Mesurer les usages d'internet

Valérie Beaudouin

France Télécom R&D
38-40, rue du Général Leclerc
92794 Issy-les-Moulineaux Cedex 9
valerie.beaudouin@francetelecom.com

Résumé. Nous rendons compte d'une démarche mise en place pour construire une représentation fine des usages d'internet et de leur évolution, en procédant à du traitement secondaire de données de trafic, provenant de panels représentatifs d'internautes. Après avoir présenté les caractéristiques des cohortes étudiées et les différents modes d'enrichissement des données de trafic mis en place, nous présentons quelques résultats construits à partir de ces données enrichies, et en particulier une segmentation des internautes construite sur la base de l'entrelacement des pratiques de communication et de navigation.

1. Introduction

Dans le laboratoire de sciences humaines de France Télécom R&D, ont été développées des méthodes d'analyse fine des usages du téléphone fixe et mobile en articulant l'analyse du trafic avec des enquêtes, qui permettent de qualifier les correspondants des foyers ou individus observés [Smoreda & Licoppe, 1999]. Avec l'apparition et la diffusion d'internet, il nous a paru nécessaire de mettre en place des méthodologies nouvelles pour étudier les usages du réseau, qui viennent s'inscrire et transformer le panorama des pratiques de communication.

L'objectif visé était donc d'acquérir dans le domaine des usages d'internet une compétence similaire à celle développée pour l'analyse des usages du téléphone. Il nous fallait décrire avec fiabilité la réalité des pratiques d'internet, en dépassant les limites des innombrables enquêtes en ligne qui ne touchent que les "accrocs du net", des enquêtes quantitatives qui ne reposent que sur des données déclaratives et les entretiens qui eux parviennent à une description fine et compréhensive des pratiques mais se heurtent aux limites de leurs échantillons. S'appuyer sur des données de trafic de panels représentatifs paraissait la démarche la plus appropriée. Données de trafic plutôt que données déclaratives car les activités qu'autorise internet se font toutes dans la même situation : devant un écran, avec un clavier et une souris ce qui rend difficile une appréhension et description de la diversité des pratiques par l'utilisateur. L'analyse des traces d'usage recueillies sur le poste des utilisateurs permet en effet de *décomposer l'activité derrière l'écran*, pour *recomposer des figures d'internautes* construites sur la base de leurs pratiques. Panels représentatifs parce qu'on cherche à rendre compte d'une pratique dans toute sa diversité.

C'est dans ce contexte que nous avons monté un partenariat avec NetValue, société de mesure d'audience, pour procéder à du traitement secondaire de leurs données de panel. Ce

partenariat permettait d'avoir des données très fines d'internautes représentatifs de la population connectée à domicile en 2000. Ce premier partenariat (auquel était associé HEC) a été suivi par le montage d'un projet RNRT, SensNet, qui associe France Télécom R&D, Nielsen/Netratings (ex NetValue), le LIMSI et Paris III. Ce dernier projet vise à mettre en place un système de catégorisation sémantique des parcours sur le Web qui sera appliqué à des données de panels plus étendues (trois années : 2000 à 2002 ; trois pays : France, Angleterre, Espagne).

Sur les données de trafic, nous avons développé un ensemble de méthodes de mise en forme et de traitement des données.

Ces données enrichies ont été utilisées pour montrer l'évolution des pratiques, les caractéristiques des utilisateurs de telle pratique spécifique et enfin pour construire une typologie des internautes sur la base de leurs pratiques. Nous présentons quelques résultats établis sur une cohorte de 1140 individus tirée du panel NetValue et suivie sur l'année 2000[1].

2. Cohortes construites sur des panels

Mesurer les usages d'internet a été une préoccupation quasi concomittante à la naissance du média : suivi d'expérimentations, questionnaires en ligne, enquêtes quantitatives et qualitatives… Toutes les méthodes d'enquête ont été mobilisées. Plus rares ont été les approches qui enrichissaient les données d'enquêtes de mesures objectives de trafic, et plus rares encore les enquêtes ayant une visée de représentativité de la population connectée à internet. Dans ce contexte, les sociétés de mesure d'audience ont joué un rôle fondamental. Il y en avait trois en France : MMXI, NetValue et Netratings, qui suite à des faillites ou rachats se réduisent à une aujourd'hui, Nielsen/Netratings en partenariat avec Médiamétrie.

