

Ceci n'est pas une pipe

Lou Burnard





La modélisation : clef des "humanités numériques"



avenue centrale

rendez-vous en sciences humaines

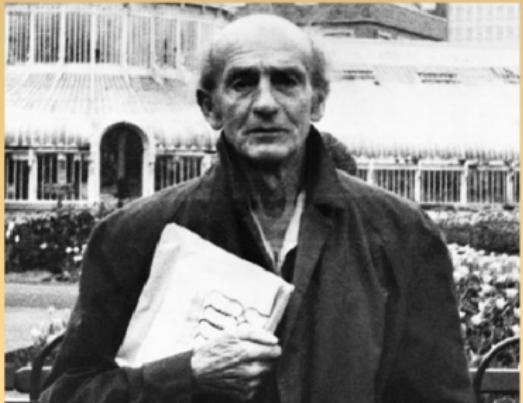
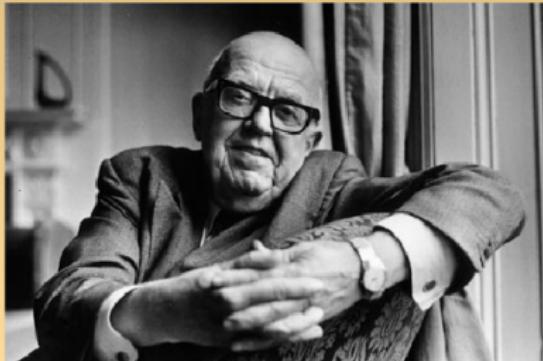
- Quelques oppositions
- De *Literary and Linguistic Computing* à *Humanities Computing* à *Digital Humanities*
- L'avenir de l'aventure DH



Quelques oppositions

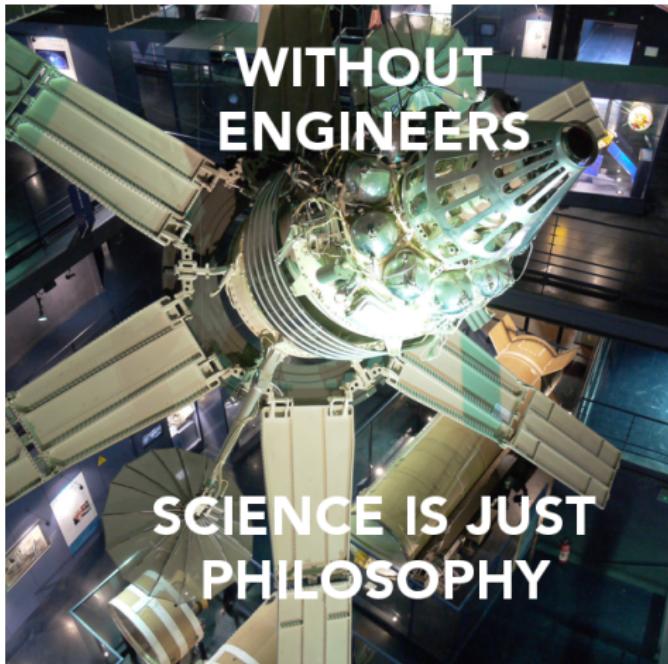
- ① sciences humaines, sciences dures
- ② disciplines scientifiques, ingénierie
- ③ textes, données
- ④ lecture rapprochée, lecture distante

Sciences humaines vs: sciences dures



- La grande querelle "deux cultures" C.P.Snow vs F.R. Leavis (1960)
- Un "rematch" d'un débat victorien sur l'importance de la culture à la société

Science vs. ingénierie



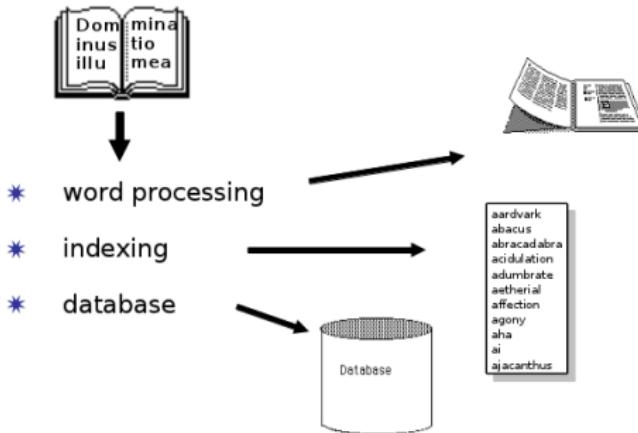
'Sans les ingénieurs, la science ne serait que de la philosophie'

Textes numériques vs. données numériques

- Le traitement informatisé des données concerne les chiffres, les quantités, les tendances statistiques...
- Le traitement informatisé des textes concerne les mots, l'écriture, la langue...
- L'informatique a donc systématiquement opposé les "données" aux "textes"
 - en traitant les textes comme si elles étaient des données
 - en traitant les données comme si elles n'étaient pas des textes

(cf Burnard, 1984)

Three kinds of software



Nous possédons actuellement des technologies (par ex XML) qui permettent de franchir toutes ces frontières

