

LE REFERENTIEL D'EVALUATION, UN OBJET DE RECHERCHE COMPLEXE

Christian Bélisson

Laboratoire CIVIIV, Université de ROUEN
christian.belisson@jeunesse-sports.gouv.fr

Mots-clés : référentiel d'évaluation, obstacle épistémologique, critère, analyse des discours

Résumé. Le référentiel est un objet complexe. Il est difficile de l'appréhender sans une approche phénoménologique et constructiviste. Mais si on occulte la démarche empirique, le débat s'enferme régulièrement dans des tautologies. C'est le risque des modélisations prescriptives si elles ne sont pas confrontées à des approches descriptives. Dans les recherches sur l'évaluation, l'empirisme a longtemps fait référence aux traitements statistiques, au moins jusqu'à l'apparition de l'ethnographie scolaire. Mais l'une et l'autre de ces méthodes sont peu adaptées à un objet comme le référentiel qui est une construction conceptuelle. L'analyse de discours offre une perspective pour interroger les critères indépendamment des indicateurs opérationnels et les processus cognitifs mis en jeu, en d'autres termes pour étudier les processus de la référence.

1. Quelques questionnements épistémologiques :

Le colloque de l'AREF nous propose, cette année, de traiter de trois questions fondamentales pour les sciences de l'éducation : la première concerne les enjeux épistémologiques, conceptuels et méthodologiques, la seconde les objets délaissés en raison des paradigmes dominants, la troisième la dimension contextuelle, historique et culturelle qui a favorisé la construction de l'objet de recherche. Bien entendu, ces questions sont génériques et ne sont que des fils directeurs. Mais l'exercice n'est pas simple. Sa difficulté ne reflète-t-elle pas, d'ailleurs, la complexité de la recherche dans nos disciplines ? Comment aborder ces questions sans une approche théorique ? Mais on s'éloigne alors de la recherche empirique, et on tombe dans certains travers : illusion globalisante, confusion conceptuelle, tautologies, etc. Si, au contraire, on approfondit la méthodologie, en raison de la complexité de l'objet, celle-ci s'avère lourde, difficile à résumer, et l'aspect technique de l'appareillage prend alors le pas sur l'accessibilité de la communication. Elle est alors reléguée dans un coin (sur un affichage ou dans les annexes d'une thèse).

Malgré l'espace réduit de cette communication, on a relevé le défi : articuler ces questions théoriques et l'exposé d'une expérience, sur un objet complexe, les processus cognitifs de référentialisation lors des évaluations. Dans un premier temps, on abordera quelques questions épistémologiques, propres à nos disciplines, suffisamment discutées par ailleurs pour n'être évoquées ici que de façon succincte. A l'appui de cette discussion, on s'inspirera des travaux de J.L. Le Moigne (1995) et de T.S. Kuhn (1962/1983) pour soulever quelques questions fondamentales. Dans un second temps, on évoquera le champ problématique qui a conduit, au 20^{ème} siècle, à définir le référentiel comme objet de recherche. Ces paragraphes serviront ainsi d'introduction aux questions méthodologiques qu'on approfondira dans la seconde partie.

1.1 Approches prescriptive et descriptive :

Le débat a été assez vif, au sein des sciences de l'éducation entre les approches prescriptives et descriptives. C'est en septembre 1985 que J.M. De Ketele, avec J. Cardinet et M.C. Dauvisis (J.M. De Ketele ; 1986) organise un colloque au cours duquel celles-ci sont confrontées. Ce débat n'est pas nouveau. L'approche prescriptive a jalonné le parcours de la réflexion pédagogique, au fur et à

mesure que l'enseignement s'est affirmé comme pratique sociale, et ceci avec d'autant plus de vigueur quand celle-ci s'est institutionnalisée au 19^{ème} siècle. L'approche descriptive a émergé, à partir de ces préoccupations, mais elle s'en est distanciée sur plusieurs points. Sans prétention à l'exhaustivité, on notera :

- La mise en place de méthodes systématiques d'observation ;
- Le souci d'apporter des connaissances utiles aux praticiens, pour leurs pratiques ;
- La construction de modèles adaptés pour interpréter les situations, etc.

Le positionnement des chercheurs en sciences sociales s'est ainsi précisé : répondre aux préoccupations de la société, tout en définissant des champs disciplinaires autonomes. E. Durkheim et J. Piaget illustrent bien ce type de parcours. En tant que pratique sociale, le référentiel d'évaluation, s'est ainsi généralisé, et son étude en est à sa phase prescriptive, avec des synthèses bien construites et argumentées de certains professionnels (par exemple, B. Porcher, C. Letemplier & C. Rak ; 1992). Première remarque : ces nouvelles pratiques correspondent à une évolution de la société contemporaine, la démocratisation de l'enseignement qui a rendu celui-ci accessible à tous les jeunes ; mais elles ont aussi été induites par des travaux scientifiques (ceux de la docimologie), qui ont mis en valeur les écarts importants de notation sur les mêmes productions d'élèves. Cette double dynamique n'a-t-elle pas induite des confusions entre les deux types de modélisation, descriptive et prescriptive ? Lorsque B.S. Bloom et ses associés proposent leurs taxonomies en 1956, l'ont-ils fait en tant que praticiens (c'est à dire en tant que professeurs d'université) ou en tant que chercheurs, c'est à dire à partir d'observations scientifiques ? Certes, tous les participants à ces travaux étaient des universitaires, mais leurs préoccupations étaient avant tout pragmatiques, et non axées sur la définition d'un objet de recherche. Par ailleurs, la transposition de ces taxonomies, à vocation universitaire, dans les lycées, les collèges et à plus forte raison, l'enseignement élémentaire et maternel, pose quelques questions. Cela ne remet pas en cause leur valeur et leur intérêt. Mais ces taxonomies n'ont pas été conçues, a priori, comme des modèles scientifiques, plutôt comme des propositions pédagogiques/didactiques pour évaluer des étudiants.

