

Analyse des erreurs et acquisition des structures interrogatives du français

1 Le Centre de linguistique appliquée de l'Université de Neuchâtel a entrepris un travail de recherche sur l'analyse des erreurs, auquel participent Mme Françoise Redard, MM. René Jeanneret et Gérard Merkt, sous la direction du soussigné. Il s'agit, plus précisément, de poursuivre les objectifs suivants:

- 1.1 Elaborer une méthode d'analyse adéquate.
- 1.2 Reconstituer la compétence linguistique des étudiants à un moment donné de leur acquisition des formes interrogatives du français.
- 1.3 Etudier l'évolution diachronique de cette compétence, dans le but de découvrir les modalités linguistiques de l'acquisition.
- 1.4 Proposer une description de l'interrogation française utile à la fois au linguiste et au professeur de français.
- 1.5 Fournir à ce professeur un inventaire des difficultés que rencontrent les étudiants dans leur acquisition des formes interrogatives.

2 Avant de reprendre ces points, voici comment nous avons formé notre corpus. Dans une première étape, nous avons utilisé un matériel qui avait été recueilli auparavant dans le cadre d'une expérience plus ancienne sur les formes de présentation des exercices de laboratoire. En dépouillant les réponses des étudiants, nous avons obtenu un premier corpus de phrases interrogatives erronées. C'est sur elles que nous avons travaillé à mettre au point une méthode d'analyse. Toutefois, pour toute une série de raisons sur lesquelles je ne m'arrêterai pas (forme des tests, conditions de passation, public, etc), nous avons été amenés à préparer de nouveaux tests, mieux en rapport avec nos objectifs. Ces tests se caractérisent de la manière suivante:

- 2.1 Ils sont groupés en quatre batteries. Chaque batterie en comporte quatre. Chaque test est formé de quinze items.
- 2.2 On retrouve à peu près les mêmes structures grammaticales dans chacune des quatre batteries, bien qu'avec un vocabulaire différent et dans un ordre variable.
- 2.3 Les quatre batteries sont destinées à être passées par le même public, avec un intervalle d'un mois environ entre chacune d'entre elles.
- 2.4 Les quatre tests de chaque batterie ont une présentation différente. Cependant, on retrouve ces quatre présentations dans chacune des batteries. Les tests se distinguent essentiellement par l'effort de création que l'on demande aux étudiants. Voici, sommairement décrites, les tâches qu'on leur demande de remplir:
 - 2.4.1 Choix multiples: choisir la forme correcte parmi un ensemble de trois formes, dont deux sont plus ou moins agrammaticales.

2.4.2 Reconstruire une phrase à partir d'un ensemble non-ordonné de mots ou de groupes de mots (système du puzzle).

2.4.3 Produire une phrase interrogative à partir d'une phrase affirmative.

2.4.4 Imaginer des questions qui doivent prendre place dans un dialogue dont on ne donne à l'étudiant que certaines répliques.

2.5 Nos tests sont destinés à un public assez hétérogène. Le seul point commun des candidats est qu'il s'agit d'étudiants étrangers de niveau moyen faible (c'est-à-dire qui se situent — grosso modo — au niveau des dernières leçons de *Voix et Images de France*, ou dans les catégories 3 à 4 du test CGM), que ces étudiants vivent en milieu francophone, et que la grande majorité d'entre eux sont adolescents. Ils sont au nombre d'une centaine répartis en six classes. Les langues suivantes sont représentées (par ordre décroissant d'importance numérique): allemand et dialectes alémaniques, anglais, hollandais, espagnol, suédois, danois, turc, iranien, arabe, hindi et ourdou. Ces étudiants ont été en grande partie initiés au français dans leurs pays d'origine, donc par des méthodes sans doute très diverses; certains ont utilisé, à Neuchâtel, *Voix et Images de France*. Actuellement, la plus grande partie d'entre eux travaille avec le manuel *2ème année de français*, édité par l'Ecole supérieure de commerce de Neuchâtel. Le nombre d'heures d'enseignement hebdomadaires consacrées au français varie entre huit et vingt. Toutefois les étudiants qui ont moins de vingt heures suivent d'autres cours donnés en français.

Précisons enfin que leur programme actuel de grammaire française ne prévoit aucun enseignement systématique de l'interrogation, mais que, du fait qu'ils ont de nombreuses heures de conversation, ils l'utilisent sans doute assez fréquemment.

