



HAL
open science

Bilan ((en guise de conclusion du Colloque Consensus ex machina, Sorbonne, 1994))

Etienne Brunet

► To cite this version:

Etienne Brunet. Bilan ((en guise de conclusion du Colloque Consensus ex machina, Sorbonne, 1994)). Consensus ex machina, Apr 1994, Paris, France. hal-01296471

HAL Id: hal-01296471

<https://hal.science/hal-01296471>

Submitted on 4 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bilan

(en guise de conclusion du Colloque
Consensus ex machina, Sorbonne, 1994)

Étienne Brunet

Université de Nice

1 - Je ne suis pas le pompier de service chargé de vous avertir que le feu a pris une nouvelle fois dans les greniers de la Sorbonne; je ne suis pas davantage un agent de sécurité venu pour faire évacuer la salle parce que les étudiants menaceraient, comme naguère, d'occuper les lieux. Hélas, la Sorbonne n'est pas en feu, ni en ébullition. C'est beaucoup plus grave: ceux qui ont lu le programme primitif et qui s'attendent à voir maintenant la figure imposante de Bernard Quemada, à entendre sa belle voix de baryton et à profiter de son expérience exceptionnelle vont être déçus, frustrés, désespérés. Bernard Quemada, retenu ailleurs, dans un autre pays, un autre continent, peut-être une autre planète, ne viendra pas aujourd'hui dans cette salle Richelieu qui, il y a quelque mois, accueillait, à propos de l'Europe, le Président de la République et la télévision. Et la silhouette peu flatteuse qui s'offre à votre vue n'appartient pas à Quemada mais à son remplaçant.

2 - Il y a d'autres absents de marque: le Père Busa, Joseph Raben, Louis Milic, Charles Muller, Louis Delatte, d'autres encore, qu'on a désignés à un précédent colloque sous le nom de prophètes. Ces prophètes, à en juger par le succès de ce colloque et la vénération de leurs disciples, n'auront pas prêché dans le désert. De façon paradoxale, on peut mesurer au nombre des absents la réussite d'une manifestation, et la maturité d'une discipline. Quand s'écartent ses fondateurs, c'est le signe qu'une discipline peut voler de ses propres ailes. La génération des pionniers laisse la place alors à celle des fermiers et l'exploitation prend le relais de l'exploration.

3 - Un jour viendra peut-être où les absents ne seront plus comme aujourd'hui les fondateurs éloignés ou les amis empêchés, mais des adversaires, réfractaires ou rebelles. Quand une discipline se développe, les dissensions, les querelles, voire les schismes, finissent par s'y produire, qui

engendrent les guerres de sécession. Le temps de la division n'est heureusement pas encore arrivé. On en est au stade de la multiplication et le colloque qui s'achève m'a paru largement consensuel, comme son titre le laissait prévoir. Il est encore trop tôt pour remercier les organisateurs, ce qu'on fera dans une heure, quand, une fois terminé mon exposé, les risques de catastrophe seront définitivement écartés. Qu'il me soit tout de même permis de rendre hommage à Christian Delcourt, Maurice Tournier et André Salem, non pas seulement pour l'excellence de l'organisation, mais aussi pour la justesse de l'orientation et leur volonté d'allumer un foyer de concorde. Ils ont veillé à ce que ce colloque ne ressemble pas à une constellation d'étoiles dans la nuit, chacune diffusant son éclat solitaire, mais plutôt à une réaction en chaîne qui multiplie l'énergie et la lumière. Mieux même, il me semble que le consensus, sollicité par les organisateurs, a été obtenu de façon naturelle, par l'accord spontané de tous sur l'essentiel, et non pas par artifice, comme les feux du même nom, non par l'effet d'un happy end de dernière heure, d'un "deus ex machina" qui tomberait des nues au moment du dernier acte, c'est-à-dire maintenant au moment où s'amorce la scène finale du cinquième jour.

- I -

Ce constat initial me rassure pour accomplir la tâche qui m'est allouée et qui consiste à établir un bilan. J'aurais eu de l'inquiétude si ce bilan avait dû être le bilan d'un Liban (permettez-moi ce premier calembour, en espérant que ce sera aussi le dernier), c'est-à-dire le bilan d'un éclatement, d'une balkanisation ou libanisation des forces en présence. Heureusement ces forces poussent sinon dans le même sens, du moins à l'unisson, et, si les rayons divergent, ils viennent d'une même source. Je ressens toutefois quelque embarras pour remplir mon contrat. D'une part ce colloque s'inscrit dans une longue série et son objectif est précisément de faire le point des dernières avancées et de baliser le terrain nouvellement conquis. Ces jalons périodiques sont eux-mêmes des bilans, qu'il n'est pas très aisé de résumer en quelques mots. Et pour avoir une vue d'ensemble de la chaîne, il serait utile d'avoir assisté à tous les colloques de la série, ce dont je ne puis me vanter. Plusieurs manifestations m'ont laissé un souvenir impérissable, à Dartmouth College, à Pise, à Louvain, à Jérusalem, à Goeteborg, à Toronto, et bien sûr aussi à Nice. Mais certains maillons manquent et certaines étapes dans cet itinéraire moderne qui va, à rebours, de Jérusalem à Paris. D'autre part le bilan d'une manifestation comme celle-ci impliquerait des dons que je n'ai pas encore, comme l'omniscience et l'ubiquité. Pour rendre compte des séances parallèles, il aurait fallu que j'installe une oreille dans chacune des trois salles différentes où le colloque a distribué ses mer-

veilles. Or ceux qui me connaissent savent qu'on ne peut guère se fier à mes oreilles et qu'elles sont si infidèles que j'ai de la peine à accorder leurs témoignages. Heureux le cyclope qui certes n'avait qu'un oeil mais disposait de deux oreilles sensibles et d'un "tympan si sonore qu'on l'entendait entendre", comme le prétend Giraudoux.

1 - Pour être l'"écho sonore" de ce colloque, j'ai donc capté le son à la source, comme font les commentateurs de télévision qui rendent compte des programmes avant qu'ils n'aient été diffusés sur l'antenne. J'ai donc réfléchi au bilan de l'évènement avant que l'évènement ait lieu. Évidemment il fallait prévoir que l'évènement pût tourner mal. J'ai donc préparé un discours de rechange, pour les cas de catastrophe, et je le tiens prêt dans ma poche gauche - celle que les latins appelaient sinistre - un tremblement de terre pouvant survenir à l'improviste, dans la minute même (ce n'est pas une permission donnée aux fidèles du dernier rang de se mettre à l'abri dans le bistrot du coin, en attendant la fin de la messe!). Au reste les séismes de la nature sont improbables à Paris mais on peut y craindre de temps à autre les séismes sociaux, les révolutions.

Sur ma demande me fut donc envoyé par anticipation le recueil des résumés des communications, ce livre dense, de 250 pages, que chacun des congressistes a reçu au moment de l'inscription et qui lui a servi de guide pour s'orienter vers les exposés de son choix. Je n'avais aucun choix douloureux à décider, mais simplement une lecture à faire, attentive et exhaustive. Pour plus de garantie j'ai demandé à l'ordinateur de lire ce texte avec moi, par-dessus mon épaule. En deux jours, lui et moi, nous avons tout enregistré¹, et tout contrôlé, en pratiquant la correction mutuelle, comme font les moines. Et, une semaine avant l'ouverture du colloque, l'équipe avait les éléments nécessaires pour faire le résumé des résumés. Je me garderai bien pourtant de le produire devant vous, car la machine m'a lâchement abandonné en cours de route, au moment de la synthèse. Du moins m'avait-elle permis de localiser chacun des mots de ce corpus, en y ajoutant quelques informations quantitatives et collectives, dont l'intérêt peut sembler mince: 28 472 occurrences dans les résumés en français et 61473 dans les abstracts en anglais (et respectivement 4983 et 6930 formes différentes). Nul besoin de la machine pour établir la proportion entre les deux langues. Un simple relevé manuel aurait suffi: sur 119 exposés 41, soit le tiers, sont proposés dans la langue du pays d'accueil. Nos visiteurs ignorent peut-être que le gouvernement national tente de sauver l'usage du français

¹ Pas exactement tout, car le scanner s'est trouvé désarmé devant quelques communications (une demi-douzaine) dont l'encre, la typographie ou la reproduction laissaient à désirer.

comme langue internationale, au moins dans les manifestations scientifiques qui se déroulent dans l'hexagone. Il ne sera pas fâché d'apprendre que ses directives ont été respectées dans ce haut lieu de la Sorbonne, dont les murs ont renvoyé plus souvent les échos du français que de l'anglais, et, au reste, plus souvent ceux du latin que du français. Remarquons au passage que les organisateurs du colloque, en recourant sournoisement au latin pour le titre du colloque, ont évité habilement de prendre parti entre les deux langues rivales. Il eut été plaisant d'entendre quelques communications exposées en latin, comme ce fut la coutume pour les thèses en Sorbonne jusqu'au XIX^e siècle. Mais le Père Busa n'étant pas aujourd'hui parmi nous, personne ne pouvait relever ce défi.

