

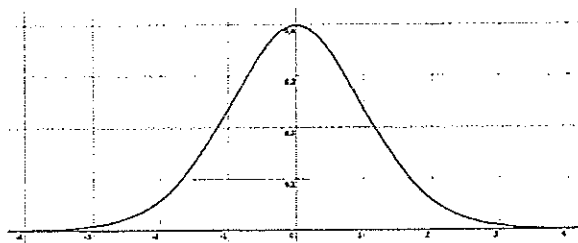
Quelques réflexions sur l'évaluation en langue: comment faire mieux ?

par Patrick Riba¹

1. Il n'y a pas de quoi se gausser

On attribue bien à tort au mathématicien Gauss (Carl Friedrich Gauß, 1777-1855) la volonté d'avoir modélisé le monde dans une répartition statistique des choses et des êtres, et l'histoire, toujours ingrate quand elle quitte le champ de l'épistémologue, ne lui rend pas grâce qui ne retient de lui que sa fameuse courbe.

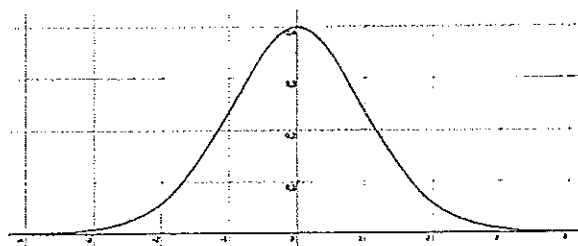
Ce mathématicien et physicien né à Brunswick, Allemagne, en 1777, a en effet joué un rôle fondamental dans le développement des mathématiques, de la géométrie euclidienne, de l'algèbre, de l'optique, de la géodésie (le gauss est aujourd'hui une unité de mesure des champs magnétiques) et de l'optique. Mais c'est en 1823 qu'il publia sa fameuse *Theoria combinationis observationum erroribus minimis obnoxiae* consacrée à la statistique et à la distribution normale des probabilités qui apparaissent fréquemment, en relevant que celles-ci se basaient sur un double principe de densité (dérivée de la fonction de distribution de la probabilité) et de propriétés mathématiques (espérance mathématique, coefficient de symétrie, etc.).



Exemple de courbe de Gauss

Le monde de l'éducation, les enseignants, nous-mêmes, utilisons beaucoup ce concept cartésien dans les classements explicites ou non que nous faisons de nos étudiants.

Nombre d'apprenants



Echelle de notation : note minimale

note maximale

¹ Directeur des cours, Centre culturel et de coopération - Institut français d'Amérique latine (IFAL), Ambassade de France au Mexique, priba@francia.org.mx

Il semble en effet naturel d'estimer que sur un groupe d'individus donnés un petit nombre se révèle incapable d'atteindre les objectifs annoncés, qu'une grande majorité y parvienne moyennement, et que seule une petite minorité puisse réussir parfaitement. On peut toutefois se poser des questions sur le caractère naturel d'un tel postulat qui, pour résumer, classerait l'humanité en un petit groupe d'incapables, une majorité médiocre et une petite minorité d'élite, et l'on doit s'en poser sur les conséquences sociales qu'il véhicule en termes d'égalité, de mobilité et de valorisation des individus.

Ramenée à l'échelle de notre salle de classe, cette approche théorique repose sur une volonté de classement des individus les uns par rapport aux autres dans ce qu'il est convenu d'appeler l'évaluation normative. De fait, la sélection de l'excellence est l'une des fonctions du système scolaire qui doit identifier et préparer les élites de la nation, ses cadres, scientifiques, éducateurs et dirigeants. C'est celle que le système applique dès qu'il met en place la règle du *numerus clausus* pour sélectionner les impétrants à des postes en nombre limité, à l'université ou à une « grande école », voire à une formation rémunérée. Elle n'est pas sans poser de problèmes, théoriques d'une part puisque le rapport aux 3 parties constatées de la répartition gaussienne renvoie toujours à des groupes limités (et Rosenthal – in Nimier, 2006 – a démontré le souci des enseignants de retrouver systématiquement ces 3 groupes d'individus alors même que le groupe total aurait été constitué d'étudiants issus eux-mêmes de 3 groupes « excellents » de 3 classes différentes), moraux d'autre part, car la sélection de l'excellence n'est pas la seule fonction de l'école qui doit également faire entrer la société dans une démarche de démocratisation du savoir.

