

Première partie

La mise en ligne d'un questionnaire : fondements théoriques et empiriques

Introduction de la partie I

Les taux de participation aux enquêtes sont en nette diminution, et les enquêtes ménages déplacements suivent la même tendance. Même si certaines procédures permettent de réduire le taux de non-réponse et que les méthodes de redressement permettent de limiter le biais introduit par la non-réponse totale, il reste nécessaire de postuler que les comportements de mobilité des ménages qui refusent de participer à l'enquête sont identiques à ceux des ménages ayant des caractéristiques socio-économiques similaires. Or, de nombreux travaux (Bonnel, 2003; Ampt, 1997; Sammer, 1997) ont montré que les comportements étaient généralement corrélés au fait de ne pas répondre, ou au nombre de tentatives nécessaire pour atteindre un répondant. C'est pour essayer de limiter ce biais de non-réponse, et ainsi la représentativité statistique des résultats obtenus, que le Laboratoire d'Economie des Transports a proposé de réaliser une enquête par le web auprès des non-répondants à l'enquête ménages déplacements réalisée à Lyon en face-à-face en 2006.

Pour réaliser ce test, nous avons adapté le questionnaire papier, traditionnellement utilisé dans les enquêtes ménages en face-à-face, en tenant compte de plusieurs contraintes. D'une part, des contraintes méthodologiques, imposées par le Certu (1998) pour toute réalisation d'enquêtes ménages déplacements (thèmes, vocabulaire, ordre et formulation des questions...). D'autre part, des contraintes techniques, imposées par le cahier des charges de l'enquête de Lyon. Le web étant un module additionnel à l'enquête en face-à-face, nous n'avons pu choisir ni le prestataire en charge de l'administration de l'étude en-ligne, ni le logiciel utilisé pour le développement du questionnaire.

L'objectif de cette partie est d'évaluer le potentiel du web pour les enquêtes ménages déplacements, principalement dans le cadre d'une méthodologie d'enquête mixte (Chapitre 1). Suivent des réflexions sur la construction du questionnaire web, les choix opérés par rapport à sa version papier et l'administration de l'enquête en ligne (Chapitre 2). Nous présenterons finalement les premiers résultats de l'option web de l'enquête ménages déplacements de Lyon, en termes de taux et de comportement de réponse (chapitre 3). Grâce à la méthodologie employée, il est en effet possible d'étudier la progression des internautes dans le questionnaire, et comprendre pourquoi certaines personnes ont préféré répondre en ligne plutôt que de participer à l'enquête face-à-face.

Chapitre 1 : Le potentiel du web pour les enquêtes ménages déplacements

"Technology by itself is not inherently good or bad. It is how technology is harnessed in the service of human endeavor that determines its effect."
(Couper, 2002).

Un des principaux enjeux de l'économie des transports est de comprendre comment les comportements de mobilité varient selon les caractéristiques socio-économiques des individus. Des études récurrentes, comme les enquêtes ménages déplacements, viennent alimenter de larges bases de données qui permettent aux analystes de rédiger des recommandations pertinentes. Avec la démocratisation de l'informatique et de l'accès à Internet, les enquêtes web sont promises à un rapide développement. Couramment utilisées dans certains domaines comme le marketing (historiquement, les premières enquêtes en ligne abouties sont apparues aux USA en 1997), il semble intéressant de s'interroger à présent sur la pertinence de ce nouveau média pour les enquêtes de mobilité, et de déterminer sous quelles conditions l'utilisation du web peut augmenter la précision des données produites, en tenant compte des non-répondants aux enquêtes traditionnelles. Si les enquêtes web présentent de nombreux atouts (faible coût, rapidité d'administration, personnalisation du questionnaire...), leurs principales limites résident essentiellement dans sa diffusion et son niveau d'appropriation. Le taux de pénétration d'internet dans la population française progresse régulièrement, mais ce média n'est pas encore accessible à tous et les internautes ont des caractéristiques sociodémographiques particulières, susceptibles d'impacter leurs comportements de mobilité. L'utilisation exclusive du web pour mener des enquêtes auprès de l'ensemble de la population ne permet donc pas une généralisation des résultats par les techniques d'inférence statistique.

La qualité des décisions en politique des transports dépend fortement de la qualité des informations disponibles. La difficulté croissante pour obtenir des données d'enquêtes représentatives de la population visée et la complexité croissante des données nécessaires à l'alimentation de modèles de plus en plus sophistiqués ne permettent généralement plus de recueillir toutes les données

selon une méthodologie unique. Une des solutions consiste aujourd’hui à intégrer le web dans un protocole d’enquête plus large. L’idée est de réussir à capter de nouveaux individus à l’aide d’un média attractif, tout en conservant la méthode de recueil de données traditionnellement utilisée dans les enquêtes ménages déplacements.

Nous revenons d’abord sur les apports du web comme méthode d’enquête, avant de nous intéresser aux limites de ce nouveau média (section I). Nous étudions plus spécifiquement le potentiel du web dans les protocoles d’enquêtes mixtes, ainsi que l’impact sur la qualité des données recueillies (section II). Enfin, l’analyse des caractéristiques sociodémographiques de la population d’internautes et de l’évolution du taux de pénétration d’internet dans les ménages français clôturent ce chapitre (section III).

I Apports et limites du web pour les enquêtes de mobilité

I.1 Une grande diversité d’enquêtes

Les enquêtes électroniques ont pris des formes très variées, suivant les développements des technologies de la communication¹³. Il est utile de distinguer les enquêtes conduites sur Internet, de celles diffusées par mail (message incluant des questions, ou e-mail avec lettre avis et questionnaire en pièces jointes). Les enquêtes diffusées par e-mail sont peu coûteuses, simples, rapides, personnalisées, et encouragent la réponse (Tse, 1998). Mais les messages peuvent être rapidement effacés, ne sont pas matérialisés ni anonymes et retiennent rarement l’attention des destinataires (Kittleson, 1995). Une enquête web est un questionnaire interactif parfois très complexe, auquel les individus peuvent répondre en se connectant sur internet (Stanton, 1998). Les personnes qui disposent d’un ordinateur, équipé d’un navigateur et d’une connexion internet, doivent se rendre sur un site web, compléter le questionnaire et renvoyer leurs réponses. La base de données est implantée sur un serveur et les données sont transmises électroniquement à l’analyste (Alvarez et Van Beselaere, 2003).

Les enquêtes web peuvent être très diverses, suivant le mode d’échantillonage, l’accessibilité du questionnaire, sa présentation et les modes de recueil alternatifs éventuellement proposés. Leur qualité est par conséquent très inégale. Des classifications ont été données dans la littérature. Couper (2000), Bradley (1999) et Hojman *et al.* (2004) discriminent les enquêtes web selon le mode de recrutement des répondants (échantillon probabiliste ou non probabiliste), qui n’offre pas les mêmes possibilités de contrôle des données. Les méthodes probabilistes reposent sur deux principes : la nécessité de disposer

¹³Les premières enquêtes électroniques consistaient en une disquette envoyée aux répondants. Ces derniers introduisaient la disquette dans l’ordinateur pour lire le fichier et répondre au questionnaire, puis enregistraient leurs réponses et renvoyaient le support aux administrateurs de l’enquête (Couper et Nichols, 1998).

d'une base de sondage, c'est-à-dire d'une liste complète et actualisée des individus qui appartiennent à l'univers de l'étude, et une probabilité connue de désignation de ces unités de sondage. Ces méthodes ne garantissent pas la représentativité de l'échantillon, puisque les non-réponses peuvent menacer la valeur inférentielle de l'étude, mais il est possible d'évaluer la non-réponse et d'ajuster les données. Il y a deux moyens de mener des enquêtes probabilistes en ligne, en restreignant l'échantillon à ceux qui possèdent un accès internet ou en utilisant plusieurs modes de recueil de données, afin de toucher l'ensemble de la population (tableau 2). Dans la plupart des cas, les commanditaires d'études ne disposent pas de base de sondage et utilisent des méthodes empiriques, qui nécessitent certains arbitrages pour constituer l'échantillon. L'inférence statistique est impossible, bien que la représentativité de l'échantillon puisse être améliorée en s'efforçant de reproduire au mieux les conditions du hasard pour la sélection des répondants. Schillewaert *et al.* (1998) et Schaefer et Dillman (1998) passent en revue les différents moyens disponibles pour collecter des données sur le web et proposent quelques pistes d'amélioration des procédures de recrutement.

Alsnih (2005) considère l'accessibilité au questionnaire, qui peut être ouvert à tous, sécurisé par un code d'accès personnel ou caché sur un site ('popup'). Dillman et Bowker (2001) insistent sur la présentation, et distinguent les questionnaires interactifs (contenant une question par page) des questionnaires passifs (permettant de visualiser l'ensemble des questions sur une seule page, grâce à l'ascenseur)¹⁴. Dans les questionnaires interactifs, le répondant doit valider chaque réponse en cliquant sur un bouton, ce qui permet de récupérer les réponses sur le serveur au fur et à mesure qu'elles sont saisies, mais rend difficile le retour en arrière et quasi impossible la correction des réponses déjà validées (Alvarez et Van Beselaere, 2003). Si proposer plusieurs questions sur une même page est plus motivant pour l'enquêté, le temps de chargement des pages web est élevé et les biais de mesure plus importants, les répondants ayant tendance à cocher systématiquement la même modalité (Couper *et al.*, 2001). Dans un questionnaire passif, les réponses sont transmises à la fin de la saisie. Il existe donc un risque de perte totale des données, en cas de rupture de la connexion internet, mais les réponses sont facilement modifiables par l'internaute et les difficultés techniques moins prononcées.

Malhotra (1999) oppose la constitution d'un échantillon 'interne', où les répondants sont sélectionnés sur internet, à celle d'un échantillon 'externe', où les répondants ne sont pas recrutés sur le web (par exemple dans l'annuaire), mais sont invités à remplir un questionnaire en ligne. Enfin, des enquêtes sont exclusivement réalisées grâce à Internet, notamment lorsqu'il s'agit d'appréhender le comportement ou l'opinion d'une population d'internautes, alors que d'autres combinent plusieurs modes de recueil de données (Bonnel, 2004).

Il existe une multitude d'enquêtes web, qui se différencient par leur niveau de qualité (Couper, 2001). Le choix du type d'enquête et son implication sur la méthode d'échantillonnage dépend au final de l'objectif de l'étude. Dans

¹⁴Dans ce cas, le temps de réponse est plus court et les non-réponses partielles moins nombreuses (Gunn, 2002).

ce rapport, nous considérerons uniquement les enquêtes menées sur Internet, conformément à la méthodologie utilisée pour l'enquête ménages déplacements de Lyon. Les répondants web sont des ménages initialement recrutés de façon aléatoire, pour participer à l'enquête-en-face à face. Cependant, le questionnaire en ligne n'est proposé qu'à ceux qui refusent de recevoir une enquêteur à domicile ou qui restent injoignables. L'enquête web n'est donc pas représentative de la population.

Méthode	Type d'étude	Mode de recrutement
Méthodes non probabilistes	Sondage 'divertissement'	Recueil d'opinion, non représentatif
	Enquêtes ouvertes et non restreintes	Participation à une étude offerte sur plusieurs sites
	Panels de volontaires	Individus recrutés pour répondre à des études ciblées
Méthodes probabilistes	Enquête systématique	Invitation de tous les visiteurs d'un site web
	Enquête par sondage basé sur des listes	Individus recrutés à partir d'une liste d'e-mail
	Enquêtes mixtes	Possibilité de répondre en ligne ou par d'autres modes
	Panels d'internautes prérecrutés	Recrutement probabiliste et proposition d'étude en ligne
	Echantillon aléatoire, accès web fourni	Recrutement probabiliste et étude en ligne (équipement fourni)

TAB. 2 – Les différentes méthodes de sélection de l'échantillon pour réaliser une enquête web

Source : D'après Couper (2000)

I.2 Pourquoi un tel engouement pour les enquêtes web ?

Les enquêtes déplacements sont généralement longues et coûteuses à mettre en oeuvre. Il s'agit le plus souvent d'interroger un large échantillon de ménages, ou d'individus, au sujet de leurs déplacements quotidiens, à l'aide d'un questionnaire relativement lourd et administré en France en face-à-face, à tous les membres du foyer. Les implications en termes de disponibilité des répondants et du nombre d'enquêteurs sont importantes. L'augmentation du taux de non-réponse a apporté de nombreux questionnements sur cette méthodologie, les comportements de mobilité des non-répondants ne pouvant être considérés comme identiques à ceux des individus interrogés. Dans ce contexte, l'utilisation d'internet est apparue comme une voie intéressante. L'absence de registre officiel d'adresses e-mail et la difficulté d'administration des enquêtes par e-mail (problèmes de téléchargement, d'affichage ...) ont laissé la place aux enquêtes web (Dillman *et al.*, 1998a).

I.2.1 Un coût marginal quasiment nul

Le coût des études est directement lié à la main d'oeuvre employée, c'est-à-dire au nombre d'enquêteurs (Bonnel, 2003). Les coûts générés par une enquête web ne sont donc pas très élevés par rapport à une enquête traditionnelle en face-à-face (Cleland, 1996). Il n'est pas nécessaire de recruter des enquêteurs, ni de dupliquer les questionnaires sur un support papier. Les coûts engendrés par l'affranchissement des lettres, les communications téléphoniques ou les déplacements des enquêteurs au domicile des interviewés sont inexistant. Il est également possible de modifier aléatoirement l'ordre des questions, sans supporter les coûts relatifs à la création de documents papier. De plus, le codage et la saisie des réponses se font par l'enquêté, directement sur l'ordinateur, sans recours à des opérateurs de saisie. Il en est de même pour le tri, la classification et la vérification des questionnaires. Le stockage physique des questionnaires n'est plus nécessaire, ce qui représente encore des économies en termes d'espace et de manutention (Keesling, 2002). Enfin, les relances, qui sont un moyen efficace de réduire la non-réponse, peuvent être automatisées sans générer de dépenses importantes. Il faut relativiser les avantages financiers des enquêtes web, et garder en mémoire que leur coût varie selon le nombre de relances effectuées auprès des non-répondants et le mode de recrutement mis en place. S'il existe une base de sondage disponible (liste d'adresses e-mail), alors le coût de mise en œuvre de l'enquête web est marginal. En revanche, si la base d'échantillonnage doit être construite, le coût de réalisation de l'enquête web est quasi-identique à celui des autres modes.

L'utilisation de ce nouveau média permet ainsi d'augmenter le nombre de contacts à moindre frais, et de diminuer le coût moyen par contact, le coût des enquêtes web augmentant de façon non linéaire avec la taille de l'échantillon (Crawford *et al.*, 2002). Les coûts variables sont faibles et correspondent essentiellement au support technique mis à la disposition des répondants. Les coûts fixes, non négligeables, varient selon la longueur et la complexité de la programmation du questionnaire. La trame peut toutefois être stockée, puis réutilisée ultérieurement. Le web est actuellement le média le moins coûteux¹⁵, en particulier pour de gros échantillons (Schonlau *et al.* (2001), Couper (2001) et Anderson *et al.* (2006)), et le développement de logiciels spécifiques ces dernières années devrait contribuer à réduire ces coûts fixes (Bonnel, 2004).

I.2.2 L'interactivité du questionnaire

L'interactivité du questionnaire, traditionnellement utilisée dans les enquêtes administrées par ordinateur au téléphone ou en face-à-face (CATI et CAPI), est également possible dans des enquêtes auto-administrées sur le web.

¹⁵En Australie, environ 350\$ sont dépensés par ménage répondant en face-à-face à une enquête de mobilité (Stopher et Greaves, 2006). Selon Crawford *et al.* (2002), le coût pour une étude auprès de 3500 répondants est 40% plus cher par la poste que par le web. Une expérience menée par Braunsberger *et al.* (2007) montre que le coût d'une enquête web est environ égal à 30% de celui d'une enquête téléphonique. Enfin, Cobanoglu *et al.* (2001) mettent en évidence que le web est moins cher que le fax et la poste quand le nombre de personnes dépasse les 200, d'autant plus si on ajoute le coût de codage des réponses.

L’assistance fournie par l’ordinateur décharge l’enquêté du travail de sélection des questions et de décryptage de la structure du questionnaire, généralement nécessaire dans le cas d’enquêtes auto-administrées. Il s’agit d’un mode dynamique, qui sollicite le répondant en cas d’incohérence des réponses, renseigne sur le temps restant pour terminer le questionnaire et fournit des informations précises avant chaque question. Le web autorise une grande interactivité au niveau de la personnalisation des questions et des réponses. Grâce aux nombreux filtres, les individus ne se voient pas poser des questions qui ne les concernent pas (Pitkow et Recker, 1995) ; la clarté du questionnaire venant réduire leur niveau de frustration (Gunn, 2002; Alvarez et Van Beselaere, 2003). Il est ainsi possible de s’adapter à la situation de chaque répondant et de cibler le questionnaire sur les questions pertinentes au regard de cette situation. La personnalisation du contact est également un atout intéressant des enquêtes web. En ne posant que quelques questions jugées ‘qualifiantes’ aux répondants, avant de soumettre l’intégralité du questionnaire aux individus sélectionnés, le gestionnaire d’étude peut segmenter l’échantillon de façon très précise. Par ailleurs, le web offre la possibilité de rendre aléatoire l’ordre des questions, sans complexifier l’administration du questionnaire. Enfin, diverses avancées technologiques voient le jour, comme l’intégration d’animations sonores ou visuelles (photos, vidéos, couleurs) et l’ajustement d’échantillon en cours d’étude. En ce qui concerne plus précisément les enquêtes de mobilité, des programmes permettent de géocoder instantanément les adresses saisies en liant le questionnaire web à des bases de données externes (Kreitz, 2001).

I.2.3 Des données de qualité

Selon Zmud (2003), augmenter la qualité d’une enquête consiste à réduire la non-réponse totale et partielle, ainsi que l’erreur de mesure. Dans les enquêtes web, les risques d’erreur sont plus faibles, qu’il s’agisse des erreurs de saisie ou du biais que peuvent générer les tierces personnes sur les réponses de l’enquêté (Krosnick et Chang, 2001). Des contrôles automatiques et dynamiques tout au long du questionnaire garantissent la validité des réponses, en limitant les erreurs et les incohérences dans la saisie (Adler *et al.*, 2002). L’enquêté est relancé, afin de corriger les réponses détectées comme suspectes, avant de pouvoir progresser dans le questionnaire (Saris, 1991). Les questionnaires obtenus sont ainsi apurés et aucun traitement ad-hoc n’est nécessaire¹⁶.

Le répondant, libéré de sa relation avec l’enquêteur et les autres membres du ménage, peut plus facilement se livrer. Les données saisies sont transcris dans un fichier sans intervention humaine et ne souffrent pas du biais de l’interviewer. Les normes de comportement, les préoccupations des jugements et la désirabilité sociale sont donc réduites en ligne (Mc Daniel et Gates, 2004)¹⁷.

¹⁶Certains programmes vont plus loin dans la personnalisation du contact et proposent des corrections au répondant (Franceschini, 2000).

¹⁷Le biais de désirabilité sociale se définit par la tendance pour une personne à donner des réponses selon les espérances culturelles perçues des autres. Ceci se produit surtout dans les interviews en face-à-face, puisque l’enquêteur est présent (Robinson *et al.*, 1991; De Leeuw et Van Der Zouwen, 1992).

Des études montrent que les réponses fournies par internet sur des sujets sensibles ou des comportements peu socialement désirables sont plus franches que lorsque l'étude est menée par internet (Joinson, 1998; Tourangeau et Smith, 1996). Selon Cobanoglu *et al.* (2001), les enquêtes web donnent des réponses de meilleure qualité que les enquêtes menées par fax ou par voie postale. Les réponses fournies aux questions ouvertes sont plus complètes et plus pertinentes sur le web (Yun et Trumbo, 2000). Lozar Manfreda et Vehovar (2002a) montrent que la fréquence d'utilisation du web impacte positivement la qualité des réponse, en particulier pour les questions difficiles. Enfin, Stanton (1998) utilise la variance des items comme indicateur de la qualité des données d'enquête. Ses recherches mettent en évidence que le schéma de covariance interne des données web est similaire à celui des données collectées par des modes plus traditionnels.

I.2.4 Des données collectées rapidement

Les études en ligne sont simples à mener, qu'il s'agisse de la diffusion du questionnaire, des relances ou de la phase de suivi, car le mode de communication est quasiment instantané (Schmidt, 1997; Weible et Wallace, 1998). Le répondant entre directement ses réponses, le support informatique supprimant tout délai dû renvoi du questionnaire, par rapport à une enquête postale (Alsnih, 2004). La centralisation de l'enquête sur un serveur rend possible un suivi régulier des réponses et la mise en place d'ajustements en temps réel, le gestionnaire de l'étude pouvant accéder à tout moment et de façon sécurisée aux données collectées. De plus, les données saisies sont déjà sous format électronique, et sont accessibles 'immédiatement' par les analystes. Certains logiciels d'enquêtes en ligne permettent même de générer automatiquement des analyses statistiques simples pour visualiser rapidement les premiers résultats d'une étude. Le caractère dynamique du web permet de générer des statistiques intermédiaires et d'envoyer un retour aux répondants (Schmidt, 1997). La diffusion de l'enquête, la collecte des données et le traitement de l'information sont donc plus rapides que dans les modes traditionnels.

Les protocoles mixtes montrent que les enquêtes web génèrent des réponses plus rapides que les autres médias. Dans la plupart des cas, la différence de temps de réponse entre une enquête web (envoyée par mail) et une enquête postale est de 6 à 10 jours, expliquée pour beaucoup par l'efficacité du service postal (Cobanoglu *et al.*, 2001; Kwak et Radler, 2000; Schaefer et Dillman, 1998; Bates, 2001). Franceschini (2000) rapporte que plus de deux tiers des réponses web parviennent à l'analyste avant la réception d'un questionnaire par voie postale. Potoglou et Karanoglou (2008) précisent que l'avantage du web en termes de vitesse de réponse dépend également beaucoup du mode de recrutement des individus. Si ces derniers sont contactées par un média technologique (internet, e-mail ...), alors la réponse est très rapide, mais si le recrutement a été fait par un mode non technologique (poste, téléphone ...), alors la réponse est plus longue et le coût plus élevé.

I.2.5 Un mode moins contraignant

Les enquêtes administrées, particulièrement dans le cas du face-à-face, sont souvent perçues par les répondants comme une intrusion dans la sphère privée, ce qui pèse sur le taux de réponse. Le caractère peu intrusif et non contraignant du web en terme de disponibilité temporelle permet de toucher davantage d'individus (Alsnih (2005) et Ampt (2003)). La connexion en ligne est souvent personnelle, et habituelle, aussi bien dans le foyer qu'au travail et la technologie utilisée permet au répondant de remplir le questionnaire par partie, les données recueillies étant stockées dans une base et rappelables ultérieurement par l'intermédiaire d'un mot de passe. L'enquêté est donc libre de répondre à l'enquête à son domicile ou au bureau, à un horaire qui lui convient.

A la suppression de la barrière temporelle s'ajoute une flexibilité spatiale. Les enquêtes web autorisent le recueil de données dans des zones géographiques relativement étendues, et permettent d'enquêter des personnes ne pouvant pas être jointes facilement. Qu'il s'agisse d'individus peu libres en journée, ne souhaitant pas recevoir un enquêteur, ou de personnes fortement contraintes comme les professions libérales, le web semble mieux adapté que les modes traditionnels. Les jeunes sont également attirés par ce nouveau média, qui leur est familier (Stopher *et al.*, 2004b), et l'utilisation du web est un moyen d'augmenter le taux de réponse auprès de ces classes d'âge. Enfin, mener une enquête en face-à-face ou par téléphone génère des coûts salariaux importants, qu'il est nécessaire de prévoir en amont et de borner en limitant la durée du recueil de données. Il est en revanche possible de proposer une enquête en ligne de façon récurrente ou quasi-permanente¹⁸ sans alourdir le budget de l'étude, la base de données s'alimentant automatiquement des nouveaux questionnaires remplis par les internautes. Le couplage du web avec les nouvelles technologies telles que le GPS facilite également le recueil de données de mobilité sur une longue période.

I.2.6 Le suivi du comportement de réponse

Les enquêtés diffèrent non seulement par leurs réponses aux variables d'intérêt, mais aussi par leur attitude de compléction du questionnaire. Grâce aux enquêtes web, il est à présent possible d'analyser le comportement de réponse des enquêtés (Bosnjak et Tuten, 2001). Il s'agit par exemple du temps passé sur chaque question, des moments d'interruption partielle ou totale du questionnaire, ainsi que de la navigation à travers les pages¹⁹.

Certains auteurs, intéressés par le comportement des internautes lorsqu'ils remplissent un questionnaire en ligne, distinguent les 'optimizers', qui portent une grande attention aux questions et fournissent des réponses très précises, des

¹⁸Par exemple, l'observatoire Sierra (programme de recherche initié par les responsables des transports en France) a mis au point une enquête automatisée sur Internet, afin de suivre les comportements des français en termes de mobilité quotidienne.

¹⁹Des conditions techniques doivent être respectées, comme la présentation d'une question par page, le téléchargement des pages sur le serveur et le calcul du temps passé sur les différentes parties du questionnaire

'satisficers', qui cherchent à répondre brièvement, en limitant leur effort (Krosnick et Alwin, 1991).

I.3 Des limites importantes

L'utilisation du web dans les protocoles d'enquête semble intéressante. Les expériences sont positivement évaluées selon leur coût, le temps de réponse, le design du questionnaire et la qualité des informations recueillies. Toutefois, comme le mode postal et téléphonique au temps de leur reconnaissance comme médias d'enquête, le web présente quelques limites méthodologiques, qui modèrent son intérêt (Sills et Song, 2002). Dans la littérature, les auteurs distinguent quatre types de biais potentiels pour les enquêtes (Dillman et Bowker, 2001; Alsnih, 2005) : le biais de couverture, le biais d'échantillonnage, le biais de non-réponse et le biais de mesure. Certains auteurs (Couper, 2001) sont allés plus loin, prenant également en considération les biais d'imputation des valeurs manquantes, technique largement utilisée dans les enquêtes auto-administrées, afin de palier à la non-réponse partielle, ou les biais de pondération des individus, puisque les enquêtes web ne sont généralement pas représentatives de l'ensemble de la population (figure 1).

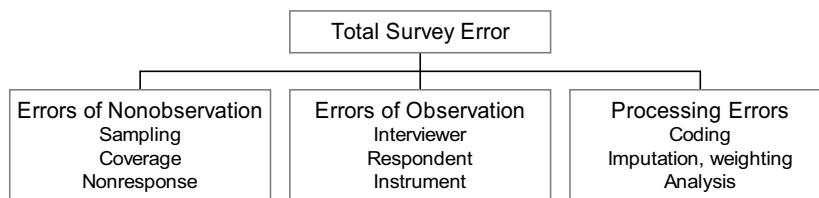


FIG. 1 – Les différentes sources de biais dans les enquêtes web

Source : M.P. Couper (2001)

I.3.1 Le biais de couverture

Afin d'effectuer une enquête de qualité, dont les résultats seront généralisables à l'ensemble de la population, il est nécessaire de sélectionner l'échantillon en effectuant un tirage aléatoire dans la population de référence, chaque individu devant avoir exactement la même chance que les autres de participer²⁰. Cette condition n'est pas toujours respectée dans les enquêtes web, puisque répondre à un questionnaire en ligne presuppose de disposer d'un ordinateur et d'une connexion Internet. Bien que l'équipement informatique des ménages français et le taux de pénétration d'Internet sur le territoire soient en forte progression²¹, tous les ménages de l'agglomération lyonnaise n'ont pas accès à internet (tableau 3), et il est impossible de leur attribuer, a priori, une probabilité connue et non nulle de participer à l'enquête.

²⁰Bien qu'il soit possible de constituer un échantillon stratifié

²¹Il est très difficile d'établir la pénétration réelle d'internet, car les données varient selon que l'on considère les individus ou les ménages, la connexion au domicile ou sur le lieu de travail, ou même la capacité individuelle à naviguer sur le web.

Zones géographiques	Population (est. 2009)	Accès Internet (27 sept 2009)	Croissance (2000-2009)	Pénétration (% pop.)
France	62 150 775	42 050 465	394,70%	67,70%
Union Européenne	803 850 858	402 380 474	282,90%	50,10%

TAB. 3 – Accès à internet en France et au sein de l’Union européenne

Source : <http://www.internetworkstats.com/stats.htm>
 (site visité le 27/09/09)

Au vu des chiffres, la proportion de ménages équipés d’une connexion web reste encore faible pour permettre le lancement d’études de mobilité à l’aide de ce seul média, si la représentativité statistique de l’ensemble de la population est visée ²². L’erreur de couverture, biais le plus important rencontré lors d’une enquête web à grande échelle, est fonction du taux de pénétration d’internet dans la population et des différences entre les internautes et les individus n’ayant pas accès au web sur les variables d’intérêt (Groves, 1989). Ce biais est présent, même pour une enquête générale auprès des internautes, puisqu’il n’existe pas de liste officielle et exhaustive des adresses e-mail des individus. Couper (2001) propose une écriture formalisée de l’erreur de couverture. Avec Y_t , Y_c et Y_{nc} les valeurs de la variable d’intérêt sur l’ensemble de la population, sur la population couverte par internet et sur la population n’ayant pas accès au web, N_t et N_{nc} le nombre d’individus de la population totale et le nombre d’individus ayant accès au web, on obtient :

$$Y_c = Y_t + \frac{N_{nc}}{N_t} (Y_c - Y_{nc}) \quad (1)$$

Plus encore, il est possible de différencier les internautes selon leur équipement multimédia (navigateur, type de connexion ...) et leur maîtrise de l’usage d’Internet. Bradley (1999) distingue ainsi 13 classes d’internautes selon l’équipement informatique et l’usage d’internet (fréquence de lecture des e-mails, nombre de messages en attente dans la boîte aux lettres, fréquence de visites de sites web, connexion internet à domicile ou sur le lieu de travail, nombre d’adresses e-mail utilisées...). Un rapport du Ressource System Group (2002) définit la capacité de navigation des internautes par le nombre de pages web visualisables sur une période de temps. Cet indicateur dépend notamment du matériel disponible (ordinateur, navigateur ...) et du serveur sur lequel les requêtes sont lancées. Même si les ménages auxquels nous nous intéressons possèdent une connexion Internet, il n’est pas établi que l’ensemble des individus qui composent ce ménage soient à l’aise avec l’utilisation du web (Gunn, 2002; Lazar et Preece, 1999). Les enquêtes web ne sont accessibles qu’aux individus familiers d’internet (Abdel-Aty, 2003). Pour les internautes peu expérimentés, la capacité de répondre aux questions à l’écran avec un clavier et une souris est limitée. Selon le degré de complexité du questionnaire, certains ne pourront pas télécharger les pages, d’autres auront des difficultés pour répondre

²²Certaines catégories de la population sont difficiles à joindre par le web (personnes âgées, peu qualifiées ...) (Roos et Marchand, 2006).

aux questions (utilisation de menu déroulant, validation des réponses, saisie des données dans des champs spécifiques, accès aux pages suivantes ...). Il est donc possible que non seulement les caractéristiques sociodémographiques des internautes diffèrent de celles de l'ensemble de la population, mais qu'elles varient également selon le type d'équipement possédé et le degré de familiarité avec le web (Alsnih, 2005).