Lecture rapprochée vs lecture à distance

- La tradition d'**explication de texte** (Lanson) valorise surtout une attention focalisée sur les mots, le contexte, les implications d'un nombre restreint de document/texte/oeuvres
- Elle est actuellement souvent opposée aux méthodes statistiques visant la contrôle et la compréhension des milliers voire des millions de documents numérisés : "**Distant Reading**" (Moretti)
- Mais les deux sont complémentaires : la lecture rapprochée serait enrichie par une compréhension du contexte vu de loin

Lecture des eBooks

```
div rend="linenumber" xml:id="A4.1">
<head>Beowulf</head>
<bibl>Dobbie, 1953 3-98; Dobbie, E.V.K., Beowulf and Judith, A
<l>Hwæt! We Gardena <caesura/> in geardagum, </l>
<l>þeodcynninga, <caesura/> prym gefrunon, </l>
<l>hu ða æbelinges <caesura/> ellen fremdon. </l>
<l>Oft Scyld Scfeing <caesura/>
<l><unclear>sceapena</unclear> preatum, </l>
<l>monegum magnum, <caesura/> meodosetla ofteah, </l>
<l><clear>orlas</unclear>, <caesura/> Syððan arest v
<l>feascaft funden, <caesura/> he þas frofre gebad, </l>
<l>weox under wolcnum, <caesura/> weorðmyndum þah, </l>
<l>oðþbat him aghwylc <caesura/> para ymbssittendra </l>
<l>ofer hronrade <caesura/> hyran scolde, </l>
<l>gomban gyldan. <caesura/> þet was god cyning! </l>
<l>ðæm eafera was <caesura/> after cenned, </l>
<l>geong in geardum, <caesura/> pone god sende </l>
<l>folce to frofre; <caesura/> fyrendearfe ongeat </l>
<l>þe hie ar drugon <caesura/>
<l><unclear>aldorlease</unclear>
```

137. Beowulf

Dobbie, 1953 3-98; Dobbie, E.V.K., Beowulf and Judith, ASPR 4 (New York).

Hwæt! We Gardena in geardagum,
þeodcynninga, prym gefrunon,
hu ða æbelinges ellen fremdon.
Oft Scyld Scfeing sceapena preatum,
monegum magnum, meodosetla ofteah,
egode orlas. Syððan arest wearð
feascaft funden, he þas frofre gebad,
weox under wolcnum, weorðmyndum þah,
oðþbat him aghwylc para ymbssittendra
ofer hronrade hyran scolde,
gomban gyldan. þet was god cyning!
ðæm eafera was after cenned,
geong in geardum, pone god sende

- Les textes numériques se présentent comme des livres imprimés... mais on ne doit pas se laisser séduire par les métaphores !
- Est-ce qu'on numérise les textes juste pour le plaisir de les distribuer aux lecteurs ?
- Qu'est ce qu'on pourrait en faire d'avantage?.

Le numérique reste incontournable ...

- Même les objets de recherche des SHS sont devenus numériques
- Les méthodes des SHS ne peuvent pas échapper à l'environnement technologique qui nous englobe tou(te)s
- Une transition du “web de documents” vers le “web de données” s’effectue
- Les questions politiques et culturelles restent, mais leur contexte évolue dans un monde de plus en plus “ouvert”
- Comment sommes-nous arrivés à ce point? voici un peu de l’histoire...

Literary & Linguistic Computing



1950-1980

- L'ordinateur est une immense machine partagée, gérée par des ingénieurs sérieux en blouse blanche
- On y lance ses travaux en **batch** ... ils sont transmis à une **file d'attente** puis **exécutés** pour produire des **sorties** imprimées sur papier
- Qu'est-ce qu'on peut faire avec?

Quelques pionniers

- 1949-56: Initiation des travaux de Roberto Busa SJ sur l'*Index Thomisticus*, une investigation du lexique d'Aquin sur cartes perforées, édités en 1975 *Thomae Aquinatis Opera Omnia cum hypertextibus* en CDROM; actuellement disponible à <http://www.corpusthomisticum.org/>
- 1967 : Kucera et Francis produisent le *Computational Analysis of Present-Day American English* dérivé d'un corpus d'un million de mots
- 1968: Maurice Tournier et d'autres à Saint Cloud travail sur l'analyse statistique des documents politiques (*Les mots de mai 68*)
- 1970: Trevor Howard-Hill produit des concordances des quartos de Shakespeare pour soutenir une thèse doctorale sur la production du First Folio

LLC: ce qui compte, c'est de compter

- Les objets et les résultats principales sont
 - Les concordances (un objet en lui même)
 - Des statistiques riches et complexes
- Il y a une hypothèse plus ou moins explicite que le "style" ou "registre" seraient comme l'auctorialité identifiable de manière statistique
- Mais les modèles statistiques sousjacentes de la fonctionnement de la langue n'ont guère avancé depuis le 19ème siècle

Ces méthodes ont évolué et sont toujours vivant et bien vivant en France ...

