

Chapitre 6

Navigation en contexte

Nous avons identifié cinq profils-type de sessions sur la base de leur forme et de leur temporalité, et constaté que chacun d'entre eux est corrélé à certains types de contenus ; pour comprendre ces différences, nous devons maintenant replonger les parcours au sein des pratiques individuelles. Cette mise en contexte de la visite de sites et de pages éclaire les parcours sous l'angle des territoires personnels des internautes : elle permet de mettre à jour les activités normées au niveau individuel, et les modes d'appréhension prototypiques au niveau collectif. Elle nous amène également à voir comment les contenus sont appréhendés en situation et leur sens est construit au sein des parcours.

6.1 La session à l'aune de l'utilisateur

Les études sur les comportements de navigation en situation de recherche d'information ont montré la double primauté de l'expertise de l'utilisateur dans la manipulation des outils de recherche, et de la connaissance du domaine de recherche dans la topologie des parcours sur le Web (par exemple : [Jenkins *et al.* 2003]). Même si notre étude dépasse le cadre de la recherche sur le Web pour considérer les activités de navigation dans leur ensemble, on cherchera légitimement à expliquer les différences constatées entre les sessions en les rattachant à l'individu et à ses pratiques.

6.1.1 Profils d'usages et profils de sessions

Types de panélistes, types de sessions ?

Dans [Catledge & Pitkow 1995], Catledge et Pitkow scindent leur panel de cent sept utilisateurs en trois classes, basées sur la longueur et la fréquence de séquences de navigation : « serendipitous browsers », « general purpose browsers » et « searchers ». [Ghitalla *et al.* 2003] observent également parmi les vingt-trois internautes qu'ils étudient des stratégies de navigation bien distinctes, qui impliquent

des différences dans la manipulation des interfaces autant que dans les sites accédés et les ressources mobilisées au sein de recherches documentaires.

On s'attendra à bon droit à trouver dans nos données des éléments similaires. Un exemple sur sept panélistes tirés au hasard dans les données SN2002 va d'ailleurs dans ce sens : la part, dans l'ensemble des sessions de chaque panéliste, de chacun des cinq types de parcours que nous avons identifiés montre des différences notables entre les individus (voir Figure 6.1). Pour le panéliste A, la répartition est assez équilibrée à l'exception des parcours éclatés qui sont très minoritaires ; pour l'utilisateur C, au contraire, parcours à pivots et parcours éclatés représentent les trois quarts des sessions.

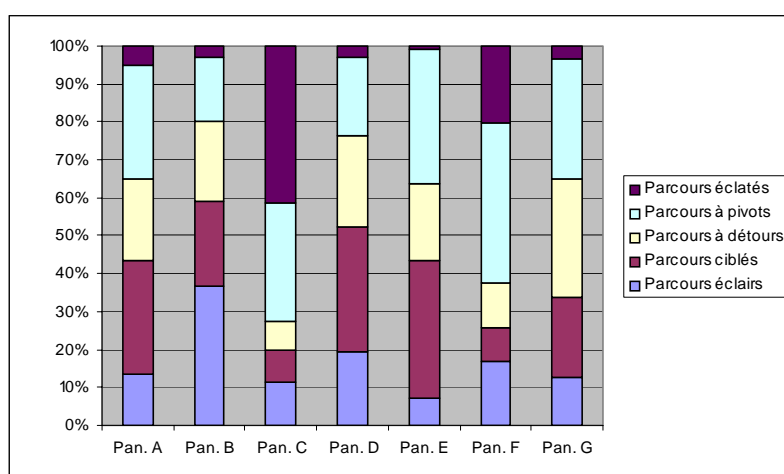


Figure 6.1. SN2002 – Part des différents types de sessions pour sept panélistes

De manière générale, à l'exception des très faibles utilisateurs qui font moins d'une dizaine de sessions sur la période d'observation, aucun des panélistes observés dans les trois jeux de données n'a de comportement exclusivement dédié à tel ou tel type de parcours. Cependant, certains types de comportements de navigation sont comparativement plus présents chez certains : pour le panel SN2002, par exemple, les parcours éclatés représentent en moyenne 13,3 % des sessions d'un internaute, mais pour un quart des individus, ils forment moins de 2 % des sessions (voir Tableau 6.1).

Tableau 6.1. Répartition moyenne des types de parcours pour les panélistes SN2002

	Part des parcours éclairs	Part des parcours ciblés	Part des parcours à détours	Part des parcours à pivots	Part des parcours éclatés
Moyenne	18,0 %	20,4 %	21,8 %	26,5 %	13,3 %
Quartile 1/2	7,3 %	8,9 %	13,5 %	16,7 %	1,9 %
Quartile 2/3	15,0 %	17,5 %	20,8 %	25,3 %	8,0 %
Quartile 3/4	24,6 %	28,1 %	28,6 %	34,0 %	18,7 %

Clef de lecture : les panélistes de SN2002 ont en moyenne 18 % de parcours éclairs ; pour un quart d'entre eux ces parcours moins de 7,3 % de leurs sessions.

Nous avons donc pratiqué une classification des internautes sur la base de la part de chaque type de parcours dans l'ensemble de leur corpus de sessions, afin de voir si des profils d'internautes peuvent être dégagés à partir des profils de sessions. Les résultats de la classification montrent des profils relativement différenciés (voir Figure 6.2 ci-dessous). On ne s'étonnera pas de ce que la partition en six classes ainsi obtenue recoupe globalement les types de sessions ; deux éléments intéressants sont ici à noter. En premier lieu, alors qu'on avait vu s'opposer dans l'examen global des sessions un groupe de sessions courtes et simples (parcours éclairs et ciblés) à un ensemble plus complexe formé des trois autres types de sessions, on constate, en ramenant les sessions aux individus, que trois pôles principaux émergent :

- d'un côté, un ensemble « parcours éclairs / ciblés »
- directement opposé à celui-ci, un groupe « parcours éclatés / à pivots »
- le groupe des parcours à détours, qui se pose comme spécifique par rapport aux deux autres.

D'autre part, l'examen des effectifs relatifs à chaque classe et à chacun de ces trois groupes montre que s'il existe certains panélistes « spécialisés » dans des comportements de navigation particuliers, un peu moins de la moitié des individus se rattache à une classe centrale et affecte un comportement médian.

Ce dernier constat relativise l'affirmation selon laquelle les internautes se comportent chacun de manière très particulière dans leurs modes de navigation : si pour une moitié d'entre eux une certaine forme de spécialisation dans un type de navigation est observable, pour l'autre moitié, les comportements de navigation sur le Web se répartissent globalement en sessions longues et courtes, simples et complexes, structurées ou décousues.

Types de sessions et types d'usages d'Internet

Pour éclairer cette relative concordance entre profils de sessions et profils d'internautes, on tentera de croiser la répartition des types de sessions de chaque panéliste avec des variables relatives à l'utilisateur. On laissera de côté les éléments socio-démographiques : on imagine difficilement dire qu'un cadre fait plus de parcours ciblés qu'un artisan – tout du moins une telle différence, si elle devait être observée, n'est pas explicative. Les variables relatives aux usages d'Internet sont plus pertinentes, en ce qu'elles positionnent les situations de navigations dans un contexte global d'usage de l'outil : l'ancienneté de la pratique renvoie à l'appropriation des TIC et pourrait expliquer l'apparition de routines et de navigations plus ciblées ou plus structurées autour de ressources identifiées ; le profil général d'usage d'Internet que nous avons présenté au Chapitre 5, qui croise intensité d'utilisation et diversité des services utilisés, peut également éclairer les comportements de navigation ; enfin, le type d'abonnement, notamment en termes de débit, peut s'avérer très structurant pour les types de sessions pratiqués.

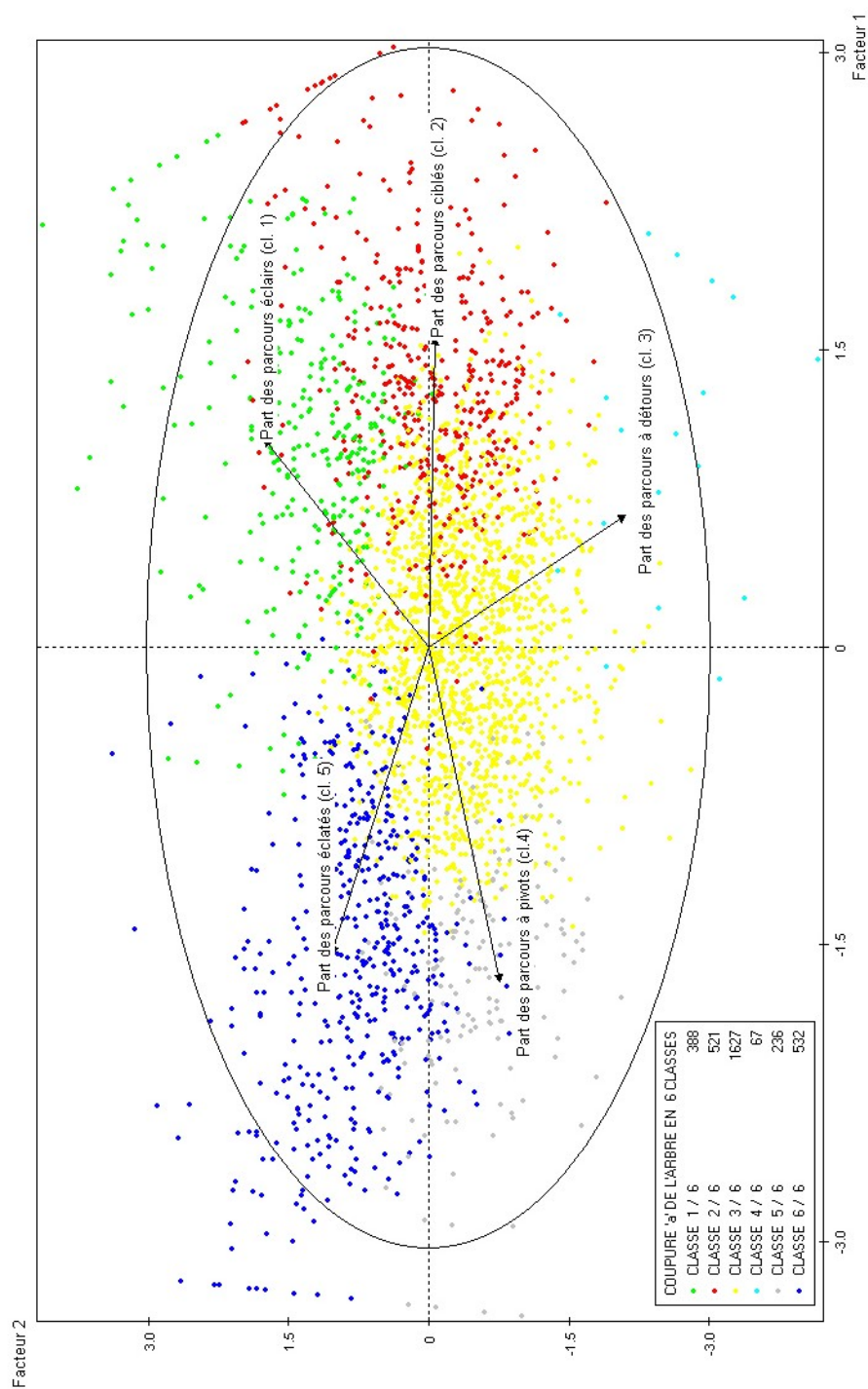


Figure 6.2. SN2002 – Classification des panélistes en fonction des types de sessions réalisés

La confrontation de ces variables et de la typologie des sessions vient pourtant démontrer ces hypothèses. L'ancienneté de la pratique, tout d'abord, ne semble que très marginalement liée aux types de sessions (voir Tableau 6.2 ci-dessous) : les différents groupes d'internautes du panel SN2002, anciens ou récents, ont en moyenne la même répartition des types de parcours sur les dix mois d'observation. On notera tout au plus une présence décroissante des parcours éclatés à mesure que la pratique est ancienne (11 % chez les plus anciens, contre 15 % chez les plus récents), tandis que les parcours éclairs sont au contraire plus faibles chez les nouveaux internautes (16 % des sessions en moyenne) que chez les plus anciens (20 % en moyenne). Ce léger effet d'apprentissage ne suffit pourtant pas à expliquer la diversité des types de sessions, et l'on conclura à l'indépendance globale des deux variables.

Tableau 6.2. SN2002 – Ancienneté de la pratique et types de parcours

Première connexion	Part des parcours éclairs	Part des parcours ciblés	Part des parcours à détours	Part des parcours à pivots	Part des parcours éclatés	Total
Avant 1997	20%	22%	22%	26%	11%	100 %
1997	19%	22%	23%	25%	12%	100 %
1998	18%	20%	23%	26%	13%	100 %
1999	18%	21%	22%	26%	13%	100 %
2000	18%	20%	21%	27%	15%	100 %
2001	16%	20%	22%	27%	15%	100 %

Clef de lecture : chez les panélistes connectés pour la première fois avant 1997, 20 % des sessions en moyenne sont du type « parcours éclairs ».

Il en va de même pour le type de connexion dont est équipé l'internaute (voir Tableau 6.3) : entre ceux équipés d'un modem RTC classique et ceux disposant d'un accès à haut débit (ADSL, Câble), on ne note pas de différence particulière dans la répartition des cinq classes de sessions.

Tableau 6.3. SN2002 – Type de parcours et type de connexion (répartition moyenne)

Type de connexion	Parcours éclairs	Parcours ciblés	Parcours à détours	Parcours à pivots	Parcours éclatés	Total
Bas débit (3035 ind.)	17,7 %	20,8 %	21,9 %	26,4 %	13,2 %	100 %
Haut débit (240 ind.)	20,6 %	17,2 %	20,5 %	27,7 %	14,0 %	100 %
NSP (96 ind.)	21,8 %	16,3 %	20,7 %	25,6 %	15,7 %	100 %

Le croisement entre types de sessions et typologie d'usages d'Internet amène à une conclusion similaire : ici encore, les différents parcours-type sont répartis de manière équivalente à travers les quatre groupes d'usagers que nous avons identifiés (voir Tableau 6.4). On notera, certes, que les usagers « ludiques » ont un profil quelque peu différent des trois autres groupes, avec une part plus importante de parcours éclatés et à pivot, au détriment des parcours ciblés : chez ces internautes intensifs, l'augmentation du nombre de sessions se fait au profit de parcours longs et complexes.

Tableau 6.4. SN2002 – Usages d’Internet et types de parcours

Types d’usages d’Internet	Part des parcours éclairs	Part des parcours ciblés	Part des parcours à détours	Part des parcours à pivots	Part des parcours éclatés	Total
Usages occasionnels	18 %	22 %	23 %	26 %	11 %	100 %
Web ordinaire	20 %	21 %	21 %	26 %	12 %	100 %
Comm. interpersonnelle	15 %	20 %	22 %	27 %	16 %	100 %
Usages ludiques	18 %	15 %	20 %	29 %	18 %	100 %

Clef de lecture : dans le groupe des internautes tournés vers les « usages ludiques », 18 % des sessions en moyenne sont du type « parcours éclairs ».

Ces éléments nous amènent à conclure à des déterminations beaucoup plus locales pour expliquer les différents profils de sessions : ce n’est pas tant le profil général de l’utilisateur, son intensité d’utilisation, sa maîtrise des interfaces ou sa connaissance des contenus et services disponibles sur le Web qui déterminent son comportement de navigation, mais plutôt les types d’activités dont la navigation est le support, et la connaissance locale des sites visités. C’est ainsi que nous interprétons les différences observables entre nos trois jeux de données (voir Tableau 6.5 ci-dessous) : les « gallicanautes » se différencient fortement des panels généralistes, avec une sur-représentation des parcours à pivot (33 % en moyenne, contre 26 % pour SN2002) au détriment des parcours ciblés (11 % pour BibUsages, contre 20 % pour SN2002).

Tableau 6.5. Part moyenne des types de sessions chez les panélistes pour chaque panel

	Part des parcours éclairs	Part des parcours ciblés	Part des parcours à détours	Part des parcours à pivots	Part des parcours éclatés
BibUsages	19 %	11 %	21 %	33 %	17 %
SN00-02	16 %	21 %	22 %	29 %	12 %
SN2002	18 %	20 %	22 %	26 %	13 %

Cette opposition est à mettre en parallèle avec les différences que nous avons déjà soulignées au cours du Chapitre 5 : les sessions plus courtes, plus denses et plus rythmées observées pour le panel BibUsages nous amenaient à conclure à une plus grande efficacité dans la navigation, que l’on était tenté de relier à un effet de l’ancienneté de la pratique. Sur ce dernier point, l’examen de l’ancienneté dans les données SensNet en 2002 montre que ce n’est pas la bonne hypothèse. On expliquera plutôt par les tâches effectuées localement la spécificité du panel BibUsages : panel mixte résidentiel/professionnel, composé essentiellement de chercheurs, amateurs ou professionnels, qui utilisent intensivement – mais pas exclusivement – le Web comme une ressource documentaire. Les correspondances observées entre forme de parcours et contenus visités vont également dans ce sens : si l’on ne peut attacher une thématique particulière à un type de parcours, on constate que certains types de services et d’outils sont sur-représentés dans certains contextes de navigation.