1.2 Quelques repères épistémologiques pour poser la problématique :

Le référentiel est un objet difficile à cerner. Le traitement des données collectées et l'interprétation des situations sont d'autant plus complexes que les objets de l'évaluation sont aussi des catégories de notre pensée : critère, indicateur, référence, etc. La confusion entre ces deux niveaux conduit aisément à des tautologies et à discourir sans parvenir à conceptualiser objectivement (à travers des attributs observables) les éléments des systèmes soumis à l'étude. Pour sortir de ces cercles vicieux, j'ai fait appel à quelques réflexions et modèles épistémologiques. A commencer, bien entendu, par celui de T.S. Kuhn, qui dans sa postface de 1969 (1962/1983), précise le concept de « paradigme » scientifique. Il définit ainsi « les éléments constitutants » de la « matrice disciplinaire », qui « forment un tout et fonctionnent ensemble » : « les généralisations symboliques (...) expressions employées sans question ou dissension par les membres du groupe », « la partie métaphysique (...) le fait d'adhérer collectivement à certaines croyances (...) le fait de croire à certains modèles particuliers », « les valeurs, elles sont en général plus partagées (...) Elles contribuent beaucoup à donner aux spécialistes des sciences de la nature, comme ensemble, le sentiment d'appartenir à un groupe » et « les exemples (...) solutions concrètes de problèmes » (1969/1983 : 248 à 255). A partir d'une réflexion globale sur la notion de système (Durand ; 1979), il m'a semblé intéressant de requalifier ces catégories : « modalités d'expérience et d'observation » (MEO) pour les « exemples » ; « modélisations cognitives et prévisionnelles » (MCP) pour les « croyances » et « prédictions » ; « relations formelles entre concepts » (RFC) pour les « généralisations symboliques » ; « critères de validité » (CV) pour les « valeurs », du moins celles qui ont une valeur intrinsèque à la discipline – les « valeurs sociétales » ayant un autre statut, déontologique. Bien entendu, ces nouvelles appellations induisent une conception différente de la notion de « paradigme » : elles ont pour intérêt d'identifier plus précisément les éléments du système matriciel et leur imbrication, mais surtout d'appréhender ce qui relève de la modélisation prescriptive, la réflexion propédeutique qui nous conduit à l'élaboration de nos méthodes (MEO) et ce qui relève de la modélisation descriptive, la réflexion heuristique qui nous

conduit à découvrir les modes de fonctionnement de notre objet (MCP). Les relations entre concepts (RFC) et les critères (CV) articulent les rapports logiques entre ces deux types de modélisation, de deux façons différentes : les premières (RFC) expriment, de façon algorithmique, les rapports entre les éléments du sous-système-objet (MCP) à partir des observations réalisées grâce au sous-système-expérience (MEO) ; les secondes (CV) justifient l'utilisation du sous-système-expérience (MEO) pour apporter des connaissances sur le sous-système-objet (MCP). Il n'y a rien, ici de particulièrement original, si ce n'est une clarification du rapport entre les deux modélisations. Il reste seulement à définir, de façon plus précise, les critères de validité adaptés à notre objet, le référentiel, et surtout les modes de relations observables entre les éléments du système.

Pour ce faire, je me suis inspiré des recherches philosophiques - historiques de J.L. Le Moigne (1995), qui a analysé les divergences entre les approches positivistes et constructivistes : les premières privilégient une méthode analytique et une aperception ontologique et déterministe des réalités objectives, les secondes une conception phénoménale de celles-ci et une méthode systémique. (N.B. : la conception de l'objet est nommée par l'auteur « hypothèse gnoséologique »)

Principes méthodologiques	Modélisation analytique Raison suffisante	Modélisation systémique Action intelligente
Hypothèses gnoséologiques		
Ontologique / Déterministe	I - Paradigmes des épistémologies positivistes et réalistes classiques	III
Phénoménologique / Téléologique	II	IV - Paradigmes des épistémologies constructivistes

Figure 1. Modèle paradigmatique de J.L. Le Moigne (1995 : 119)

Ce « produit cartésien » entre les hypothèses et les principes, pour reprendre l'expression de l'auteur, induit une réflexion propédeutique sur la construction de nos outils d'investigation, tant en termes de méthode que de conception de l'objet. A partir de ce schéma, j'ai appréhendé certaines propositions prescriptives (MEO) et descriptives (MCP) de quelques auteurs, dans diverses disciplines, selon les quatre grands axes paradigmatiques suivants :

- I - Les approches positivistes : Descartes, Kant, Comte, Durkheim, etc. ;
- II – Les approches phénoménologiques : Freud, Husserl, Schutz, etc. ;
- IV – Les approches constructivistes : Piaget, De Saussure, Mead, Lévi-Strauss, etc. ;
- III – Les approches cognitivistes : psychologie cognitive, ethnométhodologie, connexionnisme de Varela, etc.

