

## LES CLASSEMENTS D'UNIVERSITÉS

*Discours prononcé par Monsieur le Recteur Philippe VINCKE à l'occasion de la séance d'ouverture de l'année académique 2007-2008, le 21 septembre 2007.*

Depuis quelques années, l'enseignement supérieur a connu un développement important qui se traduit, notamment, par deux caractéristiques essentielles :

- une augmentation significative du nombre d'acteurs de l'enseignement supérieur et de la recherche : de nouvelles universités richement dotées apparaissent dans les pays émergents, des instituts privés se créent autour de nous, des universités prestigieuses installent des antennes aux quatre coins du monde, des programmes universitaires diplômants sont proposés sur Internet, ...
- une mobilité croissante des étudiants et des enseignants-chercheurs, mobilité encouragée, à juste titre d'ailleurs, par les pouvoirs publics de tous les pays.

Les étudiants et les chercheurs souhaitent donc désormais disposer d'informations sur la qualité des institutions d'enseignement supérieur où ils sont susceptibles de séjourner ou avec lesquelles ils sont amenés à collaborer.

Ce souci fort légitime va de pair avec une demande de plus en plus forte de la collectivité de voir les universités justifier l'utilisation des moyens mis à leur disposition et rendre des comptes à leurs pouvoirs subsidiaires.

L'évaluation est une tradition dans le monde scientifique. Les chercheurs sont habitués à discuter de leurs résultats avec leurs pairs. Leurs propositions de communications dans les revues scientifiques ou les congrès internationaux sont soumises à la critique d'autres spécialistes qui peuvent les accepter, les amender ou les refuser. En général, les chercheurs connaissent les ténors de leur discipline et savent où se trouvent les équipes de qualité. En l'absence même de classement dans des journaux à sensation, l'évaluation de la qualité fait partie intégrante de la vie de l'enseignant-chercheur, du moins pour ses activités de recherche. C'est encore peu le cas pour les activités d'enseignement.

La nouveauté – conséquence directe de l'augmentation de l'offre d'enseignement supérieur et de la mobilité croissante des étudiants et des chercheurs –, c'est la concurrence accrue que se livrent les universités pour attirer les meilleurs, étudiants ou chercheurs et, bien entendu, les moyens financiers. D'où l'apparition de comparaisons entre les institutions, de classements, de hit-parades, de *rankings*.

Deux classements font régulièrement la une des médias (peu enclins cependant à se livrer à une analyse sérieuse des méthodologies utilisées) : le classement de la Jiao Tong University de Shanghai (dit « classement de Shanghai ») et le classement annuel publié dans le *Times Higher Education Supplement* (ci-après, le « classement du *Times* »).

Plus aucune université ne peut les ignorer. Y figurer en bonne place est devenu un argument de promotion, la carte de visite de l'université, sinon même une condition *sine qua non* pour établir des collaborations internationales intéressantes. Ces classements vont donc marquer de plus en plus l'évolution du paysage universitaire mondial.

Toute évaluation est le résultat d'une comparaison et il convient, à ce stade, de distinguer deux situations radicalement différentes.

- **Première situation** : vous fixez des objectifs ou des normes *a priori* et vous comparez vos activités (vos performances) à ces objectifs ou à ces normes. On parle alors d'*évaluation absolue*, même si le résultat de l'évaluation dépend des normes fixées *a priori*. Ainsi, si vous considérez des normes de sécurité pour les voitures et que vous comparez les caractéristiques de

votre automobile à ces normes, vous pourrez en conclure qu'elle est très bonne, bonne, moyenne ou médiocre sur le plan de la sécurité.

- **Deuxième situation** : vous comparez vos performances à celles de vos voisins ou de vos concurrents. On parle alors d'*évaluation relative*. Si vous comparez la sécurité de votre voiture à celle de votre voisin, vous pourrez en conclure que votre véhicule est plus fiable, aussi fiable ou moins fiable que celui de votre voisin. Pareil constat ne vous donne aucune information sur son niveau de sécurité absolu, c'est-à-dire par rapport à un système de normes. Votre voiture peut être plus fiable que celle de votre voisin tout en étant médiocre. Elle peut être moins fiable tout en étant excellente.

Le système d'accréditation Equis, qui consiste à attribuer aux Ecoles de commerce un label international de qualité, repose sur une évaluation absolue, c'est-à-dire sur la comparaison des paramètres des Ecoles à des normes prédéfinies. Ce n'est pas le cas des classements d'universités de Shanghai et du *Times*, fondés sur une évaluation relative, et dont nous nous proposons d'analyser les méthodes.

## I. Le classement de Shanghai

Le classement de Shanghai est basé sur quatre, cinq ou six critères, selon les universités considérées.

La qualité de l'éducation d'une université est mesurée par un premier critère : le nombre d'anciens étudiants qui ont obtenu un prix Nobel ou une médaille Fields. Un ancien étudiant est défini comme une personne ayant obtenu au moins un diplôme dans l'université considérée ; il compte pour une unité s'il a obtenu ce diplôme après 1990, pour 0,9 s'il l'a obtenu entre 1980 et 1990, etc. ; il compte pour 0,1 s'il a obtenu son diplôme entre 1900 et 1910.