Dès lors, la cohérence des sessions d’un même individu tient plus à la variété des situations d’usage dans lesquelles il se trouve : l’un va consulter sa messagerie à

chaque session, l'autre n'utilisera le Web que pour rechercher des documents sur un thème précis, un troisième sera majoritairement tourné vers le WebChat. La notion de territoire revêt alors une importance particulière pour comprendre les parcours sur la Toile ; rattaché à un utilisateur donné, le territoire s'inscrit dans l'histoire de ses pratiques et de ses savoir-faire, et c'est dans ce double contexte que se déploie le sens des activités de navigation sur le Web.

Synthèse. À l'exception des très faibles utilisateurs, aucun des panélistes observés dans les trois jeux de données n'a de comportement exclusivement dédié à tel ou tel parcours-type. Certains types de comportements de navigation sont comparativement plus présents chez certains internautes, mais ni la segmentation des utilisateurs sur la base de l'usage général d'Internet, ni les variables d'ancienneté ou de type de connexion ne permettent d'expliquer ces différences.

6.1.2 Territoires sur le Web

Les données de trafic ne nous permettent pas de reconstruire véritablement les tâches des utilisateurs : on peut tout au plus supposer, en examinant manuellement les données, qu'un utilisateur est dans telle ou telle configuration, mais son intention, et corrélativement la grille interprétative qui guide son appréhension des contenus visités, nous échappent. Par contre, ces données centrées-utilisateur nous permettent de connaître les sites visités par chaque panéliste sur l'ensemble de la période et le contexte dans lequel chacun apparaît. On peut ainsi dresser pour chaque internaute une cartographie des ressources qu'il mobilise sur le Web, et observer sur la durée l'envergure et la structure de ce territoire personnel. Pour cela, le panel SN00-02 nous apporte des données inédites de par la taille de l'échantillon autant que la durée d'observation, et c'est principalement sur ces données que nous allons nous baser pour approfondir cette problématique.

Intensités variables, habitudes constantes

L'ensemble des 597 personnes du panel SN00-02 a consulté au cours des 34 mois d'observation 192 000 sites/portails différents. Pour autant, tous ces sites n'ont pas été vus par l'ensemble du panel, loin s'en faut (voir Tableau 6.6) : les deux tiers des sites n'ont été visités que par un seul internaute du panel, tandis qu'un autre quart l'a été par 2 à 5 panélistes différents.

Tableau 6.6. SN00-02, nombre de panélistes différents par site/portail

	Nombre de sites	Part des sites
1 visiteur	127 357	66,3 %
2 à 5	49 446	25,8 %
6 à 99	14 818	7,7 %
Plus de 100	378	0,2 %

Au total, seule une trentaine de sites ont été vus par plus de la moitié du panel, présentés au Tableau 6.7 ci-dessous : on retrouve dans cette liste les grands portails généralistes et les fournisseurs d'accès (MSN, Wanadoo, Yahoo, Club-internet) qui

s'imposent comme des carrefours et des points de passage pour aller « ailleurs » sur le Web (*via* les moteurs de recherche en particulier), et qui proposent en outre des services de communication qui attirent un grand nombre d'utilisateurs et les fidélisent. Par ailleurs, deux autres catégories de sites sont ici représentés : d'une part les sites proposant des outils nécessaires à la navigation (www.real.com, et www.macromedia.com, qui proposent des *plugins* multimédia souvent indispensables à la visite de certains sites ou de certains contenus), et d'autre part des sites de vente ou de services en ligne (Fnac, Alapage, SNCF, Pages Jaunes).

Tableau 6.7. SN00-02 – sites/portail vus par plus de la moitié du panel de 597 individus au cours des 34 mois

Site	Nb pan.	Site	Nb pan.
Wanadoo (fr)	553	SNCF	385
Yahoo (fr)	544	ibazar.fr	361
microsoft.com	543	chez.com	357
MSN	539	Fnac	356
multimania.fr	528	passport.com	356
Voila (fr)	522	online.fr	344
Yahoo (com)	521	libertysurf.fr	336
Lycos	515	wedoo.com	324
ifrance.com	480	Altavista (fr)	313
real.com	465	tiscali.fr	313
Club-internet	464	aol.com	311
macromedia.com	456	windowsmedia.com	309
Google	443	swisstools.net	308
Nomade	407	aol.fr	305
Amazon	406	Alapage	303
pagesjaunes.fr	400	TF1	300
Altavista (com)	396		

L'audience des sites Web apparaît ainsi très éclatée : hormis les quelques points de passage communs à la majorité des internautes, chacun semble aller chercher sur le Web les contenus qui l'intéressent spécifiquement. Ceci ne doit pas nous surprendre outre mesure : en n'observant que 597 personnes, fût-ce durant 34 mois, nous avons peu de chances que l'hétérogénéité du panel en termes de centres d'intérêt amènent ses individus à visiter les mêmes contenus.

Le même calcul appliqué aux données BibUsages le confirme : ici, si la part des sites vus par un seul visiteur est de 81 % (contre 67 % pour SN00-02), ce sont 66 sites qui ont été vus par plus de la moitié du panel. Ce résultat est d'autant plus important qu'avec ses 76 utilisateurs, le panel BibUsages est bien plus petit, et que la durée d'observation n'est que de six mois. Par contre, il est beaucoup plus homogène : constitué par des internautes intéressés par les textes et les contenus culturels en ligne, il montre qu'une communauté d'intérêt se retrouve autour de certains sites particuliers et investit des territoires similaires.

À l'échelle de l'utilisateur, le nombre de sites et de portails différents visités par chaque individu du panel SN00-02 au cours des 34 mois d'observation est très variable, allant de 11 pour le plus faible utilisateur à 8 900 pour le plus intensif, avec une moyenne globale de 943 sites (médiane : 569). Pour autant, tous les sites d'un

utilisateur ne sont pas vus de manière égale : l'examen pour chaque panéliste de la répétition des sites dans ses différentes sessions (voir Tableau 6.8) montre qu'en moyenne, les trois quarts des sites visités par les panélistes SN00-02 au cours des 34 mois d'observation n'apparaissent que dans une seule session.

Tableau 6.8. SN00-02 – Répartition des sites visités en fonction du nombre de sessions où ils apparaissent chez un panéliste

	1 session	2-4 sessions	5-6 sessions	10-99 sessions	Plus de 100 sess.
Moyenne	74,4 %	18,1 %	3,9 %	3,3 %	0,3 %
Médiane	74,5 %	18,3 %	3,8 %	3,1 %	0,2 %
Minimum	50,8 %	0,0 %	0,0 %	0,0 %	0,0 %
Maximum	100,0 %	40,0 %	12,1 %	12,3 %	6,2 %

Clef de lecture : en moyenne, 74,4 % des sites vus par un panéliste donné ne sont vus que dans une seule session, 18,1 % dans 2 à 4 sessions.

Cette dispersion des sites visités est-elle liée à l'intensité d'usage de chaque panéliste ? En d'autres termes, les utilisateurs intensifs du Web ont-ils tendance à voir plus de nouveaux sites, à avoir des pratiques de « surf » plus fréquentes par rapport aux utilisateurs modérés, cantonnés à quelques sites bien définis ? Le Tableau 6.9 scinde, par quartile du nombre total de sessions, la fréquence d'apparition des sites de chaque panéliste dans ses sessions sur les 34 mois, et permet de voir si les utilisateurs intensifs ont plus d'hapax dans leur corpus de sites que les autres.

Tableau 6.9. SN00-02 – Intensité d'usage et présence des sites visités dans les sessions

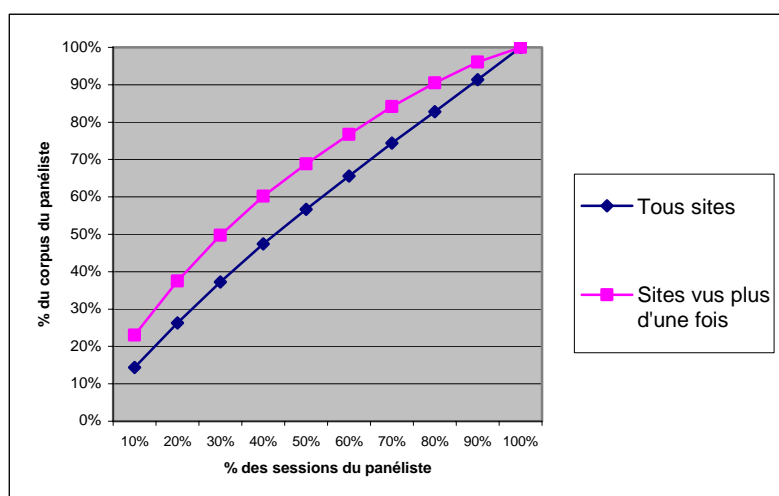
Quartile du nombre de sessions total	Sites vus dans 1 session du pan.	Sites vus dans plusieurs sessions du pan.			Total
		0 à 19 % de ses sessions	20 à 39 % de ses sessions	40 à 100 % de ses sessions	
1 ^{er} quartile	78,0 %	18,8 %	1,7 %	1,5 %	100 %
2 ^e quartile	75,7 %	23,3 %	0,5 %	0,5 %	100 %
3 ^e quartile	73,7 %	25,8 %	0,3 %	0,2 %	100 %
4 ^e quartile	71,3 %	28,4 %	0,8 %	0,1 %	100 %

Clef de lecture : chez les faibles utilisateurs (1^{er} quartile), 78 % des sites d'un utilisateur n'apparaissent que dans une session en moyenne, et 1,5 % de sites sont présents dans plus de 40 % de ses sessions sur les 34 mois.

Malgré une légère sur-représentation des sites présents dans une seule session chez les faibles utilisateurs (78 %), et l'inverse chez les intensifs (71 %), les différences nous semblent trop faibles d'un groupe à l'autre pour que l'on puisse conclure à un lien direct entre intensité d'usage et concentration des sites visités sur le Web. Tout au plus pourra-t-on remarquer que les internautes ayant effectué le plus grand nombre de sessions semblent, paradoxalement, être ceux qui revisitent le plus de sites, alors que l'on pouvait supposer une fréquentation plus « explosée » du Web de leur part. Aussi suggèrera-t-on que les utilisateurs intensifs connaissent mieux les ressources sur le Web, utilisent mieux les outils de recherche et ciblent mieux les contenus qui les intéressent, ce qui les amène à moins d'errements et moins de sites vus une seule fois.

Un corpus en constante expansion

Sur cette base, on pourrait faire l'hypothèse que chacun a sur le Web des territoires bien déterminés, et qu'une fois passée une phase de découverte des ressources intéressantes pour lui, son « corpus » de sites se stabilise autour de quelques-uns. Dans les faits, le corpus de sites de chaque panéliste est en constante augmentation au cours des 34 mois d'observation : en moyenne, lorsqu'il a effectué la moitié des sessions observées sur la période, ce sont seulement 57 % de son corpus de sites qu'il a exploré (voir Figure 6.3), et à 90 % des sessions observées, il a vu 91 % seulement de son corpus de sites.



Clef de lecture : un panéliste ayant effectué 20 % des sessions observées sur les 34 mois a visité en moyenne 26 % de l'ensemble des différents sites qu'il voit sur la période.

Figure 6.3. SN00-02 – Évolution du nombre de sites différents visités au fil des sessions

Ce lien quasi linéaire entre nombre de sessions et nombre de sites différents visités indique qu'au fil des sessions, l'internaute continue à explorer de nouveaux sites, au premier comme au dernier mois d'observation : la part importante des sites vus dans une seule session ne peut donc pas être imputée à une période initiale de découverte des ressources Web, mais semble faire partie intégrante des pratiques de navigation au quotidien.

Si l'on restreint l'analyse aux sites vus dans deux sessions différentes au moins, on cerne des sites qui sont d'un certain intérêt pour l'internaute, et qui peuvent entrer dans le cadre de son espace familier et connu : on s'attendra plutôt à ce que ces sites soient tous vus dans une période plus ou moins courte, et qu'au début des 34 mois d'observation, on en ait balayé une bonne partie. Ici encore, les données viennent démentir cette approche intuitive : comme le montre la Figure 6.3, au fil des sessions de l'internaute, celui-ci découvre de nouveaux sites sur lesquels il va revenir par la suite, et à la moitié des sessions observées sur la période, ce sont seulement 70 % des sites vus plus d'une fois qui ont été visités.

En définitive, ce n'est pour chaque internaute qu'une poignée de sites qui suscitent une réelle fidélité. Cette notion ne doit pas être confondue avec la

régularité et la présence d'un site dans toutes les sessions du panéliste : il est nécessaire de relier la session à une activité bien précise, et dans ce cadre, la régularité peut être très différente de la fréquence. Un panéliste peut visiter systématiquement le même site pour une tâche précise, cette tâche n'intervenant qu'occasionnellement : la déclaration d'impôts en ligne fournit un exemple typique de cette situation, où l'internaute ne visite le site du Ministère de finances qu'une fois par an, mais peut le faire tous les ans à la même époque. Pour analyser la fidélité aux sites, on examinera alors l'empan d'un site, c'est-à-dire le nombre de jours qui séparent la première et la dernière visite d'un site vu plusieurs fois par le panéliste au cours des 34 mois d'observation : dans un tiers de cas, cette durée ne dépasse pas trente jours, et elle n'est supérieure à un an que pour un quart des observations. Le nombre moyen de sites dont l'empan dépasse une année est de 72 en moyenne par panéliste, avec moins de 16 sites pour les internautes les moins actifs, et plus d'une centaine pour les plus intensifs. L'élément le plus remarquable ici est que, malgré cette diversité dans l'intensité d'usage, le nombre de sites vus plus d'un an est toujours proportionnel au nombre total de sites visités par l'internaute : pour 90 % des internautes du panel SN00-02, ces sites représentent entre 25 et 35 % de l'ensemble des sites visités plusieurs fois sur la période, et 7,5 % de l'ensemble des sites qu'il voit.

En définitive, rares sont les sites qui sont à la fois vus fréquemment et sur une longue période : pour l'ensemble des sites vus plus d'une fois sur les trois ans par un utilisateur donné, on compte en moyenne 83 jours entre deux visites (médiane : 34 jours), et pour un quart des sites, cet écart est de moins d'une semaine. Fréquence et fidélité ne se confondent donc pas, et le territoire sur le Web apparaît délimité en trois zones :

1. les sites qui ne sont vus que dans une seule session, et qui représentent les trois quarts des sites vus par un internaute en moyenne ;
2. les sites vus plus d'une fois, mais qui ne fidélisent pas l'internaute. Ces sites sont actifs sur des courtes et moyennes périodes, 80 jours en moyenne (médiane : 34 jours), et pour la moitié des cas apparaissent dans deux sessions seulement. Ils forment en moyenne 17 % du corpus de sites d'un internaute, et semblent correspondre à des activités passagères pour l'individu, ou à une forme de « rejet après essai ».
3. les sites familiers : vus sur plus d'un an, ils sont visités préférentiellement par l'internaute dans un contexte donné. On retrouve bien évidemment dans cette catégorie les portails généralistes et les moteurs de recherche, ainsi que les services qui fidélisent les utilisateurs comme le WebMail, mais également des sites spécialisés sur tel ou tel sujet, pour lequel l'internaute a un intérêt particulier : tomshardware.com (informations sur le matériel informatique), finances.gouv.fr (site du Ministère des finances), allocine.fr (films et séances de cinéma), etc. Ils n'apparaissent pas systématiquement dans beaucoup de sessions – moins de 5 pour la moitié des cas, mais semblent avoir la préférence de l'utilisateur pour un domaine ou un service donné.

Territoires thématiques

La notion de thématiques et de services préférentiels sur le Web fait écho à la définition de territoires et de lieux habituels en termes de sites visités. Sur la base des

descriptions de contenu que nous avons rattachées aux pages visitées, nous pouvons établir pour chaque utilisateur une cartographie des thèmes et des services auxquels il accède, et examiner la fréquence et la régularité de l'accès à ces contenus dans les sessions.

Pour cela, nous décrivons chaque session par la catégorie *CatService* ou la catégorie d'annuaire de premier niveau sur laquelle l'internaute passe le plus de temps dans la session, comme nous l'avons déjà fait précédemment (voir Chapitre 5). On a pu voir que dans la plupart des cas, cette catégorie majoritaire couvre l'essentiel de la durée des sessions, celles-ci étant plutôt monothématiques ou monofonctionnelles. On dresse ainsi un profil des contenus visités par chaque internaute, correspondant à la part de ses sessions relatives à tel ou tel descripteur de contenu.