3 Au moment où nous rédigeons cet article, la correction des tests est loin d'être terminée. Nous ne pouvons donc pas donner de réponses exhaustives aux questions 1.2 à 1.5. Nous insisterons sur l'aspect méthodologique de notre travail plus que sur ses résultats matériels. Ceux-ci feront l'objet d'un article ultérieur.

4 Nous sommes partis de l'idée que l'acquisition d'une langue (et, plus modestement, d'un secteur quelconque d'une langue, ici, l'interrogation) passe par une série d'étapes, et que chaque étape peut se définir par la possession d'une *compétence intermédiaire* de la langue. Nous prenons le terme *compétence* dans le sens restreint qu'il a en grammaire générative, c'est-à-dire comme un algorithme fini capable d'engendrer un nombre infini de phrases. Dans le cas d'un francophone, par exemple, cet algorithme peut engendrer, en principe, n'importe quelle phrase française; lorsqu'on présente à un francophone une phrase dont l'algorithme ne peut pas rendre compte, il la

ressent comme fautive, ou *agrammaticale*. On définira donc la *compétence intermédiaire* d'un étudiant étranger comme un autre algorithme, qui permet, lui, d'engendrer une partie seulement des phrases que le francophone reconnaît comme correctes, mais aussi une classe de phrases incorrectes. Nous appelons *erreurs* ces phrases incorrectes, en prenant à notre compte la définition de Corder¹.

5 Le point de vue que nous venons de présenter appelle un certain nombre de remarques.

5.1 Nous partons non pas d'une évidence absolue, mais d'une hypothèse qu'il s'agira de vérifier et, probablement, de modifier. Il n'est en effet pas prouvé que les connaissances partielles que l'étudiant a de la langue étrangère soient systématiques. Les théories behavioristes de l'apprentissage nous induiraient à penser plutôt le contraire, c'est-à-dire à décrire l'acquisition de la langue comme une série de tentatives en vue d'imiter des modèles. Certaines de ces tentatives seraient couronnées de succès, d'autres aboutiraient à des échecs, c'est-à-dire à des phrases incorrectes. Ces résultats variables s'expliqueraient par des facteurs tels que les dispositions psychologiques momentanées de l'étudiant, la façon de présenter les modèles, les conditions extérieures de l'apprentissage, etc. L'objectif principal de l'enseignement serait donc de diminuer, ou si possible de supprimer, tous les facteurs qui empêchent une bonne imitation du modèle. Pour prouver le bien-fondé de notre hypothèse, nous devons donc montrer que les performances des étudiants ont un caractère au moins partiellement systématique, c'est-à-dire qu'aussi bien les productions correctes que les erreurs peuvent s'expliquer au moyen d'un algorithme unique. Il s'agit, en d'autres termes, de démontrer l'existence du *niveau systématique* dont parle Corder².

5.2 Dans cette hypothèse, la compétence intermédiaire se définira par une série de divergences entre le système de l'étudiant et celui du locuteur francophone moyen (système de référence). Ces divergences se présenteront comme des infractions au système de référence. Elles se manifesteront dans les erreurs de l'étudiant. Nous parlerons plus bas de la nature de ces infractions.

5.3 Toujours dans notre hypothèse, l'acquisition de la langue étrangère se ramènera à des modifications intérieures du système. En principe, ces modifications doivent rendre le système de l'étudiant toujours plus semblable au système de référence, et l'acquisition sera parfaite lorsque il y aura coïncidence totale. Autrement dit, l'analyse des erreurs nous place dans une situation méthodologique tout à fait comparable à celle de l'analyse

1 cf. Bulletin CILA, 14, 1971, 6-15

2 Idem

contrastive: le problème à résoudre est celui de la comparaison de plusieurs systèmes entre eux; d'une part, le système de la compétence intermédiaire d'un étudiant à un moment donné et le système de référence, d'autre part les différents systèmes de cet étudiant considérés dans leur succession diachronique. La difficulté est de taille si l'on sait que, jusqu'à maintenant, c'est un des problèmes sur lesquels se sont achoppées les études contrastives. Il est probable toutefois que l'existence de nombreux points de contact entre les états successifs de la compétence intermédiaire d'un même étudiant facilite notre tâche jusqu'à un certain point.

6 A l'heure actuelle, c'est sans doute la grammaire générative (version 1965) qui répond le mieux à nos besoins. Rappelons en effet qu'elle possède entre autres les caractéristiques suivantes:

6.1 Elle constitue un véritable algorithme au sens défini sous 4: Un nombre relativement réduit de règles permet d'engendrer n'importe quelle phrase interrogative considérée comme correcte. Ces règles constituent donc une norme et permettent, par conséquent, de porter un jugement sur la grammaticalité d'une phrase donnée.

