

La carte de randonnée vue par les randonneurs

Priscille Baldit-Schneller ¹, Catherine Dominguès ²

¹ TIM/INaLCO – 49 bis avenue de la Belle Gabrielle – 75012 Paris – France

² Institut géographique national – 73 avenue de Paris – 94160 Saint-Mandé – France

Résumé

Cet article présente l'exploitation d'une enquête réalisée auprès de différents groupes de randonneurs sur leur pratique de la randonnée. Il s'agit de contraster ces différents entretiens collectifs afin de dégager, pour chaque type de randonnée, les points importants pour la préparation et le déroulement de la randonnée, ainsi que les critiques et les attentes des randonneurs. L'exploitation de cette enquête combine des outils statistiques simples : recherche des termes significativement fréquents, et des outils linguistiques : patrons morpho-syntaxiques pour la recherche de marqueurs modaux. Les limites de l'enquête et de son exploitation sont discutées.

Abstract

A survey has been carried out about hikers, cyclists and mountain bikers in order to define an adapted sport map. This paper presents the way these collective interviews have been analysed in order to bring out relevant cartographical objects as well as critics and expectations of questioned people. The study combines simple statistical tools (significantly more present words) with linguistic tools (morpho-syntactical patterns for modality markers). The survey and tools limits' are discussed as a conclusion.

Keywords: open questions interview, cartography, excursion, lexicometrical analysis, opinion analysis

1. Introduction

L'Institut Géographique National (IGN) dispose de bases de données géographiques, à différentes échelles de référence et résolutions, qui lui permettent de sélectionner des données géographiques pertinentes pour répondre à un besoin spécifique ¹ et de réaliser, à différentes échelles et sur différents supports, des cartes topographiques, des plans de villes et certaines cartes thématiques : cartes aéronautiques, routières, touristiques ; cartes de randonnée, de promenade ; etc.

Une carte se singularise à la fois par les données qu'elle cartographie et par les choix de représentation qu'elle affiche ; ainsi une même base de données géographiques permet d'élaborer différentes représentations adaptées à des besoins identifiés et spécifiques. L'IGN projette, en particulier, d'éditer une carte au 1:50 000 plus spécialement destinée aux utilisateurs randonneurs. L'objectif est de réaliser une carte à la fois riche, lisible et adaptée à son lecteur. Les règles de sémiologie graphique (Bertin, 1967) mises en œuvre lors de la traduction, sous

¹ Par exemple : composition d'un fond cartographique adapté à la cartographie de données thématiques ; données 2D et 3D permettant de mesurer l'évolution d'une zone urbaine ; données 3D pour l'implantation d'un ouvrage d'art ; etc.

forme graphique, du message censé être délivré par la carte garantissent la compréhension du message par le lecteur. Cependant richesse, lisibilité et adaptation sont le plus souvent antagonistes. La richesse d'une carte augmente avec la quantité d'informations qu'elle fournit, les informations cartographiées devant aussi être pertinentes par rapport à la fonction de la carte définie par son utilisateur. La lisibilité de la carte diminue avec la densité des informations représentées ; elle dépend aussi de l'homogénéité, quant au niveau de détail, des données affichées. Enfin l'adaptation doit prendre en considération la fonction de la carte, les goûts de l'utilisateur et son ressenti quant à l'aspect visuel de la carte, ces différents paramètres étant définis par l'utilisateur lui-même.

Afin de créer une carte adaptée aux besoins des randonneurs, une enquête a donc été menée. Elle a été réalisée sous forme d'entretiens auprès de différents groupes de randonneurs, et visait à leur permettre de préciser leurs besoins et leurs attentes quant aux spécifications de la carte de randonnée.

Nous avons souhaité, dans le cadre d'un stage, exploiter cette enquête à l'aide d'outils statistiques et linguistiques. L'objectif consiste à définir, à partir des commentaires des utilisateurs et de l'expression de leur ressenti, les spécifications de la carte de randonnée adaptée à leur pratique. L'exploitation de cette enquête se rattache au traitement des questions ouvertes dans le cadre d'entretiens semi-directifs². Il s'agit de combiner une analyse qualitative du contenu, en principe dépourvue d'a priori quant à la détection des besoins (Roux and Lefèvre, 2004), avec une analyse quantitative fondée sur les données lexicales (vocabulaire employé, richesse lexicale, cooccurrences, ...) contenues dans le texte des entretiens (Lebart and Salem, 1994). Le paragraphe 2 précise la méthode de recueil des entretiens et la mise en variables des données afin de constituer un corpus exploitable avec des outils statistiques et linguistiques. Le paragraphe 3 expose les résultats obtenus quant à la mise en évidence des spécificités de la carte de randonnée, des critiques et des attentes des différents groupes de randonneurs concernant les cartes utilisées pour la randonnée. Dans le paragraphe 4 sont discutées les hypothèses qui sous-tendent cette étude et ses perspectives.

2. Constitution du corpus

L'enquête réalisée auprès de randonneurs a permis de construire un corpus constitué par la retranscription des entretiens. Avant toute exploitation à l'aide d'outils linguistiques ou statistiques, ce corpus nécessite d'être prétraité : correction des fautes de frappe et d'orthographe, normalisation de la graphie (Heitz, 2006 ; Baldit, 2008). Les prétraitements qui ont été effectués ne sont pas décrits ici. Dans ce paragraphe, nous présentons le contexte de l'enquête et la méthode de recueil des entretiens ainsi que la mise en variable des données textuelles contenues dans le corpus. Quelques caractéristiques numériques décrivant le corpus sont listées à la fin du paragraphe.