I.3.2 Le biais d'échantillonnage

L'objectif d'une enquête est d'estimer avec précision la distribution d'une variable dans la population de référence, à partir de données obtenues sur une petite proportion d'individus. Afin de permettre l'inférence statistique des résultats, l'échantillon utilisé pour l'enquête doit être représentatif de la population étudiée. Le biais d'échantillonnage vient des différences, en termes de caractéristiques socio-économiques mais aussi sur les variables d'intérêt, entre les personnes qui constituent l'échantillon et celles de la population de référence (Couper, 2000).

Certaines méthodes de recrutement, telle que celle utilisée pour les enquêtes ménages déplacements en France, s'appuient sur une procédure d'échantillonnage de la population très pointue, alors que d'autres méthodes ne contrôlent pas l'échantillon au-delà du respect de certains quotas. Lors de la réalisation d'une enquête web à grande échelle, il est quasiment impossible d'utiliser les méthodes d'échantillonnage probabilistes classiques, puisqu'à la différence d'autres médias²³, aucune base de sondage n'est disponible (Alvarez et Van Beeslaere, 2003; Andrews *et al.*, 0003). Un certain pourcentage d'internautes changent régulièrement de fournisseur d'accès et donc d'adresse mail (Lang, 2002), d'autres conservent leur ancienne adresse, mais ne l'utilisent plus et ne vérifient pas régulièrement leurs messages (Bradley, 1999) et d'autres possèdent de multiples adresses, professionnelles et personnelles. Au final, un document répertoriant l'ensemble des adresses mail serait de faible recours pour mener des enquêtes, puisque difficilement actualisable et que les e-mail considérés comme intrusifs sont directement effacés par les internautes (Cho et LaRose, 1999). De plus, les internautes ont un profil socio-économique particulier (Solomon, 2001; Romano et Himmelmann, 2002). Ils sont plus jeunes, actifs, davantage masculins et avec un haut niveau de diplôme et de revenus (Myles et Tibert, 1998; Braunsberger *et al.*, 2007; Kwak et Radler, 2000; Bates, 2001; Deutskens *et al.*, 2006; Haraldsen *et al.*, 2002). Vehovar et Lozar Manfreda (2002) et Badagliacco (1990) précisent que la fréquence d'utilisation de l'ordinateur et d'internet, corrélée au sexe et au niveau d'éducation sont les plus importants facteurs qui influencent la participation aux enquêtes web. Ces dernières sont plus facilement acceptées chez ceux qui sont familiers avec l'ordinateur (à l'école, au travail ou à la maison), et parmi les répondants qui savent l'utiliser 'correctement' (Wright *et al.*, 1998)²⁴. Certains types de ménages, les plus modestes ou ceux qui ne disposent pas d'un accès internet au domicile ou sur

²³Fichier des abonnés à France Telecom pour les enquêtes téléphoniques ou fichier des adresses du recensement de l'INSEE pour les enquêtes postales.

²⁴La non familiarité avec l'outil génère un effet 'big brother' amenant plus de refus (Lo-

leur lieu de travail, se trouvent mécaniquement exclus d'une enquête web. Ils représentent pourtant une part non négligeable de la population et ont parfois des comportements spécifiques qu'il est délicat d'ignorer²⁵. Myles et Tibert (1998) précisent qu'une enquête web est une enquête auprès d'individus qui s'auto-sélectionnent et que toute représentativité de la population de référence est accidentelle et non le reflet d'une méthode d'échantillonnage²⁶.

Si la sélection des répondants est aléatoire, la précision des résultats est directement liée à la taille de l'échantillon²⁷. Les conducteurs d'études en ligne ont été tentés de compenser le biais de couverture et d'échantillonnage en interrogeant de larges échantillons²⁸. L'idée est qu'un large échantillon diminue les chances d'exclure systématiquement certains segments de la population. Mais un taux de réponse important, obtenu sur un échantillon non probabiliste, ne garantit pas l'obtention de données de qualité (Dillman et Bowker, 2001). Selon Fowler (1993) la pertinence des données dépend de la façon dont l'échantillon est construit, et non de sa taille. Ainsi, il ne sert à rien d'augmenter le nombre de répondants afin d'augmenter le degré de précision des résultats, si le mode de sélection des individus est biaisé. Dans certains cas, un plus petit échantillon bien sélectionné conduit à des résultats de meilleure qualité (Couper, 2001).

Lors d'une enquête web, les répondants ne sont donc généralement pas représentatifs de la population étudiée, ce qui interdit toute généralisation des résultats à l'ensemble de la population par inférence statistique (Ressource System Group, 2002). La construction d'un échantillon probabiliste nécessite alors le recours à d'autres médias pour sa construction, par exemple un recrutement par téléphone où l'enquêteur encourage le répondant à remplir le questionnaire en ligne (Alsnih, 2005).

I.3.3 Le biais de non-réponse

Même si l'échantillon est totalement aléatoire, certains individus ne veulent ou ne peuvent répondre à l'enquête, d'autres ne terminent pas le questionnaire. Le biais de non-réponse est fonction du taux de non-réponse à l'enquête et des différences entre les répondants et les non-répondants sur la variable d'intérêt (Keeter *et al.*, 2000; Groves et Couper, 1998)²⁹. Dans une enquête

zar Manfreda et Vehovar, 2002b).

²⁵Par exemple en termes d'utilisation de la voiture particulière pour les enquêtes de mobilité.

²⁶Au mieux, les chargés d'études vont comparer les données socio-économiques agrégées déclarées par les répondants web avec celle de la population de référence, pour vérifier que l'échantillon web peut être considéré comme le reflet possible de la population (Lang, 2002).

²⁷L'erreur d'échantillonnage diminue de moitié, lorsque la taille de l'échantillon est quadruplée.

²⁸Cependant, les déterminants de la taille de l'échantillon sont plus souvent d'ordre budgétaire (consommation d'argent, de temps, d'espace et d'énergie) que statistique (Roscoe, 1975; Alreck et Settle, 1995).

²⁹Il n'est pas démontré que le taux de réponse et le biais de non-réponse soient corrélés, des taux de réponse hauts et bas pouvant amener à un biais de non-réponse (Baines *et al.*,

web, certains groupes sociodémographiques sont sur ou sous-représentés dans l'échantillon (Witte *et al.*, 2000). Par exemple, les hommes jeunes refusent fréquemment de répondre aux enquêtes traditionnelles, alors que les enquêtes web touchent majoritairement des individus jeunes, masculins, citadins, diplômés et ayant des revenus élevés (Alsnih, 2004; Conn, 2005; Murakami et Morris, 2003). Alsnih (2005) précise que les familles nombreuses et les personnes au statut socio-économique élevé qui échappent aux modes traditionnels répondent plus volontiers sur le web. L'analyse de la littérature reste assez évasive sur les facteurs pouvant diminuer le taux de réponse des enquêtes web. Il est certain que le sujet de l'enquête et les caractéristiques sociodémographiques des individus sélectionnés pour y répondre vont largement impacter le taux de réponse (Sills et Song, 2002). Des problèmes techniques, tels que la basse vitesse du modem, la lenteur du navigateur, les interruptions survenant en cours de remplissage du questionnaire ou des temps de chargement parfois longs amènent le répondant à abandonner la saisie du questionnaire, sans que les raisons précises de ces comportements soient connues de l'administrateur (Dillman et Bowker, 2001). Il faut également prendre en compte des facteurs plus personnels, comme la familiarité des individus avec le média internet et leur attitude face à la sécurité et à la confidentialité des données délivrées (Couper, 2000).

Solomon (2001) suggère deux moments cruciaux, où les internautes risquent d'abandonner la saisie des réponses : à l'apparition d'une grille complexe de questions - réponses, ou lorsqu'on leur demande de saisir leur adresse e-mail ou des données trop personnelles. Dans le cas spécifique des enquêtes de mobilité, les abandons risquent de survenir lors du recueil des déplacements, fastidieux pour les hyper-mobiles, et lors du recueil du revenu, question toujours sensible dans les enquêtes françaises. Il n'est pas évident d'établir que ces comportements de non-réponse, totale ou partielle, ne soient pas liés aux pratiques de mobilité (Bonnel, 2004) et tout non-répondant à une enquête est en effet une source possible de biais qui doit être étudiée (Couper *et al.*, 1999). Ce biais reste toutefois difficile à estimer, dans la mesure où on ne peut identifier la population totale d'internautes, ni connaître les caractéristiques ou le comportement de mobilité de ceux qui restent 'hermétiques' à l'enquête³⁰. Il a par ailleurs peu de sens si la méthode d'échantillonnage retenue est non probabiliste.

I.3.4 Le biais de mesure

L'erreur de mesure correspond à une déviation des réponses par rapport à leur vraie grandeur (Couper, 2000). En cas de faible motivation à répondre au questionnaire ou de mauvaise compréhension des questions, les interviewés peuvent aisément omettre certaines réponses ou 'cliquer' sans réellement réfléchir à la signification d'une question un peu vague. Des répondants 'professionnels' peuvent également biaiser les réponses. Il s'agit d'individus qui répondent systématiquement aux enquêtes, de façon très rapide, sans réellement lire les

2007).

³⁰(Kalfs et Van Evert, 2003) conclu que les enquêtes sur les non-répondants sont utiles pour la recherche, mais trop coûteuses pour être mises en oeuvre à grande échelle.

questions ni penser à leurs réponses (Nielsen, 2005).

La deuxième source du biais de mesure, commune aux enquêtes auto-administrées, vient de l'absence de la relation enquêteur-enquêté, qui ne permet pas au répondant d'obtenir des compléments d'information sur le sens des questions ou les termes employés. Elle empêche également toute analyse du comportement non-verbal du répondant, souvent riche d'enseignements pour l'enquêteur et l'analyste (soupirs en cas de lassitude, expressions du visage marquant l'incompréhension...). Les conséquences sur la pertinence des données recueillies ne sont pas neutres.

Enfin, de par sa construction, le questionnaire peut conduire à un biais de mesure, les stimuli utilisés étant susceptibles de modifier le comportement des répondants. Le design du questionnaire (placement des questions, typographie, couleur, échelles de réponse...) peut motiver l'enquêté à saisir ses réponses, ou au contraire le décourager dans sa démarche (Couper *et al.*, 1999; Dillman *et al.*, 1998b). L'apparence du questionnaire en ligne varie parfois fortement selon le matériel informatique et certains dysfonctionnements techniques, non anticipés, peuvent également causer des erreurs dans la collecte des données par internet (temps de chargement des pages, non disponibilité du serveur, problème d'affichage ...) (Dillman et Bowker, 2001; Alsnih, 2004).

I.3.5 Les diverses sources d'incertitudes

Trois autres contraintes doivent être considérées par l'administrateur d'une enquête web : le coût, les problèmes techniques et le contrôle du répondant. Le temps nécessaire pour télécharger les pages web a en effet un coût pour l'enquêté, qu'il s'agisse du temps passé devant l'écran ou du montant de la communication, pour ceux qui ne bénéficient pas de forfaits illimités³¹. En règle générale, si les capacités des individus face à l'ordinateur diffèrent substantiellement, il existe une durée limite du questionnaire à ne pas dépasser pour éviter les abandons. En ce qui concerne le commanditaire, la mise en place d'une enquête web génère des coûts fixes parfois importants. Le développement du questionnaire nécessite du temps et des compétences spécifiques, sa gestion en ligne occasionne des coûts de maintenance du réseau et d'assistance des enquêtés (Schonlau *et al.*, 2001; Watt, 1999).

Les problèmes techniques constituent une autre limite des enquêtes web (Truell, 2003). Les coupures de réseau peuvent interrompre la saisie des réponses sans que le gestionnaire de l'enquête n'en soit informé, et conduire à la perte définitive des répondants. La présentation du questionnaire peut varier selon les paramètres de l'ordinateur, et apparaître fort désagréable à certains internautes, qui vont quitter le questionnaire avant d'avoir terminé la saisie des réponses.

Enfin, dans une enquête web, il est difficile de contrôler 'celui' qui se cache derrière l'ordinateur. Cette limite est propre au caractère auto-administré du questionnaire. Aucun enquêteur ne valide l'identité de la personne interro-

³¹Les enquêtes web longues à télécharger augmentent le coût pour l'enquêté (Swoboda *et al.*, 1997).

gée, qui peut facilement mentir sur ses caractéristiques sociodémographiques principales, telles que le sexe et l'âge (Nielsen, 2005) ou reporter des données relatives à un autre membre du ménage (Lang, 2002). Le répondant contrôle le temps et la façon de remplir le questionnaire. Il peut être endormi, en colère, intoxiqué, contrarié ou dans un autre état d'esprit l'empêchant de déclarer des réponses justes et honnêtes. Stanton (1998) montre que cette variation de motivation génère des réponses manquantes et biaise les résultats de l'enquête. Les internautes peuvent également 'zapper' entre différents programmes et avoir plusieurs pages web ouvertes, ce qui génère des processus cognitifs plus superficiels (Alvey, 2004). Le contrôle des réponses et du profil du répondant est donc incertain et complexifie l'analyse des données. Si la connexion n'est pas sécurisée par un mot de passe, l'administrateur s'expose par ailleurs à une duplication possible des questionnaires, si plusieurs personnes répondent successivement.

I.3.6 La sécurité des données

Les études sérieuses garantissent la confidentialité des réponses (seuls les analystes ont accès aux données), mais l'anonymat n'est pas toujours respecté (les analystes peuvent identifier les répondants). En France, la CNIL impose des règles très strictes sur la collecte des données et leur utilisation. Des techniques sont mises au point, afin de conserver l'anonymat des répondants. Par exemple, les adresses exactes ne figurent généralement pas dans les enquêtes de mobilité, et la codification géographique se fait à un niveau suffisamment élevé pour que les répondants ne soient pas identifiables. Mais les modèles sophistiqués actuellement utilisés pour prévoir la demande de transport nécessitent des données désagrégées, et les enquêtés déclarent des informations de plus en plus personnelles, comme l'adresse de leur lieu de travail (Murakami et Morris, 2003). Bien que des auteurs aient montré que certains internautes valorisaient la personnalisation du contact, l'intrusion de l'informatique dans la vie privée n'est pas toujours rassurante (Yun et Trumbo, 2000; Conn, 2005). Chaque ordinateur laisse des traces sur les sites visités, puisqu'il est possible d'enregistrer l'adresse IP de la machine utilisée par le répondant à l'aide d'un programme informatique adapté (Berry, 2004). Les données sont également susceptibles d'être interceptées pendant leur transfert de l'ordinateur du répondant vers le serveur dédié à l'enquête et Cho et LaRose (1999) ont montré que les chercheurs commettent davantage de violations de la vie privée des répondants sur le web qu'avec les modes de recueil plus traditionnel. Il semble donc préférable de garantir la confidentialité des données en recourant à des serveurs sécurisés³². Les réponses anonymes semblent préférables pour obtenir la confiance des répondants, mais leurs croyances concernant l'anonymat sur internet, avec ou sans accès contrôlé, ne sont pas connues (Stanton, 1998).

³²Les sites web sécurisés ont une adresse qui commence par 'https' (et non par 'http', comme la plupart des sites). Si une lettre est utilisée pour avertir les répondants que le questionnaire peut être rempli en ligne, alors ces derniers sont susceptibles de ne pas saisir correctement l'adresse du site sécurisé. Lors d'une expérience menée en laboratoire, Berry (2004) a montré que ce type d'erreur concerne 13% des individus.

L'utilisation de codes d'accès et de mots de passe permet de restreindre l'accès aux données, mais peut augmenter la suspicion des répondants. Enfin, il peut être utile de mettre en place des pare-feux pour garantir la sécurité des fichiers. Certains internautes refusent en effet de se connecter sur des sites inconnus ou de répondre à un questionnaire en ligne,

I.4 Un avenir prometteur

Les modes d'enquêtes traditionnels tels que le téléphone, le postal ou le face-à-face sont plus populaires que le web (Huang et Liaw, 2005). Si les études en ligne sont encore modestes, puisqu'elles ne représentent que 10% du chiffre d'affaires total du secteur des études, la diffusion des technologies de l'information et de la communication dans les entreprises, le monde universitaire et les particuliers, font du web un outil prometteur pour les commanditaires d'études. Les expériences montrent que le web peut remplacer avantageusement un certain nombre de modes traditionnels, en arrivant à des résultats identiques (Roos et Marchand, 2006; Pitkow et Recker, 1995; Stanton, 1998). En 2005, AC Nielsen a mené outre-Atlantique une enquête d'Usages et Attitudes sur deux échantillons, web et téléphone, qui a abouti à des résultats comparables³³. Abdelwahab et Abdel-Aty (2003) ont comparé la qualité et la précision des données provenant d'une étude postale, avec celles d'une étude web, et a conclu à la supériorité de ces dernières³⁴. Selon Dillman et Bowker (2001), il existe peu de littérature sur la variation de la qualité des réponses d'un individu, selon qu'il réponde sur le web ou par un mode plus traditionnel (postal, face-à-face). Plusieurs auteurs (Benfield et Szlemko, 2006; Schillewaert et Meulemeester, 2005) ont montré que le mode d'enquête, qu'il repose sur une version papier ou électronique, ne modifie pas en théorie la qualité des données fournies par le répondant. Les compétences informatiques des individus, associées à la connaissance du sujet d'étude et à leur aisance relative au remplissage d'un questionnaire peuvent cependant altérer leur capacité à répondre en ligne et donc impacter la qualité des données recueillies.

Le web est un média restreint par les capacités informatiques des répondants (Conrad *et al.*, 2003). Huang et Liaw (2005) ont mené une expérience auprès d'étudiants, pour analyser les perceptions des répondants web. Ils postulent que les déterminants de l'intention de répondre à une enquête en ligne sont : la capacité des individus à saisir leurs réponses, leur appréhension concernant l'utilisation future des données, leur motivation intrinsèque pour répondre en ligne et la pertinence du web pour l'étude. Les résultats montrent qu'une fois que les enquêtés estiment qu'ils sont en mesure de remplir le questionnaire web, le nombre de questions, la vitesse de la connexion et le design de l'interface sont les principaux facteurs qui déterminent la saisie des réponses (figure 2). L'anxiété face à l'utilisation de l'ordinateur impacte en revanche négativement la propension à répondre en ligne et la performance individuelle dans la saisie

³³Le web a généré des réponses de qualité, même en ce qui concerne les questions ouvertes.

³⁴Pourtant, de nombreuses études rapportent un taux de réponse par voie postale supérieur au taux de réponse en ligne (Schaefer et Dillman, 1998; Morris et Adler, 2003).

Le potentiel du web pour les enquêtes ménages déplacements

des réponses. Il est difficile aux commanditaires d'étudier de contrôler certains déterminants tels que les compétences des répondants, la vitesse de connexion des internautes ou l'accessibilité des ménages à une connexion internet. En revanche, une réflexion sur la prise de contact et la construction du questionnaire peut influencer positivement l'intention des enquêtés de répondre en ligne.

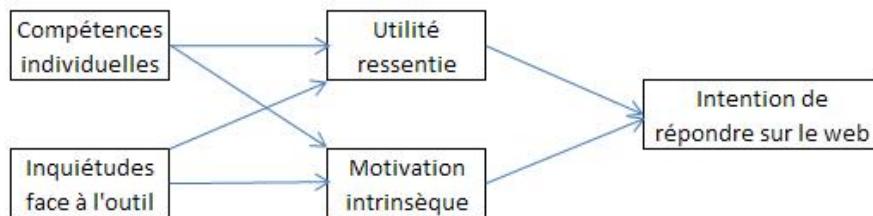


FIG. 2 – Modèle conceptuel des déterminants de l'intention de répondre en ligne

Source : Huang and Liaw (2005)

Si le coût et la rapidité sont les deux avantages les plus souvent cités par les partisans des enquêtes web (Witt, 1998), ce nouveau mode de collecte présente d'autres atouts intéressants, comme la souplesse du questionnaire et une réactivité inégalée des gestionnaires d'études³⁵. Les problèmes de l'utilisation du web sont de trois sortes : échantillonnage, consistance des réponses et la motivation des participants (tableau 4). Afin de réduire ces biais, une attention particulière doit être portée à la conception des questionnaires web, les développeurs étant davantage des techniciens du web que des professionnels des méthodes d'enquête. L'enquête idéale, support au développement d'un modèle de prévision, permettrait sans doute de couvrir l'ensemble de la population, pour un rapport coût-qualité intéressant, et prévoirait des stratégies pour limiter les non-réponses et assurer la cohérence des données recueillies (Morris et Adler, 2003). Mais il est difficile de réunir l'ensemble de ces conditions à l'aide d'un seul mode d'enquête.

Avantages (+)	Limites (-)
Fort potentiel à faible coût	Visualisation différente selon l'équipement
Réponse plus rapide	Sensible au niveau d'expertise informatique
Envoi de rappels	Sécurité des données sur le serveur
Vérification dynamique des incohérences	Confidentialité des réponses
Questions aléatoires	Comportement des Internautes
Personnalisation	Méthodes de sélection de l'échantillon
Dispersion géographique	Pas de contrôle de l'identité du répondant
Affichage progressif, interactivité	Biais de compréhension des questions

TAB. 4 – Avantages et limites du média web pour la réalisation d'enquêtes

Source : Gunn (2002)

³⁵En comparaison avec les enquêtes auto-administrées sur papier, les questionnaires web sont plus conviviaux, mieux structurés, plus personnalisés, moins chers, et permettent de limiter les non-réponses partielles et les erreurs (Nowack, 1997; Weible et Wallace, 1998).

II Vers une méthodologie d'enquête mixte

Face à la résistance de certains segments de la population à participer aux enquêtes, aux réponses incomplètes ou peu fiables et aux difficultés croissantes pour établir le premier contact avec un enquêteur, la combinaison de méthodologie différente au cours d'une même enquête semble être une solution intéressante. Il s'agit de combiner le meilleur des modes, l'augmentation du taux de couverture permettant d'améliorer la qualité des données, sans grever le coût de l'étude (Dillman et Tarnai, 1988; Murakami et Morris, 2003; Wermuth *et al.*, 2001). La question de la mixité des modes d'enquête n'est pas récente³⁶, mais elle mérite d'être reconsidérée au regard des récents développements technologiques qui impactent le domaine des études (Stopher et Jones, 2003). Par exemple, la facilité avec laquelle le script d'un questionnaire peut être transféré d'un média à l'autre encourage l'utilisation de différents modes au cours d'une même enquête (Dillman *et al.*, 2001). Si les technologies de l'information et de la communication actuellement disponibles (internet, téléphones mobiles, GPS, PDA ...) offrent la possibilité d'augmenter la qualité des données d'enquête et de réduire la lassitude des répondants, elles ne peuvent cependant pas se substituer totalement aux modes plus traditionnels, tels que le face-à-face, le postal ou le téléphone (Murakami et Morris, 2003).

II.1 Une diversité de protocoles d'enquête

Au cours d'une étude, il y a trois phases de contact possibles avec les répondants : la phase de recrutement, la phase d'interview et la phase de suivi. Dans les enquêtes transport, il arrive fréquemment que des médias différents soient employés pour chacune de ces phases, par exemple le postal pour le recrutement et l'information des répondants et le web pour l'enquête. L'utilisation de plusieurs modes pour le recueil de données suscite en revanche plus d'interrogations. Cette mixité peut prendre différentes formes, suivant l'objectif poursuivi : réduire la lassitude des enquêtés, en leur laissant le choix du mode de réponse, ou faire face à une diminution du taux de réponse, en proposant un mode alternatif aux non-répondants. Ces deux protocoles sont schématisés dans la figure 3.

Chaque mode peut être appliqué à un segment particulier de la population cible, un mode pouvant être optimal pour un segment mais pas pour un autre. Par exemple, dans l'enquête nationale transport allemande, l'enquête est conduite par CATI, pour la partie de l'échantillon pour laquelle il est possible d'obtenir un numéro de téléphone, et par voie postale pour le reste (Bonnel et Armoogum, 2005). Cette pratique n'est toutefois pas sans risque sur la qualité des données, puisque les effets de modes peuvent créer des différences dans les

³⁶Les enquêtes postales et face-à-face sont les plus anciennes, et leur combinaison constitue les premières formes d'enquêtes mixtes (De Leeuw, 2005). Pour réduire le biais de couverture au démarrage du téléphone, une enquête mixte permettait de bénéficier des coûts attractifs des enquêtes par téléphone, tout en maintenant une bonne couverture de la population étudiée (Groves et Lepkowski, 1985). En recherche marketing, les hybrides du téléphone et du web sont devenus très populaires (Oosterveld et Willem, 2003) et (De Leeuw, 2005).

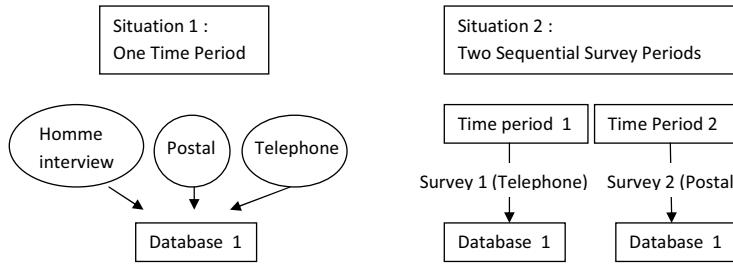


FIG. 3 – Deux situations de mixité du recueil de données

Source : J. Morris and T. Adler (2003)

réponses des groupes. Il peut également y avoir des vagues séquentielles d'enquêtes sur la même population, à l'aide de modes différents (Morris et Adler, 2003; Dillman et Christian, 2005). Dans l'enquête de mobilité Belge MOBEL, les questionnaires sont envoyés par courrier et des contacts téléphoniques sont réalisés auprès des ménages répertoriés dans les annuaires pour les motiver à participer (Hubert et Toint, 2003). L'intérêt du protocole d'enquête séquentiel mixte est que ceux qui répondent au second mode sont différents de ceux qui répondent au premier. L'espérance est que cette seconde population rassemble les non-répondants restants (Baines *et al.*, 2007). Le choix peut également être laissé à chaque enquêté du média utilisé pour répondre à l'enquête. C'est le cas dans le protocole mis en place par Keesling (2002), où une mention accompagnait le questionnaire envoyé par voie postale, précisant qu'il était possible de remplir le même questionnaire en se connectant sur le site web dédié au projet. Cette méthode est souvent utilisée pour réduire l'erreur de couverture, mais génère d'autres problèmes comme l'auto-sélection des individus et l'erreur de mesure due aux modes utilisés (Lozar Manfreda *et al.*, 2001). Enfin, lors d'enquêtes longitudinales, des répondants sont interrogés à plusieurs reprises à l'aide de différents médias.

Dans notre étude, le processus est séquentiel, puisque l'enquête web n'est proposée qu'aux individus qu'il n'aura pas été possible de joindre durant la première vague d'interviews en face-à-face.

II.2 Pour diminuer le taux de non-réponse

Face à la réticence de plus en plus vive de la population à répondre aux enquêteurs, il devient difficile de réaliser une enquête de mobilité en face-à-face de bonne qualité, les non-répondants ayant probablement des comportements de déplacements différents des ménages qui acceptent de se soumettre à l'interview. De plus en plus d'enquêtes reposent sur des protocoles complexes, associant plusieurs modes ou méthodologies, pour augmenter le taux de réponse global, accroître le taux de couverture de la population cible ou améliorer la qualité des données produites (Couper, 2000; Gunn, 2002; Dillman *et al.*, 2001; Alsnih, 2004; Groves et Kahn, 1979). Qu'il s'agisse de la phase de recrutement ou d'interview, multiplier les modes de contact est un moyen sûr de toucher davantage d'unités statistiques (Goyder, 1987). Non seulement les individus

apprécient de disposer de plusieurs moyens pour communiquer sur leurs habitudes³⁷, mais tous les modes ne permettent pas de toucher les mêmes segments de la population (les 'hypers mobiles' par exemple sont rarement présents à leur domicile aux créneaux horaires pendant lesquels les enquêteurs prennent les rendez-vous et viennent administrer le questionnaire. Les répondants à un média ne sont donc pas forcément les mêmes que les répondants à un autre média (Bonnel, 2003; Bayart *et al.*, 2009; Shettle et Mooney, 1999).

L'utilisation d'internet dans un processus séquentiel apparaît comme un moyen efficace de capter des répondants qui échappent aux modes traditionnels, et de réduire le biais de non-réponse si les répondants web sont représentatifs de l'échantillon restant qui n'a pas répondu au premier mode (Baines *et al.*, 2007).

II.2.1 En limitant le niveau d'effort requis

Avant de participer à une enquête, l'individu va évaluer le niveau d'effort requis. Il va notamment s'interroger sur le temps pris par l'interview, l'importance du sujet, les difficultés physiques, intellectuelles ou émotionnelles ressenties pour répondre au questionnaire, le bien être procuré par ses réponses, sa volonté de livrer des informations personnelles et les éventuelles pressions externes qu'il peut subir (Ampt, 1997). A cela peut s'ajouter la nature de l'étude, les renseignements factuels (données socio-économiques) étant plus facilement fournis que les réponses de nature comportementale ou attitudinale, qui réfèrent davantage à l'émotionnel (Richardson *et al.*, 1995). L'évaluation de ces différents paramètres va influencer les comportements individuels de réponse / non-réponse. Mais l'enquêté va également prendre en compte ses expériences passées, allant des sollicitations aux participations effectives. Pour améliorer le taux de réponse, il est donc nécessaire de considérer l'ensemble des éléments qui affectent la perception de difficulté de l'enquête par le répondant, et donc son processus de choix d'y participer ou non. Cette étape est d'autant plus délicate que le niveau de difficulté ressenti dépend des individus et du moment du contact.

La mise en place d'une enquête sur Internet va jouer essentiellement sur les dimensions temporelles, physiques, intellectuelles et interpersonnelles de l'étude. Le temps dont disposent les individus pour répondre au questionnaire est en effet limité et fortement contraint par le type d'activités dans lesquelles ils sont impliqués (travail, école, loisirs, occupations diverses au domicile ...). De manière générale, les questionnaires auto-administrés sont plus flexibles, puisque l'individu choisit lui-même le moment le plus approprié pour y répondre. Les entretiens en face-à-face essaient de lever cette contrainte temporelle, en laissant choisir à l'interrogé le créneau horaire du rendez-vous, mais les possibilités sont limitées, et dégager une plage d'une heure trente dans un emploi du temps surchargé n'est pas toujours une tâche facile.

³⁷Une étude sur les non-répondants menée en Australie par l'Institute of Transport Studies montre que les personnes préfèrent avoir le choix du mode de réponse et que cette option impacte positivement le taux de réponse et la qualité des données.