Ces sociétés de mesure d'audience contribuaient à la définition des tarifs publicitaires et à une époque où la valeur d'une entreprise était liée à son nombre de visiteurs-clients, elles jouaient sur la valeur boursière des entreprises. Pour pouvoir occuper cette position sur le marché, leurs données devaient répondre à deux exigences : la représentativité de la population et la complétude des données de navigation. La représentativité du panel est assurée mois par mois par une enquête téléphonique sur vaste échantillon qui permet d'évaluer la population connectée à internet et ses caractéristiques. Le panel est ajusté pour être conforme à la répartition nationale et à son évolution (ce qui est particulièrement nécessaire dans un marché à forte croissance). La qualité des informations sur les parcours internet est garantie par le recueil sur le poste utilisateur de toutes ses traces de navigation. En effet, la diversité des sites et des activités possibles sur le réseau ne permet pas de s'appuyer sur la bonne volonté de l'internaute pour décrire ses pratiques. La mise en place d'un dispositif technique de mesure des pratiques s'est avéré très tôt indispensable : chaque société a développé son propre outil avec des caractéristiques propres. Ces dispositifs de captation des flux ne sont pas transparents : ils peuvent cesser d'enregistrer pour des raisons diverses (le système est désactivé momentanément et l'utilisateur oublie de le remettre, un reformatage de disque sans réinstallation…). Certaines sociétés comme NetValue détectaient les clients silencieux et les relançaient par téléphone.

[1] Les éléments présentés ici sont le résultat d'un travail collectif dont rend compte [Beaudouin et *al*., 2002] et le numéro 116 de la revue *Réseaux*, "Parcours sur Internet".

Les panels de mesure d'audience sur internet constituent donc une source essentielle de connaissance des pratiques d'internet à la fois représentatives et exhaustives.

Le panel de NetValue nous a paru constituer la source la plus complète. En effet, les autres opérateurs ne mesuraient que la fréquentation des sites web et ne tenaient aucun compte des pratiques autres que le web qui occupent cependant en volume comme en temps passé une part de plus en plus conséquente des usages d'internet. Netmeter, le système de captation d'audience de NetValue, enregistre en effet tous les protocoles autres que *http* : aussi bien le mail que les logiciels de peer-to-peer, les messageries instantanées que les jeux en réseau, le chat que le "streaming". Dans une perspective d'étude des usages d'un point de vue sociologique, l'articulation entre les pratiques de communication interpersonnelle et les pratiques d'information via le web nous paraissait centrale.

2.1. Cohorte : une représentativité qui se déforme au fil des mois

A partir du panel France de NetValue, nous avons défini une cohorte constituée par les internautes présents sur les deux derniers mois de l'année 99 et actifs au moins une fois en 2000. Cette cohorte a été suivie sur une année entière 2000, et la même démarche a été adoptée pour constituer une cohorte en 2001, puis en 2002. Nous avons une sous-cohorte de 600 individus environ suivie pendant trois ans.

Les panels d'audience servent à produire des chiffres de fréquentation des sites mois par mois mais n'ont pas pour mission d'étudier les évolutions. La démarche entreprise dans le cadre du partenariat avec NetValue, puis dans SensNet, a visé au contraire à étudier les transformations longitudinales des pratiques pour une population fermée. Nous avons ainsi simulé ce que pourraient être les évolutions des pratiques d'internet une fois le marché arrivé à saturation. Certains de nos résultats ont paru contre-intuitifs, en raison d'une confusion classique entre coupe transversale et analyse longitudinale : nous avons ainsi montré que l'usage des moteurs diminuait progressivement, alors que les offreurs de moteurs voyaient au contraire l'usage de leurs outils augmenter fortement (grâce à l'augmentation du nombre d'internautes).