Il n'en reste pas moins que cette proportion du simple au double est exactement l'inverse de celle qu'on avait observée à Nice, à l'occasion d'un colloque qui s'était tenu en 1985 sous l'égide de l'ALLC et où 58 communications sur 83 (soit les deux tiers) avaient choisi la langue du pays d'accueil. Mais avant de déplorer précipitamment le déclin du français, il convient d'observer que l'audience internationale n'est pas semblable à dix ans de distance. Si la variété des nations représentées est la même dans les deux cas (17 exactement), l'équilibre des masses est fort différent: la présence américaine, fort discrète à Nice en 1985, est ici vigoureusement affirmée (avec un tiers des communications à l'actif des États-Unis et du Canada). On ne peut que se féliciter de ce progrès et, si le français y perd un peu, la discipline y gagne puisqu'elle ignore de plus en plus les frontières. L'essentiel est que chacun puisse s'exprimer dans la langue où il se sent le plus à l'aise - ce que je fais présentement en m'excusant auprès du public anglophone qui est ici le plus nombreux et le plus patient.

La comparaison avec le colloque de Nice permet d'autres conclusions plus intéressantes. Certains se souviennent peut-être avec indulgence de ce premier colloque ALLC organisé en France, dans une Université jeune, qui n'avait que vingt ans, et dans un palais neuf, qui n'avait qu'un an. Ici, du haut de cette pyramide inversée qu'est l'amphithéâtre Richelieu, huit siècles vous contemplent, dont l'empreinte est profondément gravée dans les pierres, sinon dans les murs, et plus profondément encore dans les esprits, puisque s'est créé ici, dès le 13^e siècle, le moule ou la matrice de la pensée occidentale. Quel progrès dans ce recul, quelle profondeur dans cette perspective de l'histoire! Les propos qu'on a entendus ici depuis quatre jours s'ajoutent à des milliers de discours tenus dans ce lieu depuis Robert de Sorbon et leurs échos se mêlent à ceux des grandes voix qui ont retenti dans cette enceinte.

2 - Mais comme la solennité des lieux risque de paralyser l'orateur, un brin de dérision peut lui permettre de retrouver son aplomb. Comme les circonstances ont fait de lui l'organisateur du colloque de Nice, il disposait dans ses archives des communications faites il y a dix ans et publiées presque aussitôt dans deux gros volumes d'Actes. Il n'y a pas de bilan en effet sans comparaison, sans retour en arrière. Et le colloque de 1985, qui s'est tenu dans des conditions voisines de celui-ci, a paru un jalon adéquat pour mesurer le chemin parcouru. Le texte des Actes de 1985 a donc été réintroduit dans la mémoire électronique afin d'être confronté à celui des résumés de 1994. Mais avant de livrer les résultats de cette confrontation, j'ai voulu vérifier, par une expérience dérisoire, si l'ordinateur avait le don des langues et savait par exemple reconnaître le français de l'anglais. J'ai feint d'être sourd à la différence en mêlant dans le même corpus et dans le désordre une vingtaine d'exposés de chacune des deux langues, dont l'ensemble représente 150 000 mots. Chacun de ces mots a apporté son témoignage pour établir la distance des exposés pris deux à deux. Le principe est que deux textes sont d'autant plus proches qu'ils ont plus de mots en commun, compte tenu de leur étendue respective. Une partie des distances calculées (il y en a près d'un millier) est donné dans le tableau 1, qui reste assez opaque à l'oeil humain.

Tableau 1. Distances lexicales des exposés de 1985 pris deux à deux

| HYPERbaseALLC | | | | | | | | | | | | | | | | | | | |
|--|------|------|------|--------|------|----------|------|----------|------|--------------|------|--------|------|------|------|---|--|---|--|
| Edition Graphiq. Dict. Classe Sélect | | | | | | | | | | DICTIONNAIRE | | | | | | | | | |
| cliquer sur un mot pour voir le contexte | | | | | | | | | | | | | | | | | | | |
| D | | C | | Hiérar | | Richesse | | Distance | | Courbe | | Factor | | P | | K | | K | |
| 1851 | 1901 | 1878 | 1912 | 1877 | 1475 | 1494 | 1421 | 1418 | 1430 | 1423 | 1406 | 1427 | 1452 | 1460 | 1434 | | | | |
| | | | | | | | | | | 1403 | 1373 | 1515 | 1542 | 1409 | 1510 | | | | |
| Dugast | | Duga | 1906 | 1813 | 1887 | 1903 | 1910 | 1863 | 1886 | 1880 | 1884 | 1863 | 1903 | 1820 | 1901 | | | | |
| 1859 | 1888 | 1871 | 1903 | 1895 | 1437 | 1440 | 1467 | 1490 | 1519 | 1466 | 1476 | 1433 | 1483 | 1413 | 1511 | | | | |
| | | | | | | | | | | 1507 | 1515 | 1335 | 1404 | 1439 | 1481 | | | | |
| Evrard | | Evra | 1922 | 1842 | 1901 | 1912 | 1896 | 1879 | 1910 | 1897 | 1899 | 1855 | 1905 | 1823 | 1921 | | | | |
| 1883 | 1874 | 1874 | 1903 | 1916 | 1449 | 1433 | 1462 | 1400 | 1512 | 1437 | 1506 | 1426 | 1457 | 1398 | 1513 | | | | |
| | | | | | | | | | | 1485 | 1542 | 1404 | 1330 | 1419 | 1419 | | | | |
| Farah | | Fara | 1891 | 1836 | 1870 | 1902 | 1906 | 1846 | 1889 | 1888 | 1887 | 1897 | 1900 | 1789 | 1886 | | | | |
| 1871 | 1892 | 1888 | 1876 | 1890 | 1395 | 1458 | 1461 | 1409 | 1463 | 1420 | 1484 | 1416 | 1435 | 1416 | 1470 | | | | |
| | | | | | | | | | | 1467 | 1409 | 1439 | 1419 | 1354 | 1427 | | | | |
| Fortier | | Fort | 1876 | 1795 | 1859 | 1877 | 1854 | 1828 | 1876 | 1834 | 1861 | 1843 | 1857 | 1742 | 1874 | | | | |
| 1820 | 1854 | 1851 | 1879 | 1862 | 1392 | 1423 | 1461 | 1446 | 1386 | 1455 | 1455 | 1450 | 1461 | 1407 | 1492 | | | | |
| | | | | | | | | | | 1474 | 1510 | 1481 | 1419 | 1427 | 1356 | | | | |
| Frautschi | | Frau | 1903 | 1832 | 1893 | 1900 | 1899 | 1882 | 1895 | 1877 | 1892 | 1840 | 1891 | 1802 | 1900 | | | | |
| 1888 | 1890 | 1882 | 1887 | 1910 | 1460 | 1451 | 1493 | 1482 | 1508 | 1497 | 1522 | 1466 | 1490 | 1451 | 1513 | | | | |
| | | | | | | | | | | 1513 | 1509 | 1413 | 1380 | 1466 | 1479 | | | | |
| Catach | | Cata | 1901 | 1880 | 1896 | 1907 | 1907 | 1879 | 1906 | 1864 | 1903 | 1889 | 1906 | 1821 | 1904 | | | | |
| 1866 | 1916 | 1882 | 1893 | 1907 | 1392 | 1464 | 1422 | 1437 | 1459 | 1460 | 1409 | 1472 | 1431 | 1398 | 1464 | | | | |
| | | | | | | | | | | 1412 | 1437 | 1455 | 1510 | 1422 | 1414 | | | | |
| Flikeid | | Flik | 1902 | 1859 | 1887 | 1901 | 1892 | 1864 | 1900 | 1861 | 1876 | 1898 | 1898 | 1816 | 1898 | | | | |
| 1839 | 1878 | 1870 | 1906 | 1878 | 1452 | 1497 | 1435 | 1436 | 1489 | 1481 | 1468 | 1466 | 1512 | 1449 | 1491 | | | | |
| | | | | | | | | | | 1435 | 1429 | 1502 | 1490 | 1492 | 1479 | | | | |

