » peut poser des problèmes de catégorisation. En effet, d'une manière générale, l'âge, le sexe et le domicile sont des données facilement définissables, sur la base desquelles on peut établir des catégories bien délimitées⁵⁰. En revanche, les différentes catégories de profession ne peuvent pas faire l'objet d'une division aussi nette. Depuis longtemps, l'établissement des classes socio-professionnelles ou socio-économiques fait l'objet de nombreuses réflexions en sociologie, et plusieurs systèmes différents ont été mis en place pour déterminer le statut social des locuteurs. Coveney explique que le concept de classe sociale est controversé, et que plusieurs études

⁴⁹ *Ecouter*, *entendre*, *voir* et *regarder* sont donnés par Gaétane Dostie (2004, 110-111) comme exemples de verbes de perception et de cognition.

⁵⁰ Des questions se posent tout de même au niveau de la création des différents groupes d'âge et des limites jugées pertinentes pour les distinguer. Selon Virginie Quillard, les sociologues considèrent aujourd'hui que la limite entre la jeunesse (caractérisée par une instabilité professionnelle et personnelle) et l'âge adulte (caractérisée par la stabilité) se situe autour des 35 ans (Quillard, 2001, 67).

sociolinguistiques prennent en compte différents groupes de facteurs pour assigner une classe sociale aux locuteurs (Coveney, 2002, 18). Cependant, il ne s'appuie que sur la profession pour établir les trois classes sociales à l'intérieur desquelles il range ses informateurs. Elsig se base également sur la profession de ses informateurs, mais il mentionne aussi leur lieu de domicile comme indicateur de statut social (toutefois, ces deux aspects sont mesurés séparément).

Les facteurs stylistiques sont souvent difficiles à classer, à moins qu'ils ne concernent un critère clairement identifiable, tel l'identité de la personne à qui s'adresse le locuteur (c'est un facteur analysé par Coveney). Nous avons déjà mentionné les difficultés liées à l'établissement de différentes catégories de styles ou de contextes stylistiques au point 1.2.3. L'arbre de décision de Labov, dont nous avons reproduit le schéma au point 2.3.4.6, se base sur l'assomption que le type de sujet abordé et l'identité de l'interlocuteur suffisent à déterminer le degré de formalité du discours. Or, les limites d'une telle approche sont évidentes : en effet, il semble peu probable que seuls les contextes à droite du tableau fassent apparaître un langage relâché et encore moins que cet effet opère systématiquement (par exemple, que les récits soient toujours liés à un style relâché). Ceci dit, les facteurs liés au style occupent une place secondaire dans les études de Coveney et d'Elsig, du fait que les contextes stylistiques d'où proviennent leurs données varient peu.

Finalement, les nombreux et divers effets regroupés sous le nom de «facteurs pragmatiques» constituent probablement l'aspect de la variation le plus difficile à identifier. Le compte rendu du système mis en place par Coveney pour déterminer les fonctions communicatives des questions (point 2.3.4.3) nous a donné un aperçu de la complexité de cette tâche. Pour créer son système de traits distinctifs destiné à attribuer des fonctions communicatives aux questions de son corpus, Coveney s'est inspiré de trois sources d'analyses différentes, liées à différents aspects de la pragmatique des questions. Il existe en effet plusieurs manières d'aborder le sens ou la portée pragmatique particulière d'un énoncé. Dans tous les cas, il faut mettre en rapport le contenu propositionnel de l'énoncé avec son contexte d'énonciation, en sachant que ce dernier peut être caractérisé par une multitude d'éléments différents (croyances et connaissances des interlocuteurs, énoncés qui précèdent, environnement physique, etc.). On pourrait donc imaginer de nombreuses alternatives au système mis en place par Coveney. Ce système nous semble toutefois assez complet, puisqu'il recouvre un champ pragmatique large (des actes tels que l'offre ou l'assertion y sont inclus), tout en permettant une description détaillée des particularités de certaines questions.

Nous l'avons vu au point 2.3.4.7, certains aspects pragmatiques de nature impondérable ne peuvent faire l'objet d'aucune mesure objective. Mais au-delà de ces phénomènes, il reste que certaines relations entre le contexte pragmatique et les structures interrogatives ne sont pas saisissables à travers la définition de fonction communicative de Coveney. Pour preuve, l'existence d'une relation entre la nature de l'information (connue *versus* nouvelle) contenue dans le groupe sujet-verbe-complément(s) et l'emploi de la structure SVK. Parmi les différents paramètres qui servent à définir la fonction communicative des questions dans l'étude de Coveney, aucun ne concerne le caractère nouveau ou connu de l'information contenue dans le syntagme verbal. C'est en manipulant ses données, pour chercher à expliquer l'inadéquation de certaines structures dans certains contextes, que Coveney a découvert qu'une information déjà donnée ou connue favorisait l'utilisation de SVK. L'émergence de ce type de découvertes montre que la définition de fonction communicative élaborée par Coveney, aussi précise et complexe soit-elle, ne suffit pas à traquer toutes les relations qui peuvent exister entre les éléments du contexte

pragmatique et les structures interrogatives. Nous verrons plus loin comment la notion de stratégie peut également servir à explorer ces relations.

Une autre difficulté méthodologique liée à la prise en compte de facteurs pragmatiques concerne la taille des corpus et le nombre de tokens à analyser. Après avoir éliminé les questions catégoriques de son corpus, Coveney a gardé un total de 220 questions (totales et partielles confondues). Il a ensuite analysé chaque token un à un pour en déterminer la fonction communicative. Il faut rappeler que Coveney a lui-même participé à tous les entretiens qui constituent son corpus, et qu'il conseille d'en faire de même à tout chercheur désireux d'analyser l'influence de facteurs pragmatiques sur la variation. Cette manière de faire semble difficilement applicable sur un nombre très élevé de tokens. Or, à plusieurs reprises dans son livre, Coveney regrette que sa faible quantité de questions ne lui permette pas de constituer des preuves quantitatives convaincantes. La conciliation d'une vaste étude quantitative avec une analyse précise de la valeur pragmatique de chaque token demande donc de grands moyens, qui ne sont peut-être pas à la portée d'un chercheur travaillant seul. Contrairement à Coveney, Elsig a travaillé sur la transcription d'un corpus qu'il n'a pas composé lui-même. La taille de ce corpus et le nombre de questions qu'il analyse sont plus élevés que dans l'étude de Coveney. Cette situation représente un avantage sur le plan quantitatif, mais elle complique sérieusement la tâche de l'attribution des fonctions communicatives, à cause du nombre élevé de tokens, et surtout, du manque de familiarité avec les énoncés étudiés. En effet, Elsig n'a pas participé aux enregistrements du corpus sur lequel il travaille, et il n'en utilise vraisemblablement que la version écrite. Dans cette situation, l'interprétation précise de la valeur sémantico-pragmatique poserait certainement plus de problème que dans la situation de Coveney.

3. 2 Objections au traitement variationniste des questions

Comme nous l'avons vu dans la première partie de ce travail, les objections liées à l'utilisation de la variable sociolinguistique en syntaxe relèvent principalement de deux types: elles concernent soit le principe d'interchangeabilité et d'équivalence des variantes, soit le recours aux facteurs sociaux (plutôt qu'aux facteurs pragmatiques) pour expliquer l'usage des variantes. Nous ne reviendrons pas sur ce deuxième type d'objections, d'une part parce qu'elles ont été moins souvent exprimées que les premières, et d'autre part, parce que le conditionnement social des variantes n'est pas au cœur des études de Coveney et d'Elsig. C'est donc essentiellement sur la première objection, liée à l'équivalence des variantes, que nous nous pencherons, pour en examiner l'application au niveau des structures interrogatives.