2.1. Recueil des entretiens

Les randonneurs interrogés sont regroupés en fonction de leur mode de randonnée : randonneurs fréquents à vélo, en vtt (au moins cinq randonnées par an pour les participants de chacun de

² Un entretien semi-directif est une méthode d'étude qualitative fondée sur la réalisation d'entretiens individuels ou collectifs durant lesquels l'animateur dicte uniquement les différents thèmes devant être abordés sans pour autant pratiquer un questionnement précis. Ce type d'entretien vise généralement à comprendre en profondeur des attitudes ou comportements. (*Glossaire de marketing* : <http://www.definitions-marketing.com/>).

ces groupes), et pédestres (sans aucune indication du nombre de randonnées). Les randonneurs/promeneurs occasionnels (moins de huit randonnées par an quel que soit le mode de randonnée) forment le dernier groupe. Quatre groupes de neuf ou dix participants sont ainsi constitués.

Chaque groupe est réuni pour un entretien collectif d'environ trois heures conduit par un animateur, le même pour les quatre groupes. Les entretiens sont orientés par un guide d'animation fourni à l'animateur, qui récapitule l'ensemble des questions auxquelles l'entretien est censé apporter des réponses et les techniques d'entretien qui peuvent être mises en œuvre pour obtenir ces réponses. Après un tour de table permettant à chacun de se présenter et de raconter une expérience marquante de randonnée, les questions posées visent à recueillir les commentaires spontanés des utilisateurs relatifs à la carte de randonnée idéale.

Les entretiens sont enregistrés (vidéo et audio) et l'anonymat des participants est garanti. Questions et réponses sont ensuite retranscrites par l'animateur du groupe. Le corpus d'étude est constitué des retranscriptions de cette enquête qualitative à partir de questions ouvertes (Brugidou et al., 2004 ; de Peretti, 2005 ; Gélinas-Chebat et al., 2004 ; Urdapilleta et al. 2008).

2.2. Mise en variable des données textuelles

Nous souhaitons analyser ces données qualitatives avec des outils statistiques (Gavart-Perret and Moscarola, 1998). Cette approche statistique permet de comparer les réponses des locuteurs sur le plan lexical (lexicométrie) et facilite la recherche d'information par le repérage de marqueurs linguistiques dans le corpus. Pour mettre en œuvre ces méthodes et établir des relations entre les informations collectées, ou entre ces informations et des variables externes, il est nécessaire de mettre en variable les données textuelles (Peretti, 2005).

La mise en variable des données a pour objectif de segmenter le corpus afin de faire correspondre des réponses à des questions, et donc permettre la comparaison entre les réponses à la même question des locuteurs des différents sous-corpus. Des jalons textuels correspondant à des variables pertinentes sont ajoutés dans le corpus pour caractériser chaque partie (Lamalle et al., 2006). Ces balises correspondent à une variable externe (numéro de question) ou à une variable interne, issue du texte (numéro du locuteur, groupe de randonneur). Elles sont insérées de manière automatique en exploitant les caractéristiques typographiques de la retranscription (les lignes en gras décrivent le thème de la question ; celles commençant par un chiffre donnent le numéro attribué à chaque locuteur). Fig. 1 montre les annotations ajoutées à un extrait du texte brut pour obtenir le même extrait annoté.

Texte brut :

Supports utilisés pour préparer

1 La carte IGN 1/25 ou 1/50, après on se renseigne sur le topoguide de la région, du coin, on va à l'office de tourisme, à l'office on récupère tout ce qu'ils veulent bien nous donner, tout ce qui peut nous intéresser.

Texte annoté :

<question=2A> Supports utilisés pour préparer

<locuteur=loc1> *1 La carte IGN 1/25 ou 1/50, après on se renseigne sur le topoguide de la région, du coin, on va à l'office de tourisme, à l'office on récupère tout ce qu'ils veulent bien nous donner, tout ce qui peut nous intéresser.*

Figure 1 : Exemple d'insertion de balises (le format choisi est celui du logiciel Lexico3 ³)

³ Lexico3 est réalisé par l'équipe SYLED-CLA2T – Université de la Sorbonne Nouvelle Paris 3.

2.3. Caractéristiques quantitatives du corpus

Les tailles des différents sous-corpus diffèrent : le corpus vététiste contient 18 % moins de mots que le corpus des randonneurs pédestres. Pourtant, le discours des vététistes étant lexicalement plus riche, le nombre de lemmes contenus dans ces deux sous-corpus sont quasi identiques. Le tableau de Fig. 2 permet aussi de mettre en évidence que le nombre d'interventions de l'animateur n'est pas lié à la taille du corpus. En particulier, le corpus des cyclistes est le plus important en taille mais il est aussi celui pour lequel l'animateur est le moins intervenu.

