

## A partir et autour de l'évaluation de l'oral

Pour la préparation du colloque organisé par l'Ecole de langue et de civilisation françaises de l'Université de Genève en décembre 1990 (voir la présentation) sur le thème: «L'évaluation de l'oral: transparence et cohérence?», cinq questions trivialement majeures avaient été posées, comme en exergue, aux différents intervenants:

- Que veut-on évaluer à l'oral?
- Que peut-on évaluer à l'oral?
- Quels critères et quels niveaux pour l'évaluation?
- Quelles relations entre une conception modulaire des objectifs et les certifications officielles?
- Transparence et cohérence pour qui?

Ce menu était sans aucun doute trop ambitieux pour une rencontre de deux jours, mais, en fin de parcours, il devenait intéressant de s'interroger sur les questionnements qui avaient effectivement retenu l'attention des participants, dans les contributions programmées aussi bien que dans les discussions auxquelles ces dernières avaient donné lieu. Les textes qui précèdent offrent aussi une image significative, dans les choix qu'ils font séparément et dans la sélection qu'ils constituent, pris ensemble.

Entre le colloque de 1990 et le présent recueil s'est encore inscrit le Symposium européen réuni à Rüschlikon en novembre 1991, où évaluation et certification ont été au coeur des débats. Les perspectives sont, dans ce dernier cas, beaucoup plus vastes et les enjeux, aussi bien pratiques que théoriques, d'une tout autre ampleur, mais certaines des lignes de force qui s'en sont dégagées ne paraissent pas aller à l'encontre des remarques qu'on pouvait faire à l'issue du colloque de 1990 et qui, avec application particulière à l'évaluation de l'oral, peuvent se résumer en quelques constats principaux.

### *1) Des objectifs devenus plus discrets?*

La question du «Que veut-on évaluer?» ne revient pas comme leitmotiv de premier plan. Alors qu'il y a une dizaine d'années le slogan majeur était qu'on ne pouvait dissocier objectifs et évaluation, la caractérisation de l'un des volets n'ayant de sens que dans la mesure où l'autre se trouvait également et aussi précisément défini, tout se passe désormais comme si la réflexion sur les objectifs paraissait moins incontournable que naguère. Certes,

dès lors qu'on focalise l'attention sur les modalités d'évaluation et les conditions de certification, le projecteur se trouve en partie déplacé, mais il reste que la spécification des objectifs ne passe plus, au-delà du rappel de principe, pour un préalable obligé et une condition sine qua non de bonne démarche.

Ceci méritant d'autant plus d'être souligné que les modèles naguère proposés pour la construction d'objectifs (par exemple celui qui, avec des variantes, a prévalu pour l'élaboration des niveaux-seuils) sont loin de faire l'unanimité et demanderaient plus que des révisions cosmétiques, à tout le moins pour d'autres étapes et d'autres finalités de l'apprentissage. Même s'il y a un relatif consensus sur la nécessité de repenser la manière dont on détermine et explicite les finalités d'apprentissage, le pendule s'est plutôt déporté vers la face évaluation du couple obligé. C'est l'inverse qu'on observait naguère, quand la caractérisation communicative des objectifs paraissait importer plus que la technicité d'une appréciation des résultats.

Il est par exemple assez remarquable que les études intervenues depuis lors sur la communication exolingue et la dissymétrie communicationnelle entre natif et alloglotte engagés dans une même conversation n'aient pas conduit à redéfinir plus profondément certaines des spécifications d'objectifs.

## 2) L'évaluable et l'évalué

La question du «Que peut-on évaluer?» tend, elle aussi, à être esquivée. Et là encore le constat ne va pas sans quelque paradoxe rétrospectif. On soulignait, au moment d'une insistance sur la capacité à communiquer, que l'aptitude à convaincre, le succès perlocutoire importaient plus, au bout du compte, que telle ou telle finesse grammaticale. On se demandait aussi ce qui, sur des terrains de cet ordre, relevait d'un apprentissage second et ce qui pouvait être conceptualisé en termes de transfert de compétences acquises en langue maternelle (ou relevant de la «personnalité» du locuteur). Or, il ne semble pas qu'on porte aujourd'hui un intérêt direct à ce que serait l'appréciation précise d'une force de conviction ou d'un effet perlocutoire.

