

UMR ADEF

JOURNAL DU SEMINAIRE TAD/IDD

Théorie Anthropologique du Didactique & Ingénierie Didactique du Développement

Toda pregunta implica la pérdida de una intimidad o el extinguirse de una adoración. María Zambrano (1904-1991)

There is a phrase I learned in college called, "having a healthy disregard for the impossible." That is a really good phrase. Larry Page (1973-?)

Le séminaire TAD & IDD est animé par Yves Chevallard au sein de l'équipe 1 de l'UMR ADEF, dont le domaine général de recherche s'intitule « École et anthropologie didactique des savoirs ».

Ce séminaire a, solidairement, une double visée. D'une part, il vise à mettre en débat des recherches (achevées, en cours ou en projet) touchant à la TAD ou, dans ce cadre, à des problèmes d'ingénierie didactique du développement, quel qu'en soit le cadre institutionnel. D'autre part, il vise à faire émerger les problèmes de tous ordres touchant au développement didactique des institutions, et notamment de la profession de professeur de mathématiques.

Deux domaines de recherche sont au cœur du séminaire : un domaine en émergence, la didactique de l'enquête codisciplinaire ; un domaine en devenir, la didactique des savoirs mathématiques.

La conduite des séances et leur suivi se fixent notamment pour objectif d'aider les participants à étendre et à approfondir leur connaissance théorique et leur maîtrise pratique de la TAD et des outils de divers ordres que cette théorie apporte ou permet d'élaborer.

Sauf exception, les séances se déroulent le vendredi après-midi, de 15 h à 17 h puis de 17 h 30 à 19 h 30, cette seconde partie pouvant être suivie en visioconférence.

→ Séance 5 – Vendredi 21 mars 2008

SUR LA TAD

1. Deux problématiques de recherche

a) La didactique peut être définie comme *la science des conditions et des contraintes de la diffusion sociale des praxéologies*. Cette définition demanderait bien des précisions complémentaires, par exemple sur ce fait que les *mécanismes et autres processus* de diffusion sont ici subsumés sous les « conditions et contraintes » évoquées ; sur le fait aussi que la « diffusion », c'est aussi la *rétenion*, éventuellement délibérée (il y a alors, si l'on peut dire, *intention anti-didactique*) de praxéologies. Bien entendu, la notion même de praxéologie, centrale, devrait être précisée. Mais je n'irai pas plus loin dans ces rappels.

b) Voilà dont la définition de *la didactique* qu'on peut aujourd'hui adopter dans le cadre de la TAD. Cette dernière se propose comme cadre théorique (et technologique, et pratico-technique) au développement de la science didactique entendue en ce sens. Cela rappelle, de quels *objets* nous parle la

TAD ? Comme on le sait, il y est question de *personnes* et d'*institutions*, ainsi que de la *dialectique des institutions et des personnes*. Il y est question aussi d'*organisations praxéologiques* vivant en tel ou tel « habitat » ou ne parvenant pas à y vivre – je veux dire en telle personne ou en telle institution. Il y est question encore d'*équipements praxéologiques* (institutionnels ou personnels), de *conditions et de contraintes* relatives à l'*écologie praxéologique*, et qui gouvernent les *dynamiques praxéologiques* éventuelles ou, au contraire, à ce qu'on peut appeler les *stases préxéologiques*, en rangeant parmi ces dernières tous les faits de régression praxéologique ou de dégradation de l'équipement praxéologique d'une personne ou d'une institution.

c) D'une façon très générale, une première *problématique de recherche* en didactique peut être repérée par la question suivante :

Étant donné les contraintes pesant actuellement sur elle, sous quelles conditions – existantes ou à créer – une institution ou une personne donnée peut-elle intégrer à son équipement praxéologique une organisation praxéologique désignée ?

→ Il s'agit là sans doute du questionnement le plus classique en didactique, tout au moins quand on l'applique à « la classe » : lorsque, en une civilisation et une société données, une école est établie, une pédagogie réglée, un programme fixé, le didacticien peut entrer en scène : sous des contraintes auxquelles il n'est pas supposé toucher, il est alors censé pouvoir démêler les conditions « didactiques » – les seules qu'on l'autorise chichement à manipuler, au moins sur le papier – qui permettront ou faciliteront une certaine diffusion dans la classe de praxéologies que le programme désigne.

→ Les travaux de recherche conduits dans cette problématique se donnent souvent pour objectif de mieux connaître quelques-uns des objets partiels qui y interviennent. C'est ainsi par exemple qu'on étudiera les praxéologies désignées comme « à diffuser », leur composition, leurs cohérence, leur complétude, leur degré d'intégration, etc. De même pourra-t-on examiner leur transposition – dans les manuels, dans les classes, etc. – sous les *contraintes imposées*, étude qui devrait se prolonger par l'examen des effets transpositifs des *conditions envisagées*, que celles-ci soient établies dans l'institution observée ou que le didacticien imagine simplement de les y voir établies. On s'attachera encore, classiquement, à explorer plus ou moins finement l'équipement praxéologique *préalable* des acteurs de l'institution où la diffusion desdites praxéologies doit se faire – c'est-à-dire de mettre au jour certaines des *contraintes* (en l'espèce, d'origine sociale ou scolaire) sous lesquelles ladite diffusion devrait s'opérer ; etc.

d) Ce qui toutefois reste largement absent du champ de la conscience didacticienne commune, ce sont ces *contraintes* dont les effets ne sont pas supposés être « portés » par les acteurs directs ou très proches de l'institution – de la classe, par exemple. Cette indifférence à ce qui *peut* faire la différence procède d'un schéma largement répandu, bien qu'il soit antinomique du paradigme scientifique pourtant le plus assuré : selon ce schéma, le comportement d'un « système » s'expliquerait par des propriétés

propres au système, sans qu'on ait à se préoccuper des contraintes *environnementales*. Dans une classe, il y aurait ainsi les propriétés du professeur, les propriétés de tel ou tel élève, celles des praxéologies étudiées, voire celle de « la classe » saisie comme un tout. C'est là bien entendu un point de vue restrictif, qui oblige à prêter au professeur, à l'élève, au savoir enseigné *beaucoup plus* qu'il n'est nécessaire et juste, et dont il me semble que bien peu de courants de recherche en didactique ont cherché pourtant à s'affranchir d'une façon autre qu'erratique.

➔ Je prendrai ici un exemple récemment porté à mon attention. Dans des travaux poursuivis depuis des années, nous avons identifié une « variable » dont la « valeur » nous paraît lourde de conséquences dans les classes (et dans la formation des professeurs) : « la profession ». De quelle profession s'agit-il ?, demande, me dit-on, un commentateur mal avisé. S'attend-il à ce qu'on lui réponde : la profession de serrurier ? Celle de marchand ambulant de crème glacée ? Ou celle de moniteur de ski ? Bien entendu, il s'agit tout simplement de la profession *de professeur de mathématiques* dans l'enseignement secondaire – profession que, lorsque le contexte ne laisse pas d'ambiguïté sur ce point (ce qui est presque toujours le cas dans les travaux susmentionnés), nous appelons « la profession », tout court, en empruntant un raccourci qu'on nomme en rhétorique une *antonomase d'excellence* (voyez cela dans l'article « Antonomase » de l'encyclopédie Wikipedia).

➔ Donc, il y a des professions, et il y a en particulier *la* profession (de professeur de mathématiques). Je saisis l'occasion pour préciser quelques aspects de la notion cardinale de profession. Je distingue, à propos de chaque activité sociale régulière, réglée, généralement rémunérée, le *métier* et la *profession*. Par exemple, lorsqu'un serrurier est appelé en urgence, prend ses outils, se rend chez le client, ouvre la porte qui résistait aux efforts de celui-ci, il *exerce son métier*. Mais par ailleurs ce serrurier appartient à une *profession*, qui a ses écoles de formation (il existe un CAP de serrurerie métallerie, par exemple) ; qui a aussi, comme on dit, ses « organisations professionnelles », etc. D'une façon générale, *la profession* joue un rôle essentiel dans la « fabrication » *du métier*, de ses critères, de ses exigences, de sa « qualification » : elle peut faire qu'il s'agisse durablement d'un *quasi-non-métier* (tout le monde pourrait l'exercer, même s'il existe un malthusianisme corporatiste traditionnel qui, en filtrant les entrants, prétend s'assurer de leur *qualité*), ou, à l'apposé, d'un *métier hautement qualifié* (même si l'on peut y entrer « par la petite porte »). La profession crée un système de conditions et de contraintes qui façonnent et régulent le métier, le faisant être ce qu'il devient. Dans le métier de professeur de mathématiques, par exemple (je répète là des exemples fort anciens pour plusieurs d'entre nous), « la profession » impose aux gens de métier de corriger (en les annotant) les travaux des élèves – obligation dont bien peu d'entre eux s'exonèrent *motu proprio*. Elle oblige aussi à « préparer ses cours ». Mais elle n'impose pas *d'analyser* les séances de classe *après qu'elles se sont déroulées* – ce que d'autres professions font pourtant avec la dernière énergie, comme il en va dans le métier de footballeur (professionnel) par exemple.

→ J'ai dit plus haut que « la profession » est une « variable », et qui peut prendre diverses « valeurs ». Dans les travaux réalisés jusqu'ici, la modélisation est, à cet égard, restée des plus sommaires. Tout d'abord, on y pose qu'il y a bien, qu'il y a toujours *une profession*, émanation de la noosphère du métier plus ou moins forte, plus ou moins développée, plus ou moins intégrée. C'est cela que l'on nie quand on prétend que le métier n'est rien d'autre que ce que font les gens de métier – et qu'il serait donc un archipel fait de myriades d'îles qu'aucune profession ne saurait pour cela représenter ni, à plus forte raison, gouverner. (Nous ne sommes sans doute pas loin de cet état de choses s'agissant du métier de professeur de mathématiques : chacun est une île, seul compte à mes yeux ce que *je* fais, etc.) Au-delà de ce principe, nous n'avons distingué jusqu'ici que *deux* valeurs de la variable « profession » : ou bien la profession considérée est une... profession, ou bien c'est une *semi-profession* – notion popularisée autrefois par Amitai Etzioni dans son livre *The semi-professions and their organizations: teachers, nurses and social worker* (New York Free Press, 1969). Au passage, on notera ce genre de télescopage terminologique qu'adorent dénoncer ceux qui ne peuvent s'attaquer au fond : une profession ne serait donc pas toujours une profession, et patati et patata. C'est là le statut actuel de *la* profession, qui demeure aujourd'hui encore, selon moi, une semi-profession. Une profession, par exemple, n'aurait pas permis que, pendant des années, ses membres enseignant en collège se trouvent aux prises avec l'ineptie qui s'est glissée dans le programme de mathématiques de 5^e à propos des angles alternes-internes, programme qui, à la fois, porte à ne pas regarder comme alternes-internes les angles alternes-internes formés par une sécante à deux droites *non nécessairement parallèles*, et à établir que l'égalité desdits angles alternes-internes entraîneraient le parallélisme de ces droites. (Sur cette question, on peut pour le moment consulter la thèse de Gisèle Cirade.) De même, pour prendre un exemple bien connu ici, une profession qui en serait une ne permettrait pas qu'aucun de ses membres disent – à l'adresse de ses élèves ! – des choses comme : « La calculatrice nous propose le même affichage pour les nombres $\sqrt{48}$ et $4\sqrt{3}$, certes ; mais cela ne nous permet pas de conclure, car il se pourrait que les deux nombres diffèrent par une décimale située au 20^e rang, ou au 50^e, ou, pourquoi pas, au milliardième rang ! » Je rappelle en passant le minuscule travail mathématique qui justifie cette vertueuse indignation. S'il était vrai que $\sqrt{48}$ est différent de $4\sqrt{3}$, et donc que $|\sqrt{48} - 4\sqrt{3}| \geq 1$, on aurait aussi

$$|\sqrt{48} - 4\sqrt{3}| = \frac{|(\sqrt{48})^2 - (4\sqrt{3})^2|}{\sqrt{48} + 4\sqrt{3}} \geq \frac{1}{\sqrt{48} + 4\sqrt{3}} \geq \frac{1}{7 + 4 \times 2} = \frac{1}{15}$$

soit (ici) $|\sqrt{48} - 4\sqrt{3}| > 0,06$, ce qui se verrait à l'affichage !