Deux critères mesurent la qualité du corps enseignant d'une université :

- le nombre de prix Nobel et de médailles Fields décernés aux enseignants durant leur activité dans l'université considérée ; là aussi le poids décroît avec l'ancienneté et un système complexe de fractionnement vient pondérer les résultats, si le scientifique était actif simultanément dans plusieurs universités et/ou si plusieurs scientifiques se sont partagé un même prix ;
- le nombre d'enseignants – chercheurs de l'université qui font partie des 250 auteurs les plus cités dans une période donnée (pour le classement 2006, il s'agissait de la période 1981-2003) et ce, dans 21 grandes disciplines (nous y reviendrons). Les données relatives à ce critère sont extraites d'une base de données commerciale (*the Essential Science Indicators database*), distribuée par une société basée à Philadelphie, la Thomson Scientific. Faute d'information précise sur la méthodologie utilisée pour élaborer ces données, il nous est impossible de les reconstituer et donc de les vérifier.

Le quatrième critère est le nombre d'articles, issus de l'université considérée, publiés dans les revues *Nature* et *Science*, au cours des cinq dernières années, avec un système de pondération pour les articles écrits par plusieurs auteurs. Dès lors que ces deux revues scientifiques prestigieuses privilégient les sciences exactes ou les sciences à laboratoires, ce critère n'est pas pris en compte si une université n'est pas active dans ces domaines : son poids est alors redistribué sur les autres critères. Tel est le cas, par exemple, de la London School of Economics, institution renommée qui se spécialise dans les sciences humaines et sociales.

Le cinquième critère est le nombre d'articles, issus de l'université considérée, répertoriés, dans une période donnée, dans le *Science Citation Index* et dans le *Social Science Citation Index*. Ces deux répertoires, produits par Thomson Scientific, établissent des statistiques sur le nombre de citations des articles dans d'autres articles et donc, d'une certaine manière, sur la renommée ou l'impact des articles scientifiques.

Dans la première version du classement de Shanghai, seuls ces cinq critères étaient pris en compte (ils étaient ramenés à quatre pour les institutions de sciences humaines). Les auteurs se sont alors rendu compte que leur classement était fortement corrélé à la taille des universités : tous les critères s'exprimaient en termes de nombres absolus (nombres de prix, de citations, d'articles, ...) indépendamment des nombres de chercheurs travaillant dans chaque université.

Ils ont essayé de corriger ce défaut en ajoutant un sixième critère construit de la façon suivante : ils effectuent, pour chaque université, la somme des scores des cinq (ou des quatre) premiers critères et divisent cette somme par le nombre d'enseignants – chercheurs de l'université. Le résultat obtenu est le score de l'université pour le sixième critère. Néanmoins, comme les auteurs ne disposaient pas des données pour toutes les universités, ce critère supplémentaire n'a été ajouté que pour certaines d'entre elles : certaines universités sont donc évaluées sur quatre critères, d'autres sur cinq, d'autres encore sur six.

Pour chacun de ces quatre, cinq ou six critères, l'université classée première reçoit un score de 100 ; les autres obtiennent un score qui est une note sur 100, calculée par une simple règle de trois.

Le score global d'une université est la somme pondérée des scores obtenus sur les différents critères. Chaque critère intervient dans le résultat final pour 25% lorsqu'il y a quatre critères, et pour 20%, lorsqu'il y en a cinq. Lorsqu'il y a six critères, le premier (nombre de prix Nobel attribués à d'anciens étudiants) et le sixième (introduit pour atténuer l'effet taille) comptent pour 10% tandis que les quatre autres interviennent chacun pour 20%.

Signalons encore, pour être complet, que les auteurs disent avoir procédé à des corrections de certains scores lorsqu'ils y trouvaient une anomalie, mais sans nous fournir d'explication précise sur ce qu'ils considèrent comme une anomalie ni sur les modalités des corrections apportées.

## **II. Quels commentaires formuler sur le *ranking* de Shanghai ?**

**Qui sont les auteurs, quelles sont leurs compétences ?** « *En réalité* », reconnaissent-ils eux-mêmes, « nous ne nous considérons pas comme des spécialistes de la scientométrie ou de la bibliométrie. Néanmoins, les classements d'universités ont de plus en plus d'influence sur le développement de l'enseignement supérieur dans le monde. C'est pourquoi nous espérons que des personnes compétentes apporteront leur aide aux débutants ».

Autrement dit, les auteurs avouent ne pas être des experts et lancent une sorte d'appel aux personnes compétentes pour s'occuper sérieusement du problème.

D'après son *curriculum vitae*, le principal auteur est un professeur de chimie (mes collègues de chimie y verront un signe de plus que la chimie mène à tout) spécialisé dans les polymères qui, en 2002, a subitement arrêté cette activité pour s'intéresser à la comparaison des universités. Un des aspects les plus gênants de son travail réside dans les zones d'ombre de sa méthodologie (j'en ai cité quelques-unes), qui ne permettent pas au lecteur de reconstituer le classement obtenu, et donc de le vérifier : une lacune à corriger si les auteurs veulent, un jour, voir leur travail qualifié de scientifique.

**Quels sont les objectifs déclarés de ce travail ?** Au départ, il s'agit d'une entreprise patriotique visant à élever le niveau de la recherche dans les universités chinoises en suscitant l'émulation avec les universités des autres pays. L'énoncé des critères en témoigne, la comparaison concerne la production scientifique, c'est-à-dire la recherche. L'enseignement, l'encadrement des étudiants ou le coût des études, par exemple, ne sont pas pris en compte. Les auteurs ne font d'ailleurs aucun commentaire sur les utilisations qu'ils considèrent comme pertinentes de leurs résultats, ni sur les précautions à prendre dans leur interprétation.

Et pourtant, l'examen attentif des données utilisées et de leur traitement incite à la plus grande prudence.

Commençons par examiner les critères sur lesquels se base le classement de Shanghai.