L'évaluation assume alors de nouvelles fonctions de régulation, régulation des apprentissages, des enseignements et des systèmes éducatifs eux-mêmes, et la seule comparaison des individus les uns par rapport aux autres ne suffit plus.

2. Des évolutions notables

Les docimologues (l'histoire française retient, sans doute à tort le nom de Piéron comme fondateur de ce courant) ont depuis longtemps identifié des biais qui pervertissent la prétendue objectivité de l'évaluation et rendent inconsistantes les notes en tant que mesure de la performance des élèves (la fatigue qui nous rend plus « gentils » ou plus « sévères » à mesure que nous corrigeons un stock de copies, l'effet « goutte d'eau » qui nous rend intolérants à une même erreur entendue plusieurs fois de suite, l'effet de halo qui nous rend sensibles aux charmes non linguistiques d'un candidat, etc.).

Les sociologues, d'autre part, nous alertent depuis longtemps sur les effets des curricula cachées, celles du métier d'apprenant et des rapports à la norme. Bourdieu et Passeron ont ainsi démontré dans diverses études (1964, 1970) le côté bourgeois d'une culture humaniste qui convient aux enfants qui en sont les héritiers, mais qui déconcerte et défavorise celles et ceux qui viennent de classes sociales plus défavorisées.

Les didacticiens des langues, enfin, se sont peu à peu centrés sur une approche par objectifs programmés qui met en évidence la notion de compétence et son corollaire la performance. Comme le rappelle Rosen (in Beacco, Bouquet, Porquier, 2004 :17), Chomsky est l'un des premiers à avoir proposé un déplacement des préoccupations du linguiste, abandonnant l'étude centrale de la langue pour s'intéresser au locuteur : « *L'objet premier de la linguistique est un locuteur idéal, appartenant à une communauté linguistique homogène, qui connaît parfaitement sa langue, et qui, lorsqu'il applique en une performance effective sa connaissance de la langue, n'est pas affecté par des conditions grammaticalement non pertinentes, telles que la limitation de mémoire, distractions, déplacement d'intérêt ou d'attention, erreurs (fortuites ou caractéristiques. (...)) Nous établissons donc une distinction fondamentale entre la compétence (la connaissance que le locuteur-auditeur a de sa langue) et la performance (l'emploi effectif de la langue dans des situations concrètes) ».*

Pour Chomsky « la compétence d'un sujet parlant français est donc l'ensemble des possibilités qui lui sont données de construire et de reconnaître l'infinité des phrases grammaticalement correctes, d'interpréter celles d'entre elles qui sont douées de sens, de déceler les phrases ambiguës, de sentir que certaines phrases, éventuellement très différentes par le son, ont cependant une forte similitude grammaticale, et que d'autres, proches phonétiquement, sont grammaticalement très dissemblables » (Ducrot, Todorov, 1972 :158).

Cette notion sera précisée par l'approche pragmatique, issue des travaux de philosophes du langage tels que Austin ou Grice qui conceptualisèrent le concept d'actes de langage (utilisation du langage comme une action, et non plus seulement comme un message) et des actes de parole (l'énoncé effectivement réalisé par un locuteur déterminé dans une situation donnée) développant l'idée que le langage dans la communication n'a pas principalement une fonction descriptive mais une fonction actionnelle.

Puis Hymes élaborera une approche fonctionnelle de la communication bientôt complétée par les travaux de Bernstein sur les codes, de Halliday sur la syntaxe fonctionnelle ou de Labov en sociolinguistique, qui, sur la base de son étude sur la stratification sociale de l'anglais parlé à New York, établit une référence méthodologique qui lui permit d'identifier des phénomènes d'insécurité linguistique, ou d'hypercorrection, mettant en évidence les rapports de force sociaux que le langage représente et véhicule. Goffman, et avec lui les ethno-linguistiques, allaient compléter cette approche en développant à travers l'étude des conversations une analyse des implicites sociaux et l'émergence d'une grammaire des conversations et de l'analyse interactionnelle.