Tableau 3. Décantation automatique de l'anglais et du français

| CORRÉLATION CHRONOLOGIQUE | | | | | |
|-------------------------------|------|-------|------------------------------|------|-------|
| EN PROGRESSION (hiérarchique) | | | EN REGRESSION (hiérarchique) | | |
| +0.84 | 1310 | les | -0.84 | 1370 | to |
| +0.83 | 1252 | des | -0.83 | 2446 | of |
| +0.83 | 1158 | à | -0.83 | 1899 | a |
| +0.82 | 3092 | de | -0.82 | 4060 | the |
| +0.82 | 1365 | et | -0.82 | 1361 | and |
| +0.82 | 673 | une | -0.82 | 581 | be |
| +0.81 | 504 | qui | -0.82 | 415 | by |
| +0.80 | 1172 | le | -0.80 | 941 | is |
| +0.80 | 773 | en | -0.80 | 443 | that |
| +0.80 | 679 | du | -0.80 | 377 | it |
| +0.80 | 457 | pour | -0.80 | 236 | not |
| +0.79 | 1581 | la | -0.79 | 501 | as |
| +0.79 | 494 | par | -0.79 | 480 | are |
| +0.79 | 297 | ce | -0.78 | 282 | an |
| +0.78 | 548 | que | -0.78 | 184 | one |
| +0.78 | 79 | dont | -0.77 | 618 | for |
| +0.76 | 649 | dans | -0.77 | 189 | have |
| +0.75 | 237 | se | -0.77 | 161 | more |
| +0.74 | 1261 | l' | -0.76 | 416 | this |
| +0.73 | 267 | sont | -0.76 | 342 | with |
| +0.73 | 167 | ces | -0.76 | 152 | has |
| +0.72 | 621 | est | -0.75 | 1351 | in |
| +0.72 | 319 | plus | -0.75 | 396 | which |
| +0.72 | 211 | s' | -0.75 | 110 | its |
| +0.72 | 178 | entre | -0.73 | 134 | other |
| +0.70 | 1193 | d' | -0.71 | 312 | or |
| +0.70 | 158 | comme | -0.71 | 73 | so |

Voilà la première constatation heureuse que m'inspire le présent colloque. Le consensus n'y est pas limité aux humanités, c'est-à-dire aux linguistes et aux littéraires. Il s'étend d'un côté à l'ensemble des sciences sociales et de l'autre aux disciplines qui ont à voir avec la logique, avec les nombres et avec la technologie et qu'on nomme sciences de l'ingénieur.

Le couple formé par les deux organisateurs locaux de la présente manifestation donne d'ailleurs l'illustration de cette nouvelle alliance: d'un côté Maurice Tournier donne la garantie que la chair humaine du discours sera préservée dans son intégralité textuelle et historique. De l'autre André Salem qui vient de l'horizon mathématique apporte l'assurance méthodologique et la sécurité technique. Il est regrettable qu'on n'ait guère entendu de sa bouche que des messages d'intendance. Il m'appartient, ayant assisté à la soutenance de sa thèse dans une salle voisine, de forcer sa modestie en signalant la parution, la semaine dernière, d'un ouvrage magistral qui est dérivé de sa thèse et dont il partage la paternité avec Ludovic Lebart. Cette *Statistique textuelle* (c'est le nom de l'ouvrage) m'apparaît comme la synthèse la plus solide des techniques quantitatives offertes aux chercheurs d'aujourd'hui dans nos disciplines. Et si l'on a besoin d'établir un bilan, c'est là qu'on le trouvera - ce qui simplifie grandement ma tâche aujourd'hui.

- II -

Mais je ne voudrais pas donner l'impression de m'en tirer à trop bon compte en détournant l'attention sur autrui. Je vais donc revenir à mes données, en en faisant deux lots homogènes. Le premier englobe tous les textes anglais publiés dans les Actes du Colloque de 1985 et tous les résumés anglais de la présente manifestation. Le second opère le même regroupement pour les textes français. Il s'agit dans les deux cas de mesurer l'évolution, en l'espace de dix ans, des thèmes traités.

1 - Comparons d'abord les deux listes opposées des tableaux 4 et 5. À gauche c'est le vocabulaire spécifique de 1985, à droite celui de 1994. Les deux listes sont triées selon la valeur de l'écart réduit, ce qui ne favorise pas les regroupements thématiques. Pourtant il n'est pas difficile de reconnaître les orientations dominantes de la recherche il y a dix ans. Il s'agit manifestement de statistique lexicale, comme l'annonçait d'ailleurs le titre du Colloque: *méthodes quantitatives et informatiques dans l'étude des textes*. Les unités de compte y sont soigneusement délimitées, depuis les éléments infra-lexicaux: *characters, letter, letters, item, items*, jusqu'aux entrées du dictionnaire: *word* (qui est le premier mot de la liste), *words, lemmas, vocabulary, entries, lexical, list, string*. Et les opérations de calcul y sont longuement développées: *frequency* (c'est le deuxième mot), *freq, occurrence, occurrences, distribution, distributions, sampling, rank, mean, divided, expected, formula, formulae, value, values, thousand, table, diagram, figure*, etc.. L'aspect informatique est également évoqué, l'accent étant mis sur ce qui était à la mode à cette époque, notamment les bases de données: *database, databank, field, fields, record, records, file, base*, et aussi sur les difficultés liées alors aux limites techniques des machines et à l'étroitesse de la mémoire disponible: *bits, code, codes, size, length, storage, stored, package, stream, price, marketing*.

Tableau 4. Excédents 1985

1985 total écart

| | | | | | | | | | | | | | | | |
|-----|-----|---|-----------|----|----|---|---------------|----|-----|---|----------------|----|----|---|------------|
| 277 | 338 | 9 | word | 23 | 24 | 3 | Brainerd | 31 | 35 | 3 | letters | 14 | 14 | 3 | skewed |
| 98 | 107 | 7 | frequency | 62 | 84 | 3 | case | 70 | 101 | 3 | level | 19 | 19 | 3 | skews |
| 56 | 57 | 6 | code | 36 | 46 | 3 | characters | 79 | 114 | 3 | lexical | 14 | 14 | 3 | span |
| 70 | 75 | 6 | items | 29 | 35 | 3 | codes | 26 | 29 | 3 | low | 41 | 51 | 3 | special |
| 66 | 76 | 5 | file | 12 | 12 | 3 | collocate | 44 | 58 | 3 | machine | 30 | 35 | 3 | stored |
| 42 | 43 | 5 | price | 12 | 12 | 3 | collocates | 20 | 20 | 3 | marketing | 23 | 24 | 3 | string |
| 56 | 60 | 5 | records | 16 | 16 | 3 | databank | 16 | 17 | 3 | match | 18 | 18 | 3 | systemic |
| 47 | 51 | 5 | storage | 19 | 21 | 3 | definitions | 12 | 12 | 3 | mean | 19 | 20 | 3 | thousand |
| 40 | 42 | 5 | table | 13 | 13 | 3 | derivatives | 63 | 83 | 3 | meaning | 13 | 13 | 3 | tomes |
| 184 | 259 | 5 | words | 16 | 16 | 3 | diagram | 42 | 54 | 3 | method | 27 | 32 | 3 | tree |
| 141 | 205 | 4 | database | 41 | 52 | 3 | distribution | 20 | 20 | 3 | microcomputers | 25 | 30 | 3 | value |
| 57 | 69 | 4 | field | 13 | 13 | 3 | distributions | 12 | 12 | 3 | Muller | 40 | 50 | 3 | values |
| 23 | 23 | 4 | formula | 15 | 16 | 3 | divided | 32 | 40 | 3 | name | 56 | 72 | 3 | vocabulary |
| 24 | 25 | 4 | hebrew | 33 | 39 | 3 | entries | 19 | 20 | 3 | node | | | | |
| 29 | 30 | 4 | item | 28 | 31 | 3 | expected | 21 | 22 | 3 | nominal | | | | |
| 54 | 65 | 4 | list | 33 | 38 | 3 | expert | 24 | 26 | 3 | occurrence | | | | |
| 33 | 34 | 4 | prolog | 37 | 48 | 3 | fields | 25 | 30 | 3 | occurrences | | | | |
| 27 | 28 | 4 | rank | 22 | 24 | 3 | figure | 12 | 12 | 3 | outer | | | | |
| 65 | 78 | 4 | record | 17 | 18 | 3 | formulae | 15 | 16 | 3 | package | | | | |
| 32 | 36 | 4 | stream | 13 | 13 | 3 | freq | 13 | 13 | 3 | predication | | | | |
| 31 | 39 | 3 | 1985 | 14 | 14 | 3 | functor | 19 | 19 | 3 | ratings | | | | |
| 34 | 42 | 3 | base | 26 | 31 | 3 | lemma | 20 | 20 | 3 | sampling | | | | |
| 12 | 12 | 3 | bits | 36 | 42 | 3 | length | 33 | 42 | 3 | size | | | | |
| 12 | 12 | 3 | bootstrap | 27 | 30 | 3 | letter | 13 | 13 | 3 | skew | | | | |