- Analyse factorielle
- Fouille de données (text mining)
- <http://textometrie.ens-lyon.fr>



... et en train d'être réinventés aux Etats Unis

L'importance de ne pas lire

- “What can you do with a million books?” (Greg Crane)
- Une synthèse de méthodes :
 - linguistique de corpus
 - statistiques de co-occurrence
 - topic Modelling
 - visualisation
- ou une réappropriation des techniques anciens ?



Humanities Computing



Années 80 : décennie d'une foi illimitée dans les technologies

- Les ressources et méthodes numériques bien que perçues comme étranges et difficiles s'installent dans les SHS
- Financement public important pour des activités d'infrastructure : à la fois national et européen : au Royaume Uni par ex le *Computers in Teaching Initiative* et le *Arts and Humanities Data Service*
- Une nouveauté ou une amélioration du passé?
- L'arrivée du centre HC universitaire, et forcément des cours

The image consists of two parts. On the left, a photograph of a young man with glasses, wearing a green t-shirt, sitting at a desk and reading a book. To his left is a stack of books. Above him is a quote: "I used to read texts. Now I am learning the tools to play with them." Below the photo is his name, ANTHONY English. On the right is a promotional graphic for the M.A. in Humanities Computing at the University of Alberta. It features the text "M.A. IN HUMANITIES COMPUTING AT THE UNIVERSITY OF ALBERTA" above a blue stylized logo. Below the logo is a detailed description of the program: "The Masters of Arts in Humanities Computing is an interdisciplinary programme of the Faculty of Arts at the University of Alberta. The program integrates computing methods and theories with research and teaching in the Humanities. It addresses the demand for Arts graduates proficient in computing skills, able to work either in the realm of humanities research and teaching or in the emerging job markets of information management and content delivery over the Internet."

Apparition de la *quellenkritik*

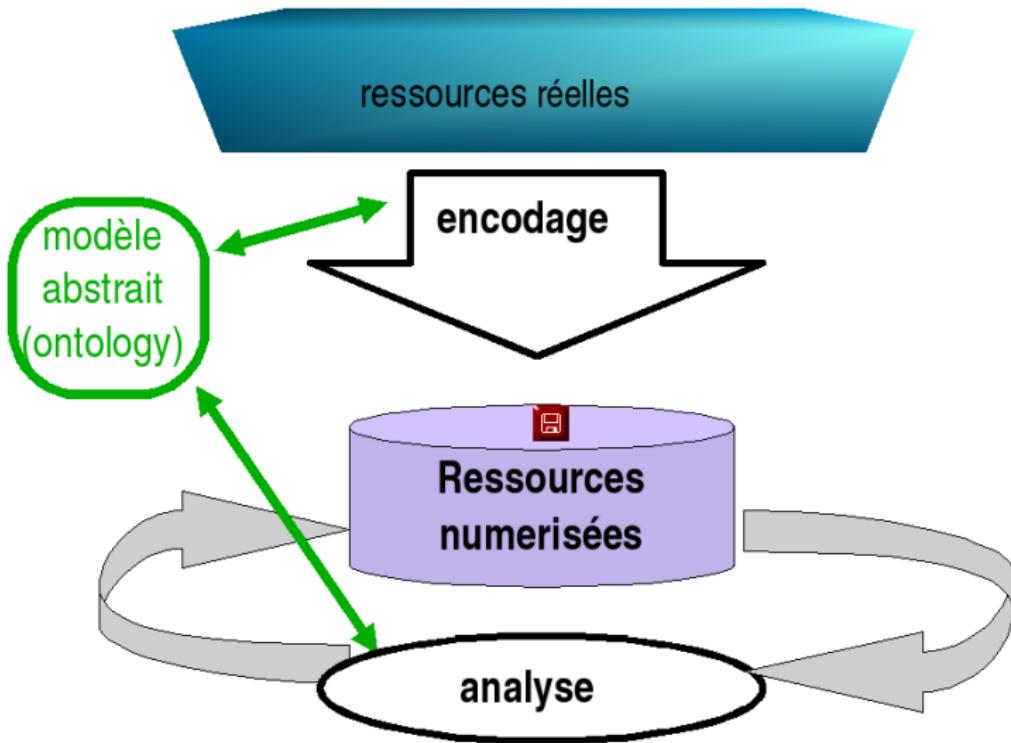
- En France, J-P Genet et d'autres proposent que les données historiques une fois numérisées pourraient servir à enrichir une analyse
- Encore systematisée en Allemagne par Manfred Thaller avec le logiciel kleio, un sgbd textuel avant la lettre
- Thaller (en 1989) lance un défi: si HC serait une discipline scientifique, où est sa théorie sousjacente ?

Théorification de HC

- Quels principes peut on identifier pour permettre la mise en relation des outils employés par le HC ?
- Unsworth (entre autres) évoque les traditions scientifiques ("scholarly primitives")
 - **La recherche** effectuée selon des traits externes
 - **L'analyse** selon des traits internes
 - **Les associations** selon des perceptions partagées
- Mais c'est surtout la **modelisation des données textuelles** qui caractérise le DH

Ceci n'est pas un arbre





Les inconvénients des ressources numériques (circa 1989)

- Elles sont difficiles à trouver : ni disponibles en BU, ni mentionnées dans les revues
- Elles ne sont pas cataloguées de manière cohérente, ou pas du tout
- Elles ne sont pas fiables
- Elles émergent d'un contexte inconnu, pas (toujours) très scientifique
- Elles disparaissent, se transforment, ou deviennent inutilisables sans préavis
- Elles ne fonctionnent pas : ou il faut bien choisir son ordinateur pour les faire fonctionner
- Et surtout...