<i>Sous-corpus</i>	<i>Nombre de locuteurs</i>	<i>Nombre d'interventions de l'animateur</i>	<i>Nombre de mots (avant annotation)</i>	<i>Nombre de lemmes</i>	<i>Richesse lexicale (nb lemmes / nb mots)</i>
pédestres	9	135	11 863	1 352	11%
vététistes	9	116	9 716	1 361	14%
cyclistes	10	86	12 972	1 573	12%
occasionnels	9	120	12 296	1 487	12%
Total	37	457	46 847	5 773	12%

Figure 2 : Caractéristiques quantitatives des différents sous-corpus

3. Caractéristiques des cartes de randonnée idéales

(Reinert, 1993) formule l'hypothèse selon laquelle un énoncé traduit un point de vue particulier qui implique en son centre « l'existence d'un *sujet* dans une certaine modalité du faire ou de l'être ». Il précise aussi que cette hypothèse « consiste justement à considérer le vocabulaire d'un énoncé particulier comme une trace pertinente de ce point de vue. Il est à la fois la trace d'un lieu référentiel et d'une activité cohérente du sujet énonciateur ».

Dans ce contexte, définir les spécifications de la carte de randonnée idéale dépend de la pratique de la randonnée des utilisateurs. Il s'agit alors de mettre en évidence :

- les objets, notions et concepts importants pour les randonneurs, selon leur pratique ;
- les critiques formulées par les utilisateurs sur les cartes existantes et leurs attentes concernant la carte de randonnée à construire.

Une analyse lexicométrique permet, pour un corpus constitué d'une suite de réponses à des questions ouvertes, de faciliter l'accès aux informations et d'établir des relations entre les informations collectées ou entre ces informations et des variables externes. (Lebart and Salem, 1994) précisent d'ailleurs que les analyses de réponses regroupées seront en fait assez voisines des analyses de « vrais textes », contrairement aux analyses de *réponses individuelles*. (Gavart-Perret and Moscarola, 1998) distinguent deux types d'analyses lexicales : la première, qualifiée d'analyse lexicale descriptive, permet d'explorer le corpus ; la seconde, analyse lexicale interprétative, repère des modèles d'énonciation. Nous utilisons successivement ces deux approches et les résultats sont ensuite synthétisés dans un tableau à la fin de ce paragraphe.

3.1. Objets, notions et concepts importants pour les randonneurs

Les contraintes liées à l'exercice de la randonnée varient selon le type de randonnée pratiqué. Il est donc pertinent d'observer le lexique employé par chaque type de randonneurs et donc d'étudier chaque sous-corpus séparément afin d'extraire les objets, notions et concepts importants pour construire des cartes de randonnée adaptées à une certaine pratique de la randonnée. Nous mettons en place une analyse contrastive. Pour chaque type de randonneurs (i.e. dans chaque

sous-corpus), nous recherchons les mots « significativement fréquents » par rapport à un corpus de référence. Cette méthode relève de l'analyse lexicale descriptive (Labbé, 2000 ; Beaudoin, 2000). Elle s'appuie ici sur l'outil Lexico3. Il s'agit de quantifier les fréquences des unités lexicales⁴ et leur fréquence d'association. Une forme est dite spécifiquement positive d'une partie (ou forme caractéristique de cette partie) si sa fréquence est anormalement (pour un seuil de spécificité fixé) élevée dans cette partie par rapport à un corpus de référence. Ici, les parties étudiées sont successivement chacun des sous-corpus ; le corpus de référence, identique pour la comparaison avec chacun des quatre sous-corpus, est formé de l'ensemble des quatre sous-corpus.

Le corpus soumis à l'analyse est non lemmatisé afin d'éliminer les erreurs éventuelles du lemmatiseur : par exemple, le nom *dénivelé* est homographe du participe passé de *dénivelé* ; en cas d'erreur du lemmatiseur, l'évocation du marquage des dénivelés sur la carte serait perdue au profit du verbe d'action, moins pertinent, *dénivelé*. Les résultats sont présentés en Fig. 3 (le signe + indique que sont également prises en compte les occurrences de ce mot au pluriel : à partir de la forme extraite, singulier ou pluriel, les occurrences de la forme duale ont été extraites et comptabilisées manuellement). Les nombres d'occurrences sont indiqués après chaque terme : dans le sous-corpus/dans l'ensemble des 4 sous-corpus.

<i>Pédestres</i>		<i>Vététistes</i>		<i>Cyclistes</i>		<i>Occasionnels</i>	
<i>[carte] 1:50 000</i>	45/108	<i>[carte] 1:25 000</i>	28/70	<i>chevron+</i>	12/12	<i>balade+</i>	31/49
<i>chargée</i>	7/10	<i>cartes</i>	51/163	<i>équipements</i>	6/6	<i>compliqué+</i>	13/17
<i>chemin</i>	14/27	<i>CD-rom</i>	13/22	<i>forêt+</i>	22/26	<i>facile</i>	9/16
<i>dénivelé+</i>	12/17	<i>clarté</i>	6/7	<i>Internet</i>	17/23	<i>fouillis</i>	6/7
<i>kilométrage</i>	10/15	<i>confiance</i>	5/5	<i>kilomètres</i>	14/26	<i>itinéraire+</i>	18/27
<i>[une vue]</i>	6/6	<i>courbe+</i>	19/44	<i>logiciel</i>	11/15	<i>jeune</i>	5/6
<i>macro</i>		<i>[de niveau]</i>					
<i>pédestre</i>	7/10	<i>GPS</i>	21/44	<i>Michelin</i>	18/28	<i>légende</i>	20/40
<i>randonne</i>	85/174	<i>GR⁵</i>	10/14	<i>pente</i>	5/6	<i>massif</i>	5/6
<i>repérage</i>	7/8	<i>information</i>	7/10	<i>route+</i>	39/66	<i>nous</i>	25/54
<i>sentier+</i>	18/37	<i>Libris</i>	5/5	<i>vélo</i>	52/94	<i>points</i>	28/79
<i>terrain</i>	5/26	<i>végétation</i>	7/10	<i>ville+</i>	13/25	<i>sympa</i>	9/16
<i>Vous</i>	53/111	<i>VTT</i>	46/70	<i>visible</i>	5/5	<i>trop</i>	41/116