Les deux premiers critères concernent les prix Nobel et les médailles Fields obtenus par les anciens étudiants et par les chercheurs. Savoir à quelle université attribuer un prix est loin d'être simple. Les scientifiques, surtout à ce niveau, ont l'habitude de changer d'institution au cours de leur carrière et un prix récompense souvent des travaux bien antérieurs à la date d'attribution.

Cette difficulté a d'ailleurs conduit à des situations inextricables. Ainsi, les deux grandes universités de Berlin (l'Université libre et l'Université Humboldt) ont-elles été exclues du classement de Shanghai faute d'arriver à déterminer à quelle institution attribuer le prix Nobel obtenu par Albert Einstein : l'affectation ou non du prix à l'une des deux universités aurait en effet modifié son classement de plus de cent places dans le résultat final.

D'autres critères font intervenir des nombres de citations d'auteurs ou d'articles. Il existe une très abondante littérature sur les avantages et les inconvénients des indices de citation et des analyses bibliométriques. Sans entrer dans les détails de ces études qui mériteraient, à elles seules, tout un exposé, nous noterons que, dans la situation actuelle, ces outils semblent assez pertinents dans les disciplines qui les ont intégrés dans leur mode de fonctionnement (en particulier, les sciences de la vie et de la santé). Ils le sont beaucoup moins, à ce jour, dans d'autres.

De manière générale, les experts en bibliométrie admettent que des domaines tels que les sciences de l'ingénieur, les sciences de l'environnement, les sciences sociales, le droit et, plus globalement les sciences humaines (à l'exception de certaines parties de la psychologie et de l'économie) ne peuvent, pour le moment du moins, être traitées de façon satisfaisante par les techniques habituelles d'analyse bibliométrique. Le recours exclusif à ces techniques risque donc d'introduire un biais important dans la comparaison des universités, en fonction des disciplines qui y sont représentées.

Nous l'avons vu aussi, un des critères se base sur le nombre de chercheurs qui font partie des 250 auteurs les plus cités dans les 21 disciplines retenues par la Thomson Scientific. Je vous en livre rapidement la liste : Mathématiques, Physique, Chimie, Biologie – Biochimie, Informatique, Géosciences, Sciences de l'espace, Ingénierie, Science des matériaux, Agriculture, Environnement, Médecine clinique, Médecine vétérinaire, Pharmacologie, Biologie moléculaire et génétique, Microbiologie, Immunologie, Neurosciences. Nous en sommes à 18 disciplines sur 21 et aucun domaine des sciences humaines n'a encore été cité. Les trois dernières disciplines sont Psychologie – Psychiatrie, Economie et gestion, Sciences sociales. Aucune trace de la Philosophie, des Lettres, de l'Histoire, de l'Archéologie, du Droit, des Sciences politiques, ....

L'examen du nombre de revues publiées dans chaque domaine révèle également un déséquilibre gigantesque entre les disciplines. La méthodologie utilisée ne tient pas compte du fait que les volumes de production scientifique peuvent varier fortement d'un domaine à l'autre et, par conséquent, favorise certaines universités par rapport à d'autres, selon les disciplines qu'elles couvrent.

Elle ignore aussi les spécificités ou les traditions des différents domaines. Ainsi, dans certaines disciplines, il n'est pas rare de voir des articles signés par six, sept ou dix co-auteurs, alors que dans d'autres, on mettra un point d'honneur à publier seul. Se baser sur le nombre de fois que des auteurs apparaissent dans des listes risque alors de biaiser les évaluations.

Deux critères concernent les nombres d'articles publiés dans *Nature* et *Science*, d'une part, ou indexés dans les bases de données de la Thomson Scientific, d'autre part. Précisons que lorsqu'il y a plusieurs auteurs, ceux-ci sont comptabilisés pour 1,  $\frac{1}{2}$  et  $\frac{1}{4}$  pour les trois principaux auteurs (notion qui, dans beaucoup de disciplines, n'a pas de sens) et  $\frac{1}{10^e}$  pour les autres.

Autrement dit, plus le nombre d'auteurs est important, plus l'article contribue positivement à l'évaluation de l'université dont ils sont issus : en poussant le raisonnement à l'extrême, nous pourrions suggérer que, dorénavant, tous les collègues de l'ULB co-signent tous les articles produits à l'ULB.

Enfin, malgré l'adjonction du sixième critère, l'effet taille continue à jouer un rôle important, excepté, bien entendu, pour les quelques universités qui collectionnent les prix Nobel. Il suffirait donc que les universités belges fusionnent pour progresser dans le classement de Shanghai. Ce qu'ont d'ailleurs décidé de faire certaines universités de grandes villes françaises : le président de l'Université Pierre et Marie Curie (Paris 6) me signalait récemment que la fusion de son institution avec l'Université Paris-Sud (Paris 11) les amènerait aux environs du 8<sup>e</sup> rang mondial, sans rien changer à leur politique d'enseignement ou de recherche.

Les auteurs du classement de Shanghai ne disent rien du poids respectif des critères qui a pourtant un impact majeur sur le classement final. Leur site Web propose au visiteur de fixer lui-même des poids, sans référence aucune aux unités des échelles utilisées ou au mode de normalisation, ce qui est contraire aux principes élémentaires de ce type de travail.

Indépendamment des commentaires sur les critères choisis et les biais qu'ils entraînent, qu'en est-il de la précision des données numériques utilisées ?

Les analyses bibliométriques à grande échelle sont nécessairement entachées d'une foule de petites erreurs matérielles dues notamment aux fautes d'orthographe dans les noms des auteurs, au mauvais encodage de leur affiliation, des numéros des revues ou des pages des articles concernés.