Elles utilisent vraiment trop de formats d'encodage mutuellement incompréhensibles !!



L'effet Babel

La numérisation fige une interprétation quelconque – une lecture – d'un documents

Bien sûr il existe plusieurs lectures possibles pour la plupart des documents...

I

Loomings

Call me Ishmael. Some years ago – never mind how long precisely – having little or no money in my purse,
... et (malheureusement) plusieurs manières d'exprimer ces lectures!

Encodage ou babel?

```
|chap1
<C 1> Loomings
\chapter[1]{Loomings}
:h1.1. Loomings
MOBY001001LOOMINGS
|c1
.chapter Loomings
.cp;.sp 6 a;.ce .bd 1. Loomings
<h1>Loomings</h1>
<p class="h1">Loomings
~x
```

- Bonne nouvelle : il existe des logiciels capables de traduire entre 500 formats divers
- Mauvaise nouvelle : on en a besoin



Encodage ou babel?

```
|chap1
<C 1> Loomings
\chapter[1]{Loomings}
:h1.1. Loomings
MOBY001001LOOMINGS
|c1
.chapter Loomings
.cp;.sp 6 a;.ce .bd 1. Loomings
<h1>Loomings</h1>
<p class="h1">Loomings
~x
```

- Bonne nouvelle : il existe des logiciels capables de traduire entre 500 formats divers
- Mauvaise nouvelle : on en a besoin

Encodage ou babel?

```
|chap1
<C 1> Loomings
\chapter[1]{Loomings}
:h1.1. Loomings
MOBY001001LOOMINGS
|c1
.chapter Loomings
.cp;.sp 6 a;.ce .bd 1. Loomings
<h1>Loomings</h1>
<p class="h1">Loomings
~x
```

- Bonne nouvelle : il existe des logiciels capables de traduire entre 500 formats divers
- Mauvaise nouvelle : on en a besoin

La science repose sur une continuité des connaissances

- Conserver les “bytes” d’un encodage ne suffit pas : il faut aussi une continuité de compréhension
- L’encodage devrait être auto-descriptif
- L’archive devient un vecteur pour transmettre nos interprétations

D'où l' importance de la TEI (Text Encoding Initiative)

<http://www.tei-c.org>



TEI: le produit le plus significatif de HC?

- D'origine une réponse aux problèmes posés par l'incohérence des formats et le manque des standards numériques
- La TEI est devenue un seul modèle encyclopédique des "particularités significatives" des ressources textuelles
- Et une infrastructure capable de répondre aux besoins et priorités évolutifs de tous les tribus de la communauté scientifique



Par exemple :

AN ANSWER TO WILLI ALABLAST his MOTIVES.

By ROGER FENTON F
Grays Inne.