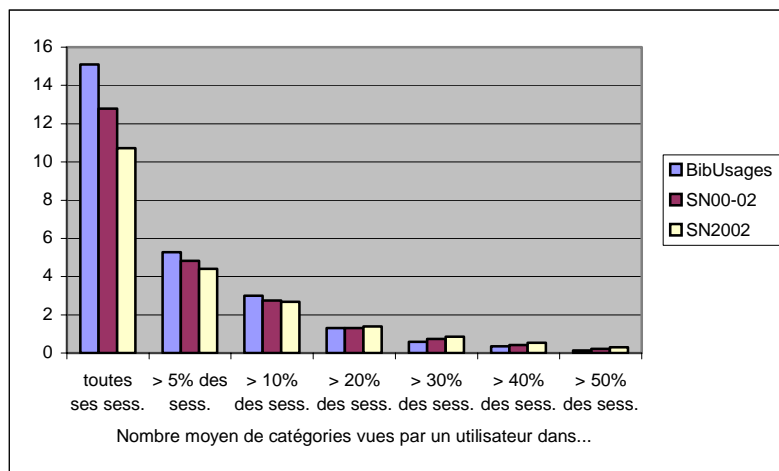
L'examen des profils nous révèle que les internautes ont chacun des centres d'intérêt bien délimités sur le Web, qui occupent la majorité de leur vie sur la Toile : l'exemple donné au Tableau 6.10 pour deux internautes du panel SN00-02 montre chez le premier un intérêt très prononcé pour les sites pornographiques et un fort usage des moteurs de recherche ainsi que, plus modérément, des sites de supermarchés en ligne, tandis que le second panéliste est plus orienté vers les sites consacrés au sport et aux loisirs, ainsi qu'au service de WebMail.

Tableau 6.10. Profil thématique de panélistes – exemples pour SN00-02 avec Nomade¹

Profil du panéliste 21166		Profil du panéliste 296	
Catégorie	% sess.	Catégorie	%sess.
PORNO	39,3 %	Sport et détente	22,9 %
<i>Aucune description</i>	19,5 %	<i>Aucune description</i>	20,0 %
Moteur	9,5 %	WebMail	10,0 %
Mes Courses	6,3 %	Culture et loisirs	7,1 %
Nouvelles technologies	5,8 %	Espace B to B	7,1 %
PG - Page Accueil	4,8 %	Mes Courses	7,1 %
Espace B to B	4,5 %	Moteur	7,1 %
WebMail	2,5 %	Actu, médias	4,3 %
Société, Vie pratique	2,3 %	Éducation, formation	4,3 %
Culture et loisirs	1,5 %	PG – Personnalisation	2,9 %
Actu, médias	1,3 %	Forme et Santé	1,4 %
Sport et détente	1,0 %	Nature et sciences	1,4 %
Voyage, géographie	0,8 %	PG - Non catégorisé	1,4 %
Éducation, formation	0,3 %	Société, Vie pratique	1,4 %
Nature et sciences	0,3 %	WebChat	1,4 %
PG - Achat	0,3 %	<i>Total</i>	100 %
PG - Non catégorisé	0,3 %		
PG - Personnalisation	0,3 %		
<i>Total</i>	100 %		

¹ Nous avons appliqué ce calcul à l'ensemble des sessions, d'où une catégorie « Aucune description » forte, correspondant aux sessions mal décrites par les annuaires. Cela étant, les résultats sont similaires si l'on restreint le calcul aux sessions bien décrites.

Un fort effet de concentration est observable pour l'ensemble des internautes des trois panels, et joue à deux niveaux : d'une part, alors que le nombre de catégories descriptives va de 31 à 34 selon l'annuaire considéré, on en compte beaucoup moins pour chaque panéliste pris individuellement. Pour le panel SN2002, ce sont en moyenne 10,7 catégories différentes qui décrivent l'ensemble des sessions de chaque panéliste, 12,8 pour l'échantillon SN00-02 et 15 pour les internautes de BibUsages. D'autre part, sur cette dizaine de catégories auxquelles le panéliste a consacré au moins une session, les deux tiers font l'objet d'une visite très occasionnelle et figurent dans moins de 5 % des sessions, tandis qu'en moyenne une seule catégorie décrit plus de 35 % des sessions du panéliste (voir Figure 6.4).



Clef de lecture : pour chaque internaute de BibUsages, 15 catégories différentes décrivent l'ensemble de ses sessions en moyenne, mais 5 catégories seulement sont représentées dans plus de 5 % des sessions.

Figure 6.4. Nombre de catégories décrivant les sessions des panélistes

Le nombre de catégories différentes sur lesquelles un panéliste porte son intérêt est, assez logiquement, fonction de l'intensité d'usage du panéliste : plus celui-ci réalise de sessions, plus il est amené à voir des sites sur des contenus variables. Pour le panel SN00-02, le quart le moins actif voit au total 5 types de contenus différents en moyenne sur les 34 mois d'observation, tandis que le quart le plus intensif en consulte plus de 18. Pourtant, fait remarquable, le nombre de catégories représentatives de plus de 5 % des sessions est indépendant de l'intensité d'usage et demeure autour de 5 pour l'ensemble du panel (voir Tableau 6.11).

Tableau 6.11. SN00-02 – Intensité d’usage et variété des contenus visités (moyennes)

Quartile du nombre de sessions	Nb total de catégories vues par le panéliste	Nb cats vues dans plus de 5% des sessions	Nb cats vues dans plus de 10% des sessions	Nb cats vues dans plus de 30% des sessions	Nb cats vues dans plus de 50% des sessions
1 ^{er} quartile	5,5	4,4	2,9	1,0	0,3
2 ^e quartile	11,7	5,2	2,8	0,6	0,2
3 ^e quartile	15,3	5,0	2,7	0,6	0,2
4 ^e quartile	18,6	4,7	2,6	0,7	0,2

Clef de lecture : pour les panélistes les moins actifs, 5,5 catégories décrivent l’ensemble de leurs sessions, mais une seule catégorie en moyenne décrit plus de 30 % des sessions.

Ces chiffres confortent l’idée de la constitution d’un espace personnel sur le Web par les internautes : ce qui était vrai au niveau des sites visités l’est également sous l’angle des contenus. Si un individu peut être amené à visiter des sites sur des thèmes très variés, la majorité de ses pratiques reste centrée sur un nombre limité de services et de thématiques. Ce constat invite à relativiser, ou du moins à préciser, la notion de « surf » telle qu’elle peut être employée couramment : si le Web peut permettre d’accéder à toutes sortes de services, de types de documents, de thématiques, dans les pratiques, cette possibilité n’est pas utilisée à l’échelle macroscopique, chacun se créant les niches relatives à ses centres d’intérêt.

Ces informations sont importantes pour la compréhension des comportements de navigation : ramenées à l’échelle de la session, elles doivent permettre d’éclairer les comportements de navigation à l’aune des pratiques et des territoires individuels.

Synthèse. L’audience des sites Web apparaît ainsi très éclatée : hormis les quelques points de passage communs à la majorité des internautes, chacun se constitue sur le Web un territoire spécifique. La structure de ces territoires personnels peut être scindé en trois zones : les sites vus une fois seulement, qui constituent la grande majorité du corpus de sites de chaque panéliste, et qui ne cessent de l’alimenter au fil du temps ; les sites vus plusieurs fois mais de faible empan, correspondant à un besoin borné dans les temps ; et les sites réguliers, de fort empan, qui constituent le cœur des parcours de l’internaute. Dans cette dernière catégorie, on distingue une poignée de sites qui constitue le territoire familier, et les sites visités moins fréquemment correspondant à des contextes particuliers apparaissant peu souvent. Une structuration similaire est observée sur le plan des contenus : une grande diversité globale se lit dans les pages visitées par chaque panéliste, mais seuls quatre à cinq thèmes et services font l’objet de visites régulières et fréquentes.

6.2 Sessions en contexte

Pour chaque parcours, on peut maintenant replacer les sites visités dans le corpus général du panéliste d’une part, et les thèmes et services accédés dans la cartographie des contenus habituels ou occasionnels de l’internaute d’autre part. Cette double mise en contexte des parcours éclaire et explique les comportements observés.

6.2.1 Types de parcours et territoires personnels

Sessions fermées vs. sessions ouvertes

Les sites vus par un panéliste dans une seule session au cours des 34 mois d'observation constituent en moyenne les trois quarts du corpus de sites du panéliste. En revanche, ils ne figurent pas de manière égale dans toutes les sessions : seuls 35 % des sessions du panel SN00-02 contiennent des sites de ce type, et leur présence ou non dans un parcours est significative. En effet, les cinq parcours-type que nous avons identifiés ne sont pas représentés de la même manière dans les sessions comprenant ces sites à visite unique et dans celles qui n'en contiennent pas (voir Tableau 6.12).

Tableau 6.12. SN00-02 – Présence de sites vus une seule fois par le panéliste et types de parcours

Classe de session	Sessions ne contenant pas de sites vus une seule fois	Sessions contenant un ou plusieurs sites vus une seule fois	Toutes sessions
Parcours éclairs	21,5 %	3,4 %	15,0 %
Parcours ciblés	28,5 %	6,3 %	20,6 %
Parcours à détours	22,1 %	22 %	22,1 %
Parcours à pivots	19,0 %	47,4 %	29,1 %
Parcours éclatés	8,9 %	20,9 %	13,2 %
<i>Total</i>	100 %	100 %	100 %

Dans le premier cas, parcours à pivots et parcours éclatés sont sur-représentés, et constituent les deux tiers des sessions de la classe, au détriment des parcours éclairs et ciblés ; si la session ne contient que des sites vus plus d'une fois, au contraire, les parcours rapides et linéaires sont plus présents que dans l'ensemble des sessions (la moitié des sessions, contre 35 % dans l'ensemble). Enfin, les parcours à détours échappent à ce comportement, leur part étant la même dans les deux cas (21 %), ce qui confirme le profil mixte de ce type de parcours, à cheval entre la classe des parcours à pivots et celle des parcours ciblés.

Ces éléments se retrouvent également sous l'angle des thèmes et des services accédés : pour croiser ces deux variables, nous avons évalué pour chaque panéliste la fréquence des thèmes et services qu'il visite sur l'ensemble de la période, que nous avons recodée en cinq catégories allant de « contenu très rare » à « contenu habituel » (voir Tableau 6.13).

Tableau 6.13. Recodage de la fréquence des thèmes/services de chaque panéliste

Fréquence codée	Présence dans les sessions du panéliste
thème / service très rare	majoritaire dans aucune session
thème / service rares	majoritaire dans moins de 5 % des sessions
thème / service occasionnels	majoritaire dans 5 % à 20 % des sessions
thème / service fréquents	majoritaire dans 20 % à 40 % des sessions
thème / service habituels	majoritaire dans plus de 40 % des sessions

Les sessions contenant des thèmes peu fréquents chez le panéliste sont plus longues, plus complexes, et s'apparentent à des parcours de découverte (voir Tableau 6.14). Les sessions comportant des contenus rares et très rares pour un panéliste sont ainsi quasiment absentes des parcours éclairés et ciblés (1,5 % et 6,1 %), alors qu'elles sont très présentes dans les parcours à pivots (50 %) ; au fur et à mesure que l'on sélectionne des sessions avec des contenus plus fréquemment visités, ces sessions sont de plus en plus tournées vers les parcours éclairés et ciblés, et de moins en moins vers les parcours à pivots et éclatés.

Tableau 6.14. SN00-02 – Types de parcours et fréquence des contenus pour le panéliste (sessions bien décrites)

Classe de session	Sessions avec des contenus très rares	Sessions avec des contenus rares	Sessions avec des contenus occasionnels	Sessions avec des contenus fréquents	Sessions avec des contenus habituels	Toutes sessions
Parcours éclairés	1,5 %	6,1 %	8,7 %	10,3 %	14,1 %	13,8 %
Parcours ciblés	10,7 %	20,7 %	22,2 %	25,9 %	26,3 %	24,9 %
Parcours à détours	19,3 %	24,3 %	23,6 %	22,4 %	21,2 %	22,2 %
Parcours à pivots	49,2 %	35,9 %	33,2 %	30,2 %	25,6 %	28,3 %
Parcours éclatés	19,4 %	13,0 %	12,2 %	11,1 %	12,9 %	10,8 %
<i>Total</i>	100 %	100 %	100 %	100 %	100 %	100 %

La visite de nouveaux sites et l'appréhension de contenus inhabituels sont donc résolument liées à des parcours longs, plus complexes, plutôt orientés vers la recherche, en particulier avec les parcours à pivots. Une opposition nette se dégage entre deux contextes particuliers de navigation : d'un côté, les sessions routinières, plutôt fermées, qui amènent l'internaute à visiter des sites connus, et de l'autre des sessions ouvertes et exploratoires, où l'utilisateur découvre de nouveaux sites et de nouveaux contenus.

Du familier vers l'inconnu

Si la catégorie des sites habituels est très minoritaire dans l'ensemble des sites visités, elle constitue pourtant le noyau central des territoires personnels sur le Web. En effet, bien qu'aucun site de cette catégorie pris isolément n'apparaisse dans toutes les sessions, ces sites considérés dans leur ensemble sont présents dans 91 % des sessions de chaque panéliste en moyenne (voir Tableau 6.15).

Tableau 6.15. SN00-02 – Empan des sites et présence dans les sessions

% de sessions contenant au moins un site d'empan..	
Empan nul (site vu dans une seule journée)	43%
1 à 9 jours	13%
10 à 30 jours	12%
Un à cinq mois	35%
Six à 12 mois	36%
Plus d'un an	91%

Clef de lecture : pour un panéliste donné, les sites de son corpus dont l'empan est supérieur à un an sont présents en moyenne dans 91 % de ses sessions sur les trois ans.

À l'inverse, les sites qui ne sont vus qu'au cours d'une seule journée (au sein d'une ou plusieurs sessions dans la journée) sont présents dans moins de la moitié des sessions d'un panéliste. Le croisement de ces deux variables amène à penser que la découverte de nouveaux sites, qu'ils soient vus une fois ou dans une période allant jusqu'à douze mois, se fait quasi-systématiquement à partir d'un terrain connu : sites d'information proposant des liens, outils de recherches préférentiels. Ceci se voit confirmé lorsque l'on examine le temps passé dans chaque session sur ces sites habituels : dans 70 % des sessions, c'est sur ces sites que l'internaute passe le plus de temps dans la session (voir Tableau 6.16). Les autres catégories de sites n'entrent en première ligne que dans moins d'un tiers des sessions, les sites vus dans l'espace d'une journée n'étant dominants que dans 10 % des cas.

Tableau 6.16. SN00-02 – Empan des sites et importance dans les sessions

Catégorie de sites la plus présente dans la session en durée	% sessions
Sites vus dans une journée (empan nul)	9,5 %
Sites occasionnels, très faible durée de vie (empan de 1 à 9 jours)	2,0 %
Sites occasionnels, faible durée de vie (empan de 10 à 29 jours)	1,6 %
Sites occasionnels, durée de vie moyenne (empan de 1 à 5 mois)	8,2 %
Sites occasionnels, durée de vie importante (empan de 6 à 12 mois)	8,7 %
Sites habituels (empan de plus d'un an)	70,1 %
<i>Total</i>	100 %

Clef de lecture : dans 9,5 % des sessions, c'est sur les sites à visite unique que l'internaute passe le plus de temps dans la session.

Les sites habituels sont donc le support privilégié de la navigation à double titre : non seulement ils sont présents dans la grande majorité des sessions, mais en plus, ils occupent souvent la part la plus importante du temps de navigation. Cela est compréhensible pour les sessions routinières où un seul site est visité, mais pour les sessions plus longues, où l'internaute découvre de nouveaux sites, cela signifie que le territoire nouveau et occasionnel est toujours structuré autour du territoire connu et maîtrisé, et que les écarts de parcours restent limités, voire timides. Ce constat rejette définitivement l'idée de surf débridé au gré des hypertextes au rang des mythes : une bonne partie des sessions n'amène l'internaute qu'à des endroits connus et fréquents, et même lors de sessions ouvertes, il ne s'éloigne que peu de temps de l'espace personnel qu'il s'est forgé.

Typologie

Pour confirmer ces éléments, les observer de manière systématique et analyser leur combinaison, nous avons pratiqué une classification des sessions sur la base des indices relatifs à la place des sites de chaque session dans l'ensemble des sessions de l'individu. En travaillant toujours sur les données SN00-02, nous décrivons chaque session par trois familles de variables : fréquence des thèmes ou des services accédés, fréquence des sites visités et empan des sites visités. Chaque famille de variable est décomposée en valeurs significatives, dont on évalue la place dans la session ; nous ramenons ces trois éléments au temps passé dans la session, et obtenons un total de quatorze variables présentées au Tableau 6.17. Du fait de l'inclusion de variables de

contenu, nous ne travaillons que sur les sessions bien décrites par le duo *CatService - Nomade*, l'annuaire retenu ici pour caractériser les contenus visités.

Tableau 6.17. Variables retenues pour décrire les territoires dans les sessions

Famille de variable	% de la durée de la session sur des...
Fréquence des sites	sites n'apparaissant que dans cette session sites apparaissant dans 0 à 9 % des sessions sites apparaissant dans 10 à 39 % des sessions sites apparaissant dans 40 à 100 % des sessions
Empan des sites	sites n'apparaissant que dans une journée (empan nul) sites d'empan entre 1 et 29 jours sites d'empan entre 1 et 6 mois sites d'empan entre 7 et 12 mois sites d'empan supérieur à un an
Fréquence des thèmes et services	thèmes / services très rares thèmes / services rares thèmes / services occasionnels thèmes / services fréquents thèmes / services habituels

À l'issue de l'analyse en composantes principales des sessions sur ces variables, nous avons pratiqué une classification sur les six premiers facteurs (74 % de l'information), et retenu une partition en trois classes (voir Figure 6.5 ci-dessous). Comme le laisse entrevoir la projection sur les axes 1 et 2, le détail de la composition des classes oppose assez nettement un groupe (classe 1) aux deux autres.