Figure 3 : Liste ordonnée des termes spécifiques du vocabulaire de chaque sous-corpus (les mots outils : prépositions, déterminants, ont été supprimés de cette liste)

Le calcul des spécificités effectué ici permet d'extraire les termes spécifiques du vocabulaire de chaque catégorie de randonneurs. Ces termes sont ensuite regroupés thématiquement (en fonction de connaissances en cartographie, externes à l'enquête) ; par exemple *1:25 000*, *1:50 000*, *1:100 000* sont des termes relatifs à l'échelle de la carte.

3.2. Critiques des cartes de randonnée existantes et attentes des randonneurs

Des cartes existent pour aider les randonneurs à préparer et réaliser leur randonnée. Un des objectifs de l'enquête est d'évaluer ces outils : il s'agit de recueillir et d'analyser, par type de pratique et donc dans chaque sous-corpus, les critiques des randonneurs sur les cartes existantes

⁴ Une unité lexicale est entendue ici comme une forme graphique comprise entre deux blancs.

⁵ GR est l'abréviation normalisée de *sentier de grande randonnée*.

et leurs attentes quant à un support de randonnée adapté. La démarche développée ici relève de l'analyse lexicale interprétative. Nous avons utilisé deux méthodes pour repérer les indices linguistiques désignant les passages pertinents : l'extraction de marqueurs modaux et le repérage d'associations lexicales. Ces deux méthodes utilisent des patrons morphosyntaxiques dont des exemples sont donnés.

3.2.1. Extraction de marqueurs modaux

Les modalités élocutives « précisent la manière dont le locuteur révèle sa position vis-à-vis du propos qu'il énonce » (Charaudeau, 1992 : 599). La méthode consiste alors à repérer les marqueurs modaux qui indiquent une prise de position des locuteurs sur des notions concernant ou bien les supports de randonnée existants (une critique) ou bien la carte idéale (une attente). Les marqueurs modaux recherchés dans ce corpus sont des verbes psychologiques : *agacer, contrarier, décevoir, déranger, ennuyer, étonner, excéder, inquiéter, intéresser, passionner, rassurer, satisfaire, surprendre* (Mathieu, 1997), ainsi que des verbes d'opinion : *apprécier, demander, désirer, manquer, nécessiter, préférer, satisfaire, souhaiter* (Généreux and Santini, 2007). La liste des segments répétés ⁶ issus du corpus a permis aussi d'orienter la réalisation des patrons morphosyntaxiques. Fig. 4 montre quelques marqueurs modaux étudiés dans ce corpus.

<i>Marqueur modal indiquant :</i>	<i>Orientation axiologique</i>	<i>Exemple</i>
Une opinion	positive / négative	- <i>je pense que</i> - <i>je trouve que</i> - <i>pour moi</i>
Une préférence	Positive	- <i>je préfère</i> - <i><avantage></i>
Un souhait	Positive	- <i>j'aimerais bien que</i> - <i>si ... ce serait ...</i>
Une critique	Négative	- <i><être> trop</i> - <i>pas assez</i>
Une évaluation	- satisfaction	- <i>j'aime bien</i> - <i>c'est bien</i>
	- satisfaction moyenne	- <i>c'est pas mal</i> ⁷
	- pas satisfaisant	- <i>il n'y a pas</i> - <i>(il) manque</i> - <i>j'aime pas/je n'aime pas</i>

Figure 4 : Classement en fonction de la valeur modale et l'orientation axiologique des marqueurs modaux les plus fréquents extraits du corpus. Les lemmes sont notés entre chevrons

Les marqueurs du tableau précédant permettent d'extraire 337 occurrences ; en ajoutant les verbes psychologiques et les verbes d'opinion cités au-dessus 435 occurrences ont été repérées. Ceci illustre le resserrement du vocabulaire utilisé par les randonneurs autour de syntagmes très courants : 57 occurrences pour *j'aime* et *j'aimerais* ; 40 occurrences pour *c'est bien* et *ce qui est bien*. Les commentaires suivants ont été extraits du corpus :

⁶ Les segments répétés sont des suites de formes dont la fréquence est supérieure ou égale à 2.

⁷ A distinguer du syntagme *pas mal*, qui renvoie à une notion de quantité.