II.2.2 En conservant la pertinence des réponses obtenues

Si l'interactivité du web permet de relancer le répondant en cas de non-réponse à une ou plusieurs question, des effets en termes de comportements de réponse ont également été remarqués. Nous considérerons plus particulièrement dans cette section les effets de modes qui pèsent sur la qualité des données (Gunn, 2002). Dans le cas de questionnaires administrés, la présence de l'enquêteur génère un biais, puisque le répondant risque de positionner son discours en fonction de l'impact positif qu'il espère produire sur son interlocuteur ('social compliance') (De Leeuw, 1992; Schuman et Presser, 1981). Bien que ce phénomène s'observe surtout lors de question d'opinion ou d'attitude, il est possible de le remarquer également dans les enquêtes de déplacements, le domaine des transports étant par nature très sensible³⁸. Ce biais est fortement réduit dans le cas d'une enquête web, puisque l'interaction entre l'enquêteur et l'enquêté est virtuelle.

Par ailleurs, lorsqu'une liste de modalités est lue par l'interviewer, le répondant a généralement tendance à sélectionner le dernier choix ('recency effect'), car il n'a pas le temps de fixer chaque choix dans une mémoire de long terme avant que le prochain ne soit donné. Dans les modes auto-administrés, le processus cognitif est plus profond, et les premiers items sont davantage choisis (Krosnick et Alwin, 1987; Schuman et Presser, 1981). Cet effet survient surtout dans les modes oraux tels que le téléphone ou le face-à-face, mais est peu présent dans les enquêtes web. Enfin, ce mode de recueil assisté par ordinateur permet une vérification instantanée des réponses fournies (Schonlau *et al.*, 2001).

Si l'utilisation exclusive du web dans les enquêtes déplacements semble pré-maturée, il est en revanche possible de combiner efficacement ce média avec un mode de recueil plus traditionnel, au risque de modifier les mécanismes cognitifs sollicités chez les répondants.

II.3 Le web : un mode complémentaire du face-à-face ?

Quatre grandes générations de modes d'enquête se sont succédées, s'adaptant au fil des années aux évolutions technologiques : le face-à-face, le postal, le téléphone et l'internet. Chacun offre des avantages incontestables, mais également des limites non négligeables, dont le commanditaire doit tenir compte avant de faire son choix. De nombreux pays tels que le Canada, la Suisse et les USA utilisent depuis plusieurs années le téléphone³⁹, moins coûteux, pour réaliser leurs enquêtes de mobilité (Richardson *et al.*, 1995). En France, le mode dominant reste le face-à-face. Il semble opportun de s'interroger à présent sur le potentiel du web, comme complément aux enquêtes traditionnelles.

³⁸Par exemple, l'enquête étant financée par le Sytral, les individus peuvent avoir tendance à sur valoriser les transports en commun.

³⁹Aux USA, le mode postal est également utilisé.

II.3.1 Deux modes aux atouts différents

Le face-à-face est un mode d'enquête relativement flexible, puisque les enquêteurs fixent à l'avance des rendez-vous avec les répondants, en fonction de leurs disponibilités. Il s'agit d'un mode adapté aux enquêtes lourdes et complexes. L'enquêteur facilite la tâche du répondant, montre certains éléments (utilisation possible de visuels et de sons dans les protocoles d'enquête), et relance éventuellement l'entretien. Il peut s'assurer que les répondants comprennent les questions et que toutes sont identiquement interprétées, (Conrad et Schober, 2000; Dillman et Christian, 2005), sa voix et ses gestes agissant comme un paralangage qui donne des informations au répondant. La couverture de l'échantillon peut être totale, mais le coût horaire des enquêteurs augmente le budget de l'étude et limite souvent son étendue géographique. D'un autre côté, l'influence de l'enquêteur, propre aux enquêtes administrées, peut venir biaiser les réponses. Le face-à-face est également qualifié d'intrusif, puisque l'enquêteur vient au domicile des personnes interrogées. Ces dernières ne sont pas toujours disponibles aux créneaux proposés, selon leur activité professionnelle, ni intéressées par le sujet de l'étude, et réunir l'ensemble des personnes du ménage pour recevoir un enquêteur est souvent un exercice difficile. Enfin, des problèmes de sécurité limitent la réalisation d'enquêtes à domicile dans certaines zones urbaines⁴⁰.

Les avantages et les limites du web comme média d'enquête ont été détaillés dans les sections précédentes. Concernant le mode d'administration du questionnaire, les enquêtes web sont à mi-chemin entre le mode administré et le mode auto administré. L'environnement s'apparente aux enquêtes postales et le questionnement met en scène un enquêteur virtuel (Haraldsen *et al.*, 2002). Le web et le face-à-face sont donc des modes d'enquête aux caractéristiques très diverses (média utilisé, forme et longueur du questionnaire ...), qui impactent directement leurs performances (tableau 5).

Critères	Mode d'enquête	
	face-à-face	web
Couverture	++++	+
Taux de réponse	++++	+++
Qualité des données	++++	++
Langage	++++	++
Complexité des questions	+++	+++
Coûts	+	++++
Qualité du contrôle	++	++++

TAB. 5 – Comparaison des enquêtes en face à face et par web

Source : A partir de Morris et Adler (2003), adapté de Ettema *et al.* (1996)

⁴⁰Les enquêtes ménages déplacements en face-à-face ont été remplacées par des enquêtes téléphone aux USA, car trop chères et trop dangereuses à mener dans certaines aires urbaines (Stopher et Metcalf, 1996).

II.3.2 Intérêt de la mixité du web et du face-à-face

Nous avons par exemple montré dans la section I qu'il n'est pas recommandé de réaliser une enquête web exclusive auprès de l'ensemble de la population, dans le but de recueillir des données représentatives (Bonnel, 2004). En effet, les nouvelles technologies ne sont pas à la disposition de l'ensemble des habitants, et employer seulement le web élimine la possibilité pour certaines personnes d'être sélectionnées (Dillman *et al.*, 1999). Par ailleurs, le problème des spams et les mesures préventives utilisées pour y mettre fin rendent difficile le contact d'un répondant sur le web (Rose *et al.*, 2004).

Dillman (2000) et Dillman *et al.* (1995) montrent en revanche que l'utilisation du web dans un protocole d'enquête est un moyen efficace d'augmenter le taux de réponse global, et concluent que les enquêtes web peuvent être utilisées de façon complémentaire aux méthodes traditionnelles. Il est donc intéressant de réfléchir sur la possible intégration du web dans les enquêtes de mobilité et d'explorer plus largement la combinaison de plusieurs modes de collecte de données, telles que le web et le face à face (Couper, 2000; Gunn, 2002; Comley, 1996). Introduire le web dans un protocole d'enquête contribue à la diffusion de ce média, et à la conversion des ménages autrefois fidèles aux modes traditionnels tels que le face-à-face (Cobanoglu *et al.*, 2001). En proposant le web en complément du face-à-face, la population générale est mieux cernée, puisque certains individus se reconnaissent davantage dans l'utilisation du web et sont par conséquent plus disposés à livrer des informations⁴¹.

Dans les enquêtes ménages déplacements, il est demandé aux répondants de se souvenir des déplacements effectués les jours précédents la passation du questionnaire. Cet important effort de mémoire ne permet pas toujours de collecter de manière exhaustive les petits déplacements ou de restituer les chaînes de déplacements. Si des interviews en face-à-face permettent de recueillir des informations très précises grâce aux relances et à la vigilance de l'enquêteur, leur coût ne rend pas possible un fréquent renouvellement de l'enquête. A contrario, les enquêtes web souffrent des biais propres aux enquêtes administrées, une défaillance de la mémoire du répondant pouvant fausser les réponses, mais sont peu coûteuses. Une méthodologie faisant appel au web et du face-à-face permet d'exploiter les atouts des deux modes. Il est donc envisageable de continuer à mener des enquêtes ménages déplacements en face-à-face pour recueillir des informations précises sur l'évolution des comportements de mobilité des ménages, et de compléter ces données avec une enquête web afin d'améliorer le taux de réponse et de toucher des individus qui ne répondent pas en face-à-face.

⁴¹Les technologies de l'information et de la communication se diffusent progressivement dans la population. Les 'adopteurs précoce' ont un profil particulier, souvent sous-représenté dans les enquêtes traditionnelles. L'utilisation du web est donc un moyen de réduire certains biais d'échantillonnage (Cf. section I.3.3).

II.3.3 Des problèmes de comparabilité

Proposer plusieurs modes de recueil de données n'est pas sans risque sur la validité de l'enquête, des variations dans les questionnaires utilisés pouvant influencer les réponses (Christensen, 2004). Il est donc préférable de construire un questionnaire simple, qui sera adaptable facilement sur le web sans nécessiter de changements majeurs, et induira peu de différence sur les réponses (Cobanoglu *et al.*, 2001). A l'inverse, un questionnaire web contenant de nombreux filtres risque d'aboutir à une version papier très complexe et difficilement administrable (Alsnih, 2004). Il existe des différences culturelles sous-jacentes au choix du mode d'enquête (Morris et Adler, 2003; Jeavons, 1999), et les ménages choisissant de répondre par Internet peuvent avoir des comportements de mobilité différents de ceux qui répondent en face-à-face. Selon une étude de Ressource System Group (2002), les répondants web sont plutôt jeunes, perçoivent des revenus élevés et ont une mobilité supérieure aux répondants par voie postale ou par téléphone (entre 7.4 et 10.1 déplacements quotidiens par ménage, ce qui représente une mobilité plus élevée de 18% à 45% par rapport aux autres modes.). Les ménages qui répondent sur internet sont de plus grande taille et ont un statut socio-économique plus élevé que les répondants par voie postale ou par téléphone (Arentze *et al.*, 2004), ce qui impacte positivement leur niveau de mobilité.

Dans le cas d'une enquête en face-à-face, la communication est essentiellement auditive, alors que l'utilisation du web favorise le langage visuel, sous contrôle du répondant (Dillman, 2000) (textes, graphiques et informations numériques (Couper *et al.*, 2001). La taille et le type des lettres, mais aussi leur couleur, les animations et les symboles sont autant de détails de présentation qui ajoutent de la compréhension au texte, mais qui peuvent éloigner du sens des questions. Ces différents signaux, qui envoient des messages conflictuels aux répondants, affectent le taux de réponse et la qualité des informations saisies (Dillman *et al.*, 1999). L'effet est particulièrement important dans les questionnaires web, où le potentiel du multimédia autorise la combinaison de plusieurs langages. Ainsi, la combinaison d'une enquête web et d'une enquête en face-à-face pose des problèmes de comparabilité, le support de communication entre l'enquêteur et l'enquêté étant très différent (Dillman *et al.*, 2001). Les effets de modes varient selon le type de questions (faibles pour les questions socio-démographiques et élevés pour les questions d'attitude et d'opinion), puisque certaines appellent des réponses préformatées, alors que d'autres amènent le répondant à écouter la question, l'échelle de réponse et à réfléchir (Dillman et Christian, 2005).

Dans les protocoles d'enquête mixte, la question de l'intégration des données se pose : les données collectées avec des modes différents peuvent-elles être combinées et comparées ? La structure des questions change quelquefois, et le temps des enquêtes n'est pas fixe. Après une sérieuse réflexion, le chercheur doit décider si les effets de modes prévus sont assez sérieux pour ne pas recourir à la mixité des modes d'enquête ou si les avantages d'une telle démarche nécessitent d'en prendre le risque. Le but est d'optimiser la procédure de recueil de données et de réduire les biais pour un temps et un budget donné. La

mixité des modes donne une opportunité de compenser la faiblesse de chaque mode à un coût acceptable. Il y a donc moins de coût et d'erreur qu'avec un seul mode. Le trade-off se fait entre les coûts et les erreurs, particulièrement l'erreur d'échantillonnage (De Leeuw, 2005), afin de sélectionner le meilleur protocole d'enquête.

III Les caractéristiques des internautes français

Internet est l'adoption technologique la plus rapide de l'histoire, et l'utilisation de ce nouveau média s'inscrit pleinement dans la logique de l'évolution des modes de communication des individus. Internet est en effet devenu progressivement un outil d'expression, de consommation et de divertissement (commerce en ligne, 'chat' ...). En 2006, environ six adultes sur six possédaient un ordinateur à domicile (57%), et 43% étaient équipés d'une connexion internet chez eux (Bigot et Croutte, 2007) ⁴². Ces taux sont passés respectivement à 64% et 53% en 2007, les 10 points de gagnés en taux de pénétration d'internet représentant plus de 4 millions de personnes équipées (annexe I). Si les internautes ont un profil de plus en plus varié, il est encore légitime de supposer que certaines catégories de la population, davantage concernées que d'autres par la 'fracture numérique', n'ont pas un accès direct à ce nouveau moyen de communication ⁴³. Afin de mieux comprendre quels segments de la population sont potentiellement les plus concernés par les enquêtes web, il est nécessaire de considérer trois composantes : la possession d'un ordinateur, la capacité à se connecter à internet et le niveau de familiarité des individus avec le web.

III.1 L'ordinateur, un bien de consommation largement diffusé

Dans cette section, les données concernent l'année de réalisation de l'enquête ménages déplacements sur le web, c'est-à-dire 2006. La progression des indicateurs observée depuis confortent nos hypothèses sur la légitimité des enquêtes web.

Près des deux tiers de la population était équipée d'un ordinateur à domicile (60%) en 2006, et une personne interrogée sur 10 déclarait en posséder plusieurs (Bigot et Croutte, 2007). Le rythme de progression, faible durant des années, s'est nettement accéléré depuis 1997, grâce notamment à l'attrait d'internet (figure 4) ⁴⁴.

⁴²Enquête réalisée en juin 2007, auprès de 2015 personnes de 18 ans et plus et 215 individus de 12 à 17 ans, en face-à-face, au domicile. Les échantillons sont représentatifs et sélectionnés selon la méthode des quotas.

⁴³Dans une étude comparative menée sur le web et par téléphone, Flemming et Sonner (1999) ont montré que les personnes qui choisissaient de répondre en ligne étaient plutôt des hommes, jeunes et diplômés, dont les réponses sur des questions d'opinion publique variaient significativement par rapport aux répondants par téléphone.

⁴⁴Il y a donc un effet d' entraînement très net, puisque dès qu'une personne possède un équipement technologique (de type téléphone mobile, GPS, ordinateur, internet...), sa probabilité d'en acquérir d'autres s'accroît. Dans ce contexte, internet apparaît comme le prin-

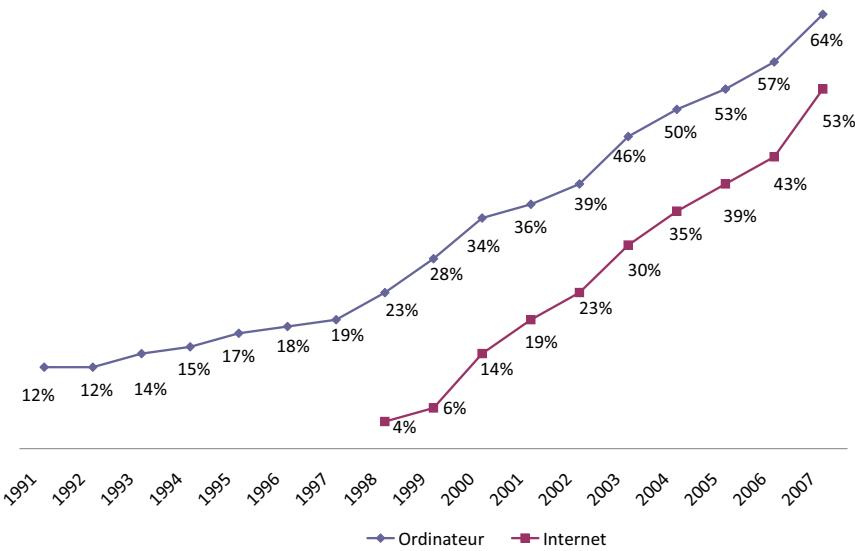


FIG. 4 – Evolution du taux de pénétration de l'ordinateur et d'internet dans les ménages français

Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Des chercheurs ont montré la relation entre les variables sociodémographiques et la peur de l'ordinateur (Igbaria et Parasurman, 1989). Globalement, le niveau de diplôme est le critère le plus déterminant pour expliquer la possession d'un ordinateur (86% des diplômés du supérieur), suivi du nombre de personnes dans le foyer (80% des ménages de 4 personnes et plus) et du revenu⁴⁵. L'accès à l'ordinateur est quasiment généralisé chez les cadres supérieurs (89%) et les professions intermédiaires (86%). Les personnes de plus bas statut social tendent à avoir moins d'expérience et d'attitude positive avec l'ordinateur (Shashaani, 1994). Vient ensuite l'âge, puisque 83% des jeunes de 12-17 ans disposaient d'un ordinateur, contre seulement 17% des plus de 70 ans (annexe IX). Si les répondants âgés sont moins confortables dans l'utilisation des applications informatiques, les jeunes ont une attitude positive et une certaine familiarité avec l'ordinateur. Ils ont davantage confiance dans la technologie (Wright *et al.*, 1998). L'effet de l'âge va cependant s'estomper, avec les cohortes de personnes familières avec l'informatique qui progressent (Couper et Rowe, 1996).

Depuis 2006, il est intéressant de noter que les progressions les plus importantes du taux de pénétration de l'ordinateur se font au sein des catégories les plus démunies, et que les inégalités d'équipement se réduisent (annexe IX).

cipal levier dans l'équipement informatique des ménages.

⁴⁵C'est dans les foyers les plus favorisés que le taux de multi équipement est le plus élevé, puisque 41% des cadres sup possèdent plus d'un ordinateur.

III.2 Une progression rapide des connexions internet dans les ménages français

Lors de la réalisation de l'enquête web de 2006, 43% des personnes de 18 ans et plus étaient connectées à internet à leur domicile, soit environ 21 millions d'adultes sur l'ensemble du territoire, et ces chiffres n'ont cessé de progresser depuis⁴⁶. De façon de plus en plus systématique, la présence au domicile d'un micro ordinateur se traduit par une connexion à internet (75% des personnes qui disposent d'un ordinateur ont également une connexion internet) et on observe les mêmes disparités en termes de catégories socodémographiques (figure 5).

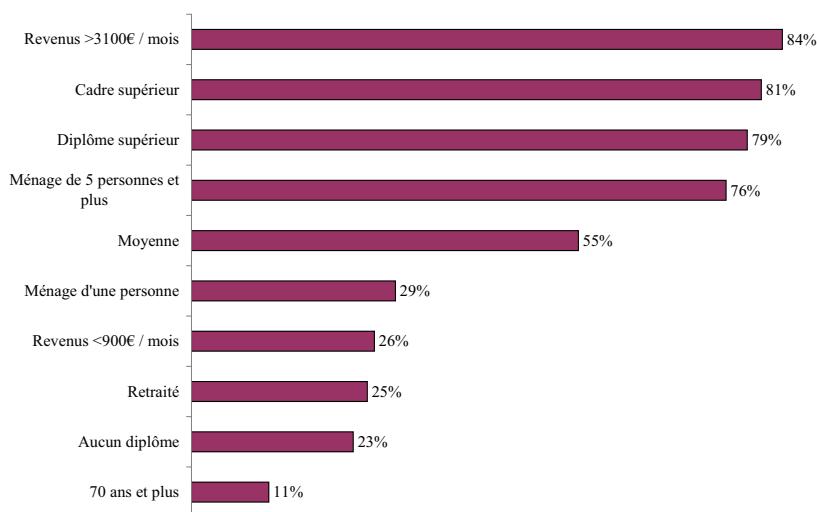


FIG. 5 – Taux d'équipement à internet au domicile
Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Le web reste peu accessible aux personnes âgées (6% des plus de 70 ans et 14% des retraités seulement sont connectés chez eux), aux individus peu qualifiés (un non diplômé sur 5 seulement dispose d'une connexion internet à domicile, contre près des trois quarts des diplômés du supérieur), et aux personnes disposant d'un faible niveau de revenu (22% sont connectées) (annexe IX).

Cependant, un effet de rattrapage s'opère, car les progressions les plus importantes du taux de pénétration d'internet sont à mettre au profit de catégories plutôt sous-équipées, comme les 60 ans et plus, les personnes au foyer et les revenus moyens. Nous assistons à une normalisation du sexe et de l'âge, et les personnes âgées sont de plus en plus sensibilisées. Cette évolution rapide des internautes génère une meilleure représentativité des enquêtes web (Mediametrix, 2000).

⁴⁶En juillet 2008, la proportion de français de plus de 11 ans qui se sont connectés au moins une fois à internet a passé la barre des 60%, soit 31,77 millions de personnes (Médiamétrie).

III.2.1 Des français au bord de la route du haut débit

En 2006, 88% des connexions à internet des particuliers étaient à haut débit, alors qu'en 2004 cela ne concernait qu'une connexion sur deux (55%). Ces connexions se font à 95% grâce à l'ADSL. Le taux d'accès au haut débit le plus bas est enregistré chez les retraités, alors que les travailleurs indépendants et les habitants des communes de plus de 100 000 habitants ont un accès quasiment total au haut débit (95%). Cette différence d'accès résulte de l'inégalité d'équipement en ordinateur, mais également de l'architecture du réseau téléphonique français, puisque 2% des lignes téléphoniques sont inélégibles à l'ADSL, 8% ne permettent pas un accès à plus de 512Kbits/seconde et 27% ne permettent pas un accès à plus de 2Mbits/seconde⁴⁷. Des travaux sont en cours, et c'est selon le critère géographique que le rattrapage est le plus spectaculaire : les ruraux souffraient en 2005 d'un retard d'accès au haut débit de 27 points par rapport à l'agglomération parisienne. Il était de 18 points en 2006 et est seulement de 8 points aujourd'hui.

III.2.2 Quels accès hors du domicile ?

Internet est un outil de communication largement utilisé par les entreprises. L'accès personnel au web sur le lieu de travail n'est pas linéaire. Son évolution a connu une pause en 2005 et 2006 (38% des actifs), puis est repartie en 2007 (44% des actifs). La situation reste fortement contrastée entre les groupes sociaux, puisque plus des trois quarts des cadres supérieurs ont accès à internet sur leur lieu de travail en 2006 (74%), contre seulement 17% des ouvriers. Seule une minorité d'actifs utilisent internet à des fins personnelles sur le lieu de travail, mais les individus concernés sont de plus nombreux et assidus (connexions quotidiennes). Plus de la moitié des étudiants utilisent internet sur leur lieu d'études, mais ces connexions sont ponctuelles.

D'autres solutions existent pour naviguer sur internet, notamment dans des lieux publics (cybercafé, wi-fi, liaison avec le téléphone mobile ...). Il serait néanmoins illusoire de penser que les individus qui n'ont pas accès à internet à leur domicile ou sur leur lieu de travail/études profitent de ces évolutions technologiques pour naviguer régulièrement sur internet. En 2006, les connexions dites 'nomades' sur le web ne concernait que 17% de la population (principalement des étudiants de moins de 25 ans, des personnes diplômées du supérieur et ayant des forts revenus), et ne progressent pas. Le domicile reste donc le lieu de connexion privilégié à internet, suivi du lieu de travail (tableau 6).

III.3 Des usages très variés, selon les caractéristiques des internautes

Sur l'ensemble du territoire, 41% des individus utilisent quotidiennement un ordinateur chez eux, 41% n'ont pas d'ordinateur ou ne l'utilisent jamais

⁴⁷Ces foyers, très nombreux en zones rurales, mais l'arrivée de la fibre optique et du très haut débit, telle qu'annoncée par les trois plus gros opérateurs de télécommunication français, va réduire cette fracture.

	2005	2006	2007
Connexion au domicile	37%	42%	51%
Connexion au travail ou études	25%	27%	31%
Dans cyber-café, bibliothèques	11%	13%	12%
Dans lieu public (wi-fi)	2%	5%	4%
Sur le téléphone mobile	7%	6%	6%
Total	527%	55%	62%

TAB. 6 – Lieux de connexion des internautes français

Source : CREDOC (population de 12 ans et plus)

(personnes âgées, non diplômées, retraités et possédant de faibles revenus) et 19% en ont un mais l'utilisent plus rarement. Si le sexe n'est pas prédictif, l'âge, le niveau de diplôme (80% des diplômés du supérieur l'utilisent tous les jours, vs. 13% des non diplômés) et le revenu semblent déterminants. Le lieu de résidence révèle quelques écarts, puisqu'à Paris et dans les grandes agglomérations il ya deux fois plus d'utilisateurs quotidiens que de personnes qui ne l'utilisent jamais⁴⁸. Chez les actifs, la polarisation est très grande : 71% des cadres utilisent l'ordinateur tous les jours au travail, alors que 79% des ouvriers ne l'utilisent jamais. Les étudiants se trouvent dans une situation intermédiaire, puisque 45% d'entre-eux utilisent un ordinateur sur leur lieu d'études en 2006.

Parmi les personnes qui disposent d'un ordinateur et d'une connexion à internet au domicile, le taux d'utilisation dépasse 90% depuis 2005. La part de connexion quotidienne est très importante (34% des français de 12 ans en 2006) et, depuis 2007, le nombre d'internautes qui se connectent chaque jour dépasse celui des non internautes. La fréquence de connexion à internet varie selon les caractéristiques sociales, l'âge et le diplôme étant les facteurs les plus discriminants : plus on est jeune diplômé, plus on est assidu. Il est également intéressant de noter une corrélation positive entre les différents types de connexion, puisque 76% des individus qui se connectent dans la sphère professionnelle se connectent également chez eux.

La compétence web des internautes dépend largement des services utilisés (navigateur, e-mail, chat, création de blog ...) (Haraldsen *et al.*, 2002). La diversité des accès aux services et activités proposés par internet dépend souvent de l'âge. Ainsi, les adolescents (12-17 ans) créent beaucoup de sites personnels ou blogs (47%), jouent en réseau (43%), et écoutent la radio (47%) ou regardent la télévision sur le net (23%), alors qu'une faible proportion des adultes sont familiers avec ces activités. En revanche, les achats sur le web progressent beaucoup et touchent tous les internautes (47% ont fait au moins un achat en ligne en 2006). Mais ce sont les démarches administratives ou fiscales par internet qui rencontrent le plus de succès (49% des internautes).

La crainte concernant la protection des données individuelles limite cependant considérablement la navigation sur le web (24% des internautes en 2006,

⁴⁸Cela s'explique en partie par la corrélation du lieu de résidence avec d'autres variables sociodémographiques précédemment citées.

et 29% en 2008 sont sensibles à ce problème). Trois internautes sur quatre souhaiteraient pouvoir s'opposer à la collecte des traces de leur passage. Cette crainte concerne surtout les cadres supérieurs, les 40-59 ans et les diplômés du supérieur, les jeunes y étant moins sensibles. Viennent ensuite la complexité de l'outil et son manque d'utilité, pour les plus de 70 ans, les retraités et les non diplômés. Enfin, malgré les offres commerciales des opérateurs, toujours plus nombreuses, le coût reste un frein pour les personnes à faibles revenus et les jeunes (12-17 ans) qui ne disposent pas d'une connexion illimitée à internet.

III.4 Vers une réduction des inégalités

Les différents taux d'équipement en ordinateur et en internet connaissent un rythme de progression soutenu. Mais des inégalités perdurent, et l'ensemble des groupes n'ont pas identiquement accès à ces technologies. Selon la figure 6, les inégalités sont non négligeables pour l'équipement en ordinateur et internet à domicile.⁴⁹. Les indicateurs d'inégalité évoluent toutefois à la baisse, et en ce qui concerne le taux de pénétration d'internet, les inégalités sont deux fois moins importantes aujourd'hui qu'en 2000. Si le 'fossé numérique' s'est réduit, l'accès à un ordinateur et à internet ne garantit pas une utilisation fréquente de ce nouveau média.

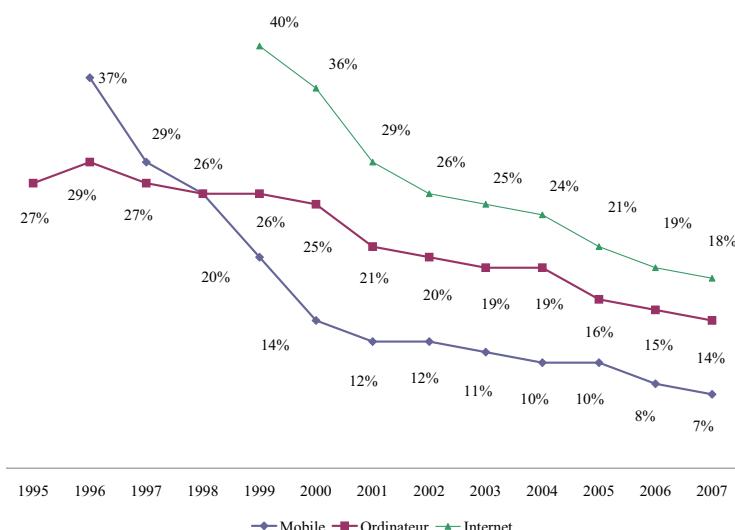


FIG. 6 – Evolution des inégalités d'équipement des ménages français - Indice de Gini

Source : CREDOC, enquête "Conditions de vie et Aspirations des Français" (2007)

Il faut cependant relativiser ces évolutions. Aux USA par exemple, le taux de pénétration d'internet est très élevé (74%). Ceci n'empêche pas qu'un biais

⁴⁹L'indice de Gini est un indicateur synthétique d'inégalités qui varie entre 0 et 100%. Un coefficient de Gini égal à zéro signifie qu'il n'y a pas d'inégalité. Un coefficient égal à 100% signifie qu'un seul groupe détient tout l'équipement et que les autres en sont privés. Entre 0 et 100%, l'inégalité est d'autant plus forte que l'indice de Gini est élevé.

de couverture subsiste, des personnes (bas revenus, faible niveau d'études...) n'ayant toujours pas accès à internet. Si les enquêtes web permettent de toucher des individus qui ne répondent pas aux modes traditionnels, elles excluent toutefois certains groupes sociodémographiques.

IV Conclusion

Les méthodes d'enquête doivent tenir compte de l'évolution des modes de communication entre les individus, puisque la réalisation d'une étude implique l'existence de communications interpersonnelles (Murakami et Morris, 2003). Internet est un espace naturel de communication pour les jeunes générations. C'est un média hautement interactif, dont les récentes applications ont remis l'expression écrite au goût du jour. Ainsi, le développement des blogs, de la messagerie électronique et des forums font que les internautes sont habitués à s'exprimer par l'intermédiaire de leur ordinateur. Les enquêtes web apparaissent comme un moyen de palier à la crise de légitimité que semblent subir le secteur des études, en offrant aux répondants un mode de collecte de données plus proche de leur mode de vie et de communication et en autorisant une grande interactivité, malgré l'absence d'enquêteur pour introduire les questions.

L'adoption rapide d'internet par les foyers français en fait outil intéressant pour les gestionnaires d'enquêtes, qui peuvent interroger rapidement et à moindre coût un grand nombre d'individus. Ils sont capables de concevoir des questionnaires dont le design des questions et des zones de réponses s'adaptent parfaitement aux besoins de l'étude, de lancer quasi-instantanément l'enquête à une large échelle et d'automatiser une grande partie des traitements statistiques (Rose *et al.*, 2004). Le développement du haut-débit en France laisse présager que des enquêtes pourront être menées à grande échelle ces prochaines années, même si internet ne sera peut-être jamais totalement accessible (contraintes de coût, incapacité à naviguer...). En effet, au fil des mois, davantage de foyers sont équipés d'une connexion internet, et les internautes deviennent de plus en plus familiers avec les transactions sécurisées sur le web (qu'il s'agisse de données financières ou personnelles). Avec l'augmentation du nombre des enquêtes en ligne, la population cible s'élargit (Cho et LaRose, 1999) et des études complexes, utilisant des interfaces toujours plus conviviales, vont pouvoir se développer. Par ailleurs, le faible coût des études en ligne permet d'en augmenter la fréquence et de limiter l'impact d'événements spéciaux comme les 'violences urbaines', les grèves ou les attentats sur les résultats (Sharp *et al.*, 2004). Un accent particulier devra cependant être mis sur la confidentialité des données, avec des procédures de contrôle sur le point clé de l'identité du répondant (Michałowska, 2000).