Bien entendu, la limite de l'approche par cohorte tient au fait que pour un marché en forte croissance, la représentativité de la cohorte diminue de mois en mois. C'est pour cela que nous avons choisi de redéfinir une nouvelle cohorte chaque année.

2.2. Quelles données

Grâce aux données des panels de NetValue, nous disposons d'informations détaillées sur les internautes et d'une description à grain fin de leurs pratiques d'internet.

Le premier point qu'il faut rappeler, c'est que les traces de connexions sont attribuées à un individu du foyer et non pas à une machine. Contrairement aux approches site-centric qui ont de grandes difficultés à identifier une personne derrière les adresses des machines ou les "cookies", les panels d'audience donnent accès à l'individu, puisque toute connexion nécessite de s'authentifier parmi une liste d'utilisateurs (les membres du foyer panélisés). Les individus sont décrits par les variables socio-démographiques classiques liées à eux et à leur foyer : âge, sexe, occupation, PCS du chef de foyer (conforme à la nomenclature de l'INSEE), composition du foyer, lieu d'habitation, niveau de revenu… Nous avons beaucoup regretté l'absence de la variable "niveau de diplôme" qui d'après de nombreux travaux joue un rôle plus décisif que la PCS ou le revenu [DiMaggio *et al.*, 2001]. Les foyers sont

également décrits par leur niveau d'équipement audiovisuel et de communication. Enfin, des éléments sur leur connexion internet sont recueillis : ancienneté, type de connexion, lieux de connexion.

En termes de pratiques d'internet, les bases de données sont extrêmement détaillées. En ce qui concerne la navigation sur le web, pour chaque individu sont enregistrées les informations suivantes sur les pages vues : l'url, l'horodatage, la nature de la requête, éventuellement le "referer", c'est-à-dire la page précédente qui a conduit à la page vue. Ceci amène plusieurs commentaires. Ne sont pas retenues comme pages vues les images et autres composantes de la page : seul le fichier qui articule les éléments multimédia est conservé. Mais quand les pages qui s'affichent à l'écran sont constituées de plusieurs cadres, il est difficile de reconstituer la page telle qu'elle s'est présentée à l'utilisateur. Plus généralement, entre la définition technique de la page vue et la réalité de ce qu'a consulté l'internaute, les écarts peuvent être sensibles. Les enregistrements ne sont qu'une approximation de la notion de page vue. On notera aussi que le système ne permet pas de distinguer les fenêtres de navigation quand il y en a plusieurs. Pour le courrier électronique, sont enregistrés tous les messages envoyés et reçus par l'internaute avec expéditeur et destinataires anonymisés, horodatage, taille du message, présence de pièces jointes. Les autres protocoles donnent chacun lieu à des enregistrements particuliers liés à leur spécificités techniques. Ainsi pour les pratiques de peer-to-peer, sont enregistrés le logiciel utilisé, la durée et le volume de chaque action sur le système. Quelle que soit l'activité, elle est toujours reliée à un individu du panel, clairement identifié.

3. Des modules d'enrichissement des données

Les données dont nous disposons sont d'une qualité et d'une richesse inégalable au vu des difficultés que représente un objet aussi divers et changeant qu'internet. Cependant, ces données pour être finement exploitables nécessitent différentes procédures d'enrichissement.

Internet est un media hybride qui autorise des activités de types très différents : rechercher ou consulter de l'information, échanger des mails, converser en direct, télécharger de la musique, jouer… Ces différentes activités peuvent se faire soit en passant par des protocoles particuliers dédiés à une activité précise (comme le smtp pour le mail, l'irc pour le chat…), soit *via* le Web. La première démarche de qualification des données a consisté à identifier parmi les usages du web ceux qui correspondaient à des activités spécifiques comme la communication interpersonnelle ou la recherche d'information. Un module de catégorisation des services a été ainsi mis en place.