De même, et ceci intéresse l'expression orale, gestualité et attitudes proxémiques ne semblent pas vraiment prises en compte dans la réflexion didactique sur l'évaluation (alors que ces dimensions ne sont pas absentes des entraînements fonctionnels que proposent des sessions de formation sur le mode «Apprenez à parler en public» ou «Triomphez de votre timidité» ou encore «Conduire efficacement une réunion»). De même encore, bien que tout le monde s'accorde pour admettre que la compétence encyclopédi-

que joue un rôle déterminant dans la communication linguistique (elle affecte par exemple la manière dont un apprenant en situation d'évaluation de l'expression orale abordera et gèrera les thèmes sur lesquels on lui demande de présenter un bref exposé ou de prendre part à une discussion), elle ne fait pas l'objet d'une considération approfondie.

## 3) Une compréhension sans problèmes?

Comme toujours par le passé, les capacités d'expression sont réputées d'appréciation plus délicate que celles de compréhension. Il fut un temps où le problème paraissait d'autant plus flagrant que, d'une part, on ne se préoccupait pas souvent d'évaluer la compréhension en tant que telle et que, d'autre part, quand on s'y intéressait, on cherchait souvent à la traquer par des épreuves de statut ambigu exigeant le recours à une maîtrise de l'expression. La situation a paru ensuite se décanter dans la mesure où la compréhension, tant écrite qu'orale, a donné lieu à travaux particuliers, mise au point d'épreuves diversifiées, plus facilement considérées comme valides et fiables que celles portant sur la production. D'où peut-être aujourd'hui l'impression qu'on y voit à peu près clair pour ce qui concerne les modalités d'évaluation de la compréhension mais que le bât continue à blesser pour la production.

Faut-il s'en tenir là ? Est-ce que la question est vraiment réglée ou est-ce que la compréhension est en passe de redevenir une fois encore le parent pauvre de la réflexion didactique, malgré des mises au point comme celle que présente le Bulletin CILA 53)? N'a-t-on pas donné une représentation par trop simplifiée de ce qui est mis sous «comprendre»?

Plus fondamentalement, doit-on, comme se le demandent NORTH et quelques autres, revoir le découpage canonique des quatre capacités et introduire plutôt une distinction entre interaction (activités communicatives exigeant du même sujet la mise en oeuvre et l'articulation de ses capacités de production et de compréhension), production (activités communicatives telles que la rédaction d'un rapport ou la présentation d'un exposé oral mobilisant avant tout les capacités d'expression du sujet) et compréhension (activités communicatives telles que la lecture ou la fréquentation des médias audio-visuels faisant appel surtout aux capacités de réception et d'interprétation des messages)?

## 4) Plus de transparence que de cohérence?

Dans les débats actuels, à l'initiative notamment de RICHTERICH et SCHNEIDER, la recherche d'un maximum de transparence et de cohérence est à

l'ordre du jour et est devenue un mot d'ordre unificateur. On doit évidemment d'autant plus s'en féliciter pour ce qui concerne l'évaluation que c'est, par excellence, un domaine où, quels que soient les courants méthodologiques dominants à tel ou tel moment de l'histoire, une certaine opacité et divers types de discrédit ont été de mise.

Cela dit, il convient de relever que, des deux termes, c'est celui de transparence qui paraît actuellement le plus récurrent. Sans doute parce qu'il faut que tout le monde (tous les partenaires et instances concernés) y voie un peu plus clair avant qu'on ne s'interroge plus avant sur les incohérences éventuelles. Mais peut-être aussi parce que la cohérence n'est pas une priorité aussi ardente dans un secteur où les pesanteurs et intérêts institutionnels sont complexes et lourds et où les usagers eux-mêmes (au premier chef les apprenants) s'accommodent à l'occasion fort bien, pourvu en effet qu'ils les connaissent, de certains défauts de cohérence.