2 - Combien apparaît plus large le champ ouvert dans le tableau 5, qui correspond aux préoccupations actuelles. Au lieu d'être confiné dans le bureau exigu de la statistique, on est introduit dans la grande salle où les questions fondamentales de la recherche sont abordées et où les grands mots ne sont pas récusés: *research, analysis* (qui se trouvent en tête de la liste), *project, développement, work, study, studies, scientific*. L'aspect théorique ne fait plus peur: *theory, concept, concepts, conceptual, knowledge*, non plus que les moyens concrets à mettre en oeuvre: *use, way, tools*, et les débats à engager: *question, questions, interpretation, responses, dialogue, discuss, errors*³. La machine, présente dans le titre du colloque, occupe le devant de la scène sous la forme substantive (*computer* est le premier mot spécifique, devant *computing, information, technology, software, program*) ou adjectivale (*electronic, digital*). Plus rien n'y évoque les limites ou les contraintes de l'instrument et le préfixe micro a disparu. La puissance est venue avec la miniaturisation et les grands écrans accompagnent les petits boîtiers. Mais ce qui frappe le plus c'est le changement de cap qui offre à la discipline un horizon plus vaste, centré autour du texte. Quoique le texte fût au coeur de la problématique de 1985 (il figurait dans le titre même), il apparaît bien davantage en 1994, sous des aspects plus riches et plus variés, documentaires et linguistiques: *text, texts, hypertext, corpus, language, linguistic, speech, book, dictionary, grammar*, ou plus spécifiquement littéraires: *literature, literary, style, stylistic, poetry, narrative, play, chinese, greek, authors, humanities*. Il faut saluer cette rencontre des linguistes et des littéraires, assis à la même table, autour du texte. On a vu dans le passé les premiers s'écarter, parce qu'ils n'avaient pas besoin de corpus (le sentiment linguistique suffisant à juger de la grammaticalité) et

³ Il est toutefois possible que la loi du genre intervienne ici. On compare en effet deux séries d'énoncés qui n'ont pas exactement le même statut. Le résumé d'une communication future tend à utiliser des termes plus généraux que la communication elle-même, une fois rédigée. Le fait que les résumés du présent livret sont souvent tronqués complique encore la situation. En de tels cas on a affaire à l'introduction de l'exposé, laquelle montre en principe une problématique plus ouverte et plus large que le cors du sujet.

les seconds se tenir au loin, parce qu'ils n'avaient pas besoin de machines. Que les retrouvailles autour de l'ordinateur se fassent à la Sorbonne, voilà un signe des temps qui ne peut laisser personne indifférent. Je ne connais qu'un autre lieu où la symbolique puisse être aussi forte: c'est l'Académie française, où Richelieu - encore lui - accueillera peut-être un jour le congrès de l'ALLC. Cette présence grandissante du texte, toutes formes regroupées, est visible dans la distribution des mots de la famille où l'on dénombre 1200 occurrences. Les déficits sont plus nombreux dans le premier colloque et les excédents dans le second.