Ensuite, l'enrichissement des données a consisté dans la définition d'une session internet multiprotocoles. Deux manières de définir les sessions ou visites sont généralement utilisées : soit la session est définie par rapport à la période de connexion (ce qui ne présume pas d'usages réels, puisqu'on peut rester connecté sans activité aucune), soit par rapport au simple usage du web. Ainsi, une nouvelle session est définie après une interruption d'activité sur le réseau de plus de trente minutes. L'observation des pratiques des internautes et les travaux qualitatifs et ethnographiques menés sur les usages d'internet nous ont plutôt conduit à penser qu'il y avait un entrelacement fort entre les activités sur internet : navigation sur le web, courrier électronique, forums, chats, messageries instantanées [Beaudouin et Velkovska, 1999]. Nous avons bâti une session internet qui intègre toutes les formes d'activité sur le réseau.

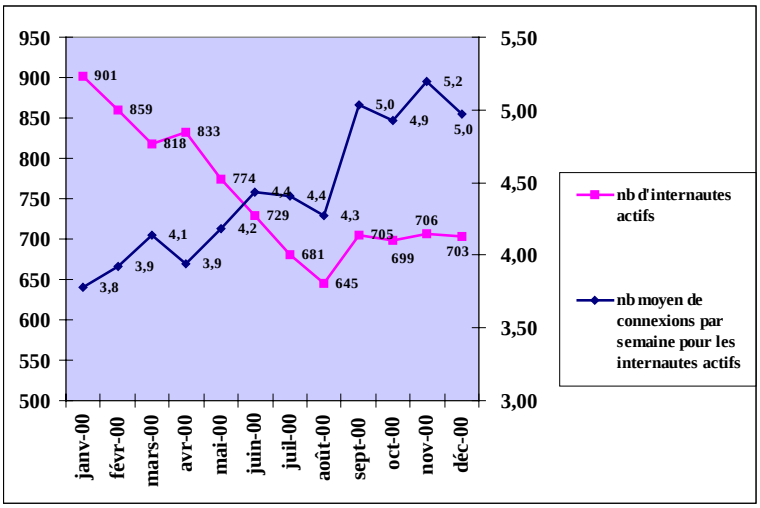
Cette notion de session est pour nous centrale : en effet, elle est l'unité de mesure que nous avons retenue pour comparer des formes d'activités qui laissent des traces très hétérogènes sur le réseau. Ni le nombre de pages vues, ni le nombre d'événements sur le réseau ne nous paraissent être de bons indicateurs. La notion de session est utilisée pour décrire si telle activité (surf, mail, chat, messagerie instantanée…) a été pratiquée ou non au cours de la session. Ainsi avons-nous pu réduire les tables de navigation gigantesques à une matrice de sessions, où les sessions sont décrites par des batterie de traits (présence ou absence de mail, de l'utilisation d'un moteur, de web…).

En complément, d'autres démarches d'enrichissement des données de navigation sur le web ont été mises en place : une des pistes explorées consiste à exploiter les annuaires de sites web présents sur le réseau [Assadi et Beauvisage, 2002]. Ainsi le travail de qualification des sites par des "surfeurs" peut-il être projeté sur les données de parcours. L'autre piste consiste à analyser le contenu et la structure des pages vues, autrement dit à appliquer des méthodes de fouille de textes aux corpus de pages web [Beaudouin et al, 2001]. Cette piste a été explorée en collaboration avec le Limsi et Paris III et se poursuit aujourd'hui dans le projet RNRT SensNet.

4. Vers une représentation des usages

Ces procédures d'enrichissement des données permettent ensuite de construire une représentation fine des usages, qui en partant de traces techniques cherche à reconstruire les pratiques des utilisateurs.

4.1. Données de cadrage



Clef de lecture : en janvier 2000, 901 internautes de la cohorte (sur 1140) se sont connectés au moins une fois et ont fait en moyenne 3,8 sessions par semaine. En décembre, ils n'étaient plus que 703 actifs avec 5 sessions en moyenne.

FIG *1. – Evolution des internautes actifs et du nombre moyen de sessions par internaute*

Internet, comme bien d'autres pratiques, se caractérise par une distribution très inégale de l'activité. On note une très forte dispersion dans l'intensité d'utilisation d'Internet. En effet la quasi-totalité du trafic (90% des sessions) a été générée par la moitié du panel. Alors que le nombre d'internautes actifs diminue de mois en mois (même quand on ne tient pas compte des mois d'été), le nombre de sessions (tous protocoles confondus) est quant à lui assez stable. Pour les internautes qui restent actifs, la pratique du réseau semble donc s'intensifier.