Il ne s'agit pas, ici, de s'engager dans une étude historique et philosophique, mais seulement de poser quelques repères, de discerner certains grands principes (CV), des modes opératoires (MEO), des relations significatives entre phénomènes (RFC) afin d'en déduire des modélisations de l'objet (MCP). Identifier les ruptures épistémologiques et utiliser les différents modes d'approche des objets est un moyen, en passant de l'un à l'autre, de se distancier et d'interroger à chaque moment nos résultats et nos conceptions. Ainsi, les interprétations sont discutées au regard des méthodes qui ont conduit aux résultats (phénoménologie), les concepts sont opérationnalisés pour être observables (positivisme), les relations structurelles sont appréhendées en rapport avec la signification qu'elles ont pour les acteurs (constructivisme), etc. La démarche est prescriptive (à finalité décisionnelle), elle a pour but d'orienter la recherche (l'élaboration des MEO). Dans le cadre de mes recherches, on peut discerner deux orientations majeures induites par cette propédeutique 1) Elle a permis de structurer le champ problématique et de préciser l'intérêt des différents travaux invoqués pour définir les objets de recherche. Cet aspect est brièvement abordé au paragraphe suivant (&. 1-3). 2) Elle a introduit une réflexion transversale sur les disciplines sollicitées pour les étudier et les observer, en particulier la sémiolinguistique et la sociologie.

Le tableau suivant synthétise ce second aspect.

Epistémè privilégié	Principe	Méthodologie	Sémio-linguistique	Sociologie
Phénoménologie	Réflexion sur les processus - analyse des phénomènes intuitivement perçus - et des catégories utilisées pour appréhender l'objet	Analyse des raisonnements logiques mis en œuvre par les acteurs	Analyse des processus de référenciation et référentialisation et des systèmes de signification	Approche et concepts ethnographiques pour aborder les modes de communication sociale des acteurs
Constructivisme	Dégager les structures - des modèles cognitifs de l'objet - des cadres de référence pour étudier l'objet	Etude structurale des conceptions des acteurs et de la genèse de leur formation.	- Les principales structures du discours - Les modalités de positionnement dans les situations d'énonciation	Formation des conceptions sociales Représentations sociales lors des situations de communication
Connexionnisme	Générer de la signification en rapprochant - les modélisations de l'objet - les méthodes de recherche appliquées	Etude des phénomènes les plus significatifs : - processus cognitifs généraux - caractères stat. significatifs	Récurrence des cooccurrences et des structures syntaxiques Etude des contextes de l'énonciation	Traitements statistiq. corrélationnels entre variables sociolog. et variables sémiolinguistiques pour identifier les porteurs de conceptions
Positivisme	Discerner des éléments : - observables dans les matériaux collectés - fonctionnels pour expliquer l'objet	Analyse systématique des objets / attributs et traitements statistiques	Analyse syntaxique et sémantique des unités du langage (APD, AAC)	Traitements statistiques adaptés à l'expérience et en fonction de l'hypothèse

Figure 2. Opérationnalisation des différents paradigmes

On peut ici envisager la transversalité à plusieurs niveaux, en particulier entre l'étude des structures du discours et de la référenciation et les contraintes de la communication sociale, ou encore entre l'utilisation des méthodes sociologiques et l'observation systématique des contextes de l'énonciation des discours, ou encore entre l'étude systématique des unités du langage et les analyses statistiques (par exemple, certains logiciels comme ALCESTE), etc. Ainsi, problématiques et objets de recherche (observés systématiquement) peuvent s'articuler.

1.3 Les problématiques de l'évaluation, en sciences de l'éducation :

En 1936, la commission Carnegie utilise les méthodes statistiques pour questionner l'équité des notes d'examens, et analyser les écarts sur une même production écrite. Ces travaux sont repris par H. Pieron (1963), puis par J.J. Bonniol (1972) et par G. Noizet et J.P. Caverni (1978). Ces deux derniers auteurs introduisent le débat sur les critères : « ce qui intéresse fondamentalement une psychologie de l'évaluation, et donc un modèle de l'évaluateur, c'est le cheminement qui conduit à la notation. Autrement dit, il s'agit de décrire l'évaluation comme un processus et de préciser quelles sont les causes susceptibles de l'influencer » (1978 : 136). En 1956, les taxonomies sont mises au point par des universitaires américains : il s'agit de classer les objectifs de l'enseignement, pour étudier les façons d'évaluer les apprentissages, voire de constituer des banques de données d'outils. On peut y voir les prémisses, voire l'origine, de la notion de « référentiel ». Mais la démarche est essentiellement prescriptive, le modèle décisionnel. Il faut attendre les travaux de J.M. Barbier (1985 et 1991) pour voir s'introduire la rupture phénoménologique avec une analyse des processus mis en œuvre par les acteurs, soit au cours de l'évaluation (1985), soit lors de la construction de projets (1991). A la même époque, à partir d'une approche microsociologique, P. Perrenoud (1984) étudie les rapports dans les classes entre

les enseignants et les élèves et la façon dont se fabrique, souvent dans l'implicite, la notion « d'excellence scolaire ». Les travaux de P. Merle (1996) apportent un point de vue complémentaire, en analysant les contraintes et les modes d'adaptation du jugement professoral au contexte. Enfin, la didactique (G. Brousseau, Y. Chevallard) propose l'étude des processus cognitifs, lors des interactions entre les élèves et l'enseignant, mais sans approfondir la question de l'évaluation, si ce n'est en introduisant la notion de « contrat didactique », apport non négligeable.

