

Etude de la perception de la qualité des sites Web par les usagers en vue de dimensions d'évaluation et résultats utiles à la conception

Ghuilaine Clouet

FranceTélécom R&D, DIH/EQS, 4 rue du Clos Courtel - BP 91226, 35512 Cesson Sévigné,
Université Paris 8, g.clouet@rd.francetelecom.com

Résumé :

Cette étude comparative des deux situations d'entretien individuelle et collective sur la perception de la qualité par les usagers des sites Web commerciaux permet de mettre en évidence l'apport des entretiens-utilisateurs par rapport à une grille de dimensions élaborée par des évaluateurs. Elle aborde également la question de la nécessité de la mise en corrélation des besoins en conception et de l'évaluation et par là-même, des dimensions d'évaluation ou de perception des usagers. Enfin, elle met à jour l'utilité des deux situations d'entretien pour l'élaboration des dimensions de perception des usagers et pour la production de résultats directement réutilisables dans la conception (au sens large, marketing, développement, etc.).

Mots-clefs : Internet, perception des usagers, entretien individuel, entretien collectif, besoins en conception.

1. Introduction

Avec l'apparition des sites Web, il s'agit d'analyser non plus des activités « liées exclusivement au travail mais aux loisirs, à la culture et à l'acquisition de biens » (Leulier, Bastien, Scapin, 1999). L'acteur observé dans cette activité n'est pas seulement celui qui interagit avec l'interface mais également celui qui va choisir un produit ou service Internet. Le champ de l'ergonomie « traditionnelle » s'élargit pour prendre en compte les divers facteurs d'appréciation des usagers et clients. La notion de qualité perçue englobe ainsi des dimensions qui touchent à l'utilité et à l'utilisabilité aussi bien qu'à des notions de préférences.

Afin d'évaluer au plus juste la perception des usagers (d'un point de vue quantitatif), il faut élaborer des questionnaires adaptés et d'abord mettre en évidence les critères de jugement ou encore les dimensions de qualité qui rentrent en compte dans la perception des usagers. Par ailleurs, les résultats d'évaluation selon ces dimensions sont destinés à être redéployés vers les différents domaines de conception. En vue de produire des résultats d'évaluation adaptés à ceux qui vont les utiliser pour reconcevoir les sites, il s'agit d'analyser leurs besoins. Ceci concerne aussi bien le marketing opérationnel et stratégique que le développement de l'interface et des structures techniques qui véhiculent les services (réseaux).

2. Description de l'expérimentation :

Notre étude compare la situation d'entretien individuelle et la situation collective tout d'abord pour l'élaboration des dimensions de qualité des usagers et aussi pour la mise en évidence de résultats directement utiles aux services de conception destinataires. Elle concerne le commerce par Internet.

Pour faire élaborer précisément aux usagers leurs critères d'appréciation, notre méthode est basée sur le recueil de verbalisations et sur la mise en situation d'internautes en interaction avec des produits ou services, ici, des sites Web commerciaux. Il s'agit de corréler l'analyse de leurs verbalisations et de leur activité, sans pour autant évaluer leurs performances.

Une première étape consiste à faire verbaliser l'utilisateur grâce à un entretien semi-directif, hors-usage. Ce dernier est amené à verbaliser sur sa perception des sites marchands sur la base de ses expériences antérieures, puis il remplit des « fiches-critères ». Une seconde étape plonge l'internaute dans une situation d'utilisation et de jugement comparatif de deux sites marchands. Après avoir verbaliser librement lors de son utilisation des deux sites (Nosulenko, 1997), l'utilisateur écrit une nouvelle série de fiches.

A quelques aménagements près, le même protocole d'« entretien-mise en situation » est utilisé pour les deux situations d'entretien. Ainsi, les deux entretiens collectifs menés, d'une durée de 2 heures ½, comptaient 8 participants des deux populations de sujets décrites ci-après, tandis que 8 entretiens individuels ont été réalisés avec la première population et 8 autres avec la seconde. Les résultats détaillés de l'entretien collectif avec la seconde population n'étant pas encore élaborés, nous nous basons seulement sur des observations, le concernant.

La première population comprend des sujets qui sont des internautes plus confirmés (1 à 7 ans contre 1 à 2 ans d'utilisation d'Internet, avec un usage plus intensif) et des usagers plus expérimentés des sites marchands (plusieurs achats avec paiement en ligne).