Une étude parue en 2002 dans la revue *Nature* (une étude sérieuse donc !) estime à environ 30% l'erreur globale due à cette accumulation de scories matérielles. Cette remarque ne concerne d'ailleurs pas le seul classement de Shanghai.

Les auteurs de ce classement, de leur côté, prétendent que leurs données sont correctes à 2% près mais sans étayer en rien leur affirmation. En outre, toute idée d'erreur ou d'imprécision disparaît complètement dans la présentation de leurs résultats. Or, 2%, c'est, par exemple, l'écart entre les 45<sup>e</sup> et 60<sup>e</sup> places dans le classement de Shanghai. Et 30% – l'erreur estimée dans l'étude de *Nature* –, c'est l'écart entre la 12<sup>e</sup> et la 100<sup>e</sup> place dans le même classement.

L'affiliation des auteurs d'articles scientifiques aux universités est un critère tout aussi problématique. Nombre d'entre eux ne désignent pas leur appartenance institutionnelle de manière univoque : alors qu'« Université Libre de Bruxelles » est le nom officiel de notre institution, beaucoup de collègues utilisent d'autres désignations, surtout – et le cas est fréquent – s'ils écrivent en anglais : University of Brussels, Free University of Brussels, Brussels University, ...

Nous l'avons vérifié sur un petit échantillon d'éminents collègues de notre université, l'expression « Université Libre de Bruxelles » apparaît dans moins de 10% de leur production scientifique. Ce phénomène – qui ne concerne pas uniquement notre université – entraîne un biais important en faveur des universités anglo-saxonnes.

Parfois, au lieu du nom de l'Université, l'auteur d'un article mentionne sa Faculté, son département ou son service, avec des adresses qui vont – pour l'ULB –, de l'avenue Roosevelt au boulevard du Triomphe, en passant par la route de Lennik à Anderlecht ou la rue des professeurs Jeener et Brachet à Gosselies. La personne ou plutôt le logiciel qui, à la Thomson Scientific ou à Shanghai, traite toutes ces données, ignore évidemment qu'il s'agit de la même université.

Dans le même ordre d'idées, la recherche médicale est particulièrement mal prise en compte : une grande partie de cette recherche se déroule dans les hôpitaux académiques et n'apparaît pas toujours dans la production scientifique des universités. Nous avons d'ailleurs découvert récemment que les

publications émanant de notre hôpital académique, l'hôpital Erasme, étaient rattachées, dans les bases de données de la Thomson Scientific, à la Erasmus Universiteit de Rotterdam.

Enfin, permettez-moi de terminer l'analyse du classement de Shanghai par un aspect technique qui, à lui seul, incite à la plus grande prudence dans le traitement des résultats de ce classement.

Le classement de Shanghai utilise une **normalisation** de 0 à 100 des échelles, en attribuant, sur chaque critère, la note 100 à l'université classée première sur ce critère, et en appliquant une règle de trois pour les autres universités. Chaque université reçoit ensuite, comme score global, la moyenne pondérée de ses résultats. Illustrons cette technique à partir d'un exemple numérique.

 UNIVERSITÉ LIBRE DE BRUXELLES. UNIVERSITÉ D'EUROPE			
	0.4	0.4	0.2
	C1	C2	C3
<b>a</b>	<b>2000</b>	<b>500</b>	<b>5</b>
<b>b</b>	<b>1360</b>	<b>440</b>	<b>10</b>
<b>c</b>	<b>1600</b>	<b>375</b>	<b>10</b>

Le tableau présente les évaluations de trois universités, a, b et c, sur trois critères. Sur le premier des trois critères, les évaluations sont, respectivement, de 2 000, 1 360 et 1 600 (les chiffres désignent, par exemple, les nombres de chercheurs). Sur le deuxième critère, les évaluations sont respectivement de 500, 440 et 375 (soit, par exemple, les nombres de thèses de doctorat sur une période donnée). Sur le troisième critère, les évaluations sont de 5, 10 et 10 (nous aurions ici les nombres de prix prestigieux). Les nombres indiqués au-dessus du tableau représentent les poids des critères : 0.4, 0.4 et 0.2.

La première étape consiste à normaliser les critères. Sur le premier, l'évaluation la plus élevée est 2 000. On la ramène à 100, ce qui revient à la diviser par 20. En faisant de même avec les autres éléments de cette première colonne, on obtiendra 68 pour b et 80 pour c. Sur le deuxième critère, l'évaluation la plus élevée est 500. On la ramène à 100 en la divisant par cinq. En faisant de même avec les autres éléments de cette deuxième colonne, on obtiendra 88 pour b et 75 pour c. Sur le troisième critère, les évaluations doivent être multipliées par 10 pour amener la plus élevée à 100, ce qui donne 50 pour a et 100 pour les deux autres.

 UNIVERSITÉ LIBRE DE BRUXELLES. UNIVERSITÉ D'EUROPE			
	0.4	0.4	0.2
	C1	C2	C3
<b>a</b>	<b>100</b>	<b>100</b>	<b>50</b>
<b>b</b>	<b>68</b>	<b>88</b>	<b>100</b>
<b>c</b>	<b>80</b>	<b>75</b>	<b>100</b>

On calcule ensuite les moyennes pondérées des universités. Pour a :  $100 \times 0.4 + 100 \times 0.4 + 50 \times 0.2 = 90$ . De manière analogue, on obtient 82,4 pour b et 82 pour c.

Supposons maintenant que l'on modifie l'évaluation de a sur le premier critère, en remplaçant 2 000 par 1 700. Tout le reste est inchangé, ce qui nous donne le tableau suivant.