2. Les démarches empiriques et la fonction de la recherche :

2.1 Des problématiques aux hypothèses :

Les travaux exposés dans le paragraphe précédent soulèvent diverses problématiques, voire des paradoxes, qui ont orienté mes hypothèses de recherche.

1. Si les systèmes de notation sont « arbitraires », pourquoi les maintient-on ? Ou, en d'autres termes, quelles sont les contraintes organiques, les forces institutionnelles ou les habitudes des praticiens, qui en favorisent le maintien ?
2. Le modèle des taxonomies ont apporté des propositions pour résoudre cette contradiction. Les référentiels se sont développés, dans le but de pallier aux écueils révélés par la docimologie. Mais les expériences de P. Merle montrent que l'adoption de barèmes ne modifie guère les écarts de notation entre correcteurs (1998 : 11-13). Certes, un barème n'est pas un référentiel, mais ces résultats ne sont-ils pas aussi valables pour les référentiels, qui restent assez généraux dans leur définition des apprentissages à évaluer ?

Ces questions ouvrent la voie à diverses hypothèses plus ou moins contradictoires. Pour P. Perrenoud (1998 : 36), par exemple, une des raisons du maintien de l'arbitraire des évaluations est une méconnaissance des acteurs. Mais cette explication constitue-t-elle une raison suffisante à notre époque où les travaux docimologiques ont reçu une large publicité ? Cet auteur, par ailleurs (1998 : 31-35), analyse de façon beaucoup plus convaincante, les raisons organiques qui conduisent les institutions à entretenir le « flou » sur les critères d'évaluation. Il étudie aussi les contraintes qui conduisent les professionnels à privilégier des modes d'évaluation traditionnels (1998 : 74-86). Une nouvelle hypothèse est cependant plausible : l'évaluation n'est pas totalement arbitraire. Il existe un référentiel implicite au sein du milieu professionnel qui conduit à certaines convergences des points de vue. Il existerait alors un aléa important du système de notation (mis en valeur par la docimologie), mais aussi des convergences non explicitées. En particulier, les observations de P. Merle (1996), de P. Perrenoud (1998) et des didacticiens ont mis en valeur un ensemble de critères implicites communs, induits par la situation didactique et les exigences intrinsèques au contrat didactique : travail régulier, ponctualité, respect des consignes de travail, modes de communication avec l'enseignant, etc. Mais n'y a-t-il que ces critères comportementaux ? P. Merle (1998 : 14, 15) cite les travaux d'E. Chatel, à partir de statistiques appliquées aux résultats scolaires français, qui montrent que les élèves dont les résultats sont satisfaisants toute l'année, réussissent de façon très significative au baccalauréat dès les premières épreuves, alors que la réussite est beaucoup plus aléatoire pour les élèves moyens. Or ces épreuves sont anonymes. L'expérience qu'on proposera ci-dessous apporte une nouvelle contradiction.

2.2 Une collecte de données sur le jugement de productions écrites :

Le protocole de cette expérience s'inscrit dans la logique des recherches de la docimologie : faire produire des stagiaires en situation de formation, puis communiquer leur production écrite à des examinateurs et comparer les réponses de ces derniers. Mais, la demande formulée à l'encontre des ceux-ci a été différente de la tradition docimologique, puisqu'on ne leur a pas demandé d'attribuer une note, mais d'identifier dans le groupe la prestation la « meilleure » et la plus « insatisfaisante ». Dans le cadre d'une séance de formation ordinaire, au début d'un stage BAFA, les stagiaires ont décrit leur représentation de leur future mission : « quel est, pour vous, le rôle de l'animateur ? ». Parmi eux, une stagiaire avait deux ans d'expérience professionnelle ; deux autres ont éprouvé des difficultés, une n'a pas eu son stage validé, l'autre