Mais des doutes sur la représentativité des échantillons ont alimenté la polémique autour des enquêtes en ligne, et les interrogations des chercheurs et des professionnels des études rappellent celles posées en leur temps par le développement du recueil de données par téléphone ou par minitel. Les inéga-

lités en termes d'équipement informatique et d'accès au haut-débit (bien que ce dernier soit de plus en plus choisi par les internautes, au détriment du bas débit) compromettent l'utilisation exclusive du web comme mode de recueil de données sur l'ensemble de la population, la représentativité de l'échantillon n'étant plus assurée. Les expériences d'enquêtes web menées jusqu'à aujourd'hui reposent essentiellement sur des méthodes non probabilistes, ce qui limite la qualité des données.

L'objectif des enquêtes ménages déplacements est de fournir régulièrement des données récentes et détaillées pour alimenter les modèles de plus en plus sophistiqués des planificateurs, et permettre aux chercheurs de réaliser des analyses comportementales. Ces enquêtes ne sont pas seulement restreintes à une population d'internautes et nécessitent la construction d'un échantillon qui permette d'appliquer les techniques d'inférence statistique. Il semble que la solution optimale consiste aujourd'hui à intégrer le web dans un protocole d'enquête plus large. Il peut par exemple être utilisé en complément, pour améliorer la performance des modes traditionnels (téléphone, voie postale ou face-à-face). L'utilisation des nouvelles technologies peut en effet potentiellement diminuer la lassitude des répondants et augmenter le taux de réponse global de l'étude, sans grever le coût de l'enquête⁵⁰. Le couplage du web avec un mode de recueil plus traditionnel, tel que le face-à-face, est une solution intéressante pour améliorer la représentativité et la qualité des enquêtes ménages déplacements, en produisant une image plus fidèle du comportement de mobilité de l'ensemble de la population. L'idée est de réussir à capter de nouveaux individus, tout en conservant la méthode traditionnelle de recueil des données, qui permet de toucher une large partie de la population de référence, et d'étudier en profondeur les comportements de déplacements des ménages.

Dans le chapitre suivant, nous présentons la méthodologie retenue pour tester la réalisation d'une enquête web en complément de l'enquête ménages déplacements, effectuée en face-à-face.

⁵⁰Sauf si l'enquête web nécessite la construction d'un échantillon qui autorise la généralisation des résultats à l'ensemble de la population

Chapitre 2 : La mise en place d'un module web dans l'enquête ménages déplacements de Lyon

"Ne jamais oublier que l'enquête, à un moment donné, échappe à ses concepteurs. Elle n'est dès lors que le résultat de la relation établie entre enquêteur et enquêté et la qualité de cette relation est grandement facilité par la mise à disposition d'un bon outil de recueil [le questionnaire]."
(Certu, 1998), à propos des enquêtes ménages déplacements..

Notre recherche consiste à tester puis à proposer une méthodologie d'enquête web adaptée au recueil de données sur les déplacements quotidiens. L'enjeu est conséquent, puisqu'il n'existe pas d'expérience similaire en France en ce qui concerne les enquêtes de mobilité, et que la comparabilité des données avec l'enquête standard réalisée en face-à-face n'a pas été éprouvée à ce jour. La conception du questionnaire web est délicate, puisqu'il s'agit non seulement de respecter les contraintes de la version standard des enquêtes ménages déplacements élaborée par le Certu (1998), mais également de tenir compte des spécificités du web en tant que nouveau média d'enquête. En effet, si le web permet d'améliorer la qualité de l'enquête, il est en revanche important de ne pas accroître la charge pour l'enquêté (Bonnel, 2004). Le design affecte non seulement les taux de réponse et d'abandon, mais aussi la qualité des réponses obtenues (Couper *et al.*, 2001). Un arbitrage doit donc être fait entre convivialité et difficulté d'administration du questionnaire.

La première partie de notre recherche a consisté à modifier et clarifier le modèle de questionnaire utilisé pour les enquêtes ménages déplacements standard, afin de s'assurer que chaque question génère une réponse valide et appropriée au format de l'enquête web. Plusieurs sociétés ont été impliquées dans ce projet (Sofreco, en charge de l'enquête traditionnelle en face-à-face et responsable de l'enquête web, Affimétrie, pour la gestion, la mise en place et le suivi de l'enquête en ligne, et Conversoft, développeur du questionnaire web, à partir du logiciel Converso) et ont imposé certaines contraintes techniques et méthodologiques.

Nous analysons d'abord les difficultés relatives à l'élaboration d'un questionnaire web, et les principales recommandations suggérées par la littérature

(section I). Puis, nous présentons en détail la méthode suivie pour la rédaction des questions (section II), avant d'expliquer d'administration de l'étude en ligne (section III).

I Considérations méthodologiques

I.1 Quelques recommandations

La plupart des recommandations concernant la structure du questionnaire web sont adaptées de celles qui s'appliquent aux autres modes d'enquête, telles que la teneur de l'introduction, la présence de sections et de filtres, l'usage limité des questions ouvertes, la mise en place de pré-tests et le rôle des incitations matérielles (Gaddis, 1998). L'utilisation d'Internet génère toutefois des préoccupations nouvelles. Dillman *et al.* (1998a) recensent par exemple pas moins de onze principes à respecter pour concevoir des questionnaires web de qualité. Andrews *et al.* (0003) présentent cinq éléments relatifs à la présentation du questionnaire et l'administration de l'enquête, essentiels pour obtenir des résultats de bonne qualité : la capacité d'accès avec différents équipements, l'accès sécurisé afin d'empêcher les soumissions multiples, la personnalisation des questions, les sauvegardes partielles et les remerciements en fin de questionnaire. Nous proposons dans cette section une revue de la littérature, suivie de l'application concrète dans l'enquête ménages déplacements web réalisée à Lyon.

I.1.1 Un questionnaire accessible au plus grand nombre

Pour limiter les problèmes techniques et augmenter le taux de réponse, il est nécessaire de faciliter l'accès au questionnaire (Morrel-Samuels, 2003). Du point de vue technique, il est évident que le questionnaire doit être conçu de façon à rester identique quel que soit le type de matériel utilisé par l'internaute, et à pouvoir être facilement adaptable sur d'autres supports. Les questions doivent être simples, ne pas nécessiter de compétences particulières pour y répondre, et doivent être conçues de façon à être téléchargeables par différents systèmes d'exploitation (Mac, Windows) et avec différentes versions de navigateurs (Netscape, Internet Explorer...), des problèmes techniques pouvant impacter négativement le taux de réponse (Bishop, 1997). Les internautes les moins expérimentés ont des navigateurs plus anciens. Les plus expérimentés ont des explorateurs 'libres', dernière génération, ou plus complexes et quelquefois non compatibles avec la version optimisée du questionnaire (Haraldsen *et al.*, 2002). Le concepteur doit également tenir compte des limitations liées à la taille de l'écran, qui varie selon l'équipement des internautes et donne des présentations différentes du questionnaire (Couper *et al.*, 2001; Dillman et Bowker, 2001). Des recherches montrent que les questionnaires simples et rapides à télécharger génèrent de meilleurs taux de réponse (Solomon, 2001).

Lors de l'administration du questionnaire, il est probable que des erreurs de saisie de l'identifiant ou du mot de passe, ou plus en amont une confusion

dans le nom du site hébergeant le questionnaire, contraignent le répondant à l'abandon. Il est recommandé d'utiliser des identifiants et mots de passe courts, afin de limiter le risque d'erreur dans la saisie. Selon Schleyer et Forrest (2000) et Couper *et al.* (2001), les répondants dont les identifiants ou mots de passe contiennent des caractères ambigus ont significativement moins de chance de réussir à accéder au questionnaire web que les autres (la lettre 'O' peut être confondue avec le chiffre '0', la lettre 'I' avec le chiffre '1'). Suivant la même logique, l'adresse du site qui héberge le questionnaire doit être simple et ne pas contenir de caractères spéciaux.

Enfin, beaucoup d'enquêtes internet requièrent Java pour fonctionner. Une application est chargée et lancée sur l'ordinateur du répondant, puis le serveur reçoit les réponses du navigateur du répondant et lui envoie les données appropriées au format HTML. Mais les temps de chargement sont relativement longs, ce qui génère des frustrations auprès des répondants. Le désavantage d'utiliser un programme complexe pour une enquête web est que les répondants ne possèdent pas forcément la bonne version de l'application, et que d'autres peuvent vite se retrouver face à un bug (Stopher *et al.*, 2004a). Ce risque est fort sur un environnement internet, où ces facteurs sont difficiles à contrôler (pare-feu).

Pour l'enquête web de Lyon, nous n'avons eu aucun contrôle sur l'attribution des codes personnels (identifiants et mots de passe) générés par la société Sofreco. En revanche, nous avons choisi un nom de site simple et non ambigu (www.enquete.let.fr). Nous n'avons pas opté pour le langage Java, compte-tenu des nombreux inconvénients développés ci-dessus, et nous nous sommes assurés par l'intermédiaire de tests techniques de l'accessibilité et de la présentation du questionnaire pour différents types d'équipement informatique.

I.1.2 Une introduction de qualité

Lorsqu'il accède au questionnaire, le répondant doit trouver une introduction accueillante, motivante et rassurante, en ce qui concerne la confidentialité des données et la sécurité du site (Dillman *et al.*, 1998b; Lazar et Preece, 1999; Zmud, 2003). Dès les premières pages, des informations sur les manipulations requises pour naviguer dans le questionnaire sont fournies (par exemple le moyen d'accéder à la page suivante), et restent accessibles par la suite. L'objectif est de motiver le répondant, non d'accroître sa lassitude en dressant une liste importante d'instructions. Les instructions concernant des questions précises sont divulguées au fur et à mesure de l'avancement dans le questionnaire, le moment le plus approprié étant celui où l'internaute va en avoir besoin (Dillman *et al.*, 1998a)⁵¹.

Le temps nécessaire pour répondre à l'étude est également indiqué, pour permettre aux internautes pressés de reporter leur participation à l'étude à un moment plus opportun, ce qui réduit leur frustration (Alsnih, 2005). Une barre de progression indique le temps restant avant la fin du questionnaire, bien

⁵¹ Le rapport du Ressource System Group (2002) précise que les instructions doivent figurer sur le même écran que la question concernée, afin de simplifier la tâche des répondants.

que son bénéfice soit largement discuté dans la littérature. Certaines études n'y sont pas favorables et Couper *et al.* (2001) met en garde les concepteurs d'enquête. Si cet outil informe le répondant sur sa progression dans le questionnaire et entretient la motivation, l'utilisation d'éléments graphiques trop sophistiqués peut élever inutilement le temps de chargement des pages web et générer au final des abandons. Enfin, il est important de préciser à quelle date les résultats seront disponibles et qui aura accès aux données (Conn, 2005), pour lever les éventuels doutes sur la confidentialité des données.

Dans l'enquête web de Lyon, les deux premières pages donnent des indications concernant la navigation dans le questionnaire (*'Pour sélectionner vos réponses, il suffit de cliquer sur la puce.'*) et rassurent le répondant sur la confidentialité et la sécurité des données (*'Vos réponses resteront strictement confidentielles, et leur exploitation sera réalisée de façon anonyme.'*). Le temps nécessaire pour répondre au questionnaire est également précisé (*'Cette enquête prend moins de 20 minutes.'*) et une barre de progression permet à l'enquêté de suivre l'avancement du questionnaire (Cf. VI).

I.1.3 Une présentation adaptée au média web

Lorsque le contact avec le répondant a été établi, le questionnaire joue un important rôle de motivation et son design est un élément clé du succès de l'enquête (Andrews *et al.*, 0003). Le questionnaire doit être accueillant et ergonomique, pour donner envie aux internautes d'y répondre (Hojman *et al.*, 2004). Les internautes sont familiers de ce nouveau média et ont des attentes précises en termes de design des pages web. Si l'étude ne fait pas rapidement une bonne impression aux répondants, ils peuvent quitter le questionnaire en fermant la fenêtre web par un simple 'clic'. La mise en page, les questions et les modalités de réponse vont affecter le taux de réponse, le taux d'abandon et les réponses elles-mêmes (Dillman *et al.*, 1998a; Zmud, 2003). Selon Alvarez et Van Beselaere (2003), la validité et la pertinence des résultats d'une enquête web dépendent de la manière dont l'enquête est conçue et administrée. Les éléments du questionnaires qui sont difficiles à comprendre, qui nécessitent un temps de téléchargement important ou qui semblent intérressants pour les enquêtés diminuent les chances de réponse. Ainsi, trop de questions ouvertes, une présentation complexe ou au contraire trop simpliste découragent les répondants. Suivant le type d'étude et le public visé, certains designs sont plus appropriés (Couper, 2000).

Des auteurs suggèrent que le questionnaire web propose les questions d'une manière différente de la version papier, puisque les individus appréhendent différemment les questionnaires lorsqu'ils sont face à un ordinateur (Dillman et Bowker, 2001). Les concepteurs d'enquêtes web tirent partie des logiciels performants et de la capacité de traitement des ordinateurs pour créer des questionnaires très sophistiqués. Par ailleurs, les capacités graphiques des explorateurs de type 'Internet Explorer' ou 'Netscape' permettent d'agrémenter

les enquêtes web de documents multimédia⁵². Les conséquences sont toutefois assez paradoxales. D'un côté, les internautes sont séduits par l'attractivité des questionnaires, de l'autre la complexité de l'outil rend son accès et son utilisation difficile pour certains répondants (Dillman *et al.*, 1998b; Couper *et al.*, 2001; Dillman *et al.*, 1998a). La plupart des possibilités techniques requièrent en effet des ordinateurs performants, des logiciels récents et un temps de téléchargement conséquent pour pouvoir fonctionner. Avec le développement des capacités multimédia, l'écart entre ce que les concepteurs d'enquête peuvent mettre en ligne et ce à quoi les internautes peuvent accéder et répondre se creuse⁵³. Les possibilités multimédia du web ont donc parfois un effet paradoxal, puisqu'elles rendent les questionnaires en ligne plus difficiles à administrer.

Selon Smith (1997), l'aspect visuel du questionnaire est très important, puisque ce sont les aspects non verbaux qui influencent les réponses aux questionnaires web. Or, ce que voit le répondant sur son ordinateur ne correspond pas forcément à ce que voit le concepteur de l'enquête. Andrews *et al.* (0003) montrent que les phrases courtes sont plus facilement lisibles à l'écran, les internautes ne lisant pas tout le texte, mais balayant du regard l'écran, à la recherche de mots clé. Il est également important que les concepteurs d'enquête ne se focalisent pas seulement sur la manière dont opèrent les ordinateurs, mais également sur la façon dont les répondants espèrent que l'enquête va se dérouler. Dillman *et al.* (1998a) précisent que lorsqu'un répondant est sollicité, il va se remémorer ses expériences passées. Les questionnaires en rupture totale avec les méthodes traditionnelles risquent de dérouter les internautes et de conduire à un échec⁵⁴. Enfin, la standardisation de la présentation donne un rythme et assure une certaine continuité du questionnaire. A contrario, de fréquents changements de formats rendent la lecture et la compréhension du questionnaire plus difficiles. Il est donc préférable de faire varier la police de caractères avec parcimonie, pour créer un contraste sans dérouter les répondants. La forme des réponses peut également impacter les résultats. Couper *et al.* (2001) précise que sélectionner une valeur à l'aide d'un 'bouton radio' requiert moins d'effort que saisir une valeur dans une case vide. Les non-réponses sont moins importantes, puisque le répondant est obligé de choisir une valeur, et le traitement des données est simplifié, puisque les réponses sont prédéfinies par le concepteur de l'étude.

Les contraintes techniques que nous détaillons en section I.2.5 ne nous ont pas permis d'optimiser la présentation du questionnaire web lors de l'enquête ménages déplacements de Lyon. Ce dernier étant développé à partir d'un logi-

⁵²Selon (Bachmann *et al.*, 2000), des couleurs, des images de bonne qualité et des graphiques complexes peuvent être utilisés sur le web, ainsi que des commentaires audio et vidéo.

⁵³Le risque est de décourager les personnes les moins familières avec l'informatique, de produire un biais de non-réponse important et de recueillir des données de faible qualité.

⁵⁴Il est recommandé que la forme des questions, tout en respectant les contraintes propres à ce média, se rapproche du format conventionnel que les internautes sont habitués à trouver dans les enquêtes papier (Dillman *et al.*, 1998a).

iciel CATI, il n'a pas été possible de bénéficier de tous les avantages des logiciels CAWI (ergonomie, interface graphique...).

I.1.4 Une progressivité des questions

Une enquête web est par définition auto-administrée. Le rôle essentiel du créateur de l'enquête est d'anticiper au mieux la façon dont les individus vont apprécier le questionnaire, afin de les encourager à saisir l'ensemble des réponses souhaitées (Dillman *et al.*, 1998a). Selon Mertler (2002), les répondants doivent être en mesure de remplir le questionnaire web avec la même aisance que la version papier. La proximité des éléments est importante, et il faut veiller à grouper les questions et les items qui sont proches, pour faciliter le travail de l'interviewé (Conn, 2005). Le questionnaire web doit être concis, structuré par parties (Gunn, 2002; Alsnih, 2005), reliées par quelques phrases de transition. Des termes tels que 'finalement' peuvent être employés en introduction de la dernière partie, pour indiquer au répondant qu'il est proche de la fin du questionnaire, et éviter ainsi les abandons tardifs, grande source d'insatisfaction pour l'enquêté et l'administrateur de l'enquête (Gunn, 2002; Alsnih, 2004).

Concernant le contenu des questions, une progressivité doit être respectée (Frary, 1996). Il est judicieux de commencer par des questions simples et en relation avec le sujet de l'étude, afin d'encourager le répondant à rentrer dans le questionnaire (Ampt, 2003)⁵⁵. Puis, apparaissent les questions plus complexes, et en dernier lieu les questions personnelles. La connaissance des caractéristiques individuelles est nécessaire pour permettre le traitement statistique des réponses, mais ce type de question peut inquiéter l'individu, surtout dans le cas d'un questionnaire auto-administré, et conduire à l'abandon ou à des non-réponses partielles. Si nombreux d'auteurs recommandent de les placer en dernier, Andrews *et al.* (2003) mettent en garde l'effet de surprise que peut provoquer l'apparition de questions personnelles à la fin du questionnaire, ce sentiment de 'trahison' pouvant conduire à un abandon.

Le questionnaire web de Lyon est structuré en quatre parties (Cf. section II.1). Afin de ne pas inquiéter le répondant au point que ce dernier abandonne l'étude, nous avons choisi de commencer le questionnaire par des questions simples sur la motorisation, le lieu d'habitation, et la composition du ménage. Suivent des questions concernant les principales caractéristiques socio-économiques des répondants. Ceci permet d'établir plus rapidement un lien avec les pratiques de déplacement et de renvoyer les questions relatives à l'équipement et aux ressources du ménage dans la dernière partie du questionnaire.

⁵⁵Ainsi, la première question, qui concerne directement le sujet de l'étude, doit être facilement compréhensible et s'adresser à l'ensemble de l'échantillon (Gaddis, 1998; Dillman *et al.*, 1998a).

I.1.5 Limiter les non-réponses

Les internautes ouvrent généralement plusieurs pages web et exercent différentes tâches simultanément (courrier électronique, chat, messengers ...). Leur niveau de concentration et d'attention n'est pas optimal, la durée de réponse au questionnaire plus longue et la proportion de non-réponses partielles parfois élevée (Lozar Manfreda et Vehovar, 2002b). Le problème de la non-réponse partielle est plus important lorsque l'enquête est réalisée sur le web qu'en face-à-face, car aucun enquêteur ne vient relancer le répondant⁵⁶. S'il est possible, en cas de non-réponse, d'avertir le répondant par un message d'alerte ou de forcer la réponse en empêchant le passage à la question suivante, ces choix méthodologiques ne sont pas neutres. D'un côté, nous risquons d'obtenir des réponses fausses ou incohérentes, voire des abandons, puisque le questionnaire est auto-administré et que l'enquêteur n'est pas présent pour motiver l'interviewé (Dillman *et al.*, 1998b; Dillman, 2000; Abraham *et al.*, 1998). Les répondants font davantage d'erreurs quand la pénibilité augmente, et les messages d'erreur renforcent cette pénibilité. De l'autre, nous pouvons penser que les individus qui font la démarche de répondre comprennent plus facilement la rigidité de la structure du questionnaire. Au final, il est utile de décider d'un niveau de tolérance, afin d'éviter de demander des confirmations trop fréquentes ou des corrections qui sèment le doute et augmentent la pénibilité et le caractère inquisiteur des questions (Bonnel, 2003)⁵⁷.

La solution intermédiaire consiste à laisser un échappatoire par la présence d'une modalité 'Ne Sait Pas' ou 'Autre, à préciser' aussi souvent que possible. Les recherches sur le rôle de ces échappatoires dans les enquêtes sont nombreuses (Krosnick et Fabrigar, 1997; Schuman et Presser, 1981; Schuman et Scott, 1989). Elles montrent que proposer une modalité 'ne sait pas' augmente la part des répondants qui la choisissent et nuit à la qualité des données puisque l'analyste est privé d'information. Par ailleurs, la réponse 'Ne Sait Pas' est peu pertinente, si les modalités sont exhaustives. Elle reste en revanche utile pour réduire la frustration d'individus qui ne souhaitent pas s'exprimer sur une partie du questionnaire, quelle qu'en soit la raison. Dans les modes administrés, l'enquêteur peut avoir comme consigne de relancer l'individu qui ne veut pas répondre, et de ne proposer la modalité 'ne sait pas' qu'en dernier ressort. Cette démarche est moins évidente lors des enquêtes administrées, bien que le web autorise une utilisation modérée de ces 'non-réponses'. Il est en effet possible de relancer les répondants à l'aide d'une alerte, avant de proposer explicitement un échappatoire ou de proposer la modalité 'ne sait pas' (éventuellement en surbrillance, pour diminuer son impact visuel) (De Rouvray et Couper, 2002). L'impact de ces différentes formulations sur le comportement de réponse des internautes n'est pas encore clairement démontré. De Rouvray et Couper (2002) rapportent que le taux de données manquantes est plus faible

⁵⁶Haraldsen *et al.* (2002) montrent, dans une expérience où les réponses ne sont pas forcées, 98.6% des questionnaires contiennent au moins une erreur et 82.6% au moins une non-réponse à une question principale.

⁵⁷Des relances simples sont plus efficaces. Elles incitent le répondant à fournir des réponses complètes, sans accroître la pénibilité du questionnaire (Zukerberg *et al.*, 1999).

lorsqu'une relance est prévue, combinée avec une modalité 'ne sait pas' et il semblerait que la proportion de réponse 'ne sait pas' reste inchangée, selon que les répondants visualisent ou non la relance. On peut donc supposer que cette technique permette aux individus qui auraient évité la question de fournir des réponses pertinentes.

En encourageant les enquêtés à donner une réponse à chaque question, améliorons-nous la qualité des données ou augmentons-nous le 'bruit' dans les résultats ? Converse (1974) rappelle que '*instead of browbeating our respondents into giving opinions they did not feel they possessed ...we explicitly invited respondents who had no opinion to report this fact directly*'.

Dans l'enquête web de Lyon, nous avons choisi de proposer des modalités 'Ne sait pas' ou 'Non-réponse' aux questions complémentaires concernant l'équipement en moyens de communication (téléphone fixe et portable, connexion internet à domicile et type de connexion internet à domicile) et le revenu des ménages. Ces questions sont situées à la fin du questionnaire, et risquent d'apparaître trop curieuses auprès des répondants qui viennent de renseigner le cycle de déplacements, sujet central de l'étude.

I.1.6 Les sauvegardes partielles des réponses

Nous savons que les problèmes techniques pénalisent souvent les enquêtes web, et que des perturbations de la connexion peuvent se produire aux dépens du gestionnaire d'enquête. La navigation dans le questionnaire s'interrompt et la saisie des réponses n'est plus possible (Alsnih, 2005). Les individus sont également susceptibles d'être dérangés momentanément au domicile ou au travail, et contraints de suspendre la saisie des réponses. Il est donc indispensable que les répondants puissent reprendre le questionnaire ultérieurement, sans que l'ensemble des données initialement saisies soient perdues. En structurant le questionnaire, il est possible d'enregistrer les réponses au fur et à mesure⁵⁸. Ces sauvegardes partielles demandent peu de temps de chargement et, en cas d'interruption de la saisie, l'internaute est en mesure de rentrer à nouveau dans le questionnaire à l'endroit où il l'avait quitté.

Une des solutions consiste à affecter un mot de passe personnel à chaque individu, qui reste valide durant toute la durée de l'enquête. Pour accéder au questionnaire en ligne, le répondant doit saisir à chaque connexion son identifiant et son mot de passe. Cette identification permet par ailleurs de contrôler l'échantillon, puisqu'un même individu ne peut remplir plusieurs questionnaires (Conn, 2005; Couper *et al.*, 2001). Le risque est que les interviewés ne comprennent pas le caractère aléatoire de l'affectation des mots de passe et perçoivent cette mesure comme une atteinte à la confidentialité de l'étude, ce qui peut modifier leurs réponses (Stanton et Rogelberg, 2001).

⁵⁸Certains questionnaires proposent même un bouton 'click-save-exit' sur chaque page, permettant au répondant de quitter le questionnaire à tout moment sans perdre les informations déjà saisies (Gaddis, 1998; Morrel-Samuels, 2003).

Dans l'enquête web de Lyon, un identifiant et un mot de passe figurent sur la lettre avis envoyée aux ménages sélectionnés. Ces informations sont nécessaires pour accéder au questionnaire et permettent d'enregistrer les réponses saisies. Ces sauvegardes interviennent à plusieurs endroits du questionnaire (Cf. section III.3).

I.1.7 Les filtres et tests de cohérence : personnalisation des questions et qualité des données recueillies

De nombreux filtres sont présents dans les questionnaires web. Ils permettent d'éviter la redondance des questions ou les questions non pertinentes, qui allongeraient inutilement la durée du questionnaire et limiteraient sa crédibilité. Leur rôle est encore plus fort que dans le cas d'un questionnaire administré en face-à-face ou par téléphone, puisque l'interviewé ne connaît pas a priori le questionnaire et ne sait pas comment naviguer à l'intérieur de celui-ci. Par ailleurs, il est fréquent que des tests de cohérence opèrent au fur et à mesure de la présentation des questions. L'objectif est de vérifier la structure du fichier et de détecter des contradictions éventuelles dans les réponses saisies. Soit l'internaute valide les informations saisies et est dirigé vers la suite de l'enquête, soit il modifie sa réponse. La forme de la relance, priviliege la souplesse et l'ergonomie pour l'enquêté.

Dans l'enquête web de Lyon, nous avons choisi de placer des filtres aussi souvent que possible, afin de personnaliser le questionnaire. Des tests de cohérence sont également utilisés, surtout dans la partie relative aux déplacements, pour s'assurer de la pertinence des réponses. Par exemple, si le répondant déclare ne pas être sorti de chez lui durant la période d'enquête ou si l'heure de fin d'un déplacement est postérieure à celle de fin de l'enquête, une alerte est envoyée pour demander des corrections. Il est également important de s'assurer que les personnes qui ne déclarent pas de déplacements soient des vrais immobiles. Les utilisateurs de données assurent en effet qu'une large proportion de non mobiles témoigne de la faible qualité d'une enquête⁵⁹. Pour certaines questions, l'utilisation de tests de cohérence est plus délicate. Ainsi, lorsqu'un répondant ne déclare pas son domicile comme lieu de destination du dernier déplacement, il est probable qu'il n'ait pas correctement assimilé la définition du terme 'déplacement'. Mais il se peut également qu'il n'ait pas passé la nuit chez lui. Dans ce cas, nous n'avons pas placé de relance. L'objectif est d'obtenir des données de qualité, pas de ralentir inutilement la durée de l'enquête.

I.1.8 La mise en oeuvre du pré-test

Avant de lancer une enquête, il est fortement conseillé de vérifier sur un petit échantillon de la population si les principaux objectifs de l'étude sont atteints et si les résultats sont bien ceux espérés (Zmud, 2003). Selon Oppenheim

⁵⁹Au moment du codage, il peut être intéressant de différencier les personnes qui ne se sont pas déplacées de celles qui refusent de dévoiler des informations sur leurs comportements de mobilité.

(1992), '*Survey piloting is the process of conceptualizing and re-conceptualizing the key aims of the study, and making preparations for the fieldwork and analysis so that not too much will go wrong and nothing will have been left out*'. Il s'agit de tester la bonne compréhension des questions, la faisabilité technique de l'étude par rapport aux différents équipements existants sur le marché, d'estimer de façon réaliste le temps de réponse et le bon positionnement des filtres (Alsnih, 2004). Les fréquentes erreurs rattrapées grâce au pré-test sont : le biais dans la formulation des questions⁶⁰, les erreurs de rédaction, les questions inappropriées, les instructions manquantes, le vocabulaire technique non défini, le manque de place pour répondre aux questions ouvertes, et le manque de techniques pour maintenir la motivation du répondant (Andrews *et al.*, 0003). Les tests techniques aident au contrôle des procédures de stockage de l'information, à la gestion et à la récupération des données collectées et à la compatibilité des navigateurs (Benfield et Szlemko, 2006). Nous avons vu dans la section I.1.1 qu'il est recommandé de vérifier la présentation du questionnaire pour des versions anciennes de navigateurs (Dillman et Bowker, 2001). Il est également souhaitable de tester la capacité du serveur sur lequel va être hébergé le questionnaire, en envoyant simultanément plusieurs requêtes sur le site web dédié à l'enquête.

Les pré-tests peuvent dans un premier temps concerner des personnes familières avec l'étude, pour déceler les principaux problèmes, puis être menés sur une plus grande échelle, auprès d'individus de la population étudiée. Selon Morris et Adler (2003), au moins trente questionnaires doivent être remplis pour pouvoir tirer des conclusions du pré-test, les individus concernés étant ensuite exclus de la base de sondage⁶¹. Afin de tester diverses versions du questionnaire, il arrive que plusieurs pré-tests soient menés successivement.