On peut faire l'hypothèse de deux types de trajectoires opposées dans les usages d'Internet. Pour une fraction des internautes, les usages sont rares et tendent à se raréfier voire à disparaître au fil des mois, pour les autres au contraire, surtout pour les gros utilisateurs, la pratique tend à s'intensifier. Le profil des sessions internet permet sans doute d'expliquer ces différences de trajectoires.

Il y a en effet d'importantes disparités dans les usages. Si tous les internautes naviguent sur la toile et presque tous utilisent le courrier électronique, ils ne sont qu'un quart à s'être connectés à un *chat*, à avoir utilisé une messagerie instantanée ou à avoir fréquenté un forum en 2000. Il ne s'agit évidemment pas toujours du même quart. Si l'on ne regarde que la question de la durée, la figure suivante montre bien que le type d'activité a un effet très notable sur la durée des sessions.

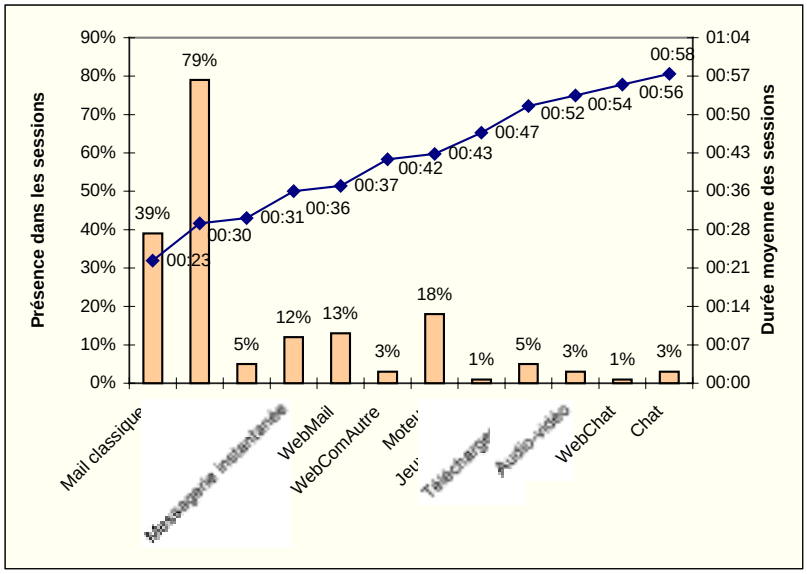
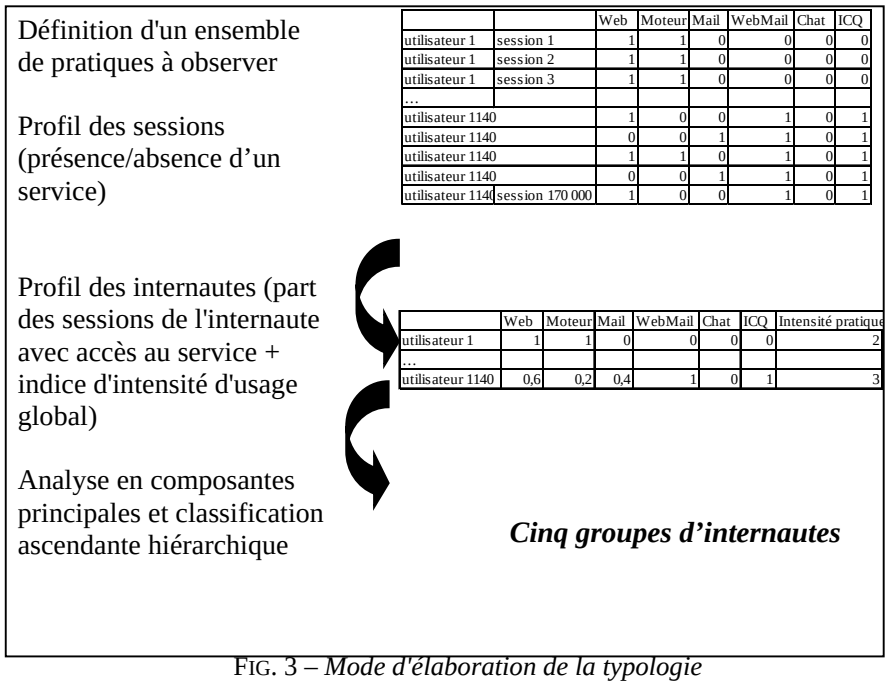