avec des réserves. On qualifiera ici les réponses écrites par les stagiaires de « lettres ». On a communiqué ces onze lettres à 22 directeurs de centres, professionnels de l'animation ou formateurs. Le but était d'analyser leur perception intuitive de ces représentations langagières : quel est le sens que confère l'interlocuteur expérimenté à la représentation proposée par l'animateur en formation ? Il était demandé à ces professionnels d'identifier le « stagiaire professionnel » et celui dont le « stage est insatisfaisant ». Afin d'éviter d'introduire certains biais, les écrits ont tous été dactylographiés et les fautes d'orthographe corrigées. Seule la syntaxe a été préservée telle qu'elle. Par ailleurs, aucune indication sur les candidats n'a été évoquée, en particulier sur le sexe et l'âge. Bien que les réponses divergent, elles se concentrent toutes, sans exception, sur quatre représentations pour qualifier la représentation « professionnelle » (lettres N° 2, 6, 7 et 12) et sur quatre autres pour qualifier la représentation jugée « insatisfaisante » (lettres N° 3, 5, 8 et 11). Quand ils hésitaient, les enquêtés donnaient deux réponses. Pour la représentation « professionnelle », il n'y a eu que quatre doubles réponses, dont deux en deuxième choix. En revanche, il y a eu 39 réponses pour choisir le candidat « insatisfaisant », soit 17 réponses en sus du premier choix. Les deux lettres qui se disputaient le rejet, étaient les deux stagiaires qui ont eu des difficultés, celle dont le stage a été insatisfaisant (N° 3) et celle qui a eu des réserves (N° 8).

Lettres N°	1	2	3	4	5	6	7	8	10	11	12	
"professionnelle"	0	5	0	0	0	14	3	0	0	0	4	26
"insatisfaisante"	0	0	14	0	7	0	0	13	0	5	0	39

Figure 3. Résultats de l'expérience

Nous constatons que jamais un choix « insatisfaisant » n'a recoupé un choix « professionnel ». Certes les effectifs sont faibles au regard de la tradition statistique et l'expérience demande à être généralisée. Mais l'application du t de Student nous conduit à estimer que la probabilité pour que l'hypothèse nulle soit infirmée est bien supérieure à 999‰, pour les lettres N°6, N°3 et N°8. Ainsi, à partir de six à dix lignes écrites, des professionnels de l'animation ont jugé les représentations de stagiaires « animateurs », et sans jamais les avoir rencontrés, ont répondu dans le même sens qu'une équipe de trois formateurs qui les ont suivis, observés et évalués pendant huit jours. L'organisation syntaxique du langage a donc bien une signification, dans ce milieu professionnel, qui privilégie certaines représentations du « rôle de l'animateur ».

2.3 La relativité des fonctions mathématiques :

Ces résultats ne sont pas forcément contradictoires avec ceux de la docimologie. S'il y avait eu notation, il est fort probable que des divergences conséquentes seraient apparues, ne fut-ce qu'en raison des différences de choix entre examinateurs qui ont privilégié tel ou tel stagiaire. L'analyse des appréciations, exposée au paragraphe 3, montre que les positionnements n'ont pas été tout à fait identiques entre les évaluateurs. Par ailleurs, les discussions qui ont eu lieu entre les professionnels à l'issue de l'expérience ont fait état de certains désaccords assez vifs. Il y a donc bien à la fois convergence des appréciations, mais aussi des divergences significatives.

Mais ces résultats contradictoires entre la docimologie et l'expérience ci-dessus ont surtout pour intérêt de nous interroger sur nos méthodes de traitement des données. Selon la façon dont on pose les questions aux examinateurs et selon l'application des fonctions statistiques aux données de l'expérience, il est possible de produire des interprétations différentes qui privilégient soit les divergences (docimologie), soit les convergences (ci-dessus) – le verre à moitié vide ou le verre à moitié plein. La docimologie a privilégié l'analyse sur les écarts à la moyenne, à partir de la courbe des notes, selon la « loi normale » : elle a donc survalorisée la fonction des « écarts » dans ses interprétations. Dans l'expérience ci-dessus, c'est l'écart-type réduit calculé à partir des pourcentages moyens, qui a été exploité pour infirmer l'hypothèse nulle. Par exemple, la probabilité moyenne pour que la lettre N°6 soit citée est de 2,36 fois, soit 9% de probabilité. Or elle est citée 14 fois (54% des réponses), soit un écart de 11,62 points par rapport à la moyenne. La probabilité pour que cet écart soit dû au hasard est inférieure à 1‰. En d'autres termes, en docimologie, les réponses sont considérées convergentes - dans la « norme » - lorsque les notes se

rapprochent de la moyenne de la courbe « normale », alors que les choix sont considérés convergents dans l'étude ci-dessus, d'autant plus que leur probabilité d'apparition s'éloigne de la moyenne « normale ». Le choix de la fonction mathématique (statistique) est étroitement lié à la fonction téléologique de la recherche – c'est-à-dire à ce que l'on veut montrer -. L'interprétation des résultats demande donc à être toujours réinterrogée en fonction des modes de raisonnement du chercheur (phénoménologie). Il n'y a pas lieu de critiquer des résultats qui ont été maintes fois démontrés par la docimologie, mais seulement de faire remarquer que les phénomènes ont été abordés sous un certain angle qui a longtemps occulté d'autres phénomènes tout aussi conséquents. N'est ce pas un « point aveugle » ? Un « obstacle épistémologique » au sens défini par G. Bachelard (1938) ? On peut donc maintenant formuler une hypothèse : le système de notation / sélection perdure malgré ses imperfections, non seulement en raison de certaines contraintes organiques et certains jeux institutionnels, mis en valeur par les approches ethnographiques des milieux scolaires (P. Merle ; P. Perrenoud), mais aussi en raison d'un certain nombre de critères implicites des milieux professionnels qui les justifient, donc des systèmes de référence fondés sur ces critères. Il existe donc un enjeu, pour la recherche en sciences de l'éducation, à pouvoir identifier ces critères en tant que processus cognitifs mis en œuvre lors des situations d'évaluation. C'est l'objectif que je me suis fixé à travers l'analyse des discours.