e) Deux remarques complèteront le bref panorama précédent. D'une part, lorsqu'on s'interroge à propos de la diffusion de telle ou telle praxéologie en telle ou telle institution, toute *contrainte* dont on peut montrer *l'existence* dans la société ou la civilisation n'est pas nécessairement *sensible* dans le

cas que l'on envisage ; en sorte que l'on peut alors, à bon droit, *ignorer* tel ou tel paquet de contraintes, sous réserve d'avoir de bonnes raisons de le faire. (Dans le cas des angles alternes-internes comme dans celui des liens entre développements décimaux et comparaison de nombres, nous avons de bonnes raisons de penser que *l'état actuel de la profession* est une contrainte clé.) D'autre part, et en sens inverse, il faut en règle générale aller identifier comme sensibles des contraintes dont le siège paraît *éloigné*, dans l'univers cognitif culturellement contraint de nombre de didacticiens, de la scène principale de l'action : il faut ainsi prendre en compte de multiples *autres scènes*, dont les contenus se co-déterminent sans qu'il soit d'emblée évident d'y découper un système quasi isolé. Dès lors que l'on souscrit à cette exigence scientifiquement irréfutable, on est amené à sortir du « petit monde » du métier de professeur de mathématiques *tel que le peint la profession dans son état actuel* – et tel que l'envisagent à sa suite les didacticiens qui n'ont pas approfondi suffisamment la rupture épistémologique avec le point de vue du monde social et professionnel qu'ils étudient. Vous n'avez pas été sans observer que la chose surprend presque toujours, quand elle ne choque pas. En bien des cas, en fait, elle choque, et l'on entend alors ce vieux refrain selon lequel nous n'aurions pas à connaître de ceci ou de cela, que prétendre le faire serait presque obscène, etc.

f) S'affranchir des assujettissements qui font tenir debout l'institution qu'il étudie, et notamment répudier ces assujettissements qui permettent qu'existe au sein de cette institution telle ou telle *position* institutionnelle constitue une *exigence première* du chercheur. Le didacticien n'est pas le professeur ; or la problématique « classique » énoncée plus haut ressemble étrangement à celle du professeur ! D'autres problématiques existent, qui peuvent au reste fort bien démarquer la perspective spontanément attachée à telle ou telle position, autre que celle de professeur, au sein des institutions d'enseignement. Je m'arrêterai dans ce qui suit sur l'une d'elles, que je vais m'efforcer de formuler dans les termes les plus généraux, les moins particularistes (et donc les plus réalistes).

➔ Je pars pour cela d'une notion clé : celle de *projet*. (Cette notion n'est pas nouvelle en TAD : elle m'a été indispensable, par exemple, pour articuler l'esquisse d'une théorie de l'évaluation.) Une institution a un projet ; une personne a un projet. Voilà le point de départ. Par exemple, une institution de formation professionnelle a le projet de créer en son sein une nouvelle filière de formation (qui conduise à des métiers anciens ou nouveaux). Une personne a un projet professionnel : elle veut créer une société de création de sites Web dont elle sera d'abord l'unique salarié. Un petit groupe d'amis a un projet de loisir de longue durée : faire le tour du monde à la voile. Des élèves d'une classe de 3^e ont un projet d'étude et de recherche : découvrir l'origine des relations mathématiques « données » dans une fiche disponible sur Eduscol qui leur ont permis de calculer des distances de freinage (voir http://eduscol.education.fr/D0160/07assr_3maths_distancearret.pdf). Etc.

➔ Pour concevoir et mettre en œuvre un projet *P*, un équipement praxéologique Π sera *utile*, et sans doute indispensable en certaines de ses

parties. C'est alors que s'élève une question cruciale : de quel équipement praxéologique Π a-t-on *besoin* ? Que sera, en d'autres termes, la composition de cet équipement ? De quelles praxéologies sera-t-il fait ? Cette question, au vrai, est mal posée, car les critères de pertinence et d'adéquation manquent en général : le *besoin* est chose incertaine, et qui ne s'inscrit qu'inexactement dans la *demande*. Une voie d'attaque possible est alors la suivante – il s'agit là d'un schéma que je crois éclairant, même si sa faisabilité n'est jamais complètement assurée. La conception et la réalisation du projet P s'inscrivent dans une histoire où les acteurs X du projet vont rencontrer des questions $Q_1, Q_2, \dots, Q_p$. Face à chacune de ces questions Q_ℓ , X ouvrira un *parcours d'étude et de recherche* E_ℓ visant à fabriquer une réponse R_ℓ (ou plutôt $R_\ell^\forall$) au sein d'un système didactique qu'on peut noter

$$S(X_\ell; Y_\ell; Q_\ell)$$

où $X_\ell \cap X \neq \emptyset$. En d'autres termes, chaque fois, une partie des porteurs du projet P va à l'école – à propos de la question Q_ℓ . Le bilan d'un tel parcours peut se représenter comme l'on sait par la formule suivante (on y verra que j'ai retouché un peu les notations usitées jusqu'ici) :

$$[S(X_\ell, Y_\ell; Q_\ell) \Rightarrow R_1^\diamond(\ell), R_2^\diamond(\ell), \dots, R_n^\diamond(\ell), O_{n+1}(\ell), \dots, O_m(\ell)] \Rightarrow R_\ell^\forall.$$

La réception des réponses $R_j^\diamond(\ell)$ suppose un certain équipement praxéologique – toujours mal défini – qu'apporte en principe l'étude de certaines des œuvres $O_k(\ell)$, que d'autres de ces œuvres compléteront pour permettre la production de $R_\ell^\forall$. Tout cela ($R_j^\diamond(\ell), O_k(\ell), R_\ell^\forall$) viendra au moins partiellement s'ajouter à l'équipement praxéologique antérieurement disponible, produisant ainsi un certain équipement praxéologique Π_ℓ . La répétition de ce processus à propos des différentes questions Q_ℓ ($1 \leq \ell \leq p$) conduit finalement à un équipement praxéologique $\Pi_P = \Pi_p$. On notera que, bien entendu, dans la suite $Q_1, Q_2, \dots, Q_p$, il se peut que l'on ait $Q_\ell = Q_{\ell+k}$: en d'autres termes, l'étude de la question Q_ℓ pourra être *reprise*, peut-être même *plusieurs fois*, si l'avancement du projet P en venait à faire apparaître la réponse $R_\ell^\forall$ construite jusque-là comme inadéquate sous les contraintes ou dans les conditions où le projet P serait désormais amené à se développer.

➔ L'étude du schéma précédent est au cœur de la *didactique de l'enquête codisciplinaire*. La question soulevée initialement – quelles sont les praxéologies *utiles* à la conception et à la réalisation du projet P ? – y trouve un principe de réponse : ce sont les praxéologies qui se révéleront *utiles* à la fabrication de réponses $R_\ell^\forall$ adéquates à P . On notera que, bien entendu, le problème n'est, formellement, que reconduit à un autre niveau, mais dans une formulation qui apparaît plus précise : l'utilité praxéologique a maintenant pour critère d'adéquation le projet d'enquête E_ℓ , non le projet plus vaste P – encore que l'adéquation de $R_\ell^\forall$ elle-même se juge à son utilité vis-à-vis de P .

→ Tout cela se situe d'abord au niveau de l'*économie* didactique : décider d'étudier Q_i , en venir à décider, un peu plus tard, de « se rendre disponible » telle œuvre $O_k(\ell)$ sont autant de décisions en matière d'*économie didactique*, dont, comme toujours, la justesse n'est jamais certaine : il n'apparaîtra que plus tard, peut-être, que telle œuvre, $O_r(\ell)$, était décidément incontournable, tandis que telle autre, $O_s(\ell)$, aurait pu, sans que cela nuise en rien à la conception et à la réalisation du projet P , être ignorée.

e) Dans le schéma précédent, lorsque, pour tout ℓ , on a $Y_\ell = \emptyset$ (et $X_\ell = X$), le collectif X (parfois réduit à une personne) se forme en autodidaxie. Chose plus importante peut-être, X se forme alors généralement *sur le tas*, c'est-à-dire au plus près de son « travail » – concevoir P et le réaliser. La distance est alors minimale entre le besoin praxéologique éprouvé et l'effort pour le satisfaire ; la *skholê*, ici, est proche, au plus près de ce qui l'engendre et la motive. En ce cas, l'équipement praxéologique supposé utile est déterminé *au fur et à mesure*, avec de rarissimes anticipations, de façon *fonctionnelle*, en se trouvant soumis continuellement à l'épreuve de cette pierre de touche quasi unique qu'est l'heureux avancement de P .

→ Paradoxalement, cette manière de faire soulève un problème d'adéquation de l'équipement praxéologique ainsi dégagé : elle privilégie en effet la *phénoménologie* des difficultés rencontrées, des questions soulevées, ou, disons, leur *apparence* (c'est-à-dire la manière dont elles apparaissent à X), au risque d'oublier ce qui devient alors une *nouménologie*, qui, derrière les phénomènes apparents, donnerait leur place à des réalités « transcendantes » par rapport à l'immédiateté de l'action de X et des obstacles que X croit y déceler. Bien entendu, la théorie kantienne des phénomènes et des noumènes est prise ici en un sens métaphorique et, surtout, de manière *relative* : car un « noumène », lorsqu'il apparaît, devient phénomène. Cette remarque, au reste, n'a rien de philosophiquement hétérodoxe – dans son *Vocabulaire philosophique* (Armand Colin, 1901), Edmond Goblot écrivait (p. 367) par exemple : « ... s'il se révèle à la pensée, [le noumène] devient phénomène. »

→ On constate un « écrasement phénoménologique » de l'équipement praxéologique acquis *sur le tas*, souvent *à la dure*, alors même que cet équipement praxéologique semble bien adapté aux besoins *visibles* de l'activité, ce qui, à juste titre, ne manque pas de faire impression sur qui ne peut s'en prévaloir : l'un « sait faire », l'autre pas. Notons en passant que cette réduction phénoménologique du problème des praxéologies utiles ne concerne pas seulement le *logos* mais aussi la *praxis* elle-même : les critiques émanant de points de vue « nouménologiques » existants par ailleurs dans la société pourront ainsi mettre en cause non seulement ce que X *sait* (ou plutôt *ne sait pas*), mais encore ce qu'il *fait*, et *comment* il le fait, etc.

→ Un changement structurel dans le schéma pris plus haut pour point de départ apparaît quand quelque Y propose à des x, x', x'' ... intéressés une *formation à la conception et à la réalisation de projets du type de P* , soit ce

que je nommerai de façon plus compacte *une formation* à $[P]$, et qu'on peut donc noter ainsi :

$$S(\{x, x', x'' \dots\} ; Y ; [P])$$

Ici, c'est Y qui définit (ou semble définir) l'équipement praxéologique $\Pi_{[P]}$ supposé adéquat : le problème de déterminer $\Pi_{[P]}$ n'en est donc plus un pour $x, x', x'' \dots$, qui s'en remettent à Y . (Celui-ci est en général mandaté par une école Σ , responsable *en dernier ressort* du choix de $\Pi_{[P]}$.) Le choix proposé par Y découle-t-il alors d'un processus de détermination *fonctionnelle* ou, à l'opposé, d'une détermination plus ou moins *formelle*, avec tous les degrés imaginables d'extériorité par rapport à l'enjeu didactique $[P]$? À observer certaines formations existantes, il semblerait que le travail d'*exploration fonctionnelle* qui devrait inspirer (puis rectifier) le choix des contenus praxéologiques ne soit guère fait, et que ces choix procèdent plus souvent d'un pur et simple *recopiage curriculaire*, marqué par un certain conformisme, jusque dans l'emprise observable des modes du moment. On retrouve là, paradoxalement, la situation de l'enseignement secondaire, où une instance supérieure, quasi transcendante, fait entendre sans appel ses verdicts – que l'on n'a plus alors qu'à suivre. Dans un cas, l'obéissance est statutaire ; dans l'autre, le souci parfois éperdu d'être conforme ou de se mettre en conformité l'emporte sur l'exigence d'utilité vraie.