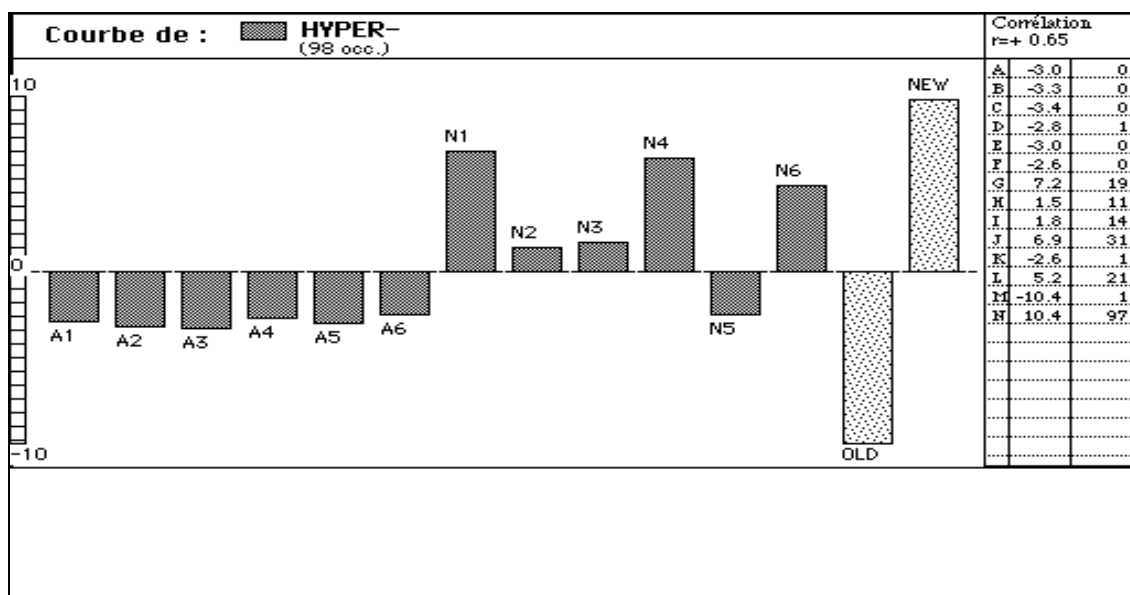
Tableau 5. Excédents 1994

| 1985 | total | écart | | 63 | 75 | 6 | study | 43 | 50 | 5 | theory | 19 | 20 | 4 | historical |
|------|-------|-------|-------------|----|-----|---|----------------|----|-----|---|---------------|----|-----|---|---------------|
| | | | | 29 | 29 | 6 | tag | 25 | 27 | 5 | turn | 17 | 18 | 4 | ideas |
| | | | | 29 | 29 | 6 | TEI | 27 | 28 | 5 | verbal | 18 | 20 | 4 | instance |
| 212 | 271 | 11 | computer | 35 | 37 | 6 | translation | 27 | 29 | 5 | visual | 16 | 17 | 4 | institute |
| 207 | 267 | 10 | analysis | 20 | 20 | 5 | 1993 | 32 | 35 | 5 | ways | 14 | 14 | 4 | interaction |
| 136 | 160 | 10 | research | 60 | 79 | 5 | ? | 47 | 57 | 5 | world | 29 | 36 | 4 | interactive |
| 125 | 146 | 10 | work | 30 | 34 | 5 | authors | 19 | 19 | 4 | 1992 | 30 | 35 | 4 | interest |
| 118 | 144 | 9 | based | 21 | 21 | 5 | checker | 24 | 26 | 4 | abstract | 14 | 14 | 4 | interested |
| 76 | 78 | 9 | electronic | 25 | 25 | 5 | chinese | 16 | 17 | 4 | accent | 13 | 13 | 4 | interpretive |
| 93 | 106 | 9 | humanities | 27 | 29 | 5 | community | 21 | 22 | 4 | aided | 19 | 21 | 4 | issues |
| 105 | 123 | 9 | project | 42 | 51 | 5 | concept | 47 | 62 | 4 | approach | 27 | 33 | 4 | just |
| 68 | 72 | 8 | conceptual | 23 | 23 | 5 | dialogue | 25 | 31 | 4 | areas | 27 | 34 | 4 | kind |
| 79 | 91 | 8 | development | 26 | 27 | 5 | dictionaries | 18 | 19 | 4 | arguments | 15 | 16 | 4 | know |
| 59 | 59 | 8 | images | 24 | 24 | 5 | digital | 20 | 21 | 4 | attention | 39 | 51 | 4 | languages |
| 142 | 186 | 8 | information | 24 | 24 | 5 | discuss | 32 | 40 | 4 | automatic | 25 | 31 | 4 | lemma |
| 199 | 276 | 8 | language | 28 | 30 | 5 | documents | 35 | 44 | 4 | basic | 39 | 51 | 4 | linguistics |
| 105 | 128 | 8 | software | 41 | 49 | 5 | edition | 31 | 40 | 4 | basis | 13 | 13 | 4 | linguists |
| 407 | 647 | 8 | text | 31 | 35 | 5 | encoding | 18 | 18 | 4 | center | 15 | 15 | 4 | male |
| 103 | 127 | 8 | time | 33 | 37 | 5 | formal | 20 | 22 | 4 | central | 17 | 18 | 4 | manuscript |
| 67 | 78 | 7 | computing | 38 | 43 | 5 | greek | 22 | 25 | 4 | changes | 86 | 133 | 4 | many |
| 56 | 60 | 7 | concepts | 75 | 108 | 5 | his | 20 | 23 | 4 | component | 56 | 83 | 4 | meaning |
| 59 | 67 | 7 | grammar | 31 | 33 | 5 | history | 16 | 17 | 4 | creation | 16 | 16 | 4 | medium |
| 48 | 48 | 7 | hypertext | 53 | 68 | 5 | important | 16 | 16 | 4 | datbank | 32 | 39 | 4 | methods |
| 73 | 85 | 7 | knowledge | 29 | 33 | 5 | including | 20 | 22 | 4 | date | 13 | 13 | 4 | metrical |
| 55 | 61 | 7 | material | 39 | 45 | 5 | interpretation | 17 | 17 | 4 | describe | 14 | 14 | 4 | mind |
| 107 | 143 | 7 | new | 97 | 143 | 5 | linguistic | 30 | 38 | 4 | differences | 19 | 19 | 4 | modeling |
| 76 | 92 | 7 | paper | 27 | 28 | 5 | links | 94 | 146 | 4 | different | 17 | 17 | 4 | multimedia |
| 38 | 38 | 7 | sgml | 27 | 30 | 5 | materials | 18 | 19 | 4 | document | 30 | 37 | 4 | names |
| 61 | 71 | 7 | style | 34 | 38 | 5 | modern | 17 | 18 | 4 | documentation | 13 | 13 | 4 | navajo |
| 186 | 279 | 7 | texts | 28 | 30 | 5 | narrative | 22 | 24 | 4 | domain | 38 | 50 | 4 | necessary |
| 66 | 78 | 7 | tools | 25 | 27 | 5 | period | 37 | 48 | 4 | elements | 46 | 62 | 4 | need |
| 151 | 217 | 7 | use | 39 | 46 | 5 | place | 19 | 20 | 4 | engineering | 15 | 16 | 4 | never |
| 94 | 121 | 7 | way | 34 | 40 | 5 | play | 20 | 21 | 4 | established | 16 | 16 | 4 | novel |
| 40 | 44 | 6 | book | 60 | 78 | 5 | present | 42 | 57 | 4 | etc | 14 | 14 | 4 | odd |
| 57 | 66 | 6 | content | 23 | 24 | 5 | presented | 23 | 26 | 4 | events | 51 | 73 | 4 | over |
| 97 | 137 | 6 | corpus | 45 | 56 | 5 | provide | 16 | 17 | 4 | examples | 70 | 105 | 4 | particular |
| 30 | 30 | 6 | errors | 32 | 34 | 5 | question | 30 | 35 | 4 | feature | 26 | 29 | 4 | people |
| 34 | 34 | 6 | hypermedia | 27 | 29 | 5 | questions | 48 | 63 | 4 | features | 17 | 18 | 4 | philosophical |
| 33 | 34 | 6 | image | 21 | 21 | 5 | responses | 13 | 13 | 4 | female | 22 | 25 | 4 | plays |
| 97 | 131 | 6 | literary | 30 | 33 | 5 | role | 17 | 18 | 4 | focus | 14 | 14 | 4 | poet |
| 35 | 35 | 6 | poetry | 31 | 34 | 5 | scientific | 87 | 132 | 4 | form | 14 | 14 | 4 | poetic |
| 85 | 109 | 6 | program | 28 | 30 | 5 | sections | 46 | 63 | 4 | forms | 15 | 15 | 4 | political |
| 40 | 43 | 6 | speakers | 34 | 39 | 5 | show | 48 | 66 | 4 | french | 13 | 13 | 4 | possibilities |
| 53 | 62 | 6 | specific | 41 | 51 | 5 | special | 57 | 78 | 4 | general | 88 | 134 | 4 | possible |
| 50 | 59 | 6 | speech | 24 | 25 | 5 | stylistic | 15 | 16 | 4 | german | 16 | 16 | 4 | presentation |
| 42 | 45 | 6 | students | 44 | 53 | 5 | support | 13 | 13 | 4 | graduate | | | | |
| 55 | 68 | 6 | studies | 36 | 41 | 5 | technology | 25 | 31 | 4 | help | | | | |

3 - On notera en particulier l'émergence d'un concept nouveau qui est absent en 1985 et qui apparaît 60 fois dans le présent colloque. C'est celui de l'hypertexte. Rares sont les communications qui n'évoquent pas cette façon libre de naviguer dans le texte et de créer des chemins virtuels que le lecteur est libre d'emprunter ou non. La chose non plus que le mot ne sont tout-à-fait nouveaux et je me souviens d'une Table ronde consacrée à ce sujet en 1988, au Colloque de Jérusalem (l'année précédente une confé-

rence sur le même sujet avait eu lieu en Caroline du Nord). Il s'agit là certainement d'une mode, qui est bâtie sur un préfixe valorisant et qui fait tache d'huile. Sur le même modèle on a construit hypermedias, Hypercard, Hypertalk et celui qui vous parle s'est servi du même moule pour son Hyperbase. C'est précisément ce logiciel Hyperbase qui a produit l'histogramme ci-dessous, consacré au préfixe hyper-.

Graphique 6. Progrès du préfixe HYPER-



Avec ce mot à la mode dit-on la même chose qu'on exprimait jadis avec les bonnes vieilles concordances? Il semble que non. L'hypertexte n'est pas seulement une technique documentaire et ne se réduit pas à l'indexation du texte intégral. Dans plusieurs des communications qu'on a entendues ces jours derniers, il semble bien qu'on ait affaire à un nouveau mode de lecture et même de création, puisque la lecture devient créative, et s'engage dans des voies virtuelles et vierges, que l'auteur même n'a pas déflorées. En certains cas limites, ce n'est plus le chemin qui est virtuel, mais le texte même, dont la rédaction est confiée à la machine. Comme l'admission d'un ordinateur-poète a paru difficile en Sorbonne, un lieu plus moderne a été trouvé à Jussieu pour accueillir, parallèlement au présent colloque, les journées d'études sur la littérature générée par ordinateur, dont il ne m'appartient pas de faire le bilan. Il ne m'appartient pas non plus de donner le droit de cité à de telles expériences. On peut les justifier, au moins pour leur vertu pédagogique et démonstrative. Quand l'émulation, à force de contraintes et de combinatoires mêlées, parvient à la qualité du pastiche, elle jette un jour nouveau et utile sur le mystère de la création.

Tableau 7. Les mots-clés des Conférences ALLC

| fréqu. | forme | réduit | fréqu. | forme | réduit | fréqu. | forme | réduit |
|--------|--------------|--------|--------|--------------|--------|--------|---------------|--------|
| 209 | textes | 202.9 | 75 | permet | 38.8 | 44 | structure | 29.9 |
| 235 | analyse | 122.1 | 30 | and | 38.4 | 1818 | des | 29.4 |
| 204 | texte | 120.4 | 40 | of | 38.1 | 62 | base | 29.0 |
| 142 | données | 95.3 | 48 | littéraires | 38.1 | 51 | règles | 29.0 |
| 80 | traitement | 75.2 | 29 | indice | 36.8 | 31 | modèles | 28.9 |
| 58 | dictionnaire | 71.1 | 117 | système | 36.0 | 68 | mode | 28.8 |
| 53 | information | 70.5 | 48 | description | 35.9 | 37 | programme | 28.3 |
| 78 | université | 64.6 | 40 | unités | 35.7 | 33 | thème | 27.6 |
| 55 | étudiants | 63.5 | 32 | outils | 35.5 | 30 | documents | 27.4 |
| 63 | méthodes | 61.9 | 67 | différents | 34.9 | 30 | interprétatio | 27.3 |
| 41 | structures | 58.3 | 53 | langues | 34.5 | 30 | permettent | 26.8 |
| 110 | recherche | 54.2 | 26 | catégories | 34.3 | 55 | domaine | 26.6 |
| 262 | mots | 52.4 | 50 | enseignement | 34.2 | 126 | cas | 26.3 |
| 41 | informations | 52.1 | 29 | syllabes | 34.2 | 44 | recherches | 26.1 |
| 35 | vocabulaire | 46.6 | 58 | résultats | 32.9 | 40 | valeurs | 25.4 |
| 145 | formes | 46.5 | 71 | type | 31.2 | 28 | distribution | 25.3 |
| 33 | commentaire | 45.3 | 38 | modes | 30.8 | 20 | outil | 25.1 |
| 36 | index | 43.8 | 27 | techniques | 30.3 | 36 | systèmes | 25.1 |
| 116 | discours | 40.5 | 57 | littéraire | 30.3 | 38 | problèmes | 24.7 |
| 63 | niveau | 39.9 | 37 | the | 30.0 | 37 | application | 24.5 |