Ce groupe rassemble 51,6 % des sessions, et s'apparente aux parcours exclusivement routiniers : 93 % de la durée de ces sessions se fait sur des sites dont l'empan dépasse l'année (global : 73 %), 58 % sur des sites figurant parmi les plus fréquents, et les deux tiers sur des thèmes et services fréquents ou habituels (voir Tableau 6.18).

Tableau 6.18. SN00-02 – Variables discriminantes des sessions routinières (cl. 1)

	Moy. pour la classe	Moy. globale	Variable
Corrélation positive	95 %	73 %	% durée sur sites d'empan de 350 jours et plus
	58 %	36 %	% durée sur sites vus dans 40 à 100% des sessions
	29 %	20 %	% de durée sur cats Nomade/CS fréquentes
	31 %	24 %	% durée sur sites vus dans 10 à 39% des sessions
	20 %	13 %	% de durée sur cats Nomade/CS habituelles
	27 %	29 %	% de durée sur cats Nomade/CS occasionnelles
Corrélation négative	0 %	3 %	% durée sur sites d'empan entre 1 et 29 jours
	12 %	20 %	% de durée sur cats Nomade/CS rares
	2 %	8 %	% durée sur sites d'empan entre 180 et 349 jours
	1 %	7 %	% durée sur sites d'empan entre 30 et 179 jours
	1 %	9 %	% durée sur sites vus dans une seule session
	2 %	9 %	% durée sur sites d'empan 0
	10%	32%	% durée sur sites vus dans 0 à 9% des sessions

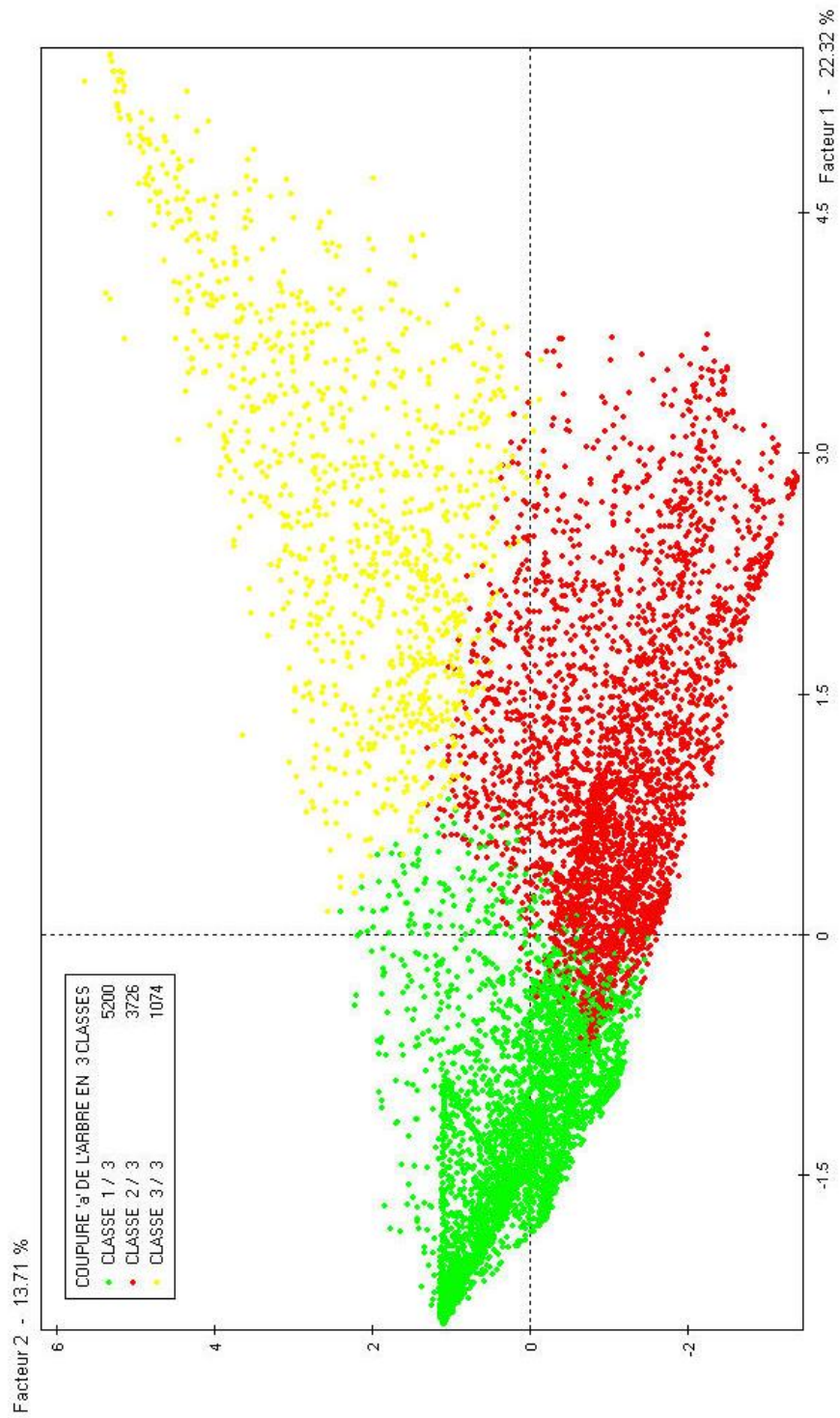


Figure 6.5. SN00-02 – Classification des sessions relative aux territoires sur le Web

À ces sessions routinières, s'opposent deux groupes de sessions plus ouvertes, orientées vers des contenus moins habituels. Le premier groupe, qui représente 36,5 % des parcours, est centré sur des contenus occasionnels : les sites présents dans peu de sessions et ceux d'empan moyen (entre un mois et un an) y sont sur-représentés, ainsi que les catégories Nomade / *CatService* rares (voir Tableau 6.19). Pour autant, ces sessions ne sont pas orientées vers la découverte, et semblent plutôt correspondre à des contenus ou des services que le panéliste visite relativement régulièrement sur une période donnée.

Tableau 6.19. SN00-02 – Variables discriminantes des sessions « contenus occasionnels » (cl. 2)

	Moy. pour la classe	Moy. globale	Variable
Corrélation positive	67 %	32 %	% durée session sur sites vus dans 0 à 9% des sessions
	20 %	8 %	% durée session sur sites d'empan entre 180 et 349 jours
	16 %	7 %	% durée session sur sites d'empan entre 30 et 179 jours
	28 %	20 %	% de durée session sur cats Nomade/CS rares
	7 %	3 %	% durée session sur sites d'empan entre 1 et 29 jours
Corrélation négative	33 %	29 %	% de durée session sur cats Nomade/CS occasionnelles
	17 %	24 %	% durée session sur sites vus dans 10 à 39% des sessions
	7 %	13 %	% de durée session sur cats Nomade/CS habituelles
	5 %	9 %	% durée session sur sites d'empan nul
	5 %	9 %	% durée session sur sites vus dans une seule session
	11 %	20 %	% de durée session sur cats Nomade/CS fréquentes
	52 %	73 %	% durée session sur sites d'empan de 350 jours et plus
	11 %	36 %	% durée session sur sites vus dans 40 à 100% des sessions

Tableau 6.20. SN00-02 – Variables discriminantes des sessions « découverte » (cl. 3)

	Moy. pour la classe	Moy. globale	Variable
Corrélation positive	56 %	9 %	% durée session sur sites d'empan nul
	53 %	9 %	% durée session sur sites vus dans une seule session
	35 %	20 %	% de durée session sur cats Nomade/CS rares
	1 %	0 %	% de durée session sur cats Nomade/CS très rares
Corrélation négative	5 %	7 %	% durée session sur sites d'empan entre 30 et 179 jours
	5 %	8 %	% durée session sur sites d'empan entre 180 et 349 jours
	23 %	32 %	% durée session sur sites vus dans 0 à 9% des sessions
	3 %	13 %	% de durée session sur cats Nomade/CS habituelles
	7 %	20 %	% de durée session sur cats Nomade/CS fréquentes
	10 %	24 %	% durée session sur sites vus dans 10 à 39% des sessions
	14 %	36 %	% durée session sur sites vus dans 40 à 100% des sessions
	32 %	73 %	% durée session sur sites d'empan de 350 jours et plus

Le dernier groupe de sessions, représentant 11,9 % de l'ensemble, est au contraire orienté vers la découverte et l'exploration de nouveaux territoires (voir Tableau 6.20). Les sites apparaissant dans une seule session, sous-représentés dans les autres groupes, sont ici majoritaires dans la session, dont ils représentent en moyenne 56 %

de la durée. Corrélativement, les thèmes et services rares et très rares sont fréquents dans ce groupe, au détriment des thèmes fréquents et habituels.

L'activité de navigation se partage donc globalement entre deux grands types de sessions : d'un côté, les sessions routinières où l'internaute accède à des pages qu'il voit très fréquemment, et qui font partie de son territoire restreint. De l'autre, on trouve des parcours plus ouverts, décomposés en sessions de semi-routine, aux contenus occasionnels que l'utilisateur sera néanmoins amené à revisiter, et en sessions de découverte au sein desquelles l'internaute visite beaucoup de sites qu'il ne reverra jamais par la suite.

Synthèse. Sur le plan des thèmes et services accédés, une opposition nette se dégage entre deux contextes particuliers de navigation : d'un côté, les sessions routinières, plutôt fermées, qui amènent l'internaute à visiter des sites connus, et de l'autre des sessions ouvertes et exploratoires, où l'utilisateur découvre de nouveaux sites et de nouveaux contenus. Cette découverte suit un mouvement allant du connu vers l'inconnu : les sites habituels constituent le noyau central des territoires personnels sur le Web, et forment le support privilégié de la navigation à double titre. Non seulement ils sont présents dans la grande majorité des sessions, mais en plus, ils occupent souvent la part la plus importante du temps de navigation. La classification des sessions sur la base de la fréquence et de l'empan des sites et des contenus affine cette description et met en évidence deux grands types de sessions : d'un côté, les sessions routinières renvoyant à un territoire restreint fréquemment balayé ; de l'autre, des parcours plus ouverts, décomposés en sessions de semi-routine, aux sites et contenus occasionnels mais réguliers en contexte, et sessions de découverte dont la plupart des sites ne seront jamais revus par la suite.

6.2.2 Navigation routinière et parcours exploratoires

Cette description tripartite en sessions routinières, à contenus occasionnels et tournées vers la découverte recouvre sur le fond l'opposition constatée sur la base de la topologie des parcours entre sessions courtes et linéaires (parcours éclairs et ciblés), et sessions longues et complexes (parcours à détours, à pivots et éclatés). Le croisement de ces deux typologies montre le lien qui existe entre elles, et permet de l'affiner (voir Tableau 6.21 ci-dessous). Au sein des sessions routinières, parcours éclairs et parcours ciblés représentent plus de la moitié des sessions, contre 28 % pour les sessions « contenus occasionnels » et 16 % pour les sessions « découverte » : il y a bien un lien fort entre la visite de contenus nouveaux et l'allongement et la complexification des parcours. Pour autant, même au sein des sessions routinières, les parcours complexes demeurent importants, tandis que dans les sessions « découverte », les parcours à pivots sont majoritaires : ces deux exemples appellent à considérer la combinaison des comportements de navigation et des territoires comme le signe de contextes de navigation différenciés, où plusieurs activités peuvent être imbriquées.

Tableau 6.21. SN00-02 – Topologie des sessions et territoires personnels

	Répartition globale	Territoires sur le Web		
		Sessions routinières	Sessions « contenus occasionnels »	Sessions « découverte »
Parcours éclairs	13,8 %	18,2 %	10,4 %	5,2 %
Parcours ciblés	24,9 %	33,3 %	17,6 %	10,7 %
	} 38,7 %	} 51,5 %	} 28 %	} 15,9 %
Parcours à détours	22,2 %	20,2 %	23,7 %	26,2 %
Parcours à pivots	28,3 %	19,4 %	34,9 %	47,0 %
Parcours éclatés	10,8 %	8,9 %	13,4 %	10,9 %
	} 61,3 %	} 48,5 %	} 72 %	} 84,1 %
<i>Total</i>	100 %	100 %	100 %	100 %

L'examen détaillé des différents groupes va nous permettre d'approfondir ces éléments et de voir comment sont articulés et valorisés au sein des sessions le familier et l'inconnu au sein de cours d'action particuliers.

Navigation routinière

Les deux tiers des parcours ciblés et des parcours éclairs relèvent de sessions routinières, principalement occupées par des contenus visités fréquemment et sur de longues périodes ; réciproquement, les sessions routinières sont composées pour la moitié d'entre elles de parcours ciblés et de parcours éclairs. Au sein des sessions routinières, on distinguera donc deux sous-classes : d'un côté, les parcours routiniers directs, linéaires et rapides, et de l'autre des parcours routiniers complexes amenant l'utilisateur à revoir des pages et des sites au sein de la session. Entre ces deux extrêmes, c'est la pluri-activité qui semble être déterminante : au sein des parcours routiniers, on compte en moyenne 1,5 descriptifs de contenu différents (catégories Nomade et *CatService*) pour les parcours éclairs, contre 4,2 pour les parcours à pivot ou les parcours éclatés. L'examen manuel des sessions le confirme : les sessions routinières directes sont dominées par des contenus à fort renouvellement, ou qui amènent l'utilisateur à y revenir fréquemment. Au premier rang de ceux-ci, on trouve les portails généralistes, souvent ceux de fournisseurs d'accès, les services de messagerie, ainsi que des sites d'informations généraux (services dédiés sur les portails généralistes, ou journaux en ligne comme le Parisien, Libération, le Monde) ou boursière (Boursorama, sites de banques, etc.).

Lorsque la session s'allonge et se complexifie, ces contenus sont soit juxtaposés, soit entrelacés. Dans le premier cas, l'internaute balise son territoire familier sur le Web en passant sur chacun des sites familiers plus ou moins de temps mais toujours en une seule séquence, et n'y revient pas par la suite. La session présentée Figure 6.6 en est un exemple : les sites Club Internet, Le Monde et education.gouv.fr apparaissent dans une grande partie des sessions du panéliste, et sont ici visités séquentiellement.

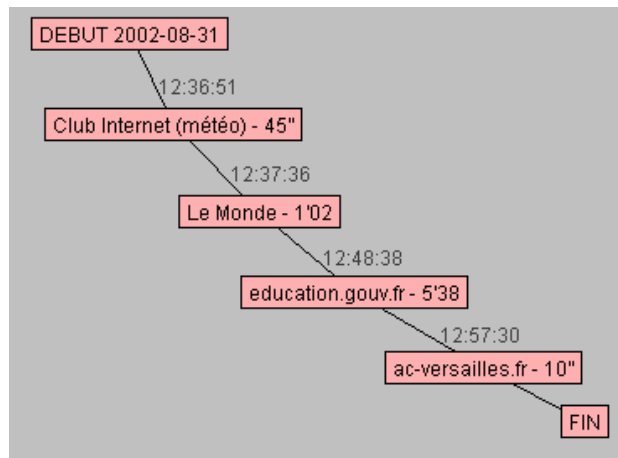


Figure 6.6. SN00-02 – session 107972 (pan 16632)

Ceci n'exclut pas une cohérence entre les différents sites de la session : dans cet exemple, si la visite initiale du service d'informations météorologiques sur Club-Internet semble indépendante du reste de la session, par la suite l'internaute consulte sur le site du Monde un article relatif à l'augmentation du salaire des ministres, puis la page d'accueil de la rubrique « Éducation » où il y suit un lien vers le site de l'Éducation Nationale pour y consulter des informations sur les avantages sociaux réservés aux enseignants.

Dans le cas de l'entrelacement, la navigation peut être alternée et s'apparenter à du multifenêtrage. Toutefois, ce cas reste minoritaire : on observe bien souvent le recours à un site-pivot autour duquel s'organisent et se distribuent les différents contenus visités, qu'ils aient ou non un rapport entre eux ou avec ce site central. Celui-ci est dans la plupart des cas un portail généraliste, dont l'utilisateur exploite le moteur de recherche ou suit les liens pour accéder aux sites qui l'intéressent. Un exemple de cette organisation particulière de la navigation est donné Figure 6.7 et Tableau 6.22 : la rubrique « Informations » de Yahoo fait office de nœud central, même si l'utilisateur n'y passe que peu de temps à chaque fois.