➔ Il est un cas qu'il faut noter encore parce qu'il rompt avec une fiction reçue dans l'enseignement secondaire (en dépit des IDD et autres TPE) : il peut se faire que, face à la demande de $x, x', x'' \dots$, une école Σ demande à une équipe Y de proposer une « formation » concrétisée dans la mise en fonctionnement d'un système

$$S(\{x, x', x'' \dots\} ; Y ; [P])$$

alors même que Y ignore à peu près tout du domaine d'activité $[P]$ et n'a pas grande idée quant à l'équipement praxéologique $\Pi_{[P]}$ qu'il pourrait mobiliser utilement ! Une telle situation interdit seulement à $y \in Y$ de *professer*, et même d'*enseigner* – de « montrer » ce que lui-même n'a jamais vu. Mais Y peut efficacement jouer le rôle d'*aide à l'étude* et, surtout, de *directeur d'étude* – l'étude, ici, consistant d'un même mouvement à tenter et d'*explorer*, et de *satisfaire* les besoins praxéologiques suscités par $[P]$, c'est-à-dire à s'efforcer de *déterminer* et de *maîtriser* un équipement praxéologique $\Pi_{[P]}$.

f) Le problème du choix des contenus praxéologiques d'une formation donnée, précisée d'abord par son objet déclaré, $[P]$, a été jusqu'ici, je crois, *fort peu étudié*. Corrélativement, un tel choix participe encore aujourd'hui, très largement, d'une « cuisine » institutionnelle (ou personnelle) qui échappe et entend échapper à l'observation démocratique.

➔ Une illustration de ce fait massif a été récemment fournie par un article dû à deux anciens ministres de l'Éducation nationale, Jack Lang et Luc

Ferry, qui y critiquent vivement les programmes du primaire rendus publics par l'actuel ministre Xavier Darcos : on en reproduit ci-après un extrait révélateur (voir <http://hebdo.nouvelobs.com/hebdo/parution/p2262/articles/a369288-.html>).

S'agissant des nouveaux programmes, nul ne parvient à savoir, pas même les anciens ministres de l'Éducation que nous sommes, comment et par qui ils ont été rédigés ! Et pour cause. Les groupes d'experts, présidés et composés par des personnalités identifiables et reconnues, ont disparu. Le Conseil national des Programmes a été supprimé, et l'Inspection générale elle-même n'a pas été saisie du dossier ! Est-il raisonnable de laisser de simples conseillers du ministère ou de l'Élysée élaborer dans l'opacité la plus totale des textes voués à régir l'école de la nation pour dix ans au moins et qui concernent des millions de familles et de citoyens ?

→ On ne s'étonnera donc pas que, fréquemment, la *genèse* et la *motivation* des contenus praxéologiques d'une formation, proposée par une « institution enseignante » Y mandatée en général par une institution « scolaire » Σ , apparaissent des plus obscures, y compris aux $y \in Y$ eux-mêmes ! Cette remarque attire plus encore l'attention sur une problématique de recherche que l'on peut formuler ainsi.

Étant donné les contraintes pesant sur elle et un ensemble de conditions existantes ou à créer, à quel équipement praxéologique est-il possible qu'une institution ou une personne donnée accède pour opérer dans un domaine d'activité donné ?

→ Ce *schéma de question* appelle un commentaire portant sur son apparente « évidence », ce qui est un symptôme – à mes yeux – du fait que, en règle générale, *nous ne nous posons pas ce genre de questions*, parce que la réponse serait réglée d'avance – au double sens du terme – par des institutions jugées « compétentes ». Or, quand on examine un tant soit peu ce que, pour tel domaine d'activité, pourrait être la réponse, on s'aperçoit que la situation est on ne peut plus complexe, variable, inattendue. On peut par exemple le voir si l'on imagine un instant les variations dans le temps historique et dans l'espace géographique d'une « formation à... ». Que peut-on apprendre en matière de TAD, aujourd'hui, en Suisse, par exemple ? Et en Belgique ? Et au Royaume-Uni ? Et en Italie ? Et que pourra-t-on y apprendre dans dix ans ? Plutôt que de mentionner rapidement divers exemples, je me centrerai, dans ce séminaire, lors des séances qui suivent, sur trois sujets privilégiés, auxquels d'autres pourront éventuellement venir s'ajouter.

→ Voici le premier sujet. On peut observer dans le cadre de certaines formations en sciences de l'éducation à Lambesc que des étudiants évoquent avec quelque insistance le projet de se former à la *recherche d'informations sur Internet*, ou comme je dirai, à la RII. (D'autres au contraire considèrent ce qu'ils font à cet égard comme ne pouvant guère être amélioré, les progrès attendus passant seulement par l'amélioration des outils de recherche – qui seraient par exemple responsables de l'*information overload* dont d'aucuns disent souffrir.) En cette affaire, l'institution « scolaire » concernée est

apparue, jusqu'à présent, relativement silencieuse, ce qui peut suggérer une certaine indifférence actuelle à l'objet de la demande des étudiants.

➔ Voici maintenant le deuxième des trois sujets annoncés. Le cadre institutionnel est le même que précédemment : les sciences de l'éducation, à Lambesc. Le projet, en l'espèce, est de se former à *l'analyse de données* : beaucoup d'étudiants, semble-t-il, s'y engagent de façon sincère, même s'il est vrai qu'ils y sont peu ou prou incités, voire contraints, par l'institution à laquelle ils sont venus s'assujettir. Car cette fois, l'institution scolaire n'est nullement indifférente ; elle propose une formation dont les contenus praxéologiques sont sans doute déterminés *en partie* de façon fonctionnelle.

➔ Voici enfin le troisième sujet : il a trait, non à des étudiants en sciences de l'éducation, mais à celles et ceux qui suivent ce séminaire, et porte sur le fait de s'instruire en matière de « style éditorial » ou, plus précisément, d'orthotypographie. En ce cas, on peut dire que la fragile institution qu'est le séminaire TAD/IDD propose timidement une « formation » (où *y* ne saurait se permettre de *professer*, même s'il se risque parfois à *enseigner*). Je ne me prononcerai pas sur le fait de savoir si ce projet de formation est partagé par les participants à ce séminaire ; je me contenterai ici de m'interroger sur l'écho qu'a eu parmi eux l'invitation à se pencher sur l'ouvrage *Toute l'orthographe*, que je m'étais permis de vous recommander...

➔ Dans chacun de ces trois cas, la question est donc : à quel équipement praxéologique (respectivement en matière de RII, d'analyse de données, de typographie), les personnes et, avec elles, telle institution à laquelle elles se seraient assujetties, pourraient accéder, et cela sous quelles conditions existantes ou à créer ? Par rapport à cette question, on peut formuler une question plus large encore, qui désigne un horizon problématique : dans tel système d'institutions considéré, quelle est *l'offre praxéologique* en telle matière (RII, analyse de données, typographie) ? Par exemple, si l'on prend pour système institutionnel les institutions en charge de la scolarité obligatoire en France, on voit aisément que l'offre y est quasi nulle en matière de typographie, alors qu'elle est simplement très faible ou faible dans les deux autres domaines. On pourra de même s'interroger sur l'offre praxéologique à l'adresse des étudiants en sciences de l'éducation de Lambesc en matière de RII, d'analyse des données ou de typographie. Etc.

➔ Bien entendu, on peut accéder, sous des conditions appropriées à préciser, à un équipement praxéologique qui ne fait pas partie de l'offre disponible dans tel système d'institutions, mais qui fait partie de l'offre disponible en tel ou tel autre système d'institutions. On a ainsi l'idée d'une *totalité de l'offre disponible*, totalité qu'on peut explorer et cartographier localement ou régionalement, sinon globalement. Une telle exploration peut être suscitée par le projet de satisfaire des besoins praxéologiques précis, auquel cas elle aura pour critère d'arrêt la satisfaction de ces besoins (ou la conviction que, sous les contraintes existantes, cette satisfaction n'est pas possible). Ce peut être aussi une exploration plus systématique, participant

d'une *clinique praxéologique à visée didactique*, partie essentielle de la *clinique didactique*. C'est à ce dernier thème que je m'attacherai maintenant.

2. Clinique didactique

a) La clinique didactique soulève des problèmes cruciaux auxquels je voudrais consacrer un peu de place dans la suite des séances de ce séminaire. Les objets de la clinique sont ceux-là même de la didactique. À titre d'illustration, j'esquisserai aujourd'hui un petit nombre de considérations sur la question de l'équipement praxéologique d'une personne ou d'une institution. Comment connaître cet équipement ? Comment « prendre de l'information » à son sujet ?

→ Un premier postulat est que toute prise d'information est *partielle*, et que l'on peut seulement espérer multiplier les angles de vue et les situations d'observation. Il convient donc d'emblée de rejeter le fantasme de *tout voir* et de *tout connaître*. Une conséquence de cette limitation (dont le tracé concret peut varier, être déplacé, etc.) est que toute conclusion sera toujours *en partie conjecturale*, avec des parties très incertaines et des parties beaucoup plus sûres, selon l'étendue de la *dialectique des médias et des milieux* à laquelle elles auront été soumises.

→ Un deuxième postulat énonce le primat de l'*observation* sur l'*interrogatoire* (par entretien ou par questionnaire) : à l'instar de la connaissance du monde naturel non humain (minéral, végétal, animal), la connaissance du monde social suppose l'*observation* des personnes et des institutions, quand bien même leur *état* aurait été volontairement « perturbé » par l'observateur, comme il en va dans une *expérience* (laquelle est une perturbation en principe « contrôlée »). Comme en médecine, l'interrogatoire est un adjuvant de l'observation clinique : qu'il la précède ou la complète pour la relancer, l'interrogatoire est là pour orienter l'observation, en suscitant notamment un questionnement et un travail sur les objets à observer et sur les façons de le faire.

→ Un troisième postulat, bien connu, est que toute observation est susceptible de modifier *l'engagement praxéologique des personnes observées*. Cette remarque est à étoffer par plusieurs considérations.

1° Lorsqu'une telle modification est importante, il se peut que l'observation qui la suscite ne permette plus guère qu'une prise d'information sur le comportement des personnes *dans une situation perturbée par l'observation* – ce qui, au reste, peut être en quelques cas *l'enjeu même de l'observation* !

2° Certains équipements praxéologiques, à l'inverse, sont dotés d'assez d'*inertie* pour que l'observation ne puisse guère les modifier ou modifier leur mise en œuvre dans la situation observée.

3° Dans l'ensemble des éléments praxéologiques mobilisés par les personnes observées, il en est par rapport auxquels ces personnes se trouvent dénuées d'*intention* vis-à-vis de l'observateur, en sorte que cette source importante de modifications potentielles n'existe pas.

4° Les notations précédentes, formulées par rapport à l'observateur (lorsque celui-ci n'a pas le statut d'acteur de la situation), peuvent l'être tout aussi bien par rapport à *n'importe quelle personne* engagée dans la situation observée : l'engagement praxéologique d'une personne dans une situation *peut* dépendre de la présence de telle autre personne, les remarques formulées ci-dessus en termes d'*inertie* (des éléments praxéologiques) ou d'*intention* (des personnes) s'appliquant alors tout autant à ce type de cas.

