

MAGUY BAUHR – JAN HELLEKANT

Comment évaluer l'expression orale au lycée?

Maguy Bauhr och Jan Hellekant tjänstgör på Avdelningen för språkpedagogik vid Göteborgs universitet. Avdelningen, som leds av doc. Torsten Lindblad, ansvarar bl.a. för olika nationella språkprov. Ett av de nyare projekten har som huvuduppgift att ta fram muntliga prov i engelska för grundskolan och i engelska, franska och tyska för gymnasieskolan. Här redogör författarna för ett försök att utveckla muntliga prov i franska och drar på basis av detta några tentativa slutsatser rörande elevernas muntliga språkfärdighet.

Introduction

Depuis 1968, date à laquelle le baccalauréat suédois a été supprimé, la Direction Nationale de l'Enseignement Public met au point des tests standardisés et obligatoires – les épreuves centralisées – dans certaines matières, dont l'allemand, l'anglais et le français. Ces épreuves sont administrées en deuxième année du lycée. Elles visent essentiellement à établir un système de notation équitable à l'échelle nationale.

Bien que l'expression orale soit en général considérée, aussi bien par les professeurs que par les élèves, comme essentielle et bien que les instructions officielles soulignent son importance, elle n'est pas prise en compte dans les tests nationaux.

Au début des années 60, les tests qui étaient alors utilisés pour évaluer les compétences des élèves avaient été passés au crible et sévèrement critiqués pour leur manque de fiabilité. Les spécialistes préconisaient de ne retenir parmi les épreuves possibles que celles qui pouvaient donner lieu à une évaluation strictement objective. Confrontés au problème posé par l'évaluation des compétences orales et ne trouvant pas de manière absolument fiable de les tester, les constructeurs finirent par y renoncer.

Vers la fin des années 70, la DNEP a voulu susciter une nouvelle réflexion dans ce domaine. Un test d'expression orale a été élaboré pour le français. Cette tentative est restée isolée et sans grandes conséquences pratiques.

Une dizaine d'années plus tard, en 1987, l'Institut de Recherche pour la Pédagogie des Langues de l'Université de Göteborg a entrepris de mettre au point des épreuves d'expression orale pour l'allemand, l'anglais et le français. Cette mission lui a été confiée pour une période de cinq ans, avec un objectif précis, celui d'étudier la possibilité d'intégrer ces épreuves à celles qui constituent actuellement le test centralisé.

Le projet

La nature du problème reste inchangée: il s'agit toujours de trouver pour

une activité libre et difficilement canalisable un test qui permette une évaluation suffisamment fiable. Ce qui peut-être est en train de changer, c'est l'attitude à l'égard du problème et la volonté d'essayer de le résoudre, dans la mesure du possible.

Effets escomptés

Dans le courant des années 80, les spécialistes de l'évaluation avaient attiré l'attention sur l'effet de reflux, c'est-à-dire l'effet qu'a le contenu d'un test sur l'enseignement et sur l'apprentissage. Ils portaient de la constatation qu'une compétence qui fait l'objet d'une évaluation acquiert par là même un certain prestige auprès des usagers, c'est-à-dire, dans notre contexte, auprès des élèves et de leurs professeurs. Une bonne manière d'améliorer la maîtrise de la langue orale serait donc de la tester.

de Landsheere (1988) souligne par ailleurs l'importance qu'il y a à modifier les épreuves si l'on veut modifier les pratiques scolaires: «Si certaines réformes de programme sont demeurées pratiquement sans effet, c'est parce que les examens externes n'ont pas évolué parallèlement». (p. 201)

Dans une certaine mesure, la décision prise par la DNEP d'initier le projet qui nous occupe procède donc de la volonté de développer une meilleure maîtrise de la langue parlée.

En outre, les professeurs savent par expérience qu'il n'y a pas toujours concordance entre la manière dont leurs élèves s'expriment oralement en français et leur niveau tel qu'il est établi par d'autres types d'épreuves. La plupart d'entre eux tiennent compte de cette distorsion au moment d'attribuer à leurs élèves les notes de fin de semestre. Il paraissait probable que si l'on arrivait à établir un niveau moyen à l'échelle nationale, le jugement qu'ils portent sur les prestations de leurs élèves serait facilité.