6.2 Une grammaire générative est "falsifiable", c'est-à-dire que, en principe, il est toujours possible d'expliquer la production d'une forme agrammaticale par une malformation définie de l'algorithme. Par conséquent, nous serons en mesure, non seulement de proposer une description correcte de la norme interrogative, mais aussi de formuler des hypothèses explicites sur la manière dont l'étudiant étranger enfonce la grammaire de la langue cible lorsqu'il produit une erreur.

6.3 En ramenant tous les choix sémantiques à la base syntagmatique (en considérant donc que les règles de transformation ne modifient pas le sens des phrases qu'elles engendrent), la grammaire générative que nous avons élaborée permet de distinguer d'une part l'*acceptabilité* grammaticale d'une phrase, d'autre part son *adéquation* à la situation ou au contexte³. La première met en jeu les règles de transformation, la seconde les règles syntagmatiques. En outre, nous avons supposé que les choix sémantiques de la base sont identiques en français et dans la langue maternelle des étudiants. C'est dire que, une fois admis que l'étudiant connaît le sens de la question qu'on lui demande de poser, une erreur de sa part se situe nécessairement au niveau des règles de transformation.

6.4 Si la grammaire générative nous donne des directives rigoureuses concernant la forme des règles, il n'en est pas de même de leur contenu ni de

3 cf. Corder, in Nickel, G., *Fehlerkunde*, Berlin, Cornelsen-Velhagen & Klasing, 1972, 38-50

la façon dont elles s'ordonnent. Il y a donc, en principe, un nombre infini de façons d'engendrer les phrases interrogatives du français. Cette liberté est importante pour nous, car elle nous permet (et nous incite) à trouver une solution qui soit adéquate à deux points de vue: la description du français et l'analyse des erreurs. Nous avons rencontré des cas où ces deux critères entrent en conflit⁴.

7 La place nous manque pour présenter ici le détail de notre algorithme⁵. Disons seulement que nous utilisons une version considérablement simplifiée de la base syntagmatique de PACEF (Projet d'analyse contrastive espagnol-français-italien)⁶ et une série de vingt-trois règles de transformation. Nous engendrons ainsi l'ensemble des structures interrogatives qui font partie, à notre avis, de ce qu'on appelle parfois le *français standard*⁷. Autrement dit, nous acceptons comme correctes les structures utilisables à tous les niveaux de langue. Nous excluons, par exemple, des phrases comme *tu fais quoi* ou *tu vas où*, caractéristiques de la langue familière, et *qu'est-ce* ou *ton frère a-t-il téléphoné*, caractéristiques, respectivement, d'une langue recherchée et de la langue écrite en général. Remarquons d'ailleurs qu'il suffirait de quelques petites modifications de l'algorithme pour les "grammaticaliser". Précisons à ce propos que la notion d'erreur est toute relative. Pour nous, le terme *erreur* a une signification purement technique: il désigne toute divergence entre la production de l'étudiant et la grammaire de référence. Nous ne nous prononçons pas sur la question de savoir si la phrase agrammaticale est absolument inacceptable, relativement acceptable, ou tout à fait acceptable, mais seulement à un niveau de langue particulier. Dans la première étape de notre analyse, nous ne prétendons en aucune façon porter un jugement qualitatif nuancé sur la compétence de l'étudiant. Un tel jugement pourrait s'appuyer, plus tard, sur la nature des infractions: certaines d'entre elles seraient déclarées *stylistiques*, et n'auraient donc aucune conséquence négative sur l'appréciation du travail de l'étudiant.

4 cf. Py, B. *Un problème de linguistique appliquée: grammaire générative et contraintes pédagogiques*, in *Studi italiani di linguistica teorica ed applicata*, 3, 1973

5 On trouvera une présentation détaillée dans Py, B., *Analyse des erreurs et grammaire générative: la syntaxe de l'interrogation en français*, Québec, CIRB, Université Laval, 1972

6 cf. Borel, J.-P., *Investigaciones lingüísticas previas a la confección de material didáctico*, in *Boletín de la Asociación europea de profesores de español*, 4, 1971, 5-54

7 Nous acceptons ici la définition qu'en donne J. Dubois dans *Grammaire structurale du français: nom et pronom*, Paris, Larousse, 1965, p. 5.