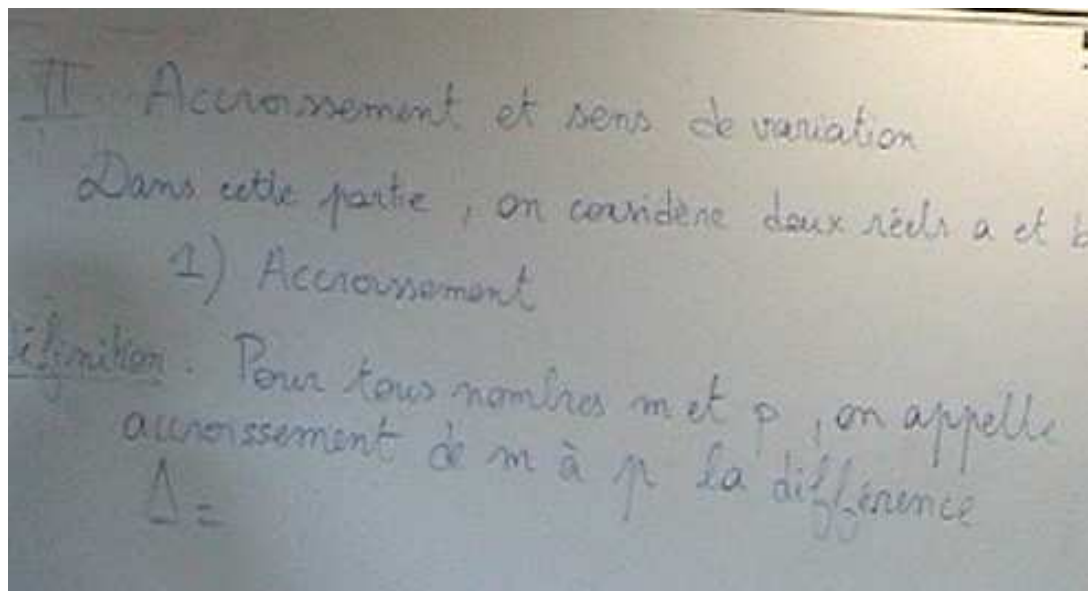
b) Je donne ici une illustration simple de ces considérations. Lors de la visite en classe d'un professeur stagiaire, il est vraisemblable que le professeur observé s'efforcera de contrôler divers aspects de son engagement praxéologique dans la situation, notamment ceux sur lesquels il s'attend à être jugé, ou dont il pense qu'ils pourraient déplaire au visiteur ou simplement gêner l'observation.

➔ Lorsqu'ils sont identifiables par l'observateur, de tels changements praxéologiques *occasionnels* doivent d'abord être regardés comme manifestant chez l'observé une *intention de conformité ou de connivence praxéologiques* avec l'observateur. En même temps, pourtant, la réalisation de cette intention révèle chez l'observé un engagement praxéologique qui, en tant que tel, n'est l'objet d'*aucune intention* de sa part, relativement à l'observateur. Ainsi en va-t-il par exemple lorsque, soucieux d'hypercorrection lexicale ou notationnelle, l'observé met en œuvre des formulations ou des écritures qu'il tient objectivement pour « soignées », mais dont le contenu spécifique, quoique mis au service de cette intention, *va pour lui de soi* et ne dépend donc pas, en tant que contenu tenu pour « correct », de la présence ou de l'absence de l'observateur. Ainsi pourra-t-on découvrir que ce professeur stagiaire pense qu'il faut – ou plutôt qu'il faudrait, si un tel niveau de correction était exigé de lui – lire « proposition *une* » ce qu'il écrit au tableau « proposition 1 » et qu'il lirait peut-être, dans une situation « ordinaire » de classe, « proposition *un* », ou même qu'il désignerait familièrement comme « la *un* ».

➔ Dans la situation idéalisée évoquée ci-dessus, j'ai supposé *en partie au moins repérable* par l'observateur ce qui *change* lorsque le visiteur est là. Mais ce qui s'observe aussi, ce sont des éléments praxéologiques à *forte inertie*, parce que quasi *naturalisés*, et donc peu susceptibles d'être remaniés de façon occasionnelle, qui sont ici *mobilisés* pour *servir* une certaine intention certes, mais qui le sont *sans modification*, n'étant pas eux-mêmes l'objet d'une intention particulière puisqu'ils vont de soi pour l'observé. D'autres éléments à forte inertie praxéologique donneront lieu encore à une observation semblablement assurée, en particulier parce que l'observé n'a, à leur endroit, aucune intention de modification engendrée par le fait de l'observation.

➔ Voici un exemple récemment observé lors d'une visite dans une classe de seconde. Si la professeure stagiaire visitée s'attend sans doute à être jugée – entre autres choses – à la façon dont elle « gère le tableau », elle ne s'attend certainement pas à être observée (sinon jugée) dans ses habitudes

« typographiques » – qui, vraisemblablement, *n'existent guère pour elle comme telles*, alors même qu'elle s'efforce tout aussi sûrement de ne pas laisser traîner de « fautes d'orthographe » au tableau. Voici alors ce qu'on peut observer, et qu'on peut supposer non adultéré par quelque intention que ce soit – sinon peut-être par la volonté d'écrire de façon « soignée ».



On aura noté, donc, que chaque virgule est ici précédée d'une espace (très) « forte ». Il ne s'agit pas là d'un geste aléatoire : l'examen de la vidéo montre qu'il est *systématique*. Voilà donc, peut-on déjà penser, une idiosyncrasie typographique de l'observée. Mais s'agirait-il là d'une pratique *plus large*, qui toucherait, disons, l'ensemble des professeurs, ou du moins une partie significative d'entre eux ? Si la réponse était positive, elle montrerait alors qu'il y a là une manière « commune », voire « orthodoxe » d'écrire dans la profession, et donc une manière à *forte inertie*, qui ne vient donc pas au bout de la craie de cette professeure à l'occasion de la visite !

→ Lors d'une visite dans un autre lycée, j'ai pu observer le texte figurant sur une feuille A4 placardée sur un mur de la salle des professeurs. J'en reproduis ci-après les deux premiers paragraphes : le collectif des professeurs de lettres de l'établissement s'y adresse apparemment à l'ensemble des professeurs. Bien entendu, on peut s'attendre ici à rencontrer une syntaxe et une orthographe voulues correctes. Par contraste, ainsi qu'on le voit (ci-après), ce texte révèle (ou plutôt confirme) les faiblesses de la culture *commune* des professeurs de lettres en matière de typographie : espace (systématique) avant une virgule ou avant un point, espace (aléatoire) avant une parenthèse fermante ou après une parenthèse ouvrante, écriture fautive de *etc.*, affectée de variations anarchiques, etc. Il se confirme au passage que l'usage de séparer la virgule du mot qui la précède semble *vivace*. Et sans doute peut-on apercevoir dans une certaine volonté pédagogique de clarté l'origine de cet usage qui serait propre d'abord à l'écriture manuscrite pratiquée en classe.

Les professeurs de lettres ont annoncé à leurs collègues dès le 04/09/07 les dates retenues pour les devoirs communs de français en seconde et en première ; plusieurs modifications ont dû avoir lieu pour tenir compte de contraintes diverses (nombre de surveillants disponibles , voyages scolaires , etc ...) et chaque fois qu'un aménagement du calendrier a dû être fait , un affichage d'information a été mis en place .

Peu avant les vacances d'hiver , ces mêmes collègues ont dû une nouvelle fois reprogrammer les derniers devoirs de l'année mais depuis de nouvelles contraintes sont apparues (évaluation des AID des 1ST2S , inspections de collègues en maths , etc...) : résultat le seul devoir ils sont à peu près certains aujourd'hui est le **bac blanc des premières prévu le mercredi 19 mars** .

Il reste à trouver des dates pour caser le dernier devoir des secondes (placé dans la mouture précédente le 3 mars , il devrait être reporté) et le dernier devoir des premières (prévu dans la mouture précédente le 2 mai il devrait lui aussi être reporté) .

➔ Je voudrais faire ici une remarque plus générale sur la clinique en général – en didactique comme ailleurs. Ce qu'on nomme « l'expérience clinique » est, en l'espèce, essentiel : mon expérience constante depuis des décennies, c'est que trop de didacticiens sont étrangers à la clinique des univers institutionnels et personnels qu'ils sont censés étudier et donc, à force, connaître un tant soit peu ; en sorte qu'à la longue ils tendent à ne rien savoir des faits les plus évidents. Lorsque, dans la séance 4, j'ai introduit comme un exemple en apparence tombé du ciel l'écriture *n* fois fautive

Il y a , non seulement des savoirs , mais aussi des savoirs - faire

transcrite au clavier par

Il y a , non seulement des savoirs , mais aussi des savoirs - faire

je tirai profit de mon expérience clinique de l'activité scripturale des professeurs, là où le béotien, qui ne veut rien savoir, s'exclamerait peut-être : « Ah bon, les profs écrivent comme ça ? Ça m'étonne, j'ai jamais vu ça ! » Une vie d'ignorance !

c) Dans la perspective ouverte par l'exemple précédent, où l'on passe d'une manière d'écrire que l'on pouvait croire *personnelle* à la conjecture de sa probable nature *institutionnelle*, je voudrais m'arrêter un instant sur la question de l'étude de l'équipement praxéologique *d'une institution I*. Plusieurs remarques liminaires sont ici encore de mise.

➔ Il se peut qu'un type de tâches n'existe pas pour *I* comme *objet institutionnel*, en sorte qu'il n'existera pas non plus de technique(s)

institutionnelle(s) associée(s). En ce cas, face à des tâches du type considéré, les personnes sujets de *I* devront alors user de leurs assujettissements allogènes à des institutions *I'*, *I''*, etc., pour lesquelles le type de tâches serait objet institutionnel. Par delà les divergences praxéologiques personnelles constatées, l'exploration sur ce point devrait alors révéler une situation de *vide institutionnel* au moins provisoire. (Notons que la même description vaudrait, à un autre niveau, pour une *personne x* : au lieu du *bricolage* praxéologique *institutionnel* observable dans le cas précédent, on observera alors un *bricolage personnel*.)

➔ On a évoqué précédemment des « éléments praxéologiques », sans préciser davantage la chose ; l'exemple donné (la manière de placer les signes de ponctuation) relève d'abord de la *praxis*, même si l'on a pu invoquer un élément de *logos* pour l'expliquer. Mais l'ensemble des préceptes énoncés jusqu'ici vaut pour *tout type* d'éléments praxéologiques, qu'il s'agisse par exemple d'éléments de *technique* (donc de l'ordre du faire) ou de *technologie* (donc de l'ordre du « discours »). Ainsi, pour observer un discours technologique (ou technico-technologique), on n'ira pas nécessairement le susciter soi-même en *interrogeant* la personne observée : on tentera d'observer, en les suscitant éventuellement, des situations dans lesquelles des fragments d'un tel discours ont chance d'apparaître, à l'adresse d'autres *acteurs* de la situation observée.

➔ L'exemple que nous examinerons maintenant est celui d'une exploration clinique (ou « paraclinique ») à l'aide d'un *différenciateur sémantique*. Comment peut-on situer ce geste simple par rapport aux développements précédents ? On a demandé à 22 étudiants d'une licence professionnelle de sciences de l'éducation, dont le port d'attache est Lambesc (près d'Aix), d'attribuer un score sur une échelle laid/beau de 0 à 6 à un certain nombre de mots, par exemple « Internet ». En faisant cela, on explore les zones les plus évanescences – mais non les moins « effectives » – de la théorie personnelle de chacun des « juges » à propos des réalités auxquelles renvoie le mot soumis à leur appréciation, et peut-être, du même coup, de la théorie d'une certaine institution, celle qui réunit là ces 22 « juges ». Comme dans la clinique médicale la plus commune (« Toussez, monsieur ! » « Dites “Ah...” », continuez ! »), on fait « faire quelque chose » aux personnes concernées ; comme dans ce cas encore, la personne – le « patient » ici, le « sujet » là – ignore ce que l'observateur – médecin ici, chercheur là – tente de saisir à travers ce qu'il lui demande de faire. Bien entendu, de même que le patient peut refuser de tousser, de tirer la langue, etc., le sujet peut refuser d'assigner des scores, sur les différentes échelles proposées, au mot qu'on lui communique.

➔ Voici les scores, totalisés sur l'ensemble des juges, obtenus par « Internet » sur les 15 échelles qui étaient proposées.