4 - Le corpus français a été pareillement étudié, en mettant en parallèle les communications de Nice et celles de Paris. Même faveur actuelle de l'hypertexte, même poussée des sciences sociales⁴, même progrès des ambitions et des projets, même appui sur le texte. C'est d'ailleurs ce mot *texte*, au singulier comme au pluriel, qui apparaît comme la clé distinctive de notre communauté scientifique, si on prend pour toile de fond l'ensemble des textes qui ont été assemblés dans FRANTEXT et dont Jacques Dendien, créateur de cette base de données, vous a parlé, à Nice comme à Paris. Le vocabulaire spécifique apparaît dans le tableau 7, avec une évidence telle qu'on dirait une liste de mots-clés, constituée par un rapporteur intelligent. Cette liste est fort circonstanciée et comporte des centaines de mots, auxquels il faut ajouter les termes absents du modèle⁵, comme *corpus*, *ordinateur*, *informatique*, *automatique*, *linguistique*, *fréquence*, *contexte*, *occurrences*, *logiciel*, *lemmatisation*, *statistique*, *concordance*, *lexique*, etc... Qu'est-il donc besoin d'établir un bilan chiffré, puisque la machine peut le faire toute seule?

- III -

Reste à prolonger dans l'avenir la droite de tendance. Ici la courbe ne peut être qu'en pointillés.

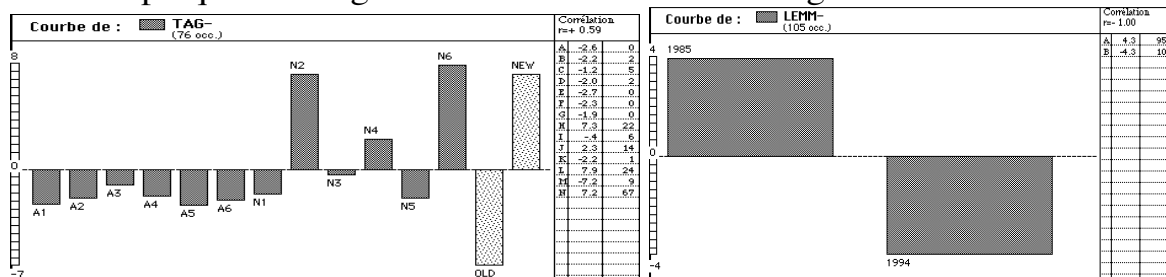
⁴ C'est l'adjectif *sociale* qui arrive en tête des spécificités du présent colloque, suivi à quelque distance du mot *politique*.

⁵ Pour des raisons d'encombrement (et aussi de légitimité statistique), le fichier de référence de notre logiciel Hyperbase ne propose à la comparaison que les 10000 formes les plus fréquentes.

1 - Ce qu'on peut supposer - et souhaiter - c'est un renforcement des normes d'enregistrement, afin qu'un objet stable et solide puisse permettre les comparaisons et être à la base de traitements ultérieurs, sans doute spécialisés et diversifiés. La recherche ne peut pas ne pas se ramifier et se particulariser. Encore faut-il qu'elle s'y emploie le plus tard possible, en utilisant une matière première indifférenciée, contrôlable, standardisée. On a beaucoup entendu parler dans cette enceinte du standard SGML (cité 39 fois) ou des recommandations du groupe TEI (évoqué 29 fois). Tant mieux. Je crains pourtant que cette norme ne se soit pas encore imposée dans les faits (ainsi on ne trouve qu'une seule mention de SGML dans les textes français et aucune de TEI). Tant pis. J'ai cherché en vain une allusion au projet Genelex qui prétend établir une norme pour les dictionnaires électroniques et éviter la multiplication des mêmes efforts et l'incompatibilité des données et des traitements dans les industries de la langue. Certains chercheurs de cette assemblée oeuvrent à cette rationalisation. Puissent-ils réussir, là où tant de tentatives ont échoué qui visaient il y a dix ou vingt ans à faire admettre les mêmes principes de lemmatisation. Cette lemmatisation aux mille têtes bourgeonnantes n'est ni morte, ni enterrée, même si le mot s'emploie moins souvent: 95 emplois pour la racine *lemm-* dans le corpus français en 1985 contre 10 seulement en 1994 (voir graphique ci-dessous). Il y a en effet des substituts qui expriment la même exigence et qui en anglais ont recours à la racine *tag-*.

Graphique 8 Progrès de TAG-

et Régression de LEMM-



Cette technique du marquage syntaxique ou sémantique des mots, souvent associée à l'intelligence artificielle, tend de plus en plus à remplir la fonction de désambiguïsation et de caractérisation qui était attachée jusqu'ici à l'opération de lemmatisation. Mais elle est plus riche en possibilités, puisque les marqueurs sont le plus souvent inscrits dans le texte même, alors que le résultat de la lemmatisation ne pouvait guère s'observer que dans des produits dérivés, index ou dictionnaires de fréquence. Un texte ou un dictionnaire ainsi enrichi se prête à des interrogations beaucoup plus fines et beaucoup plus sûres et l'on en a vu ici