Imprinted
W.Af
y2

a Day.	a Week.	a Month.	a Year.
1 d.	2 d.	13 d.	365 d.
2 d.	3 d.	26 d.	730 d.
3 d.	4 d.	39 d.	1095 d.
4 d.	5 d.	52 d.	1460 d.
5 d.	6 d.	65 d.	1825 d.
6 d.	7 d.	78 d.	2190 d.
7 d.	8 d.	91 d.	2555 d.
8 d.	9 d.	104 d.	2920 d.
9 d.	10 d.	117 d.	3285 d.
10 d.	11 d.	130 d.	3650 d.
11 d.	12 d.	143 d.	4015 d.
12 d.	13 d.	156 d.	4380 d.
13 d.	14 d.	169 d.	4745 d.
14 d.	15 d.	182 d.	5110 d.
15 d.	16 d.	195 d.	5475 d.
16 d.	17 d.	208 d.	5840 d.
17 d.	18 d.	221 d.	6205 d.
18 d.	19 d.	234 d.	6570 d.
19 d.	20 d.	247 d.	6935 d.
20 d.	21 d.	260 d.	7300 d.
21 d.	22 d.	273 d.	7665 d.
22 d.	23 d.	286 d.	8030 d.
23 d.	24 d.	299 d.	8395 d.
24 d.	25 d.	312 d.	8760 d.
25 d.	26 d.	325 d.	9125 d.
26 d.	27 d.	338 d.	9490 d.
27 d.	28 d.	351 d.	9855 d.
28 d.	29 d.	364 d.	10220 d.
29 d.	30 d.	377 d.	10585 d.
30 d.	31 d.	390 d.	10950 d.
31 d.	1 d.	403 d.	11315 d.
1 d.	2 d.	416 d.	11680 d.
2 d.	3 d.	429 d.	12045 d.
3 d.	4 d.	442 d.	12410 d.
4 d.	5 d.	455 d.	12775 d.
5 d.	6 d.	468 d.	13140 d.
6 d.	7 d.	481 d.	13505 d.
7 d.	8 d.	494 d.	13870 d.
8 d.	9 d.	507 d.	14235 d.
9 d.	10 d.	520 d.	14600 d.
10 d.	11 d.	533 d.	14965 d.
11 d.	12 d.	546 d.	15330 d.
12 d.	13 d.	559 d.	15695 d.
13 d.	14 d.	572 d.	16060 d.
14 d.	15 d.	585 d.	16425 d.
15 d.	16 d.	598 d.	16790 d.
16 d.	17 d.	611 d.	17155 d.
17 d.	18 d.	624 d.	17520 d.
18 d.	19 d.	637 d.	17885 d.
19 d.	20 d.	650 d.	18250 d.
20 d.	21 d.	663 d.	18615 d.
21 d.	22 d.	676 d.	18980 d.
22 d.	23 d.	689 d.	19345 d.
23 d.	24 d.	702 d.	19710 d.
24 d.	25 d.	715 d.	20075 d.
25 d.	26 d.	728 d.	20440 d.
26 d.	27 d.	741 d.	20805 d.
27 d.	28 d.	754 d.	21170 d.
28 d.	29 d.	767 d.	21535 d.
29 d.	30 d.	780 d.	21900 d.
30 d.	31 d.	793 d.	22265 d.
31 d.	1 d.	806 d.	22630 d.
1 d.	2 d.	819 d.	23005 d.
2 d.	3 d.	832 d.	23370 d.
3 d.	4 d.	845 d.	23735 d.
4 d.	5 d.	858 d.	24100 d.
5 d.	6 d.	871 d.	24465 d.
6 d.	7 d.	884 d.	24830 d.
7 d.	8 d.	897 d.	25195 d.
8 d.	9 d.	910 d.	25560 d.
9 d.	10 d.	923 d.	25925 d.
10 d.	11 d.	936 d.	26290 d.
11 d.	12 d.	949 d.	26655 d.
12 d.	13 d.	962 d.	27020 d.
13 d.	14 d.	975 d.	27385 d.
14 d.	15 d.	988 d.	27750 d.
15 d.	16 d.	1001 d.	28115 d.
16 d.	17 d.	1014 d.	28480 d.
17 d.	18 d.	1027 d.	28845 d.
18 d.	19 d.	1040 d.	29210 d.
19 d.	20 d.	1053 d.	29575 d.
20 d.	21 d.	1066 d.	29940 d.
21 d.	22 d.	1079 d.	30305 d.
22 d.	23 d.	1092 d.	30670 d.
23 d.	24 d.	1105 d.	31035 d.
24 d.	25 d.	1118 d.	31400 d.
25 d.	26 d.	1131 d.	31765 d.
26 d.	27 d.	1144 d.	32130 d.
27 d.	28 d.	1157 d.	32495 d.
28 d.	29 d.	1170 d.	32860 d.
29 d.	30 d.	1183 d.	33225 d.
30 d.	31 d.	1196 d.	33590 d.
31 d.	1 d.	1209 d.	33955 d.
1 d.	2 d.	1222 d.	34320 d.
2 d.	3 d.	1235 d.	34685 d.
3 d.	4 d.	1248 d.	35050 d.
4 d.	5 d.	1261 d.	35415 d.
5 d.	6 d.	1274 d.	35780 d.
6 d.	7 d.	1287 d.	36145 d.
7 d.	8 d.	1300 d.	36510 d.
8 d.	9 d.	1313 d.	36875 d.