Pour l'enquête web de Lyon, nous avons réalisé deux séries de pré-tests, tant pour évaluer la faisabilité technique du questionnaire que tester sa bonne compréhension par les internautes. L'objectif était de pouvoir améliorer sensiblement la qualité du questionnaire, avant de lancer l'étude auprès des non-répondants à l'enquête standard. Nous avons d'abord réalisé un test de validation en interne auprès de dix personnes familières des enquêtes ménages déplacements, afin d'évaluer la pertinence de la méthodologie retenue pour le recueil des déplacements individuels. Dans un deuxième temps, la version finale du questionnaire a été diffusée auprès de personnes non familières des enquêtes de déplacements, aux caractéristiques socio-économiques diverses. Des entretiens semi-directifs (téléphoniques ou en face-à-face) ont été menés auprès de ces répondants, suivant des thèmes déterminés à l'avance (annexe III) :

- *Evaluation de la durée du questionnaire* : minuter le questionnaire, pour avoir une estimation fiable du temps de réponse, évaluer l'impact de la barre de progression, s'interroger sur la 'lourdeur' du questionnaire et vérifier si certaines questions doivent être supprimées.

⁶⁰Par exemple si on obtient une grande proportion de réponse 'Autre' (Pratt, 2003).

⁶¹Un pré-test diffère d'une enquête pilote. Cette dernière est menée sur un petit échantillon de la population qui sera enquêtée, afin de tester chaque élément de l'enquête.

- *Evaluation du contenu du questionnaire* : tester la compréhension des questions et la pertinence des réponses, vérifier l'utilisation du bon vocabulaire, vérifier l'impact des instructions sur le bon déroulement du questionnaire, identifier les causes possibles d'abandon.
- *Evaluation de la structure du questionnaire* : vérifier que les textes d'accroche initiale, de conclusion et de liaison sont compréhensibles et suffisants, contrôler la cohérence du cheminement du questionnaire et valider la présentation du questionnaire.

D'une façon générale, aucun problème de fonctionnalité du questionnaire web n'a été formulé dans les pré-tests (Les verbatims recueillis sont présentés en annexe IV). Ils ont permis de valider la durée moyenne du questionnaire (15 minutes en moyenne), temps acceptable par les répondants ('*Un peu plus de 10 minutes pour remplir le questionnaire. Honnêtement, cela ne semble ni long, ni difficile à remplir*'), malgré une certaine lourdeur des questions ('*Le questionnaire n'est pas long, à condition bien sûr de ne pas avoir une vie trépidante avec plein de déplacements en journée*'). L'ajout de la barre de progression n'est visiblement pas un atout du questionnaire ('*La présence de la barre est attrayante au début, puis nous ne l'utilisons plus*'). Les sauvegardes partielles des réponses semblent en revanche appréciées, bien qu'un répondant précise qu'*'il faudrait renseigner l'internaute sur les enregistrements intermédiaires des données, afin de ne pas le décourager en cas d'interruption de la saisie'*. Concernant la compréhension des questions, on constate une '*bonne compréhension du questionnaire*'. Si la formulation semble claire, elles '*demandent des réponses super détaillées des fois, ce qui est un peu fastidieux*'. La saisie des adresses est globalement lourde et certains précisent qu'*'il est difficile de se souvenir des adresses exactes des lieux, et nous avons tendance à ne pas renseigner tous les champs'*. La pénibilité de la partie sur les déplacements ('*La partie où il faut décrire l'ensemble des déplacements effectués est assez rébarbative et il est certainement difficile de se souvenir avec exactitude des horaires exacts...*') peut générer des abandons ou l'omission de déplacements ('*Lorsque nous connaissons le temps nécessaire pour saisir un déplacement en totalité, nous avons tendance à omettre certains déplacements*'). Les textes d'accroche, de liaison et de conclusion ne posent pas de problèmes. A contrario, certaines questions apparaissent comme trop personnelles ('*Que viennent faire les questions sur les téléphones et l'Internet ? L'enquête porte-t-elle bien sur les transports ou recherche-t-on à nous vendre quelque chose derrière ?*').

L'évaluation du design du questionnaire montre que '*les aides visuelles et les instructions sont claires dans l'ensemble*', mais que le nombre de validations et les relances sont trop nombreuses. Dans l'ensemble, '*c'est fonctionnel et ça se remplit facilement au moins*'. Ces remarques ont conduit à certaines simplifications du questionnaire, mais ont surtout mis en évidence un manque d'ergonomie ('*Pourquoi l'exemple avec les symboles (maison, travail ...) ne sont ils pas repris ensuite, ce serait plus attractif ?*', '*Lorsque nous saisissons un déplacement, il faudrait proposer la saisie du retour de façon automatique. Cela éviterait de tout ressaisir*') et des problèmes de mise en page du question-

naire ('*Mauvaise adaptation des pages du questionnaire à la taille de l'écran, et nécessité d'utiliser les ascenseurs horizontaux et verticaux*'). L'utilisation d'un logiciel adapté aux enquêtes web semble incontournable pour que l'internaute ait l'impression de se trouver face à n'importe quel site web lui demandant des informations pour traiter sa demande (comme les sites d'information/réservation de compagnies ferroviaires, aériennes ou d'agences de voyages, ou des sites de vente en ligne...). Au final, les pré-tests ont permis de mettre en ligne un questionnaire efficient et attractif, et ont révélé quelques problèmes techniques qui auraient pu entraver la réalisation de l'étude (la lenteur du serveur internet ou le temps d'attente relativement long pour le téléchargement des pages du questionnaire, à cause du grand nombre de répondants essayant d'accéder à l'enquête en même temps).

I.2 Les principales difficultés rencontrées

Les internautes ont des attentes concernant le graphisme ou l'ergonomie des questionnaires web auxquels ils acceptent de se soumettre, et il est fort probable qu'un design 'au rabais' ait un impact très négatif sur le taux de réponse final (Morrel-Samuels, 2003). Une fois prise en compte les recommandations détaillées ci-dessous, la mise en ligne de l'enquête ménages déplacements réalisée sur le web pose quatre problèmes majeurs : la longueur et la sophistication du questionnaire, le nombre de personnes sollicitées et de fortes contraintes légales.

I.2.1 Des interviews en face-à-face traditionnellement longues

Les études de mobilité réalisées traditionnellement en face-à-face sont généralement longues, car les réponses doivent être très précises et concerner l'ensemble du ménage. Les interviews durent une heure et trente minutes en moyenne, chaque personne du ménage de 5 ans et plus étant enquêtée personnellement. L'utilisation du web nécessite d'abord de considérer certaines caractéristiques propres à ce média, comme le temps de téléchargement des pages, proportionnel à la durée du questionnaire, ou la durée de concentration du répondant. Par ailleurs, les personnes concernées par les enquêtes en ligne sont souvent pressées et ne veulent pas passer trop de temps à télécharger le questionnaire ni à saisir leurs réponses⁶². Il a été démontré que de longs questionnaires auto-administrés produisaient des taux de réponse plus faibles (Dillman *et al.*, 1993), même si certains auteurs doutent de l'existence d'une relation significative entre la longueur du questionnaire et le taux de réponse (Bogen, 1996). D'autres (Michałowska, 1999) ont insisté sur la nécessité d'utiliser le média web en tant que tel et de ne pas transposer directement en ligne les versions de questionnaires papier ou téléphone.

Nous avons allégé le questionnaire standard, afin d'adopter une longueur

⁶²Les internautes ont une capacité variable à utiliser leur ordinateur, mais il existe un temps 'limite' (entre 15 et 20 mn) à ne pas dépasser pour une enquête web (Gunn, 2002).

raisonnable, en supprimant les questions d'opinion, qui ne paraissaient pas indispensables à notre recherche⁶³. En revanche, nous avons conservé les questions d'ordre sociodémographiques, nécessaires pour identifier le profil des répondants, et évité les simplifications abusives, sources de biais dans les réponses. Les intitulés de certaines questions ont également été modifiés, pour les rendre plus compréhensibles. En effet, si dans un questionnaire administré en face-à-face ou par téléphone l'enquêteur peut expliciter certains termes non compris par les répondants, dans le cas d'une enquête web, l'internaute se retrouve seul face à son ordinateur et doit être en mesure de comprendre rapidement la question posée. Pour ce travail, nous nous sommes inspirés du questionnaire CERTU Enquête Déplacements Villes Moyennes, réalisé par téléphone. Ce questionnaire est relativement court, puisqu'il est administré en vingt minutes en moyenne, et présente l'avantage d'être facilement transposable sur le web, puisque les interviews téléphoniques sont assistées par ordinateur (CATI).

I.2.2 Un questionnaire standard relativement complexe

Pour encourager le maximum de réponses, les atouts du web (interactivité, possibilité graphiques et sonores ...) ne doivent pas être utilisés de façon abusive, des expériences comparatives montrant un taux de réponse plus élevé pour un questionnaire simple que pour un questionnaire plus sophistiqué (Dillman *et al.*, 1998b). Un questionnaire complexe peut dérouter les internautes, notamment ceux qui ne maîtrisent pas encore suffisamment l'utilisation d'Internet. Par ailleurs, en présence de plusieurs modes d'enquêtes, la similarité des questionnaires est importante. Une construction trop complexe, qui peut apparaître comme conviviale et personnalisée en ligne est difficile à retranscrire sur papier, que ce soit pour le répondant dans le cas d'une enquête postale, ou pour l'enquêteur dans le cas d'une enquête en face-à-face non assistée par ordinateur. Les questionnaires trop disparates rendent très difficile la comparabilité des enquêtes et il est préférable de conserver une version électronique proche de la version papier.

Concernant la présentation du questionnaire, proposer l'ensemble du questionnaire sur une seule page, l'ascenseur permettant aisément le déplacement du curseur sur l'écran, permet d'obtenir un temps de réponse plus court et moins de non-réponses partielles, mais peut affecter la qualité des données, en produisant davantage de réponses 'similaires' (Gunn, 2002). D'un autre côté, faire figurer une question par page permet d'étudier a posteriori les comportements de réponse des individus et surtout de personnaliser le questionnaire en fonction des réponses déjà fournies. Cela permet d'écourter le questionnaire, car il est possible d'utiliser des filtres pour ne retenir que les questions pertinentes pour l'individu interrogé⁶⁴. L'utilisation d'une seule question par

⁶³Le questionnaire web a ensuite été validé par l'ensemble des acteurs impliqués, le CERTU, le SYTRAL et le CETE fin septembre 2005.

⁶⁴Si toutes les questions sont visibles, le répondant qui ne maîtrise pas l'usage des filtres risque de surestimer le temps de réponse et de ne pas comprendre pourquoi certaines questions ne lui sont pas posées.

page nécessite toutefois de donner un peu de flexibilité dans la navigation du questionnaire, à l'aide de boutons 'précédent' et 'suivant', et de permettre une correction/validation rapide des réponses saisies à l'aide de boutons 'Effacer' et 'Envoyer' (Smith, 1997). La mise en page du questionnaire, la rédaction des questions, le vocabulaire utilisé et la qualité du graphisme sont autant de détails qui permettent de conserver l'attention des internautes jusqu'à la fin du questionnaire (Michałowska, 1999).

Le questionnaire de l'enquête ménage est relativement complexe, des données très précises sur les déplacements et leur enchaînement étant demandées aux répondants. Notre objectif a été de développer un questionnaire convivial, pour attirer de nouveaux répondants, tout en restant simple, pour ne pas alourdir le travail de réponse et décourager les internautes inexpérimentés. Les contraintes techniques n'ont pas permis de choisir la présentation des questions. Une seule question figure sur chaque écran, l'internaute devant choisir une modalité et valider sa réponse pour accéder à l'écran suivant.

I.2.3 Un seul répondant sélectionné dans le ménage

Pour analyser la cohérence des déplacements, qui s'inscrivent dans une logique au sein du ménage, il faut interroger tous les membres et déterminer un jour de référence pour le recueil des données (Certu, 1998). Les grands ménages répondent moins facilement à ce type d'enquête, ce qui génère un biais puisque leur mobilité et leurs caractéristiques sociodémographiques diffèrent de ceux des autres répondants (De Heer et Moritz, 2000). Pour certains modes de recueil, notamment auto-administrés, il est par définition difficile d'imposer le jour de référence des déplacements et d'obtenir les réponses pour l'ensemble des membres du ménage (Bonnel, 2003). Par ailleurs, les recherches montrent que le nombre de déplacements est sous-estimé lorsqu'il est rapporté par une tierce personne (Sharp *et al.*, 2004; Richardson *et al.*, 1995) et que ces pratiques ne permettent pas d'obtenir les réponses détaillées (Stopher et Jones, 2003).

Techniquement, il aurait été possible de préciser ces contraintes aux individus dans la lettre de présentation de l'étude, puis de leur laisser le choix de suivre ou non les directives. Concrètement il semble difficile de demander à l'ensemble des personnes du ménage de se connecter le même jour pour remplir tour à tour le questionnaire web. Cette restriction peut avoir un effet repousoir sur les répondants potentiels (Bonnel et Le Nir, 1998), et générer des erreurs de mesure (réponses rapportées, ne concernant pas le même jour de référence ...). Enfin, il ne faut pas oublier que l'enquête web est destinée aux personnes non joignables ou refusant de recevoir un enquêteur à domicile, et qu'un des principaux attraits du média web est sa souplesse. Dans notre démarche exploratoire, nous avons opté pour le remplissage d'un seul questionnaire par foyer (par toute personne de 11 ans et plus), quelle que soit la composition de celui-ci. Cette souplesse contraste fortement avec l'enquête en face-à-face et ne permet pas de connaître avec certitude l'identité du répondant.

I.2.4 Un cadre législatif précis

Avant le lancement d'une enquête auprès du grand public, il est nécessaire d'effectuer des déclarations à la C.N.I.L., qui s'intéresse à la constitution de l'échantillon et à l'exploitation des données.

Dans la mesure où le fichier relatif à l'échantillon utilisé pour l'enquête web existe déjà (interviews prévues en face-à-face, mais non réalisées), une nouvelle déclaration C.N.I.L. de constitution du fichier des enquêtés n'a pas été nécessaire. En revanche, le questionnaire web contient des questions spécifiques et a fait l'objet d'une déclaration C.N.I.L. de traitement des données. De plus, lors des enquêtes ménages déplacements précédentes, il était impossible de connaître les adresses précises des lieux visités par les enquêtés, puisque l'agglomération Lyonnaise était découpée en zones. Or, dans le cadre de notre recherche sur la faisabilité d'un questionnaire web, nous souhaitions obtenir l'intégralité des informations sur les origines et les destinations des déplacements. Légalement, il est possible de disposer des coordonnées précises d'un individu, si ce dernier n'est pas identifié, c'est-à-dire de gérer deux bases de données, une avec les adresses et une avec les noms et numéros de téléphone. Il est également intéressant de demander aux individus qui le souhaitent d'indiquer leur adresse mail à la fin du questionnaire, pour pouvoir les recontacter en cas de besoin (phase de suivi). Les questions posées n'étant pas relatives à l'état de santé du répondant, la C.N.I.L. ne s'est pas opposée à cette demande. Enfin, la loi informatique et liberté est mentionnée dans la lettre avis envoyée aux enquêtés, et rappelée en bas de la première page web du questionnaire, où un lien permet d'accéder directement au site de la C.N.I.L.

I.2.5 Contraintes de réalisation

Tous les logiciels d'enquête proposent désormais un module web, dont la qualité est très inégale selon les prestataires. Les interface CATI réécrites pour les enquêtes CAWI sont souvent peu performantes et les meilleurs produits sont ceux conçus spécifiquement pour administrer des enquêtes en ligne. Ils permettent notamment de visualiser les réponses saisies en temps réel, de traduire le script du questionnaire en plusieurs langages et d'exercer un réel contrôle sur le design des pages du questionnaire et sur la façon dont elles apparaissent aux enquêtés (Burke et James, 2006). Par ailleurs, il est important de définir dès la conception de l'étude le format sous lequel les données saisies seront enregistrées et vers quels logiciels elles pourront ensuite être exportées, en vue d'analyses statistiques et économétriques.

Dans l'enquête ménage déplacements de Lyon, le web constitue une annexe au cahier des charges. Ainsi, nous n'avons pas pu choisir le prestataire chargé du développement du questionnaire en ligne. SOFRECO, qui gère la réalisation de l'enquête face-à-face, a développé le questionnaire web en utilisant l'interface CATI du logiciel 'Converso', adapté aux enquêtes téléphoniques. Cette contrainte n'a pas permis d'apporter au questionnaire web toute la souplesse

que ce média peut offrir aux enquêtes de mobilité. Les données validées par les enquêtés alimentent une base de données au format 'Converso', peu répandu, et une importante phase de mise en forme du fichier fut nécessaire avant de l'exporter vers des logiciels de gestion de bases de données. Les arbitrages technologiques sont un des facteurs clés de succès d'une enquête web. Faire l'impasse d'un logiciel approprié pour l'administration du questionnaire et ne pas assez anticiper les besoins en termes de traitement des données nuisent au bon déroulement de l'étude.

II La rédaction des questions

Face à un ordinateur, les individus n'appréhendent pas une enquête de la même manière que si elle leur est soumise en face-à-face (Dillman et Bowker, 2001). Ce résultat est particulièrement important dans les enquêtes de mobilité urbaine, où le recueil des données de déplacements reste difficile à mettre en oeuvre. La perspective de comparabilité des données ne doit pas conduire à figer le questionnaire, mais au contraire à l'adapter aux comportements des enquêtés face à chaque type de média et à la présence ou non d'un enquêteur. Sur le web, il est indispensable que le questionnement apparaisse comme le plus 'naturel' possible pour l'enquêté. Afin de faciliter la tâche des enquêtés, et de pouvoir comparer les résultats d'enquêtes menées à des horizons temporels différents ou par plusieurs modes de recueil, il est préférable de conserver le mode de rédaction des questions et des modalités de réponses couramment utilisées (Stopher *et al.*, 2004b). Enfin, même si l'enquête web est simplifiée par rapport à la version face-à-face, il faut maintenir un noyau de questions permettant de recueillir des données de mobilité précises et de les croiser avec des caractéristiques sociodémographiques pertinentes (Pratt, 2003).

II.1 Présentation du questionnaire web

Notre questionnaire s'intitule 'Enquête web sur les déplacements dans l'agglomération lyonnaise'. L'introduction reprend des éléments de la lettre avis envoyée aux enquêtés, en insistant sur l'importance des informations recueillies et sur la confidentialité des données. La charte graphique du questionnaire web est basée sur celle que le Sytral a utilisé dans sa campagne de communication, pour conserver une certaine homogénéité⁶⁵. Le principal changement concerne la structure du questionnaire. En effet, le questionnaire standard CERTU est organisé en quatre parties, et recueille successivement :

- les caractéristiques du ménage et du logement ;
- les caractéristiques socio-économiques de chaque personne composant le ménage ;
- les déplacements de chaque personne du ménage âgée de 5 ans et plus, effectués la veille du jour de l'enquête ;

⁶⁵Toutefois, le logo du Sytral n'apparaît pas à l'écran, pour ne pas faire croire aux répondants que l'enquête s'intéresse uniquement aux déplacements effectués en transports en commun.

- des opinions auprès d'une personne du ménage âgée de 16 ans et plus, après tirage au sort.

Le questionnaire web est structuré en quatre parties, ou 'blocs', conformément au questionnaire C.A.T.I., utilisé pour les enquêtes téléphoniques. Nous distinguons d'abord les questions relatives au ménage, celles concernant la personne interrogée, puis celles concernant les données de déplacements, avant de terminer par des questions complémentaires sur l'équipement multimédia du ménage et le niveau de revenu (annexe VI).

II.2 La collecte d'informations sur les caractéristiques du ménage et du répondant

II.2.1 Les caractéristiques générales du ménage

Par définition, l'unité statistique des enquêtes ménages déplacements est le ménage. Dans les interviews en face-à-face ou par téléphone, plusieurs personnes répondent au questionnaire, et fournissent des renseignements sociodémographiques. Ce n'est plus le cas pour l'enquête web, où un seul individu remplit le questionnaire. Il est important que l'analyste connaisse les caractéristiques des différentes personnes du ménage et leur relation avec le répondant, pour appréhender les comportements de mobilité (Pratt, 2003). La première question permet donc de définir l'âge, le sexe et l'occupation des personnes qui composent le ménage, en se basant sur les renseignements fournis par le répondant. Les prénoms sont également demandés, bien qu'ils n'aient pas d'utilité directe, afin de personnaliser le questionnaire et de faire le lien avec les questions individuelles. L'adresse du domicile est recueillie avec précision, car c'est un élément explicatif fort des comportements de mobilité, mais n'aura pas à être ressassé par le répondant dans la partie relative à ses déplacements. Concrètement, un cadre détaillé est proposé à l'enquêté (n°, rue, lieu dit, code postal, commune et département) et une relance apparaît lorsqu'il manque des informations essentielles.

Certains renseignements concernant le ménage sont reportés à la fin de l'interview. Il s'agit de questions sur l'équipement téléphonique et les revenus annuels nets du ménage. Les premières ne sont pas directement liées à la mobilité et peuvent être ressenties comme une intrusion dans la sphère privée. La seconde est traditionnellement très sensible, et risque d'effrayer les répondants. Connaître l'équipement des ménages en moyens de communication est particulièrement important dans cette enquête, qui s'adresse à des internautes. Nous avons donc rajouté une question par rapport au questionnaire standard, afin de connaître le type de connexion Internet possédée au domicile. La distinction se fait entre haut et bas débit. La notion de forfait limité est également intéressante, puisque les internautes qui en bénéficient sont moins contraints, mais reste peu discriminante, puisque les offres haut débit proposées sur le marché permettent majoritairement un accès illimité au web.

Par analogie avec l'enquête standard, nous avons placé la question concernant les revenus annuels du ménage à la fin de l'étude. Cette variable, souvent

utilisée dans les enquêtes ménages déplacements, est plus utile pour appréhender la mobilité des membres du ménage que le niveau de ressources personnelles du répondant (Pratt, 2003). Les classes de revenus sont similaires à celles utilisées dans le questionnaire en face-à-face, afin de faciliter les comparaisons⁶⁶.

II.2.2 Les caractéristiques du répondant

Avant de passer aux renseignements individuels, il est utile de vérifier quel individu 'se cache' réellement derrière l'ordinateur. Ayant demandé à l'internaute de saisir les prénoms des membres du ménage, en commençant par lui-même, il est peu délicat et inconsistant de lui redemander son identité. L'ordinateur génère donc automatiquement le premier prénom saisi, et demande simplement une confirmation.

Une fois pris en compte le nombre de voitures possédées par le ménage, le renseignement individuel le plus pertinent dans le cadre d'une enquête de mobilité est la possession du permis de conduire. Puis, viennent des questions concernant le niveau d'étude, le type d'emploi occupé, et l'adresse exacte du lieu de travail ou d'étude. Cette adresse est enregistrée dans le système et est utilisable de nouveau dans le recueil des déplacements pour les trajets impliquant le lieu de travail. Contrairement au lieu de résidence, il est possible que certaines personnes ne connaissent pas les coordonnées du lieu de travail avec exactitude. Une zone de saisie 'en clair' permet aux individus d'indiquer des renseignements divers (nom d'établissement, lieu public le plus proche, croisement de rues ...), qui seront utilisés pour une codification a posteriori. Enfin, nous recueillons la profession de l'individu enquêté. Dans un souci d'adaptation du contenu des questions au média utilisé, nous proposons un menu simplifié pour le choix de la profession principale, complété si besoin par une liste détaillée.

II.2.3 La fréquence d'usage des modes de transport

Diverses questions concernant la fréquence d'utilisation des modes de transport dans les déplacements quotidiens complètent le questionnaire individuel. Comme dans l'enquête face-à-face, nous distinguons la voiture particulière, conducteur ou passager, les transports collectifs et les deux roues. La distinction de puissance entre les deux roues à moteur est maintenue (bien que le nombre de réponses pour chacune de ces modalités soit trop faible pour être exploité statistiquement), puisque les usages de ces deux types de véhicules ainsi que leurs impacts environnementaux sont différents.

Des questions concernent l'utilisation de la voiture pour se rendre sur le lieu de travail ou sur le lieu d'étude, ainsi que les éventuelles difficultés de stationnement rencontrées par le conducteur. En effet, les déplacements quotidiens pour motif travail ou étude sont nombreux, et la possibilité de pouvoir

⁶⁶Dans notre étude, la modalité 'Ne Sait Pas' est largement choisie, quel que soit le mode d'enquête, de nombreux individus refusant de dévoiler des informations liées à leur pouvoir d'achat (Cf. Chapitre 3 section III.3).

garer ou pas un véhicule à proximité de son lieu d'activité principale impacte fortement l'utilisation de la voiture particulière.

Au final, les questions du bloc 'Ménage' et 'Individu' restent relativement proches de celles utilisées dans l'enquête ménages déplacements réalisée par téléphone dans les villes moyennes et en face-à-face (thèmes, ordre d'apparition, formulation ...). La différence principale réside dans le tableau de composition du ménage, que les internautes doivent remplir au début du questionnaire.

II.3 La collecte d'informations sur les déplacements individuels

Les enquêtes ménages déplacements requièrent des données précises concernant les lieux fréquentés par le répondant durant la période de référence de l'étude. Dans le cas de Lyon, cette période ne dure que 24h, mais la mémoire des individus est fortement sollicitée. La partie du questionnaire dédiée aux déplacements doit donc être conviviale et motivante, et permettre le recueil de données précises.

II.3.1 Une approche mixte déplacements-activités

Deux approches sont possibles pour le recueil des données de mobilité : une basée sur les déplacements, l'autre sur les activités. Dans l'approche basée sur les déplacements, les répondants doivent renseigner tous les champs relatifs à un déplacement, avant de passer au suivant. Dans l'approche par activité, l'individu liste les activités effectuées dans la journée, ainsi que leur horaire de départ, puis renseigne pour chaque activité les données de localisation et les caractéristiques des déplacements entre ces activités, lorsque la localisation change (le mode de transport, la durée, et éventuellement le nombre de personnes participant au déplacement). Selon Ressource System Group (2002), les enquêtés préfèrent les questionnaires qui proposent une approche par activité. Proposer un cheminement des activités effectuées la journée de référence permet en effet de recueillir les déplacements d'une manière naturelle pour l'interviewé et l'effort de mémoire est moins important que dans le cas d'un recueil chronologique des déplacements. Des travaux anciens (Jones *et al.*, 1980) ont montré que l'approche par les activités générait davantage de déplacements que celle par les déplacements. D'autres, plus récents, confirment cette hypothèse⁶⁷. Stopher et Jones (2003), précisent que l'approche par activité fournit des informations plus justes et doit être mise en oeuvre même si l'objectif de l'enquête est le recueil de données concernant les déplacements. Nous avons élaboré le questionnaire web de l'enquête ménages déplacements de Lyon en nous basant notamment sur des expériences australienne⁶⁸ et autrichienne⁶⁹.

⁶⁷Sous-estimation des déplacements (environ 20 à 30%), en particulier les petits déplacements, lors d'un recueil à l'aide d'un journal de déplacements (Wolf *et al.*, 2003; Pendyala, 2003).

⁶⁸Postal Survey, Two-Day Travel Diary, Department 'Infrastructure, Planning and Natural Resources', Institute of Transport Studies, University of Sydney, NSW 2006.

⁶⁹Postal Traffic survey, Household and individual questionnaires, Austria, 2002.

d'enquêtes auto-administrées et sur l'acquis des travaux conduits sur les schémas d'activités depuis plus de trente ans (Jones *et al.*, 1980; Bonnel, 1985, 1995; Stopher, 1998). Le cas de l'enquête postale réalisée par l'université de Sydney illustre cette approche. Un cahier a été envoyé aux répondants, pour qu'ils notent durant deux jours l'ensemble de leurs déplacements, en raisonnant en termes d'activités. L'idée sous-jacente est que les déplacements ne sont pas une fin en soi, mais un moyen de réaliser une activité (loisir, travail, accompagnement ...). Toutefois, cette approche est en partie en rupture avec celle qui est utilisée en France dans les enquêtes ménages déplacements en face à face, où le recueil porte sur l'enchaînement des déplacements. Afin de pouvoir comparer les résultats des deux enquêtes, nous avons retenu une solution intermédiaire, avec une entrée activité et une saisie de l'enchaînement des déplacements. Pour simplifier le travail du répondant, nous lui proposons un cheminement dans ses déplacements quotidiens, à l'aide d'une série de questions (horaires, lieux, activités, moyens de transport ...). A posteriori, l'analyse des résultats nous conduit à remettre en cause ce choix. L'oubli de certains déplacements, notamment le retour au domicile en fin de journée, par quelques internautes aurait probablement pu être évité par une approche centrée sur les pratiques d'activités, les déplacements découlant des changements de localisation.

II.3.2 Le déplacement, une notion difficile à appréhender

Au début de la partie sur les déplacements, nous indiquons sous forme de liste les consignes les plus importantes pour la saisie des données. L'objectif n'est pas d'être exhaustif, mais de valoriser les réponses de l'internaute, d'insister sur la richesse et la précision des données fournies, et d'expliquer le moyen d'obtenir rapidement de l'aide. Puis un exemple de cycle de déplacements est donné, à titre purement illustratif. Malgré le risque de suggérer des réponses, sa présence rend compréhensible le questionnaire, prend peu de place (1 page) et exploite les possibilités multimédia du web. Nous avons développé un exemple, en prenant garde à varier les modes et les motifs de déplacement, afin de ne pas influencer l'enquêté (figure 7).

Concrètement, le répondant doit renseigner précisément l'endroit où il se trouve au début de la période de référence, ainsi que l'activité effectuée. Ensuite, les questions suivent le cheminement suivant, jusqu'à ce que l'heure de départ du dernier déplacement soit postérieure à l'heure de fin de la période d'enquête (figure 8)⁷⁰.

D'un côté, Internet est un média qui nécessite des questionnaires concis. De l'autre la forme auto-administrée requiert des instructions suffisamment détaillées, aucun enquêteur ne venant apporter de précision au répondant ou organiser la saisie des réponses. La notion de 'déplacement', au cœur de l'étude n'est pas facilement comprise et assimilée par les répondants, et les résultats des enquêtes de mobilité ne peuvent pas être plus précis que la mémoire des

⁷⁰S'agissant de recueillir des déplacements, les heures de début et les heures de fin doivent être renseignées. Si nous avions opté pour un recueil basé sur les activités, seule l'heure de début aurait été demandée, l'heure de fin d'une activité étant par définition égale à l'heure de début de l'activité suivante.