#### *5) Des modalités pas tout à fait innocentes*

Si la concentration de fait s'opère, avec réalisme, sur les modalités concrètes de l'évaluation, à rendre plus transparentes tant dans le quotidien d'une classe que dans le cadre des grandes certifications donnant lieu à reconnaissance internationale, ces modalités elles-mêmes ne sont pas indifférentes ni innocentes. Elles disent en creux ce que l'on entend, de facto, évaluer et ce que l'on estime, de facto, évaluable. Quelques observations peuvent donc être faites à ce propos, sans surprise notable, mais probablement trop partielles:

a) Ce qu'on veut évaluer, par telle ou telle modalité, reste à dominante linguistico-linguistique, même au service de la communication. Il a été noté que, pour l'oral, le gestuel (voire l'intonation) n'entre guère en ligne de compte, lorsqu'il s'agit de mesure. De fait on ne cherche à évaluer que ce qu'on croit savoir / pouvoir enseigner. En un sens, cela relève de l'honnêteté déontologique. Il n'en résulte pas pour autant que cet état de fait soit pleinement satisfaisant à terme.

b) On est sorti des évaluations de simple «correction» et on a dépassé les oppositions binaires, un temps utiles mais bancales si on les prend à la lettre, du type «accuracy/fluency» (correction/aisance) ou «compétence linguistique/compétence de communication». On a plutôt essayé de multiplier les critères d'appréciation (comme ici, pour l'expression orale: l'exactitude, la correction, la fluence, l'adéquation communicationnelle, la qualité de la prononciation, etc). Cet affinement, souhaitable en soi, peut poser des problèmes de délimitation ou de recoupement entre les différents critè-

res. Il exige aussi qu'on décide, pour chacun de ces angles d'appréciation, quels sont les indices posés comme pertinents dans les productions à évaluer. Ce qui n'est pas toujours le cas.

c) Dans les faits, quand des indices réputés pertinents n'ont pas été mis en évidence, l'appréciation, même sur critères multiples, risque bien de se faire à partir de représentations pas souvent explicitées et non toujours partagées. X ne se fait pas nécessairement la même idée que Z de ce qu'est une réalisation langagière adéquate dans telle ou telle situation de communication orale, ni de ce qui atteste de la diversité des moyens linguistiques mobilisés par un apprenant. Ceci d'autant plus que l'oral (comme l'écrit, mais d'une autre manière) est un lieu où les représentations sociales sont fortement à l'œuvre et donnent lieu, constamment et avec des variations notables suivant les groupes sociaux considérés, à évaluation spontanée (ou naïve) continue, dans l'usage même de la parole entre natifs, aussi bien que dans la relation avec l'alloglotte. On évalue en fonction aussi des représentations qu'on a de l'objet à évaluer. Et, pour l'oral, il y a à la fois défaut de descriptions et excès de représentations. Seule une meilleure (et mieux partagée) connaissance descriptivo-théorique des discours oraux et écrits permettrait de complexifier et corriger les représentations qui en circulent habituellement, y compris chez les enseignants et évaluateurs que nous sommes. Des progrès ont été accomplis pour les formes de l'écrit. Ils sont moindres, mais non négligeables, pour ce qui concerne la diversité des productions orales.

#### *6) Diversifier les procédures d'évaluation*

Un des constats majeurs quand on s'interroge sur l'évolution récente des pratiques en matière d'évaluation, et par exemple pour ce qui concerne l'oral, est en effet une diversification plus marquée des procédures, des tâches communicatives proposées aux apprenants, des types et genres de productions textuelles qui servent de support à l'évaluation.