L'introduction d'une épreuve de production orale correspondait donc à une double attente:

- un effet de reflux, c'est-à-dire un effet bénéfique sur l'apprentissage et l'enseignement.
- une plus grande équité et une meilleure adéquation de la note attribuée aux élèves.

Définition

Pour une période expérimentale de cinq ans, l'objectif essentiel était de développer des instruments appropriés pour tester la maîtrise de l'expression orale. Cet objectif était décrit en trois points:

- Produire différents modèles d'épreuves et en analyser les mérites comparés.
- Mettre au point des méthodes et des critères d'évaluation efficaces.
- Etudier comment ces épreuves pouvaient être mises en œuvre dans les écoles.

Principes de construction

En principe, une épreuve, pour être efficace, doit répondre à certaines exigences fondamentales de validité, de fiabilité et d'économie.

1. En vue d'assurer un degré satisfaisant de validité au test que nous proposons de construire, notre premier soin a été de préciser ce que nous voulions mesurer.

Notre intention n'est pas de contrôler le résultat d'une séquence d'apprentissage, mais d'évaluer un savoir-faire général, ce que Mothe (1990) caractérise comme la capacité des élèves «à se tirer d'affaire plus tard, lorsqu'ils seront privés des béquilles du professeur» (p. 133). Il s'agit donc d'un test de niveau (*proficiency test*) et non d'un test de contrôle (*achievement test*).

Ce savoir-faire correspond à l'un des objectifs fixés dans les instructions officielles: la maîtrise de l'expression orale dans certaines situations simples de la vie courante. Notre intention était donc d'élaborer un test qui soit aussi réaliste que possible. Nous partons du principe que plus la situation que nous proposons aux élèves est crédible, plus nous avons de chances qu'ils s'y intéressent et qu'ils s'expriment avec spontanéité. En appliquant ce principe, nous espérons aussi que notre test sera considéré comme valide par les élèves.

2. Le talon d'Achille de tout test de production est la fiabilité. Une fiabilité parfaite suppose que le résultat soit constant quel que soit l'examineur. Il est donc essentiel de réduire autant que possible l'impact de la personnalité de l'examineur. Or dans le test oral, l'examineur intervient à deux reprises, d'abord comme interlocuteur puis comme évaluateur. En outre, comme évaluateur, il est, à ce jour, relativement inexpérimenté.

Pour essayer de pallier ces difficultés, nous avons pris certaines précautions:

–Nous avons recommandé au professeur d'enregistrer la conversation qu'il a avec son élève, ce qui peut lui permettre de se consacrer à ses deux rôles successivement.

–Nous avons proposé une évaluation factorielle, pour que tous les professeurs tiennent compte dans la mesure du possible des mêmes critères.

–Nous avons proposé des enregistrements modèles sur vidéocassettes et sur audiocassettes. Différentes prestations d'élèves, accompagnées de commentaires établis par un groupe de professeurs, sont soumises à la réflexion des futurs examineurs. Ces enregistrements ont été soigneusement choisis sur un échantillon représentatif obtenu grâce à un prétest.

–Enfin, nous proposons aux professeurs qui le désirent un programme de formation.

3. Un test oral au niveau national pose des problèmes d'organisation et des problèmes de financement. Ces problèmes ne relèvent pas de notre compétence. Nous pouvons toutefois faire remarquer qu'un examineur expérimenté a rarement besoin de plus de 10 minutes pour tester oralement

un élève. Dans la majorité des cas, c'est moins de temps qu'il ne lui en faut pour corriger une copie. Les problèmes d'organisation ne sont pas insolubles. Le lycée suédois est actuellement en pleine période de réorganisation. Il est probable que dans une organisation rénovée du temps scolaire, ces épreuves trouveront leur place, surtout si l'on considère la tendance générale, au niveau européen, à introduire et à développer ce genre d'épreuve.

Description du test

La première contrainte était donc de partir d'une situation qui ait des chances d'être considérée par les élèves comme plausible, c'est-à-dire d'une situation dans laquelle ils pouvaient imaginer se trouver dans un avenir relativement proche. Par exemple, l'élève pourrait avoir à raconter quelque chose qu'il vient de lire dans un journal, ou avoir à aider un touriste français à acheter des cadeaux pour les membres de sa famille restés en France, etc.