Tableau 9: Concordance de electronic (78 occ.) Tri du contexte droit

| | | | |
|---------|----------------------------------|-------------|------------------------------|
| N3 331h | our manual search with the | electronic | and explore new avenues that |
| N4 391a | studies . The combination of | electronic | and printed form should in |
| N6 461d | discussions about the design of | electronic | archives have centered |
| N2 278a | that SV' s intent - - to present | electronic | artifacts for examination in |
| N1 260c | be converted into a coherent | electronic | collection with full SGML |
| N1 262e | structures , containing a) an | electronic | CONST - internal thesaurus , |
| N1 260b | Jesuit Plantation Project is the | electronic | conversion of the Maryland |
| N3 330d | much less than its distant | electronic | cousins . ¶ |
| N4 367d | have begun to consider | electronic | data handling a useful |
| N4 367c | . First attempts to use | electronic | data processing for |
| N5 439b | which will be housed in an | electronic | database , built on |
| N6 446e | . ¶There is the mushrooming of | electronic | datasets , and the |
| N1 262f | , b) an adapted version of the | electronic | dictionary ' Van Dale 12' (|
| N5 424a | the absence of a comprehensive | electronic | dictionary with accent - |
| N4 389b | of a genuinely innovative | electronic | edition by considering a |
| N3 316b | will have access to other | electronic | editions and reference works |
| N2 304e | texts , not only contemporary | electronic | editions but electronic |
| N1 263d | many ways , writing spaces are | electronic | embodiments of paper notes . |
| N5 445c | texts that we are acquiring in | electronic | form) . ¶The Center staff |
| N5 425d | work of our literature in | electronic | form . ¶ |
| N3 348b | in speech writing and the | electronic | format . At a much lower |
| N1 268c | analysis , lexicography , | electronic | glosses , or references lo |
| N1 267a | The Electronic Hamlet " ¶ " The | Electronic | Hamlet " is a much more |
| N1 268b | . ¶What is unique about " The | Electronic | Hamlet " is that it makes |
| N1 267a | Bacchae) ? " 2 . About " The | Electronic | Hamlet " ¶ " The Electronic |
| N4 387h | . 3 . In an age of increasingly | electronic | information , graduate |
| N1 266b | such as electronic linking , an | electronic | lexicon , and instantaneous |
| N6 453d | on the promised mantle of the | electronic | library . ¶ |
| N1 266b | to hypertext studies such as | electronic | linking , an electronic |
| N1 252c | presentation coincide (fax , | electronic | mail , floppy disc , |
| A3 102a | Network (JANET) to use | electronic | mail facilities to make |
| N3 353f | modality offered by the | electronic | medium destabilizes the text |
| N6 465e | face - to - face meetings in an | electronic | meeting room where each |
| N4 401g | encoded in the growing medium of | electronic | narrative ? What are the new |
| N4 401e | framework . Similarly with | electronic | narrative forms we are in |
| N2 304f | may be excerpted , saved in | electronic | notebooks , and reconfigured |
| A2 57b | by several users CARDBOX - an | electronic | personal filing cabinet with |
| N1 261d | " concepts between the | electronic | project and the course |
| N3 323b | , consists almost completely of | electronic | renditions of late - |
| N3 322e | ¶Many texts began as responsible | electronic | reproductions but were |
| A4 151d | NROFF Using - ME " . Berkeley : | Electronic | Research Laboratory |
| N2 306e | placing the new possibilities of | electronic | scholarship in the context |
| N2 301g | they can become the subjects of | electronic | search . This is done by |
| N1 274c | such index , ¶essential for any | electronic | search operations , can thus |
| N1 273e | performed either manually or by | electronic | spelling checkers , would |
| N5 410p | represented by large number of | electronic | text . ¶ |
| N3 321a | David L . Gants | Electronic | Text Center , University of |
| N5 443b | , text . ¶Seaman David M . ¶ | Electronic | Text Center , University of |
| N5 443c | University of Virginia' s | Electronic | Text Center and on - line |
| N5 445a | . ¶A principal aim of the | Electronic | Text Center has been to help |
| N6 462a | . 1 Scanning from paper The | Electronic | Text Center here at the |
| N1 260a | Plantation Project " is an | electronic | text project being conducted |
| N5 443d | from the first stage of our | electronic | text service , and will |
| N6 461d | around the connection of | electronic | text to images , and images |
| N3 323e | of print sources from which the | electronic | text will be constructed , |
| N6 461c | to be more oriented towards | electronic | texts . ¶In the first year , |
| N3 321e | for the construction of | electronic | texts and textual databases |
| N5 445a | community , and to establish | electronic | texts as a mainstream |
| N3 322a | cursory survey of the | electronic | texts available from |
| N3 321e | , and posits three classes of | electronic | texts for use in computer - |
| N4 365c | ? ¶Proposal¶The 1993 Center for | Electronic | Texts in the Humanities |
| N3 348c | to assist their study of | electronic | texts now realize that the |
| N1 255d | Text Archive , a total of 44 | electronic | texts ranging from children' |
| N3 324a | A rationale of | electronic | textual criticism must also |
| N3 321a | Virginia Toward a Rationale of | Electronic | Textual Criticism¶1 . |
| N5 415g | work . ¶Blind trust in the | electronic | tools (hardware and |
| N3 330a | s Odyssey¶With the advent of | electronic | tools , such as the |
| N3 330b | the benefits of other types of | electronic | tools that are available . |
| N2 304e | electronic editions but | electronic | transcripts of early Quarto |
| N3 315e | of cover designs . ¶An | electronic | variorum text . The texts |
| N1 261a | (2) The | electronic | version of the Maryland |
| N5 445b | that happens to work with | electronic | versions of books . From the |
| N3 323a | listings of errors found in the | electronic | versions of the Riverside |
| N6 465e | involving also interactive | electronic | whiteboards (LiveBoards) |
| N3 331g | , we were able to conduct an | electronic | word search by using the |
| N1 262c | Dpartment of Linguistics¶ | ELECTRONIC | WRITING : THE CONST - |
| N1 263e | from literal metaphor whenever | electronic | writing offers assistance |
| N1 260a | Research and Pedagogy through an | Electronic¶ | Archive Project in an |

quelques exemples fort convaincants, comme la version 2 de l'Oxford English Dictionary ou le prototype du TLF informatisé.

Ce qu'on voit se dessiner dans un proche avenir et dont les prémices sont déjà fort apparentes dans les communications entendues ces derniers jours, c'est l'invasion multiforme du texte électronique, à commencer par le mot lui-même qu'on rencontre 75 fois en 1994 contre 3 seulement en 1985. Voir tableau 9.

Tout devient électronique dans le discours et la réalité modernes: le courrier, les images, les sons, les bases de données, structurées ou non, les modes d'emploi ou manuels de référence, les livres de toute farine, la production littéraire, et l'on a même évoqué ici un Hamlet électronique, qui n'est pas robot devenu acteur, mais le recueil de toutes les éditions, de toutes les études et de toutes les représentations auxquelles la pièce a pu donner lieu. Même les conférences comme celle qui nous réunit présentement peuvent se dispenser de la présence physique et prendre la forme de Forums électroniques ou de téléconférences. La communication qui emprunte le canal électronique a en effet de grands avantages: la disponibilité, la rapidité, l'interactivité, la fiabilité, la stabilité, le faible coût et l'universalité. Ah! si Erasme avait disposé il y a 500 ans du réseau Internet! Là où des mois étaient nécessaires à l'universalisme de la Chrétienté pour aller de Bologne à Cambridge ou de Salamanque à Uppsala, une fraction de seconde suffit à l'universalisme universitaire contemporain.

2 - On n'a pas beaucoup parlé des réseaux lors de ce colloque, comme si la chose allait de soi. Ceux à qui, comme à moi, l'âge permet le recul ne peuvent pas ne pas s'étonner de cette facilité nouvelle et instantanée qui est offerte par Internet pour communiquer aux quatre coins du monde les informations, les textes, les idées, le graphisme, la voix, les outils. Je n'ai trouvé que deux allusions à Internet, deux à Janet et aucune à Transpac, aucune aux outils si commodes que sont Telnet, Fetch, Gopher et Mosaic. Sans doute le mot *électronique* résume-t-il tout cela. Sans doute aussi les littéraires, majoritaires dans cette assemblée, sont-ils moins attentifs au vecteur qu'au contenu véhiculé.

De là vient peut-être aussi le peu d'écho rencontré par un autre vecteur qui est pourtant appelé à devenir le support dominant de l'information. Je n'ai trouvé que 8 mentions ce que l'on appelle partout le CD-ROM et que la loi française recommande de désigner sous le nom de DOC (Disque Optique Compact). Voilà une invention - déjà ancienne - qui s'adapte fort bien aux données textuelles, lesquelles sont en général massives, durables et linéaires. Les premières applications de cette technologie ont pris en compte effectivement les besoins documentaires qui reposent sur le texte: brochures techniques de l'industrie, base de données bibliographiques,

encyclopédies de toutes sortes, recueil de textes juridiques, administratifs ou littéraires. On a cité déjà le CD-ROM OED2 qui sert de modèle aux dictionnaires électroniques, et on a eu droit à la démonstration de Discotext où l'on peut voir un bon exemple à suivre dans le domaine du texte intégral. Mais les obstacles au livre électronique sont loin d'être levés. Et l'on assiste dans les milieux de l'édition traditionnelle au même désarroi que l'invention de Gutenberg provoqua chez les moines copistes. On voit certains éditeurs brandir le drapeau du copyright et enrôler dans leurs légions la valetaille des auteurs, morts ou vivants, et les cohortes des ouvriers du livre. Pourquoi cette peur du CD-ROM, support économique qui offre l'immense avantage d'éviter le papier et de ménager les forêts? Il est certain que la profession éditoriale est en crise et que les expériences passées de l'audio-visuel et de la photocopieuse l'ont rendue méfiante à juste titre. Et si les éditeurs les plus puissants ont déjà franchi le pas, comme Warner, Mac Graw Hill, Wilson, Maxwell ou Hachette, chez la plupart les réticences demeurent qui tiennent à la difficulté de contrôler la diffusion. On craint le piratage, sous les formes multiples qu'il peut revêtir: par exemple la sortie sur papier des textes enregistrés sur CD-ROM, voire même une réimpression à moindre coût, sans frais de saisie. Ces craintes ne sont pas vaines, mais on peut trouver à cet égard des parades techniques plus efficaces que celles qu'on peut opposer au scanner ou à la photocopieuse, qui offrent les mêmes opportunités de piratage. Le danger le plus grand n'est pourtant pas celui-là, mais bien la possibilité de copier le CD-ROM lui-même. Il existe en effet des enregistreurs de DOC (ou CD-ROM) qu'on peut se procurer, logiciels compris, pour moins de 50000 frs et qui peuvent dupliquer les CD-ROM du commerce sur un support optique dont le prix ne dépasse pas 100 frs.