FIG. 2 – *Durée des sessions en fonction de la présence d'un type d'activité*

4.2. Segmentation

Ces premières observations nous ont conduit à construire une typologie des internautes en partant de la matrice décrivant les sessions en fonction d'un certain nombre de critères, en particulier toutes les pratiques de communication interpersonnelle. La figure suivante explicite la démarche : caractérisation des sessions, puis caractérisation des internautes en

fonction de leur profil de session sur toute l'année. La segmentation se fait sur les internautes en fonction de leurs profils.



		Web	Moteur	Mail	WebMail	Chat	ICQ
utilisateur 1	session 1	1	1	0	0	0	0
utilisateur 1	session 2	1	1	0	0	0	0
utilisateur 1	session 3	1	1	0	0	0	0
…							
utilisateur 1140		1	0	0	1	0	1
utilisateur 1140		0	0	1	1	0	1
utilisateur 1140		1	1	0	1	0	1
utilisateur 1140		0	0	1	1	0	1
utilisateur 1140	session 170 000	1	0	0	1	0	1

	Web	Moteur	Mail	WebMail	Chat	ICQ	Intensité pratique
utilisateur 1	1	1	0	0	0	0	2
…							
utilisateur 1140	0,6	0,2	0,4	1	0	1	3

FIG. 3 – *Mode d'élaboration de la typologie*

Sur le millier d'internautes, nous avons construit des profils d'utilisation d'Internet, en fonction des services utilisés (Web, courrier électronique, messageries instantanées, *chats*…). Nous avons trouvé deux profils d'usages majeurs : d'une part, les faibles et très faibles utilisateurs d'Internet, qui représentent 47% de la cohorte mais ne génèrent que 15% des sessions sur l'année, et d'autre part les utilisateurs plus assidus (53% de la cohorte et 85% des sessions enregistrées). Cette opposition reflète des trajectoires d'appropriation d'Internet bien différenciées : les résultats montrent que l'utilisation d'Internet décline au fil des mois pour les premiers alors qu'elle augmente pour les autres. Ainsi se dessinent deux trajectoires d'usage d'Internet, avec une décroissance des usages pour les faibles utilisateurs, et une croissance d'usage pour les autres. Ce qui caractérise les faibles utilisateurs d'internet en termes de pratiques est la très faible présence du courrier électronique. [Lelong & Thomas, 2001] a montré que l'utilisation du courrier électronique, et donc l'accès à un réseau de correspondants sur internet était le facteur principal de l'ancrage et de la routinisation des pratiques d'internet. On voit ici que l'absence de correspondants s'accompagne d'un faible usage du web et préfigure des abandons.