3. Résultats :

3.1. Apport des « entretiens –mise en situation » avec les usagers par rapport à une grille de dimensions de référence :

Une première grille de référence a été élaborée par quatre évaluateurs de formations complémentaires (2 ergonomes IHM, 1 ergonomiste plus spécialisé dans les sites commerciaux, 1 linguiste) sur la base de leurs expériences conjuguées et des heuristiques de Nielsen (1993 et Nielsen & Molich, 1990). L'extrait suivant donne un petit aperçu de la grille de référence :

12 : Sécurisation de la transaction

- ⇒ L'internaute qui commande en ligne est assuré qu'il peut saisir et transmettre ses coordonnées bancaires en toute sécurité et en toute fiabilité.
- ⇒ Des informations sont données aux moments stratégiques (ex : saisie n° CB, validation de la commande) qui garantissent intégrité et confidentialité de la transaction et qui informent du traitement qui va être fait derrière (enregistrement du n° de CB pour éviter de le faire transiter sur le réseau à chaque transaction, protection des renseignements personnels ...).

Tableau 1 : extrait de la grille de référence

Chaque élément de définition, écrit par les usagers dans les fiches-critères, a d'abord été enrichi de toutes les significations données dans les verbalisations lors de la première et de la seconde étape respectivement et des explications identifiées dans l'analyse de leur activité d'utilisation. Puis nous avons spécifié le degré de « nouveauté » de chacun de ses éléments au regard de la grille de référence. Certains éléments-utilisateurs dits « référents » font déjà partie de cette grille. Par exemple, un usager ayant écrit « *la sécurité du paiement* » nous apporte un élément déjà connu en rapport avec la dimension 12 de la référence. D'autres éléments-utilisateurs, nommés « spécifiant », viennent compléter des éléments de référence incomplets. Par exemple, « *le cryptage des données par serveur SSL* » vient spécifier l'élément suivant de la dimension 12 : *traitement qui va être fait derrière (enregistrement du n° de CB pour éviter de le faire transiter sur le réseau à chaque transaction)*. Enfin, d'autres éléments-utilisateurs sont véritablement nouveaux. « *L'assurance du paiement par un organisme extérieur* » ou « *assurance du label fia-net* » apportent des informations nouvelles qui ne sont pas données par la référence.

Nous avons ainsi pu dégager les quantités de « référent », de « spécifiant » et de « nouveau » apportées par les entretiens-utilisateurs par rapport à la grille de référence.

Sur la base des résultats de l'entretien collectif de la première population et des 16 entretiens individuels des deux populations, les entretiens-utilisateurs apportent en moyenne 35% d'éléments « nouveaux » et 14,5% d'éléments qui viennent spécifier ceux de la référence. Autrement dit, la moitié (50,5%) des éléments apportés par les usagers viennent enrichir la grille élaborée par les évaluateurs. L'utilité voire la nécessité des « entretiens-mise en situation » avec les usagers est indéniable.

3.2. Elaboration d'une nouvelle grille :

Les éléments de la grille de référence étaient, en fait, structurés en 25 dimensions en tout. Les éléments des usagers et la prise en compte des besoins en conception nous ont amenés à un redécoupage de ces dimensions. Nous avons ainsi obtenu 31 dimensions. Ceci ne signifie pas pour autant que 6 dimensions soient nouvelles. L'ensemble des éléments des 25 dimensions de la grille de

référence a été étalé, en quelque sorte, de façon à mieux prendre en compte l'activité des usagers. En effet, du fait de l'apport des éléments verbaux des usagers et de l'analyse de leur activité, le découpage de la nouvelle grille reflète davantage cette activité. Elle met en évidence tout le contenu des critères de jugement des usagers et constitue une grille de dimensions de perception, destinée dans un premier temps aux évaluateurs. En reprenant précisément les termes et les formulations les plus communes des usagers, cette grille permet de produire le questionnaire d'évaluation.

Par ailleurs, dans le contenu de ces dimensions de perception, sont mis en évidence les éléments (de l'interface ou autres) sur lesquels portent le jugement des usagers, qui sont donc les éléments à évaluer. Cette grille est ainsi une grille de dimensions d'évaluation et révèle des éléments en rapport avec les besoins de la conception (au sens large).

Ainsi, le redécoupage a également été réalisé dans le sens de considérer les besoins du marketing et du développement technique. Cette prise en compte préalable n'entrave pas le découpage, précédemment décrit, adapté à l'activité des usagers. Par exemple, les notions de « niveau de prix » et de « la clarté de l'affichage des prix » des produits disponibles à l'achat ont été séparées. Ceci permet de tenir compte du marketing qui sera directement concerné par l'appréciation du niveau des prix par les usagers et de considérer aussi le développement technique, davantage concerné par l'affichage des prix. Ceci n'entrave pas la façon dont les usagers ont parlé de leur perception des sites marchands étant donné que peu d'entre eux ont regroupé ces deux notions sur une même fiche-critère. Le redécoupage, dans cette perspective, concerne quelques autres dimensions et demeure, pour l'instant, partiel voire « simpliste ».