Dans la mesure, par exemple, où on distingue la capacité à interagir à l'oral de la capacité à construire un discours monologique long, on proposera ici un exposé à préparer en temps limité, là un échange en face à face avec ou sans temps de préparation. Pour éviter le déséquilibre et la tension qui peuvent s'établir dans une conversation entre examinateur et examiné, on pourra avoir recours à une discussion sur thème entre plusieurs apprenants ou à des échanges de messages (consignes, récit, description...) entre examinés ou groupes d'examinés, voire à diverses formes de jeux de rôles. L'enregistrement au laboratoire de langues de «bouts d'essai» pour une presta-

tion orale en forme n'est pas exclue, non plus que des exercices plus formels, où la diction et la prononciation seraient particulièrement sollicitées.

En la circonstance, ce qui importe est la pertinence relative d'une tâche proposée par rapport à ce que l'on entend mesurer et par rapport à ce que l'apprenant est censé savoir faire. Retour à la relation nécessaire entre objectifs et évaluation, retour aussi à une certaine interprétation de la notion de cohérence.

### 7) Une géométrie variable dans l'évaluation

La diversification des tâches et des modes d'évaluation prend d'autant plus de sens qu'elle peut s'accompagner d'une «géométrie variable» dans la prise en compte de divers critères d'appréciation des performances. Plutôt que d'évaluer une seule performance en fonction de X critères (voir plus haut), on peut fort bien envisager qu'une batterie d'épreuves autorise une répartition ou une pondération privilégiée de tel critère (par exemple la fluence relative de l'expression orale) sur telle tâche communicative (par exemple un exposé monologal). Cette variation dans le choix ou le mode d'application des critères permet aussi de mieux déterminer les indices qui seront considérés comme significatifs au regard du critère sélectionné. Encore faut-il alors que les règles (variables) du jeu soient claires pour tous (et singulièrement pour les apprenants). Retour à la transparence.

### 8) La nécessaire faisabilité

Encore faut-il aussi que cette diversification/distribution reste viable en termes de faisabilité à l'intérieur de conditions et d'un contexte donnés. Retour au coût et au temps de l'évaluation. Porter l'attention sur les modalités d'une évaluation fiable, comportant une bonne validité d'apparence et de contenu, diversifier les tâches et multiplier les critères, tout en renforçant la transparence et la cohérence de l'ensemble, voilà en effet qui paraît souhaitable. Mais on devra veiller à ne pas alourdir et raffiner les procédures à un point tel que, les principes étant respectés et les modalités techniques adéquates, on aboutisse à un dispositif ingérable ou inaccessible à la plupart des apprenants ou faisant ressembler le parcours d'apprentissage et l'éventuelle certification des résultats à une très longue course d'obstacles.

### 9) La grille et le cadre

Deux questions centrales semblent aujourd'hui se poser dans l'évolution en cours. Elles sont liées par plus d'un biais et il n'est pas étonnant que,

présentes dans les discussions du colloque genevois de 1990, elles aient constamment refait surface au cours du symposium de Rüslikon.

a) La première est celle des grilles ou échelles auxquelles on a recours pour situer les acquis ou le niveau des apprenants ou pour comparer entre elles différentes certifications existantes. Il est clair qu'on a besoin, notamment dans la perspective de circulations internationales accrues et d'une valorisation professionnelle et personnelle des acquisitions, de définir des cadres de référence, des tables d'équivalence, diverses formes d'étalonnage, qui, sans prétendre mettre fin à une diversité saine des réalités institutionnelles et des projets et cheminements individuels, permette d'en rendre raison, chacun (acteur institutionnel ou individuel) pouvant mieux savoir ainsi où il est et où il en est.