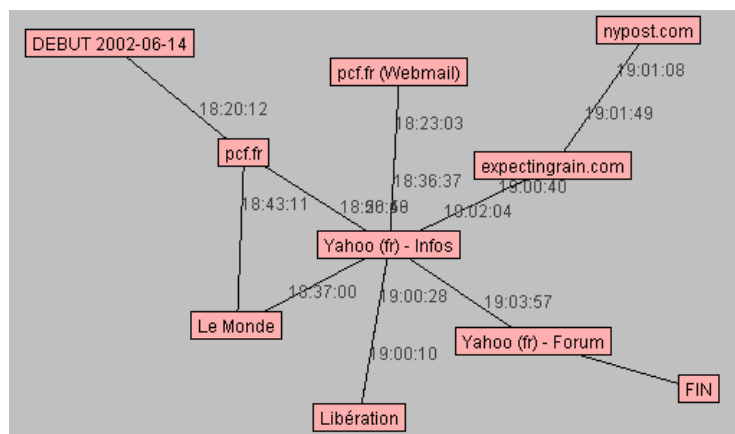


Figure 6.7. SN00-02 – session 48806 (pan 1731)

Tableau 6.22. SN00-02 – session 48806, vue séquentielle

Site / Portail	Date	Nb pages	Durée	Empan	Présence dans les sess. du pan.
www.pcf.fr	18:20:12	1	36"	930	331 (17 %)
Yahoo (fr) - Informations	18:20:48	7	2'15	1031	840 (43 %)
pcf.fr - WebMail	18:23:03	36	13'34	900	404 (21 %)
Yahoo (fr) - Informations	18:36:37	1	23"	1031	840 (43 %)
Le Monde	18:37:00	5	6'11	641	191 (10 %)
www.pcf.fr	18:43:11	28	13'39	930	331 (17 %)
Yahoo (fr) - Informations	18:56:50	11	3'20	1031	840 (43 %)
Libération	19:00:10	2	18"	641	319 (16 %)
Yahoo (fr) - Informations	19:00:28	1	12"	1031	840 (43 %)
expectingrain.com	19:00:40	1	28"	921	172 (9 %)
nypost.com	19:01:08	3	41"	0	1 (0 %)
expectingrain.com	19:01:49	1	15"	921	172 (9 %)
Yahoo (fr) - Informations	19:02:04	5	1'53	1031	840 (43 %)
Yahoo (fr) - Forum	19:03:57	20	1'01	488	102 (5 %)

Dans ce type de configuration, il n'est pas rare que l'internaute « déborde » de ses contenus habituels, mais cet écart demeure ponctuel et borné à un seul site, et occupe peu de temps dans la session. Ces écarts correspondent bien souvent au suivi de liens proposés dans les sites habituels : leur contenu est déjà décrit dans la page source, et ces éléments de découverte de nouvelles pages et de nouveaux sites sont motivés et maîtrisés, et ne donnent pas lieu à de longues digressions. Dans l'exemple ci-dessus, le site du journal nord-américain New York Post (nypost.com) est vu ici pour la première et unique fois par le panéliste, et correspond à un lien présent dans une page du site expectingrain.com.

En définitive, il apparaît que lors de ces balayages récurrents des terrains connus, les internautes se connectent à des flux. Ce peut être des flux d'information – informations générales, boursières, sites spécialisés – ou des flux de communication – WebMail, forums, *chat* ; dans les deux cas, le mode de réception des contenus ne semble pas si éloigné de celui des médias traditionnels : une fois que l'internaute a identifié les canaux qui l'intéressent, c'est par ce biais qu'il « consomme » les contenus, comme il le ferait pour des émissions de télévision ou de radio, ou des quotidiens et des magazines. La différence réside ici d'une part dans la disponibilité des contenus Web et la possibilité de les mobiliser au moment qui convient le mieux à l'utilisateur, et d'autre part dans l'ouverture que ces contenus routiniers offrent vers l'inhabituel *via* l'hypertexte.

L'occasionnel : régularité en contexte

L'appréhension de contenus occasionnels est différente, et implique des parcours plus complexes : on a pu constater que les parcours à pivots et à détours sont plus présents dans ce cas (voir Tableau 6.21, p. 258) que pour les contenus habituels. Cependant, pour accéder à ces contenus peu fréquents, l'internaute a peu recours aux moteurs de recherche : seul un tiers des sessions « contenus occasionnels » comportent un accès à un moteur, contre 12 % pour les sessions routinières, et 60 % pour les sessions « découverte » (voir Tableau 6.23 ci-dessous).

Tableau 6.23. SN00-02 – Part des sessions avec moteur par type de session

	Sessions routinières	Sessions « contenus occasionnels »	Sessions « découverte »
Parcours éclairés	3,5 %	8,5 %	20,6 %
Parcours ciblés	5,3 %	11,2 %	31,6 %
Parcours à détours	13,8 %	27,8 %	56,4 %
Parcours à pivots	23,5 %	42,3 %	72,6 %
Parcours éclatés	22,9 %	35,8 %	62,5 %
Ensemble	11,8 %	29 %	60,2 %

Il s'agit ici de sites connus et visités plusieurs fois par le panéliste, mais dans des contextes qui se produisent rarement. Le contenu de ces sites est variable selon les intérêts de chaque internaute ; toutefois, de manière générale, on retrouve fréquemment dans ce genre de configurations des activités d'achat en ligne ou d'organisation de voyages. En outre, lorsqu'un moteur de recherche est mobilisé dans ce contexte, il s'agit bien souvent de requêtes ponctuelles correspondant à la recherche de l'adresse d'un site connu, par exemple « fnac », « alapage », « nouvelles frontières », forme d'utilisation des moteurs comme pseudo-*bookmarks*.

L'exemple de session présenté Figure 6.8 et Tableau 6.24 ci-dessous illustre ces éléments : l'internaute cherche visiblement à connaître les films qui passent dans les cinémas près de chez lui. Pour cela, il va dans un premier temps sur le service dédié du portail Yahoo ; après cette recherche infructueuse, il visite les catégories « Art et culture → Cinéma » et « Art et culture → Cinéma → Annuaires et guides » dans Yahoo, puis interroge le moteur de recherche avec la requête « allocine ». Cet usage du moteur s'apparente à un pseudo-signet : l'internaute connaît l'existence du site et souhaite y aller, mais il n'en a pas l'adresse. Après cela, une longue séquence de plus de 7 minutes sur allocine.fr conduit l'utilisateur à trouver les horaires des films pour les salles près de son domicile, et il poursuit sa recherche en visitant le site dédié à un film particulier figurant au programme.

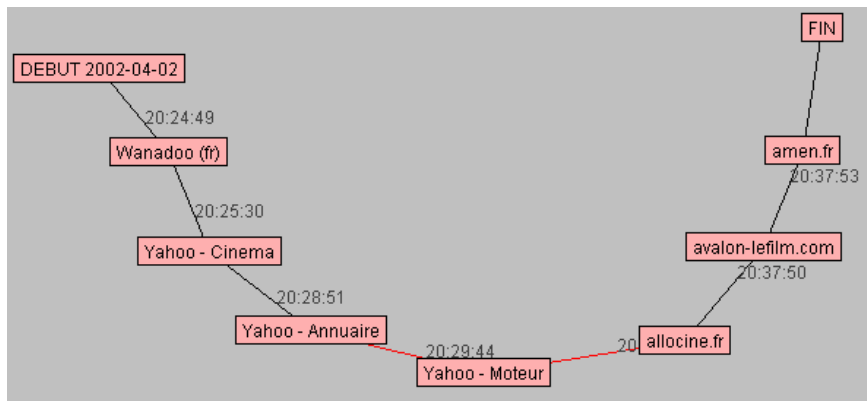


Figure 6.8. SN00-02 – session 328 (pan. 4)

Tableau 6.24. SN00-02 – session 328 (pan. 4), vue séquentielle

Site / portail	Date	Nb pages	Durée	Empan	Présence dans les sess. du pan.
Wanadoo (fr)	20:24:49	6	41"	818	109 (50 %)
Yahoo - cinéma	20:25:30	15	3'21	106	5 (2 %)
Yahoo - annuaire	20:28:51	2	55"	816	14 (6 %)
Yahoo - moteur	20:29:44	1	31"	791	11 (5 %)
allocine.fr	20:30:15	21	7'35	865	7 (3 %)
avalon-lefilm.com	20:37:50	1	3"	0	1 (0 %)
webserveur91.amen.fr/avalon	20:37:53	11	11'57	0	1 (0 %)

Dans un autre registre, la session présentée ci-dessous (Figure 6.9 et Tableau 6.25) est structurée autour de la recherche de billets d'avion : plusieurs sites de voyagistes sont visités, qui ont pour la plupart un empan très important bien qu'ils n'apparaissent que dans peu de sessions (Travelprice, Degrifour, Lastminute, Anyway). Pour y accéder, l'internaute n'utilise pas un moteur de recherche : soit il connaît assez bien l'adresse pour l'entrer dans la barre d'adresse du navigateur, soit il l'a stockée dans ses Favoris. Dans les deux cas, l'internaute est en terrain connu pour la tâche qui l'occupe, et il parcourt les différentes offres dans une logique de comparaison.

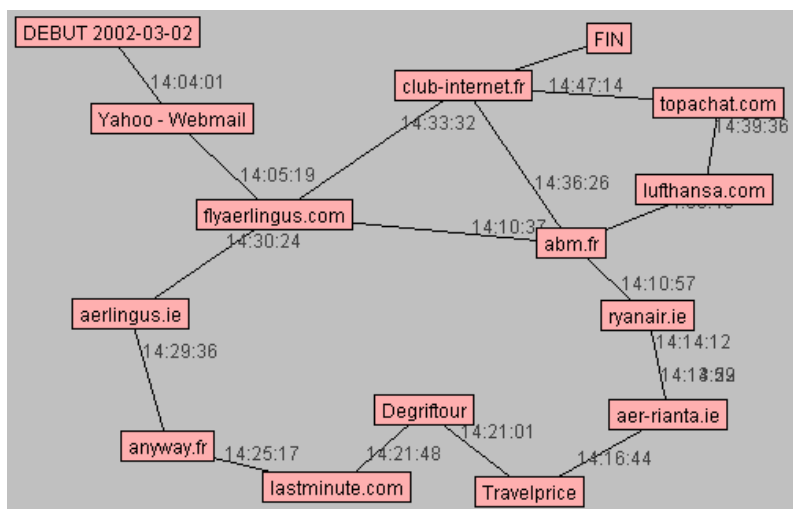


Figure 6.9. SN00-02 – session 17512 (pan. 334)

Dans ce type de parcours, l'appréhension des contenus Web est plus proche d'un format « guichet », « pages jaunes » ou « galerie marchande » : les activités induites sont plutôt rares pour chaque internaute, mais celui-ci adopte des comportements relativement stables dans ce contexte, et visite souvent les mêmes sites ou groupes de sites.

Tableau 6.25. SN00-02 – session 17512, vision séquentielle

Site	date	Nb Pages	Durée	Empan	Présence dans les sess. du pan.
Yahoo - WebMail	14:04:01	1	1'18	774	677 (47 %)
flyaerlingus.com	14:05:19	9	5'18	8	4 (0 %)
abm.fr	14:10:37	3	20"	0	2 (0 %)
ryanair.ie	14:10:57	19	3'02	0	1 (0 %)
aer-rianta.ie	14:13:59	5	13"	0	1 (0 %)
ryanair.ie	14:14:12	1	10"	0	1 (0 %)
aer-rianta.ie	14:14:22	1	2'22	0	1 (0 %)
Travelprice	14:16:44	8	4'17	799	12 (1 %)
Degriftour	14:21:01	2	47"	799	18 (1 %)
lastminute.com	14:21:48	1	3'31	839	15 (1 %)
anyway.fr	14:25:17	1	4'19	663	16 (1 %)
aerlingus.ie	14:29:36	2	48"	4	4 (0 %)
flyaerlingus.com	14:30:24	4	3'08	8	4 (0 %)
Club Internet	14:33:32	1	2'54	1006	319 (22 %)
abm.fr	14:36:26	3	22"	0	2 (0 %)
lufthansa.com	14:36:48	5	2'48	0	1 (0 %)
topachat.com	14:39:36	15	7'38	960	91 (6 %)
Club Internet	14:47:14	1	32"	1006	319 (22 %)

Corrélativement, les thèmes et services les plus visités dans ce contexte sont orientés vers le Web marchand et les activités hors Web (voir Tableau 6.26) : les catégories Nomade « Mes Courses » et « Vie pratique » sont sur-représentées dans ces sessions à contenus occasionnels, de même que les sites pornographiques.

Tableau 6.26. SN00-02 – Thèmes et services des sessions « contenus occasionnels » (Nomade)

Catégorie dominante dans la session	Sessions « contenus occasionnels »	Ensemble des sessions
Mes Courses	12,5 %	7,2 %
Société, Vie pratique	11,6 %	9,6 %
PORNO	10,2 %	5,9 %
Sport et détente	9,6 %	7,3 %
Espace B to B	8,9 %	11,3 %
Actu, médias	8,3 %	5,0 %
CS – WebMail	6,1 %	12,7 %
Nouvelles technologies	5,0 %	2,8 %
CS – Moteur	4,0 %	3,7 %
Culture et loisirs	3,7 %	3,0 %
PG - Page Accueil	3,4 %	12,5 %
Voyage, géographie	3,4 %	2,1 %
Autres	13,5 %	16,9 %

Clef de lecture : dans 12,5 % des sessions « contenus occasionnels », la catégorie Nomade « Mes Courses » occupe le plus de temps dans la session.

La démarche comparative de l'utilisateur induit un allongement global des sessions et une augmentation du nombre de sites visités ; ce phénomène peut être atténué lorsqu'un site connu répond aux attentes de l'utilisateur, comme pour celui

de la SNCF pour les horaires et billets de train, pagesjaunes.fr pour les fonctions d'annuaire, ou service-public.fr pour les démarches administratives. À l'inverse, les sites de vente en ligne de produits touristiques, culturels ou technologiques amènent souvent l'internaute à visiter plusieurs sites différents dans un même champ d'activité.

Prédateur vs. flâneur : usages différenciés des moteurs de recherche

La présence dans les sessions de sites occasionnels et de faible empan dénote un tout autre usage du Web, apparenté à une ressource d'informations ponctuelle mobilisée pour des besoins précis et limités dans le temps. Dans ce contexte, les moteurs de recherche sont hautement mis à contribution : on a pu voir que six sessions « découverte » sur dix contiennent une requête moteur, et que ces sessions sont pour 85 % d'entre elles des parcours complexes, avec la moitié de parcours à pivots et un quart de parcours à détours (voir Tableau 6.21, p. 258).

C'est au sein de ces sessions longues et complexes que l'utilisateur découvre de nouveaux sites, ouvre de nouveaux territoires, explore des contenus inédits ; pour autant, l'usage majoritaire des moteurs dans ce contexte nous rappelle qu'il ne s'agit pas là d'une navigation au gré du vent et des liens hypertexte, mais d'une recherche sur un sujet précis, guidée par un objectif explicite par l'utilisateur. Au sein de ces parcours ouverts, on distingue deux types de comportements, que l'on qualifiera de « prédateur » et de « flâneur » : dans le premier cas, l'internaute cherche des informations précises et ciblées sur un thème donné, et feuillette rapidement les différentes ressources qu'il trouve pour aller au plus vite au résultat qui l'intéresse. Dans l'autre cas, les détours sont plus nombreux, les requêtes plus ouvertes, et l'internaute semble de manière générale disposer de plus de temps pour visiter les « à côté ». Au cours des entretiens menés avec les panélistes de BibUsages, ces éléments de différenciation étaient soulignés par les interviewés :

« Je n'utilise jamais Kartoo comme moteur de recherche au niveau professionnel, j'utilise Wanadoo, Google, Voila, et puis c'est tout. Et pour mes recherches personnelles, ça va être Kartoo par exemple parce que je ne sais pas exactement comment formuler ma recherche. Au niveau professionnel, le fil directeur, c'est l'efficacité ; c'est-à-dire qu'il faut assurer une rentabilité ou une mesure de son travail qui soit réelle. Donc une recherche personnelle, c'est une somme et une variété de documentation qui vont permettre de bâtir quelque chose autour d'un concept ou d'une idée. C'est de la curiosité pure ; mais ce n'est pas sanctionné de la même manière. » (Utilisateur H)

On retrouve ces deux types de comportements dans les données au gré de l'examen manuel des sessions. Deux exemples illustreront cette dichotomie : dans le premier, l'internaute effectue une recherche sur un sujet précis, les encres fiduciaires. Il est amené à reformuler sa requête de nombreuses fois, et à consulter de nombreuses pages de résultats : le Tableau 6.27, qui présente l'ensemble des mots-clés recherchés sur les différents moteurs dans cette session, montre les stratégies de précision (ajout de mots), de déplacement (« encres petrel » modifié en « encres fiduciaires » puis en « encres gommables »), de sophistication (ajout de guillemets) ou de changement d'outil (passage de Google à Altavista). Au total, ce sont trente-

quatre pages de résultats consultées dans la session qui amènent l'internaute à visiter quinze sites différents relatifs à sa recherche.

Tableau 6.27. Requêtes moteur pour la session 41772 (SN00-02)

Date	Moteur	Nb pages de résultats	Requête
22:31:05	Google	2	petrel
22:32:03	Google	3	encres pétrel
22:34:58	Google	2	encres fiduciaires pétrel
22:35:11	Google	10	encres fiduciaires
22:47:40	Google	1	encres fiduciaires pétrel
22:47:44	Google	2	encres fiduciaires
22:48:15	Google	2	encres petrel
22:50:19	Altavista.fr	1	"encres petrel"
22:50:28	Altavista.fr	6	petrel
22:53:11	Google	1	encres gommables et delebiles
22:53:17	Google	3	encres gommables
22:53:55	Google	1	encres delebiles

La forme du parcours est en ce cas centrée sur l'outil de recherche : l'internaute ne sort pas des sites proposés dans les résultats du moteur, et bien souvent ne consulte que la page indiquée par le moteur, sans naviguer dans le reste du site (voir Figure 6.10). Dans l'exemple, la séquence de recherche occupe la quasi-totalité de la session : vingt-quatre minutes sur la demi-heure que dure la session, les six dernières minutes étant consacrées à une toute autre activité, la consultation de la météo pour des stations de sports d'hiver.