*ce n'est pas attractif, c'est trop pastel
j'aime bien savoir où je vais
les petits circuits de 4 kms, j'aime bien
par rapport au dénivelé, à la végétation, je préfère travailler sur les cartes*

3.2.2. Extraction d'associations lexicales concernant les objets cartographiques

Le repérage des marqueurs modaux permet d'identifier les opinions des locuteurs. Pour préciser les critiques et attentes des randonneurs concernant les objets de la carte et les concepts associés, nous avons recherché des associations lexicales entre un marqueur modal indiquant une opinion, une préférence, un souhait, une critique, une évaluation du locuteur (Fig. 4) et un terme significatif pour le randonneur. Pour chaque sous-corpus, ces termes significatifs appartiennent à la liste des termes les plus significatifs présentée dans le paragraphe 3.1. Cette liste a été augmentée des termes concernant les objets cartographiques issus d'une ontologie de concepts cartographiques (Dominguès et al., 2009). Ces extractions ont été réalisées à l'aide des patrons morpho-syntaxiques d'Unitex⁸ auxquels ont été rajoutées des annotations sémantiques (par exemple, [mqClarté]) qui permettent ensuite de synthétiser les résultats obtenus avec plusieurs transducteurs repérant différents marqueurs modaux. Ici, le corpus est lemmatisé (avec Cordial Analyseur⁹) afin d'utiliser au mieux l'expressivité de ces patrons morphosyntaxiques. Un exemple de transducteur est montré en Fig. 5.

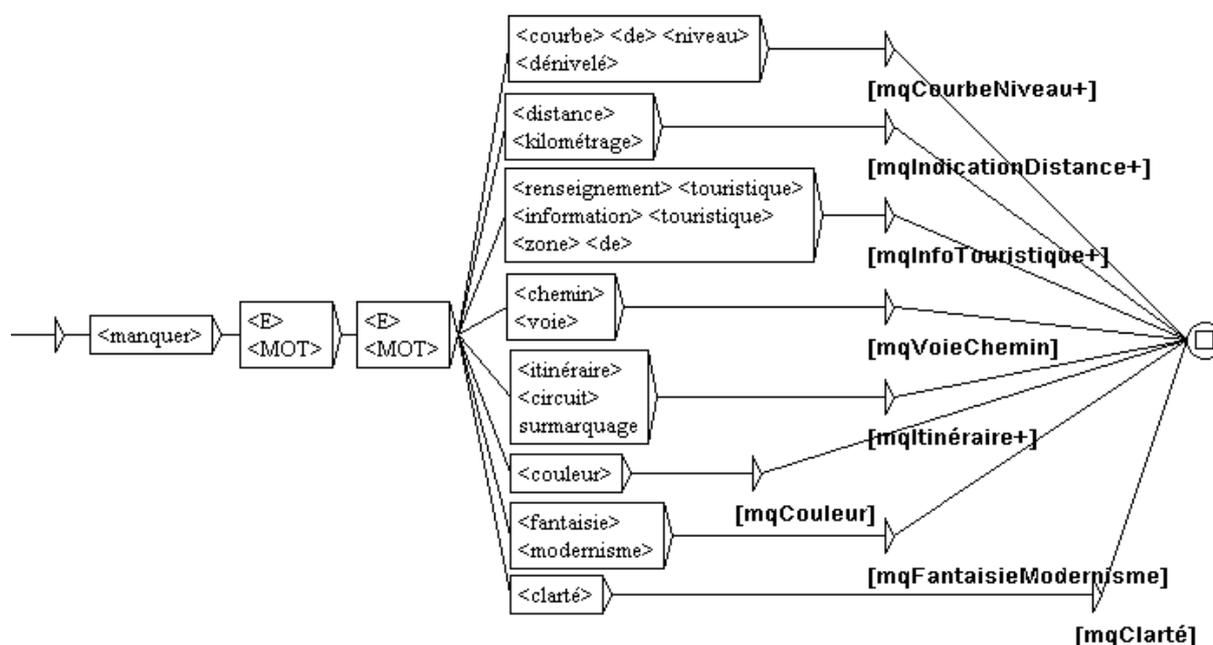


Figure 5 : transducteur permettant d'annoter les occurrences extraites par le patron :
<manquer> + <objet géographique ou caractéristique de la carte >

3.3. Synthèse des résultats : notions importantes, critiques et attentes des randonneurs

Les critiques des utilisateurs comme leurs attentes concernent à la fois le contenu informationnel de la carte (les données géographiques à sélectionner) et la manière de le cartographe (la représentation cartographique de ces données). Elles sont présentées dans Fig. 6. La structuration

⁸ www-igm.univ-mlv.fr/~unitex/.

⁹ http://www.synapse-fr.com/Cordial_Analyseur/Presentation_Cordial_Analyseur.htm.

du tableau est fournie par les questions posées aux différents groupes, ou les thèmes évoqués spontanément par un randonneur et repris par les autres randonneurs du groupe. Les termes utilisés par les locuteurs sont indiqués en regard de la notion correspondante. Pour les questions concernant l'intérêt de la représentation sur la carte d'un objet géographique ou d'une notion, les signes ++, +, -, -- quantifient l'importance que les randonneurs accordent à la notion, et donc l'intérêt de sa représentation sur la carte.