Charaudeau dira en résumé (2000 :36) : « il nous faut désormais tenir compte d'une compétence pragmatique qui nous oblige à regarder les faits de langage comme des actes tournés non point vers le monde pour en rendre compte (théorie représentationnelle) mais vers l'autre, pour le mettre en position d'avoir à penser et à faire (théorie actionnelle) ».

Threshold Level (Trim, Ek, 1975) constituera en ce sens la première manifestation de cette approche communicative qui inspire largement le format actuel des « référentiels pour le français » du Conseil de l'Europe. Les actes de parole y sont analysés en contenus propositionnels appelés notions et en rôles illocutifs appelés fonctions.

« Dans son essence, le modèle est très simple, parce qu'il analyse le comportement langagier en deux composantes seulement : la réalisation d'intentions énonciatives et l'expression de (ou la référence à des) notions. Ce qu'on fait en se servant du langage peut être décrit comme la verbalisation de certaines intentions. Au moyen de la langue, on affirme, on pose des questions, on donne des ordres, on fait des remontrances, on persuade, on présente des excuses, etc. En réalisant de telles intentions, on exprime certaines notions où l'on s'y réfère. On veut, par exemple, s'excuser d'être en retard, d'être en retard à une réception, d'être en retard à la réception qui a eu lieu hier, etc. » (Van Ek 1977 :5). On peut donc construire un catalogue d'objectifs qui contiennent les composantes suivantes :

- les situations dans lesquelles on se sert de la langue étrangère ;
- les activités langagières dans lesquelles s'engagera l'apprenant ;
- les intentions énonciatives que l'apprenant réalisera ;
- ce dont l'apprenant sera capable en ce qui concerne chaque sujet à traiter ;
- les notions générales que l'apprenant pourra manier ;
- les notions spécifiques que l'apprenant pourra manier ;
- les formes linguistiques dont l'apprenant pourra se servir ;
- le degré de maîtrise que l'apprenant pourra atteindre.

Van Ek, 1979 :105)

Cette approche théorique redonne au sujet actant sa véritable place et n'est pas sans conséquence sur les méthodes évaluatives anglo-saxonnes de l'époque. Le sujet reprend une place active dans le processus évaluatif. Qui parle à qui ? Dans quels

buts ? Peut-on se contenter d'évaluer les aspects linguistiques même s'ils sont plus faciles à isoler ?... Le même Van Ek (1973 in Bolton, 1987 :37) définit « la démarche nécessaire pour rendre les objectifs opératoires : fixer

1. le contenu de ce qu'il faut apprendre ;
2. le comportement qu'on attend du bon élève ;
3. les circonstances dans lesquelles ce comportement aura à se manifester ;
4. le critère d'acceptabilité de la performance ».

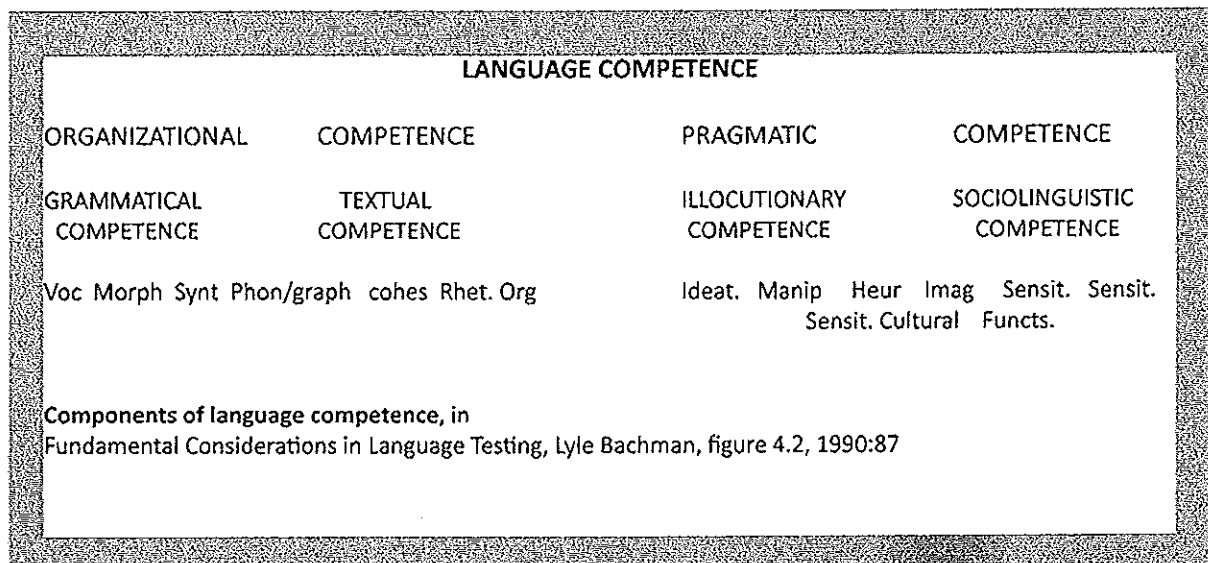
Morrow (1979) fut l'un des premiers à souligner la nécessité d'évaluer la performance dans des situations quotidiennes quand il affirma que les tests de langage devaient refléter les usages suivants du langage (in Brown, Hudson, 2002: 20) :