Jusqu'à présent, l'interlocuteur a toujours été le professeur qui, lui, était contraint de jouer un rôle un peu moins plausible: une logeuse irascible, un étudiant étranger, un commissaire de police. Par contre le professeur avait l'avantage d'être muni d'une feuille de rôle sur laquelle était inscrit le texte intégral de ses interventions. Cette précaution ne reflétait aucune méfiance à l'égard des compétences du professeur. Elle avait été prise pour répondre à la deuxième contrainte que nous nous étions fixée: que tous les élèves soient placés, autant que possible, dans la même situation, pour permettre que leurs prestations soient facilement comparables.

La troisième contrainte tient bien entendu au niveau des élèves suédois en cinquième année de français. Il fallait trouver des exercices qui permettent à tous les élèves, même les plus faibles, de ne pas rester muets, mais aussi qui permettent aux meilleurs d'entre eux de s'exprimer au maximum de leurs possibilités. C'est pourquoi les tests que nous avons construits se composent de deux parties. La première est constituée d'un jeu de questions-réponses qui est à la portée de tous les élèves, mais dont ils se tirent avec plus ou moins d'aisance. Dans la seconde partie, qui est plus complexe, les élèves doivent faire preuve d'une plus grande maîtrise de la langue, être capables d'élaborer leur discours et d'évoluer entre le présent, le passé et le futur.

L'évaluation

La notation

Notre ambition première avait été, étant donné le système de notation actuellement en vigueur dans les écoles suédoises, de distinguer trois groupes qui représenteraient chacun un tiers de la population. Ces trois groupes furent dénommés comme suit:

–Au-dessus de la moyenne.

–Moyenne

–Au-dessous de la moyenne

Il semblait peu satisfaisant, au premier abord, de condamner le tiers des élèves à avoir une «mauvaise note». Et il semblerait, au vu des premiers résultats, que les professeurs aient eu une certaine répugnance à décourager leurs élèves.

Dans un article où il présente le projet dans sa totalité, Lindblad (1992), en analysant les résultats obtenus pour les trois langues, donne les chiffres suivants pour l'épreuve qui nous occupe: Au-dessus de la moyenne 35%, Moyenne 53%, Au-dessous de la moyenne 12% (p. 286). Ces chiffres pourraient être aussi dus au fait que, l'épreuve n'étant pas pour l'instant obligatoire, les élèves faibles en ont souvent été dispensés.

Les critères d'évaluation

Pour des raisons de fiabilité, il était préférable que l'évaluation soit factuelle: les professeurs, au lieu de se fier à leur impression générale, essaient d'analyser la prestation de l'élève en fonction de critères préétablis. Une fois ce point admis, restaient un certain nombre de décisions à prendre. Quels critères choisir et comment assurer leur pondération?

Dans un premier temps et pour que nos résultats puissent être comparables à ceux obtenus au sein de notre institution, pour l'anglais et l'allemand, nous avons adopté les critères suivants qui étaient sensiblement les mêmes pour les trois langues:

- Efficacité dans la communication (la capacité de l'élève à se faire comprendre, la facilité qu'il a à s'exprimer, et la variation et la richesse de sa langue).
- Degré de correction (correction dans le choix des mots et correction grammaticale).
- Degré de correction dans la prononciation et l'intonation.

L'examineur, après avoir accordé une note en fonction de chacun de ces critères était invité à décider d'une note globale. S'il estimait que cette décision était difficile à prendre, il lui était suggéré que l'efficacité dans la communication devait être considérée comme primordiale.

A l'usage, ces critères n'ont pas donné entière satisfaction et nous avons été amenés à les modifier. Il semblerait que les professeurs choisissent souvent de pas enregistrer leurs élèves, à la fois pour gagner du temps et pour éviter la présence perturbante du magnétophone. Ce choix est regrettable, car en cas de contestation ou d'hésitation il est souvent nécessaire de se reporter à un document. Quoi qu'il en soit, il nous a paru nécessaire de simplifier le travail des examinateurs, en leur demandant de ne tenir compte que de deux facteurs:

- la capacité à se faire comprendre et le degré d'aisance dans l'expression.
- le degré de correction, en matière de prononciation, de grammaire et de vocabulaire.