<i>sous-corpus</i> : → <i>notion</i> : ↓	<i>pédestres</i>	<i>vététistes</i>	<i>cyclistes</i>	<i>occasionnels</i>
Type de randonnée	itinéraire	parcours	parcours	balade parcours
Infrastructure utilisée	sentier chemin	GR sentier	route piste chemin	sentier
Repérage		GR chemin végétation	route circuit forêt ville	point de repère
Déclivité	dénivelé	courbe [de niveau]	chevron courbe de niveau relief pente	
Echelle utilisée	1:25 000 1:50 000	1:25 000 1:50 000 1:100 000	1:100 000 1:50 000 1:25 000	1:25 000 1:50 000
Distance	+	++	++	
Quadrillage	-	-	-	-
Lisibilité	++	++	++	+

Figure 6 : Synthèse thématique des informations importantes pour les randonneurs, leurs critiques et leurs attentes

Infrastructure utilisée : les termes identifiés sur ce thème ne renvoient pas au même objet mais sont sémantiquement assez proches. La sémantique interprétative donne des pistes pour les différencier. En effet, le sème inhérent ¹⁰ /balisé/ est contenu dans des termes comme *circuit*, *GR*, *circuit cyclotouristique*, *piste*, alors que d'autres termes sont simplement associés au modifieur *balisé* : *sentier balisé*, *chemin balisé*. L'utilisation de ces termes est donc significative et permet de distinguer le balisage inhérent aux itinéraires sur le terrain, des indications sur la carte renseignées par la légende.

L'ensemble des randonneurs insiste sur la nécessaire mise en valeur des circuits et sentiers de randonnée pour adapter les cartes à la randonnée.

Repérage : le syntagme *point de repère* est particulièrement utilisé chez les randonneurs occasionnels, ce qui s'explique par le peu d'expérience qu'ils ont de la lecture des cartes.

¹⁰ Un sème inhérent est définitoire et donc associé par défaut à l'occurrence si le contexte ne le contredit pas.

Ils cherchent alors à avoir des repères cartographiques qui soient clairement reconnaissables sur le terrain. Ainsi, un randonneur occasionnel parle des *grands randonneurs* en disant : *[ils ont] besoin de moins de points de repère que nous*. Le repérage sur la carte est utilisé par les randonneurs aussi bien pour la préparation que pendant la randonnée.

Déclivité : cette information constitue un des éléments les plus importants pour les cyclistes. Elle est signalée par la pente (pourcentage ou chevrons). La pente n'est pas indiquée sur les cartes IGN. Si ce terme apparaît dans le corpus des cyclistes, c'est justement parce que ceux-ci déplorent cette absence. Les cyclistes font aussi référence à la carte Michelin qui utilise les chevrons pour figurer la pente : *sur les cartes Michelin les chevrons sont marqués*.

Echelle de la carte : les randonneurs pédestres utilisent la carte au 1:25 000 et éventuellement celle au 1:50 000. Ils font, beaucoup plus que les autres randonneurs, allusion à ces deux échelles, sachant que la référence fréquente au 1:50 000 est probablement due au fait que l'enquête s'oriente principalement sur les usages de cette carte. Ce type de randonneur a l'habitude d'utiliser les cartes et le fait savoir : trois fois plus d'occurrences concernant l'échelle dans leur sous-corpus par rapport à celui des randonneurs occasionnels.

Distances : les randonneurs qui pratiquent sur de grandes distances, cyclistes, vététistes et quelques randonneurs pédestres, soulignent le besoin d'indiquer des distances sur les cartes, ce qui n'est pas pour le moment mentionné sur les cartes IGN.

Quadrillage : les quatre catégories de randonneurs pensent que le quadrillage est inutile.

Lisibilité : cette propriété de la carte est évoquée par tous les groupes. D'après les travaux de (Bertin, 1967), la lisibilité est améliorée par le choix des couleurs, des contrastes de couleurs et d'un caractère adapté (taille et police) ; elle est amoindrie si la quantité d'informations est trop importante. Cependant l'expérience des lecteurs influence aussi cette perception. Pour des randonneurs occasionnels (qui ont donc peu l'habitude de lire ces cartes), améliorer la lisibilité signifie le plus souvent diminuer la quantité d'information cartographiée. Au contraire, des lecteurs plus expérimentés apprécient une plus grande quantité d'information à condition que les choix de représentation mettent en évidence (grâce aux règles de Bertin) les objets cartographiques importants pour un randonneur.

4. Discussion et perspectives

Cette enquête avait pour objectif de définir, selon les pratiques de la randonnée identifiée au préalable, les points importants, critiques et attentes des randonneurs. Pour cela nous avons analysé les transcriptions des entretiens organisés autour de questions ouvertes, en utilisant conjointement des techniques d'analyse lexicale (repérage des mots significatifs) et des outils linguistiques (repérage des marqueurs modaux). Cette méthode mixte était particulièrement bien adaptée à notre analyse. Néanmoins le traitement de ce corpus présente plusieurs difficultés d'ordres méthodologique et technique.