9 d.	10 d.	1326 d.	37240 d.
10 d.	11 d.	1339 d.	37605 d.
11 d.	12 d.	1352 d.	37970 d.
12 d.	13 d.	1365 d.	38335 d.
13 d.	14 d.	1378 d.	38700 d.
14 d.	15 d.	1391 d.	39065 d.
15 d.	16 d.	1404 d.	39430 d.
16 d.	17 d.	1417 d.	39795 d.
17 d.	18 d.	1430 d.	40160 d.
18 d.	19 d.	1443 d.	40525 d.
19 d.	20 d.	1456 d.	40890 d.
20 d.	21 d.	1469 d.	41255 d.
21 d.	22 d.	1482 d.	41620 d.
22 d.	23 d.	1495 d.	41985 d.
23 d.	24 d.	1508 d.	42350 d.
24 d.	25 d.	1521 d.	42715 d.
25 d.	26 d.	1534 d.	43080 d.
26 d.	27 d.	1547 d.	43445 d.
27 d.	28 d.	1560 d.	43810 d.
28 d.	29 d.	1573 d.	44175 d.
29 d.	30 d.	1586 d.	44540 d.
30 d.	31 d.	1599 d.	44905 d.
31 d.	1 d.	1612 d.	45270 d.
1 d.	2 d.	1625 d.	45635 d.
2 d.	3 d.	1638 d.	46000 d.
3 d.	4 d.	1651 d.	46365 d.
4 d.	5 d.	1664 d.	46730 d.
5 d.	6 d.	1677 d.	47095 d.
6 d.	7 d.	1690 d.	47460 d.
7 d.	8 d.	1703 d.	47825 d.
8 d.	9 d.	1716 d.	48190 d.
9 d.	10 d.	1729 d.	48555 d.
10 d.	11 d.	1742 d.	48920 d.
11 d.	12 d.	1755 d.	49285 d.
12 d.	13 d.	1768 d.	49650 d.
13 d.	14 d.	1781 d.	50015 d.
14 d.	15 d.	1794 d.	50380 d.
15 d.	16 d.	1807 d.	50745 d.
16 d.	17 d.	1820 d.	51110 d.
17 d.	18 d.	1833 d.	51475 d.
18 d.	19 d.	1846 d.	51840 d.
19 d.	20 d.	1859 d.	52205 d.
20 d.	21 d.	1872 d.	52570 d.
21 d.	22 d.	1885 d.	52935 d.
22 d.	23 d.	1898 d.	53300 d.
23 d.	24 d.	1911 d.	53665 d.
24 d.	25 d.	1924 d.	54030 d.
25 d.	26 d.	1937 d.	54395 d.
26 d.	27 d.	1950 d.	54760 d.
27 d.	28 d.	1963 d.	55125 d.
28 d.	29 d.	1976 d.	55490 d.
29 d.	30 d.	1989 d.	55855 d.
30 d.	31 d.	1999 d.	56220 d.
31 d.	1 d.	2012 d.	56585 d.
1 d.	2 d.	2045 d.	56950 d.
2 d.	3 d.	2078 d.	57315 d.
3 d.	4 d.	2111 d.	57680 d.
4 d.	5 d.	2144 d.	58045 d.
5 d.	6 d.	2177 d.	58410 d.
6 d.	7 d.	2210 d.	58775 d.
7 d.	8 d.	2243 d.	59140 d.
8 d.	9 d.	2276 d.	59505 d.
9 d.	10 d.	2309 d.	59870 d.
10 d.	11 d.	2342 d.	60235 d.
11 d.	12 d.	2375 d.	60600 d.
12 d.	13 d.	2408 d.	60965 d.
13 d.	14 d.	2441 d.	61330 d.
14 d.	15 d.	2474 d.	61695 d.
15 d.	16 d.	2507 d.	62060 d.
16 d.	17 d.	2540 d.	62425 d.
17 d.	18 d.	2573 d.	62790 d.
18 d.	19 d.	2606 d.	63155 d.
19 d.	20 d.	2639 d.	63520 d.
20 d.	21 d.	2672 d.	63885 d.
21 d.	22 d.	2705 d.	64250 d.
22 d.	23 d.	2738 d.	64615 d.
23 d.	24 d.	2771 d.	64980 d.
24 d.	25 d.	2804 d.	65345 d.
25 d.	26 d.	2837 d.	65710 d.
26 d.	27 d.	2870 d.	66075 d.
27 d.	28 d.	2903 d.	66440 d.
28 d.	29 d.	2936 d.	66805 d.
29 d.	30 d.	2969 d.	67170 d.
30 d.	31 d.	3002 d.	67535 d.
31 d.	1 d.	3035 d.	67900 d.
1 d.	2 d.	3068 d.	68265 d.
2 d.	3 d.	3101 d.	68630 d.
3 d.	4 d.	3134 d.	69005 d.
4 d.	5 d.	3167 d.	69370 d.
5 d.	6 d.	3200 d.	69735 d.
6 d.	7 d.	3233 d.	70100 d.
7 d.	8 d.	3266 d.	70465 d.
8 d.	9 d.	3300 d.	70830 d.
9 d.	10 d.	3333 d.	71195 d.
10 d.	11 d.	3366 d.	71560 d.
11 d.	12 d.	3400 d.	71925 d.
12 d.	13 d.	3433 d.	72290 d.
13 d.	14 d.	3466 d.	72655 d.