8 Sous 6.4 nous avons laissé entendre qu'il pouvait y avoir contradiction, dans l'élaboration de la grammaire de référence, entre les points de vue de la description et de l'efficacité analytique. De fait, nous avons constaté que la description la plus simple (donc, d'après les générativistes, la meilleure) ne permettait pas d'expliquer d'une façon adéquate toutes les erreurs des étudiants. Par exemple, un algorithme qui introduirait *est-ce que* au moyen de la même règle dans les phrases *qui est-ce que tu as rencontré* et *où est-ce que tu l'as rencontré* n'expliquerait pas l'absence de corrélation, que nous avons constatée chez de nombreux étudiants, entre les deux productions agrammaticales suivantes: *qui tu as rencontré* et *où tu l'as rencontré*. Si l'on veut expliquer cette non-corrélation, il est nécessaire de disposer de deux règles distinctes. Cette remarque ouvre un champ de recherches aussi vaste qu'intéressant. Elle pose tout le problème des critères d'évaluation des grammaires. Une analyse comme la nôtre permet sans doute de contribuer à ces recherches. En effet, il s'agit non seulement d'expliquer des erreurs à partir d'une grammaire de référence donnée, mais aussi de modifier cette grammaire aussi souvent que nécessaire jusqu'à ce qu'elle constitue un instrument d'analyse adéquat (cf. 1.4). Pratiquement, depuis le début de nos travaux (hiver 1970-71) nous avons procédé à plusieurs révisions plus ou moins profondes, et la version actuelle a une capacité explicative incontestablement plus grande que la première. Pour des raisons évidentes nous sommes obligés de limiter le nombre de modifications (car chaque modification nous oblige à reprendre les analyses déjà faites), mais, lorsque nous aurons examiné tout le corpus à notre disposition, l'existence de corrélations fréquentes entre différents types d'erreurs (ou, au contraire, leur non-existence), nous amènera sans doute à proposer une version "définitive" de notre grammaire de l'interrogation. Ainsi, l'idéal que nous visons est une grammaire de référence dont la constitution logique recouvre non seulement la réalité de la langue, mais explique encore avec élégance les erreurs des étudiants. De ce point de vue, nos résultats sont certainement loin d'être définitifs. Prenons un exemple où ils sont positifs: Une des erreurs les plus fréquentes est la non-application d'une règle qui sert à placer le syntagme interrogatif en tête de phrase; d'où, par exemple, *il habite où*. Or, dans notre grammaire de référence, la non-application de cette règle a pour effet de rendre possible l'application ultérieure d'une autre règle, qui sert, elle, à introduire facultativement *est-ce que* en tête des interrogations globales: *est-ce que vous partez*. Et nous avons constaté qu'il existe effectivement une corrélation de fait, chez un certain nombre d'étudiants, entre la production de *il habite où* et celle de *est-ce qu'il habite où*. Autrement dit, sur ce point précis, une simple analyse du système de référence permettait de prédire un certain type d'erreurs.

9 L'analyse des résultats d'un étudiant doit passer par une série d'étapes que nous esquissons ici.

9.1 Il s'agit d'abord d'écartier les erreurs qui ne concernent pas l'interrogation proprement dite, c'est-à-dire celles que nous ne pouvons pas expliquer par une infraction à la grammaire de référence.

9.2 On écarte ensuite les fautes qui, manifestement, se situent au niveau de la performance. On les reconnaît par le fait qu'elles sont isolées et ne constituent que des infractions secondaires. C'est le cas, par exemple, d'une production comme *où est-ce il est* lorsque, par ailleurs, l'étudiant produit correctement *où est-ce qu'il habite*, etc. Il est souvent difficile de distinguer d'emblée les erreurs systématiques des fautes de performance, et seule une analyse de l'ensemble des réponses de l'étudiant permet de prendre une décision dans les cas douteux.

9.3 On interprète chaque erreur en fonction de la grammaire de référence, en termes d'infractions. Celles-ci peuvent être de différentes natures:

9.3.1 Une règle n'est pas appliquée; c'est le cas, par exemple, dans *tu partiras quand* ou *pourquoi tu restes*; dans le premier cas, l'étudiant n'applique pas la règle d'antéposition du syntagme interrogatif, dans le second, celle qui introduit *est-ce que* lorsque le sujet précède le verbe.