3. L'analyse des discours et la définition du « critère » :

3.1 Critère et processus de référence :

Il serait trop long de reprendre ici la revue de la littérature et la réflexion heuristique qui m'ont conduit à cette définition du « critère ». Je me contenterai de la livrer succinctement :

1. Le critère est une conception qui est portée par un discours. Il se différencie de l'indicateur qui est un « outil de saisie de la réalité » (J.M. Barbier ; 1985). Observer les critères pose le problème des « indicateurs langagiers » pour analyser ces constructions conceptuelles, à différencier des indicateurs pour observer les caractéristiques de l'objet évalué, qu'on nommera « indicateurs opérationnels » ;
2. Le critère s'exprime souvent sous la forme d'un concept générique, qui signifie un référent (au sens de la linguistique), c'est-à-dire une représentation de l'objet à évaluer : critère de satisfaction, critère de vitesse, critère de prix, etc. Ce concept est en relation avec des attributs, eux-mêmes en rapport de référence avec des indicateurs opérationnels : par exemple, « vitesse lente ou rapide » (chronomètre / distance parcourue/ normes), « satisfaction nulle, moyenne ou importante » (réponses à un questionnaire), etc. ;
3. Le critère fait aussi référence à une opération méthodo-logique : critère de choix, de délimitation, de regroupement, de catégorisation, de répartition, etc. ; la méthode est cohérente avec le choix des attributs, elle traduit une opération qui s'applique à ceux-ci ;
4. Enfin le critère fait référence à un ensemble conceptuel (conception) qui justifie l'opération méthodologique et les indicateurs opérationnels en fonction de l'objectif de l'évaluation, de la « fonction de l'évaluation » au sens des travaux sur cette question.

Nous apercevons, ici, à quel point la notion de « référence » est complexe, puisqu'elle fait référence au moins à trois caractéristiques,

- La référence au sens de la linguistique : l'objet auquel il est fait référence ;
- La référence au sens de la logique : l'opération méthodique qui établit une relation significative entre les indicateurs opérationnels ;
- La référence au sens du « cadre de référence » de la sociologie interactionniste (E. Goffman ; 1974/1991), c'est-à-dire le contexte auquel il nous faut faire référence pour donner du sens, à l'interaction, à l'action ou à la situation d'évaluation.

Bien entendu, ces opérations de référence sont en cohérence systémique pour constituer le « critère » : la référence – objet détermine les attributs qui sont en relation avec les indicateurs opérationnels ; la référence – méthode structure ces attributs à partir du discours (concept), en fonction des opérations de traitement réalisées sur les indicateurs opérationnels ; la référence –

cadre justifié, par un discours (conception), l'utilisation des indicateurs et l'application des méthodes (des actions organisées, des opérations successives, etc.), en fonction du contexte.

3.2 L'identification du critère dans le discours des acteurs :

La difficulté consiste maintenant à repérer, dans le discours des acteurs, ces caractéristiques du « critère ». En d'autres termes, quels sont les indicateurs langagiers qui nous conduisent à élaborer (au sens du constructivisme) un critère, à partir du discours de l'acteur ? Ou, en termes linguistiques, quels sont les processus langagiers qui signifient les processus de référence caractéristiques du critère ? Ou encore les liens entre les processus de la signification (le sens qui est produit par le langage) et ceux de la référence (les rapports avec le monde sensible) ? Pour F. D'Armengaud (1985) qui reprend les thèses de Morris, il existe trois dimensions qui donnent du sens au langage : la relation des signes aux objets, c'est la dimension sémantique ; la relation des signes aux interprètes, c'est la relation pragmatique ; la relation formelle des signes entre eux, c'est la dimension syntaxique. Ainsi, la structure de la référence dégagée à partir de l'analyse du critère correspond aux trois dimensions de la signification. Il est donc possible d'observer, dans le discours, les indicateurs langagiers qui signifient les opérations de la référence :

- Par rapport au référent – objet, le mot qui l'exprime (signifiant) est en relation avec d'autres mots, adjectifs et attributs (relations attributives), verbes factifs (relations fonctionnelles), etc. ; ces relations sont qualifiées de « sémantiques ». L'observation systématique de ces relations entre les mots (cooccurrences) permet d'analyser celles qui reviennent régulièrement (récurrence), et par là même d'inférer comment l'interviewé a construit le « concept ». C'est le principe de base de l'analyse de contenu et de la statistique lexicale ;
- Par rapport au référent – méthode, l'argumentaire de l'interviewé s'observe à travers l'appareil argumentatif (P. Charaudeau ; 1983). On peut en inférer la structure à partir de certains indicateurs comme les connecteurs (conjonctions, prépositions, etc.), mais aussi des processus comme les anaphores, les phénomènes d'inclusion d'une proposition dans une autre, les implications induites par la succession des propositions, les modalisations verbales (par ex., pouvoir, devoir, savoir faire) et adverbiales, autant d'indicateurs qui s'inscrivent dans la construction d'un argumentaire ;
- Par rapport au cadre de référence, il existe deux façons de l'appréhender : à l'intérieur du texte (on parle alors de « cotexte ») ou dans les rapports du texte et de son environnement, c'est le contexte. A.J. Greimas et J. Courtes (1979) différencient ainsi la référentialisation, processus de référence qui se construisent à l'intérieur du texte, et la référenciation, ceux qui se construisent dans le rapport avec le contexte de l'énonciation. On peut observer les premiers à travers l'appareil narratif, quand l'interviewé explique ce qui l'a conduit à bâtir son jugement. Dans ce cas, ces conceptions sont en rapport avec les processus syntaxiques et sémantiques. Les seconds s'observent plutôt à travers l'appareil énonciatif, la façon dont le discours s'ancre dans la situation : les déictiques (par ex., les pronoms personnels), mais aussi certains phénomènes d'objectivation, comme l'utilisation du substantif des verbes (en « tion »), ou encore l'utilisation de phrases nominales, etc. Mais l'interprétation de ces processus exige aussi une bonne connaissance des codes culturels des interviewés, en référence à la situation (par exemple, tutoiement ou vouvoiement), c'est-à-dire de leurs représentations sociales (E. Goffman).