ℓ/L	S/P	F/C	T/B	u/U	I/A	M/B	f/F	A/C	S/O	L/B	P/A	I/S	A/N	F/D
54	50	54	80	108	57	83	93	74	80	68	94	45	33	45

On voit que le score du mot « Internet » sur l'échelle laid/beau, 68, se situe presque exactement à mi-chemin entre le minimum théorique (0) et le maximum théorique ($22 \times 6 = 132$). À juger selon ce critère numérique, « Internet » ne serait ni laid ni beau pour notre groupe de juges – même si l'un d'eux lui attribue le score maximal, 6, le trouvant donc « très beau » ! Les échelles sur lesquelles les scores de « Internet » sont les plus extrêmes sont mises en évidence ci-après.

ℓ/L	S/P	F/C	T/B	u/U	I/A	M/B	f/F	A/C	S/O	L/B	P/A	I/S	A/N	F/D
54	50	54	80	108	57	83	93	74	80	68	94	45	33	45

Le score de 108 sur l'échelle u/U, c'est-à-dire sur l'échelle peu utile/utile, correspond à 81,8 % du maximum. (Sur l'ensemble des mots et des échelles, le plus haut score total est 123, soit 93,2 % du maximum.) À l'évidence, « Internet » est *fort utile* : le plus bas score attribué à ce mot est 3, la « moyenne », qui est obtenue deux fois. Le score de 33 obtenu sur l'échelle A/N, soit l'échelle artificiel/naturel, représente 25 % du maximum : en lisant l'échelle dans l'autre sens (naturel/artificiel), on arrive à un score total de 99, soit donc 75 % du maximum. C'est le plus bas score total obtenu sur l'ensemble des mots : « Internet » est ainsi *fortement artificiel*. Les échelles sur lesquelles « Internet » obtient un score total supérieur à 65 % du maximum ou inférieur à 35 % du maximum, c'est-à-dire supérieur à 85,8 points ou inférieur à 46,2 sont mises en évidence ci-après.

ℓ/L	S/P	F/C	T/B	u/U	I/A	M/B	f/F	A/C	S/O	L/B	P/A	I/S	A/N	F/D
54	50	54	80	108	57	83	93	74	80	68	94	45	33	45

Il s'agit, de gauche à droite, des échelles faible/fort, passif/actif, instable/stable et facile/difficile : « Internet » est donc jugé dans l'ensemble plutôt *fort, actif, instable* et *facile*. En fait, les deux scores de 45 dissimulent des situations un peu différentes. Sur le fait que « Internet » serait plutôt *instable*, il y a un accord *assez large* du groupe des juges, comme le montre le tableau suivant.

Score	0	1	2	3	4	5	6
Effectif	3	4	5	9	1	0	0

En revanche, les jugements sur le fait que « Internet » serait ou non *difficile* sont davantage contrastés, ainsi qu'en témoigne le tableau suivant.

Score	0	1	2	3	4	5	6
Effectif	3	6	5	4	3	1	0

Tout se passe en effet comme si cinq scores moyens (3) avaient été redistribués d'une part vers le bas de l'échelle (deux scores 1 de plus), d'autre part vers le haut de l'échelle (deux scores 4 et un score 5 de plus).

➔ On peut comparer les scores obtenus par « Internet » et « Web ». Sur le tableau suivant, la deuxième ligne rappelle les scores totaux de « Internet » ; la troisième présente ceux relatifs à « Web ».

ℓ/L	S/P	F/C	T/B	u/U	I/A	M/B	f/F	A/C	S/O	L/B	P/A	I/S	A/N	F/D
54	50	54	80	108	57	83	93	74	80	68	94	45	33	45
74	55	52	76	105	61	79	86	78	88	71	83	55	34	55

On voit que les différences ne semblent pas profondes – « Web » apparaît par exemple un peu plus *lourd* (ℓ/L, léger/lourd) que « Internet » mais un peu moins *actif* (P/A, passif/actif), etc. On peut se demander si « Internet » et « Web » ont la particularité d'avoir seuls le « profil » que nous avons sommairement décrit jusqu'ici. Voici donc le tableau précédent complété par des lignes relatives aux autres mots proposés aux 22 juges, à savoir, après « Internet » et « Web », et par ordre croissant des numéros des lignes du tableau : « Statistique », « Didactique », « Université », « Éducation », « Lambesc ».

ℓ/L	S/P	F/C	T/B	u/U	I/A	M/B	f/F	A/C	S/O	L/B	P/A	I/S	A/N	F/D
54	50	54	80	108	57	83	93	74	80	68	94	45	33	45
74	55	52	76	105	61	79	86	78	88	71	83	55	34	55
85	73	38	60	95	47	79	80	76	107	56	85	66	45	80
85	100	80	84	106	73	100	102	88	91	78	88	79	71	83
95	108	75	112	112	67	100	107	79	96	88	105	74	73	98
94	115	95	91	123	69	109	105	91	85	102	114	78	81	94
62	109	84	98	117	77	108	104	85	88	101	108	84	79	84

On observe ici, dans plusieurs colonnes, un changement qui se produit lorsqu'on passe du bloc des trois premières lignes à celui des quatre dernières. Ainsi, sur l'échelle S/P, superficiel/profond, on découvre que « Internet » et « Web » sont plutôt *superficiels*, que c'est un peu moins vrai pour « Statistique », et que « Didactique », « Université », « Éducation » et « Lambesc » sont regardés davantage comme *profonds*. Un schéma analogue s'observe à propos de l'échelle F/C, froid/chaud : « Statistique » surtout, suivi par « Internet » et « Web », sont froids, tandis que « Didactique », « Université », « Éducation » et « Lambesc » sont plutôt chauds. Sur l'échelle T/B, terne/brillant, une étoile se détache : « Université » ; lui fait pendant un astre, lui, relativement terne, « Statistique ». Sur l'échelle u/U, tous les mots sont rangés franchement du côté de l'*utile*, y compris « Statistique », avec un pic, « Éducation », dont le score représente 93,2 % du maximum. La moindre appréciation de « Statistique » s'exprime encore sur l'échelle suivante, I/A, irritant/apaisant, où il est le moins bien noté : ici, en vérité, aucun mot n'a de score impressionnant puisque le maximum est atteint par « Lambesc », dont le score ne représente pourtant que 58,3 % du score maximal. On retrouve le contraste entre les trois premières lignes et les quatre suivantes avec l'échelle M/B, mauvais/bon : « Internet », « Web » et « Statistique » ne sont pas mauvais, mais « Didactique », « Université », « Éducation » et « Lambesc » sont meilleurs. On laissera le lecteur poursuivre ce rapide examen, sachant que les colonnes suivantes du tableau correspondent successivement aux échelles faible/fort (f/F), abstrait/concret (A/C), spontané/organisé (S/O), laid/beau (L/B), passif/actif (P/A), instable/stable (I/S), artificiel/naturel (A/N).

➔ Je n'irai pas plus loin quant à la prise d'information réalisée ainsi qu'on l'a dit plus haut. Mais l'information recueillie et analysée (sommairement !) nous éclaire, de façon bien sûr conjecturale, sur le rapport des « sujets » examinés à travers un différenciateur sémantique à un certain nombre d'objets. Il est remarquable, ainsi, qu'il y ait un bon accord autour de *l'utilité* qu'auraient les réalités évoquées par les mots retenus pour l'examen réalisé. On peut même imaginer qu'il s'agisse là d'une attitude systématique, engendrée par l'engagement formatif de ce type de « public » : pour cette raison, on peut s'attendre (mais nous ne l'avons pas vérifié) à ce qu'un autre mot, désignant une autre réalité de l'univers de ces « sujets », par exemple « Mémoire », obtienne un score élevé sur l'échelle peu utile/utile. Mais on ne manquera pas d'observer aussi que l'effort de formation ainsi apparemment consenti va de pair avec ce qu'on peut décrire comme une imparfaite sérénité... On n'insistera pas, enfin, sur ce que révèle l'analyse ébauchée à propos de la « théorie » des sujets observés touchant la statistique – discipline froide, plutôt terne, irritante aussi, pas particulièrement mauvaise sans doute mais très organisée, dénuée de beauté, et pour finir artificielle. Un esprit moins pondéré pourrait conclure qu'il serait bon qu'on s'occupe, avec tact et doigté, du problème ainsi exposé.

3. Questions de style

a) À titre de rappel, je voudrais d'abord illustrer quelques notations de la séance 4 à propos des questions de *style*. Me référant à la culture française *non spécialisée* en la matière, j'avais indiqué qu'on y passait sans façon d'une supposée « loi d'airain » orthographique à une anomie plus ou moins affirmée touchant à la ponctuation, aux majuscules, à la typographie. Dans un ouvrage que je recommande, *La ponctuation ou l'art d'accommoder les textes*, dû à Olivier Houdart et Sylvie Prioul (Seuil, Paris, 2006), les auteurs écrivent typiquement (p. 26) : « ... la norme en matière de ponctuation est moins contraignante que pour l'orthographe : un texte donné n'a (en principe) qu'une seule orthographe, mais il peut être ponctué de multiples manières selon le tempérament et le style de chacun. » Ce point de vue, qui admet complaisamment la variation *personnelle*, fait contraste d'une part avec le postulat d'*unicité* quasi « transcendante » d'un Pierre Larousse dans son *Grand dictionnaire universel du XIX^e siècle*, pour qui les règles de la ponctuation sont « fixes, basées sur la syntaxe » et « n'admettent ni la fantaisie ni les caprices », d'autre part avec l'idée même de style *institutionnel* (style de l'APA, etc.), qui oblige ceux qui s'assujettissent à l'institution considérée, mais qui nous affranchit de la règle unique *supposée* – qui, en plusieurs cas, s'avère *introuvable*.

➔ Pour la comparaison des rapports au « style » orthographique, je reproduis ci-après un article signé de Jean-Charles Condo, publié sur le site *Branchez-vous !* le 16 août 2004 (<http://www.branchez-vous.com/actu/04-08/08-274704.html>).

Internet perd sa majuscule chez Wired

Le site Web américain bien connu des technophiles a décidé de ne plus orthographier avec des majuscules les termes « Internet », « Web » et « Net ».

« À partir de cette phrase, *Wired News* ne mettra plus de “I” majuscule dans internet », annonce le site Web du magazine avant d’ajouter que les mots « Web » et « Net » perdent du même coup leur première lettre majuscule. Une annonce qui étonnera tout de même un peu car les anglophones ont généralement la majuscule très facile.

La raison ? *Wired* estime en fait qu’il n’y a jamais eu de raison valable pour que ces termes nécessitent une majuscule, sauf dans le cas du nom d’une entité officielle telle que le *World Wide Web*. Le magazine américain ajoute dans son article qu’Internet n’est rien d’autre qu’un média qui permet d’envoyer et de recevoir de l’information, et qu’à ce titre il ne mérite donc pas plus de majuscule que la radio et la télévision.

Ailleurs sur le Web, la majuscule d’Internet se dresse encore fièrement, notamment sur les sites Web du dictionnaire d’*Encarta* et de l’*Encyclopædia Britannica*. *Wired* amorcera-t-il une nouvelle tendance dans le monde anglophone ?

En français, l’application d’une majuscule à Internet est parfois débattue. Selon l’Office québécois de la langue française, le terme Internet, formé à partir des mots anglais « INTERconnected » et « NETworks », est considéré comme un nom propre et doit donc s’écrire avec une majuscule. L’OQLF a également retenu comme synonymes les expressions « réseau Internet » et « Net ».

Pour ce qui est du Web, l’OQLF apporte les précisions suivantes :

À l’origine, la graphie avec majuscule s’est imposée pour souligner le caractère unique du Web. On le considère alors comme un nom propre (tout comme le mot Internet d’ailleurs). Plus récemment, la graphie avec minuscule a fait son apparition dans la presse francophone et dans Internet. Signe que cet emprunt à l’anglais commence à s’intégrer à la langue française (on le traite comme un nom commun). Actuellement, les deux graphies coexistent dans l’usage, que le terme soit utilisé seul ou en apposition (ex. : page Web). Mais c’est une simple question de point de vue : si l’on considère le Web comme une entité unique (nom propre), on choisit la majuscule, et si l’on considère le web comme une ressource Internet parmi d’autres, on choisit la minuscule (nom commun). Pour l’instant, l’OQLF privilégie toujours la graphie avec majuscule.