4.1. Sur la population enquêtée

Dans cette enquête, la mise en place de quatre sous-corpus a été dictée par la segmentation même de la population des randonneurs lors des entretiens. Cette segmentation qui privilégie l'hétérogénéité inter-groupes de randonneurs interrogés se justifie par l'objectif d'identifier les attentes des randonneurs selon leur pratique de la randonnée mais les sous-corpus correspondent à deux segmentations différentes de la population : l'une selon la fréquence de la randonnée,

l'autre selon son type, d'où une difficulté pour l'analyse contrastive. Il aurait aussi été possible de comparer les trois sous-corpus de randonneurs à pied, en VTT et en vélo par rapport au sous-corpus de randonneurs occasionnels. Cependant, les questions posées sont identiques pour tous les groupes de randonneurs, ce qui garantit l'homogénéité thématique des sous-corpus entre eux. Dans ce contexte, l'analyse statistique a permis d'identifier sans a priori les thèmes importants de la carte pour chaque type de randonneurs (Beaudoin, 2000) et une analyse manuelle fondée sur des outils linguistiques a été nécessaire pour mettre en évidence les opinions de chaque groupe.

Une perspective pourrait consister à vérifier que la segmentation en quatre groupes de randonneurs supposés représentatifs d'une pratique de la population, est légitime. Pour cela, il faudrait que les termes jugés significativement fréquents dans un sous-corpus soient effectivement discriminants par rapport à la pratique de la randonnée, c'est-à-dire qu'ils soient à la fois utilisés par les locuteurs de chaque sous-corpus et non utilisés par ceux des autres sous-corpus.

4.2. Sur le corpus

Ce corpus est constitué par la retranscription d'entretiens collectifs. La première difficulté réside dans la mise en relation de la question avec les réponses des différents locuteurs. En effet, la réponse d'un locuteur est influencée à la fois par les propos qu'il a tenus précédemment et par les réponses données par les autres locuteurs à la même question. De plus, l'animateur peut intervenir plusieurs fois sur la même question. Un hiatus s'installe alors entre la question initiale posée par l'animateur et les réponses des locuteurs successifs. En outre, un terme peut apparaître fréquemment dans un passage du texte parce qu'il a été introduit par un locuteur alors qu'il n'avait pas un lien direct avec la thématique de la question. Ces difficultés d'organisation du corpus se retrouvent dans son annotation cependant nécessaire à l'exploitation statistique. Ces difficultés ont été décrites (de Peretti, 2008 ; Brugidou et al., 2004). Une perspective consisterait à ajouter manuellement des balises quand une réponse d'un locuteur concerne plusieurs questions.

Enfin, le traitement lexicographique se trouve parfois désorienté à cause d'erreurs de transcription dues à la méconnaissance du domaine (*coordonnées Lambert* systématiquement retranscrit en *coordonnées lombaires* ; *point coté* en *point côté*). Un outil précieux pour améliorer la qualité de la transcription serait de fournir un glossaire du domaine au transcripateur.

4.3. Sur les mots composés

Les outils utilisés ici pour l'analyse du lexique du corpus ne permettent pas de repérer les mots composés. Ceux-ci doivent être identifiés manuellement à partir des occurrences des mots simples : par exemple *courbe* du sous-corpus des vététistes pour lesquels il faut ensuite réaliser une analyse en contexte pour retrouver les contextes droits *de niveau* afin de reconstituer *courbe de niveau*. Les différences de nombres d'occurrences entre *courbe* et *niveau* s'expliquent par l'appartenance de *niveau* aux composés *au niveau de* et *niveau de difficulté*. Pourtant, dans ce sous-corpus vététiste, sur les vingt occurrences de *courbe*, dix-huit sont des occurrences de *courbe de niveau*, une en est une reprise anaphorique et la dernière appartient à un syntagme non lexicalisé, et vraisemblablement erroné, *courbe de nivelé*. Enfin, *niveau* n'apparaît pas dans la liste des termes significatifs du corpus des vététistes parce qu'il n'y est pas significativement présent. En effet, il est aussi fréquent dans les autres sous-corpus (31 occurrences dans le sous-corpus des cyclistes, 18 chez les randonneurs occasionnels et 18 chez les randonneurs pédestres).

Le repérage des noms composés, très fréquents dans les langues de spécialité, s'avère donc laborieux et hasardeux.

4.4. Sur l'intérêt et les limites de l'enquête

La conception d'une carte obéit à des règles cartographiques et la mise en place d'une enquête pour définir une carte de randonnée adaptée aux pratiques des randonneurs devait répondre à des questions techniques concernant l'échelle, le quadrillage, l'indication des distances, etc. Ces préoccupations ont donc dicté les questions posées par l'animateur et auxquelles l'enquête a permis de répondre, ces questions constituant autant d'*a priori* sur les éléments importants de la carte pour les randonneurs.

Cependant, l'utilisation de règles techniques ne suffit pas à évaluer l'impression globale du lecteur qui fait aussi appel à ses goûts et compétences artistiques. Une carte donne une représentation visuelle de règles singulières mais appliquées simultanément pour cartographier des données particulières. La spontanéité des locuteurs est donc un élément précieux pour tenter d'atteindre leur ressenti à la lecture de la carte. Les méthodes statistiques d'exploitation de l'enquête permettent de faire ressortir, de manière très prévisible, les termes appelés par la question, mais aussi ceux prononcés spontanément par les locuteurs en association d'idée avec les termes de la question ou parce qu'ils témoignent de leurs centres d'intérêt. Ces termes peuvent alors être mis en évidence alors qu'ils n'étaient ni attendus, ni recherchés (Desmarais and Moscarola, 2004). Cet aspect constitue le principal intérêt de l'étude directe de ces entretiens semi-directifs.