3.2.1 Les particularités de chaque forme

Plusieurs auteurs s'appliquent à souligner des différences de sens ou d'usage entre les différentes formes de la question. Comme le rappelle Gadet, l'interrogation, avec ses nombreuses structures, « pose crucialement le problème de la signification à attribuer à la diversité de formes possibles » (Gadet, 1997a, 105). Selon elle, la grande diversité formelle des questions ne gagne pas à être traitée dans une optique variationniste, car « différentes remarques suggèrent que le recours à l'une ou l'autre (forme) n'est pas aléatoire » (Gadet, 1997a, 113). Ces « remarques » concernent, par exemple, le fait que certaines questions totales agissant comme demandes de confirmation ne peuvent être exprimées par une structure autre que SV, ou encore, le fait que l'usage des différentes structures peut changer

en fonction du contexte interactionnel.⁵¹ Plutôt que de considérer les similitudes d'usage et de fonctionnement des différentes structures interrogatives, Gadet préfère en souligner les « capacités d'opposition ». Ainsi, elle considère que l'usage des différentes formes peut être « modulé » par le locuteur, en fonction de différents paramètres.⁵²

Les auteurs d'études variationnistes ne nient pas l'existence de rapports exclusifs entre certaines formes interrogatives et certaines significations ou valeurs pragmatiques : nous l'avons vu, les questions dont la forme ne peut être changée sans entraîner de modification d'acceptabilité ou d'équivalence sont classées comme catégoriques, et donc exclues de l'analyse. C'est uniquement après ce tri, à l'intérieur des contextes définis comme *variables* que l'équivalence des différentes formes interrogatives est proclamée. Or, à ce niveau, on peut encore s'interroger sur la comparabilité des différentes structures. En effet, les différentes structures de l'interrogation sont autant de manières différentes de coder l'information et de conférer une valeur plus ou moins interrogative à un énoncé. Dans ce passage, Berrendonner souligne la spécificité (sémantico-pragmatique) des trois structures principales de la question totale, en expliquant par quelles opérations respectives elles permettent de faire passer une question dans la mémoire discursive des interlocuteurs⁵³ :

L'une [SV] manifeste une intention primaire verbalement, et une posture secondaire gestuellement ; l'autre dénote le couplage des deux à l'aide de l'opérateur *est-ce que* ; la troisième l'exprime indirectement par une marque d'enchâssement spécifique. Les trois ne sont donc pas synonymes. (Berrendonner, 2005, 16)

Alors que les recherches variationnistes reposent sur le principe d'équivalence, certaines études s'attachent à déceler les nuances intrinsèques qui caractérisent chaque forme interrogative et qui font qu'elles ne sont pas synonymes.

Dans un article intitulé « La position du sujet dans l'interrogation directe », Nicole Le Querler étudie les particularités des questions à inversion,⁵⁴ en comparaison aux questions à ordre SV. Sur la base d'un corpus journalistique, elle a établi deux différences majeures dans l'emploi de ces deux structures. Premièrement, elle a remarqué que la postposition du sujet favorisait une interprétation de l'énoncé comme une demande d'information, alors que l'antéposition du sujet orientait plutôt l'interprétation de l'énoncé vers une assertion, « modalisée ou non » (Le Querler, 1994, 158). Deuxièmement, elle a noté que les structures SV et VS pouvaient jouer un rôle dans la portée attribuée aux mots interrogatifs *pourquoi* et *comment*. Elle donne (67a) (forme authentique provenant de son corpus) et (67b) (forme remaniée) pour montrer l'effet de la position du sujet sur l'interprétation du sens de *comment* :

(67a) Comment vos ouvriers, ici, à Wolfsburg, acceptent que vous en licenciiez 25 000, alors que vous investissez 15 milliards de francs à l'Est ?

(67b) Comment acceptent vos ouvriers, ici, à Wolfsburg, que vous en licenciiez 25 000, alors que vous investissez 15 milliards de francs à l'Est ? (Le Querler, 1994, 160)

⁵¹ Gadet mentionne que lors d'un entretien chez une femme issue de la classe populaire, la sonnette de la maison retentit et entraîna les deux énoncés suivants, de la part de la femme interviewée : le premier, *c'est qui ?*, adressé à voix basse à sa fille, et le second, *qui est-ce ?*, adressé plus fort à la personne qui avait sonné.

⁵² La nature précise de ces paramètres n'est pas explicitée par Gadet.

⁵³ Voir point 1.3 pour le concept de « mémoire discursive » et l'explication du fonctionnement des questions donnée par Berrendonner.

⁵⁴ Plutôt que de questions à inversion, LeQuerler préfère parler de questions à postposition du sujet.

Selon Le Querler, le sujet postposé en (67b) limite l'interprétation de *comment* à sa portée intra-prédicative, (où *comment* signifie « de quelle manière »), tandis que le sujet antéposé en (67a) permet une interprétation intra-prédicative aussi bien qu'une interprétation extra-prédicative (où *comment* signifie « comment se fait-il que »). Elle observe que la position du sujet entraîne un conditionnement similaire au niveau de l'interprétation de la portée (intra ou extra-prédicative) de *pourquoi*.⁵⁵

La particule *est-ce que* est également liée à certaines spécificités sémantico-pragmatiques. Selon Anne-Marie Santin-Guettier, *est-ce que* est un « agencement » particulier, qui ne devrait pas être considéré comme un « simple outil interrogatif », ceci pour plusieurs raisons (1994, 188). D'abord, Santin-Guettier souligne les particularités des opérations qui mènent de l'assertion *il prend le train* à la question *est-ce qu'il prend le train*. Elle donne un compte rendu détaillé des relations syntaxiques et sémantiques qui unissent les trois composants *est / ce / que* entre eux et avec la proposition située à droite de *que*. En résumé, Santin-Guettier remarque que plusieurs relations métalinguistiques apparaissent entre les différents éléments d'une question en *est-ce que*, de sorte que « tout énoncé amorcé par *est-ce que* revient à la fois à poser p [la proposition, ici *il prend le train*] et à confronter p à sa propre remise en question » (1994, 192). En plus de la complexité des opérations à la base d'une question en *est-ce que*, Santin-Guettier mentionne l'évolution diachronique de la particule comme un gage supplémentaire de sa spécificité. Si *est-ce que* est aujourd'hui une particule (pratiquement) figée⁵⁶, ce n'était pas le cas avant le XII^e siècle, qui marque, selon Renchon, les débuts probables de la grammaticalisation de la tournure (Renchon, 1967, 151-152). Avant le figement d'*est-ce que*, le verbe *être* avait une valeur sémantique propre, il pouvait varier en temps et divers éléments pouvaient apparaître entre *est-ce* et *que*. La tournure avait alors une valeur emphatique. Les siècles qui suivirent le XII^e virent la particule se figer peu à peu : des variations dans le temps du verbe *être* peuvent encore être observée au XVI^e siècle.