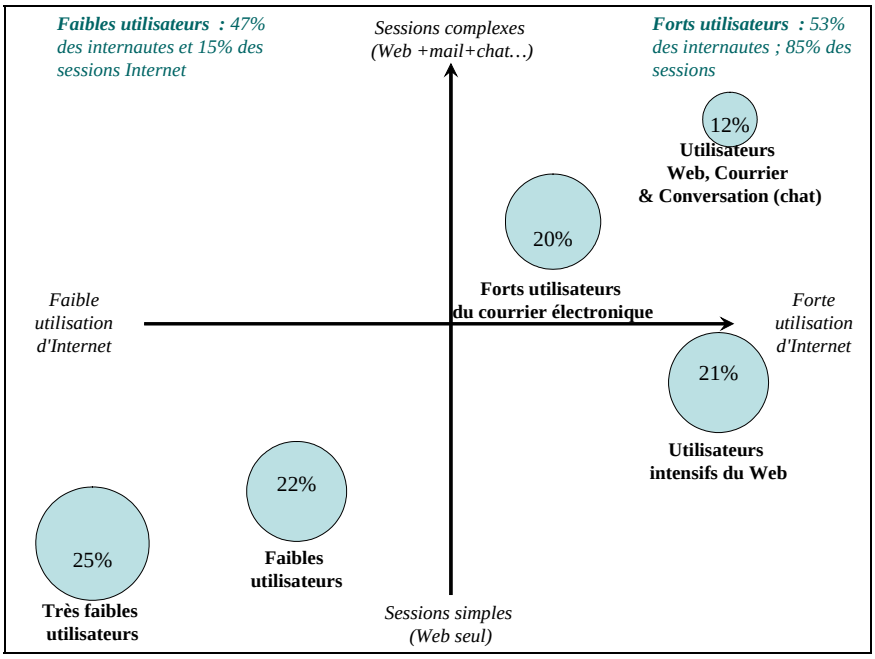


FIG 4 – *Typologie des internautes fondée sur le profil de leurs sessions internet*

Concentrons-nous maintenant sur les plus forts utilisateurs, parce que contrairement aux autres, leurs pratiques ont atteint une certaine forme de stabilisation. On y distingue trois groupes. Le premier, qui représente 22% de la cohorte, est constitué d'utilisateurs intensifs du Web pour qui la pratique de la communication interpersonnelle, et donc de l'écriture, est secondaire[2]. Pour les deux autres groupes d'internautes, les échanges interpersonnels priment sur la consultation de la toile. Le deuxième, qui représentent 20% des internautes, est constitué d'utilisateurs du courrier électronique[3] : si le courrier a une place centrale dans leurs pratiques, il n'en reste pas moins qu'ils fréquentent aussi le Web, certes nettement moins intensément que le premier groupe. Le troisième et dernier groupe (11 % des internautes) rassemble des individus qui pratiquent la conversation sur le réseau et utilisent donc les outils de communication en co-présence temporelle (chat et messagerie instantanée).

On se trouve confronté à un modèle étagé, où l'utilisation d'un dispositif supplémentaire ne s'accompagne pas de l'abandon des autres. Le premier groupe fait principalement du Web, accordant peu de place aux échanges[4], le deuxième, sans négliger la toile, accorde en plus une place très importante au courrier électronique combinant navigation et échanges "épistolaires", le troisième enfin combine navigation et courrier, comme le précédent, en y

[2] Ils consultent le Web dans presque toutes leurs sessions (85 %), n'accèdent au courrier électronique que dans 30 % de leurs sessions et ne font ni messagerie instantanée, ni chat.
[3] Ils consultent leur messagerie dans près des deux-tiers de leurs sessions et la navigation sur la Toile ne concerne que la moitié de leurs sessions. La conversation en direct est absente de leurs pratiques.
[4] Nous retrouvons ce type de profil dans les entretiens : certains utilisateurs disent explorer intensément la toile, mais ne cherchent pas à savoir qui se cache derrière un site et cherchent encore moins à entrer en contact.

ajoutant la conversation en direct. La manière d'investir les potentialités du réseau varie considérablement d'un groupe à l'autre.

Les utilisateurs du deuxième groupe marqué par une forte présence du courrier électronique, sont plutôt issus d'un milieu social élevé et les utilisateurs d'outils de conversation électronique (*chat* et messageries instantanées) de milieux plus modestes. Les foyers de cadres supérieurs se caractérisent par une désaffection prononcée pour les *chats* et les messageries instantanées. On peut avancer un certain nombre d'hypothèses : la valorisation de l'écrit et du livre dans les milieux favorisés entraînerait une dévalorisation de ces formes d'échange sans mémoire qui mobilisent un type d'écriture très éloigné des canons légitimes ; la pauvreté des contenus échangés et l'absence de finalités « informationnelle » rendraient également ces échanges suspects. Inversement, cet écrit sans mémoire, cadré par des normes locales, distinctes des normes habituelles, permettrait dans des milieux plus modestes de lever la barrière de l'écrit.