9.3.2 Les limites du domaine d'application d'une règle sont transgressées. C'est une des infractions les plus fréquentes. Elle explique les exemples agrammaticaux suivants: *avec qui est-ce qui elle est partie*, *qui est-ce que avez-vous invité*, *de qu'est-ce que vous parlez*, *est-ce que chez qui d'nez-vous ce soir*, etc.

9.3.3 Neutralisation d'une opposition à l'intérieur du domaine d'application. Exemples: *combien de cartes postales avez envoyé vous*, *à qui est-ce que tu joues*, etc.

9.3.4 Transgression d'une condition d'application. Dans notre grammaire, une seule règle en comporte: l'inversion du sujet nominal n'est en effet pas possible si le mot interrogatif est *pourquoi*. La transgression de cette condition engendre, par exemple, *pourquoi s'arrête la voiture*.

9.3.5 Déformation d'un morphème de surface. Ce genre d'erreurs est fréquent avec *qui est-ce qui*, *qu'est-ce qui*, etc. Exemples: *qui est-ce qui est Suzanne*.

9.3.6 Inversion de deux règles. C'est ainsi qu'une phrase agrammaticale comme *est-ce que les enfants où sont* devient possible dans notre système: il suffit de supposer que la règle d'antéposition du mot interrogatif précède la règle d'antéposition du sujet au lieu de la suivre.

9.3.7 Modification d'une consigne. Notre grammaire comprend deux règles dites phonologiques. L'une d'elles dit que, si on obtient une phrase terminée par la séquence *sujet nominal + est*, il faut revenir en arrière et appliquer la

règle d'inversion du sujet, laquelle est normalement facultative. Certains étudiants paraissent ressentir l'impossibilité de dire *comment est-ce que cette table est*, mais au lieu de remplacer cette phrase par *comment est cette table*, ils suppriment le verbe et produisent seulement *comment est-ce que cette table*.

Cet inventaire des infractions est très schématique, et ces interprétations ne se justifient, bien sûr, qu'en fonction des règles du système de référence. Elles ont seulement pour but de donner au lecteur une idée de notre façon de travailler.

9.4 Certaines de ces infractions sont *isolées*. Dans ce cas, il serait imprudent d'en tirer des conséquences quant à la compétence de l'étudiant; il y a tout lieu de penser qu'il s'agit seulement de fautes de performance, donc non-systématiques. Un grand nombre d'infractions sont *répétées* sans être absolument généralisées. Dans ce cas, nous estimons qu'elles sont révélatrices d'un trait systématique de la compétence intermédiaire: il ne faut pas oublier que celle-ci n'est probablement jamais un système fermé, mais plutôt un système en perpétuelle évolution. A cet égard, les infractions répétées montrent qu'une nouvelle forme de compétence intermédiaire est en préparation mais qu'elle n'a pas encore remplacé complètement la précédente. Enfin, un troisième groupe comprend des infractions *généralisées*. Il semble qu'elles sont moins nombreuses que les infractions répétées. Elles représentent des secteurs cristallisés de la compétence intermédiaire que l'étudiant, au moins provisoirement, ne paraît pas remettre en cause.

Il est probable que certaines des infractions isolées ou répétées s'expliquent par la coexistence, sous les yeux de l'étudiant, de plusieurs normes, dont l'existence est liée à celle des registres de langue et, plus particulièrement, de la différence entre langue parlée et langue écrite. Dans le cas de l'interrogation française, ce phénomène est particulièrement évident. L'étudiant vit dans un milieu où, par exemple, les règles d'introduction de *est-ce que* et d'antéposition du syntagme interrogatif sont régulièrement éludées: *quand vous partirez, vous allez où*, etc. Cette contradiction entre la langue parlée familière et la norme enseignée à l'école conduit évidemment l'étudiant à des hésitations qui se manifestent par ces infractions isolées ou répétées.

10 *Conclusion* Sans apporter de réponse précise à la question de savoir comment des étudiants étrangers acquièrent les structures interrogatives du français, cet article décrit surtout une méthode de travail, énumère certains de ses présupposés théoriques et essaie de montrer l'intérêt d'une telle étude aussi bien pour l'enseignant que pour le linguiste. Nous espérons surtout avoir montré comment un problème essentiellement pédagogique, tel que l'analyse des erreurs, peut chercher une solution en utilisant les instruments que lui

propose la linguistique théorique. D'autre part, il se peut que l'application de ces instruments à des problèmes pédagogiques suggère au linguiste des solutions nouvelles à ses propres questions.

Centre de linguistique appliquée
Université de Neuchâtel
CH 2000 Neuchâtel

Bernard Py