(68) Jusques à quand sera-ce que tu abuseras de la jeunesse de nostre Roy ? (Hotman, *Epître envoyée au Tigre de France* <Renchon, 1967, 151>)

Si Foulet considère, en 1921, que la particule *est-ce que* « s'est vidée de toute signification particulière » (Renchon, 1967, 155), Renchon n'en est pas aussi convaincu :

On peut cependant se demander si la grammaticalisation de la forme périphrastique a été complète et irréversible, la périphrase n'étant plus susceptible d'un sens plein. (Renchon, 1967, 154)

Santin-Guettier, elle, exprime clairement son désaccord face aux propos de Foulet :

Qualifier *est* de « syllabe » ayant perdu « toute attache avec le verbe *être* » revient à gommer l'existence de la mémoire linguistique que gardent les mots. (Santin-Guettier, 1994, 188)

⁵⁵ Il est intéressant de constater que l'existence d'un lien entre le sens de *pourquoi* et la structure interrogative de la phrase dans laquelle il apparaît est également mentionnée par Coveney. Ce dernier remarque que dans l'exemple « c'était une colonie pour quoi alors ça ? », seule la variante SVK est disponible, les autres variantes entraînant une interprétation de *pourquoi* dans le sens « pour quelle raison » (Coveney, 2002, 186).

⁵⁶ De courts éléments peuvent toutefois s'insérer, dans certains cas, entre *est-ce* et *que* (cf. « Est-ce **donc** qu'il est malhonnête d'être le père de ses enfants ? » Marivaux, *Le paysan parvenu* <Renchon, 1967, 153>)

Ces différentes remarques montrent que les structures interrogatives en concurrence comportent toutes des spécificités, au niveau syntaxique bien sûr, mais aussi, au niveau de leurs propriétés ou fonctionnement sémantico-pragmatiques et de leur évolution diachronique. Ces spécificités, dont nous n'avons donné ici qu'un aperçu, constituent une entrave au principe d'équivalence des variantes sur lequel reposent les approches variationnistes. Les différentes questions soulevées dans le débat sur l'existence du concept de variable (socio)linguistique en syntaxe semblent donc particulièrement pertinentes dans le domaine des structures interrogatives. Pour tenter de les résoudre au niveau de l'interrogation, il nous semble important de revenir sur la portée de la notion d'équivalence et les principes qui la sous-tendent. En effet, l'*équivalence* proclamée entre les variantes de la question est une notion discutable, dont la méthodologie variationniste use de manière contradictoire.

3.2.2 La notion d'équivalence

Le but des études de Coveney et d'Elsig est de déterminer quels facteurs exercent une influence sur l'usage des structures interrogatives en français oral non surveillé. Les deux études arrivent à la conclusion que les différentes structures en concurrence n'ont pas les mêmes probabilités d'apparition dans tous les contextes, puisque la présence ou l'absence de certains facteurs façonnent leurs schémas d'usage. En même temps, les chercheurs variationnistes affirment que toutes les structures interrogatives qu'ils analysent sont interchangeables et équivalentes. Il y a donc une contradiction apparente dans cette approche qui s'attache, d'une part, à définir sur quels critères on peut expliquer les usages de chaque forme, tout en affirmant d'autre part que toutes les formes sont équivalentes. En effet, si l'usage d'une forme résulte en partie des caractéristiques linguistiques de la phrase, des caractéristiques sociales du locuteur et/ou des aspects pragmatiques de l'énoncé et de l'énonciation, comment peut-on avancer que son remplacement par une autre forme variante entraîne un énoncé équivalent ? Cette question est d'autant plus importante au niveau de l'influence des facteurs pragmatiques : si une forme interrogative est (même subtilement) liée à un aspect sémantico-pragmatique de l'énoncé, on doit en conclure que son utilisation transmet des nuances de sens qu'une autre forme ne transmettrait pas.

La notion d'équivalence sémantique est donc toute relative et les propres résultats des études variationnistes permettent d'en mesurer les limites. Le terme *équivalence*, ancré dans la tradition variationniste, n'est guère approprié pour des variantes syntaxiquement complexes telles que les questions : il suggère une identité et une symétrie auxquelles ne peuvent prétendre les différentes structures de l'interrogation. Toutefois, ce constat ne nous semble pas suffisant pour pénaliser d'emblée toute approche variationniste des questions ; en effet, l'étude de Coveney montre que la méthodologie variationniste peut s'adapter aux complexités syntaxiques et sémantico-pragmatiques des questions.

La variable syntaxique, telle que la conçoit Coveney, est loin d'être une transposition pure et simple de la variable phonologique au domaine de la syntaxe. En fait, les modifications qu'il y a apportées répondent aux principales critiques des opposants à la variable syntaxique. L'objection selon laquelle les variables syntaxiques ne peuvent être conditionnées par des facteurs sociaux en raison de la primauté de l'influence des facteurs pragmatiques ne touche pratiquement pas l'étude de Coveney, puisqu'elle traite davantage des facteurs pragmatiques que des facteurs sociaux. Reste l'opposition basée sur l'impossibilité d'équivalence entre les variantes : Coveney dit restreindre l'équivalence au sens descriptif et fonctionnel des variantes. Si le sens descriptif semble relativement stable d'une structure interrogative à l'autre, il est difficile de garantir une équivalence fonctionnelle totale entre les structures interrogatives. Comme nous l'avons souligné, Coveney lui-même signale les

particularités pragmatiques des variantes de l'interrogation, et donc leur non-équivalence. En fait, la notion d'équivalence (qui gagnerait à être désignée autrement) a sa raison d'être dans la méthodologie variationniste dans la mesure où elle permet de dissocier deux types de questions : celles dites catégoriques et celles dites variables. Pour ceux qui cherchent à comprendre l'usage des structures interrogatives, les premières sont tout aussi intéressantes que les secondes. En effet, les questions catégoriques présentent une structure syntaxique indissociablement liée à un certain aspect (linguistique ou sémantico-pragmatique) de l'énoncé dans lequel elles apparaissent. En revanche les liens qui unissent les structures des questions variables à certains aspects intra et extra-linguistiques de l'énoncé sont moins forts. Le critère d'équivalence permet ainsi de distinguer différents types de question, et différents degrés d'affinités entre certaines structures et certaines caractéristiques intra et extra-linguistiques. De plus, il permet de résoudre les problèmes d'interchangeabilité syntaxique qui touchent les études quantitatives traditionnelles. En effet, nous avons montré au point 2.2.2.2 que la classification des énoncés interrogatifs en vue d'une quantification présentait plusieurs difficultés pour les auteurs d'études quantitatives. Le recours au critère d'équivalence prévient ces difficultés en éliminant, entre autres, les questions qui ne sont pas syntaxiquement interchangeables.

3.3 Applications du cadre théorique variationniste aux questions

Tout au long de ce travail, nous avons utilisé les termes « variable » et « variationnisme » pour référer à des méthodes d'analyses formées sur la base ou à la suite du concept de *variable phonologique* créé par Labov. Or, le paradigme labovien a connu plusieurs changements depuis son apparition dans les années soixante, de sorte qu'il connaît aujourd'hui plusieurs orientations. Ainsi, l'étude dite variationniste de Coveney se différencie en de nombreux points des premières études variationnistes menées par Labov en phonologie. Alors que ces dernières se bornaient à souligner les effets de la stratification sociale sur l'usage de la langue et son rôle dans le changement linguistique, Coveney s'applique principalement à expliquer les variations de formes syntaxiques par des facteurs linguistiques et pragmatiques. Aujourd'hui, on peut donc se demander si le terme de « variable » est toujours approprié, s'il doit être suivi de l'adjectif linguistique ou sociolinguistique, et s'il s'agit d'un concept ou d'un outil analytique. Autant de questions auxquelles il est difficile de répondre étant donné le flottement qui règne autour de cette notion. Il semble donc nécessaire que chaque auteur définisse clairement sa terminologie et les principes qu'il y associe. C'est ce que font Coveney et Elsig dans leurs études qu'ils qualifient tous deux de « variationnistes ».