Une autre opposition notable se situe entre les jeunes et les plus âgés. Dans les pratiques d'Internet des jeunes, la place des outils de communication est centrale, mais leur spécificité véritable tient à leur capacité à articuler et combiner des pratiques très diverses : cette dextérité face à l'écran dans l'enchaînement de tâches diverses semble être ce qui les distingue des utilisateurs plus âgés.

Enfin, en ce qui concerne le genre, les femmes sont quasiment absentes dans le groupe des utilisateurs intensifs du Web, elles « se défendent bien » dans les deux autres groupes, ceux où la part des échanges interpersonnels est grande. Sur Internet, comme ailleurs, on observe cet engagement fort des femmes dans l'entretien du réseau relationnel.

5. Conclusion et perspectives

Les travaux se poursuivent dans le cadre du projet SensNet dans trois directions : élaborer une plateforme de traitement qui intègre le savoir-faire accumulé, progresser dans une voie d'enrichissement des données qui consiste à exploiter le contenu et la structure des pages et enfin analyser les usages et leurs évolutions en s'appuyant sur des données de mieux en mieux qualifiées.

Finalement, les progrès à faire dans la fouille des données web tiennent davantage à cette capacité à qualifier de manière pertinente des contenus qui évoluent sans cesse qu'aux méthodes de traitements statistiques mises en œuvre.

Références

[Assadi & Beauvisage, 2002] Assadi H. & Beauvisage T. *A comparative study of six French-speaking web directories*. The Seventh International ISKO Conference, Granada, Spain, 2002.

[Beaudouin & Licoppe, 2002] Beaudouin V. & Licoppe C., Eds. *Parcours sur internet*. Réseaux, n°116, Paris, Hermès, 2002.

[Beaudouin & Velkovska, 1999] Beaudouin V. & Velkovska J. "Constitution d'un espace de communication sur internet (Forums, pages personnelles, courrier électronique...)", *Réseaux*, 17, n°97, p. 121-177, 1999.

[Beaudouin *et al.*, 2001] Beaudouin V., Fleury S., Habert B., Illouz G., Licoppe C. & Pasquier M. *TyPWeb : décrire la Toile pour mieux comprendre les parcours*. CIUST'01

(Colloque International sur les Usages et les Services des Télécommunications -- e-Usages), Paris, ENST, 492-503, 2001.

[Beaudouin et al., 2002] Beaudouin V., Assadi H., Beauvisage T., Lelong B., Licoppe C., Ziemlicki. C., Arbues L. & Lendrevie J. Parcours sur Internet : analyse des traces d'usage. Suivi d'une cohorte d'internautes du panel NetValue France en 2000. RP-FTRD7495, 2002.

[DiMaggio *et al.*, 2001] DiMaggio P., Hargittai E., Russell N. W. & Robinson J. P. "Social Implications of the Internet", Annual Review of Sociology, 27, p. 307-336, 2001

[Lelong & Thomas, 2001] Lelong B. & Thomas F. (2001). *L'apprentissage de l'internaute : socialisation et autonomisation*. CIUST'01 (Colloque International sur les Usages et les Services des Télécommunications -- e-Usages), Paris, ENST, 74-85.

[Pasquier, 2002] Pasquier D. "Les signes de soi". Enquête sur l'organisation des sociabilités en milieu lycéen, rapport FTR&D, 2002.

[Smoreda & Licoppe, 1999] Smoreda Z. & Licoppe C. *La téléphonie résidentielle des foyers : réseaux de sociabilité et cycle de vie*. 2ème Colloque International sur les Usages et Services des Télécommunications, Bordeaux, 1999.

Summary

To construct a detailed representation of internet use practices and their evolution, we process detailed traffic data by panels of internet users. This paper describes overall methodology, and details our cohorts and the data refinement and analysis techniques applied. Then some results are presented to illustrate our approach, including a segmentation of users based on interlacing of browsing and communication practices.