La mise en place d'un module web dans l'EMD de Lyon

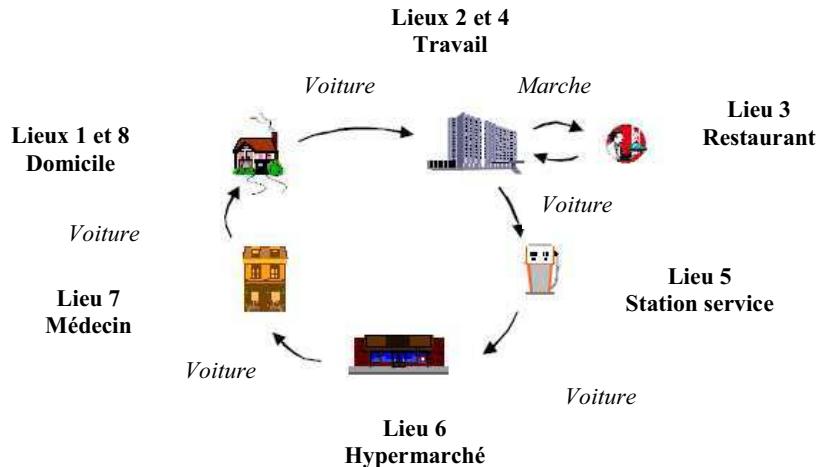


FIG. 7 – Exemple d'un cycle de déplacements quotidiens
Source : Questionnaire web (2005)

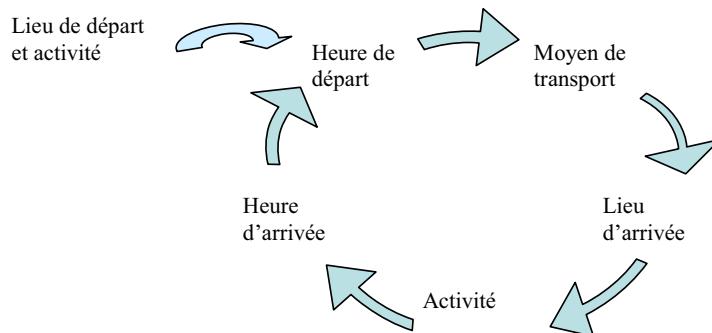


FIG. 8 – Cycle des questions relatives à un déplacement
Source : Questionnaire web (2005)

répondants. Plus les enquêtés vont pouvoir être 'aidés' dans leur effort de mémoire, plus juste vont être les réponses (Zmud, 2003). Pour permettre le recueil de données de qualité et éviter les biais de mesure dus à des interprétations différentes de ce concept, les enquêteurs renseignent les individus à l'aide de leur manuel d'instruction lors des enquêtes en face à face. Or, le caractère auto-administré d'une enquête web ne permet pas cet échange. Il est toujours possible de fournir la définition du déplacement et de l'illustrer à travers quelques exemples. Certaines enquêtes soulignent les termes pouvant prêter à confusion dans le libellé des questions et y associent une définition à l'aide d'un lien hypertext. L'internaute n'a alors qu'à cliquer pour voir s'afficher des précisions à l'écran⁷¹. Mais il n'est pas possible de fournir une aide en ligne regroupant tous les cas de figure prévus dans le manuel de l'enquêteur,

⁷¹Conrad *et al.* (2003) montrent que les répondants en ligne tirent relativement peu profit de cette possibilité, surtout si la démarche nécessite plus d'un 'clic' ou si les précisions concernent des termes courants (exemple : ménage, résidence ...). L'option est davantage utilisée lorsque les informations semblent utiles aux répondants, et les données recueillies sont de meilleure qualité.

au risque de complexifier le questionnaire, ce qui générerait l'effet inverse de celui souhaité.

La définition du déplacement est loin d'être évidente "*Il n'est donc pas si simple de donner de cette unité de mesure de la mobilité qu'est le déplacement une définition opératoire. C'est en réalité à travers les instructions qui sont données aux enquêteurs que se précise dans ses détails cette définition* (Bonnafous et Puel, 1983)". Dans les enquêtes en face à face, les enquêteurs disposent de la définition suivante : 'Le déplacement est le mouvement d'une personne, effectué pour un certain motif, sur une voie publique, entre une origine et une destination, selon une heure de départ et une heure d'arrivée, à l'aide d'un ou plusieurs moyens de transport' (Cete, 2005) ⁷². Dans l'enquête web, l'assimilation de la notion de 'déplacement' par l'internaute est difficile à contrôler et il n'est pas certain que l'aide en ligne ⁷³ soit comprise et prise en compte.

II.3.3 Le choix du mode et de l'activité

Nous avons jugé préférable de simplifier la liste des activités proposées dans le questionnaire en face à face, en effectuant d'abord des regroupements par thèmes, puis en demandant à l'enquêté des précisions dans des sous-menus pour les modalités études, achats et loisirs (figure 9). En ce qui concerne les tournées professionnelles, une note précise à l'internaute les informations à fournir.

Quelle était votre activité principale à cet endroit ? ▼		
Travail		
Etudes		
Achats		
Démarches personnelles (administratif, santé, recherche d'emploi...)		
Loisirs ou visites		
Restauration hors du domicile		
Aller chercher / déposer quelqu'un		
Tournée professionnelle		
Autre <i>Precisez</i>		

Si études ▼	Si achats ▼	Si loisirs ou visites ▼
Collège	Grand magasins et galeries marchandes	Activités sportives, culturelles ou associatives
Lycée	Hypermarchés et supermarchés	Promenade, « lèche-vitrines »
Universités et grandes écoles	Petits et moyens commerces, marchés	Visite à des parents ou à des amis

FIG. 9 – Choix de l'activité exercée
Source : Questionnaire web (2005)

⁷²En France, l'organisme central, le CERTU, assure l'homogénéité technique des enquêtes ménages.

⁷³Illustration de la définition donnée par le CERTU dans le guide méthodologique (Certu, 1998) à l'aide d'exemple.

La liste des modes de transport a été simplifiée et limitée aux modes les plus fréquemment utilisés pour les déplacements urbains⁷⁴. Une modalité autre est également nécessaire, afin de recueillir les modes plus rarement utilisés. Par ailleurs, dans la version face à face, on distingue les transports collectifs selon l'autorité organisatrice. Il n'est pas raisonnable de conserver cette typologie dans l'enquête web, puisque les internautes ne disposent pas toujours de l'information, qui peut par ailleurs être reconstruite a posteriori à partir de la localisation de l'origine et de la destination du déplacement effectué. Au final, nous avons retenu seulement 13 modes de transport, principalement en effectuant des regroupements⁷⁵ ou des suppressions⁷⁶ dans la liste des modalités proposées en face-à-face. Nous avons conservé la question relative au paiement du stationnement pour les individus déclarant un déplacement en voiture particulière, cette contrainte ayant généralement une influence sur le choix du mode de transport.

II.3.4 La chronologie des déplacements

Les enquêtes ménages déplacements commencent généralement entre 3h et 4h du matin, et se terminent à la même heure un ou plusieurs jours après (Stopher *et al.*, 2004b). A cette heure, le niveau de trafic est minimal sur les principaux axes routiers⁷⁷. Peu de déplacements seront interrompus si on choisit ce créneau pour borner la période de référence et les individus ne se retrouveront pas dans la situation délicate d'essayer de saisir un déplacement ayant commencé ou pris fin en dehors de cette tranche horaire.

L'usage est de recueillir les heures de départ et d'arrivée des déplacements sous la forme 0h/24h, afin de permettre un calcul rapide de la durée des activités (Pratt, 2003). Cette codification pose cependant des problèmes pour les déplacements dont l'aller se fait avant minuit et le retour après minuit, puisque la durée d'activité calculée est négative⁷⁸. Dans l'enquête web, nous avons demandé aux répondants de remplir deux champs, un pour le jour et un pour l'heure du déplacement (plus un champ pour les minutes). Ainsi, le début et la fin des déplacements se situent soit le jour 1 (jour de référence de l'étude) entre 04h et 24h00, soit le jour 2 (jour de fin de l'étude) entre 0h00 et 3h59. Quels que soient les horaires de début et de fin des déplacements se situant dans la période de référence, la durée calculée reste donc positive.

⁷⁴Dans la version face à face, l'avion fait partie de la liste des modalités, mais la codification est effectuée par l'enquêteur. En revanche dans une version auto-administrée, proposer explicitement l'avion comme mode de transport pour des déplacements urbains peut sembler fantaisiste (même si la modalité est tout à fait possible, puisque les déplacements ayant leur origine ou leur destination dans le périmètre d'enquête sont supposés être tous saisis).

⁷⁵Car interurbain, transport scolaire ou transport employeur.

⁷⁶Les rollers, skates et trottinettes, qui ne peuvent être assimilés aux bicyclettes, mais aussi le transport fluvial ou maritime, aérien.

⁷⁷Ce constat se vérifie en France, aux Etats-Unis et en Australie, mais peut varier entre minuit et 05h00 selon les pays.

⁷⁸Par exemple, un déplacement de 22h00 à 02h00 durerait -20 heures.

III Administration du questionnaire en ligne

III.1 L'importance de la prise de contact

Le premier contact entre l'étude et le répondant est très important, puisqu'il va déterminer en grande partie son choix de se soumettre ou non au questionnaire (Ampt, 2003). Généralement, une lettre avis est envoyée aux ménages de l'échantillon, et les recherches ont mis en évidence une augmentation du taux de réponse aux enquêtes web, grâce à ce type de pré-notification (Alsnih, 2004; Mehta et Sivadas, 1995). Cette lettre permet d'informer les individus de l'enquête à venir, de la légitimer en insistant sur l'objectif poursuivi et les instances impliquées⁷⁹, de rassurer sur le caractère confidentiel du traitement des données, de fournir une estimation du temps nécessaire pour remplir le questionnaire (20 minutes maximum), ainsi que les coordonnées d'un contact (une adresse e-mail ou un numéro de téléphone gratuit, si possible) pour les éventuelles questions⁸⁰. Dans les enquêtes web, cette lettre avis est fortement recommandée (Sills et Song, 2002), bien qu'elle soit souvent envoyée par mail lorsque l'adresse électronique des individus est connue. Contacter les individus à de multiples reprises par mail est ni coûteux ni chronophage, mais lorsque l'adresse électronique est inconnue, les méthodes classiques de prise de contact par téléphone ou voie postale restent la seule solution. La lettre mentionne précisément l'adresse du site où est hébergé le questionnaire, l'identifiant et le mot de passe attribué, et insiste sur le caractère simple et rapide de l'enquête (Annexe V). L'option web peut être mise en avant, même s'il est préférable de réduire l'anticipation de la difficulté de remplissage du questionnaire, plutôt que de développer les atouts du web (Haraldsen *et al.*, 2002).

Les individus supportent des coûts pour répondre aux enquêtes (temps, concentration, frais de connexion ...). La théorie de 'l'échange', avancée par Sudman et Bradburn (1996) montre que ces derniers sont d'autant plus motivés à se soumettre au questionnaire que le retour espéré est supérieur aux efforts engagés. (Dillman, 1978) précise que ce retour peut être de nature intangible, comme la satisfaction de faire partie d'un groupe sélectionné ou le sentiment que les réponses fournies contribuent largement au succès de l'étude. Un moyen d'augmenter le seuil de tolérance aux enquêtes est donc d'insister sur l'importance de l'étude et sur la nécessaire participation de tous pour garantir la qualité des résultats. Les techniques de personnalisation, qui donnent de l'importance au répondant vont dans ce sens. Il est également conseillé de communiquer sur le caractère ponctuel de l'étude et de fixer une date limite de présence du questionnaire en ligne⁸¹. Enfin, Alsnih (2004) recommande de

⁷⁹Les personnes sollicitées répondent plus facilement lorsque la demande émane d'une autorité publique (Zmud, 2003).

⁸⁰Une étude menée par le bureau du Census des Etats-Unis montre que la moitié des répondants ont contacté l'assistance pour obtenir de l'aide sur l'utilisation des identifiants et mots de passe ou pour résoudre des problèmes de configuration de leur navigateur (Nichols et Sedivi, 1998).

⁸¹Stopher *et al.* (2004b) rajoute que la publicité autour de l'étude (télévision, radio, presse...) motive la coopération des individus, notamment pour les enquêtes en face-à-face.

préciser qu'aucune contribution monétaire ne sera demandée, pour distinguer l'étude des démarches de télémarketing.

Dans certains protocoles d'enquête, le choix du mode de collecte est donné aux individus. Cette option est valorisante, ce qui augmente le taux de réponse et génère des questionnaires de meilleure qualité (Alsnih, 2004). En effet, les individus se sentent plus ou moins proches d'un mode particulier, et ne peuvent parfois pas être touchés par certains médias. Pour l'Enquête Ménages de Lyon, ce choix n'a malheureusement pas été possible. Il s'agit d'une enquête séquentielle, où les non-répondants au premier mode doivent être informés du nouveau média disponible par une nouvelle lettre avis (Kalfs et Van Evert, 2003). Les enquêtes transport comprennent généralement une phase de recrutement et une phase d'interview et des protocoles sophistiqués combinent des contacts off-ligne et en ligne. Au moment du repérage, les enquêteurs essayent d'obtenir un numéro de téléphone, utilisable pour les relances après l'envoi du premier courrier (si aucun numéro n'est disponible, les relances sont postales). Les individus refusant de recevoir un enquêteur à domicile se voient proposer oralement, puis par écrit, de répondre au questionnaire web (annexe V). Une lettre est également envoyée aux ménages qu'il est impossible de joindre par téléphone, après 8 tentatives à des jours et horaires différents (ménages déclarés non-répondants). Ces lettres reprennent l'argumentaire développé dans la première lettre avis (présentation de l'étude, choix aléatoire des ménages, importance des réponses, confidentialité des données ...) et insistent sur le web, nouveau média disponible qui satisfait davantage aux contraintes temporelles des ménages. Pour encourager les individus, la durée restreinte de l'enquête web (15 mn) et un numéro de téléphone permettant d'obtenir des informations complémentaires sont indiqués dans les courriers. La sélection du répondant sur le web parmi les membres du ménage s'opère grâce à l'âge (plus de 11 ans).

III.2 La codification des adresses

Les enquêtes ménages déplacements recueillent des données sur la localisation des origines et des destinations des déplacements. De façon générale, les informations géographiques obtenues sont cruciales, car la modélisation de la demande de transport évolue vers des approches complexes, telles que la microsimulation. Par ailleurs, le zonage peut changer entre deux enquêtes ménages déplacements, et l'identification précise des adresses permet de réaliser des analyses sur de nouveaux périmètres. Avant de lancer une étude, il est donc nécessaire de s'interroger en amont sur la précision des données que l'on souhaite obtenir.

III.2.1 Le recueil de données précises

Le géocodage est une étape clé du traitement des données de mobilité. Cette méthode consiste à traduire les informations saisies en données directement utilisables par les modèles de planification stratégique. Il s'agit d'affecter une zone géographique à chaque localisation fournie par le répondant, selon

une cartographie très précise (Cambridge Systematics, 1996). Dans la plupart des enquêtes, le géocodage est effectué manuellement, ce qui implique un lourd travail, une bonne connaissance du périmètre d'enquête de la part des chargés d'études et conduit inévitablement à des erreurs (Greaves, 1998, 2003). Pour simplifier la saisie des adresses et le traitement de l'information lors de la réalisation d'une enquête web, il est possible de s'appuyer sur les systèmes d'information géographique et les nouvelles technologies de l'information et de la communication. Des fonctions déployables aisément à l'aide de menus déroulants (nom de rue, de ville ...) ou de cartes peuvent être envisagées. Des bases de données interactives permettent de codifier en temps réel les lieux d'origine et de destination des déplacements, en déchargeant une large part du travail sur l'enquêté. Selon l'expérience de Adler *et al.* (2002), proposer un géocodage en direct sur le web permet d'obtenir un taux de géocodage de 100% (au final, l'individu clique sur une carte, mais renseigne la localisation). Ceci permet aux répondants peu familiers des ordinateurs de renseigner facilement les adresses des lieux visités⁸², sans recourir à l'assistance technique (Adler *et al.*, 2001). Dans une enquête déplacements menée au Michigan en 2001, les distances entre deux adresses données par les répondants étaient systématiquement calculées. Lorsque la durée du déplacement saisie ne correspondait pas à une fourchette de valeur acceptable au regard de la distance théoriquement parcourue, une alerte permettait au répondant de corriger sa réponse (Murakami, 2004). D'autres études font appel à des logiciels qui interagissent directement avec les répondants, suggérant des alternatives lorsque l'adresse saisie n'est pas reconnue (Ressource System Group, 2002). Ces outils, bien que complexes, facilitent l'effort de mémoire et de saisie des enquêtés, et permet de réduire la non-réponse partielle et les erreurs de géocodage. Mais l'usage de grosses bases de données sur le web est à éviter, à cause du risque de blocage du serveur et du temps de téléchargement. De plus, si l'utilisation d'une cartographie est relativement lourde au niveau de l'affichage des pages web, cette méthode suppose également une bonne capacité des internautes à se situer sur une carte (Alsnih, 2005; Wermuth *et al.*, 2001).

Pour l'enquête Ménages Déplacements de 2006, le périmètre du SCOT de Lyon a été divisé en 856 zones géographiques appelées zones fines. Lors de l'interview en face-à-face, les localisations des activités sont directement codées par l'enquêteur. Il est cependant difficile de mettre en place ce zonage sur le web, et de demander à l'internaute de choisir la zone correspondant à l'origine et à la destination de ses déplacements. L'abandon du zonage amène au recueil de données réelles, plus précises, le répondant saisissant l'adresse exacte des lieux visités. Ces renseignements sont ensuite transformés en coordonnées longitudinales et latitudinales, conformément à l'adresse générée par le logiciel de géobornage, qui recherche l'adresse officielle se rapprochant le plus de celle donnée par l'enquêté⁸³. Cette méthode autorise ensuite toute modification du

⁸²Se référer à Ressource System Group (2002); Adler *et al.* (2001) pour les résultats des méthodes utilisées par les répondants pour saisir les adresses des lieux fréquentés.

⁸³Le référentiel utilisé est le Lambert II étendu.

zonage du territoire (Adler *et al.*, 2001).

Dans certains cas, les internautes sont dans l'incapacité de saisir certaines adresses, soit parce que leur mémoire leur fait défaut, soit parce qu'il ne veulent pas révéler l'information (par exemple, une personne n'a pas renseigné l'adresse du domicile, élément par définition parfaitement connu). De plus, si le lieu d'origine ou de destination d'une activité se situe en dehors du périmètre de référence, l'adresse saisie est souvent très peu précise, car il s'agit d'un lieu inhabituel pour le répondant. Nous avons choisi de demander aux individus de saisir les adresses en clair, puis de les recoder *a posteriori* (géo-bornage) à l'aide de dictionnaires d'adresses. Depuis une quinzaine d'années, il est d'usage de raisonner en localisation fonctionnelle. Les lieux n'ont donc plus seulement une adresse, mais un nom, une fonction (enseigne, nom d'établissement ...) Pour permettre aux répondants web de situer le mieux possible la destination de leurs déplacements, une zone de saisie libre est proposée. Ils peuvent y indiquer notamment des noms de points d'intérêt (les gares, les universités, les écoles, les centres hospitaliers, les mairies ...) ou des intersections de rues. Une base concernant les points d'intérêt sera utilisée en deuxième recours, si le premier codage ne permet pas de définir des coordonnées précises.

III.2.2 Le géobornage : des résultats mitigés

Suite à l'administration d'une enquête à 600 individus, Adler *et al.* (2001) ont trouvé que près de 80% des adresses du domicile étaient parfaitement renseignées, les autres 20% préférant situer le lieu sur une carte afin que leur adresse précise ne figure pas dans la base de données. Environ la moitié des lieux de travail étaient identifiés à l'aide d'un nom de rue et la majorité des autres lieux (activités hors du domicile) à l'aide du nom d'établissement ou de points d'intérêts. Stopher *et al.* (2004b) recommandent que dans une enquête de qualité, 100% des adresses du domicile des répondants, 95% des adresses de leur lieu de travail habituel ou d'études et 90% des autres localisations soient correctement géocodées. Il existe plusieurs niveaux de géocodage, et le SCOT de Lyon a été divisé en 859 zones fines, qu'il est possible de regrouper en 158 zones agrégées. Ainsi, lorsque les adresses saisies ne sont pas assez précises pour qu'une zone fine soit affectée (exemple : seule la ville est renseignée), alors on fait correspondre une zone agrégée. En face-à-face, l'enquêteur peut montrer des cartes aux enquêtés, afin de les aider à situer les lieux fréquentés. Mais bien que les technologies disponibles autorisent la visualisation de cartes sur internet, les contraintes de l'expérimentation et les outils mis à disposition par Conversoft ne l'ont pas permis.

Les résultats du géobornage des adresses sont présentés dans le tableau 7. Il en ressort que l'adresse du domicile est correctement restituée. 64% des répondants indiquent les coordonnées précises de leur lieu de résidence, et 29% se contentent de livrer le nom de la rue. En ce qui concerne le lieu de travail ou d'études habituel, le niveau de précision est un peu plus faible (44% d'adresses exactes et 50% de coordonnées relatives au nom de rue), ces données étant probablement jugées plus confidentielles par les répondants. Les adresses des autres lieux fréquentés (achats, loisirs, accompagnement ...) sont moins bien

retranscrites, bien que 38% soient exactes et 43% identifiés grâce au nom de la rue. Le géobornage des adresses est donc possible a posteriori, et conduit à des données de bonne qualité. Le temps de traitement n'est pas négligeable, et il est nécessaire de desserrer progressivement les contraintes du logiciel de géobornage pour arriver à identifier une adresse précise en fonction de la qualité des renseignements saisis.

Géobornage	Lieu du domicile	Lieu de travail	Autres lieux
Numéro exact	58%	39%	38%
Numéro le plus proche	6%	5%	4%
Centroïde rue	29%	51%	43%
Centroïde arrondissement	0%	1%	1%
Ville	7%	2%	13%
Non renseigné	0%	2%	1%
Total	369	282	1075

TAB. 7 – Qualité de géobornage des adresses saisies par les répondants
Source : EMD web - Lyon 2006

III.2.3 Un calcul des durées et distances parcourues proche de la réalité

Le géobornage permet d'obtenir une adresse exacte pour chaque lieu fréquenté et donc des coordonnées longitudinales et lattitudinales. La distance de chaque déplacement est ensuite estimée à vol d'oiseau, à partir des coordonnées de l'origine ($x_{\text{départ}}, y_{\text{départ}}$) et de la destination ($x_{\text{arrivée}}, y_{\text{arrivée}}$) du déplacement, par la formule suivante :

$$Distance = \sqrt{[(y_{\text{arrivée}} - y_{\text{départ}})^2 + (x_{\text{arrivée}} - x_{\text{départ}})^2]} \quad (2)$$

La distance est fournie en mètre. Si l'adresse exacte des lieux visités n'est pas renseignée, alors la distance à vol d'oiseau en mètres entre les centroïdes des zones d'origine et de destination des déplacements est calculée. Au final, les distances calculées des déplacements recueillis dans le cadre de l'enquête web sont de bonne qualité, comme le montre le tableau 8.

Calcul de distance	Effectif	Pourcentage
Vol d'oiseau	1047	95%
Vol oiseau centroïde de zones	59	5%
Total	1108	100%

TAB. 8 – Précision du calcul de distance
Source : EMD web - Lyon 2006

Dans 95% des cas, la distance vol d'oiseau calculée des déplacements internes au périmètre d'enquête est précise, grâce au géobornage.

III.3 La validation du questionnaire

Nous avons montré dans la section I.1.6, qu'il était préférable de prévoir, des validations intermédiaires du questionnaire, afin de permettre un stockage des réponses. Dans l'enquête web, nous sollicitons quatre types de validation : une après chaque bloc ménage et individu, une après chaque déplacement, puis une validation finale. Les questions relatives au ménage et à l'individu sont peu nombreuses et rapides. En revanche, la saisie d'un déplacement est relativement longue, surtout lorsqu'il faut renseigner des adresses différentes de celle du lieu de domicile ou du lieu de travail, et demande un effort de mémoire plus important. Afin d'éviter un lourd travail de ressaisie en cas d'interruption de la connexion, une validation est requise à la fin de chaque déplacement, et les données sont sauvegardées dans la base de l'administrateur. Ces validations intermédiaires réduisent également le niveau de fatigue de l'enquêté, qui en cas d'erreur ou d'oubli peut revenir au déplacement concerné pour le corriger (Ressource System Group, 2002).

Il est impossible d'anticiper le moment de connexion des répondants, et donc le jour de référence utilisé pour la saisie des déplacements. A l'écran, le jour de référence apparaît automatiquement et correspond au dernier jour de la semaine avant le jour de la première connexion du répondant. Cette méthode pose deux problèmes. Si beaucoup de personnes se connectent durant le week-end, les déplacements du vendredi seront sur représentés et l'effet mémoire sera plus important pour les internautes qui répondent le dimanche ou le lundi. Par ailleurs, il n'est pas évident de se rappeler tous les déplacements effectués la veille ou l'avant veille, selon leur ordre chronologique, et il est fort probable que des répondants souhaitent modifier, ajouter ou supprimer des déplacements. Proposer un résumé des informations saisies sous forme graphique ou textuelle a l'effet d'une récompense pour certains répondants (Adler, 2003). Un schéma dynamique qui s'agrémente au fur et à mesure des réponses saisies permet de visualiser le cheminement des déplacements, mais pose des contraintes de faisabilité liées aux étapes de validation des données et au langage retenu par l'administrateur. Ce dernier a recommandé la production d'un tableau synthétique des réponses saisies à la fin du bloc 'déplacements', pour garantir une bonne présentation des données, quel que soit le type de matériel utilisé et le nombre de déplacements saisis. Il suffit alors à l'internaute de vérifier ces informations, puis de modifier les données ou d'insérer un nouveau déplacement. La suppression étant irréversible, une confirmation est demandée à l'enquêté lors de l'activation de cette commande, pour éviter des erreurs de manipulation pouvant conduire au découragement et à l'abandon du questionnaire. Une validation générale est également prévue à la fin du bloc 'déplacements', après que l'internaute ait vérifié l'ensemble de ses réponses.

III.4 L'imputation des données

III.4.1 Les valeurs aberrantes

Avant le traitement des données, il est nécessaire d'effectuer quelques tests pour vérifier la cohérence des réponses. Certains tests sont réalisés automatiquement, mais d'autres n'ont pas été prévus dans le script du questionnaire web. Par exemple, l'heure de début d'un déplacement doit être supérieure à l'heure de fin du déplacement précédent, mais les répondants ont la possibilité de rajouter deux déplacements 'oubliés' à la fin du questionnaire, sans que leur cohérence ne soit vérifiée (le logiciel CATI ne permet pas d'effectuer ce test).

La durée des déplacements est également un indice de la qualité des données. Une mauvaise compréhension de la définition d'un déplacement peut conduire à inclure la durée de l'activité dans celle du déplacement, et donc générer des durées aberrantes. Enfin, un nombre non négligeable d'individus ne retournent pas à leur domicile à la fin de la période d'enquête. Il semble que le dernier retour au domicile soit un déplacement fréquemment omis, car considéré comme peu important, ou par lassitude du répondant.

III.4.2 Les valeurs manquantes

La non-réponse partielle se définit comme l'impossibilité d'obtenir des données réelles et complètes de l'ensemble des répondants (Zmud et Arce, 2000). Cette non-réponse vient souvent grever la qualité de l'étude, et peut être minimisée en optimisant la construction et l'administration du questionnaire. Sur le web, il est fréquent que des alertes encouragent les répondants à répondre et dans notre étude, les réponses sont forcées. En effet, les répondants doivent impérativement sélectionner une modalité de réponse à chaque question, pour pouvoir accéder à la question suivante et donc progresser dans le questionnaire. Les répondants ont également la possibilité de saisir leur adresse e-mail à la fin du questionnaire, mais pour des raisons de délais (les résultats de l'enquête web ont été disponibles fin juillet, soit 3 mois après son lancement), nous n'avons pas utilisé cette information pour les relancer. Malgré les nombreuses alertes mises en place tout au long du questionnaire, les premiers tris effectués sur la base de données laissent apparaître un certain nombre de valeurs manquantes, qui ne permettent pas une exploitation satisfaisante des données. Ces valeurs manquantes s'expliquent par un refus du répondant de communiquer l'information (l'adresse de son domicile), un manque de connaissance ne lui permettant pas de le faire (l'adresse précise d'un magasin) ou une question filtrée en amont du questionnaire (la possession du permis de conduire pour les moins de 16 ans).

III.4.3 Les techniques d'imputation

Si les données manquantes peuvent être imputées suite à des analyses précises, il est toujours possible d'introduire des biais. En effet, le chargé d'étude peut être tenté d'effectuer des imputations 'logiques', alors que le comportement du répondant est en réalité inconsistante avec son profil ou ses déclarations

antérieures (Adler, 2003). Les procédures d'imputation permettent cependant de construire une information probable à partir d'une information incomplète, plutôt que de la rejeter en totalité (Bonnel et Le Nir, 1998). Nous avons imputé certains résultats de l'enquête web, bien que les données individuelles soient par définition difficiles à prédire par la plupart des méthodes d'imputation courantes (Stopher et Jones, 2003).⁸⁴.

Dans la plupart des cas, la durée des déplacements peut-être calculée à l'aide des heures et des minutes de début et de fin. Pour les déplacements internes au périmètre d'enquête pour lesquels la durée est inconnue ou aberrante, nous imputons la valeur à partir de la distance du déplacement⁸⁵. Pour cela, nous retenons la vitesse moyenne des déplacements réalisés avec le même mode de transport, en considérant seulement les déplacements dont la durée et la distance sont inférieures ou égales à la valeur du 90è centile de durée et de distance, afin de tenir compte de certaines valeurs probablement aberrantes. Au final, il y a 63 déplacements enquêtés pour lesquels la durée a du être imputée. Les vitesses moyennes de référence par mode figurent dans le tableau 9.

Mode	Minimum	Maximum	Moyenne
Deux-roues	12,16	33,31	17,73
Marche à pied	0,2	9,4	2,90
Transport en commun	0,24	20,94	6,03
Vélo	3,03	16,55	8,08
Voiture conducteur	1,12	66,18	14,46
Voiture conducteur + TC	4,34	22,64	12,10
Voiture passager	3,32	55,17	14,95
Voiture passager + TC	4,54	4,54	4,54

TAB. 9 – Vitesses moyennes de référence par mode en km

Source : EMD web - Lyon 2006

IV Conclusion

Les enquêtes en ligne sont parfois risquées, mais une méthodologie adaptée aux objectifs et à la population étudiée peut générer une étude moins coûteuse, plus rapide et de bonne qualité. A l'issue de cette expérience, nous pouvons tirer certains enseignements relatifs à la conduite d'une enquête ménages déplacements sur le web.

La construction du questionnaire est un facteur clé de succès d'une enquête web. Selon Dillman *et al.* (1998b), il faut considérer les limites technologiques

⁸⁴Les méthodes utilisant les régressions multivariées et la maximisation de la vraisemblance sont statistiquement performantes et permettent d'utiliser les données d'enquêtes à des fins de modélisation (Acock, 1997).

⁸⁵L'imputation n'a pas été faite pour les déplacements ayant au moins une extrémité hors du périmètre d'enquête.

du web, ainsi que celles des utilisateurs, et comprendre comment les répondants s'attendent à ce que l'enquête se déroule. Dans les protocoles d'enquête mixtes, il est illusoire de chercher une standardisation quel que soit le mode utilisé, pour garantir la comparabilité des données, et préférable d'harmoniser la perception et la compréhension du questionnaire, plutôt que de le reproduire à l'identique. Nous avons suivi cette recommandation dans la conception du questionnaire web de Lyon, mais nous n'avons pas été assez loin en ce qui concerne le recueil des déplacements. L'analyse des résultats montre qu'il faut abandonner la logique déplacements et privilégier une logique activité pour le recueil des données de mobilité, les déplacements apparaissant alors 'naturellement' à chaque changement de localisation.