Nous avons déjà rendu compte au point 2.3.3.1 de la description précise que Coveney donne des termes « variable » (concept forcément lié à des différenciations sociales) et « alternance » (outil analytique). Elsig distingue pour sa part la variable linguistique et la variable sociolinguistique, qui ne se différencient que par le fait que la seconde est conditionnée par des facteurs sociaux. On ne retrouve pas chez Elsig la distinction entre *concept* et *outil* soulignée par Coveney pour différencier les variables des alternances. On peut s'interroger cependant sur la pertinence de cette distinction, au vu des résultats et de l'orientation générale de l'étude de Coveney. En effet, même si Coveney affirme que les questions correspondent bien à la définition qu'il donne d'une variable, l'importance des facteurs sociaux est moindre dans son étude, tant dans leur influence sur la variation des questions, que dans la manière dont ils sont traités. Ainsi, Coveney a-t-il recruté des informateurs qui ne représentent pas toutes les tranches d'âge, ni toutes les classes sociales (seules trois informatrices sont âgées de plus de 37 ans et seuls deux informateurs sont issus de la classe ouvrière). Il apparaît donc clairement, dès le départ, que l'intérêt pour le

conditionnement social de la variation et la compréhension des différentes communautés linguistiques n'est pas au centre de l'étude de Coveney. Cette impression se confirme lorsque l'on observe les résultats de l'étude, peu développés au niveau de l'influence des facteurs sociaux : si quelques corrélations sont signalées au niveau de la question totale, le conditionnement social de l'usage de la question partielle n'est pas clair. Dans ce contexte, il peut sembler superficiel de vouloir rattacher les questions au concept de variable telle qu'il est défini par Coveney (c'est-à-dire comme un témoin de l'influence des caractéristiques sociales sur l'usage de la langue).

Si Coveney et Elsig s'entendent grossièrement sur la définition théorique de la variable sociolinguistique, la méthodologie par laquelle ils tentent d'expliquer l'usage des questions présente des divergences. Au point 1, nous avons vu que l'extension de la variable (socio)linguistique hors de la phonologie avait engendré des réflexions sur la nécessité ou l'utilité de mesurer l'effet de facteurs pragmatiques sur l'usage des variantes. De ce point de vue, les études de Coveney et d'Elsig correspondent à deux versants différents de l'usage de la variable. Celle d'Elsig est plus traditionnelle puisqu'elle se limite à mesurer les effets des facteurs socio-stylistiques et linguistiques, comme c'était le cas des premières études recourant au concept de variable.⁵⁷ En revanche, dans l'étude de Coveney, l'analyse de l'influence des facteurs pragmatiques sur les variantes de la question est placée au premier plan (conjointement avec l'influence des facteurs linguistiques) tandis que l'analyse des facteurs sociaux et stylistiques n'occupe qu'une place secondaire. Les deux études se différencient également par les systèmes d'analyses statistiques utilisés pour mesurer la variation. Face à ces divergences, le critère d'équivalence et d'interchangeabilité des variantes apparaît comme une constante d'une étude à l'autre : il forme, semble-t-il, l'unité principale des études issues du paradigme variationniste labovien. Ces dernières se distinguent ainsi d'autres études dont les méthodes quantitatives ou statistiques peuvent être proches, mais dont les exigences d'équivalence diffèrent de celles des études variationnistes.

C'est le cas des études quantitatives sur l'interrogation abordées au point 2.2. Du fait qu'elles ne se basent pas sur l'équivalence des structures interrogatives, les études quantitatives n'excluent (pratiquement) pas de questions de leurs études. Par conséquent, l'étendue des questions qu'elles prennent en compte est bien plus large que dans les études de Coveney et d'Elsig⁵⁸ (rappelons que ce dernier a écarté plus de la moitié des questions totales qu'il avait au départ). Nous avons mentionné plus haut l'intérêt de prendre en compte, pour la compréhension de l'usage de l'interrogation, aussi bien les structures variables que les structures catégoriques (c'est-à-dire invariables). Or, Coveney et Elsig ne réservent pas le même traitement aux questions jugées catégoriques : s'ils en donnent tous deux une description générale, seul Coveney en tire des conclusions plus précises sur le fonctionnement des structures interrogatives. Ainsi, à la différence d'Elsig, Coveney rend compte de manière détaillée de toutes les caractéristiques (linguistiques ou fonctionnelles) qui contraignent l'usage de l'une ou l'autre structure interrogative. De ce fait, les facteurs à effet contraignant sont considérés avec la même attention que les facteurs à effet variable, de même que les questions exclues pour leur non variabilité sont prises en compte de la même manière que les questions variables.

⁵⁷ Cependant, l'étude d'Elsig se différencie des premières études variationnistes dans l'importance qu'elle accorde aux facteurs linguistiques agissant sur la variation, et à la grammaire des formes interrogatives.

⁵⁸ Toutefois, nous l'avons vu, le problème de la définition des questions se pose d'autant plus crucialement que l'équivalence entre les différentes formes n'est pas requise.

D'une manière générale, on remarque donc que l'approche variationniste de Coveney est certainement plus adaptée à l'étude des structures interrogatives que celle d'Elsig. Par sa prise en compte des causes pragmatiques de la variation, par son étude qualitative minutieuse de toutes les questions de son corpus et sa considération pour les questions invariables, Coveney offre une vision plus large et en même temps plus précise de l'explication de la variation des questions que celle donnée par Elsig. Mais il reste que la diversité des formes interrogatives et les modalités de leur utilisation est un sujet extrêmement vaste, qu'une approche telle que celle de Coveney ne suffit pas à expliquer exhaustivement.

3.4 Limites des études variationnistes et suggestions alternatives

En traitant des difficultés méthodologiques des études variationnistes, nous avons déjà évoqué certaines limites propres à l'utilisation de la variable (socio)linguistique en syntaxe, telles que le caractère approximatif des résultats, l'imprécision de la notion d'équivalence ou des catégories de facteurs agissant sur la variation. Les limites que nous abordons à présent ont trait à la portée des études variationnistes dans le domaine de la question : elles apparaissent aussi bien dans l'étendue des questions prises en compte que dans les perspectives selon lesquelles elles sont observées. A ces limites, nous proposerons des alternatives qui ne sont probablement pas toutes intégrables dans une étude variationniste, mais qui en constituerait certainement des compléments utiles pour une compréhension globale de l'usage des diverses formes interrogatives.