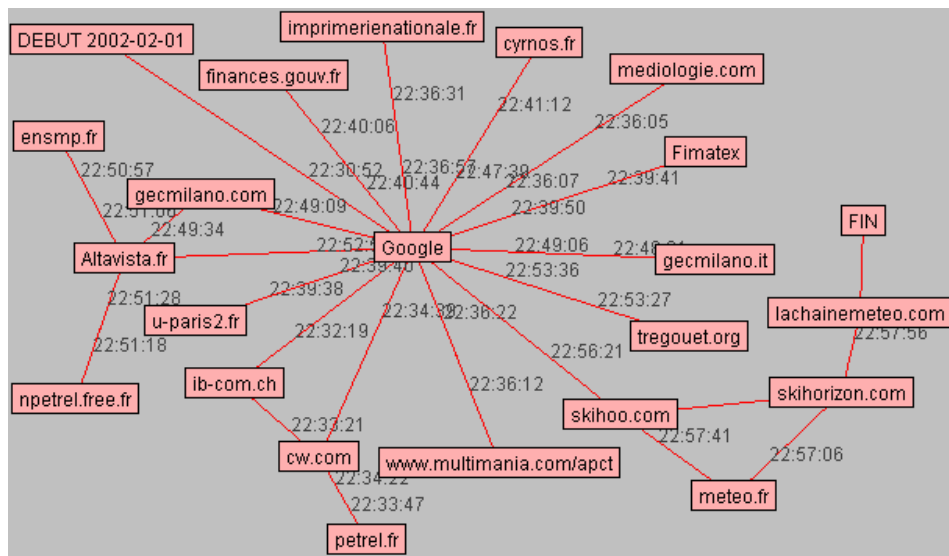


Figure 6.10. SN00-02 – session 41772 (pan 1291)

Ce premier exemple témoigne de comportements « prédateurs » : l'utilisateur a une idée relativement précise de ce qu'il cherche, dans un domaine restreint, et il a la

capacité de préciser ou de modifier le contenu de sa requête pour obtenir le résultat désiré.

Ce type de parcours s'oppose à des recherches plus ouvertes, qui embrassent une plus grande diversité, et où la succession des mots-clés ouvre et complète le champ de recherche au lieu de le restreindre. L'exemple de session donné Figure 6.11 en est une illustration : d'une durée totale d'une heure, cette session est axée autour de trois recherches successives – « la Tranche-sur-Mer », « île de Ré » et « île de Noirmoutier ».

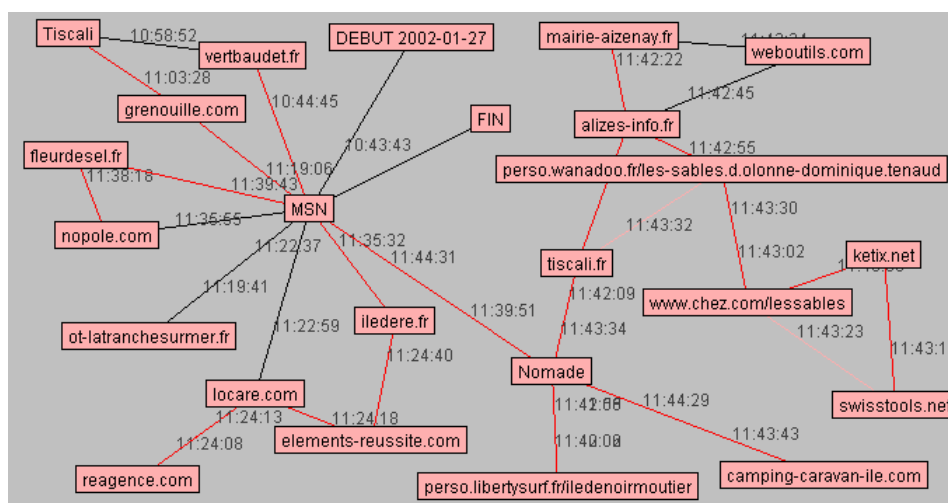


Figure 6.11. SN00-02 – session 7310 (pan 156)

Dans cette session, les requêtes ne sont pas reformulées, et peu de pages de résultats sont consultées : tout au plus l'utilisateur essaye sa dernière requête également sur Nomade après avoir utilisé MSN jusqu'alors. L'allongement et la complexification de la session vient d'un temps plus important accordé aux pages de résultats de recherche : visiblement intéressé par les éléments touristiques des trois lieux, l'internaute navigue sur les sites proposés par les moteurs, examine les listes d'hôtels et leur site le cas échéant, consulte les plans, les descriptifs des villes, etc.

Ce comportement de type « flâneur » est surtout observable dans les requêtes moteur visant des contenus pornographiques. Les mots-clés sont dans ce cas souvent vagues et généraux : « sexe », « porno », « gratuit », « pics » ou encore « xxx » sont très souvent employés et combinés dans ce contexte, amenant à naviguer longuement de site en site à partir des pages de résultat des moteurs. Hors de ce champ, la navigation au gré des liens est globalement assez rare : le fonctionnement même des outils de recherche incite l'utilisateur à préciser autant que possible ses requêtes, de sorte que c'est vers les sites eux-mêmes que se déplace la demande des utilisateurs en matière de liens pertinents. La difficulté n'est plus tant de trouver les bons moteurs, mais de trouver les bons sites sur un thème donné, qui fourniront à la fois des contenus de qualité et des ouvertures pertinentes. Le comportement des internautes de BibUsages sur le site de Gallica est révélateur à cet égard : le moteur de recherche interne est largement majoritaire dans l'accès aux documents, mais les

« Dossiers », visites guidées thématiques, sont occasionnellement consultés, lorsque l'internaute dispose de plus de temps et de curiosité.

Signification des contenus à l'aune des pratiques

En définitive, l'analyse de la forme des parcours et de leurs contenus en regard des territoires personnels sur le Web nous amène à distinguer trois types d'activité, auxquelles correspondent trois modes d'appréhension et d'interprétation des contenus. Les actions routinières valorisent les contenus de type « flux d'information » ou « flux de communication », au sein de sessions plutôt courtes où se juxtaposent les différents sites habituels ; ce territoire familier est régulièrement balisé par l'internaute, et sous-tend l'ensemble de sa navigation. Les contenus occasionnels correspondent au Web marchand ou de service, et à un mode d'usage de type « guichet de renseignements » ; ce contexte, plutôt rare, amène l'utilisateur à visiter un nombre restreint de sites connus, qui sont balayés au cours de la session selon un mode de feuilletage orienté vers la comparaison. Enfin, les parcours ouverts, qui amènent à découvrir de nouveaux sites, font massivement appel aux moteurs de recherche, mais sont orientés vers la résolution d'un problème ou d'une question particulière. Dans ce dernier cas, le Web n'est pas tant considéré comme un espace documentaire que comme une collection de sources possibles au sein desquelles il convient de trouver la plus complète et la plus fiable pour une interrogation ponctuelle.

Ces trois schémas d'action prototypiques permettent de décrire la majorité des activités de navigation sur le Web. Ils impliquent des formes de parcours et des rythmiques particulières, et conditionnent l'appréhension des sites et des contenus : un même site pourra apparaître dans ces trois contextes, mais s'insérer dans des structures interprétatives différentes. C'est ce que nous avons pu voir avec les moteurs de recherche, où l'outil peut être utilisé comme pseudo-signet ou comme outils de découverte de ressources. De manière générale, certains sites sont plus sujets aux utilisations multiples tandis que d'autres orientent vers des schémas actantiels fermés : le WebMail est un outil de communication fermé, tandis qu'un forum pourra être perçu par les uns comme une ressource pour un besoin ponctuel et accédé via une requête moteur, et par d'autres comme un flux impliquant une visite routinière. Autre exemple, les sites de vente en ligne impliquent globalement des comportements de type « galerie marchande », mais on a pu voir que dans certains cas, ils servent de réservoir à images pour la confection de CD récupérés *via* des réseaux de *peer-to-peer*.

Il est probable que l'évolution actuelle de l'offre en matière d'outils de publication sur le Web soutienne cette tendance : les systèmes de gestion de contenu¹ permettent à tout un chacun d'installer facilement des outils de publication collaborative, des forums, des blogs, etc. qui transforment un site statique en flux continu. Cet aspect protéiforme des contenus Web, que l'on peut considérer tantôt comme ressource documentaire, tantôt comme canal de communication et d'information, tantôt comme fournisseur de services, montre en définitive que

¹ Ou « CMS » : Content Management System.

textes, documents et outillages accessibles sur la Toile n'ont de signification que par le biais du parcours qu'y trace l'utilisateur, et du cours d'action dans lequel s'inscrit sa pratique.

Synthèse. En croisant la typologie des sessions fondée sur l'approche topologique et temporelle, et celle basée sur les territoires personnels sur le Web, trois comportements prototypiques de navigation sont mis à jour, qui reflètent des contextes d'usage différenciés. En premier lieu, les sessions routinières, liées aux parcours éclairs et ciblés, s'apparentent à des balayages récurrents des terrains connus, où les internautes se connectent à des flux d'information (informations générales, boursières, sites spécialisés) ou de communication (WebMail, forums, chat) ; dans les deux cas, le mode de réception des contenus ne semble pas si éloigné de celui des médias traditionnels. Ensuite, dans les parcours orientés « Web pratique », plus complexes et allongés, l'appréhension des contenus est plus proche d'un format « guichet », « pages jaunes » ou « galerie marchande » : les activités induites sont plutôt rares pour chaque internaute, mais celui-ci adopte des comportements relativement stables dans ce contexte, et visite souvent les mêmes sites ou groupes de sites. Enfin, les parcours ouverts, qui amènent à découvrir de nouveaux sites, font massivement appel aux moteurs de recherche, dans le cadre de requêtes précises. Dans ces trois modes prototypiques d'appréhension du Web, les comportements de navigation sont profondément influencés par la double dynamique de la familiarité territoriale et du contexte d'usage.

6.3 Le document numérique, l'usuel et l'œuvre¹

Dans la littérature sur les hypertextes, il est souvent fait implicitement référence à l'appréhension des contenus en contexte de recherche : en Recherche d'Information, le Web est considéré comme un réservoir d'informations et de connaissances non structuré ; dans [Ghitalla *et al.* 2003], il est envisagé comme un espace documentaire. S'il est réducteur de considérer le Web uniquement sous cet angle, cette approche est pertinente dans le cadre de son utilisation en contexte de recherche ciblée, mais elle mérite d'être revisitée à l'aune des différentes modalités d'appréhension des contenus que nous avons mises à jour. Le panel BibUsages, constitué de chercheurs et de consommateurs de fonds textuels numérisés, nous permet d'étudier plus avant ces parcours ouverts de type « recherche » et de voir comment l'accès numérique à des ressources traditionnellement imprimées en renouvelle l'appréhension.

¹ Cette partie reprend des éléments du rapport du projet BibUsages ; voir [Assadi *et al.* 2003a].

6.3.1 Internauts lecteurs, internautes chercheurs

Nous avons déjà eu l'occasion de souligner les spécificités du panel BibUsages en ce qui concerne les contenus visités et les types de sessions, plutôt orientées vers la recherche. Nous approfondissons ici cette description, pour voir la place qu'occupent les contenus culturels et les sessions de recherche chez ces internautes, et comment s'articulent les différents contenus visités dans le contexte de la session.

Intérêt prononcé pour les contenus culturels

Dans les réponses au questionnaire en ligne soumis en mars 2003 aux visiteurs de Gallica, nous pouvions déjà observer que les « gallicanautes » constituent une population de lecteurs particulièrement intéressés par les contenus culturels et l'actualité en ligne (voir Tableau 6.28). Parmi les centres d'intérêt déclarés, la quasi-totalité des répondants mentionne des contenus de type culturel ou académique : « Art et littérature » pour 62 % des répondants, « Recherche documentaire » pour les trois quarts ; les sites d'actualités intéressent également la moitié des interrogés, tandis que les sphères ludiques (jeux, sport) et économiques (information sur les entreprises, services bancaires) ou les services de communication restent minoritaires.

Tableau 6.28. Questionnaire BibUsages en ligne : centres d'intérêt sur le Web

Quels sont vos principaux centres d'intérêt sur le Web ?	
Contenus culturels, dont :	98,0 %
Recherche documentaire ou bibliographique	76,8 %
Art et Littérature	62,4 %
Sciences Humaines et sociales	50,7 %
Informatique et multimédia	31,2 %
Sciences et technologies	30,5 %
Autres informations culturelles	31,9 %
Informations économiques ou institutionnelles, dont :	64,1 %
Actualités	47,3 %
Institutions et service public	25,8 %
Banque et finances	13,3 %
Emploi, Stage (recherche ou offre)	10,3 %
Économie et entreprise	9,1 %
Autres informations économiques ou institutionnelles	10,3 %
Loisirs, dont :	44,2 %
Voyages, tourisme	27,2 %
Sorties, divertissements	16,9 %
Jeux	7,3 %
Sports	7,1 %
Autres Loisirs	7,1 %
Communication	21,3 %
Autres centres d'intérêt	10,4 %

L'identification des différents types de portails et de services avec *CatService* permet d'observer dans les données de trafic ces éléments de manière plus fine, et nous montre que les « gallicanautes » constituent un véritable panel d'internautes-lecteurs et d'internautes-chercheurs (voir Tableau 6.29).

Tableau 6.29. Audience par types de sites et de portails (juillet – décembre 2002)

Type de portail (<i>CatService</i>)	Présence dans les sessions BibUsages	Nombre de panélistes de BibUsages	Présence dans les sessions SN2002
Portail généraliste	60,3 %	71	55,2 %
Moteur	30,1 %	67	20,2 %
Site « perso »	21,1 %	69	15,7 %
WebMail	17,6 %	15	22,4 %
Bibliothèque électronique	8,0 %	59	0,2 %
Media / Presse	7,6 %	49	2,7 %
e-commerce / biens culturels	5,5 %	63	3,1 %
Généalogie	4,6 %	32	0,3 %
e-commerce / Banque Bourse	2,9 %	30	12,3 %
Media / Radio	2,0 %	33	1,1 %
Media / TV	1,5 %	38	2,5 %
e-commerce / tourisme	0,9 %	46	1,5 %
e-commerce / courses	0,1 %	4	0,2 %

En premier lieu, portails généralistes et moteurs de recherche occupent une place prépondérante dans l'activité Web, et touchent la quasi-totalité du panel : les moteurs de recherche entrent en tête des sites les plus visités et les plus présents dans les pratiques. On les retrouve dans 30 % des sessions du panel, contre 20 % pour le panel SN2002. Cet élément est confirmé dans les entretiens, où beaucoup d'interviewés évoquent des pratiques de recherche d'information au premier rang de leur activité sur le Web. D'autre part, le panel se caractérise par une très forte fréquentation des portails « culturels » : les bibliothèques électroniques, les sites de commerce de biens culturels (Alapage, Fnac, Amazon...) et les sites de médias (presse, radio, télévision) occupent une place privilégiée dans les pratiques. Le détail de la fréquentation des portails de vente de biens culturels révèle en particulier l'importance de la bibliophilie, avec une part non négligeable du trafic sur des sites comme Chapitre.com, Livre-rare-book, Galaxidion ou Librissimo (voir Tableau 6.30).

Tableau 6.30. Fréquentation des sites de vente de biens culturels

Portail	Nombre de panélistes	Nombre de sessions	Nombre de sessions par panéliste	Temps moyen dans la session (min.)
Amazon	55	386	7,0	4 min. 06
Fnac	42	299	7,1	2 min. 25
Alapage	29	107	3,7	2 min. 13
Chapitre.com	26	160	6,2	9 min. 17
Livre-rare-book	17	92	5,4	5 min. 02
Galaxidion	6	73	12,2	6 min. 24
Librissimo	6	7	1,2	4 min. 31
Numilog	5	14	2,8	4 min. 42
Eyrolles	3	4	1,3	1 min. 03
CNRS Éditions	2	8	4,0	8 min. 37
<i>Autres (cinq portails)</i>	2	2	1	0 min. 20

Enfin, les sites consacrés à la généalogie ont une place importante dans les usages, et leur présence témoigne de centres d'intérêt personnels très forts ; il faut sans doute ajouter à cela l'audience particulière des sites personnels, présents dans 21% des sessions, qui sont d'importantes sources d'informations pour les recherches généalogiques sur le Web.

Cette première vue montre un lien global entre la fréquentation des bibliothèques électroniques et les outils de recherche d'information d'une part, et les sites de « contenus à lire » et de commerce de biens culturels d'autre part. Les consommateurs de bibliothèques numériques sont de manière générale de grands consommateurs de produits culturels, que ce soit dans une optique de consultation ou de consommation.

Audience des bibliothèques électroniques en contexte

Au sein de la plateforme *CatService*, vingt-trois sites ont été identifiés comme relevant de la catégorie « bibliothèques électroniques », parmi lesquels dix-huit ont été visités par le panel entre juillet et décembre 2002. Nous avons retenu une acception large de cette catégorie en y incluant, outre les versions numériques des bibliothèques patrimoniales, des sites d'éditeurs, des bibliothèques numériques régionales ou municipales, ainsi que des collections numériques issues d'initiatives associatives. Nous constituons ainsi une catégorie assez vaste de « collections de textes d'ouvrages accessibles en ligne ». L'audience détaillée de ces différents sites montre que Gallica arrive en tête, en nombre de sessions comme de panélistes, mais que la visite des autres sites est loin d'être anecdotique (voir Tableau 6.31). Si la richesse du fonds explique la fréquentation plus forte du fond numérisé de la BnF, on peut penser à la lumière des entretiens que Gallica figure pour les internautes comme une source de textes parmi d'autres.