 UNIVERSITÉ LIBRE DE BRUXELLES. UNIVERSITÉ D'EUROPE			
	0.4	0.4	0.2
	C1	C2	C3
<b>a</b>	<b>1700</b>	<b>500</b>	<b>5</b>
<b>b</b>	<b>1360</b>	<b>440</b>	<b>10</b>
<b>c</b>	<b>1600</b>	<b>375</b>	<b>10</b>

La normalisation du premier critère se fait donc en divisant par 17 au lieu de 20 (b : 80, c : 94). Rien ne change pour les deux autres critères, et on obtient le tableau normalisé que voici.

 UNIVERSITÉ LIBRE DE BRUXELLES. UNIVERSITÉ D'EUROPE			
	0.4	0.4	0.2
	C1	C2	C3
<b>a</b>	<b>100</b>	<b>100</b>	<b>50</b>
<b>b</b>	<b>80</b>	<b>88</b>	<b>100</b>
<b>c</b>	<b>94</b>	<b>75</b>	<b>100</b>

Le calcul des moyennes pondérées conduit alors au résultat suivant : a : 90 ; b : 87.2 ; c : 87.6.

 UNIVERSITÉ LIBRE DE BRUXELLES. UNIVERSITÉ D'EUROPE		
	Avant	Après
<b>a</b>	<b>90</b>	<b>90</b>
<b>b</b>	<b>82.4</b>	<b>87.2</b>
<b>c</b>	<b>82</b>	<b>87.6</b>

Si nous comparons les scores avant et après la modification, nous voyons que les scores de b et c se sont rapprochés de celui de a, ce qui est normal puisque la performance de a sur le premier critère a

diminué. Mais – ce qui est beaucoup plus gênant –, l'ordre des scores de b et c a été modifié. L'université b était meilleure que c avant la modification. Elle devient moins bonne après la modification, alors que rien n'a changé ni dans les performances de b, ni dans les performances de c.

Voilà un exemple d'effet, qu'on pourrait qualifier de pervers, de la normalisation telle qu'elle est pratiquée dans cette méthode. Une modification d'une performance d'une université sur un critère peut bouleverser le classement des autres universités, toutes choses étant égales par ailleurs.

Ce type de considération relativise fortement les conclusions à tirer du classement de Shanghai ; il est pourtant absent de la littérature qui y est consacrée. Sans doute même, les auteurs ignorent-ils tout de ce type de phénomène et, plus généralement (car nous n'avons pris qu'un exemple), des propriétés cachées ou des effets pervers de leur méthodologie.

Toutes ces réflexions conduisent, me semble-t-il, à mettre sérieusement en doute le caractère scientifique du travail effectué par les auteurs du classement de Shanghai.

### **III. Le classement du *Times***

Le classement du *Times Higher Education Supplement* est publié par une entreprise privée, filiale de la News International Publishers Limited, société qui imprime le *Times* et le *Sunday Times*. La méthodologie repose en partie sur une enquête portant sur la renommée des universités. Plusieurs centaines de scientifiques, de pays différents, sont invités à mentionner les universités qu'ils considèrent comme les meilleures dans les régions du monde pour lesquelles ils s'estiment compétents. Récemment, les responsables ont complété cet exercice par une enquête menée auprès d'un certain nombre d'employeurs.

Du résultat de cette double enquête, les auteurs du classement déduisent les scores des universités sur le premier critère, qui interviendra pour 50% dans le résultat final. Nous ne savons rien en revanche des calculs précis effectués pour traduire les résultats de l'enquête en scores sur ce critère, ni de la façon dont ils tiennent compte des enquêtes des années précédentes.

Les quatre autres critères sont :

- l'impact, en termes de citations, des chercheurs de l'université (avec un poids de 20%) ;
- le taux d'encadrement des étudiants (avec un poids de 20%) ;
- le pourcentage d'étudiants étrangers (avec un poids de 5%) ;
- le pourcentage d'enseignants étrangers (avec un poids de 5%).

Le premier d'entre eux, l'impact en termes de citations, est obtenu, ici aussi, à partir des bases de données de la Thomson Scientific.

L'information lacunaire donnée sur la manière dont cet impact est calculé ne permet pas non plus de reconstituer et de vérifier les calculs.

Les données concernant les trois autres critères sont fournies par les universités elles-mêmes, qui sont invitées à remplir un questionnaire en ligne.

La normalisation à 100 de chaque échelle et l'agrégation par une moyenne pondérée se font comme dans le classement de Shanghai. L'exemple numérique présenté ci-dessus et la critique qu'il entraîne sur la manière de traiter les données, s'appliquent donc aussi au classement du *Times*.

#### IV. Quels commentaires apporter sur le classement du *Times* ?

Tout d'abord, une part importante du classement est basée sur des avis d'« experts ». Selon les auteurs du classement, le monde scientifique est habitué au système du *peer review* (le jugement par les collègues) et leur méthodologie s'inscrirait dans cette tradition.

Cet argument nous paraît assez superficiel. En effet, on demande ici à des scientifiques d'évaluer non un article scientifique ou un projet de recherche dans leur spécialité ou encore le *curriculum vitae* d'un collègue dans leur domaine de recherche (c'est-à-dire là où ils sont véritablement experts), mais de donner un avis éclairé et précis sur les performances de dizaines ou même de centaines d'universités dans leur globalité et leur complexité : la réponse qu'ils risquent donc de donner, s'ils souhaitent en donner une, sera une vague extrapolation de ce qu'ils savent ou l'écho d'une rumeur ou encore une information lue dans la presse.