La prononciation ne figure donc plus en tant que critère à part entière, mais comme un des éléments de la correction. Nous diminuons ainsi l'im-

portance globale accordée à la correction.

Notre décision de ne plus accorder une place privilégiée à la prononciation se fonde aussi sur une observation qui mériterait d'être analysée de plus près: il semblerait que ce critère n'ait pas été très discriminant. La prononciation des élèves, sans être parfaite, est en général considérée comme acceptable par les professeurs.

Nos conclusions rejoignent celles auxquelles aboutit Berkoff (1985) après avoir participé, en Israël, à l'élaboration d'un examen d'expression orale en langue étrangère: «Une grille d'évaluation trop analytique détourne l'attention de l'examineur. Ce qu'il convient de privilégier, c'est l'outil d'évaluation qui permet d'obtenir la plus grande dispersion». (p. 99)

Le savoir-faire des élèves

Bien que notre objectif principal ait été de développer des instruments appropriés pour tester l'expression orale des élèves, les enregistrements d'élèves que nous avons étudiés nous ont permis de faire certaines observations sur le niveau que les élèves peuvent atteindre en avant-dernière année du lycée.

Nous nous proposons de partir de deux échelles de niveaux établies aux Etats-Unis par deux organismes différents:

ACTFL Scale	ILR Scale
	5 Native or bilingual proficiency
	4+
	4 Distinguished proficiency
	3+
Superior	3 Professional working proficiency
Advanced High	2+
Advanced	2 Limited working proficiency
Intermediate High	1+
Intermediate Mid	1 Survival proficiency
Intermediate Low	
Novice High	0+
Novice Mid	0
Novice Low	No practical proficiency

Fig. 1. Correspondance entre les niveaux de l'ACTFL et ceux de l'ILR

The American Council on the Teaching of Foreign Language, l'ACTFL (1989), distingue quatre niveaux principaux: supérieur, avancé, intermédiaire et débutant. De son côté, l'Interagency Language Roundtable, l'ILR, distingue six niveaux. Pour illustrer la correspondance entre ces deux échelles de niveaux, l'ACTFL a mis au point un schéma que nous reproduisons ci-dessus (Byrnes & Thompson, ch 2, p. 15).

1. Nous sommes en présence de deux échelles de niveaux qui ont en commun d'essayer de couvrir l'une et l'autre un spectre qui va d'une maîtrise nulle de la langue orale à une maîtrise définie comme idéale. Il peut être intéressant de savoir où se situent approximativement nos élèves à l'intérieur de ce spectre.

L'ACTFL décrit son niveau intermédiaire comme suit: «L'apprenant est capable de participer à un dialogue: il est capable de poser des questions et de répondre aux questions qui lui sont posées» (op. cit., ch. 3, p. 2). Cette définition correspond assez bien à la performance des élèves qui réussissent à s'acquitter de la première partie de notre test.

Toujours d'après l'ACTFL, ce qui caractérise le niveau avancé, ce serait «la capacité à élaborer un récit et une description, en utilisant des outils simples pour exprimer le temps et l'aspect» (loc. cit.). On peut dire, en gros, que c'est ce que nous demandons aux élèves dans la seconde partie de notre test.

Il serait bien sûr intéressant de savoir quel degré de perfection est requis par l'ACTFL et l'ILR pour l'accès à ces niveaux. Nous avons aussi très peu d'informations sur ce que l'ILR entend par «intercompréhension dans la vie courante» et «intercompréhension dans la vie active». Il nous est donc difficile de tirer des conclusions d'une évaluation faite en dehors de critères purement scolaires.

2. Peut-on déjà tirer certaines conclusions à partir du matériel dont nous disposons? Dans quelle mesure les élèves de cinquième année de français sont-ils capables d'effectuer les exercices que nous leur demandons sans qu'il y ait trop d'obstacles à l'intercompréhension? En quoi consistent ces obstacles? A quelle distance les lycéens suédois sont-ils de la langue d'un jeune Français?