3.4.1 Les structures non traitées

Les études variationnistes dont nous avons rendu compte dans ce travail concentrent leurs observations sur l'usage des formes canoniques de l'interrogation.⁵⁹ En effet, ces études établissent au départ les différents types de structures (ou variantes) qu'elles veulent quantifier. De ce fait, seules les structures quantitativement significantes sont prises en compte, c'est-à-dire, celles dont on considère généralement qu'elles constituent des questions correctement énoncées, ou du moins acceptables. Toutes les questions dont la structure ne correspond à aucune des variantes établies sont écartées de l'analyse. Ainsi, les exemples (69) à (71), attestés, ne peuvent être placés dans aucune des catégories de variantes établies par les études variationnistes et quantitatives que nous avons analysées :

(69) quoi est signe de quoi ? (Gadet, 1997a, 108)

(70) pourquoi est-ce que le locuteur accumule-t-il ainsi les formules ? (*ibid.*, question écrite)

(71) - et il s'appelle ? (*ibid.*, 112)

- Grégory

Berrendonner qualifie de « marginales » les variantes (ou lectes) dont l'utilisation n'est pas centrale, et qui sont souvent associées à des erreurs ou à des imperfections de production (Berrendonner, 1988, 22). Si les structures interrogatives marginales sont insignifiantes pour les études de nature quantitative en raison de leur basse fréquence, elles font partie du système de l'interrogation et méritent qu'on s'y attarde. En effet, Berrendonner explique que l'observation des lectes marginaux est indispensable à la connaissance des propriétés du matériau variationnel (ici les questions) à disposition du locuteur. Ainsi, la prise en compte de structures interrogatives peu utilisées mais attestées pourrait amener une meilleure compréhension des utilisations possibles du système interrogatif en français.

⁵⁹ La même observation peut être faite au sujet des études quantitatives présentées en 2.2.

Un autre effet lié à l'établissement de variantes canoniques est le peu de précision dans la caractérisation des structures interrogatives qui varient. Les structures interrogatives variantes données dans les tableaux 1 et 2 (études quantitatives) puis 4 et 5 (études variationnistes) sont essentiellement établies en fonction de l'ordre des différents constituants de la phrase. Leur description n'est guère détaillée, de sorte que certaines distinctions structurales des questions n'y sont pas représentées. Ainsi, les différentes catégories de variantes ne permettent pas de distinguer les questions à sujet disloqué (voir point 2.2.2.2) des autres questions : malgré la présence d'un sujet nominal disloqué à gauche du verbe, Coveney a compté l'exemple (72) comme une occurrence de la variante SV :

(72) il est connu *Brassens* euh au en Angleterre ? (Coveney, 2002, 94)

Coveney justifie ce procédé par le fait que les phénomènes de dislocation ne sont pas spécifiques aux questions et doivent être étudiés indépendamment ; de plus, il dit n'avoir rencontré qu'un nombre très restreint de questions à sujets disloqués dans son corpus (mais il ne donne aucun chiffre). Elsig, lui, rapporte les chiffres suivants : sur les 64 questions totales à sujet nominal de son corpus, 11 sont doublées d'un pronom sujet alors que pour les questions partielles, la proportion est de 33 sur 50. Au point 2.2, nous avons vu que les phénomènes de dislocation touchaient une grande partie des questions à sujet nominal issues des corpus de Söll et de Behnstedt. Plusieurs chercheurs signalent donc la dislocation du sujet comme un phénomène touchant fréquemment les structures interrogatives à sujet nominal. Or, la décision des chercheurs variationnistes d'assimiler les structures à sujets disloqués aux structures à sujet simple est discutable. Prenons la question (73a), issue de l'étude de Behnstedt, et (73b), sa transformation en une structure KESV à sujet nominal simple :

(73a) Comment est-ce qu'elles peuvent savoir, les autorités ? (Al, 1975)

(73b) Comment est-ce que les autorités peuvent savoir ?

Dans les études de Coveney et Elsig, (73a) et (73b) auraient été considérés comme appartenant à la même variante, soit KESV. Or, les différences entre (73a) et (73b) sont notables à plusieurs niveaux. Premièrement, les deux énoncés ne présentent pas l'information de la même manière, puisqu'en (73a), le sujet dupliqué encadre le groupe verbal, mais le référent du pronom *elles* n'est donné qu'après le verbe : on peut donc supposer que chacune des deux structures présente ses propres affinités avec certains contextes pragmatiques. Deuxièmement, le statut socio-stylistique des deux questions n'est probablement pas égal, de sorte que les contextes stylistiques dans lesquelles elles apparaissent ne sont pas forcément les mêmes. Ces observations montrent que l'assimilation des structures interrogatives à sujet disloqué aux structures interrogatives à sujet simple n'est pas entièrement justifiée.

La particule *est-ce que* et ses différents « concurrents » ne sont pas non plus systématiquement différenciés les uns des autres dans les calculs des études quantitatives et variationnistes que nous avons analysées. La richesse de ces formes, attestée dans plusieurs corpus de ces mêmes études, fait apparaître une grande diversité à l'intérieur même d'une des structures canoniques de l'interrogation (KESV). Là encore, l'observation et la confrontation d'exemples montre que l'usage des différentes particules est lié à différentes caractéristiques sémantico-pragmatiques et socio-stylistiques :

(74a) Il dit qui c'est que c'est qui frappe sur la maison comme ça, bien je dis c'est eux-autres au côté tu sais, qui nous donnent du trouble, hein ? (Elsig, 2009, 150)

(74b) Il dit qui est-ce qui frappe sur la maison comme ça (...) ?

Il est évident que (74a) présente un degré d'emphasis plus important que (74b) et que leur statut socio-stylistique n'est pas comparable. On gagnerait donc certainement à différencier les nombreuses particules rivales d'*est-ce que* pour analyser les fréquences et les conditions d'utilisation propres à chacune. Dans ce cas comme dans celui des sujets disloqués, on remarque que l'établissement de variantes plus spécifiques pourrait contribuer à rendre des études variationnistes sur les questions plus complètes et pertinentes.

Enfin, il faut encore mentionner comme structures non traitées, les questions qui sont exclues à cause de leur préférence (marquée ou exclusive) pour l'une ou l'autre structure interrogative : c'est le cas par exemple des questions échos ou des questions *tags* (appelées « questions avec appendice » par Borillo). Ces dernières, très fréquentes dans le langage oral, sont systématiquement exclues des études de Coveney et d'Elsig.

3.4.2 La notion de « stratégie »

Coveney a établi que l'usage de la structure SVK dans les questions partielles est conditionné, du moins en partie, par la nature de l'information transmise par le sujet, le verbe et le(s) complément(s) éventuel(s). Selon lui, cette contrainte pragmatique sur la structure SVK permet aux locuteurs (au moins dans certains cas) de choisir avec quelle intensité ils souhaitent présupposer certaines informations (Coveney, 2002, 226). Par cette remarque, Coveney évoque le potentiel stratégique de la variation, tel qu'il est défini par exemple par Berrendonner (voir point 1.3.1). Les études variationnistes montrent que de nombreux facteurs peuvent influencer sur l'emploi des structures interrogatives, mais elles ne font pas (ou peu) cas de l'exploitation possible de cette diversité par les locuteurs. Il nous semble pourtant que la diversité des formes interrogatives peut facilement donner lieu à différentes stratégies discursives. Un autre exemple trouvé chez Coveney est la réticence de la variante KSV à apparaître avec le mot interrogatif *quand*. Ce phénomène s'explique, selon Coveney, par une « stratégie perceptuelle » mise en place pour éviter la confusion entre une phrase subordonnée affirmative et une interrogation (Coveney, 2002, 229). Il explique en effet que *quand* peut être utilisé comme subordonnant dans des phrases d'ordre *quand*+S+V, ces dernières précédant souvent la phrase principale dont elles dépendent, comme c'est le cas en (75) :

(75) En général, quand la première face ne marche pas, la seconde désarme l'adversaire.
(Pennac, *La fée carabine*, 268)

Le fait que les auditeurs soient habitués à interpréter les énoncés en *quand*+S+V comme des phrases subordonnées, pourrait expliquer, selon Coveney, la faible fréquence des structures KSV lorsque le mot interrogatif est *quand*.