3.3. Comparaison des performances de la situation d'entretien, individuelle versus collective :

En comparant l'entretien collectif et les entretiens individuels avec les sujets expérimentés, nous n'observons aucune différence significative en termes de quantité d'éléments « nouveaux » (respectivement 32% contre 36%) ni même « spécifiques » (14,5% contre 19%) par rapport à la référence. En revanche, chaque élément-utilisateur a été classé selon son degré de détail, quelle que soit la catégorie à laquelle il appartenait en termes de degré de nouveauté par rapport à la référence. Par exemple, « *le cryptage des données par serveur SSL* » et « *l'assurance d'un organisme extérieur* » sont deux éléments-utilisateurs respectivement « spécifique » et « nouveau » et ce sont aussi deux éléments de détail par rapport à l'élément plus générique : « *sécurité du paiement* ». Ainsi, la situation individuelle obtient 49% de détails et l'emporte clairement sur la situation collective qui n'en compte que 23%. De plus, la répartition de la fréquence d'énonciation des éléments, sur chacune des 31 nouvelles dimensions, semble révéler une plus forte dépendance entre les scores selon chaque dimension et les discussions « lancées » par les différents participants dans la situation de groupe. Ainsi, l'importance relative de chaque dimension obtenue dans les entretiens individuels nous semble plus objective. Il est connu que la présence du groupe peut amener les participants à exprimer des opinions plus extrêmes ou au contraire plus conformistes (Morgan, 1997, Mucchielli 1996). Dans notre cas, nous assistons à des discussions très vivantes entre les participants et trouvons qu'ils ont une tendance à exprimer des opinions plus extrêmes qu'ils ne le feraient dans un échange à deux, avec seulement l'interviewer.

Tout ceci nous amène à croire que la situation d'entretien individuelle est la plus favorable pour l'élaboration des dimensions de qualité avec une certaine précision.

Par ailleurs, la phase d'utilisation est plus performante pour déceler les difficultés rencontrées par les sujets en passation individuelle qu'en passation collective (où un seul participant navigue sur le site selon les décisions collectives du groupe). Ainsi, les entretiens individuels sont également préférables pour obtenir des détails sur l'utilisabilité, utiles pour le développement de l'interface.

Toutefois, étant donné le caractère majeur des thèmes des discussions entre les participants du groupe (sécurité, confidentialité des données, qualité des produits, etc.), nous pensons que les entretiens collectifs permettent de révéler les problèmes les plus couramment rencontrés et surtout les inquiétudes et préférences principales ressenties par les usagers. Autrement dit, la situation collective permet de dégager rapidement les grandes lignes, résultats plus utiles au marketing dans le cadre d'une étude préliminaire. En effet, cela permet de déterminer quels sont les grands points sur lesquels il

s'agit de travailler pour améliorer les sites commerciaux sans pour autant avoir d'éléments de détail précisant comment il faut les reconcevoir.

Les discussions de l'entretien collectif ont également apporté 9% d'explications détaillées que nous n'avons pas obtenues dans les entretiens individuels (ni dans les 8 entretiens avec la première population ni même dans les 8 autres avec la seconde). Nous pensons que cette part d'explications, bien que faible, n'aurait pas été élaborée ou plus difficilement en entretien individuel. Ces éléments ont été exprimés lors de confrontations de motivations divergentes dans l'utilisation des sites marchands entre les participants. La force des entretiens collectifs réside en effet dans les interactions de groupe (Morgan, 1997, Krueger, 2000) et peut certainement être utilisée à meilleur escient sur les points cruciaux des sites marchands et/ou requérant la confrontation de points de vue car en rapport avec différents usages. Autrement dit, il s'agirait de mener non pas des entretiens de groupe sur l'élaboration générale des critères de jugement mais plus spécifiquement des « focus groupes » sur, par exemple, la question des éléments et indices de sécurité des sites marchands, des moyens de recours en cas de problèmes, de l'efficacité de leur moteur de recherche, etcetera. De même, Bouch, Kuchinsky et Bhatti (2000) ont réalisé des « focus groupes », complémentaires à des études comportementales, qui se sont révélés riches en explications sur les raisons sous-jacentes au ressenti et aux agissements des sujets, aussi dans le domaine du commerce par Internet.