Description d'un des tests utilisés

L'exercice suivant qui constitue la première partie du test de 1992 (la partie «facile») a été exécuté par un certain nombre de lycéens suédois. Nous avons également mis à contribution un jeune Français, P., qui a bien voulu jouer le jeu dans les mêmes conditions que les élèves suédois. P. a 23 ans, c'est-à-dire qu'il a cinq ans de plus que les lycéens suédois. Il a fait des études de philosophie et d'économie.

La feuille de rôle qui est donnée à l'élève est rédigée *en suédois*. L'élève commence par lire un texte qui lui explique la situation: «Il y a à l'école une élève roumaine qui vient d'arriver en Suède. Votre professeur vous a

demandé de vous occuper d'elle. Vous savez qu'elle a vécu en Afrique avant de venir en Suède. Vous lui posez quelques questions pour en savoir davantage».

Après avoir lu cette courte présentation de la situation, l'élève trouve au verso de sa feuille de rôle quelques jalons qui lui permettront de converser avec la jeune Roumaine. Les jalons sont aussi rédigés *en suédois*. L'élève les a sous les yeux pendant qu'il passe le test.

- prénom?
- nom de famille?
- âge?
- habites où?
- frères et sœurs?
- arrivée en Suède quand?
- habité avant?
- combien de temps?
- pourquoi la Suède?
- resteras combien de temps?
- ce que vous savez sur la Roumanie.

Quelques réflexions sur les prestations des élèves

Les élèves les plus faibles ne se sont pas encore débarrassés de tournures telles que *«Quand as-tu venir?», *«Et pourquoi tu venir?» ou *«Et où as-tu habité?». Ces élèves sont capables de se faire comprendre, si l'on suppose qu'ils ont affaire à un interlocuteur bien intentionné.

Un certain nombre d'élèves sont parfaitement compréhensibles par n'importe qui, même un interlocuteur mal intentionné. Cependant il peut arriver que, même chez ces élèves, le courant ne passe plus. Par exemple, lorsque l'un des élèves, qui par ailleurs est tout à fait compréhensible, demande: «Où est-ce qu'il y a?», il n'est pas tout à fait évident, même si le contexte est connu, qu'il voudrait savoir où se trouve la Côte d'Ivoire. C'est le seul cas où un interlocuteur non averti serait probablement obligé de lui faire préciser sa pensée. A ce niveau, la langue des élèves se caractérise souvent par une certaine lourdeur et une manière stéréotypée de poser les questions.

Les élèves les meilleurs se tirent à merveille de cet exercice. Ils emploient à peu près les mêmes termes que P. Bien entendu, P., triplement avantage par la langue, le niveau d'instruction et l'âge, a le loisir de s'intéresser à ce que dit son interlocuteur. Il fait systématiquement la liaison avec ce qui vient d'être dit, exprime son étonnement devant ce que son interlocuteur lui apprend et donne son opinion sur l'intérêt qu'il y a à vivre en Côte d'Ivoire ou en Suède plutôt qu'en Roumanie.

Il est difficile de tirer des conclusions à partir d'un exemple quant au niveau général des élèves. Mais, ces résultats correspondant assez bien aux résultats obtenus les années précédentes, nous pouvons avancer l'hypothèse

que la quasi-totalité des élèves atteint un niveau que nous pourrions décrire comme étant le niveau qui permet l'intercompréhension dans la vie courante. Ils le font toutefois à des degrés différents.

Pour ce qui est du niveau qui permettrait l'intercompréhension dans la vie active, il ne peut s'agir, étant donné l'âge et les compétences des lycéens, que d'un niveau qui permettrait soit de suivre une formation théorique ou professionnelle, soit d'effectuer un travail relativement peu qualifié. Ce niveau supposerait des relations sociales plus suivies, c'est-à-dire la capacité de participer à des conversations moins rudimentaires, de parler de son pays, de faire un compte-rendu succinct et aussi de se défendre contre les abus éventuels d'un logeur, d'un employeur, d'un camarade ou d'un enseignant.