Si l'on reprend la citation de Berrendonner donnée au point 1.3.1, on voit que les stratégies qu'il appelle « lectales » servent des buts multiples tels que « remédier aux conflits opératifs », « prévenir les ambiguïtés » ou « favoriser la reconnaissance perceptive », (Berrendonner, 1987, 51). L'exploitation des diverses structures interrogatives pourrait donc se faire sur plusieurs plans. L'exemple de *quand*, donné plus haut, montre qu'une structure peut être préférée à une autre pour éviter des ambiguïtés interprétatives. Quant à l'exemple de la structure SVK, il montre que l'usage de telle ou telle structure peut traduire différentes attitudes du locuteur face à l'information contenue dans sa question. En fait, les spécificités de chaque structure interrogative font qu'elles peuvent chacune conférer différentes caractéristiques sémantiques, cognitives ou interactionnelles à la question énoncée. Une analyse de la variation des questions menée dans une perspective d'efficacité communicative pourrait certainement mettre en lumière de nombreux autres usages stratégiques liés à la diversité des formes interrogatives.

L'exploitation des diverses formes interrogatives par les locuteurs semble d'autant plus évidente que les questions revêtent souvent une importance majeure dans les interactions. Elles constituent un contexte « saillant », ou « proéminent », selon Cheshire, qui reprend des propos d'Ihalainen. Pour résumer, le concept de « saillance » ou « proéminence » décrit des environnements linguistiques où l'information est mise au premier plan et où l'engagement (*involvement*) du locuteur est fort (Cheshire, 1996). Selon une hypothèse d'Ihalainen, c'est à l'intérieur des contextes saillants (dont font partie la négation et l'interrogation) que les formes non standard sont le plus souvent maintenues. Ces « formes non standard » peuvent apparaître dans n'importe quelle partie du système linguistique ; la remarque d'Ihalainen ne concerne donc pas spécialement les structures syntaxiques des questions, mais elle suggère que les questions sont, d'une manière générale et par leur rôle dans l'interaction, des contextes où la variation est riche. Cette même idée sous-tend les propos de Gadet, lorsqu'elle déclare que « l'interrogation constituant un point d'enjeu dans les rapports interpersonnels, elle requiert souplesse et variété de formes » (Gadet, 1997b, 15). En effet, les questions sont des actes qui peuvent facilement cristalliser des rapports de force entre interlocuteurs. Comme l'explique Kerbrat-Orecchioni, la valeur taxémique de la question est ambivalente. D'une part, une question peut apparaître comme un ordre de répondre, qui place le locuteur dans une situation dominante et constitue un acte menaçant pour l'interlocuteur. Mais une question peut également apparaître comme l'aveu d'un manque et placer le locuteur dans une position d'infériorité par rapport à l'interlocuteur qui détient le savoir (Kerbrat-Orecchioni, 1991a, 28-29). Dans les deux cas, la question n'est pas un acte neutre pour les places et les faces des interlocuteurs. La diversité des formes interrogatives peut donc apparaître comme un moyen de modeler et de nuancer l'acte particulièrement menaçant qu'est la question.

3.4.3 Les questions dans différents registres de langue

Pour expliquer les conditions d'usage des différentes structures interrogatives, les grammaires recourent très souvent à la notion de style ou de registre. La remarque la plus constante parmi toutes les grammaires concerne la relation privilégiée entre la structure interrogative à inversion et la langue écrite. Il ne s'agit là pas tant d'une différence de style, que d'une différence de médium : les différences d'ordre diamésique sont donc placées en premier plan dans la différenciation des formes de la question. Au delà de cette première explication, les grammaires contiennent souvent plusieurs commentaires sur la valeur stylistique des structures interrogatives. Dans *Le Bon Usage*, on apprend par exemple que l'inversion est caractéristique de la « langue soignée », que la structure SVK apparaît dans « l'interrogation de type familier », tandis que KSV relève d'un « usage très familier » ou encore, que les différents tours concurrents d'*est-ce que* comme *c'est que* ou *ce que c'est que* sont « franchement populaires ». Si ces remarques restent très générales et guère exhaustives, elles mettent bien le doigt sur une facette importante de la variation des questions, à savoir, l'influence du contexte stylistique ou énonciatif sur le choix des structures interrogatives.

Plutôt que de parler de style ou contexte stylistique, Biber préfère le terme « registre », qu'il utilise pour décrire « toute variété associée à des contextes situationnels ou des buts particuliers » (Biber, 1995, 1, notre traduction). Selon lui, un registre se définit selon plusieurs dimensions, et ceci de manière graduelle : au lieu des dichotomies du type formel/informel, il préfère considérer que chaque registre a un certain degré de formalité, de planification, d'explicité, d'interactivité, etc. Biber précise que l'on peut définir des registres suivant différents niveaux de généralité. Par exemple, les écrits de fiction constituent un registre large tandis que les petites annonces immobilières correspondent à un registre plus spécifique. Selon lui, il est impossible de définir toutes les distinctions situationnelles sur la base d'un

critère unique (comme par exemple le critère de *l'attention portée à la parole* sur lequel se base Labov)(Biber, 1995, 36).

Dans leurs études sur les questions, Coveney et Elsig tirent tous deux leurs données d'un seul contexte situationnel, à savoir des entretiens menés dans un contexte le plus détendu possible. Ils souhaitent ainsi accéder au « français vernaculaire » des locuteurs, suivant la méthodologie variationniste traditionnelle.⁶⁰ Par conséquent, peu de distinctions stylistiques ou situationnelles apparaissent dans leurs études. Les seules distinctions qu'ils fournissent s'expriment en termes de formel/informel ou selon la nature de l'interlocuteur. Il semble pourtant que l'établissement de distinctions stylistiques et situationnelles plus diversifiées et subtiles, telles que les conçoit Biber par exemple, fournirait des résultats intéressants pour l'étude de l'usage des formes interrogatives.

A la différence de Coveney et d'Elsig, Behnstedt a pris en compte, dans son étude quantitative sur les questions, plusieurs contextes stylistiques différents. Les résultats des fréquences d'utilisation des formes interrogatives qu'il a obtenus varient passablement suivant qu'ils concernent, par exemple, des émissions radiophoniques, ou des conversations spontanées. Virginie Quillard mentionne aussi la diversité situationnelle comme une dimension importante de l'usage des diverses formes interrogatives. Elle a mené une étude sur l'utilisation des formes interrogatives dans cinq contextes situationnels différents, avec des locuteurs de différentes origines socio-professionnelles et d'âges variables. Selon elle, les explications pragmatiques sont essentielles dans la compréhension de l'usage des structures interrogatives, mais il lui semble nécessaire « de ne pas se limiter à cet unique courant et de ne pas perdre de vue les dimensions diastratiques et diaphasiques » (Quillard, 2001, 70). En étudiant le fonctionnement pragmatique de la structure SVK, elle a pu observer que cette structure était favorisée par les questions introductives⁶¹ mais défavorisée par les questions rhétoriques. Cependant, Quillard remarque que les conditions d'utilisation de cette structure varient beaucoup d'un contexte à l'autre : dans les débats politiques, 60% des structures SVK correspondent à des questions introductives, tandis que dans une réunion d'adolescents, seules 8,2% des structures SVK sont des questions introductives. L'explication pragmatique est donc pertinente pour rendre compte de l'usage des structures SVK dans le contexte de débats, mais elle est insatisfaisante dans le cas des conversations d'adolescents. C'est pourquoi Quillard suggère que l'observation du contexte stylistique, de l'âge et de la situation socio-professionnelle est aussi importante que les considérations pragmatiques pour expliquer l'emploi des différentes formes interrogatives. En outre, elle explique que la diversification des enregistrements lui a permis d'« obtenir le plus possible de formes interrogatives, des plus soutenues aux plus relâchées » (Quillard, 2001, 65). Les études variationnistes de Coveney et d'Elsig montrent en effet que la restriction à des données tirées d'entretiens ne permet d'accéder qu'à une partie des nombreuses structures interrogatives attestées par ailleurs.