Par ailleurs, le recueil des localisations est un élément clé des enquêtes déplacements. Il est important de pouvoir localiser précisément les déplacements et de produire des données comparables à celles de l'enquête en face à face et compatibles avec le zonage du territoire. Si la mise en place d'un géocodage en ligne n'est pas envisageable (coût des bases de données, contraintes techniques...), il est indispensable de trouver d'autres moyens de recueillir les données. L'expérience de l'enquête web montre qu'il est possible de recueillir des adresses et de les géoréfléchir a posteriori. Le traitement de l'information a toutefois nécessité un lourd travail, les adresses saisies étant parfois trop imprécises pour autoriser un traitement automatique. Il est donc souhaitable de tester une validation en temps réel de l'information.

Enfin, une enquête web est différente des autres enquêtes assistées par ordinateur. Il n'est donc pas judicieux d'utiliser des logiciels conçus pour des enquêtes CATI ou CAPI (logiciel d'aide à la réalisation d'enquêtes téléphonique ou en face à face) lors de la réalisation d'une enquête web, et préférable de recourir à un logiciel CAWI (logiciel d'aide à la réalisation d'enquêtes web). Ces logiciels proposent une ergonomie du questionnaire aussi proche que possible des environnements que les internautes ont l'habitude de trouver sur la toile et permettent d'enregistrer les réponses dans des bases de données fonctionnelles et performantes. La tâche est d'autant plus délicate que les chercheurs en méthodes d'enquête ne sont pas familiers des enquêtes de déplacements, les planificateurs de transport ne connaissent pas les problèmes relatifs aux enquêtes et les ingénieurs et concepteurs d'enquêtes ne connaissent ni les problèmes de la recherche ni ceux de la politique des transports (Adler, 2003).

Les résultats de notre étude permettent de formuler des recommandations précises concernant la conduite d'enquête de mobilité sur le web. Le caractère exploratoire de cette recherche et les délais n'ont cependant pas permis de définir un cahier des charges suffisamment précis, à l'image de ce qui est disponible pour la réalisation des enquêtes en face à face.

Chapitre 3 : Premiers résultats sur le comportement de réponse des internautes

"Les outils nouveaux (le téléphone portable, les prothèses qu'on a créées, l'ordinateur, l'Internet, la mémoire de Google) nous changent, on est déjà des mutants."

Joël de Rosnay (2007).

Lors d'une enquête auto-administrée, l'enquêté se retrouve seul face au questionnaire. Il est donc impossible d'obtenir des informations précises sur son comportement de réponse. La situation est différente dans le cas d'une enquête assistée par ordinateur, comme les enquêtes CAPI ou CAWI (web). L'interface informatique utilisée pour la saisie des réponses autorise en effet le recueil de données comportementales. Nous avons saisi cette opportunité, et disposons d'un certains nombre de résultats, qui permettent de comprendre la manière dont les répondants ont appréhendé le questionnaire et de tirer des enseignements pour la mise en ligne de prochaines enquêtes.

L'enquête web concerne les individus ayant refusé de répondre en face-à-face ou restant injoignables. Nous présentons les résultats globaux de l'enquête face-à-face, en termes de taux de réponse, afin de définir la cible des répondants potentiels sur le web (section I). Puis, nous analysons en détail le taux de réponse à l'enquête web, ainsi que les raisons qui ont poussé les individus refusant de répondre en face-à-face à remplir le questionnaire en ligne (section II).

Deux relances étaient prévues, pour motiver les individus à répondre en ligne. Il est intéressant d'analyser l'impact de ces relances sur le taux de réponse et la qualité des réponses obtenues, au regard d'expériences similaires détaillées dans la littérature (section III). Un des principaux objectifs de notre recherche consiste à diminuer le biais de non-réponse constaté dans les enquêtes auprès du grand public. Avec le web, il est possible d'analyser le profil des individus selon le degré de remplissage du questionnaire et d'identifier les questions qui génèrent de nombreux abandons. Par ailleurs, si les réponses sont forcées, certaines questions offrent une modalité 'Ne sait pas' ou 'Non-réponse', utile pour ceux qui ne veulent pas se prononcer. Ces modalités sont des échap-

I Un taux d'échec global à l'enquête en face-à-face non négligeable

patoires, et nous étudions dans quelle mesure elles pèsent sur la non-réponse partielle.

Enfin, le web est un média peu contraignant, puisque les individus ont la possibilité de se connecter plusieurs fois pour remplir le questionnaire, à tout moment de la journée. La connexion sécurisée, à l'aide d'identifiants et de mots de passe permet d'analyser l'heure, la durée et le nombre de connexions des internautes (section IV).

I Un taux d'échec global à l'enquête en face-à-face non négligeable

I.1 Le calcul du taux de réponse

Selon (Ettema *et al.*, 1996), le taux de réponse du questionnaire, ainsi que la distribution des données est un indicateur de la qualité d'une enquête. (Stopher *et al.*, 2004b) précisent qu'un taux de réponse élevé permet de réduire a priori l'incidence du biais de non-réponse. Le taux de réponse à une enquête peut être formulé comme suit (Groves, 1989) :

$$\text{Taux de réponse global} = \frac{I}{I + P + NC + R + NI} \quad (3)$$

Avec :

- I : le nombre d'interviews complètes ;
- P : le nombre d'interviews partielles ;
- NC : le nombre de ménages éligibles à l'étude, mais non joignables⁸⁶ ;
- R : le nombre de ménages éligibles à l'étude, mais refusant de répondre ;
- NI : le nombre d'autres ménages éligibles mais non-répondants (problèmes techniques, maladie ...)

Selon Ampt (1997), il est utile de distinguer l'échantillon brut, constitué de l'ensemble des ménages sélectionnés pour répondre à l'enquête, de l'échantillon net, obtenu en retirant les unités statistiques non pertinentes, sans affecter la qualité de l'enquête au regard de ses principaux objectifs. A contrario, les ménages qui ne peuvent être joints après plusieurs visites ou appels, ou dont le seul contact reste un répondeur viennent peser sur le taux de réponse, car leur absence de l'échantillon nuit à la qualité des données.

Dans l'enquête ménages déplacements réalisée en face-à-face, nous estimons un taux de réponse après contact :

$$\text{Taux de réponse} = \frac{\text{Nombre de ménages répondant à l'enquête}}{\text{Nombre de ménages contactés}} \quad (4)$$

⁸⁶Cette définition pose problème, car il est parfois difficile de faire la différence entre des ménages non éligibles et des ménages éligibles mais non joignables.

I.2 Estimation de la cible théorique de l'enquête web

Le taux d'échec global (1 - taux de réponse global) est calculé en additionnant le taux de refus et le taux de rebuts. Les refus concernent les ménages qui ont refusé de répondre à l'enquête en face-à-face. Les rebuts sont de plusieurs sortes :

- les ménages impossibles à joindre (jamais présents à leur domicile, ne veulent pas répondre au téléphone...);
- les ménages absents de longue durée (ne trouvent pas de moment pour répondre en famille pendant la durée de réalisation de l'enquête)⁸⁷ ;
- les ménages inéligibles (résidences secondaires, les logements vacants et les logements qui n'ont plus l'usage d'habitation) ;
- les autres cas de rebuts.

A la fin de la période d'enquête, 6 375 interviews ont été réalisées en face-à-face sur le SCOT de Lyon, pour un total de 11 951 ménages contactés, soit un taux de réponse global de 53%. Les résultats sont détaillés dans le tableau 10.

Statut	Effectifs	%
Refus	3 716	
	Taux de refus	31 %
Impossibles à joindre	1 531	
Absents de longue durée	44	
Autres cas de rebuts	301	
	Taux de rebuts	16 %
	Taux d'échec	47 %
Total sans autres rebuts	5 291	
Total avec autres rebuts	5 592	

TAB. 10 – Taux de réponse global enquête face-à-face sur le périmètre du SCOT de Lyon

Source : SOFRECO (suivi semaine 19 - 2006)

Seuls les ménages refusant de répondre à l'enquête standard, impossibles à joindre ou absents de longue durée ont été contactés par courrier pour répondre à l'enquête web. La cible théorique de l'enquête web est donc de 5 291 ménages. En réalité, seuls 4 335 ménages ont reçu un courrier les invitant à répondre sur le web. Les envois ont débuté avant la fin de l'enquête en face-à-face, et certains repérages complémentaires ont été réalisés sans qu'un double de la fiche adresse ne soit conservé par la société en charge de l'étude. On estime donc que 956 ménages non-répondants ou non joignables lors de l'enquête en face-à-face et pouvant potentiellement répondre sur le web n'ont pas été sollicités (18% de la cible théorique).

⁸⁷Les ménages sont considérés comme absents de longue durée, si l'enquêteur dispose d'informations explicites.

I.3 Des disparités spatiales

Le périmètre d'enquête (S.C.O.T. de Lyon) a été segmenté en 9 zones, selon la logique de couronnes et la segmentation Est-Ouest correspondant à la coupure du Rhône et à des différences sociologiques assez marquées. Nous observons que le taux d'échec global de l'enquête en face-à-face est inégalement réparti entre ces 9 zones (tableau 11 et figure 10), avec un pourcentage particulièrement élevé pour les zones 1 (Hypercentre), 3 (1ère couronne est), 4 (1ère couronne ouest) et 6 (2ème couronne ouest).

Pour les zones 1 (hypercentre), 4 (1ère couronne ouest) et 6 (2ème couronne ouest), c'est un fort taux de rebuts (21%, vs. un taux moyen de 16% sur l'ensemble du périmètre) qui fait chuter le taux de réponse à l'enquête en face-à-face. Pour la zone 3 (1ère couronne est), en revanche, c'est le nombre important de refus qui pèse sur le niveau de qualité de l'enquête (42% des personnes contactées refusent d'être interviewées). Deux communes enregistrent des résultats particulièrement mauvais : Bron et Vaulx-en-Velin (45%). Les interviews en face-à-face ont débuté à la fin de l'année 2005, lorsqu'est survenue la crise des banlieues, et les communes de Bron et Vaulx-en-Velin regroupent un certain nombre de quartiers dits 'sensibles'.

Zones	Total	Taux de refus	Taux de rebus	Taux d'échec global
Hypercentre	1 690	30 %	21 %	51 %
Lyon-Villeurbanne	3 249	29 %	16 %	45 %
1re couronne Est	1 535	42 %	10 %	53 %
1re couronne Ouest	1 851	30 %	21 %	51 %
2è couronne Est	1 148	33 %	8 %	41 %
2è couronne Ouest	890	28 %	21 %	49 %
3è couronne Est	1 091	29 %	10 %	38 %
Externe	497	23 %	16 %	39 %
Total	11 951	31 %	16 %	47%

TAB. 11 – Taux de réponse enquête face-à-face par zone sur le périmètre du SCOT de Lyon

Source : SOFRECO (suivi semaine 19 - 2006)

I.4 Les raisons de la non participation à l'enquête en face-à-face

Il est intéressant de connaître les raisons pour lesquelles des individus échappent à l'enquête standard, mais sont motivés pour participer à l'étude en ligne. A la fin du questionnaire web, les répondants avaient la possibilité de s'exprimer sur ce point. Six modalités sont proposées : l'attrait du web comme média de réponse, le refus des interrogés de recevoir un enquêteur à leur domicile, le manque de disponibilité des ménages sur les créneaux proposés, l'absence de temps pour répondre à l'étude, l'absence de contact avec un enquêteur ou d'éventuelles autres raisons. Les réponses pouvaient être multiples,

Premiers résultats sur le comportement de réponse des internautes

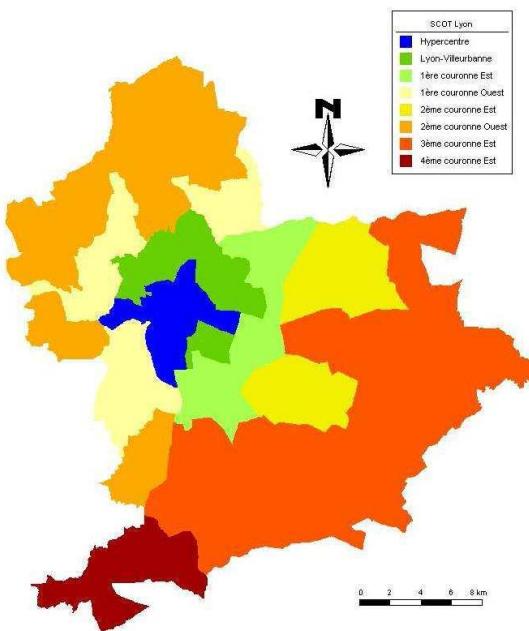


FIG. 10 – Les 9 zones du périmètre du SCOT de Lyon

Source : EMD Lyon 2006

et un champ textuel permettait aux répondants de saisir librement un commentaire. Au final, 365 personnes ont renseigné cette question et 33 ont ajouté un commentaire. Les résultats figurent dans le tableau 12.

L'engouement pour le web semble évident (*'c'est plus facile sur le web'*), même si certains enquêtés prétendent seulement avoir *'reçu un courrier me demandant une réponse par Internet'*. La mixité des modes d'enquêtes permet donc de toucher un plus grand nombre de personnes (*'vous ne m'avez pas laissé le choix. Je ne peux vous répondre que par le net'*). Les enquêtes à domicile sont en effet difficiles à réaliser (*'je n'ai pas envie d'être dérangé chez moi'*), et certaines personnes mettent en avant la *'surabondance de démarcheurs par téléphone'*. Le manque de disponibilité des répondants impacte fortement la non-réponse à l'enquête standard, qu'il s'agisse de contraintes professionnelles (*'les enquêteurs n'ont jamais réussi à me contacter compte tenu de mes horaires de travail et ont fini par m'adresser un courrier'*, *'quelqu'un est venu et je n'étais pas là. Etudiante, pas facile pour les horaires'*, *'probablement l'enquêteur n'a pas réussi à me joindre pour fixer RDV car souvent absente ou rentre tard le soir'*, *'entreprise ouverte du lundi au samedi inclus'*), de congés (*'en vacances lorsque l'enquêteur est passé'*) ou des exigences méthodologiques de l'étude (*'l'enquêteur proposait que les 4 personnes soient présentes !! mission impossible'*, *'...et en plus vous exigez que toute la famille soit présente au moment de l'enquête !!!'*, *'j'ai reçu un enquêteur à mon domicile mais j'étais seule et l'enquête n'a pas été retenue'*). Et même lorsque les personnes sont disponibles, elles sont parfois dans l'incapacité de répondre aux enquêtes (*'Maman (je réponds pour elle) est sous tutelle et ne répond pas aux enquêteurs'*, *'c'est ma mamie qui m'a demandé de le faire pour elle'*, *'ma femme rentrait de cure'*).

I Un taux d'échec global à l'enquête en face-à-face non négligeable

lorsque l'enquêtrice est passée ; elle n'était pas au courant de sa venue et j'étais absent à ce moment ; si j'avais été là , je l'aurais reçue et aurais répondu au questionnaire', 'Je suis en appartement avec mon frère ! Peut-on considérer cela comme un ménage ?').

Raisons invoquées	Occurrences	Ménages concernés
Préfère répondre en ligne	145	40%
Ne souhaite pas recevoir un enquêteur	124	34%
Peu disponible aux créneaux proposés	107	29%
Aucun contact avec un enquêteur	101	28%
Pas le temps	75	21%
Autre	26	7%
Total	578	

TAB. 12 – Les différentes raisons du choix du web comme média de réponse
Source : EMD web - Lyon 2006

Par ailleurs, selon les déclarations recueillies, quelques enquêteurs manquent de professionnalisme. Certains ne soignent pas assez la phase de recrutement (*'étudiante nous a laissé un message sur 1/2 page écolier déchirée, numéro personnel,...'*, *'J'ai eu un seul message d'un enquêteur, qui me demandait de le rappeler. C'est très bête, mais je n'avais pas l'intention de payer le téléphone pour répondre à votre enquête'*, *'L'enquêteur s'est présenté à mon domicile sans avoir pris de rendez-vous préalable'*). D'autres négligent les rendez-vous avec les ménages (*'J'avais convenu d'un rendez-vous avec un enquêteur, mais il ne s'est pas présenté, ni téléphoné pour annuler le rendez-vous'*, *'Deux rendez-vous pris avec vos enquêteurs ont été annulés par eux. J'avais rappelé pour refixer un rendez vous et vous n'avez jamais rappelé. C'est plus pratique de répondre sur internet'*, *'rendez-vous manqué avec votre représentant'*). Les sollicitations incessantes jouent négativement sur le niveau d'acceptation des études, ce qui se traduit également par des commentaires virulents (*'Gaspillage d'argent, nous avons reçu 4 courriers'*, *'je vois pas trop a quoi sert cette enquête ...on es en vacances, donc pas beaucoup de déplacements, donc vous ne comprenez pas nos galères d'attendre longtemps un bus alors qu'il ne passe pas pour X raisons !!!'*, *'les enquêtes de ...comme celle la me font ..., je perd mon temps !'*). Enfin, certains répondants ont compris tardivement l'importance de l'étude *'Après plusieurs courriers j ai compris que cette enquête était importante'*, ce qui légitimise l'utilisation des relances. D'autres ont été interpellés par le sujet de l'étude (*'je suis intéressée par la protection de l'environnement donc les déplacements urbains et je suis en train d'adhérer au PDE de mon entreprise'*) ou espèrent peser sur l'amélioration de leurs conditions de transport quotidiennes (*'volonté que Charbonnières centre soit mieux desservie dans le futur'*).

II Une expérience encourageante

II.1 Les taux de réponse constatés dans les protocoles d'enquêtes mixtes

Nous avons vu dans la section I, que le taux de réponse à une enquête se définit comme le ratio du nombre d'interviews complètes, par le nombre d'unités éligibles à l'enquête. Dans le cas de l'enquête web, il est difficile de connaître précisément le nombre de ménages éligibles, c'est-à-dire en mesure de se connecter à internet.

Les expériences d'enquêtes mixtes incluant le web étant restreintes dans le domaine des transports, nous faisons référence à des études réalisées dans d'autres domaines, tels que la santé, l'éducation ou la connaissance. Les recherches de Stanton (1998) montrent que les enquêtes web génèrent des taux de réponses relativement instables. Il peut être faible, de l'ordre de 8% (Smith, 1997) ou très élevé, atteignant les 90% (Zhang, 1999). Cependant, peu d'enquêtes web atteignent un taux de réponse de plus de 50%, et selon Falconer et Hodgett (1999) et Burke et James (2006) il semble que ce taux se situe davantage entre 10 et 35%. Schuldt et Totten (1994) et Kittleson (1995) précisent que lorsque le web est utilisé comme mode de recueil principal, le taux de réponse peut même se situer en dessous de celui d'une enquête postale. Schaefer et Dillman (1998) montrent que dans un protocole multimode, le taux de réponse à l'enquête web augmente. Cobanoglu *et al.* (2001) concluent à l'existence de grandes variations en termes de taux et de vitesse de réponse pour les enquêtes web. Comparer les taux de réponse de différents modes est complexe, puisqu'ils dépendent de la procédure et de la zone d'enquête. Dans un protocole d'enquête mixte, il n'est pas toujours possible de maintenir une procédure strictement identique entre deux enquêtes, puisque chaque mode a des caractéristiques particulières (Bonnel, 2003).

Le tableau 13 donne quelques résultats concernant le taux de retour de différents médias utilisés dans des enquêtes mixtes, et il semble que le taux de réponse soit plus faible lorsque le questionnaire est rempli sur le web.

Enquête	Description	Taux de réponse
Vehovar <i>et al.</i> 2000	Enquête sur l'usage du e-commerce en Slovénie. Répondants qui ont accès à internet.	Téléphone 52% Postal 39% Fax 32% Web 26%
Cobanoglu <i>et al.</i> 2001	Enquête auprès de 300 professeurs hospitaliers (USA)	Postal 26.39% Fax 17% E-mail/Web 44.2%
Crawford <i>et al.</i> 2002	Enquête auprès d'étudiants sur l'usage des stupéfiants	Postal 40% Web 63%
Huang 2006	Enquête auprès de collégiens	Papier 85% Web 63%

TAB. 13 – Etude comparée des taux de réponse dans des enquêtes mixtes

Source : Adapté de M. Lang, 2002

Les comparaisons sont cependant risquées, car les variables et les disciplines diffèrent. Dans certaines enquêtes le choix du média est laissé au répondant alors que dans d'autres ce choix est imposé. Le processus d'enquête varie parfois selon les modes et les taux de couverture des médias utilisés sont très différents. Il ne fait cependant aucun doute que les recherches futures permettront de rendre le taux de réponse aux enquêtes web comparable à celui des autres modes traditionnels, bien que ces procédures augmenteront substantiellement le coût de l'étude (Dillman et Bowker, 2001).

II.2 Une réactivité importante

L'enquête web s'adressant aux non-répondants à l'enquête traditionnelle, la diffusion du questionnaire en ligne était fortement liée au calendrier des interviews en face-à-face. Le lancement du terrain en face-à-face a eu lieu début novembre 2005, avant d'être suspendu pendant deux semaines, suite aux événements dans les banlieues françaises et à une perturbation du réseau de transport en commun. 4 335 courriers ont été envoyés aux répondants potentiels sur le web par les quatre bureaux d'enquêtes principaux de l'agglomération, en deux vagues, de manière à limiter le délai entre le refus et la relance. La première a débuté en avril 2006 (1 838 ménages contactés) et la seconde s'est déroulée à la mi-mai (2 497 ménages contactés). Pour chaque vague, deux relances étaient prévues auprès des ménages qui ne s'étaient pas connectés sur le site pour répondre à l'enquête en ligne⁸⁸. Comme la figure 11 le montre, la réactivité des internautes est très forte, et l'envoi des lettres avis ou des relances génère une augmentation du nombre de connexions les jours suivants.

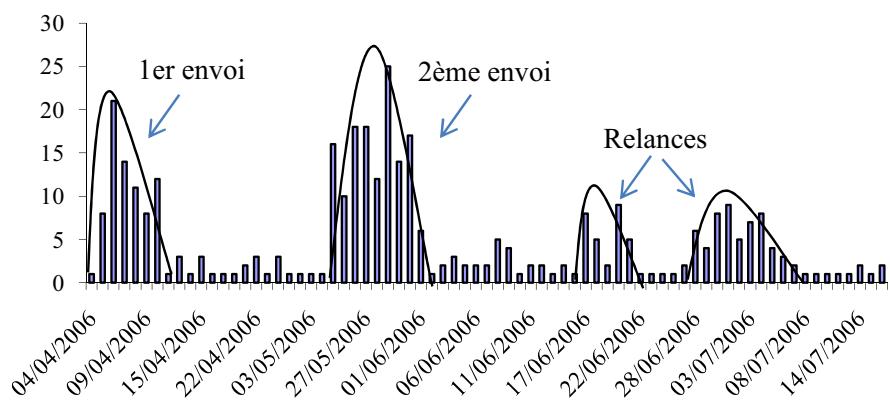


FIG. 11 – Dates de connexion des internautes

Source : Questionnaire web (2005)

II.3 Un taux de réponse satisfaisant

Au final, 536 individus ont accédé au site pour tenter de répondre à l'enquête (678 connexions enregistrées), ce qui représente un taux de connexion de

⁸⁸Les relances ont eu lieu les semaines du 22 mai et du 26 juin 2006 pour la première vague, les semaines du 12 et du 26 juin 2006 pour la deuxième vague.

Premiers résultats sur le comportement de réponse des internautes

12,4%. Ce chiffre est satisfaisant, puisque les ménages concernés par l'enquête web sont ceux qu'il n'avait pas été possible d'atteindre durant la première phase de l'enquête (face-à-face). De plus, le mode de recrutement n'est pas optimisé, puisqu'il est plus encourageant d'avoir à cliquer sur un lien pour répondre à une étude, que de suivre les instructions qui figurent sur une lettre pour pouvoir accéder au questionnaire en ligne (Conn, 2005). Parmi ces 536 contacts, 11 ont effectivement répondu au questionnaire, mais les données saisies ont été perdues suite à un problème informatique au démarrage de l'enquête. Ces 11 ménages de la vague 1 ont été relancés, puis classés en 'refus'. Cependant, tous les ménages qui se sont connectés n'ont pas terminé le questionnaire. Dans l'estimation du taux de réponse, nous ne retenons que les interviews 'exploitables', c'est-à-dire contenant au moins le cycle de déplacements de la personne interrogée. Après apurement, nous disposons de 364 interviews complètes (la personne interrogée est allée jusqu'à l'écran de validation finale du questionnaire) et de 369 interviews 'exploitables'. La population que nous allons étudier se limite alors à 369 ménages (ou individus), enregistrant un total de 1 108 déplacements, soit un taux de réponse de 8,5% (tableau 14). Ces chiffres sont cohérents avec les résultats de Lozar Manfreda et Vehovar (2002b), qui obtiennent 70% de questionnaires complets, parmi ceux ayant commencé à le remplir.

Nombre de connexions	Nombre de ménages	Bloc ménage	Bloc personne	Bloc déplacements	Validation finale
1	434	354	333	290	288
2	73	64	64	58	55
3	22	21	21	19	19
4	4	4	3	1	1
5	3	1	1	1	1
Total	536	444	422	369	364
% abandons	NA	17%	21%	31%	32%

TAB. 14 – Taux d'abandon cumulé du questionnaire web

Source : EMD web - Lyon 2006

On peut se demander si la taille de l'échantillon web est suffisante pour mener une analyse de la mobilité individuelle, c'est-à-dire si les indicateurs statistiques retenus se rapprochent de ceux réellement observés dans la population étudiée (Hill, 1998). Selon Roscoe (1975), il est important de respecter quelques règles avant de se lancer dans des analyses statistiques. Le nombre acceptable de répondants permettant de produire des analyses de qualité dépend du type d'étude⁸⁹. Dans les recherches expérimentales, où des variables explicatives sont manipulées dans le but d'évaluer la variation d'une variable

⁸⁹Selon Gay et Diehl (1992), de larges échantillons augmentent la probabilité d'obtenir des résultats significatifs aux analyses statistiques. En effet, pour de grands échantillons, une très petite différence observée entre deux moyennes peut apparaître significative, et être difficilement exploitable. Les recherches effectuées ne sont alors pas pertinentes. D'un autre côté, une différence significative obtenue entre deux petits échantillons est souvent complexe à expliquer, mais plus utile pour la recherche effectuée.

d'intérêt, les échantillons doivent contenir au moins 30 observations pour bénéficier des avantages du théorème central limite. Si des analyses sont menées sur des sous-échantillons, alors cette règle s'applique aux sous-échantillons considérés. Enfin, dans des analyses multivariées, la taille de l'échantillon doit être au moins dix fois supérieure au nombre de variables concernées. L'enquête ménages déplacements de Lyon est le fruit d'une recherche expérimentale et une expérience pilote. Au regard des règles citées ci-dessus, les 369 questionnaires recueillis en ligne représentent un effectif suffisant pour mener une analyse comparative de la mobilité.

II.4 La conversion des non-répondants

Le profil des répondants aux enquêtes web diffère de celui des non-répondants, même parmi ceux qui ont un accès individuel à internet. Les répondants, aussi nombreux soient-ils, peuvent constituer un échantillon non représentatif de la population (Stanton, 1998). Selon Crawford *et al.* (2002), différentes raisons expliquent le fort taux de non-réponse observé sur le web. Des individus ne sont pas dans la cible, n'ont pas consulté leur messagerie, ont rencontré des problèmes techniques, ont perdu la lettre avis, ont oublié de répondre, n'ont pas pris le temps de le faire, n'ont pas intérêt pour l'étude, craignent pour la confidentialité des données, ont une adresse e-mail erronée... Dans la littérature, on regroupe classiquement les refus en deux catégories (Sudman et Bradburn, 1996). Les 'hard refusals' concernent des individus réfractaires aux enquêtes, qu'il va être très difficile de faire changer d'avis. Les 'soft refusals' peuvent accepter de participer à l'étude après une ou plusieurs relances, en fonction de leurs convictions.

Couper et Rowe (1996) recensent les principales causes de refus invoquées par les répondants lorsqu'ils sont sollicités pour participer à une enquête. Le premier facteur est temporel ('je suis trop occupé', 'je n'ai pas le temps'), le second vient du manque d'intérêt des individus pour le sujet étudié ('je ne suis pas intéressé par...', 'je ne sais rien sur...'). L'auteur va plus loin et remarque des différences sociodémographiques entre les répondants, selon le motif de non participation. Ceux qui se disent trop occupés sont davantage diplômés, possèdent des hauts revenus, habitent dans de grandes agglomérations et subissent des contraintes temporelles importantes (actifs). Les individus plus vieux et moins éduqués refusent de répondre en invoquant plus souvent le manque d'intérêt pour le sujet. Si ces motifs peuvent être considérés comme une manière polie de décliner l'invitation de l'enquêteur, le fait de citer le manque de temps peut avoir un impact sur les enquêtes de mobilité, le temps étant une variable explicative du nombre de déplacements effectués.

Andrews *et al.* (0003) énumèrent quelques caractéristiques techniques qui impactent le taux de réponse aux enquêtes web : les éventuels problèmes techniques, la présence de caractères ambigus dans le mot de passe, les problèmes de confidentialité des données, les sauvegardes partielles des réponses, le placement des questions personnelles dans le questionnaire, le mode de contact (protocole en plusieurs étapes, mixité des modes, personnalisation de la lettre-

avis) les relances... Groves *et al.* (1992) complètent la liste avec les facteurs socio-économiques, les répondants ayant un niveau d'éducation et de vie plus élevés que les autres. Enfin, certains individus ont une propension à répondre aux enquêtes, alors que d'autres ne l'ont pas ou bien sont lassés d'être sollicités (Bickart et Schmittlein, 1999). Trois domaines de recherches sont particulièrement florissants dans la littérature : les leviers du taux de réponse (Yu et Cooper, 1983; Dillman, 2000), l'estimation et la correction du biais de non-réponse et la différence de comportement entre les répondants et non-répondants (Couper et Rowe, 1996).

Lors de l'enquête web de Lyon, un certain nombre de personnes ont accepté de livrer des informations personnelles. En fin de questionnaire, il est demandé aux répondants de bien vouloir saisir leur adresse e-mail. Cette question n'est pas obligatoire, mais 40% répondent positivement et livrent une adresse exploitable, s'exposant ainsi au risque d'être recontactés par le gestionnaire de l'enquête. Parmi eux, 15% communiquent leur adresse professionnelle, même s'ils déclarent posséder une connexion internet à domicile. Une investigation sur les caractéristiques socio-économiques de ces répondants montre que ce sont davantage les hommes actifs, diplômés du supérieur et occupant un emploi de cadre ou une profession intellectuelle supérieure qui saisissent volontairement leur adresse e-mail (annexe VIII). Ces personnes ne sont pas réticentes à livrer des informations personnelles, puisque 90% déclarent leur niveau de revenus (vs. 76% parmi ceux ne saisissant pas leur adresse e-mail). Les ménages concernés semblent avoir des ressources plus importantes, même si les faibles effectifs considérés ne nous permettent pas de conclure de manière significative. En revanche, il n'y a pas de différence selon l'âge.