Tout se passe comme si l'on demandait à un œnologue d'évaluer la qualité d'une liste de restaurants, non seulement pour la carte des vins, qui relève de son domaine d'expertise, mais aussi pour la qualité de la cuisine, l'originalité des préparations, l'accueil, le service, ... alors qu'il n'a jamais mis les pieds dans la plupart d'entre eux.

La méthodologie utilisée pour mener cette enquête pose d'ailleurs nombre de questions. Comment les « experts » sont-ils recrutés ? Quel est le protocole d'évaluation ? Quel est le taux de réponse à l'enquête ? Quel est le profil des personnes ayant répondu ? Quel est le profil de celles qui n'ont pas répondu ? Comment se présentent les distributions des réponses ? Comment sont traitées les divergences d'avis ? Peut-on évaluer la précision des conclusions de l'enquête ? Quel crédit un lecteur libre-exaministe peut-il accorder à une enquête pour laquelle il ne dispose d'aucune de ces informations ?

Anthony VAN RAAN, de l'Université de Leiden, a récemment calculé que la corrélation entre les réponses des scientifiques aux questions du *Times* et une analyse bibliométrique classique était de l'ordre de 0,005, c'est-à-dire assimilable à 0. Que mesure donc l'enquête du *Times* ? D'aucuns considèrent qu'elle donne, tout au plus, une information sur la compétence des personnes interrogées ...

Pour le deuxième critère, l'impact en termes de citations, je vous renvoie aux commentaires faits précédemment pour le classement de Shanghai, sur la difficulté d'avoir des données fiables et sur les nombreuses erreurs matérielles de ce genre d'exercice.

Pour les trois derniers critères, les données sont fournies par les universités elles-mêmes en réponse à un questionnaire. Les questions posées sont cependant loin d'être claires et le vocabulaire utilisé est susceptible d'interprétations diverses. Au point qu'en Communauté française, les universités ont décidé, en 2007, de s'accorder sur une interprétation commune des questions posées. Bien entendu, rien ne prouve que les autres universités du monde auront adopté la même interprétation que nous.

L'expérience l'a aussi montré, que les données soient fournies par les universités elles-mêmes n'empêche pas les erreurs. Ainsi, les données utilisées par le *Times* en 2006 étaient-elles complètement erronées pour notre université, malheureusement en notre défaveur. Cette erreur, reconnue par les responsables du classement, nous a valu un recul de 100 places entre 2005 et 2006 ; ce que la presse a monté en épingle.

D'après les auteurs du *Times*, une différence de quelques pourcents dans le score global ne serait pas significative. Mais 8% à peine séparent la 100<sup>e</sup> université de la 200<sup>e</sup> et 80% des universités se situent dans un intervalle de 20 points sur 100 (20 points sur 100, c'était, selon l'étude parue dans *Nature*, moins que la marge d'erreur due à l'accumulation des erreurs matérielles d'encodage et d'imprécisions sur les données dans ce genre d'exercice).

On peut donc regretter que les auteurs, et les journalistes qui les relaient, n'apportent pas plus de nuances dans la présentation de leurs résultats.

Je souhaiterais ajouter quelques commentaires généraux, valables pour la plupart des classements que l'on trouve actuellement dans la littérature ou dans la presse.

Tout d'abord, indépendamment des questions méthodologiques ou techniques, les classements d'universités soulèvent des questions préalables fondamentales. J'en citerai cinq :

1) Qu'est-ce qu'une université ? Place-t-on dans le même panier les universités complètes et incomplètes, les business schools privées américaines ou asiatiques et les universités publiques européennes ? Que fait-on des grandes Ecoles françaises, de l'Institut Max Planck, des laboratoires du CNRS ? Comment comparer des organisations qui évoluent dans des univers institutionnels et culturels totalement différents ? Quelle est donc la définition précise des objets que l'on souhaite comparer et classer ?

2) Comment définir « la qualité » d'une université ? Les classements de Shanghai et du *Times* semblent présupposer que la qualité d'une université est une réalité objective, claire pour tout le monde, qu'il suffit de mesurer. Peut-on toujours affirmer sans hésitation qu'une université est de meilleure qualité qu'une autre ? Cette question a-t-elle un sens ? Que signifie « de meilleure qualité » ? Peut-on mesurer la qualité d'une université comme on mesure la longueur d'une table ?

En tête de classement figurent quelques universités dont personne ne met la qualité en doute, Harvard, Stanford, Yale, le MIT, Cambridge, Oxford (universités qui ont, faut-il le dire, des moyens financiers à la hauteur de leurs ambitions), mais cela ne prouve pas que le classement, dans son entièreté, reflète une réalité objective. On lit d'ailleurs dans un éditorial des auteurs du classement du *Times* cette phrase curieuse : « *Il n'y a aucune indication qu'une université bien placée dans notre classement est meilleure qu'une université moins bien placée* ». Que signifient dès lors les autres commentaires formulés par ces mêmes auteurs à partir de leur classement ?

3) Existe-t-il un modèle absolu de ce qu'est une bonne université, un modèle vers lequel toutes les institutions du monde devraient tendre ? Est-il souhaitable d'avoir un seul modèle de bonne université ?

4) Quels aspects de l'université prendre en compte pour évaluer sa qualité ? La recherche ? L'enseignement ? La recherche et l'enseignement ? L'efficacité de l'organisation ? La capacité à obtenir des financements extérieurs ? Les contrats ? Les brevets ? Les *spin-offs* ? Le caractère international ? La participation au développement économique régional ? La formation continue ? La dimension culturelle ? Le rôle social ? Les rémunérations des professeurs ? Le coût des études ? La qualité des logements pour les étudiants ? La richesse de la bibliothèque ? La présence ou non d'une sélection à l'entrée ?