➔ Pour compléter le document précédent, voici un extrait de l’article “Internet” de l’encyclopédie Wikipedia (en anglais), qui permet de replacer la décision de *Wired News* dans un cadre plus large.

The name Internet

For more details on this topic, see Internet capitalization conventions.

Internet is traditionally written with a capital first letter, as it is a proper noun. The Internet Society, the Internet Engineering Task Force, the Internet Corporation for Assigned Names and Numbers, the World Wide Web Consortium, and several other Internet-related organizations use this convention in their publications.

Many newspapers, newswires, periodicals, and technical journals capitalize the term (*Internet*). Examples include *The New York Times*, the *Associated Press*, *Time*, *The Times of India*, *Hindustan Times*, and *Communications of the ACM*.

Others assert that the first letter should be in lower case (*internet*), and that the specific article the is sufficient to distinguish “the internet” from other internets. A significant number of publications use this form, including *The Economist*, the *Canadian Broadcasting Corporation*, the *Financial Times*, *The Guardian*, *The Times*,

and *The Sydney Morning Herald*. As of 2005, many publications using *internet* appear to be located outside of North America—although one U.S. news source, *Wired News*, has adopted the lower-case spelling.

Historically, *Internet* and *internet* have had different meanings, with *internet* meaning “an interconnected set of distinct networks”, and *Internet* referring to the worldwide, publicly available IP internet. Under this distinction, *the Internet* is the familiar network through which websites exist; however, *an internet* can exist between any two remote locations.[\[http://what-is-what.com/what_is/internet.html\]](http://what-is-what.com/what_is/internet.html) Any group of distinct networks connected together is *an internet*; each of these networks may or may not be part of the Internet. The distinction was evident in many RFCs, books, and articles from the 1980s and early 1990s (some of which, such as RFC 1918, refer to “internets” in the plural), but has recently fallen into disuse.^[citation needed] Instead, the term *intranet* is generally used for private networks, whether they are connected to the Internet or not (see also extranet).

Some people use the lower-case term as a medium (like radio or newspaper, e.g. *I’ve found it on the internet*), and first letter capitalized as the global network.

➔ On voit que la formulation « française » « Ça s’écrit comment “Web” ? Avec une majuscule ou une minuscule ? » doit être remplacée par : « *Nous l’écrivons comment “Web” ?...* », sous-entendu : *selon tel style expressément déclaré*. C’est ce changement de point de vue qu’il s’agira d’éprouver, concrètement, avec l’idée d’un « style TAD » à définir. Je rappelle seulement la définition que j’ai introduite lors de la séance 4 : « un “style” doit avoir pour ambition de proposer, de façon cohérente, raisonnée, uniforme, une “solution” pour chaque “problème” qui peut se rencontrer dans la production d’un texte imprimé ou électronique. »

b) Je vous propose aujourd’hui de commencer par le petit bout de la lorgnette, celui des *références bibliographiques*, en ne considérant sauf exception que le cas le plus simple, celui d’un ouvrage publié n’ayant qu’un auteur, lorsque les informations pertinentes sont disponibles.

➔ Voici d’abord le format de l’APA :

Author, A. A. (1994). *Title of Work*. Location: Publisher.

Cela donnerait en français, en respectant les règles usuelles de la typographie :

Auteur, A. (1994). *Titre de l’ouvrage*. Lieu : Éditeur.

Tout cela suppose une sérieuse attention « typographique » : par exemple le point à la fin du titre en italique est lui-même en romain.

➔ Voici maintenant le format de la MLA, la *Modern Language Association*. Je me réfère ici à la sixième édition de l’ouvrage signé de Joseph Gibaldi, *MLA Handbook for Writers of Research Papers* (The Modern Language Association of America, New York, 2003).

Author’s name. Title of the book. Publication information.

Les indications sont ici plus globales (elles sont détaillées dans les pages qui suivent). Bien entendu, le soulignement (dans le manuscrit supposé) équivaut à l'italique : on observera donc que le point à la fin du titre n'est pas souligné. L'exemple proposé à titre d'illustration immédiate est le suivant :

Fukuyama, Francis. Our Posthuman Future: Consequences of the Biotechnology Revolution. New York: Farrar, 2002.

Parmi les différences avec le style APA, on peut noter que le prénom de l'auteur est donné au long et que la date de publication est repoussée à la fin.

➔ Voici encore le format proposé par la deuxième édition du *Merriam-Webster's Concise Handbook for Writers* (Merriam-Webster, Inc., Springfield, Mass., 1998). Ce style tend à différencier les références relevant du champ des *humanities* de celles relevant des *sciences*. Voici un exemple du premier cas :

Maxmen, Jerrold S. *The Post-Physician Era: Medecine int the Twenty-first Century*. New York: Wiley, 1976.

Notons que le point abrégatif du second prénom tient lieu de point final du nom d'auteur. Dans ce cas, le style proposé ressemble fortement à celui de la MLA, avec date de publication après l'indication de l'éditeur. Pour les *sciences*, il en va autrement, comme le montre cet exemple :

Chandrasekhar, S. 1983. *The mathematical theory of black holes*. New York: Oxford.

Ici, la date figure avant le titre, sans parenthèses. (On aura noté aussi que, alors que le premier exemple comporte les majuscules habituelles dans un titre en anglais, il n'en va plus de même dans ce second exemple.)

➔ Voici maintenant « la présentation adoptée à l'Imprimerie nationale pour les travaux courants » (*Lexique des règles typographiques en usage à l'Imprimerie nationale*, 2002) :

DOMENACH (Jean-Luc) et RICHER (Philippe), *La Chine, 1949-1985*, Paris, Imprimerie nationale, coll. « Notre Siècle », 1987, 504 p.

Les noms des auteurs sont ici en petites capitales, les prénoms étant explicités entre parenthèses. On voit que les points et autres deux-points « anglo-saxons » sont remplacés ici uniformément par des virgules. On aura noté encore l'indication du nombre de pages de l'ouvrage.

➔ Voici enfin le « style » de l'AFNOR. (Je me réfère pour cela au document figurant à l'adresse suivante : http://www.revue-texto.net/Reperes/Themes/Kyheng_References.html.)

LE GUERN, Michel. *Sémantique de la métaphore et de la métonymie*. Paris : Larousse, 1973, 126 p. Collection Langue et Langage.

Ici, le nom de l'auteur est en grandes capitales.

c) Voilà donc pour ce petit florilège. Que proposera le style TAD sur le sujet examiné ? Il nous appartiendra d'en décider.

MATHÉMATIQUES DANS LA CITÉ

1. PageRank

a) Lors de la séance 4, on a exprimé le système d'équations censé donner le rang r_i de chaque page P_i du Web sous la forme matricielle suivante :

$$R^T = R^T \mathbf{H}$$

où R^T est le vecteur-ligne $[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n]$ et où $\mathbf{H}$ est la matrice dont l'élément H_{ij} , situé à la i -ième ligne et à la j -ième colonne, vaut 0 si la page P_i n'envoie pas de lien vers la page P_j , et $\frac{1}{|P_i|}$ sinon, où $|P_i|$ désigne le nombre de liens qui *partent* de P_i :

$$H_{ij} = \begin{cases} \frac{1}{|P_i|} & \text{s'il y a un lien de } P_i \text{ vers } P_j \\ 0 & \text{sinon} \end{cases}.$$

b) On doit donc étudier l'équation (en $\mathbb{R}$) : $R^T = R^T \mathbf{H}$. Pour que l'idée à l'origine du PageRank trouve un sens, il faudrait *a priori* que l'équation précédente n'ait pour l'essentiel qu'une solution.

→ Si, par exemple, dans le cas d'un Web à 3 pages, on avait deux solutions distinctes (ou plus !), l'une étant telle par exemple que l'on ait $r_1 < r_3 < r_2$ et l'autre telle que l'on ait $r_3 < r_1 < r_2$, comment choisir entre les deux classements : faut-il, dans une page de résultats, placer P_1 avant P_3 ou l'inverse ?

→ L'unicité de la solution ne saurait pourtant être absolue : car si

$$R^T = [r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n]$$

est une solution, il en est forcément de même de la matrice-ligne

$$[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n]$$

que l'on notera λR^T et qui s'obtient en multipliant chacun des éléments r_i par le nombre λ , lequel peut être quelconque. Seront donc également des solutions, par exemple, les matrices-lignes

$$[2r_1 \ 2r_2 \ \dots \ 2r_i \ \dots \ 2r_n]$$

ou

$$[0,6r_1 \ 0,6r_2 \ \dots \ 0,6r_i \ \dots \ 0,6r_n]$$

ou

$$\left[\frac{1}{7} r_1 \ \frac{1}{7} r_2 \ \dots \ \frac{1}{7} r_i \ \dots \ \frac{1}{7} r_n \right]$$

ou encore

$$[\sqrt{2}r_1 \ \sqrt{2}r_2 \ \dots \ \sqrt{2}r_i \ \dots \ \sqrt{2}r_n]$$

etc. Supposons en effet que l'on ait

$$[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n] \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1j} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2j} & \dots & H_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{i1} & H_{i2} & \dots & H_{ij} & \dots & H_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{n1} & H_{n2} & \dots & H_{nj} & \dots & H_{nn} \end{bmatrix} = [r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n].$$

Remplaçons dans le membre de gauche de cette égalité la matrice

$$[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n]$$

par $[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n]$; l'élément $(1, j)$, c'est-à-dire l'élément d'indice ligne 1 et d'indice colonne j , du produit matriciel

$$[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n] \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1j} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2j} & \dots & H_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{i1} & H_{i2} & \dots & H_{ij} & \dots & H_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{n1} & H_{n2} & \dots & H_{nj} & \dots & H_{nn} \end{bmatrix}$$

est donné par le produit matriciel

$$[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n] \begin{bmatrix} H_{1j} \\ H_{2j} \\ \dots \\ H_{ij} \\ \dots \\ H_{nj} \end{bmatrix}$$

et vaut donc : $\lambda r_1 \times H_{1j} + \lambda r_2 \times H_{2j} + \dots + \lambda r_i \times H_{ij} + \dots + \lambda r_n \times H_{nj}$. En mettant λ en facteur dans cette expression, on obtient

$$\lambda(r_1 \times H_{1j} + r_2 \times H_{2j} + \dots + r_i \times H_{ij} + \dots + r_n \times H_{nj}).$$

Le cofacteur de λ dans ce produit s'obtient par le produit matriciel

$$[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n] \begin{bmatrix} H_{1j} \\ H_{2j} \\ \dots \\ H_{ij} \\ \dots \\ H_{nj} \end{bmatrix}$$

et vaut donc r_j . Ainsi l'élément $(1, j)$ du produit

$$[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n] \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1j} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2j} & \dots & H_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{i1} & H_{i2} & \dots & H_{ij} & \dots & H_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{n1} & H_{n2} & \dots & H_{nj} & \dots & H_{nn} \end{bmatrix}$$

n'est-il autre que λr_j , et cela quel que soit $j = 1, 2, \dots, n$, en sorte que l'on a bien

$$[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n] \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1j} & \dots & H_{1n} \\ H_{21} & H_{22} & \dots & H_{2j} & \dots & H_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{i1} & H_{i2} & \dots & H_{ij} & \dots & H_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ H_{n1} & H_{n2} & \dots & H_{nj} & \dots & H_{nn} \end{bmatrix} = [\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n].$$

On a ainsi démontré que, si $[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n]$ est une solution de l'équation

$$[r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n] \mathbf{H} = [r_1 \ r_2 \ \dots \ r_i \ \dots \ r_n]$$

la matrice $[\lambda r_1 \ \lambda r_2 \ \dots \ \lambda r_i \ \dots \ \lambda r_n]$ est aussi une solution, et cela quel que soit le nombre λ . Ce fait ne pose pas de problème particulier : le PageRank est défini à un facteur multiplicatif près (ou, dans un langage plus élémentaire : « à une proportionnalité près »), et on pourra par exemple fixer le maximum à 10, ou à 20, etc.