Enfin, la définition de la carte de randonnée réalisée ici pourra alimenter l'ontologie de la cartographie utilisée dans le paragraphe 3.2.2. (Dominguès et al., 2009). Cette ontologie qui formalise les connaissances des experts a été utilisée dans cette étude pour préciser le contexte droit des marqueurs modaux recherchés. Cependant, une difficulté souvent rencontrée réside dans les différences entre les concepts et le vocabulaire manipulés par les experts et ceux des utilisateurs de cartes, le plus souvent non experts. Cette étude a permis d'identifier précisément les concepts importants et les attentes des randonneurs. Ces résultats pourront être intégrés aux connaissances de l'ontologie afin de rendre compte du point de vue de certains types d'utilisateurs : les randonneurs.

Références

- Baldit P. (2008). Traitement d'un corpus de réponses à des questions ouvertes pour définir les attentes des randonneurs par rapport aux cartes de l'IGN : une approche quantitative pour exploiter une étude qualitative. Mémoire de recherche M2 « Traductique et gestion de l'information ». TIM/INALCO.
- Beaudouin V. (2000). Statistique textuelle : une approche empirique du sens à base d'analyse distributionnelle. *Texte !* [En ligne], URL : <http://www.revue-texto.net/index.php?id=642>.
- Bertin J. (1967). *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Rééditions en 1973, puis en 1998. Paris : Editions de l'EHESS.
- Brugidou M., Mandran N., Moine M. and Salomon A-C. (2004). Méthode et stratégie d'analyse des questions ouvertes du panel électoral français. *Bulletin de méthodologie sociologique*, 82 [En ligne], Mis en ligne le 09 juillet 2008. URL : <http://bms.revues.org/index1068.html>. Consulté le 24 septembre 2009.
- Charaudeau P. (1992). *Grammaire du sens et de l'expression*. Paris : Hachette Education.

- Desmarais C. and Moscarola J. (2004). Analyse de contenu et analyse lexicale, le cas d'une étude en management public. In *Lexicometrica, Actes du colloque « L'analyse des données textuelles : de l'enquête aux corpus littéraires »*,
[En ligne] <http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema7/Texte-Moscarola.pdf>.
- Dominguès C., Christophe S. and Jolivet L. (2009). Connaissances opérationnelles pour la conception automatique de légendes de cartes. In *Actes des 20^{èmes} Journées Francophones d'Ingénierie des Connaissances (IC2009)*, 25-29 mai, Hammamet (Tunisie), pp. 253-264
- Gavart-Perret M.-L. and Moscarola J. (1998). De l'énoncé à l'énonciation : pour une relecture de l'analyse lexicale en marketing. *Recherche et application en Marketing*, vol 13, 2 : 31-47.
- Gélinas-Chebat C., Daoust F., Dufresne M., Gallopel K. and Lebel M-E. (2004). Analyse exploratoire d'entrevues de groupe : les jeunes Français et le tabac. In *Actes de JADT'04*, pp. 479-487.
- Généreux M. and Santini M. (2007). Défi : classification de textes français subjectifs. In *Actes de l'atelier de clôture du 3^{ème} DEFi Fouille de Textes*, Grenoble, pp. 85-94.
- Heitz T. (2006). Modélisation du prétraitement des textes. In *Actes de JADT'06*, pp. 499-506.
- Labbé D. (2000). Analyse des données textuelles et statistique lexicale. In *Actes de JADT'00*, Lausanne.
[En ligne] <http://www.cavi.univ-paris3.fr/lexicometrica/thema/thema1/spec1-texte1.pdf>.
- Lamalle C., Fleury S. and Salem A. (2006). Vers une description formelle des traitements textométriques, In *Actes des JADT'06*, Besançon, pp. 583-593.
[En ligne] <http://www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2006/PDF/II-052.pdf>.
- Lebart L. and Salem A. (1994). *Statistique textuelle*. Paris : Dunod.
- Mathieu Y.Y. (1997). Un classement sémantique des verbes psychologiques. *Cahiers du CIEL 1996-1997*. Publications Paris 7.
- de Peretti G. (2005). La « mise en variables » des textes : mythe ou réalité ?. *Bulletin de méthodologie sociologique*, 88. [En ligne], Mis en ligne le 08 juillet 2008.
URL : <http://bms.revues.org/index773.html>. Consulté le 21 octobre 2009.
- Reinert M. (1993). Les « mondes lexicaux » et leur logique. In *Langage et société*, Paris : Maison des sciences de l'Homme, 66, pp. 5-39.
[En ligne] www.cavi.univ-paris3.fr/lexicometrica/article/numero0/MRMondLex.html.
- Roux C. and Lefèvre A. (2004). Identification des besoins en information géographique. In *Actes de JADT'04*, pp. 969-976.
[En ligne] URL : www.cavi.univ-paris3.fr/lexicometrica/jadt/jadt2004/pdf/JADT_094.pdf.
- Urdapilleta I., Manetta C., Houix O. and Montet A. (2008). Mise au jour des représentations de parfums à l'aide d'une tâche de catégorisation et de l'analyse textuelle des verbalisations associées. In *Actes de JADT'08*, pp. 1139-1150.