III Analyse de la non-réponse

III.1 Non-réponse totale et non-réponse partielle

Groves et Couper (1998) distinguent les non-réponses totales ('unit non response'), lorsqu'un individu ne participe pas à l'étude, des non-réponses partielles ('item non response'), si l'individu participe à l'étude mais laisse quelques questions sans réponse. La non-réponse totale est la conséquence d'une absence de motivation du sujet, qui ne veut pas ou ne peut pas répondre à l'enquête (questionnaire non reçu, pas de temps pour répondre, incapacité à trouver les informations demandées, problèmes techniques ...). La non-réponse partielle survient lorsque les individus ne sont pas confortables avec certaines questions, mais sont motivés et en mesure de répondre au reste de l'étude. Il est parfois difficile de respecter cette césure. Des individus qui accèdent au questionnaire et prennent connaissance des questions sans y répondre sont généralement classés comme non-répondants et ceux qui abandonnent la saisie du questionnaire viennent alimenter la part des non-réponses partielles. Mais il est difficile de distinguer un abandon en cours de remplissage de la volonté d'éviter certaines questions.

Dans les enquêtes web, contrairement aux autres modes, le processus de ré-

ponse des individus est tracé automatiquement. Il est donc possible de recueillir des données qui permettent d'expliquer en profondeur les variations potentielles dans la participation aux enquêtes. Ces données, complémentaires aux questions de l'étude et récoltées passivement, sont qualifiées de méta-données (Bosnjak et Tuten, 2001). Pour recueillir des informations pertinentes, il est nécessaire que chaque question apparaisse séparément sur un écran, que l'internaute ne soit pas contraint de saisir une réponse pour accéder à l'écran suivant et que chaque page soit téléchargée indépendamment sur le serveur. Ces conditions sont difficiles à remplir et l'enquête web menée à Lyon ne les respecte pas toutes. Les méta-données collectées permettent cependant d'établir un premier diagnostic des abandons et de la non-réponse partielle.

III.2 Les causes d'abandon

Il est nécessaire de comprendre quelles sont les principales variables amenant des individus motivés pour participer à l'étude à interrompre le remplissage du questionnaire. Le taux d'attrition par section du questionnaire, ou le nombre de personnes qui commencent une enquête mais ne la terminent pas, peut révéler des jugements systématiques par certains groupes de la population et pèse sur la qualité de l'étude (Andrews *et al.*, 0003). Jeavons (1999) a mis en évidence trois principaux moments où les individus risquent d'abandonner la saisie des réponses : la première question, l'apparition d'une grille complexe de questions et la saisie de leur adresse e-mail. Certains autres écueils sont à éviter, comme les échelles de réponses trop complexes, qui demandent un grand effort de compréhension et qui aboutissent souvent à des abandons ou à des erreurs de mesure (Dillman, 1998). Il en est de même des questions ouvertes, qui alourdissent la tâche du répondant et sont difficilement analysables (Frary, 1996), ou des questions trop personnelles placées au début du questionnaire, qui effraient les répondants. Les abandons volontaires et ceux relatifs à des problèmes techniques sont souvent difficiles à distinguer.

Etant donné la longueur du questionnaire utilisé pour l'enquête ménages déplacements de Lyon, et son caractère auto-administré, certains individus n'ont pas pu (modem bas-débit, navigateur inadéquat et autres contraintes techniques...) ou pas voulu (questions parfois très personnelles, qui demandent un important effort de mémoire...) répondre entièrement à l'enquête et ont interrompu prématûrement la connexion. Le taux d'abandon progresse par paliers (tableau 14). Deux phases semblent importantes : la validation du bloc 'Ménage' et celle du bloc 'Déplacements' (respectivement 17% et 10% d'abandons). Le graphique 12 montre l'évolution du nombre d'abandons en fonction de la progression dans le questionnaire. Un répondant sur cinq interrompt la connexion après avoir saisi son identifiant et son mot de passe. Le reste des abandons est assez progressif, bien que deux moments soient critiques : la question sur le nombre de personnes composant le ménage (14%) et l'introduction de la partie relative à la saisie des déplacements (16%). Il est difficile de trouver des facteurs socio-économiques explicatifs de ces comportements, au regard des données recueillies.

Premiers résultats sur le comportement de réponse des internautes

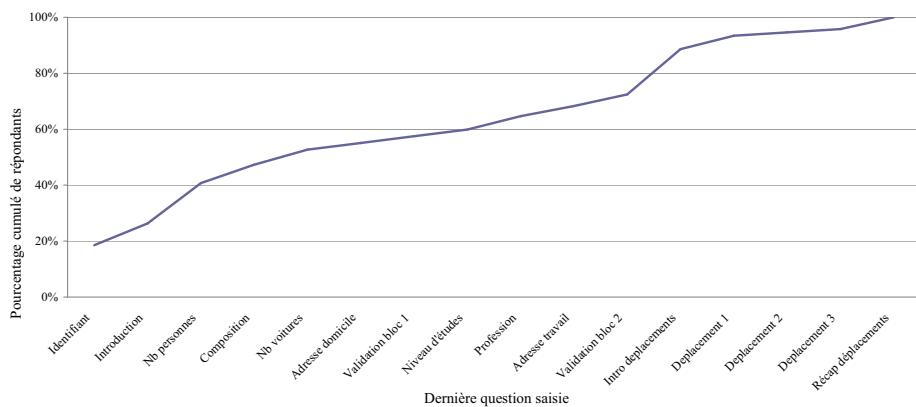


FIG. 12 – Evolution du nombre d'abandons selon la progression dans le questionnaire

Source : EMD web - Lyon 2006

Par ailleurs, le logiciel CATI utilisé pour l'enquête web impose au répondant de sélectionner une modalité à chaque question, afin de pouvoir accéder à l'écran suivant. Cette contrainte technique consiste à forcer les réponses, puisque les individus qui ne répondent pas aux questions sont relancés. Ces relances permettent de diminuer la non-réponse partielle, mais peuvent jouer sur les abandons.

Nombre de connexions	Nombre de ménages	Pourcentage de ménages	Durée moyenne par connexion (mn)
1	143	86%	12
2	16	10%	13
3	3	2%	11
4	3	2%	8
5	2	1%	10
Total	167	100%	12

TAB. 15 – Comportement de réponse des répondants partiels

Source : EMD web - Lyon 2006

Grâce au web, il est possible de recueillir des informations sur les répondants partiels au questionnaire, c'est-à-dire ceux qui n'ont pas terminé la saisie des réponses (tableau 15). La majorité d'entre-eux n'ont effectué qu'une connexion, mais 24 personnes (soit 15% des répondants partiels) ont accédé plusieurs fois au questionnaire web. Parmi les personnes qui se sont connectées plusieurs fois, seulement 42% l'ont fait le même jour et 50% dans un délai de 48h. Une hypothèse est le manque de motivation des répondants, qui n'ont pas mesuré l'importance de l'étude. Il se peut également que suite à des problèmes rencontrés pour saisir leurs réponses, ils aient tenté de joindre l'assistance technique à leur disposition. Les données disponibles ne nous permettent toutefois pas de conclure. Le temps de connexion est stable et relativement faible, les répondants partiels passant en moyenne 12 minutes devant leur ordinateur à

chaque connexion. Ce temps est moins élevé que celui observé chez les personnes ayant terminé le questionnaire (18 minutes). En revanche, aucune différence sociodémographique est notée, et l'âge moyen des répondants partiels ayant complété le tableau de composition du ménage est identique à celui des répondants (44 ans).

III.3 Un taux de non-réponse partielle modéré

Certains auteurs utilisent la non-réponse partielle comme critère d'évaluation de la qualité d'une étude (Atrostic et Burt, 1999). Le nombre de valeurs manquantes ou erronées varient selon le type de variables. Il est intéressant d'estimer la qualité d'une enquête non pas en s'intéressant au nombre de valeurs manquantes totales, mais en ne retenant que celles liées aux questions stratégiques pour une enquête déplacements⁹⁰. Cette non-réponse peut s'avérer problématique dans des enquêtes de mobilité, lorsque les informations saisies ne permettent pas au chargé d'études de localiser précisément les lieux visités. Le répondant peut ne pas être capable de se remémorer une adresse précise, surtout si ce lieu n'est pas celui de son domicile ou de son travail, et la possibilité offerte par le questionnaire, de renseigner l'intersection de rues ou le point d'intérêt le plus proche (mairie, école...) ne permet pas toujours un géobornage précis a posteriori. Il se peut également que les adresses saisies, bien que précises, ne soient pas identifiées par le logiciel de géobornage. Cela concerne notamment les adresses d'entreprises dans des zones industrielles, celles de centres commerciaux ou d'habitations dans des lieux-dits. Enfin, le répondant peut juger que les informations relatives aux lieux visités sont trop personnelles et préférer ne pas remplir certains champs du questionnaire. Les recommandations que nous avons formulées ci-dessus sur le design du questionnaire adapté, la formulation des questions, l'utilité du prétest et la prise en compte des thèmes sensibles tels que le revenu sont un moyen efficace de limiter les non-réponses partielles dans les enquêtes.

Face à des questionnaires relativement longs et à des personnes souvent pressées, certaines questions restent sans réponse. Ce phénomène est particulièrement vrai, lorsque le protocole d'enquête ne prévoit pas l'obligation de répondre aux questions ou lorsqu'un échappatoire (sous la forme 'NSP') est présent. Les questionnaires incomplets ou les réponses peu fiables se multiplient, nuisant à la qualité des données et à la pertinence des résultats obtenus (Morris et Adler, 2003). La structure du questionnaire web utilisé à Lyon impose le choix d'une réponse pour passer à la page suivante. L'analyse de la non-réponse partielle est donc complexe, bien que l'étude des réponses de type 'Autre', 'Ne sais pas' ou 'Non-réponse', assimilées à des non-réponses partielles, donne des pistes intéressantes. On peut se demander si ces modalités pèsent sur le taux de non-réponse partielle. De Rouvray et Couper (2002) postulent que proposer explicitement un choix de ne pas répondre favorise la non-réponse. Les détails des non-réponses de l'enquête web sont donnés dans le tableau 16.

⁹⁰Stopher *et al.* (2004b) définissent un indicateur statistique des valeurs imputées.

Premiers résultats sur le comportement de réponse des internautes

Nombre de non-réponses	Nombre de personnes	Pourcentage des non-réponses	Pourcentage des répondants (N=369)
0	259	-	70%
1	71	65%	19%
2	16	15%	4%
3	8	7%	2%
4	15	14%	4%
Total	110	100%	100%

TAB. 16 – Répartition des répondants selon le nombre de non-réponses partielles au questionnaire

Source : EMD web - Lyon 2006

Le tableau 17 présente le nombre de non-réponses aux questions prévoyant un échappatoire.

Intitulé de la question	Nombre de non-réponses
Possession du permis	1
Téléphone fixe	28
Téléphone portable	29
Connexion internet	21
Connexion haut débit	6
Revenus	90

TAB. 17 – Occurrence des non-réponses aux questions proposant la modalité de réponse 'Ne sait pas' ou 'Non-réponse'

Source : EMD web - Lyon 2006

Les modalités 'ne sait pas' et 'non-réponse' ont été utilisées de manière modérée par les répondants. 25% des personnes ne se sont pas prononcées sur le niveau de revenu du ménage, résultat conforme à nos attentes et qui encourage la présence d'un échappatoire dans les questions relatives aux gains financiers des individus. Les autres questions concernées par la non-réponse sont relatives à l'équipement du ménage en moyens de communication. Qu'il s'agisse de la possession d'un téléphone fixe, d'un téléphone portable, ou d'une connexion internet au domicile, le nombre de non-réponse est similaire. En revanche, les internautes connaissent le type de connexion dont ils disposent, et sont favorables à délivrer l'information.

Bien que la réponse aux questions soit forcée, il semble que quelques individus aient réussi à progresser dans le questionnaire en omettant certaines réponses. Il s'agit notamment des questions relatives aux caractéristiques du répondant (sexe, âge et occupation des personnes), aux adresses du domicile et du lieu de travail ou d'études et de celles concernant l'habitude d'utilisation des différents modes de transport. Ces oubli sont marginaux et ne concernent qu'une ou deux personnes selon les questions.

III.4 Des relances efficaces

Prévoir des relances est un bon moyen de réduire le biais de non-réponse. Keeter *et al.* (2000), Kalfs et Van Evert (2003) et Goyder (1985) montrent qu'il est possible d'augmenter le taux de réponse en déployant des efforts pour contacter les personnes difficilement joignables. Pour être efficaces, ces relances doivent être programmées dès la conception de l'étude, et effectuées à différents moments de la journée. En effet, plus les individus sont sollicités, plus ils ont des chances d'accepter le questionnaire, bien qu'il existe un seuil de 6 tentatives au-delà duquel l'impact sur la non-réponse est marginal (Harpuder et Stec, 1999) ⁹¹.

Les répondants web sont des non-répondants à l'enquête standard. Ils ont donc été sollicités plusieurs fois (8 au maximum) pour fixer un rendez-vous avec un enquêteur à domicile. Après réception du courrier les informant de la mise à disposition du questionnaire en ligne, deux lettres de relance étaient prévues, à un mois d'intervalle. Elles ont permis de récupérer certains individus qui n'avaient pas répondu au questionnaire web, suite à l'envoi du premier courrier (annexe VII).

Si les relances semblent efficaces pour augmenter le taux de réponse à l'enquête, qu'en est-il de l'impact sur la qualité des données ? Les travaux de Brög et Meyburg (1980, 1982) montrent suite à l'étude des résultats d'une enquête de mobilité réalisée par voie postale en Allemagne, que les répondants aux relances déclarent moins de déplacements que les autres et que les immobiles sont sur-représentés. L'hypothèse avancée est que les hyper mobiles répondent rapidement à une étude dont le sujet les intéresse (les déplacements), alors que les individus à faible mobilité ne se sentent pas concernés par le sujet et attendent plusieurs relances pour répondre au questionnaire. Les résultats de l'enquête web de Lyon vont dans ce sens, comme le montre le tableau 18.

		Nb de personnes	Mini	Maxi	Moyenne	Ecart-type	% mobiles
Total	Total	369	0	15	2,97	2,33	81%
répondants	Mobiles	299	1	15	3,71	2,02	100%
Répondants immédiats	Total	235	0	15	3,02	2,33	83%
	Mobiles	196	1	15	3,62	2,08	100%
Répondants relance 1	Total	60	0	12	3	2,57	75%
	Mobiles	45	1	12	4	2,18	100%
Répondants relance 2	Total	74	0	8	2,78	2,08	78%
	Mobiles	58	1	8	3,55	1,67	100%

TAB. 18 – Evolution du nombre de déplacements selon le nombre de relances effectué

Source : EMD web - Lyon 2006

⁹¹Selon Adler (2003), on observe un plus fort taux de non-réponse partielle chez les individus qui répondent au questionnaire après plusieurs relances.

IV Caractéristiques des connexions web

Au-delà du taux de réponse global, d'autres facteurs peuvent être mesurés pour évaluer la qualité de l'enquête : la dernière page complétée, le nombre de pages vues, le nombre de cases cochées, le nombre et la durée des connexions (Dillman *et al.*, 1998b).

IV.1 Des moments de connexions privilégiés

Un des avantages du web est la liberté de réponse laissée à l'internaute, qui peut choisir le lieu et le moment le plus opportun pour répondre. L'expérience menée par Hojman *et al.* (2004) montre que la répartition des connexions n'est pas uniforme sur la journée et met en évidence deux pics de connexion : un le matin entre 9h00 et 10h00 et un le soir entre 19h00 et 20h00.

En ce qui concerne l'enquête ménages déplacements de Lyon, nous remarquons une grande disparité entre les jours de semaine et le week-end⁹². La semaine, les internautes se connectent de 9h00 à 23h00, avec trois pics qui se distinguent dans la journée : le midi entre 11h00 et 13h00, l'après-midi entre 15h00 et 17h00 et le soir entre 19h00 et 22h00 (figure 13). Il est probable qu'un nombre important de connexions soit réalisé sur le lieu de travail, d'autant que 20% des répondants déclarent ne pas posséder de connexion internet à leur domicile.

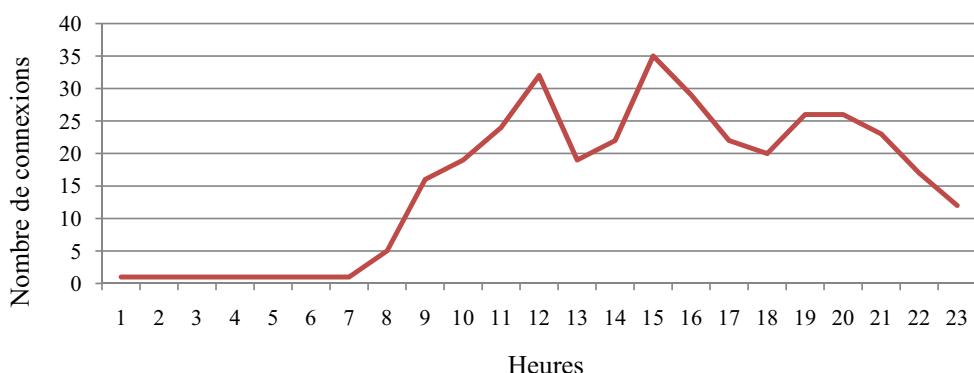


FIG. 13 – Heures de connexion des internautes la semaine

Source : EMD web - Lyon 2006

Le week-end, la répartition horaire des connexions est totalement différente (figure 14). Un seul gros pic entre 18h00 et 00h00 diffère avec le reste de la journée, où l'on constate un nombre de connexions faible mais constant (environ 5 à 10 par heure).

⁹²Bradley (1999) postule que le profil des enquêtés varient selon l'heure de remplissage du questionnaire, mais des analyses sur les enquêtes transports devraient être menées pour valider ou infirmer ce résultat.

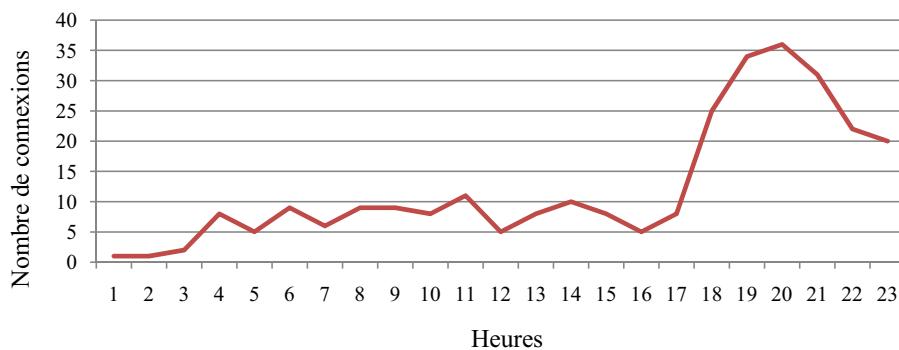


FIG. 14 – Heures de connexion des internautes le week-end

Source : EMD web - Lyon 2006

IV.2 Une durée de connexion étroitement liée au nombre de déplacements saisis

Le questionnaire web a été simplifié par rapport au questionnaire de l'enquête face-à-face, et plus ou moins basé sur celui des enquêtes C.A.T.I., réalisées par téléphone dans les villes moyennes. Cependant, le caractère peu ergonomique du site, la complexité des questions et la non familiarité supposée des individus avec le vocabulaire utilisé dans les enquêtes de mobilité laissaient penser que le temps de réponse moyen serait plus élevé sur le web. Le tableau 19 montre que le temps moyen de réponse au questionnaire web est proche de celui constaté lors des enquêtes par téléphone (23 mn, vs. 20 mn). Les disparités sont toutefois importantes (entre 5 mn et plus d'une heure et demi).

Statistique	Durée du contact (mn)
Minimum	4,9
Maximum	96,1
1er Quartile	14,1
Médiane	18,9
3ème Quartile	28,3
Moyenne	22,6
Ecart-type (n)	12,6
Effectif total	369

TAB. 19 – Durée de connexion globale du questionnaire web

Source : EMD web - Lyon 2006

Une analyse plus précise du comportement des internautes en termes de remplissage du questionnaire est possible, puisque des chronos insérés dans le questionnaire permettent de connaître le temps effectif passé sur les blocs de questions 'Ménages', 'Personnes' et 'Déplacements'. Les données du tableau 20 concernent les individus ayant terminé la saisie du questionnaire ou validé le bloc 'Déplacements'. Nous remarquons que la partie la plus consommatrice

Premiers résultats sur le comportement de réponse des internautes

de temps est celle qui concerne les déplacements, puisqu'en moyenne plus de 9 mn y sont consacrées. Ceci s'explique par le nombre et la précision des questions, qui demandent un effort de mémoire important (adresse des lieux visités, horaires de départ et d'arrivée ...). Ici encore, des disparités sont observées, puisque 25% des répondants prennent moins de 5 mn pour renseigner leurs déplacements, alors que 25% y consacrent environ 12 mn. Pour les autres parties du questionnaire, les temps de réponse sont plus courts (entre 2,5 et 4 mn en moyenne) et les différences moins marquées (écart-type d'environ 2 mn). Les blocs ménage et personne regroupent des questions plus 'habituelles' pour les interviewés, et proposent souvent des listes à cocher pour les réponses, ce qui simplifie considérablement le travail de l'internaute et minimise son temps de connexion.

Statistique	Bloc Ménage 1	Bloc Personne	Bloc Déplacements	Bloc Ménage 2
Minimum	0,2	0,2	0,6	0,8
Maximum	18,0	20,9	53,9	16,6
1er Quartile	2,4	1,9	4,4	1,5
Médiane	3,2	2,7	7,6	1,9
3ème Quartile	4,4	3,7	11,8	2,7
Moyenne	3,8	3,2	9,2	2,4
Ecart-type (n)	2,6	2,3	7,4	1,8
Effectif total	369	369	369	369

TAB. 20 – Durée de connexion du questionnaire web par bloc de réponses (en mn)

Source : EMD web - Lyon 2006

Le temps de saisie des réponses au questionnaire web est au final lié au nombre de déplacements déclarés, comme le montre la figure 15.

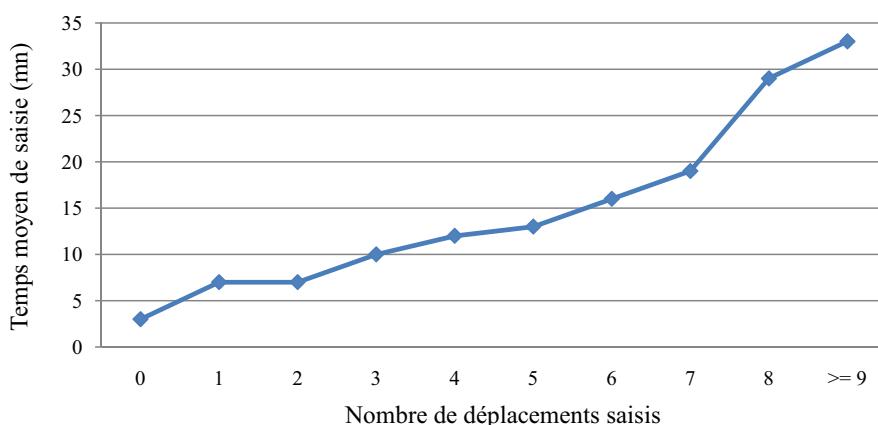


FIG. 15 – Temps moyen de saisie des réponses en ligne (mn)

Source : EMD web - Lyon 2006

IV.3 Peu de connexions multiples

Il est également intéressant de considérer l'impact du nombre de connexions sur la qualité des données saisies (tableau 21). Malgré la possibilité offerte aux ménages de remplir le questionnaire en plusieurs parties, on observe une connexion unique chez 80% des répondants. 21% des enquêtés (78 personnes) se sont connectés à plusieurs reprises avant de terminer le questionnaire, interrompus par des problèmes techniques ou un manque de disponibilité, dont près des trois quarts (73%) ne se sont connectés que deux fois seulement. Le nombre de répondants ayant effectué trois connexions ou plus est donc marginal. La sauvegarde intermédiaire des résultats, rendue possible grâce à une connexion sécurisée par un identifiant et un mot de passe rassure les répondants, mais est peu utilisée. Le temps moyen de connexion est stable, quel que soit le nombre de tentatives nécessaire pour arriver à la fin du questionnaire, et est en moyenne égal à 18 minutes.

Nombre de connexions	Nombre de ménages	Pourcentage des ménages	Durée moyenne par connexion (mn)
1	291	79%	19
2	57	15%	15
3	19	5%	15
4	1	0%	24
5	1	0%	15
Total	369	100%	18

TAB. 21 – Comportement de réponse selon le nombre de connexions

Source : EMD web - Lyon 2006

Il semble donc y avoir une durée maximum de connexion acceptable par les répondants, qui choisissent de se reconnecter à un autre moment si le questionnaire n'est pas terminé durant ce laps de temps. L'analyse des dates de connexion est également riche d'enseignement : 60% des connexions multiples ont lieu le même jour, 20% se font dans les deux jours qui suivent la connexion initiale et environ huit personnes sur 10 terminent de remplir le questionnaire en trois jours, quel que soit le nombre de connexions effectué. Ce constat rassure sur la qualité des données recueillies, puisque la mémoire individuelle concernant les déplacements est plus aléatoire au fur et à mesure que le temps passe. Il est en revanche délicat de cerner les raisons de ces connexions multiples, bien que les répondants qui terminent le questionnaire lors d'une connexion unique soient légèrement plus jeunes que les autres (42 ans, vs. 45 ans), ce qui peut traduire une certaine aisance avec l'outil internet. La littérature met en garde les chercheurs sur les différences d'équipement informatique des ménages, mais le tableau 22 montre que la proportion de ménages équipés d'une connexion internet à domicile, et parmi eux la part des connexions haut débit, est stable suivant que l'individu se soit connecté une seule ou plusieurs fois pour remplir le questionnaire.

Type de connexion	Unique	Multiple	Total
Connexion internet au domicile	82%	77%	80%
Dont connexion haut débit	93%	90%	92%

TAB. 22 – Equipement internet selon le nombre de connexions

Source : EMD web - Lyon 2006

V Conclusion

Les résultats de l'enquête web, bien que modestes, restent encourageants. Mais au-delà du taux de réponse global, il est nécessaire de s'interroger sur la composition du groupe d'internautes qui n'avaient pas accepté de répondre à la première enquête en face-à-face, ainsi qu'à la qualité des données recueillies. Certaines erreurs de la part des enquêteurs semblent avoir pesé sur le taux de non-réponse à l'enquête en face-à-face. Nombre d'individus préfèrent cependant répondre aux enquêtes en ligne. Il s'agit de personnes jeunes, ayant un haut niveau d'éducation et de revenus. Tous les internautes qui se sont connectés sur le site n'ont pas achevé la saisie des réponses. Des problèmes techniques peuvent être à l'origine de quelques abandons, mais nous avons identifié des questions clés, à partir desquelles des répondants quittent le questionnaire. Il s'agit essentiellement de questions personnelles, comme la saisie de l'adresse du domicile, ou de l'enregistrement des déplacements effectués durant la période de référence, étape fastidieuse.

Le web est un média très réactif. Les relances semblent efficaces et le temps entre la sollicitation du répondant et sa première connexion sur le site est relativement court. Bien que certains remplissent le questionnaire par partie, le temps moyen acceptable par connexion est 20 minutes. Par ailleurs, la présence d'échappatoires dans le questionnaire ne pèse pas trop sur la non-réponse partielle, et permet de ne pas frustrer les individus ne souhaitant pas se prononcer sur certaines questions. Enfin, un climat de confiance s'installe avec certains répondants, qui acceptent de livrer leur adresse e-mail personnelle à la fin du questionnaire.

Si le bilan de la conduite de l'enquête apparaît positif, il n'en reste pas moins que l'analyse des résultats doit faire face à quelques difficultés. D'une part, il convient d'analyser les caractéristiques de la population des internautes qui risquent de se différencier de l'ensemble de la population, compte tenu du taux de pénétration et d'aisance d'internet dans la population française. D'autre part, il faut s'assurer de la comparabilité des comportements de déplacements saisis quel que soit le média. Ces analyses sont conduites dans la partie suivante.

Conclusion de la partie I

Dans cette partie, nous avons mis en évidence les nombreux atouts du web en tant que mode de recueil de données (coût modéré, interactivité du questionnaire, administration rapide de l'enquête...). Cependant, si ce média est attractif et moins contraignant en termes de disponibilité temporelle, tous les individus n'ont pas accès à Internet et les compétences des internautes sont inégales. Actuellement, il est impossible de réaliser une enquête en ligne auprès d'un échantillon représentatif de l'ensemble de la population. La généralisation des réponses fournies par les méthodes d'inférence statistique reste donc problématique, même si dans le cas de l'enquête ménages déplacements de Lyon le biais d'échantillonnage est réduit, puisque les personnes qui se voient proposer un questionnaire web ont été auparavant recrutées par une méthode probabiliste.

Le questionnaire traditionnellement utilisé pour les interviews en face-à-face ne peut être diffusé en l'état sur internet, puisque la perception des questions et l'organisation du questionnaire dépendent du potentiel du média utilisé. Le questionnaire papier de l'enquête en face-à-face a donc été simplifié, afin de l'adapter à l'enquête web, de nature auto-administrée. Il était également essentiel de le rendre interactif et conforme aux attentes des internautes, habitués à participer à des enquêtes en ligne depuis plusieurs années.

Le taux de réponse global de l'enquête web est de 8.5%. Ce chiffre est faible, en comparaison du taux de réponse de l'enquête ménages standard menée en face-à-face (53%, et 50% sur le territoire du SCOT de Lyon, concerné par l'enquête web), mais reste encourageant, étant donné qu'une partie seulement des ménages ciblés peuvent se connecter à Internet, au domicile ou sur leur lieu de travail (50% des ménages français disposent d'une connexion Internet) et que l'option web a seulement été proposée aux ménages qui refusaient de répondre à l'enquête traditionnelle ou qui restaient injoignables. Par ailleurs, étant donné le caractère exploratoire de cette étude, nous nous sommes heurtés à quelques difficultés, notamment l'utilisation d'un logiciel peu approprié aux enquêtes de mobilité. Des améliorations dans le design et l'ergonomie du questionnaire nous laisseraient ainsi espérer un taux de réponse plus avantageux.

Si le web peut augmenter la qualité des enquêtes, il ne peut remplacer totalement les autres modes. La recherche doit montrer quels groupes sont plus à même de fournir des réponses de bonne qualité et par quel média (Kalfs et Van Evert, 2003). La collecte de données par différents médias est une solution intéressante, pour augmenter le taux de réponse, mais nécessite une

réflexion sur l'intégration des données, dans le but d'améliorer la représentativité de l'enquête. En effet, si la mise en place d'une enquête web en parallèle de l'enquête ménages déplacements permet d'appréhender des comportements de mobilité peu représentés jusqu'ici (individus actifs, aux horaires décalés...), la question de la comparabilité des données reste entière. Le choix du média, du mode d'administration du questionnaire et les caractéristiques des répondants ont probablement une influence non négligeable sur les comportements de mobilité. Il est donc nécessaire de comparer les caractéristiques et les comportements de mobilité des individus qui répondent en ligne avec ceux qui préfèrent recevoir un enquêteur à domicile.