14 d.	15 d.	3500 d.	73020 d.
15 d.	16 d.	3533 d.	73385 d.
16 d.	17 d.	3566 d.	73750 d.
17 d.	18 d.	3600 d.	74115 d.
18 d.	19 d.	3633 d.	74480 d.
19 d.	20 d.	3666 d.	74845 d.
20 d.	21 d.	3700 d.	75210 d.
21 d.	22 d.	3733 d.	75575 d.
22 d.	23 d.	3766 d.	75940 d.
23 d.	24 d.	3800 d.	76305 d.
24 d.	25 d.	3833 d.	76670 d.
25 d.	26 d.	3866 d.	77035 d.
26 d.	27 d.	3900 d.	77400 d.
27 d.	28 d.	3933 d.	77765 d.
28 d.	29 d.	3966 d.	78130 d.
29 d.	30 d.	4000 d.	78495 d.
30 d.	31 d.	4033 d.	78860 d.
31 d.	1 d.	4066 d.	79225 d.
1 d.	2 d.	4100 d.	79590 d.
2 d.	3 d.	4133 d.	79955 d.
3 d.	4 d.	4166 d.	80320 d.
4 d.	5 d.	4200 d.	80685 d.
5 d.	6 d.	4233 d.	81050 d.
6 d.	7 d.	4266 d.	81415 d.
7 d.	8 d.	4300 d.	81780 d.
8 d.	9 d.	4333 d.	82145 d.
9 d.	10 d.	4366 d.	82510 d.
10 d.	11 d.	4400 d.	82875 d.
11 d.	12 d.	4433 d.	83240 d.
12 d.	13 d.	4466 d.	83605 d.
13 d.	14 d.	4500 d.	83970 d.
14 d.	15 d.	4533 d.	84335 d.
15 d.	16 d.	4566 d.	84700 d.
16 d.	17 d.	4600 d.	85065 d.
17 d.	18 d.	4633 d.	85430 d.
18 d.	19 d.	4666 d.	85795 d.
19 d.	20 d.	4700 d.	86160 d.
20 d.	21 d.	4733 d.	86525 d.
21 d.	22 d.	4766 d.	86890 d.
22 d.	23 d.	4800 d.	87255 d.
23 d.	24 d.	4833 d.	87620 d.
24 d.	25 d.	4866 d.	88085 d.
25 d.	26 d.	4900 d.	88450 d.
26 d.	27 d.	4933 d.	88815 d.
27 d.	28 d.	4966 d.	89180 d.
28 d.	29 d.	5000 d.	89545 d.
29 d.	30 d.	5033 d.	89910 d.
30 d.	31 d.	5066 d.	90275 d.
31 d.	1 d.	5100 d.	90640 d.
1 d.	2 d.	5133 d.	91005 d.
2 d.	3 d.	5166 d.	91370 d.
3 d.	4 d.	5200 d.	91735 d.
4 d.	5 d.	5233 d.	92100 d.
5 d.	6 d.	5266 d.	92465 d.
6 d.	7 d.	5300 d.	92830 d.
7 d.	8 d.	5333 d.	93195 d.
8 d.	9 d.	5366 d.	93560 d.
9 d.	10 d.	5400 d.	93925 d.
10 d.	11 d.	5433 d.	94290 d.
11 d.	12 d.	5466 d.	94655 d.
12 d.	13 d.	5500 d.	95020 d.
13 d.	14 d.	5533 d.	95385 d.
14 d.	15 d.	5566 d.	95750 d.
15 d.	16 d.	5600 d.	96115 d.
16 d.	17 d.	5633 d.	96480 d.
17 d.	18 d.	5666 d.	96845 d.
18 d.	19 d.	5700 d.	97210 d.
19 d.	20 d.	5733 d.	97575 d.
20 d.	21 d.	5766 d.	97940 d.
21 d.	22 d.	5800 d.	98305 d.
22 d.	23 d.	5833 d.	98670 d.
23 d.	24 d.	5866 d.	99035 d.
24 d.	25 d.	5900 d.	99400 d.
25 d.	26 d.	5933 d.	99765 d.
26 d.	27 d.	5966 d.	100130 d.
27 d.	28 d.	6000 d.	100495 d.
28 d.	29 d.	6033 d.	100860 d.
29 d.	30 d.	6066 d.	101225 d.
30 d.	31 d.	6100 d.	101590 d.
31 d.	1 d.	6133 d.	101955 d.
1 d.	2 d.	6166 d.	102320 d.
2 d.	3 d.	6200 d.	102685 d.
3 d.	4 d.	6233 d.	103050 d.
4 d.	5 d.	6266 d.	103415 d.
5 d.	6 d.	6300 d.	103780 d.
6 d.	7 d.	6333 d.	104145 d.
7 d.	8 d.	6366 d.	104510 d.
8 d.	9 d.	6400 d.	104875 d.
9 d.	10 d.	6433 d.	105240 d.
10 d.	11 d.	6466 d.	105605 d.
11 d.	12 d.	6500 d.	105970 d.
12 d.	13 d.	6533 d.	106335 d.
13 d.	14 d.	6566 d.	106690 d.
14 d.	15 d.	6600 d.	107055 d.
15 d.	16 d.	6633 d.	107420 d.
16 d.	17 d.	6666 d.	107785 d.
17 d.	18 d.	6700 d.	