On peut en tout cas constater que la qualité de l'enseignement est très peu représentée dans les deux classements évoqués ici et que les autres aspects que je viens de citer sont quasi tous absents.

Du point de vue de l'étudiant, qui veut choisir son lieu d'études, ne serait-il pas plus utile de comparer, pour chaque discipline, les programmes offerts par différentes universités ?

Du point de vue des dirigeants de l'université, ne vaudrait-il pas mieux détecter les points forts et les points faibles de leur institution en comparant celle-ci, critère par critère, à des institutions qui poursuivent des objectifs similaires ?

Ces réflexions nous conduisent à la cinquième question de base : quelle utilisation faire de ces classements ? Rares sont les auteurs de classements qui précisent les objectifs poursuivis. Or,

comment définir une méthode scientifique de comparaison et de classement des universités si on ne définit pas *a priori* le but visé et l'usage que l'on veut faire des résultats ?

A ces cinq questions fondamentales auxquelles, me semble-t-il, tout exercice de classement devrait d'abord répondre, s'ajoutent des aspects plus méthodologiques. J'en citerai également cinq :

1) En supposant que l'on ait choisi les aspects à prendre en compte, comment les mesurer ? Avec quels outils ? Comment réaliser toutes ces mesures et rassembler toutes ces données pour les milliers d'universités du monde, ou même pour quelques centaines d'entre elles, en un temps raisonnable, surtout si le classement est publié chaque année ? Comment garantir la fiabilité et la précision des données ainsi rassemblées ? Que signifient une différence de 10, 20 ou 30 points ou une différence de 10, 20 ou 50 places dans le résultat final ? Là aussi, les auteurs des classements restent en général muets sur la signification statistique de leurs résultats.

2) Comme nous l'avons vu, une difficulté particulière est liée à l'évaluation de l'activité scientifique en sciences humaines. D'une part, les analyses bibliométriques, devenues monnaie courante dans certaines disciplines, sont, aujourd'hui du moins, inadaptées à la plupart des disciplines de sciences humaines. D'autre part, l'enquête auprès de scientifiques risque, sans doute plus en sciences humaines qu'en sciences exactes, d'être biaisée par le choix des scientifiques dans la mesure où l'on trouve, plus en sciences humaines qu'en sciences exactes, des « écoles » de pensée qui portent sur certaines recherches des appréciations diamétralement opposées. L'évaluation de l'activité scientifique en sciences humaines reste en tout cas un problème qui mérite d'être approfondi.

3) Les classements dont nous avons parlé n'évaluent que l'output des universités, sans jamais considérer l'input ou le contexte dans lequel elles doivent fonctionner. Ils ne prennent pas en compte les budgets disponibles, les contraintes en matière de recrutement d'étudiants (nos universités sont ouvertes à tous, d'autres effectuent des sélections sévères), les contraintes en matière de droits d'inscription, de rémunération des professeurs, de gestion du personnel.

4) Il existe à l'évidence, tous les auteurs le reconnaissent, un biais en faveur des universités anglo-saxonnes, l'anglais étant naturellement, dans la plupart des disciplines, la langue de communication entre les chercheurs. Des études récentes ont montré que le désavantage, pour les universités germanophones et francophones, pouvait représenter une sous-estimation de plusieurs dizaines de pourcents de leur production scientifique.

5) En supposant même que l'on ait pu rassembler toutes les données nécessaires et garantir leur fiabilité, est-il légitime de synthétiser cette masse d'informations en attribuant une note sur 100 à chaque université du monde ?

Que signifie cette note sur 100, censée intégrer des aspects aussi divers que la recherche, la pédagogie, l'encadrement des étudiants, la qualité de vie sur le campus, ... mais aussi toutes les disciplines habituellement présentes dans une université (sciences, médecine, histoire de l'art, droit, économie, psychologie, ...) ? La notion de note moyenne a-t-elle encore un sens dans un tel contexte ?

Enfin, permettez-moi d'attirer à nouveau votre attention sur un aspect purement technique commun à la plupart des méthodes de classement, et dont les conséquences sont loin d'être anodines. Il s'agit de l'habitude d'agrèger par une moyenne pondérée les performances des universités sur les différents critères.

Tout d'abord, la moyenne pondérée efface toutes les informations sur les points forts et les points faibles des universités, mettant sur un même pied une université bonne sur tous les critères et une université ayant des lacunes graves sur certains critères compensées par des aspects excellents sur d'autres.

De ce point de vue, la présentation et la comparaison des profils des universités (c'est-à-dire des vecteurs de leurs performances) seraient bien plus instructives qu'un score agrégé qui ne mesure aucune réalité objective. Certes, si une université s'améliore sur tous les critères, sa moyenne pondérée s'améliorera aussi, mais l'inverse n'est pas vrai. Une amélioration de la moyenne pondérée d'une université ne signifie pas nécessairement que cette université s'améliore. Cela signifie qu'elle s'améliore sur certains aspects, mais elle peut fort bien se détériorer sur d'autres. La qualité globale d'une université est une notion décidément trop complexe pour être mesurée correctement par un nombre unique.

Mais les auteurs des classements l'ignorent sans doute, le choix de la moyenne pondérée comme technique d'agrégation contient implicitement un choix politique.

Illustrons ceci par un exemple numérique. Voici trois universités évaluées sur deux échelles au moyen de notes comprises entre 0 et 100.