Tableau 6.31 - Audience des bibliothèques électroniques

Portail	Nombre de sessions	Nombre de panélistes	Temps moyen dans une session
BNF-Gallica	1039	59	25 min. 31
BNF-Autres	647	52	6 min. 37
ABU	39	15	2 min. 56
Bibliothèque de Lisieux	26	13	4 min. 03
Revue.org	25	13	2 min. 05
Athena	16	11	1 min. 48
Bibliopolis	16	6	7 min. 36
ClicNet	15	9	40 sec.
CNUM	13	4	6 min. 24
Mozambook	13	3	2 min. 56
Electronic Text Center	11	5	2 min. 13
Online Books Page	10	4	2 min. 05
INALF	8	2	5 min. 40
Autres (moins de 5 sessions)	24	10	2 min. 50

L'examen des types de sites et de services visités au sein des sessions comportant un accès à une bibliothèque électronique permet de voir comment l'usage des fonds numériques s'articule avec les autres contenus et services disponibles sur le Web, et

permet d'en cerner plus finement les contextes d'usage. Pour cela, nous comparons la présence des différents types de sites identifiés dans *CatService* au sein des sessions avec bibliothèque électronique et dans l'ensemble des sessions (voir Tableau 6.32).

Tableau 6.32 - *Présence des différents types de sites et services dans les sessions avec accès à une bibliothèque électronique et dans l'ensemble des sessions*

	Présence dans les sessions bib. élec.	Présence dans l'ensemble des sessions	Variation de la présence dans les sessions bib. élec.
Bibliothèque électronique	100,0 %	8,0 %	
e-commerce / biens culturels	10,4 %	5,5 %	+ 88,2 %
Sites personnels	29,6 %	21,1 %	+ 40,8 %
Moteur	40,8 %	30,1 %	+ 35,7 %
Généalogie	5,2 %	4,6 %	+ 13,4 %
e-commerce / tourisme	1,0 %	0,9 %	+ 7,7 %
Media / Radio	2,0 %	2,0 %	- 0,8 %
Portail généraliste	55,0 %	60,3 %	- 8,9 %
Media / Presse	5,5 %	7,6 %	- 27,6 %
e-commerce / Banque Bourse	1,8 %	2,9 %	- 37,2 %
Media / TV	0,8 %	1,5 %	- 43,0 %

Les moteurs de recherche sont sur-représentés dans les sessions Bibliothèques électroniques, où ils sont 1,8 fois plus présents que dans l'ensemble des sessions. Cet usage fort des moteurs montre l'importance de l'utilisation des bibliothèques électroniques dans un contexte de recherche d'information, où le fonds numérisé est une source parmi d'autres pour trouver de l'information.

Il faut sans doute ajouter à cela les sites personnels, sur-représentés dans les sessions avec Gallica, et qui s'imposent comme sources de données sur des sujets pointus, comme la généalogie. On retrouve cet engouement à travers des requêtes sur les moteurs de recherche très particulières et précises souvent adressées par un seul utilisateur, et qui consistent en des noms propres (patronymes, noms de lieu). La généalogie et le régionalisme s'imposent ainsi comme des centres d'intérêt prépondérants pour une part importante du panel, et Gallica apparaît comme une source d'information parmi d'autres pour ce type de recherches.

L'usage des sites de e-commerce de biens culturels, également corrélé à l'accès aux fonds numérisés, semble correspondre à un effet de catalogue ou de « test avant achat », ce que confirment les entretiens. Nous sommes ici probablement en présence d'une manifestation des liens vertueux pouvant exister entre Web marchand et non-marchand : les bibliothèques numériques permettent de consulter des *fac simile* d'ouvrages anciens, qu'il s'agit ici de feuilleter avant de les acquérir sur un site spécialisé de bibliophilie¹.

¹ On rejoint ici [Gensollen 1999], qui posait dès 1999, à propos de la création de valeur sur Internet, que « la partie non marchande du Web joue sans doute un rôle crucial dans l'économie du système » (p. 19).

En revanche, les sites de type « Média – presse » sont moins présents dans les sessions Bibliothèques électroniques que dans l'ensemble : si les utilisateurs des bibliothèques électroniques sont, comme nous l'avons déjà vu, globalement de gros consommateurs de journaux en ligne, les accès à ces deux types de sites ne correspondent pas aux mêmes pratiques, et se font dans des sessions et des contextes différenciés. Ce résultat illustre bien l'apport d'une analyse fine en termes de sessions, rendue possible grâce à la technologie de recueil de trafic sur le poste de l'utilisateur mise en œuvre dans le projet BibUsages ; en effet, l'analyse de l'audience globale a montré que les usages des bibliothèques électroniques et des sites de type « média / presse » étaient globalement liés, en ce qu'ils sont particulièrement importants pour la population étudiée. L'analyse fine à l'échelle des sessions apporte une information complémentaire importante : même si, globalement, la population étudiée a un intérêt et un usage fort pour ces deux types de sites, les contextes d'usage au sein de sessions de navigation sont bien distincts.

Navigation sur Gallica

Sur les 17 000 sessions enregistrées entre juillet et décembre, 6,1 % comportent un accès à Gallica, qui concernent un nombre important d'utilisateurs : 59 des 72 panélistes de BibUsages sont allés au moins une fois sur Gallica, soit huit personnes sur dix.

L'intensité d'usage de Gallica répond à une répartition similaire à celle du Web : un tiers des visiteurs de Gallica fait plus de 77% des sessions comportant un accès au site, tandis qu'un autre tiers fait seulement 4% de ces sessions. Plus encore, les utilisateurs les plus intensifs de Gallica sont aussi des utilisateurs intensifs du Web : sur la période, ils comptent 400 sessions en moyenne, pour une moyenne de 270 sessions au total pour les autres visiteurs de Gallica, et 150 sessions pour ceux qui n'y sont jamais allés. Ainsi, par son mode de recrutement à partir d'un questionnaire affiché sur le site Gallica, le panel est constitué d'utilisateurs intensifs de ce site, et ce sont eux qui tirent le trafic Web général du panel vers le haut sur l'ensemble de la période, sur Gallica comme sur les autres sites.

Les 1 039 sessions comportant un accès à Gallica sont globalement plus longues que les autres : 1 h 01 min. en moyenne, contre 28 minutes pour les autres sessions. Par ailleurs, dans une session comportant un accès à Gallica, le temps total passé sur ce site proprement dit est en moyenne de 25 minutes, soit presque la durée moyenne d'une session sans Gallica. La visite de la bibliothèque numérique de la BnF s'apparente ainsi à une activité de longue durée.

Plus encore, la consultation de Gallica s'impose comme une activité excluant la visite alternée d'autres sites. Nous avons analysé, dans les sessions, l'alternance de la visite des différents sites, et constaté que dans plus de la moitié des cas, la navigation sur Gallica occupe une seule séquence, elle n'est pas alternée avec la visite d'un autre site ; et pour 22 % des sessions seulement, on compte deux séquences distinctes sur Gallica. Le « multi-tâches » est donc rarement pratiqué, et Gallica induit une activité longue et monolithique.

Au sein de ces longues séquences de navigation, la consultation de documents (texte, image, audio) est majoritaire, et s'appuie sur le duo Recherche – Feuilletage (voir Tableau 6.33). Le téléchargement d'ouvrages, au format TIFF ou PDF, en retrait

par rapport au feuilletage, est très important : dans 38% des sessions Gallica, l'utilisateur est amené à sauvegarder une copie locale d'un document sur son poste, pour un total de plus de 2 000 ouvrages ou extraits d'ouvrages téléchargés.

Tableau 6.33. Services visités sur Gallica

Service	Nombre de sessions	Part des sess. Gallica	Nombre de panélistes
Consultation	628	76,7 %	48
Page Accueil	617	75,3 %	53
Recherche	596	72,8 %	50
Feuilletage d'un ouvrage	548	66,9 %	47
Téléchargement d'ouvrage (tout ou partie)	314	38,3 %	34
Découverte	92	11,2 %	30
Dossiers	89	10,9 %	31
Aide	20	2,4 %	13

La consultation des dossiers reste marginale en nombre de sessions, mais concerne la moitié du panel, ce qui dénote un effet de « visite par curiosité » pour cette rubrique, élément confirmé dans les entretiens :

« Je suis allé voir ce qu'il y avait dans les dossiers pour savoir si ça m'intéressait ; il y en a je sais plus combien ; pour l'instant il n'y en a pas qui sont dans mes centres d'intérêt pour le moment, mais je me suis promis un jour où j'aurais du temps d'approfondir un petit peu ça. » (Utilisateur F)

« Je trouve que c'est un peu « gadgétique » et je trouve ça un peu limité. Puis bon... Comme mon domaine est ciblé, si vous voulez... Ça porte souvent sur des petits aspects qui sont pas dans mes préoccupations générales. » (Utilisateur H)

« J'y vais systématiquement rien que pour avoir le plaisir de voir quelques belles enluminures comme ça. Si vous voulez, c'est plus de la recherche, c'est le plaisir de voir de belles choses. » (Utilisateur P)

Pour la plupart des interviewés, la visite des Dossiers dans Gallica s'inscrit explicitement dans une autre logique de parcours : la flânerie des expositions numériques est opposée à l'efficacité de la recherche ciblée *via* le moteur de recherche interne au site. Les deux contextes d'usage sont ainsi clairement différenciés, le mode « prédateur » étant le plus fréquent.

Synthèse. Chez les panélistes de BibUsages, on observe un lien global entre la fréquentation des bibliothèques électroniques et les outils de recherche d'information d'une part, et les sites de « contenus à lire » et de commerce de biens culturels d'autre part. Les consommateurs de bibliothèques numériques sont de manière générale de grands consommateurs de produits culturels, que ce soit dans une optique de consultation (fonds numériques, journaux) ou de consommation (achat en ligne). Au-delà de cette approche globale, le trafic et les entretiens montrent qu'au sein des sessions, l'usage des bibliothèques électroniques est orienté vers deux types

de contextes : la recherche documentaire, avec un recours important aux moteurs de recherche et à des sites personnels, et le test avant achat de livres anciens, lié à la visite de sites de bibliophilie et de biens culturels. Dans les deux cas, la visite du site de Gallica est effectuée sur un mode « prédateur » où l'internaute mène une recherche ciblée et feuillette les documents dans une longue période ininterrompue ; partant, flânerie et parcours au gré des liens sont de l'ordre de l'occasionnel.

6.3.2 Le document numérique dans les pratiques

L'accès aux bibliothèques numériques dans un contexte de recherche conduit à reconsidérer le statut des documents numériques proposés. S'il s'agit souvent d'œuvres littéraires dont la lecture est *a priori* linéaire et complète, les contextes d'usage, le profil des utilisateurs et les déclarations faites au cours des entretiens amènent à considérer les fonds électroniques comme des bibliothèques de recherche plus que de lecture.

Lecture parcellaire et ciblée

Lors des entretiens, la lecture en ligne a été le plus souvent évoquée dans le cas des recherches ponctuelles et précises. Elle intervient généralement pour la lecture de petits documents ou pour localiser l'information avant enregistrement. Un cas particulier est à mettre en avant, l'utilisation de *e-books* et de lecture sur PDA¹ :

« Lire sur l'écran, d'abord on s'habitue, et puis il y a un mode de saisie que je retrouve un peu sur le Palm en petit format. On a une lecture qui est d'un certain point de vue accélérée parce que le coup d'œil est synthétique. (...) On a une saisie immédiate, évidemment un petit nombre de lignes, il faut que le reader soit bon, mais s'il est bon, c'est au contraire très confortable, on a lu la page d'un coup. Je me souviens d'avoir relu la Chartreuse de Parme de cette façon, avec un plaisir considérable, parce que c'était assez voluptueux comme ça, petites pages par petites pages de saisir le texte de manière très visuelle, presque comme au cinéma. Et ça donne une sensation particulière. » (Utilisateur A)

Hormis ce cas exceptionnel, la lecture sur écran est peu pratiquée. Jugée « fatigante » par la plupart des interviewés, elle intervient dans le cadre d'une recherche très précise d'informations, de citations ou de termes définis pour des corrections orthographiques, et concerne des portions très ciblées de documents.

Malgré cela, l'impression reste minoritaire, elle n'intervient que pour de petits documents ou des extraits, le coût important étant la cause première invoquée. L'impression permet entre autres d'extraire des parcelles de document et de les traiter de manière plus aisée.

« Mais je vais imprimer le chapitre qui m'intéresse, c'est plus facile à visionner sur un papier que sur écran. » (Utilisateur L)

¹ *Personal Digital Assistant* (Palm, Pocket PC).

« Sinon j'imprime la page qui m'intéresse avec les références. » (Utilisateur I)

« Euh, au début j'imprimais, j'imprime moins, hein. Au début j'imprimais beaucoup, euh... j'imprime ce qui est utile, hein, par exemple je viens de vous donner un exemple pour Casanova ; quand j'ai repéré, je lis à l'écran, c'est fatiguant mais je lis à l'écran, et quand je repère les pages qui vont me servir, alors là je les imprime. » (Utilisateur O)

D'autres imprimeront systématiquement, mais cette pratique reste anecdotique sur l'ensemble des panélistes interrogés : en définitive, l'impression comme la lecture sont des pratiques ciblées qui concernent des portions bien délimitées de texte répondant à un intérêt particulier et momentané de l'utilisateur. La consultation de fonds numérisés se rapproche alors définitivement de la recherche ciblée : il n'est pas question de lire les œuvres proposées en ligne (à l'écran ou sur papier), mais d'y puiser des informations précises pour répondre à des besoins particuliers. Ce mode de lecture rapproche les gallicanautes des chercheurs classiques dans leur mode de lecture, où l'ouvrage n'est pas lu de manière linéaire, tandis que l'index et la table des matières sont privilégiés.

Sauvegarder et manipuler le document numérique

Ce mode d'utilisation des documents numériques a deux conséquences dans l'appréhension des contenus en ligne : d'une part la constitution de fonds personnels accumulant les sources potentiellement mobilisables sur un thème donné, et d'autre part le remaniement des documents pour en permettre une manipulation facile.

En ce qui concerne le téléchargement, nous avons déjà constaté que 38 % des sessions sur Gallica comportent une action de téléchargement d'un ou plusieurs ouvrages, ce qui représente un total de plus de 2 000 documents téléchargés par 34 des 53 personnes du panel ayant fréquenté Gallica au cours des six mois d'observation. Les entretiens ont permis de mettre en évidence l'usage régulier voire systématique du téléchargement lorsque le contenu des documents correspond aux besoins de recherche de l'utilisateur. Il s'agit à la fois d'assurer la pérennité des ressources dans l'incertitude relative au Web, et de disposer à tout moment des documents sans avoir à les télécharger à nouveau – problème particulièrement sensible pour les connexions RTC :

« Certains sites sont éphémères et je m'en rappelle une fois un site que j'ai téléchargé pour mémoire » (Utilisateur K)

« J'ai vraiment l'habitude d'enregistrer sur le disque dur, de les garder en mémoire » (Utilisateur L)

« Le plus difficile c'est ensuite d'archiver et d'avoir le temps de regarder les archives. » (Utilisateur M)

Au sein de recherches souvent très ciblées, l'information précise contenue dans un document n'est pas seule sauvegardée, et l'entour de ces informations est également objet de conservation. Le téléchargement va de l'enregistrement de

documents avec conservation des indications d'auteur et de provenance, à l'aspiration de sites entiers.

« J'enregistrerai peut-être pas forcément uniquement l'info qui me semble intéressante mais j'enregistrerai la page entière ou si je trouve que le site est intéressant, j'avoue que je copie le site sur le disque dur, ou une partie du site. » (Utilisateur L)

« Si c'est quelque chose que je vais consulter par la suite, donc là, je l'enregistre et je le garde pour après. » (Utilisateur G)

Un nombre important d'interviewés déclare constituer ainsi des sortes de fonds numériques personnels centrés sur leurs centres d'intérêt propres, dont certains les documentent et en assurent la pérennité en les gravant sur CD-Rom, ce qui atteste une véritable politique de conservation à long terme :

« Donc ensuite sous [...] Access, je mets : nom d'auteur, titre, et le numéro du Cdrom sur lequel, parce que c'est encore archivé avec des numéros, et quand par exemple je travaille actuellement sur l'évolution de la mécanique, je prends par exemple les trucs de Pierre Duhem qui sont numérisés sauf le volume titre, c'est pas grave, et après en fonction de ça je compulse, alors je compulse directement sur l'ordinateur. Je n'imprime pas. » (Utilisateur M)

"Je télécharge et puis je le stocke. Même, parfois, je ne l'édite pas. Mais je sais que ça existe et puis bon, le jour où je vais rédiger une partie de ce que je fais, bien à ce moment-là, je l'imprimerai ou je le regarderai". (Utilisateur N)

Cette politique de conservation se situe dans le cadre d'un travail plus général d'accumulation systématique de documents sur un sujet donné. Dans cette perspective, le mode image prépondérant dans le fonds proposé par Gallica est pour beaucoup un frein à la manipulation des documents, et certains procèdent à leur retranscription, manuelle ou par reconnaissance optique de caractères.