On reprendra l'expérience évoquée ci-dessus pour illustrer ces propositions. On a demandé aux examinateurs une brève appréciation pour justifier leur choix, l'objectif étant d'identifier les critères des examinateurs. Ainsi, pour la prestation professionnelle, trois univers sémantiques sont apparus pour expliquer les différences de choix entre les différentes lettres :

- La notion d' « objectifs » est récurrente dans le choix de la lettre N° 6. Nous pouvons ajouter « enrichissements à apporter » et « réflexion en termes de socialisation et d'épanouissement ». Cette notion est renforcée par la cooccurrence de termes comme : « moyens » (2 fois),

- « intégration » (2 fois), « conception » (2 fois), etc. L'ensemble de ces notions évoquent « une approche progressive », une « réflexion poussée », une « vision très réfléchie » ;
- La notion de « structure » (3 fois) est la plus fréquemment évoquée pour justifier le choix des lettres N° 2 et 7. La cooccurrence avec certaines notions précise le sens de ce choix : « très synthétique », « a fait le tour », « résume l'ensemble des qualités », la « majorité des points sont notés ». Derrière cette notion de structure, apparaît l'idée que le candidat « connaît ses devoirs », du moins « l'ensemble des qualités d'un bon animateur » ;
 - Une troisième thématique autour de « l'enfant » et de la « pédagogie » permet de comprendre le choix de la lettre N° 12.

Observons maintenant l'organisation syntaxique des lettres :

- La lettre N° 6 est argumentative : elle utilise une grande variété de connecteurs (prépositions, conjonctions) pour exprimer son raisonnement. Ceci lui donne un style « réfléchi ». Elle intègre, dans son discours, des enchâssements successifs qui montrent une progression méthodologique par rapport à des objectifs (« socialisation, épanouissement ») et des moyens. Il y a « mise en évidence d'une approche progressive, expression des enrichissements à apporter dans l'animation sur plusieurs niveaux (savoir, savoir faire...) » ;
- Les lettres N° 2 et 7 sont plus énumératives, organisées autour de la modalisation du « devoir », ce qui leur confère un style structuré : elles ont « fait le tour » de la question ;
- La lettre N° 12 fait une part importante aux enfants (cités 7 fois dont une fois par anaphore), mais surtout elle est la seule où l'« enfant » est placé en position d'« actant » et non plus d'« acté » (Ghiglione & ass. ; 1998), de « sujet acteur » vs « objet subissant l'action ».

La structure de la représentation langagière proposée par les stagiaires semble donc bien avoir été déterminante dans le choix des examinateurs.

Enfin, on remarque, en analysant les appréciations, que certains arguments sont plus étayés que d'autres. L'emploi de phrases nominales (sans verbe) apparaît plusieurs fois, essentiellement pour la lettre N°6 : « professionnel » « déclinaison logique », « responsable et mature sur son rôle d'animateur »... On ne peut les interpréter, qu'en faisant référence aux présuppositions induites par le contexte de la situation. Ainsi que l'explique E. Benveniste (1966 : 165) à partir de l'analyse des langues indo-européennes anciennes, « une assertion nominale, complète en soi, pose l'énoncé hors de toute localisation temporelle ou modale et hors de la subjectivité du locuteur ». La sentence est posée, l'existence du référent-objet a valeur d'argument en soi. La discussion entre les examinateurs à l'issue de l'expérience a apporté des éléments d'analyse complémentaire. Il semble que les évaluateurs qui ont choisi la lettre N°6 soient restés plus près de la consigne. Leur argumentation s'appuie donc sur les présupposés de la situation. Alors que ceux qui ont opté pour les trois autres lettres ont éprouvé le besoin d'argumenter plus. Ces phénomènes expriment ainsi certains enjeux sociaux qui sont en débat, derrière les divergences de cadres de référence : il existe des conceptions différentes du métier, qui interviennent lors de l'évaluation.