Ce que nous pouvons dire à ce jour, c'est qu'une partie des élèves semble accéder à ce niveau. Il reste à déterminer dans quelles proportions et à quel degré.

Conclusions

Jusqu'à présent, l'épreuve que nous venons de décrire a été facultative. Notre intention, pendant cette période expérimentale de cinq ans, était qu'à terme elle devienne obligatoire. Elle contribuerait ainsi à développer chez les élèves une meilleure maîtrise de l'expression orale, et à encourager les professeurs à tenir compte explicitement de cette compétence dans la note qu'ils attribuent à leurs élèves.

A l'avenir le contenu de cette épreuve évoluera probablement. Il nous paraît intéressant d'essayer d'explorer autant que possible d'autres domaines dans lesquels les élèves ont certaines possibilités de s'exprimer. Il est toutefois probable qu'il nous sera utile, même par la suite, de distinguer deux parties: l'une qui serait à la portée de l'ensemble des élèves et l'autre qui permettrait aux élèves les plus avancés de donner leur mesure.

La forme exacte que prendra l'évaluation au niveau national n'est pas encore établie. Les directives du Ministère de l'Éducation nationale sont toutefois nettes: A l'avenir, l'évaluation au niveau national devra «viser à donner une image complète et nuancée des connaissances des élèves». Il conviendra aussi «d'étudier la possibilité de mettre au point des épreuves orales, surtout pour les langues étrangères».

Le système de notation dans les écoles suédoises étant aussi en cours de réactualisation, il est probable que nous aurons à l'avenir non seulement à comparer les prestations des élèves entre elles, mais aussi à tenir compte de la notion de niveau. Il conviendra alors de décrire avec exactitude le niveau des élèves. C'est dans cette perspective que nous avons émis quelques hypothèses sur le niveau qu'atteignent les élèves suédois.

Références

- Berkoff, N.A. (1985) «Testing oral proficiency: a new approach». Dans Lee, Y.P. et al. (eds), *New Directions in Language Testing*. Oxford: Pergamon Press.
- Byrnes, H & Thompson Irene (1989) *The ACTFL Oral Proficiency Interview. Tester Training Manual*. New York: The American Council on the Teaching of Foreign Languages.
- de Landsheere V. (1988) *Faire réussir, faire échouer. La compétence minimale et son évaluation*. Paris: Presses Universitaires de France.
- Lindblad, T. (1992) «Oral Tests in Swedish Schools: A five-year Experiment». *System*, Vol. 20, No. 3, p. 279-292.
- Mothe, J.C. (1991) «Évaluer les compétences en langue étrangère». *Études de linguistique appliquée*, No. 79, p. 133-138.

M

PUBLICATIONS ACTUELLES

Karl Johan Danell, professeur de français à l'Université d'Umeå, a publié un manuel de linguistique destiné à l'enseignement universitaire du français aux niveaux C et D. Ce livre remplit indiscutablement une lacune. Il se recommande non seulement par son contenu, mais aussi par la nouveauté de la méthode qui consiste à présenter les notions fondamentales de la linguistique moderne en partant de quelques problèmes grammaticaux concrets, à savoir les concurrences *an-année* et *pas de vin-pas du vin* et la construction *faire + infinitif*: *La linguistique. Pratique et théorie* (Studentlitteratur, 1993, 157 p.).

En France, trois professeurs de langues et littératures scandinaves – Jean-François Battail, Régis Boyer (les deux à la Sorbonne), Vincent Fournier (Bordeaux) – ont publié, à l'intention du public français, une grande histoire des *Sociétés scandinaves de la Réforme à nos jours* (Presses Universitaires de France, 1992, 596 p.), qui témoigne de l'intérêt croissant qu'on peut constater en France pour les pays scandinaves.

Au Danemark, une nouvelle grammaire universitaire du français a vu le jour. Elle est écrite par Gerhard Boysen, professeur de langues romanes à l'Université d'Aalborg. C'est la première d'une série de «grammaires romanes», série qui comprendra en outre des grammaires de l'espagnol, de l'italien et du portugais, toutes conçues selon les mêmes principes et toutes se servant de la même terminologie grammaticale: *Fransk Grammatik* (Munksgaard, 1992, 423 p.).

Olof Eriksson