c) Comme on l'a dit lors de la séance 2 de ce séminaire, pour résoudre l'équation $R^T = R^T \mathbf{H}$, Brin et Page tentent d'utiliser une technique mathématique classique, qu'on peut maintenant décrire ainsi : ils définissent une suite de vecteurs

$$R^{(0)T}, R^{(1)T}, R^{(2)T}, \dots, R^{(k)T}, R^{(k+1)T} \dots$$

en prenant d'abord

$$R^{(0)T} = \left[\frac{1}{n} \ \frac{1}{n} \ \dots \ \frac{1}{n} \ \dots \ \frac{1}{n} \right]$$

où, rappelons-le, $n = |W|$ est le nombre de pages du Web ; si l'on écrit de façon générale

$$\mathbf{R}^{(k)T} = [\mathbf{r}_1^{(k)} \ \mathbf{r}_2^{(k)} \ \dots \ \mathbf{r}_i^{(k)} \ \dots \ \mathbf{r}_n^{(k)}]$$

on a donc : $\mathbf{r}_1^{(0)} = \mathbf{r}_2^{(0)} = \dots = \mathbf{r}_i^{(0)} = \dots = \mathbf{r}_n^{(0)} = \frac{1}{n}$. Le passage de $\mathbf{R}^{(0)T}$ à $\mathbf{R}^{(1)T}$ s'effectue en posant par définition

$$\mathbf{R}^{(1)T} = \mathbf{R}^{(0)T} \mathbf{H}$$

c'est-à-dire en effectuant le produit matriciel de $\mathbf{R}^{(0)T}$ par la matrice $\mathbf{H}$; on aura de même :

$$\begin{aligned} \mathbf{R}^{(2)T} &= \mathbf{R}^{(1)T} \mathbf{H} \\ \mathbf{R}^{(3)T} &= \mathbf{R}^{(2)T} \mathbf{H} \\ &\dots\dots\dots \\ \mathbf{R}^{(k)T} &= \mathbf{R}^{(k-1)T} \mathbf{H} \\ \mathbf{R}^{(k+1)T} &= \mathbf{R}^{(k)T} \mathbf{H} \\ &\dots\dots\dots \end{aligned}$$

et, comme on disait autrefois, « ainsi de suite à l'infini ».

d) L'espoir derrière ce « geste » est que la suite des vecteurs $\mathbf{R}^{(k)T}$ va – assez vite – *se stabiliser* (ou « converger », pour employer le terme mathématique usuel, métaphore qui vient de l'optique), en sorte que, à partir d'un certain rang k , les vecteurs ligne

$$\mathbf{R}^{(k)T}, \mathbf{R}^{(k+1)T}, \mathbf{R}^{(k+2)T} \dots$$

ne différeront plus entre eux que de façon « imperceptible », en ce sens que leurs éléments de même indice colonne i

$$\mathbf{r}_i^{(k)}, \mathbf{r}_i^{(k+1)}, \mathbf{r}_i^{(k+2)}, \dots$$

auront par exemple les mêmes 20 premiers chiffres. Or on va voir que les choses, ici, ne vont pas toutes seules. Les auteurs que je suis écrivent à ce propos ceci, où ce qu'ils désignent comme “equation (4.2.1)” est l'analogue de notre relation $\mathbf{R}^{(k+1)T} = \mathbf{R}^{(k)T} \mathbf{H}$, à ceci près qu'ils notent $\boldsymbol{\pi}$ ce que nous notons $\mathbf{R}$ (Langville & Meyer, 2006, p. 34-35) :

Equation (4.2.1) probably caused readers, especially our mathematical readers, to ask several questions. For example,

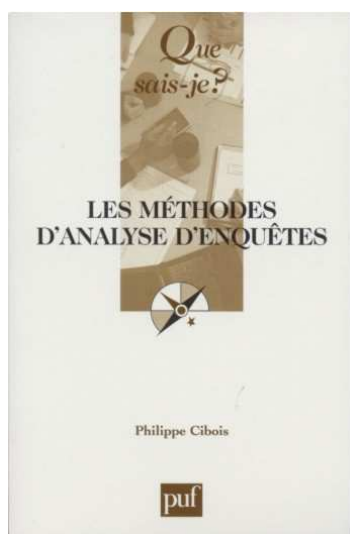
- Will this iterative process continue indefinitely or will it converge?
- Under what circumstances or properties of $\mathbf{H}$ is it guaranteed to converge?
- Will it converge to something that makes sense in the context of the PageRank problem?
- Will it converge to just one vector or multiple vectors?
- Does the convergence depend on the starting vector $\boldsymbol{\pi}^{(0)T}$?

- If it will converge eventually, how long is “eventually”? That is, how many iterations can we expect until convergence?

Ce sont ces questions que nous examinerons dans les séances à venir.

2. Analyser des données ?

a) Le besoin de mathématiques parfois simples, parfois plus complexes saisit – souvent sans qu’il en ait conscience – plus d’un quidam qui croit œuvrer pourtant dans un contexte *mathematics-proof*, l’épreuve des mathématiques. Ainsi en va-t-il, pour les étudiants de sciences de l’éducation, à propos de *l’analyse des données*.



b) Pour faire entendre la chose, je suppose ici qu’un étudiant en sciences de l’éducation a le projet d’instruire à propos de la question *Q* suivante, sans doute brutale, mais incontournable pour lui : « Comment donc analyser des données ? » Je suppose en outre que, pour tenter d’apporter de premiers éléments de réponse à la question *Q*, cet étudiant s’est procuré un exemplaire d’un petit ouvrage paru en janvier 2007 dans la collection « Que sais-je ? » (n° 3782, PUF) sous la signature de Philippe Cibois, intitulé opportunément *Les méthodes d’analyse d’enquêtes* : il s’agit là d’un exposé qui participe à l’évidence de « l’offre praxéologique » dans le domaine (voir ci-après).

c) Dans ce qui suit, nous examinerons le début du chapitre I de cet ouvrage – *Repérer les questions pertinentes* – avec en tête, toujours, la question *Q*, en nous arrêtant sur les points de difficulté que nous aurons repérés – quelle qu’en soit la nature supposée – comme sur autant de questions « auxiliaires » que l’on pourrait envisager soit d’étudier sans délai, soit, au contraire, d’accepter au moins provisoirement comme référant à des boîtes (laissées) noires, en prenant ainsi, *dans chacun des cas*, un risque quant à la réussite de l’entreprise. Nous examinerons ce texte peu à peu. En voici le début, sur lequel je ferai ensuite quelques commentaires.

Chapitre I

REPÉRER LES QUESTIONS PERTINENTES

On suppose une enquête déjà existante, soit issue d’une recherche, soit en vue de l’analyse secondaire d’une enquête disponible et rendue accessible aux chercheurs. On suppose aussi que les données de cette enquête sont utilisables par le biais soit d’un logiciel international comme SAS ou SPSS, soit d’un logiciel libre comme Trideux, développé par l’auteur et dont les exemples de ce livre sont issus. Les méthodes de dépouillement présentées ici sont indépendantes des logiciels : les aspects pratiques en dépendent évidemment et il faudra s’y reporter pour plus de détails.

→ L'auteur évoque d'abord la disponibilité des résultats d'une enquête et fait alors référence à des logiciels d'analyse de données (SAS, SPSS, Trideux). Toutefois la remarque selon laquelle « les méthodes de dépouillement présentées ici sont indépendantes des logiciels » permet au lecteur de considérer que, *pour le moment*, il peut laisser à ces logiciels le statut de *boîtes noires* – même s'il est prévenu qu'il « faudra s'y reporter pour plus de détails ».

→ À propos de la notion d'*analyse secondaire* utilisée par l'auteur, on trouve dans l'article *Secondary data* de Wikipedia l'explicitation suivante :

Secondary data analysis is commonly known as second-hand analysis. It is simply the analysis of preexisting data in a different way or to answer a different question than originally intended. Secondary data analysis utilizes the data that was collected by someone else in order to further a study that you are interested in completing.

Cette indication simple peut suffire – admettons-le, du moins – pour une première étude excriptive du texte.

d) Voici le paragraphe suivant.

On suppose donc que l'on a un fichier d'individus dont le nombre est variable, qui peut aller de quelques dizaines à plusieurs centaines de milliers : il peut sembler paradoxal d'envisager un dépouillement d'enquêtes avec moins de cent individus mais regarder attentivement le contenu d'un fichier est un objectif valable même si la possibilité d'étendre les résultats obtenus à une population de référence est faible. Quand on fait une enquête, quel que soit le nombre d'individus, on veut légitimement avoir une description de la population enquêtée : si l'effectif en est faible, on ne pourra que constater l'état de la population ; si l'effectif est important on pourra généraliser les résultats à la population dont l'enquête est issue, sous réserve que l'échantillon a été prélevé de manière raisonnée, par exemple par la méthode des quotas ou en sélectionnant des populations spécifiques. Il faut bien distinguer la description des données d'une part, des résultats qui peuvent être généralisés à l'ensemble de la population étudiée d'autre part. Pour pouvoir généraliser, on utilisera des tests statistiques, essentiellement celui du khi-deux que l'on supposera connu : on se souviendra que le khi-deux étant sensible aux effectifs, dès qu'une population d'enquêtés devient importante, il devient rare que le khi-deux d'un tableau croisé ne soit pas significatif.

→ Ce paragraphe est structuré autour des trois termes *individu*, *échantillon* (fichier), *population*. La première notation de l'auteur concerne l'*effectif* de l'échantillon : le public auquel il s'adresse est sans doute friand de « gros » effectifs (disons, plusieurs centaines) et il le met donc en garde contre le fait de s'obnubiler sur ce point : il convient de « regarder attentivement » même un « petit » fichier de données – observation que l'analyse des résultats du différenciateur sémantique esquissée plus haut illustre à propos, me semble-t-il.

→ Une deuxième notation concerne bien entendu le fait de passer des conclusions établies sur l'échantillon à des conclusions sur la population « enquêtée ». C'est rappeler d'abord que l'échantillon correspondant au fichier de données n'est là *en principe* que pour essayer de mieux connaître

la *population* dont il est issu. Ce passage se fait d'ailleurs parfois subrepticement, et bien sûr sans contrôle suffisant : ainsi, quand je disais plus haut qu'on a demandé à « 22 étudiants d'une licence professionnelle de sciences de l'éducation, dont le port d'attache est Lambesc » de remplir les grilles d'un différenciateur sémantique, on peut être tenté de conclure par exemple, sur la base de l'analyse des scores attribués par ces 22, que « les *étudiants* de Lambesc » n'ont pas une grande opinion de la statistique – ce qui, en toute rigueur ne peut être conclu sur la base *seulement* de l'enquête réalisée, quoique celle-ci, bien entendu, donne de la *vraisemblance* à cette assertion plutôt qu'elle n'en diminue le crédit.

→ Un point paraît obscur dans le texte de l'auteur : « si l'effectif en est faible, on ne pourra que constater l'état de la population... » Si l'effectif est faible, on pourra constater l'état de cette sous-population qu'est *l'échantillon*, sans pouvoir remonter jusqu'à la population. Sur ce point, l'enquête devrait continuer.

→ Pour pouvoir passer de l'échantillon à la population, précise alors l'auteur, il convient encore que l'échantillon ait été extrait « de manière raisonnée ». La mention de la *méthode des quotas* ressemble à celle d'une boîte noire. En voici tout de même une petite explication empruntée au site Internet d'Ipsos (http://www.ipsos.fr/CanalIpsos/cnl_static_content.asp?rubId=35#02).

Quels sont les avantages et les inconvénients de la méthode des quotas ?