4. La complémentarité des approches épistémologiques :

Le « référentiel », comme le « critère », sont ainsi des objets construits, mis en système (constructivisme). Ils n'existent qu'à partir du moment où le chercheur établit des rapports significatifs entre les différents phénomènes, crée le concept et le fait avaliser par la communauté culturelle (connexionnisme). Mais cette conception ne surgit pas ex-nihilo, les phénomènes sont réels. Quand nous écoutons parler une personne, nous percevons intuitivement l'opinion qu'elle a sur tel objet, son cadre de référence par rapport à telle situation, etc. Seule cette approche intuitive nous permet, dans un premier temps, de fixer notre objet, en adoptant la démarche analytique de la phénoménologie. Cette impression, nous l'avons en raison d'éléments qui existent dans le discours et qui génèrent cette sensation : ce sont eux qu'il convient d'identifier (positivisme). Ces démarches sont donc complémentaires, leur utilisation est un fait acquis dans nos disciplines : les analyses paradigmatiques ci-dessus n'apportent donc pas de nouveauté, elles clarifient seulement la fonction des différentes approches. Il ne me semble pas pertinent, en particulier, d'assimiler une méthode à un paradigme : l'analyse des discours à la phénoménologie, ou l'enquête au

positivisme. Les traitements statistiques peuvent induire des interprétations erronées, du moins excessives, sans la réflexion sur les phénomènes observés et sur la fonction des outils ; de même, si les analyses de discours ne sont accessibles, a priori, qu'avec une démarche intuitive, tant les phénomènes langagiers sont nombreux et complexes, l'analyse s'enrichit à partir du moment où les unités du langage et leurs relations sont mises en valeur. Ces approches se nourrissent respectivement. Ainsi, au fur et à mesure de la découverte des phénomènes langagiers qui créent les effets de sens « jugement », et de l'observation des connexions avec la situation d'énonciation (référenciation) et les autres structures du texte (référencialisation), la conception du critère (attributive, fonctionnelle, etc.) se précise (rapports entre les attributs du critère et les processus de référence). L'intuition première est alors enrichie de ces découvertes, qui remettent en cause certains aprioris d'origine. Cette boucle de rétroaction positive de la recherche, on l'appellera « récursion » pour reprendre l'expression d'E. Morin (1977). On se prémunit ainsi des biais fréquents des approches positivistes traditionnelles où l'hypothèse, en restreignant le champ des observations, induit certains résultats ; mais aussi des impasses tautologiques induites par des discussions philosophiques sans lien avec des approches empiriques. N'est-ce pas le risque des modélisations prescriptives, si elles ne sont pas interrogées par des approches descriptives, à partir des préoccupations professionnelles qui ont induit ces modèles ?

5. Bibliographie

- Bachelard G. (1938) *La formation de l'esprit scientifique*. Paris : Vrin.
- Barbier, J.M. (1985). *L'évaluation en formation*. Paris : PUF.
- Barbier, J.M. (1991). *Elaboration des projets d'action et de planification*. Paris : PUF.
- Benveniste, E. (1966). *Problème de linguistique générale. Tome 1*. Paris: Gallimard.
- Bloom, B.S. & ass. (1956) *Taxonomie des objectifs pédagogiques – Tome 1 – Domaine cognitif*. Trad. 1969. Montréal : Education nouvelle.
- Charaudeau, P. (1983). *Langage et discours ; éléments de sémiolinguistique*. Paris : Hachette université
- D'Armengaud, F. (1985). *La pragmatique*. Paris : PUF – Que sais-je ?
- De Ketele, J.M. (sous la direction de) (1986). *L'évaluation : approche descriptive ou prescriptive ?* Bruxelles : De Boeck.
- Durand, D. (1979) *La systémique*. Paris : PUF - Que sais-je ?
- Ghiglione R. & Landré A. & Bromberg M. & Molette P. (1998). *L'analyse automatique des contenus*. Dunod
- Goffman, E. (1974 / 1991). *Les cadres de l'expérience*. Paris : Minuit.
- Greimas, A.J. & Courtes, J. (1979). *Dictionnaire raisonné de la théorie du langage*. Paris : Université Hachette.
- Kuhn, T.S. (1962, postface de 1969 / 1983). *La structure des révolutions scientifiques*. Paris : Flammarion.
- Le Moigne, J.L. (1995). *Les épistémologies constructivistes*. Paris : Que sais-je / PUF.
- Merle, P. (1996) *L'évaluation des élèves ; enquête sur le jugement professoral*. Paris : PUF.
- Merle, P. (1998). *Sociologie de l'évaluation scolaire*. Paris : PUF – Que sais-je ?
- Morin, E. (1977) *La méthode. Tome 1. La nature de la nature*. Paris : Seuil.
- Noizet, G. & Caverni, J.P. (1978) *Psychologie de l'évaluation scolaire*. Paris : PUF.
- Perrenoud P. (1984) *La fabrication de l'excellence scolaire : du curriculum aux pratiques d'évaluation*. Genève : Droz.
- Perrenoud P. (1998) *L'évaluation des élèves. De la fabrication de l'excellence à la régulation des apprentissages. Entre deux logiques*. Paris, Bruxelles : De Boeck.

Piéron, H. (1963) - *Examens et docimologie*. Paris : PUF.

Porcher, B. & Letemplier & C., Rak, I. et ass. (1992) *Du référentiel à l'évaluation : cinq clefs pour enseigner*. Paris : Foucher.