3 - Il importe grandement de résoudre au plus tôt cette difficulté et d'établir avec les éditeurs un rapport de confiance qui permette d'aborder sereinement la question épineuse du copyright. Voici, à titre d'exemple, un CD-ROM que je viens de réaliser et qui donne accès à l'oeuvre intégrale de Julien Gracq - qui pour beaucoup apparaît comme le meilleur écrivain français vivant (si Sture Allen est dans cette salle, lui qui a présidé un Colloque ALLC à Goeteborg et qui préside aussi le jury du Nobel de littérature, qu'il voie là une intéressante suggestion). Or les négociations en cours laissent en suspens sa commercialisation, malgré l'accord de principe de l'écrivain. La dernière entrevue que j'ai eue, hier, avec l'éditeur de Gracq m'a convaincu qu'on n'envisage pas cette diffusion sans inquiétude. Je propose donc qu'un groupe d'étude soit constitué qui se charge de débayer le terrain, juridique et technique, de l'édition électronique. Cette commission serait l'analogue de la commission TEI (Text Encoding Initiative) et

pourrait s'appeler TRI (Text Recording Initiative). Le cahier des charges d'une telle étude devrait s'appuyer sur l'exemple du son et de l'image, où les problèmes du copyright se sont posés pareillement, sans enrayer le développement gigantesque du marché des cassettes audio ou vidéo. La solution a été de prélever au bénéfice des éditeurs une fraction du prix des cassettes vierges, afin de compenser le manque à gagner occasionné par la copie. Une autre solution pourrait s'inspirer de l'exemple des logiciels protégés contre le piratage. Mais cette mesure frileuse - qui entre dans la spirale défavorable de l'épée et du bouclier - a été abandonnée au moins en Amérique. Car c'est en dernier ressort l'utilisateur honnête qui fait les frais de la guérilla technique que se livrent producteurs et pirates. La mission du TRI serait de trier les mesures à prendre et les garanties à fournir pour que les données textuelles puissent être diffusées sur CD-ROM, à des fins de recherche.

La situation présente me semble favorable, parce que le prix des appareils enregistreurs de CD-ROM reste encore dissuasif et que le coût des CD-ROM vierges est dix fois supérieur au prix de pressage dans les usines de production et dépasse largement le prix de vente des disques laser du monde audio. D'autre part les lecteurs de CD-ROM se multiplient et sur les machines modernes on les voit intégrés en standard aussi souvent que le lecteur de disquettes. Enfin le catalogue des titres proposés s'étend exponentiellement, comme celui de Euro-CD qui a plus de 200 pages ou celui de Kim Stacks CD-ROM qui en propose d'autres. Certains sont de qualité exceptionnelle comme le Léonard de Vinci qui vient de recevoir le prix de la profession. Quand la réalisation atteint ce niveau technique, scientifique et pédagogique, le support influe sur le message et lui donne une liberté et une efficacité que le papier ne procure pas au même degré. Et l'on devine dans les potentialités multimédia du CD-ROM les atouts qui ont manqué jusqu'ici au développement de l'Enseignement assisté par ordinateur. Une machine capable d'expliquer non seulement par l'écriture mais aussi par la parole interactive, le dessin et l'animation, devient un substitut intéressant de l'enseignant. Son seul défaut - qui tourne parfois à l'avantage - est d'être un substitut de l'enseignant et d'exclure la présence physique et l'autorité morale de celui-ci. Voilà pourquoi tant de titres proposés sur le marché du CD-ROM appartiennent au monde de l'éducation: encyclopédies diverses, cours de langue, albums d'images ou de films et même recueil de textes. Parmi ces textes on note les Classiques de l'antiquité, la Bible, Shakespeare, Dickens, les soeurs Bronte, Jane Austen, Lewis Carroll, Oscar Wilde, Melville, Flaubert, Proust.. Hélas tous les titres retenus appartiennent au domaine public et le choix est fort restreint. Il est temps d'élargir l'éventail, sans ouvrir pour autant la boîte de Pandore.

Quel guide dans la négociation? Rien d'autre que la confiance et la loyauté. Et c'est par une leçon de morale que je voudrais terminer ce bilan. On ne peut gagner la bataille du copyright qu'en s'engageant à s'abstenir de procéder à toute copie et à réserver l'usage du CD-ROM à la recherche scientifique. Une simple garantie juridique - et non le caviardage systématique - doit garantir les droits de l'éditeur, pourvu que la preuve soit aisée à produire en cas de violation. L'expérience du logiciel non-protégé qui a réussi en Amérique et qui s'impose petit à petit en Europe invite à penser que c'est là la voie de la morale éclairée et de l'efficacité vertueuse.

Les deux associations qui nous réunissent aujourd'hui ont établi, dans leur titre et leurs objectifs, le lien entre l'ordinateur et les humanités. Il ne s'agit pas seulement d'introduire l'informatique dans les humanités mais aussi, inversement, d'introduire un peu de morale et d'humanité dans l'ordinateur. C'est d'ailleurs ici, en Sorbonne, que l'ordinateur a reçu son nom de baptême français, le latiniste Jacques Perret qui enseignait en ces lieux ayant été invité à donner l'équivalent du computer anglais. Entre ordonner et compter, je vois déjà un progrès dans l'ordre de l'humain. Mais il y a beaucoup à faire encore, pour conjurer les deux dangers qui guettent les humanités: l'exclusion stupide de l'ordinateur ou une sottise allégerie à son endroit. Les zéloteurs de l'informatique ne sont pas légion en Sorbonne, même si quelques thèses - et même des colloques - s'y sont développées dans les disciplines que nous représentons. Ici domine la méfiance devant la science sans conscience, l'enseignement sans professeur, l'intelligence sans cerveau, la virtualité sans réalité, et la terre sans homme. Mais cette inquiétude est rassurante: pour convaincre les sceptiques il suffit de montrer que l'homme anime la machine et que la main tient l'outil. On peut craindre davantage le danger opposé, qui donne tout pouvoir à la machine et à ceux qui l'utilisent non comme un outil, mais comme un paravent pour cacher des opérations incertaines ou frauduleuses. Jusqu'ici dans le domaine littéraire, l'imposture était difficile, le chercheur ayant l'obligation de citer ses sources et de permettre le contrôle des citations. Nul écran entre le texte interprété et le commentaire de l'interprète. À l'inverse les résultats qui empruntent le canal de l'ordinateur ont tendance à s'imposer à l'esprit du lecteur, et parfois aussi de l'auteur, comme parole d'évangile. Et cet argument d'autorité est détestable comme il l'était en Sorbonne au temps de Rabelais. Encore les Sorbonnicoles de la scholastique pouvaient-ils appuyer sinon justifier leur jugement à grand renfort de références. Pas de Bible au contraire dans la machine, pas de papes, pas de bulles, pas de conciles, pas de conciliabules. Mais parfois des jugements abrupts et péremptifs auxquels on est invité à souscrire. Pour qu'on puisse ajouter foi aux jugements de la machine, il faut délivrer les attendus des jugements

proférés et laisser une trace visible de la logique mise en oeuvre et des opérations effectuées, sans quoi s'installe le règne de l'arbitraire et de l'incontrôlable. Malheureusement les données sont souvent si vastes et les traitements si complexes que la place manque pour les mettre sous les yeux du lecteur. On en est donc souvent réduit à un contrat de confiance entre l'auteur et le lecteur. Et une fois de plus la morale s'impose comme le garant ultime.

Dans le domaine qui est le nôtre, le développement des moyens entraîne le renforcement de la règle morale et de la rigueur scientifique. C'est le cas aussi de quelques autres sciences, comme la physique nucléaire ou la biologie génétique. Et pour finir je m'en tiendrai à trois prescriptions empruntées à la page 185 du livret où Wilhelm Ott nous livre le fruit de sa longue et impeccable expérience dans le domaine du traitement automatique du texte:

article 1 : maîtriser a priori les méthodes et les étapes qui conduisent aux résultats

article 2 : contrôler a posteriori la validité de tels résultats

article 3 : vérifier que de tels résultats se produisent en toutes circonstances

C'est beaucoup demander au chercheur, qui doit tout ensemble garantir les données, les méthodes, les traitements et les conclusions et cela ressemble assez à la démarche cartésienne du *Discours de la méthode*. Je suis sûr que les exposés qui ont eu lieu dans la salle Descartes ont ressenti plus que d'autres cette exigence éthique. Je ne chercherai pas à démêler si la morale de Descartes est provisoirement définitive ou définitivement provisoire. Mais ce qui s'impose pareillement, dans les deux mondes auxquels notre discipline est confrontée et où Descartes s'est illustré, celui de la pensée et celui de la matière, c'est la rigueur morale et scientifique.

Voilà terminée ma leçon. Merci pour votre patience. On me demandait un bilan et, comme il arrive souvent dans les discours des personnes d'un certain âge, le bilan a tourné en leçon de morale. Mais en français bilan se dit aussi rapport moral.