« Alors quand c'est des textes, c'est facile, avec la recherche on peut voir, on va chercher un mot, hein [...] Si vous voulez parler des médecins, vous cherchez médecins, vous allez trouver quelque chose hein. Comme il apparaît un peu de tout, mais seulement on repère tout de suite, hein. En images, allez voir ! » (Utilisateur O)

« Alors, je le télécharge, puis je fais une reconnaissance de caractères, puis je lis tout le texte et je corrige ma reconnaissance de caractères qui va comporter énormément de fautes, bien sûr, puisque les ouvrages donnés en ligne par la Bibliothèque Nationale, ils ont raison de les faire suffisamment légers, il y a un compromis entre la qualité du texte et puis le poids qu'ils vont avoir. Donc ce poids ne permet pas d'avoir quelque chose de parfait, mais ça ne fait rien ! À ce moment là, cet ouvrage-là, en même temps que je le lis, dans mon texte, je corrige. Et après, l'ouvrage, ça y est, j'en fais un autre document PDF mais ça n'a rien à voir avec celui de la Bibliothèque Nationale, il doit être parfait. Je respecte la mise en page du bouquin bien sûr. » (Utilisateur P)

Ces éléments nous amènent à penser que le statut des documents électroniques est plus proche de celui de l'usuel que de l'œuvre, leur utilité étant définie dans le cadre d'une pratique ciblée et d'interrogations précises. La conservation des documents numérisés et la constitution de fonds électroniques personnels peuvent être ainsi vues comme l'accumulation de documents de référence pour un usage ultérieur, même s'ils ne sont pas effectivement consultés par la suite. L'attitude est bien ici celle de chercheurs balayant leur champ d'investigation, et il est intéressant de constater qu'elle concerne non seulement des enseignants et des chercheurs universitaires, mais également une population de « chercheurs amateurs », qui trouvent avec Gallica et les bibliothèques électroniques en général le moyen d'accéder aux fonds habituellement réservés aux universitaires et aux chercheurs professionnels¹.

Synthèse. Les entretiens renforcent et affinent les éléments mis en avant à l'aide des données de trafic : le recours aux bibliothèques électroniques dans un contexte de recherche se retrouve dans le traitement et la manipulation des documents numériques. La pratique de lectures ciblées, la constitution de fonds personnels et d'archives sur un thème donné, la documentation des corpus personnels sont le reflet d'une manipulation des documents qui les rangent du côté de l'usuel ou de la ressource documentaire plus que de l'œuvre.

6.3.3 Usages-types

Sur la base des traces d'usages issues du trafic des utilisateurs ainsi que des entretiens menés avec seize d'entre eux, trois « portraits-types » d'utilisateurs ont été identifiés :

- Le chercheur d'information.
- Le bibliophile.
- Le « lecteur à l'écran ».

Il ne s'agit nullement ici de donner une segmentation des utilisateurs, mais plutôt de dresser des profils correspondant à des usages récurrents observés dans notre échantillon.

Le chercheur d'information

Ce qui est mis en avant dans ce type d'usage, c'est bien le rôle de médiation de Gallica à l'intérieur d'Internet. Gallica est une source de documents (de référence) sur le Web, qui permet souvent d'accéder à des informations attestées et de poursuivre sa recherche. On utilise le moteur de Gallica (presque) comme tout autre moteur de recherche sur Internet et les objets visés – les contenus recherchés – sont de nature essentiellement électronique.

¹ Pour une comparaison approfondie entre « gallicanautes » et public traditionnel de la BnF, voir [Assadi *et al.* 2003b].

Dans notre panel, nous avons rencontré des utilisateurs menant des recherches dans un cadre professionnel, mais surtout, une majorité de personnes menant des recherches à titre personnel autour d'un hobby, d'une passion, un sujet de recherche très présent étant la mémoire familiale et régionale.

« J'étais financier dans des industries. [...] Je suis passé sur les recherches sur la famille de Lorraine et la Lorraine d'abord, par le biais de ma famille. Il se trouve que ma famille a été liée à la Maison de Lorraine depuis le XVIIe siècle jusqu'au XIXe, alors c'est comme ça que je suis passé... » (Utilisateur E)

Ces recherches personnelles occupent une partie importante de leur temps, et donnent parfois lieu soit à des publications, à la constitution d'archives personnelles (CD-ROM) ou à la création de sites personnels.

« Mon intention, c'est de rédiger quelque chose, alors je ne sais pas si ça intéressera mon éditeur, parce que ce sera beaucoup moins illustré que mon livre précédent. » (Utilisateur N)

« J'ai publié un petit bouquin il y a une dizaine d'années ; je vais en republier un avant la fin de l'année, un livre pour enfants, il y a le polar qui arrive et qui, je pense, sera pour le mois de mai. » (Utilisateur F)

Dans ce type d'usage, c'est le texte lui-même qui prime : « Alors le problème, c'est que j'aime bien les beaux livres anciens, c'est très beau dans une bibliothèque, mais les rééditions sont très chouettes aussi. Encore une fois je suis plus intéressé par le contenu que par la relique. » déclare un interviewé. Les questions de manipulation et de documentation des contenus numériques passent alors au premier plan : le mode texte est préféré au mode image, et l'identification des sources et des auteurs valorise les bibliothèques patrimoniales comme Gallica, même si celle-ci peut être mobilisée conjointement avec d'autres ressources sur le Web.

Le bibliophile

Contrairement au premier cas, Gallica joue ici un rôle de médiation vers le monde « réel » (y compris la sphère marchande). Pour le bibliophile, Gallica est utilisé en pré-achat, le but ultime reste l'objet-livre.

« Je vais sur le site de la BNF parce que j'y trouve quelque chose, j'ai le souvenir que je pourrai trouver quelque chose, et que je vais me consacrer un peu de temps pour voir quelques images, ou quelques documents que je ne pourrai pas consulter, ou être sûr que je vais acheter un ouvrage ; je parlais de [nom d'auteur], je me suis acheté en consultant la BNF, ça m'a beaucoup intéressé, car j'ai acheté un ouvrage qui coûtait quand même 1 200 F. » (Utilisateur H)

Dans ce cas, l'impression ne supprime pas l'ouvrage relié, et l'attrait pour les ouvrages anciens est à mettre en parallèle avec la valorisation du livre traditionnel en général :

« Il y a un côté sensuel du livre, le toucher, le prendre ; il y a la faculté de le prendre à n'importe quel moment, qu'on fasse beaucoup de Kms, prendre un bouquin et en lire trois pages, c'est aussi reposant que pendant une demi-heure,

en avion, un peu partout, dans le train parfois. Donc c'est agréable. Et puis il y a surtout le plaisir de partager, de prêter un livre, ou se le faire prêter, c'est une sorte de convivialité très importante. » (Utilisateur H)

Le format électronique ne supplante pas l'édition papier dans son statut : « Le bouquin, c'est quand même mieux ; il y a une pérennité aussi, ça se conserve, ça c'est le côté objet que l'on possède », déclare un interviewé. Ici, au-delà des arguments d'ordre pratique (maniabilité, prix, conservation), le livre reste un objet culturellement valorisé en tant que tel.

Le lecteur à l'écran

Enfin, nous avons choisi d'évoquer ici un profil plutôt atypique, correspondant à un usage rare d'après nos observations. Il s'agit d'utilisateurs ayant une pratique intensive de la lecture à l'écran. Dans ce cas, Gallica – ou tout autre site offrant des collections de textes et d'ouvrages – est considéré comme un lieu de “consommation” et non pas de médiation : « J'écume les sites où on peut télécharger des textes, c'est ma passion. » déclare l'utilisatrice interviewée qui correspond à ce profil.

« La motivation de la fréquentation est toujours le téléchargement de textes ; je recherche bien sûr des textes classiques ; je suis agrégée de lettres au départ. Donc c'est un peu normal. Donc c'est par rafales, brusquement l'idée me vient, je suis assez fantasque, l'idée me vient tout d'un coup de relire Scaron ; je l'ai dans ma bibliothèque, mais maintenant je lis de la sorte sur écran, et j'ai besoin du texte tout de suite, et donc je vais voir s'il est disponible à la BNF, et si l'ouvrage que je recherche n'est pas là, je vais chercher ailleurs. » (Utilisateur A)

On note ici un déplacement des pratiques de lecture intensive du papier vers l'écran, accompagné d'une redécouverte du plaisir lié à la lecture et à l'appréhension de l'objet-livre. L'intérêt se porte sur les œuvres littéraires, qui impliquent une lecture globale et linéaire, et dont le support de lecture renouvelle l'appréhension.

Le livre renouvelé par le Web

En somme, dans le cadre des pratiques observées, le statut du document numérique semble plus proche de l'usuel que de l'œuvre. Sa consultation et son usage s'inscrivent le plus souvent dans le cadre de recherches personnelles sur des sujets précis : les textes consultés sont majoritairement envisagés comme support de recherche bien plus que comme objet de lecture. Il est frappant de constater combien ces recherches personnelles sont menées avec rigueur et méthode : démarche de recherche systématique (avec en particulier un usage avancé des moteurs de recherche sur le Web), classement rigoureux des documents allant jusqu'à la constitution de fonds documentaires personnels, tendance à privilégier les éditions de référence (ce qui représente un des motifs du recours à Gallica). En ce sens, les fonds numériques tendent à rendre disponibles, au sein des ressources accessibles sur Internet, des documents et des possibilités de recherche nouveaux, et élargissent ainsi le public des chercheurs à l'amateur et à l'érudit. On souscrit ici à l'analyse faite par Chartier dans [Chartier 2003], qui projette une coexistence entre écriture manuscrite, publication imprimée et textualité électronique, où cette dernière

organise « de manière nouvelle la relation entre la démonstration et les sources, les modalités de l'argumentation et les critères de la preuve ».

Ce public semble assez différent de celui des bibliothèques classiques, et les chercheurs « professionnels » y sont comparativement peu représentés. Les plus de quarante ans, actifs ou retraités, sont majoritaires dans la population observée et les bibliothèques électroniques sont avant tout pour eux une source d'informations dans le cadre de recherches personnelles. L'intensité d'usage est ici bien supérieure à celle de la population générale des internautes français et va de pair avec un très fort taux d'équipement en haut débit (câble, ADSL). Nous découvrons ici une population d'internautes seniors fortement équipés en haut débit, et dont les centres d'intérêt, outre l'offre de services et de communication classique, gravitent autour des contenus « culturels ». Cette population atypique dans le paysage des internautes français constitue en elle-même un groupe d'autant plus intéressant qu'il est mal identifié dans les études d'usages à large panorama car minoritaire.

Si d'une manière générale les utilisateurs des bibliothèques électroniques sont également de forts consommateurs de « contenus à lire » (journaux en ligne en particulier), au sein des sessions, l'usage des bibliothèques numériques est fortement corrélé à celui des moteurs de recherche d'une part, et des sites de vente de biens culturels d'autre part. Deux profils se dégagent : celui du « chercheur amateur », dont les centres d'intérêt sont pointus et déjà bien connus de l'utilisateur, et celui du bibliophile pour qui Gallica fait office de catalogue avant achat. Dans les deux cas, la lecture en ligne est rare, tout autant que l'impression des documents téléchargés, et la lecture s'apparente à la recherche de fragments ciblés au sein de vastes collections laissant de côté la totalité des œuvres. Dans ce cadre, le statut des documents en ligne semble remis en cause : tandis que l'édition papier reste du côté de l'œuvre, l'édition électronique s'apparente à l'usuel.

L'étude des pratiques en contexte souligne également les passerelles entre Web marchand et non marchand pour les utilisateurs. Alors que les acteurs d'Internet (fournisseurs de contenus et d'accès) perçoivent une dichotomie forte entre sites marchands et non marchands, les internautes passent indifféremment d'un type de site à un autre, et l'on doit plutôt parler d'enrichissement mutuel entre sites marchands et non marchands dès lors qu'on les envisage sous l'angle des pratiques.

Dès lors, attirant de nouveaux publics, induisant de nouveaux modes d'appréhension des textes, s'inscrivant dans des parcours de lecture inédits, les bibliothèques électroniques, loin d'être une simple version numérisée des fonds traditionnels, s'apparentent à un nouvel espace de lecture et de consultation aux côtés des bibliothèques classiques.

Synthèse. Trois profils-type d'utilisateurs des bibliothèques électroniques se dégagent : le chercheur d'information, pour qui les fonds numérisés constituent une ressource parmi d'autres ; le bibliophile, plutôt tourné vers l'objet-livre et l'achat ; et le lecteur à l'écran, profil rare renouvelant le plaisir de la lecture sur des terminaux informatiques. Dans les trois cas, le livre-papier conserve le statut d'œuvre, tandis que l'édition électronique (fac-simile ou texte manipulable), plongée dans le contexte du Web et profondément influencée par celui-ci, s'inscrit dans de nouvelles formes d'usage où prime avant tout le texte.

Conclusion

Dans le Chapitre 5, nous avons construit une typologie des parcours sur la base de leur forme et de leur temporalité, et observé les corrélations que ces parcours-type entretiennent avec les contenus ; replacés dans le contexte des pratiques individuelles, ces modes de navigation prototypiques prennent un sens particulier. Ils ne sont effectivement liés ni à l'ancienneté de la pratique d'Internet, ni au type de connexion, ni aux types de pratiques d'Internet en général, mais semblent dépendre de contextes particuliers liés à l'activité locale de l'utilisateur. La notion de territoire personnel sur le Web nous permet de cerner ces contextes et de leur donner sens : la vue longitudinale sur trois ans qu'offrent les données SN00-02 permet de voir comment sont structurés les espaces individuels de navigation tant du côté des thèmes ou services accédés que des sites visités. Dans les deux cas, il apparaît qu'un très petit nombre de sites constitue le support principal de la navigation, quel que soit le contexte de l'utilisateur. On le retrouve dans les sessions routinières, renvoyant à un territoire restreint fréquemment balayé, ainsi que dans un deuxième ensemble de sessions impliquant des sites vus de manière occasionnelle mais régulièrement dans un contexte donné. Plus encore dans un dernier groupe de sessions tournées vers la découverte de nouveaux sites, dont la plupart ne seront jamais revus par la suite, c'est encore le noyau dur des sites habituels de l'internaute qui sont le support de la découverte et du mouvement vers l'inconnu. En définitive, et contrairement à l'idée souvent évoquée de navigation dans l'hypertexte au gré des liens, l'internaute reste le plus souvent dans un espace qui lui est familier et qui sert de pivot à l'ensemble de ses parcours.

Confrontée à la segmentation des sessions en parcours-type, cette description des sessions par le biais des territoires se révèle très productive : le croisement de ces deux typologies met à jour des comportements de navigation en situation intégrant la double dynamique des contenus et des visites par l'utilisateur. Trois modes d'appréhension prototypiques des contenus du Web sont à distinguer : en premier lieu, le mode routinier, lié aux parcours éclairés et ciblés, fonctionne sur le mode d'un balayage régulier et fréquent de ressources connues qui ont la particularité, sur le plan du contenu, d'être des flux – flux d'information (informations générales, boursières, sites spécialisés) ou flux de communication (WebMail, forums, *chat*). Ensuite, un mode « galerie marchande » ou « guichet de renseignements », qui intervient de manière moins fréquente, implique des parcours orientés vers le « Web pratique », plus complexes et plus allongés ; ces contextes particuliers surviennent peu fréquemment, mais les comportements sont très stables dans ces contextes en termes de sites visités. Enfin, les parcours ouverts, qui amènent à découvrir de nouveaux sites, font massivement appel aux moteurs de recherche, dans le cadre de requêtes précises, et sont liés aux parcours à pivots ou parcours éclatés.

Ces trois modes prototypiques d'appréhension du Web contrastent à la diversité thématique et surtout fonctionnelle des contenus Web : services proposés, outils de recherche, fréquence de mise à jour, type de publication sont autant d'éléments qui suggèrent une grande diversité dans les pratiques de la Toile et leur possible mobilisation dans des chaînes opératoires très diverses. Pour autant, la mise à jour de modes génériques d'appréhension des contenus atteste la prévalence des pratiques et des modes d'appréhension dans la valorisation des contenus : malgré des modes de

navigation parfois prédéfinis de manière très contrainte par leurs concepteurs, les sites se caractérisent avant tout par ce qu'en font les internautes.

L'usage massif des bibliothèques électroniques dans des contextes de recherche documentaire ou d'achat va dans ce sens : on rejoint ici F. Rastier qui proposait de « formuler un principe d'*architextualité* : tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent »¹. En étendant ce principe à Internet, on peut ainsi affirmer que tout contenu plongé dans le Web s'expose à en subir l'influence non seulement présentationnelle, formelle et fonctionnelle, mais aussi perceptuelle, interprétative et usuelle. La juxtaposition au sein d'une même interface de l'ensemble de l'hypertexte opère ainsi sur les sites une mise à plat, que les parcours viennent assembler au fil des pratiques individuelles et des usages collectifs ; en somme, ce n'est pas tant le contenu des sites qui donne une signification au parcours que la somme des parcours qui confère un sens individuel ou collectif aux contenus du Web au sein de pratiques stabilisées.

¹ [Rastier 2001a], p. 92.