Par exemple :

AN

<line>With a thousand <del rend="stroked">fears

<del rend="stroked">

<add place="above">

<unclear>ways</unclear>

</add>

 his MOTIVES.

<add place="above_above">

By R <ptr target="#alt4"/>

</add> that <del rend="stroked">creature

<add place="above">vision</add> s' face was grained; </line>

<line>Yet no blood <del rend="stroked">sumped

<add place="above">reached <del rend="stroked">him

</add>

<del rend="stroked">here fly comp.

<add place="above">there</add> from the upper ground, </line>

<line>And no <del rend="stroked">shell

<add place="above">guns</add> thumped, or down the flues made

moan. </line>

<line>

<del rend="stroked">But all was sleep. And no voice called for men.

</line>

<line><del rend="stroked">My

<add place="inspace">Strange</add> friend," I said, "here is no cause to mourn." </line>

<line>"None", said that other, "save the undone years, </line>

<line>The hopelessness. <add place="above">

<del rend="stroked">unachieved.

</add>

<add place="below">

<del rend="stroked">

<gap reason="unreadable"/>

<add> Whatever hope is yours, </line>

<line>Was my life also; comrade.

<add place="above">

<del rend="stroked">for

</add>

<add place="below">I went</add>

<del rend="stroked">I ran

Annotations from L. E. D. 1933. 1. By the author, and not the Bayeux Tapestry. 2. A Day is 12 hours, so 1000 days = 1000 hours. 3. A Day is 24 hours, so 1000 days = 10000 hours. 4. Years, and so of the rest from 1 to 1000.



TEI

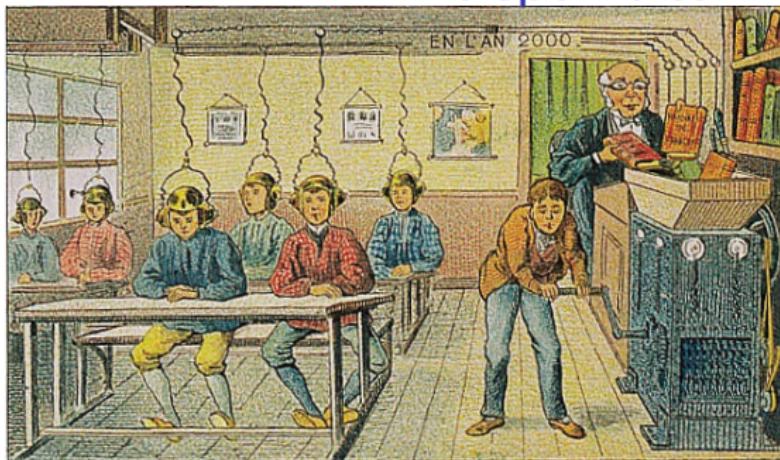
1995 - ?

- Pendant que nous théorisions...
 - Le web est arrivé !
 - Le tournant numérique transforme les archives et les bibliothèques !
 - La numérisation de masse s'effectue !
 - Les traitements numériques se déplacent sur des grilles de services, et/ou des systèmes domestiques !
 - Les réseaux sociaux emergent sur Internet !
- Convergence et travaux collectifs : méthodes empruntées des sciences dures
 - Est-ce qu'on pourrait remplacer l'édition classique en utilisant les méthodes collectives (cloud/crowd computing) ?
 - On s'aperçoit du besoin des infrastructures numériques

Les humanités numériques sont partout

- Comme M Jourdain, on fait des digital humanities sans le savoir?
- Les DH ne sont-elles qu'une gamme de technologies dans le vent ?

Le défi actuel : profiter de l'énormité des données disponibles



- Quels outils pour traiter un million de livres ?
 - <http://books.google.com/ngrams>
 - <http://www.etalab.gouv.fr/>
 - <http://rechercheisidore.fr/>
- Mais la transition de l'étude de l'oeuvre à l'étude du contexte reste problématique ...

conclusions



Quel est ce bruit dans la bibliothèque numérique ?

- Traiter un texte c'est plus que le lire, plus que l'annoter, plus que l'associer avec d'autres textes.
- Il s'agit d'exposer sa structuration afin de permettre un monde distribué, où "les livres dans la bibliothèque savent se parler entre eux"



Comment effectuer une telle démarche?

It's not rocket science (pas besoin d'avoir fait saint cyr) !

- ① Un balisage riche et sémantique (par exemple TEI-XML), dérivé d'une modelisation explicite des sources
- ② Une politique d'accès ouverte
- ③ Une infrastructure permettant l'intégration et l'archivage pérenne des données

Repenser l'édition numérique

- On est dans un monde où les documents prolifèrent, mais les textes risquent de disparaître
- Nous avons besoin de conserver nos interprétations, nos lectures, pour construire les éditions numériques
- Sans perdre les vertus traditionnelles d'un empirisme sceptique
- Sans perdre de vue l'importance du synthèse

Est-ce une édition ou une archive?

Composants de l'édition numérique nouvelle

- Images de pages (ou d'autres surfaces)
- Transcriptions, éventuellement annotées
- Edition/s synthétique
- Traduction modernes, sommaires
- Annotations paratextuelles, glossaires, préfaces, bibliographie...
- Descriptions des sources; métadonnées
- Pointeurs sur des "Factoids"

et tout cela ouvert, prêt à être ré-approprié

Un rôle majeur pour les SHS

- Nous comprenons les objets textuels
 - De quelle manière se présente ce discours?
 - Quelles sont les histoires qu'il raconte?
- Nous connaissons l'hermeneutique
 - quelle est la portée de ce discours?
 - Qu'est-ce qu'il veut dire – mais ne dit pas ?
- Voici notre contribution au web sémantique.



Merci de votre attention!

avenue centrale

Saison 2014-2015

rendez-vous en sciences humaines

www.avenue.centrale.fr

[Facebook.com/AvenueCentrale](https://www.facebook.com/AvenueCentrale)

[Twitter.com/AvenueCentrale](https://twitter.com/AvenueCentrale)



SCIENCES HUMAINES