Par rapport à la méthode aléatoire, celle des quotas a l'avantage d'être plus rapide. Avec l'aléatoire, les sondés ne sont pas « interchangeables ». Cela signifie que la personne tirée au sort doit être recontactée autant de fois que nécessaire. Grâce aux quotas, il est possible de remplacer un sondé par un autre qui a les mêmes caractéristiques socio-démographiques. Cela permet de réaliser un sondage dans des délais plus courts.

L'inconvénient majeur de la méthode des quotas est de ne pas permettre de calculer scientifiquement la marge d'erreur du sondage. Les lois statistiques qui permettent de la déterminer ne valent théoriquement que pour les sondages aléatoires. En pratique, on considère cependant que la marge d'erreur des sondages par quotas est égale ou inférieure à celle des sondages aléatoires.

Ce passage a le mérite d'attirer indirectement l'attention sur ce fait que l'auteur que nous suivons semble ignorer la « méthode aléatoire », c'est-à-dire le tirage au hasard dans la population des individus composant l'échantillon. En outre, la mention de la technique de sélection de « populations spécifiques » exigerait une enquête que l'on ne fera pas ici, ce qui, à nouveau, laisse une boîte noire dans le texte examiné.

→ L'auteur insiste à nouveau sur le travail de *description* des données, possible même quand l'effectif est faible, ainsi que sur les quelques résultats de ce travail que l'on peut envisager de généraliser à la population. L'outil pour « généraliser », ajoute-t-il, est la théorie des *tests statistiques*. L'auteur fait un sort particulier au test du χ^2 qu'il suppose explicitement être pour le

lecteur une « boîte claire », ce qui n'est rien moins que sûr ! Pour cela, nous nous arrêterons un instant sur ce point.

e) Voici tout d'abord un petit schéma de raisonnement, d'abord déterministe, ensuite aléatoire, qui éclaire ce que l'on fait dans un test statistique.

➔ Soit deux propositions p et q . On a le schéma de raisonnement suivant (que l'on peut appeler « raisonnement par contraposition ») :

Si p était vraie, alors q serait fausse. Or q est vraie. Donc p est fausse.

Ce schéma s'exprime par la tautologie suivante : $[(p \Rightarrow \neg q) \wedge q] \Rightarrow \neg p$. Cela rappelé, modifions un peu le vocabulaire et les notations. On suppose que l'on considère une certaine hypothèse H_0 (dite en statistique *hypothèse nulle*) et un certain événement E , liés par l'énoncé suivant :

Si H_0 était vraie, alors l'événement E ne pourrait pas se produire. Or on a observé la survenue de E . Donc H_0 est fausse.

Dans le cas *non déterministe*, on suppose que l'on dispose d'un *modèle probabiliste* qui fournit la probabilité de survenue de l'événement E sous l'hypothèse H_0 , ce qu'on note classiquement $P(E | H_0)$ et qu'on lit dans un langage abrégé : « probabilité de E sachant H_0 ». On a alors le schéma de raisonnement suivant, où δ est un nombre positif inférieur ou égal à 1 :

Si H_0 était vraie, alors l'événement E aurait une probabilité de survenir $P(E | H_0)$ inférieure ou égale à δ . Or on a observé la survenue de E . Donc on a observé la survenue d'un événement de probabilité inférieure δ .

Comme on le voit, le « raisonnement » est incomplet : on s'attendrait à une conclusion touchant la vérité ou la fausseté de H_0 ! Or ce qui manque est moins une conclusion qu'une *décision*. Supposons que l'on se soit fixé la règle d'action suivante : si la probabilité $P(E | H_0)$ est inférieure ou égale à, disons, $1/10\,000$, et si E se produit, je considérerai que ce résultat n'est pas « croyable », en ce sens que *je ne peux pas croire* que je viens de voir se produire sous mes yeux un événement si improbable ; je déciderai alors *de tenir H_0 pour fausse* – on dira que je *rejette H_0* (tout en étant conscient que je peux faire erreur : il est des événements plus rares encore qui se produisent !).

➔ On peut s'excentrer de ce schéma « subjectif » *en l'objectivant*. Soit un nombre ε compris strictement entre 0 et 1, par exemple $\varepsilon = 1/100$. On a observé la survenue d'un événement E tel que $P(E | H_0) \leq \varepsilon$; on dira alors que H_0 est rejeté au seuil ε . Si ε est grand, par exemple $\varepsilon = 0,5 = 50\%$, on va *rejeter H_0 à tort* en bien des cas : c'est ce qu'on nomme le risque de *première espèce*. Si ε est très petit, par exemple $\varepsilon = 10^{-6}$, on va souvent *accepter H_0 à tort* : c'est le risque de *deuxième espèce*.

→ Voici maintenant un exemple d'utilisation du test du χ^2 – une praxéologie dont on ne rendra ici claire qu'une partie. On considère une population de personnes. On suppose qu'on en a extrait au hasard un échantillon de 6800 personnes sur lesquelles on a prélevé chaque fois deux informations, la couleur de yeux, X , et la couleur des cheveux, Y . On pense que ces deux caractères *ne sont pas indépendants* – que par exemple, quand on a les cheveux blonds, on a plus souvent les yeux bleus, etc. Considérons le tableau suivant.

	Blond	Brun	Noir	Roux	Totaux
Bleu					2811
Gris-vert					3132
Brun					857
Totaux	2829	2632	1223	116	6800

On va alors formuler l'hypothèse nulle H_0 selon laquelle X et Y seraient indépendants – avec, bien sûr, l'idée que H_0 est *fausse*, et l'espoir de pouvoir *la rejeter*, même en étant exigeant, par exemple en prenant un seuil ε égal à $1/1000$. Si X et Y étaient indépendants, la distribution des différentes couleurs de cheveux parmi les personnes aux yeux bleus serait la même que dans l'échantillon tout entier : il y aurait par exemple une proportion de $\frac{2829}{6800} \approx 41,6\%$ des enquêtés qui auraient les yeux bleus, soit exactement $\frac{2829}{6800} \times 2811$ personnes ayant les yeux bleus. On aurait pu aussi calculer la proportion de personnes ayant les cheveux blonds parmi les personnes ayant les yeux bleus. Elle devrait être la même que dans l'échantillon tout entier, soit $\frac{2811}{6800}$, en sorte que le nombre de personnes ayant à la fois les yeux bleus et les cheveux blonds serait égal à $\frac{2811}{6800} \times 2829$. On retrouve ainsi le même effectif « théorique », que l'on peut écrire ainsi : $c_{11} = \frac{2829 \times 2811}{6800} \approx 1170$. On remplit de même les autres cellules du tableau :

	Blond	Brun	Noir	Roux	Totaux
Bleu	c_{11}	c_{12}	c_{13}	c_{14}	2811
Gris-vert	c_{21}	c_{22}	c_{23}	c_{24}	3132
Brun	c_{31}	c_{32}	c_{33}	c_{34}	857
Totaux	2829	2632	1223	116	6800

Voici maintenant le tableau « empirique » formés par les effectifs *observés*.

	Blond	Brun	Noir	Roux	Totaux
Bleu	1768	807	189	47	2811
Gris-vert	946	1387	746	53	3132
Brun	115	438	1223	16	857
Totaux	2829	2632	1223	116	6800

On voit que, dans le cas des personnes à yeux bleus et cheveux blonds, ce tableau observé s'éloigne du tableau théorique : l'effectif observé (1768) est assez nettement *supérieur* à l'effectif théorique (1170). Cet écart entre tableau théorique et tableau empiriquement observé et les autres écarts (en plus ou en moins) que nous pourrions calculer sont-ils explicables par la fluctuation d'échantillonnage, qui peut expliquer de petites écarts aux effectifs théoriques ? C'est là qu'intervient le teste du χ^2 .

➔ Le χ^2 est un indicateur numérique qui évalue la *distance* ou, disons, la *dissemblance* entre les deux tableaux : si on note n_{ij} les effectifs *observés*, on a de façon générale

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - c_{ij})^2}{c_{ij}}.$$

Ici, le calcul donne : $\chi^2 \approx 1075$. L'événement E évoqué plus haut est $\chi^2 \geq 1075$. Or il se trouve que la probabilité d'observer cet événement – une distance aussi grande que 1075 – sous l'hypothèse H_0 d'indépendance est *très faible*. En fait, elle est *quasi nulle*. Comment le savoir ? On peut aujourd'hui utiliser un calculateur en ligne comme celui que l'on voit utilisé ci-après (http://www.physics.csbsju.edu/stats/contingency_NROW_NCOLUMN_form.html). Voici le premier écran que l'on rencontre.

r×c Contingency Table: How many rows? columns?

You are about to enter your data for a chi-square contingency table analysis. For this to make sense you should have a table of data (at least 2x2; maximum: 9x9).

Number of rows:

Number of columns:

You must have data for each box in the table: no blanks allowed!

Le nombre de lignes (*rows*) est ici 3, le nombre de colonnes (*columns*) 4. En cliquant sur *Submit*, on voit apparaître un écran comportant une table 3×4 où l'on saisit les effectifs empiriques observés.

Data Entry: 3 × 4 Contingency Table

Your data have been doubly sorted into one of the 4 "treatments" (A, B, ...) and into one of the 3 "outcomes" (1, 2, ...). In the below table you are to enter the number of data items that fit into a particular (treatment,outcome) pair e.g., (B1 or A2). Enter your data in the below set of boxes and then click on the **Calculate Now** button when data entry is complete. Leave no empty boxes! All entries should be whole numbers!

	A	B	C	D
1	1768	807	189	47
2	946	1387	746	53
3	115	438	288	16

En cliquant sur *Calculate Now*, on obtient alors ceci.

$r \times c$ Contingency Table: Results

The results of a contingency table χ^2 statistical test performed at 06:34 on 21-MAR-2008

data: contingency table

	A	B	C	D	
1	1768	807	189	47	2811
2	946	1387	746	53	3132
3	115	438	288	16	857
	2829	2632	1223	116	6800

expected: contingency table

	A	B	C	D	
1	1.169E+03	1.088E+03	506.	48.0	
2	1.303E+03	1.212E+03	563.	53.4	
3	357.	332.	154.	14.6	

chi-square = 0.107E+04

degrees of freedom = 6

probability = 0.000

Le logiciel a calculé 1. les effectifs marginaux, 2. les effectifs théoriques, 3. la valeur de χ^2 , 4. le nombre de degrés de liberté (*degrees of freedom*, notion sur laquelle on reviendra : ce nombre vaut ici $(3 - 1)(4 - 1) = 6$), 5. La probabilité de l'événement $\chi^2 \geq 0,107 \cdot 10^4$, qu'il trouve égale à 0,000. Dans cette situation, on ne peut guère que *rejeter* l'hypothèse d'indépendance entre les caractères X et Y . On est ici, en vérité, dans un cas qui illustre cette remarque de l'auteur des *Méthodes d'analyse d'enquêtes* : « On se souviendra que le khi-deux étant sensible aux effectifs, dès qu'une population d'enquêtés devient importante, il devient rare que le khi-deux d'un tableau croisé ne soit pas significatif. » De fait, à l'aide du calculateur disponible à l'adresse <http://www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html>, on obtient par exemple ceci.

Calculate χ^2 from probability Q and d

To determine the chi-square value indicating a probability Q of non-chance occurrence for an experiment with d degrees of freedom, enter Q and d in the boxes below and press **Calculate**.

Given probability $Q=$ and $d=$

The χ^2 value is:

Autrement dit, pour que la différence entre les deux tableaux – théorique et observé – soit significative au seuil de 1 %, il suffit que χ^2 atteigne ou

dépasse la valeur 16,8118. Voici un tableau de valeurs que l'on obtient avec le même calculateur (pour 6 degrés de liberté).

Prob. $\leq$...	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
$\chi^2 \geq$...	16,8118	22,4577	27,8563	33,1070	38,2583	43,3292	48,1824	51,6164

Observer un χ^2 de plus de 52 est, sous l'hypothèse d'indépendance de X et Y , un événement dont la probabilité est déjà inférieure à *un milliardième* ; observer un χ^2 de plus 1070 a donc une probabilité infinitésimale.

That's all, folks!