⁶⁰ En effet, Labov a souvent insisté sur l'importance d'étudier le langage vernaculaire, parce qu'il présenterait moins d'hypercorrection et plus de systématisme qu'une variété de langue influencée par des formes standard (*Sociolinguistique*, 189).

⁶¹ Qui semblent correspondre à la catégorie de questions que Coveney divise en *sub-topic-introducing* et *pre-announcements*.

Ces quelques commentaires montrent que les études variationnistes sur les questions pourraient être complétées ou élargies à plusieurs niveaux. Sur le plan du matériau linguistique qui varie, les études variationnistes se limitent à certaines structures et ne rendent pas compte de toute la complexité et l'étendue de la diversité formelle des questions. Au niveau des facteurs qui entrent en compte dans le choix de ces structures linguistiques variantes, le variationnisme ne peut pas non plus prétendre à un traitement global. Ces constatations n'ont rien d'étonnant au vu de l'étendue et de la diversité des formes qui varient, d'une part, et des critères qui peuvent influencer leur usage, d'autre part.

Conclusion

La première partie de ce travail nous a permis d'aborder les fondements de la variable sociolinguistique ainsi que son développement jusqu'au domaine de la syntaxe, où son utilisation est discutée. Nous avons également traité le thème de la variation linguistique de manière plus large, pour en saisir l'amplitude et les diverses approches existantes. Deuxièmement, nous nous sommes intéressée au domaine de l'interrogation et de sa diversité, en particulier au niveau syntaxique. Nous avons vu comment cette diversité avait été traitée et expliquée par plusieurs études, quantitatives et variationnistes.

Si la variable sociolinguistique était originellement conçue pour saisir des corrélations entre usage des variantes et caractéristiques sociales des locuteurs, nous avons vu à travers les études de Coveney et d'Elsig que son application actuelle au domaine des interrogations dépassait largement son étendue originelle. Nous avons pu observer deux usages différents de la méthodologie variationniste appliquée à la diversité des questions : plus proche du concept original de variable sociolinguistique, Elsig se contente d'analyser l'influence de facteurs socio-stylistiques et linguistiques sur la variation des questions, alors que Coveney étend son analyse à la prise en compte de facteurs pragmatiques. Dans les deux cas, les facteurs sociaux ne sont pas considérés comme les principaux régulateurs de l'usage des variantes de la question. Pour Elsig, l'influence des facteurs linguistiques prime sur celle des facteurs sociaux, tandis que pour Coveney, ce sont les facteurs linguistiques et pragmatiques qui exercent une plus grande influence sur l'utilisation des diverses structures interrogatives.

Même si elles se décrivent comme des approches globales de la variation, les études variationnistes ont bien sûr une portée limitée, qui ne permet pas d'aborder tous les aspects de l'usage des structures interrogatives. En effet, à plusieurs reprises dans ce travail, nous avons souligné le grand nombre et la diversité des facettes à travers lesquelles la variation des questions peut être abordée, tant au niveau des unités linguistiques variantes que des facteurs linguistiques ou extralinguistiques susceptibles d'influencer leur usage. Les limites des études variationnistes ont été soulignées, notamment dans les modèles (insuffisants) de structures interrogatives prises en compte, et dans la mesure des phénomènes intra et extralinguistiques agissant sur leur usage. Au vu des nombreuses facettes de la variation et de ses multiples approches possibles, la volonté exprimée par Coveney et Elsig d'expliquer la variation des structures interrogatives « dans leur ensemble » peut sembler aussi disproportionnée qu'irréalisable. En effet, chaque aspect de la variation des formes interrogatives présente ses propres complexités et mérite que l'on s'y attarde de manière exclusive. Les études variationnistes ne peuvent pas prétendre à des comptes rendus minutieux et complets de chacun des aspects qui entrent en compte dans l'usage des formes de la question. Cependant, de par leur approche globale, elles ont l'avantage de montrer qu'on ne peut réduire les explications de la variation à un seul facteur. Ainsi, les études variationnistes appellent à une considération pluridimensionnelle de la variation des questions, spécialement celle de Coveney, qui aborde les causes de variation des questions de son corpus selon quatre axes différents.

En définitive, il nous semble que la méthodologie variationniste peut constituer un outil intéressant pour appréhender certains aspects de la variation des questions, mais ceci à deux conditions. Premièrement, la méthodologie doit s'adapter à l'objet étudié, c'est à dire dans le cas des structures interrogatives, prendre en compte les questions de sens et de fonction, et considérer l'influence de facteurs pragmatiques sur le choix des variantes. Deuxièmement, la portée et les limites des études variationnistes doivent être bien connues,

de sorte que les chiffres qui en résultent doivent être vus comme relatifs à certaines données et à une certaine méthodologie.

Bibliographie

- Al, B., (1975), *La notion de grammaticalité en grammaire générative-transformationnelle : étude générale et application de la syntaxe de l'interrogation directe en français parlé*, Leyde : Presse universitaire
- Béguelin, M.-J., et G. Corminboeuf, (2005), « De la question à l'hypothèse : aspects d'un phénomène de coalescence », *Les états de la question*, C. Rossari & al. (éd.), Québec : Nota bene
- Behnstedt, P., (1973), *Viens-tu? Est-ce que tu viens? Tu viens? : Formen und Strukturen des direkten Fragesatzes im Französischen*, Tübingen : G. Narr
- Berrendonner, A., M. Le Guern, G. Puech, (1983), *Principes de grammaire polylectale*, Lyon : PUL
- Berrendonner, A., (1987), « Stratégies morpho-syntaxiques et argumentatives », *Protée*, (vol.15) n° 3, 48-58
- Berrendonner, A., (1988), « Normes et variations », *La langue française est-elle gouvernable ?* sous la dir. de G. Schoeni, J.-P. Bronckart, P. Perrenoud
Neuchâtel/Paris : Delachaux & Niestlé
- Berrendonner, A., (2005), « Question et mémoire discursive », *Les états de la question*, C. Rossari & al. (éd.), Québec : Nota bene
- Biber, D., (1995), *Dimensions of register variation, a cross-linguistic comparison*, Cambridge : Cambridge University Press
- Blanche-Benveniste, C., (1997a), « La notion de variation syntaxique dans la langue parlée », *La variation en syntaxe*, F. Gadet (éd.), Paris, Langue française 115
- Blanche-Benveniste, C., (1997b), *Approches de la langue parlée*, Gap/Paris : Ophrys, L'essentiel français
- Bonhomme, M., (2005), « Flou et polyvalence de la question rhétorique : l'exemple des *Fables* de La Fontaine », *Les états de la question*, C. Rossari & al. (éd.), Québec : Nota bene
- Borillo, A., (1978), *Structure et valeur énonciative de l'interrogation totale en français*, Aix : Université de Provence, Thèse Lettres Provence
- Borillo, A., (1981), « Quelques aspects de la question rhétorique en français », *DRLAV* 25, 1-33
- Bratt Paulston C. & G. Richard Trucker (éds.), (2003), *Sociolinguistics, the essential readings*, Malden [etc.] : Blackwell Publishers
- Cedergren, H., et D. Sankoff (1974), « Variable rules : performance as a statistical reflection of competence », *Language*, vol. 50, n°2, 333-354
- Chambers, J.K. & al. (éd.), (2007), *The Handbook of Language Variation and Change*, Malden Mass : B. Blackwell