	C1	C2
<b>a</b>	<b>41</b>	<b>97</b>
<b>b</b>	<b>100</b>	<b>38</b>
<b>c</b>	<b>68</b>	<b>68</b>

Une politique possible consiste à préférer une université très forte sur un critère, même si elle est très faible sur l'autre. Cette politique conduirait à retenir l'université a ou l'université b selon le critère auquel on attache le plus de poids. Une autre politique possible consiste à préférer une université bonne sur les deux aspects considérés, c'est-à-dire ne présentant aucun point faible ou aucune lacune. Dans ce cas, c'est l'université c qui pourrait être choisie. L'outil technique utilisé pour désigner « la meilleure » université devrait permettre de traduire ces différentes politiques.

Un calcul simple, non reproduit ici, montre qu'il n'en est rien : il est mathématiquement impossible que l'université c sorte première du classement si on utilise, comme technique d'agrégation, la moyenne pondérée, quels que soient les poids attribués aux critères. Le choix de la technique d'agrégation procède donc bien d'un choix politique implicite.

## V. Conclusions

La plupart des dirigeants d'universités affirment que les *rankings* dont la presse se fait l'écho sont très discutables. Néanmoins, qu'on le veuille ou non, le public et le monde académique considèrent de plus en plus ces *rankings* comme représentant « la » vérité sur la qualité des universités. Les arguments verbaux peuvent toujours se discuter, pas une note sur 100 : c'est une information « objective » ...

Et même si les *rankings* trahissent la réalité aujourd'hui, ils finiront par la représenter demain, parce qu'ils deviendront la référence sur laquelle les universités se baseront pour mener leur politique, tant il est vrai qu'une université qui n'intègre pas les critères des *rankings* dans sa gouvernance risque fort de devenir une université de seconde zone.

Soyons néanmoins attentifs au fait qu'un développement à outrance de procédures d'évaluation et de classement peut aussi conduire à une détérioration progressive de l'espace de liberté indispensable pour créer de nouvelles connaissances. L'émulation existe déjà entre les chercheurs et il est sain qu'elle existe entre les universités, mais attention aux dérives.

Dans certaines universités chinoises aujourd'hui, des primes sont versées aux chercheurs en fonction de leur contribution au bon classement de leur université. Dans le contrat de Michael CROW, président de l'Université d'Arizona, figure une clause qui prévoit une prime de 10 000 dollars s'il améliore la position de son université. (Voilà de quoi donner des idées aux recteurs de nos universités !).

Que dira-t-on lorsque les universités se disputeront les lauréats des prix Nobel ou les champions des publications à coup de millions de dollars ou d'euros, comme les clubs de football le font pour les champions du ballon rond ?

Soyons attentifs au fait que toutes les universités du monde risquent d'adapter leur comportement aux *rankings* les plus populaires, avec tous les effets pervers que cela peut entraîner. Nous pourrions par exemple suggérer, Monsieur le Président, au prochain Conseil d'administration, de fermer les Facultés les moins rentables en terme de *ranking*. Pareille mesure aurait un effet bénéfique immédiat pour notre classement, de la même façon que la suppression d'emplois dans une entreprise fait souvent grimper le cours de ses actions en Bourse dans les jours qui suivent l'annonce des licenciements.

N'oublions pas que la publication d'un classement modifie la réalité qu'il est supposé mesurer. L'évaluation annuelle des universités pourrait conduire à des effets de mode, à un alignement progressif de toutes les institutions sur les mêmes standards, à une uniformisation de leurs profils et donc à un appauvrissement de l'offre académique. La manière de comparer les universités entre elles a inévitablement une influence sur leur politique et donc sur l'évolution du paysage universitaire.

Mon objectif n'était pas de vous amener à rejeter toute idée d'évaluation des universités mais d'essayer de vous convaincre que les principaux classements relayés par la presse ne sont pas, à ce jour, scientifiquement valables.

Evaluer la qualité de l'enseignement supérieur et de la recherche, ce n'est pas représenter une réalité objective, mesurable sur une échelle de 0 à 100. C'est une question complexe, qui doit être traitée scientifiquement, par des personnes compétentes (en bibliométrie, en matière d'enseignement supérieur et de recherche, en techniques d'analyse et de traitement de données), autrement dit par des équipes pluridisciplinaires, capables d'analyser avec précision l'interprétation et l'utilisation des résultats des évaluations. Ces compétences, nous les avons dans nos universités. Alors, pourquoi confier notre destin à des chimistes chinois ou à des entreprises commerciales anglo-saxonnes ?

Evaluer la qualité de l'enseignement supérieur et de la recherche n'a de sens qu'après avoir défini une politique d'enseignement supérieur et de recherche. Cette politique, il nous revient de la déterminer.

Je suggère donc que les universités de la Communauté française unissent leurs efforts pour expliciter leur politique d'enseignement supérieur et de recherche et que, sur cette base, elles développent les outils de son évaluation, en s'inspirant bien entendu des pratiques et des standards internationaux.

Parallèlement, il me paraît important que les responsables européens de l'enseignement supérieur et de la recherche poursuivent le même travail, afin de permettre à l'espace européen de l'enseignement et de la recherche de se développer en toute autonomie.

## **VI. Référence récente contenant l'essentiel de la bibliographie**

Jan SADLAK and Liu NIAN CAI (editors), The world-class university and ranking : aiming beyond status, Unesco-Cepes, 2007.

## **VII. Sites utiles**

- <http://www.isihighlycited.com>
- <http://ed.sjtu.edu.cn/ranking.htm>
- <http://www.thes.co.uk>