Tout ceci rejoint les propos de Nielsen (1993) qui préconise l'utilisation de « focus groupes » pour un apport informatif avant ou après la conception d'une interface.

Enfin, sur la base de simples observations, nous estimons que l'entretien collectif avec les sujets expérimentés est plus fructueux que celui avec les sujets novices du fait de pouvoir confronter des expériences plus précises et des points de vue plus affirmés.

Pour les sujets novices, nous pensons qu'il vaut mieux privilégier la situation d'entretien en face à face et particulièrement la phase d'utilisation pour comprendre leurs difficultés et leurs aspirations quant à la conception de ces sites. Dans la comparaison entre les entretiens individuels avec la première et la seconde population, nous notons que les sujets novices sont moins prolifiques dans l'écriture d'éléments de définition dans les fiches-critères. La part d'enrichissement de ces éléments avec les significations apportées par les verbalisations et par l'analyse de l'activité est aussi légèrement plus forte avec les sujets novices (164%) qu'avec les sujets expérimentés (146%). Les sujets novices semblent élaborer moins facilement de critères dans les fiches-critères et donnent plus de précisions lors des entretiens. De plus, davantage d'éléments expliquant leurs verbalisations plus approximatives sont révélés par l'analyse de leur activité.

4. Conclusion :

Premièrement, l'apport des « entretiens-mise en situation » avec les usagers des sites marchands est tout à fait significatif par rapport à la grille de référence élaborée par les évaluateurs, même si de plus amples recherches devront être entreprises pour savoir plus précisément comment cette grille elle-même a été élaborée.

Par ailleurs, afin de mieux comprendre et refléter les besoins des différents domaines de conception, il est nécessaire que la nouvelle grille évolue et soit validée par différents acteurs. Il s'agit notamment d'éviter un cloisonnement des services de conception dans cette prise en compte. La complémentarité des résultats en ergonomie et en marketing a, par exemple, déjà été étudiée et prouvée dans le champ des télécommunications (Aytéo, Sidhu, Coyle, Robinson, 1996).

Enfin, davantage de résultats sont nécessaires et seront apportés ultérieurement pour consolider les conclusions dégagées concernant les deux situations d'entretien avec les sujets expérimentés ou novices. De plus, du fait du caractère très lié entre les performances identifiables des deux situations d'entretien et du domaine de validité (sur quel thème ? des résultats servant à quoi, à qui ?), comme le souligne Morgan (1997), la question de quelle méthode est préférable dans quelles circonstances est essentiellement une question empirique et davantage de recherches sont encore nécessaires pour y répondre. Dans le domaine précis des sites commerciaux et par la suite des sites Web en général, nous espérons palier à cette faiblesse.

Références bibliographiques :

- Aytéo, M., Sidhu, C., Coyle, G., Robinson, S. (1996). Working with marketing. *Proceedings of CHI'96 Companion* (Vancouver, BC, Canada, 13-18 April), pp.313-314, ACM Press.
- Bouch, A., Kuchinsky, A., Bhatti, N. (2000). Quality is in the eye of the beholder : meeting users' requirements for Internet quality of service. *Proceedings of CHI 2000* (The Hague, Netherlands, 1-6 April), pp. 297-304, ACM Press.
- Kruegger, R.A. (2000). *Focus Group : A practical guide for applied research* (3rd ed.- 1st ed.: 1994), Sage, Thousand Oaks, CA.
- Leulier, C., Bastien, J.M.C., Scapin D.L. (1999). L'évaluation des sites Web : Apports respectifs des questionnaires de satisfaction et de l'inspection ergonomique. *Actes de IHM'99, Interaction pour tous* (Montpellier, 22-26 Novembre), pp. 49 - 52.
- Morgan, D.L. (1997). *Focus Groups as qualitative research*, Qualitative research methods vol.16, (2nd ed. - 1st ed.: 1988), Sage, Thousand Oaks, CA
- Mucchielli, R. (1996). *L'interview de groupe*, (2ème éd. - 1ère éd.: 1987), E.S.F., Paris.
- Nielsen, J. (1993). *Usability Engineering*. MA Academic Press, San Diego, CA.
- Nielsen, J., Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of CHI'90* (Seattle, WA, 1-5 April), pp.249-256, ACM Press.
- Nosulenko, V.N., Samoylenko, E.S. (1997). Approche systémique de l'analyse des verbalisations dans le cadre de l'étude des processus perceptifs et cognitifs, in *Information sur les sciences sociales*, vol.36(2), pp.223-261, Sage.