- Cheshire, J., (1996), « Syntactic variation and the concept of prominence », *Speech Past and Present, Studies in English Dialectology in Memory of Ossi Ihalainen*, J. Klemola, M. Kytö, M. Rissanen (éds.), Frankfurt am Main/ Bern [etc.] : Peter Lang
- Coveney, A., (2002), *Variability in Spoken French: interrogation and negation*, Bristol : Intellect Books
- Dostie, G., (2004), *Pragmaticalisation et marqueurs discursifs*, Bruxelles : De Boeck, Duculot
- Elsig, M., (2009), *Grammatical variation across space and time : the French interrogative system*, Amsterdam : John Benjamins Publishing Company
- Fontaney, L., (1991), « A la lumière de l'intonation », *La question*, C. Kerbrat-Orecchioni (éd.), Lyon : PUL
- Gadet, F., (1992), « Variation et hétérogénéité », *Langages 108* (vol. 26), 5-15 (disponible sur www.persee.fr)
- Gadet, F., (1997a), *Le français ordinaire*, Paris : A. Colin
- Gadet, F., (1997b), « La variation plus qu'une écume », *La variation en syntaxe*, F. Gadet (éd.), Paris, *Langue française* 115, 5-17
- Gadet, F., (2003), *La variation sociale en français*, Gap/Paris : Ophrys, L'essentiel français
- Godard, D., (1992), « Le programme labovien et la variation syntaxique », *Langages 108* (vol. 26), 51-65 (disponible sur www.persee.fr)
- Jacobson, S., (1989), « Some approaches to syntactic variation », *Language Change and Variation*, R.W. Fasold and D. Schiffrin (éds.), Amsterdam/Philadelphia : John Benjamins Publishing Company
- Kerbrat-Orecchioni, C., (1991a), « Introduction », *La question*, C. Kerbrat-Orecchioni (éd.), Lyon : PUL
- Kerbrat-Orecchioni, C., (1991b), « L'acte de question et l'acte d'assertion : opposition discrète ou continuum ? », *La question*, C. Kerbrat-Orecchioni (éd.), Lyon : PUL
- Labov, W., (1972), *Sociolinguistic Patterns*, Philadelphia : University of Pennsylvania Press
- Labov, W., (1978), « Where does the linguistic variable stop? A response to Beatriz Lavandera », *Sociolinguistic Working Paper 44*, 1-17 (disponible sur www.eric.ed.gov)
- Lavandera, B., (1978), « Where does the sociolinguistic variable stop ? » *Language in society* 7, 171-182 (disponible sur www.eric.ed.gov)
- Lavandera, B., (1988), « The study of language in its socio-cultural context », *Linguistics : The Cambridge Survey, vol. IV, Language : The Socio-cultural Context*, F.J. Newmeyer (éd.), Cambridge : Cambridge University Press
- Le Querler, N., (1994), « La position du sujet dans l'interrogation directe », *L'interrogation : des marques aux actes*, travaux linguistiques du CERLICO, Rennes : PUR

- Lyons, J., (1977), *Semantics*, Cambridge; London : Cambridge University Press
- Lyons, J., (1995), *Linguistic Semantics : an Introduction*, Cambridge : Cambridge University Press
- Milroy, J. & L., (2000), « Varieties and variation », *The Handbook of Sociolinguistics*, F. Coulmas (éd.), Oxford [etc.] : Blackwell Publishers
- Pohl, J., (1965), « Observations sur les formes d'interrogation dans la langue parlée et dans la langue écrite non littéraire », *Linguistique et philologie romanes : Xe Congrès international de linguistique et philologie romane*, Paris : Librairie Klincksieck, tome II, 4^e partie
- Quillard, V., (2001) « La diversité des formes interrogatives : comment l'interpréter ? », *Langage et société* 2001/1, n°95, 57-72
- Renchon, H., (1967), *Etudes de syntaxe descriptive, II La syntaxe de l'interrogation*, Bruxelles : Palais des académies
- Romaine, S., (1984), « On the Problem of Syntactic Variation and Pragmatic Meaning in Sociolinguistic Theory », *Folia Linguistica* Vol. XVIII, n° 3-4 (disponible sur www.eric.ed.gov)
- Sankoff, D., (1988), « Sociolinguistics and syntactic variation », *Linguistics : The Cambridge Survey, vol. IV, Language : The Socio-cultural Context*, F. J. Newmeyer (éd.), Cambridge : Cambridge University Press
- Sankoff, G., (1973), « Above and beyond phonology un variable rules », *New ways of analyzing variation in English*, C.-J. Bailey & R. W. Shuy (éds.), Washington D.C. : Georgetown University Press
- Santin-Guettier, A.-M., (1994), « Inversion et statut des opérations », *L'interrogation : des marques aux actes*, travaux linguistiques du CERLICO, Rennes : PUR
- Söll, L., (1983), « L'interrogation directe dans un corpus de langage enfantin », *Etudes de grammaire française descriptive*, F.-J. Hausmann (éd.), Heidelberg : J. Groos, Studies in descriptive linguistics 9
- Traverso, V., (1991), « Question et commentaire dans la conversation familière », *La question*, C. Kerbrat-Orecchioni (éd.), Lyon : PUL
- Winford, D., (1996), « The problem of syntactic variation ». In J. Arnold & al (éd.), *Sociolinguistic variation: data, theory, and analysis*, Stanford: Center for the Study of Language and Information, 177-192

Ouvrages de référence

- Grevisse, M., et A. Goosse, (2007), *Le Bon Usage*, Paris-Louvain-la-Neuve : Duculot, 14^e édition
- Riegel, M., J.-C. Pellat, R. Rioul, (2008), *Grammaire méthodique du français*, Paris : PUF
- Arrivé, M., F. Gadet, M. Galmiche, (2005), *La grammaire d'aujourd'hui*, Paris : Flammarion

Neveu, F., (2004), *Dictionnaire des sciences du langage*, Paris : A. Colin

Dubois, J., & al. (2007), *Linguistique et sciences du langage*, Paris : Larousse

Détrie, C., P. Siblot, B. Vernie, (2001), *Termes et concepts de l'analyse du discours : une approche praxématique*, Paris : H. Genève, H. Champion / Genève : distr. Slatkine

Œuvres littéraires

Camus, A., (1956), *La chute*, Paris : Gallimard, coll. « Folio », 2003

Koltès, B.-M., (1988), *Roberto Zucco*, Paris : Les Editions de Minuit, 2004

Pennac, D., (1987), *La fée carabine*, Paris : Gallimard, coll. « Folio », 1994