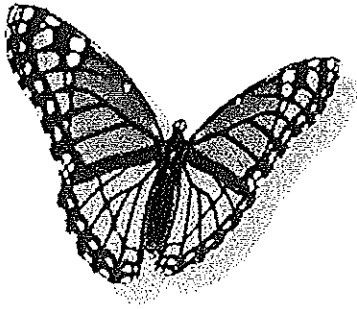
1. le langage est utilisé en interaction ;
2. l'interaction est généralement non prédictible;
3. le langage a un contexte ;
4. le langage est utilisé dans un contexte ;
5. il doit y avoir un besoin pour évaluer une performance ;
6. le langage est authentique et non simplifié ;
7. le succès du langage est basé sur des comportements.

Canale et Swain (1980) proposèrent une modélisation encore largement utilisée dans les tests de langues pour évaluer la compétence langagière qu'ils décomposent en trois parties :

- Compétence grammaticale
- Compétence sociolinguistique
- Compétence stratégique

Mais c'est Bachman (1990 :87) qui présenta la vision la plus homogène du concept d'habileté langagière de communication :





3. Un long chemin reste à parcourir

C'est donc désormais la compétence individuelle qui est prise en compte pour l'évaluation, indépendamment du groupe classe ou de toute autre comparaison.

Cette description du langage en une succession de traits latents présente pour l'évaluateur de nombreux avantages dans la mesure où elle met en évidence des éléments observables à condition de les spécifier selon les niveaux à atteindre. Le Cadre s'inspirera largement de cette théorie pour l'appliquer à un contexte de test avec une vision plus actionnelle « *dans le sens défini par Charaudeau, l'option majeure est de considérer qu'il n'y a pas là superposition ou juxtaposition de compétences toujours distinctes, mais bien existence d'une compétence plurielle, complexe, voire composite et hétérogène, qui inclut des compétences singulières, voire partielles, mais qui est une en tant que répertoire disponible pour l'acteur social concerné* » (Coste, Moore et Zarate (1998 :12), in Rosen (op ibid. : 21).

Il reste que si les systèmes éducatifs et les centres de langues développent aujourd'hui presque tous une approche basée sur un développement individuel de la compétence en langue, l'absence d'une théorie des tests adaptée aux nouvelles

finalités de l'évaluation se fait encore sentir, et l'évaluation des apprenants n'est pas toujours ni valide, ni fiable. Or, si l'évaluation peut être entendue comme un message ou un facteur motivationnel, elle est aussi une mesure sur la base de laquelle sera rendu un jugement aux conséquences sociales importantes.

« *La mesure est l'opération qui permet de passer de l'espace continu du Réel à l'espace discret du Connu* » (Perdijon, 2004 :2). Une mesure n'a de sens que si elle est consensuellement acceptée. « *Il ne saurait exister de vie sociale sans consensus sur une commune mesure* » (Perdijon, op ibid :1).