L'évaluateur, l'institution de formation et celle de certification, les instances professionnelles, les apprenants avant tout, ont besoin d'instruments référentiels les uns aux autres dès lors qu'ils peuvent difficilement être communs de part en part. Mais il faut veiller à ce que, pour des raisons de simplification, de coordination commode ou de compromis sur les apparences (ce qui serait tout le contraire de la cohérence), on n'en arrive pas un jour à se raccrocher vaille que vaille à des échelles de compétence répartissant tous les savoirs, savoir faire et capacités des apprenants sur un étagement procustéen à X ou Y niveaux.

b) La seconde est celle du cadre participationnel de l'évaluation. Pour que validité et économie de l'évaluation n'apparaissent pas comme antinomiques, mais aussi par principe et pour ne plus trop forcer sur d'autres dichotomies au demeurant nécessaires telles «évaluation formative/évaluation sommative» ou «auto-évaluation/hétéro-évaluation», il y a lieu sans doute de s'orienter dans une direction qui permettrait:

- que l'évaluation en cours d'apprentissage fasse aussi partie de l'évaluation terminale ou certifiante, de même que l'évaluation certifiante doit entrer dans un processus d'apprentissage continu;
- que l'auto-évaluation, l'évaluation par les pairs et l'hétéro-évaluation (interne ou externe au dispositif de formation) puissent opérer en complémentarité, selon des pondérations variables en fonction des lieux et des moments d'apprentissage;
- que les transitions entre systèmes, institutions de formation, cycles d'études, modes d'apprentissage soient mieux assurées (ce qui implique un minimum de points de rencontre et de recherche participative d'interfaces, pour gérer au mieux des différenciations qu'il n'y a pas nécessairement lieu de réduire).

C'est donc au moment où l'évaluation en langues, par un travail patient sur sa technique et ses modalités, se renforce en tant que domaine d'investigation exigeant une compétence particulière et donc une certaine forme de professionnalisation qu'il lui faut, dans le même mouvement, s'interdire tout nouveau cloisonnement de spécialité «pointue» et devenir ainsi, sans paradoxe excessif et comme elle aurait dû toujours l'être, l'affaire de tous.

Université de Genève  
Département de linguistique et  
Ecole de langue et de civilisation françaises  
CH-1211 Genève 4

DANIEL COSTE

### *Eléments bibliographiques*

- ANDENMATTEN, S., BRUDER, O., FAUCHERRE, A., LANGNER, M. SCHWARTZ, A. (eds.) (1991): *Verstehen im Deutschunterricht, Bulletin CILA. numéro spécial, 53.*
- BOLTON, Sybille (1987): *Evaluation de la compétence communicative en langue étrangère*, collection LAL, Hatier-Crédif.
- CANALE, M. (1981): «Communication: how to evaluate it?», in: *Bulletin de l'ACLA, 77-99.*
- CARROLL, Brendan J. (1980): *Testing Communicative Performance. An Interim Study*, Pergamon Press.
- CARROL, Brendan J., WEST, Richard (1989): *ESU Framework - Performance Scales for English Language Examinations*. Longman.
- HEATH, J.B. (1988): *Writing English Language Tests, New Edition*, Longman Handbooks for Language Teachers.
- HOLEC, Henri (1990): «Apprendre à l'apprenant à s'évaluer: quelques pistes à suivre», in: *Etudes de linguistique appliquée 80, 39-47.*
- MADSEN, Harold S. (1983): *Techniques in Testing*, Oxford University Press.
- NORTH, Brian (1991): «Un portfolio européen de langues: options pour une échelle de compétence», document présenté au Symposium du Conseil de l'Europe de Rüschlikon (10-16 novembre 1991).
- KOLLER, John W., PERKINS, Kyle (1980): *Research in Language Testing*, Newbury House.
- PORCHER, Louis (1990): «L'évaluation des apprentissages en langue étrangère», in: *Etudes de linguistique appliquée 80, 5-38.*
- RICHTERICH, René, SCHNEIDER, Günther (1991): «Transparence et cohérence: pour qui et pour quoi?», document présenté au Symposium du Conseil de l'Europe de Rüschlikon (10-16 novembre 1991).
- UNDERHILL, N. (1987): *Testing Spoken Language*, Cambridge University Press.