Lorsque je prends une règle pour évaluer la longueur d'un segment donné, le résultat de mon opération résulte en une quantification sur une échelle donnée et s'exprime par un nombre. Ce nombre, cette valeur, n'a de sens que s'il/elle, est accepté(e) par tous ; cela signifie que l'unité de mesure choisie ainsi que l'outil qui en découle doivent consensuels. Cela signifie également que le résultat obtenu est discriminant et qu'il me permet d'inférer une prise de décision. J'effectue donc une opération valide puisque je mesure ce que je prétends mesurer, à savoir en l'occurrence la dimension dudit segment d'une extrémité à l'autre, dans sa plus grande valeur, et l'information recueillie me permet d'estimer par exemple la possibilité de ranger ou non un livre dans une bibliothèque. C'est également une opération fiable puisqu'elle peut être répétée autant de fois qu'on le souhaite par autant de personnes possibles, et que l'on obtiendra toujours le même résultat. Ce sont ces deux qualités, validité et fiabilité, qui confèrent à la mesure sa valeur sociale.

La démocratisation de l'enseignement/apprentissage des langues et leur impact en termes de mobilité sociale et géographique nous imposent aujourd'hui plus que jamais une attitude éthique et juste qui ne peut plus ignorer ces deux concepts. Les certificateurs internationaux se dotent de méthodes chaque jour plus performantes pour les garantir (cf. par exemple www.alte.org), mais dans notre classe, dans notre institution, au quotidien nous ne pouvons les utiliser.

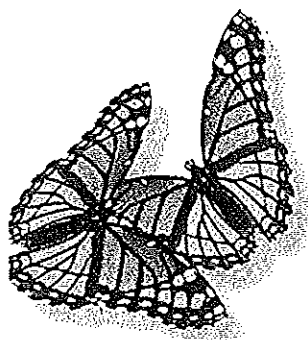
Nous avons cependant à notre disposition un

certain nombre de moyens simples pour rendre notre approche plus éthique.

1. La simple énonciation des objectifs programmatiques ne suffit pas et doit être intégrée dans un document collectif de l'équipe enseignante qui explicite les fondements théoriques et les modèles utilisés (perspective actionnelle...). Cette curricula, qui explicite les choix théoriques des différents éléments systémiques composant l'acte d'enseignement apprentissage (finalités, concepts théoriques, objectifs, méthodes, moyens...) est fondamental car on ne peut évaluer ce que l'on veut évaluer que si l'on est capable de l'expliquer.

2. L'évaluation des performances des apprenants ne peut être du ressort individuel de l'enseignant. Il doit travailler en équipe et s'assurer du respect collectif de standards d'évaluation. Une même production écrite et orale devrait ainsi toujours obtenir la même note, quel que soit l'évaluateur.

3. L'évaluation des performances des apprenants doit se faire dans le plus grand respect de leur intégrité et pour leur plus grand bénéfice. Une charte établissant les procédures d'évaluation et de notation devrait ainsi garantir l'utilisation de l'évaluation au profit de chaque apprenant : explicitation préalable des critères d'évaluation, droit à l'erreur et à la remédiation pour l'apprenant, notation valorisante, évaluation argumentée, droit d'appel de la note, etc.



4. Évaluer, ça s'apprend

En s'ingéniant à trouver la note vraie par une série de calculs pondérateurs, les docimologues français ont mis en évidence des biais induits de l'évaluation, définissant ainsi les prémices du champ de la fiabilité. Mais cet apport était dévoyé par le concept même de véridicité absolue qu'ils prétendaient défendre, et ce pour plusieurs raisons. La première fut rapidement perçue par la communauté enseignante et elle explique sans doute le rejet global de la docimologie par les enseignants. Les traitements statistiques imposés aux notations des enseignants relevaient du principe de l'étalonnage inspiré de la métrologie. En se préoccupant davantage du comment évaluer plutôt que du quoi évaluer, les docimologues imposèrent des calculs aussi complexes que vains pour atteindre une notation objective dont tout le monde pressentait la vacuité. Ce faisant, ils réduisaient également l'évaluation à une simple mesure où les concepts de constance et de mobilité prenaient une place prépondérante. L'enseignant devenait ainsi une sorte de métreur/géomètre soumis à des procédures « scientifiantes » qu'il subissait sans en être véritablement partie prenante.

La mise en évidence par la sociologie de l'impact de reproductibilité du système scolaire conduisit, en pleine contestation étudiante, à la recherche de formes alternatives d'enseignement où l'évaluation aurait été amenée à disparaître, et les apports de la docimologie furent vite relégués au rang des méthodes à oublier. De nouveaux concepts apparurent qui rendaient encore plus complexe toute recherche en évaluation, en particulier avec le passage progressif de la transmission de savoirs à la construction de compétences qui accompagna un changement apparent du statut de l'enseignant. Il s'ensuivit une profonde modification des relations de pouvoirs entre maître et élève dont l'évaluation était sou vent le nœud modal.

La situation aujourd'hui reste paradoxale. L'évaluation joue, quoi qu'on en dise, un rôle toujours aussi prépondérant dans la vie éducative. Mais si l'école sélectionne encore et toujours, explicitement et/ou implicitement, elle tente aussi de ré-

guler son propre fonctionnement, d'orienter et de remédier. Ainsi, au comment et au quoi évaluer, l'évaluateur doit ajouter le pourquoi. En reprenant les concepts de « validité » et de « fiabilité » nous ne nous inscrivons plus dans les traces à moitié effacées des docimologues, et nous n'aspérons pas à donner à l'enseignant les méthodes quantitatives qui assureront la qualité globale du système d'évaluation qu'il propose, car l'évaluation en langue n'est pas une simple mesure. Nous les invitons à avoir une réflexion collective et systémique qui devrait permettre une mise en cohérence générale du système au-delà de ses effets directs sur la qualité du processus évaluatif lui-même.

Mais si l'on peut donc louer cette recherche de justice et d'équité, comment d'ailleurs ne pas la louer, on peut aussi en craindre les conséquences, et nous rappellerons à titre de conclusion provisoire que s'il est bon de n'évaluer que ce que l'on a enseigné, il est sans doute beaucoup moins bon de n'enseigner que ce que l'on va évaluer.

Références bibliographiques

- Bachmann, Lindenfeld, Simonin, Langage et communications sociales, LAL, Hatier-Credif, Paris, 1981
- Bachman, L.F., Fundamental Considerations in Language Testing, Oxford University Press, Oxford, 1990
- Béacco, Bouquet, Porquier, niveau B2 pour le français, un référentiel, Didier, Paris, 2004
- Béacco, Bouquet, Porquier, niveau B2, textes et références, Didier, Paris, 2004
- Béacco, Porquier, niveau A1 pour le français, un référentiel, Didier, Paris, 2007
- Béacco, Lepage, Porquier, Riba, niveau A2 pour le français, un référentiel, Didier, Paris, 2008
- Bourdieu, P., Passeron, J.C., les héritiers, Edition de Minuit, Paris, 1964
- Bourdieu, P., Passeron, J.C., la reproduction, Editions de Minuit, Paris, 1970
- Brown, G., Listening to Spoken English, Longman, Harlow, 1990
- Canale M, Swain M, Theoretical Bases of Communicative Approaches to Second Language and Testing, Applied Linguistics, vol 1, n°1, 1980
- Charaudeau, P, de la compétence sociale de communication aux compétences de discours », in Didactique des langues romanes, L Collès et Al (dir), De Boeck-Duculot, Bruxelles, 2000
- Coste, Moore, Zarate, compétence plurilingue et pluriculturelle, Le français dans le monde n° spécial recherches et application « apprentissage et usage des langues dans le cadre européen, Paris, 1998
- Ducrot, O, Todorov, T, dictionnaire encyclopédique des sciences du langage, Essais coll, éd du Seuil, Paris, 1972
- Morrow, K, Communicative language testing : revolution or evolution ? in Brumfit & Johnson (eds) The Communicative Approach to Language Teaching. Oxford University Press, 1979
- Perdijon, J, la Mesure, histoire, science et philosophie, Dunod, Paris, 2004
- Rosen, Evelyne, étude relative à la répartition des compétences et des exposants linguistiques, in Beacco, Bouquet, Porquier, « niveau B2 pour le français, textes et références » Didier, Paris, 2004
- Van Ek, J.A., The Threshold level, Conseil de l'Europe, Strasbourg, 1975.

