

CARTE SUR MESURE CHERCHE RÈGLES DE CONSTRUCTION

par Catherine Dominguès

Laboratoire COGIT – Institut géographique national, 73 avenue de Paris
94160 Saint-Mandé

<http://recherche.ign.fr/labos/cogit/>
Courriel : catherine.domingues@ign.fr

Les données et les outils cartographiques disponibles sur Internet permettent à tout internaute de concevoir une carte. Cependant, la conception d'une carte reste un processus complexe, à la fois technique et créatif, et la qualité de cartes sur mesure est rarement bonne. Dans ce contexte, cet article décrit la mise en place d'un modèle de conception cartographique en vue de créer des outils d'aide à la conception de cartes sur mesure. Ces outils visent un public novice ou expert. Ils intègrent des connaissances appartenant à des domaines variés. Ils prennent en compte le contexte et les goûts de l'utilisateur. Notre démarche s'appuie sur la définition et l'exploitation d'un corpus textuel de référence afin d'en extraire des connaissances pertinentes qui sont ensuite formalisées dans une base de connaissances. Cette démarche est structurée selon trois axes principaux de travail qui sont présentés. Les sources d'information pertinentes sont d'abord identifiées : les connaissances en cartographie doivent être enrichies par d'autres permettant, en particulier, de prendre en compte le ressenti de l'utilisateur. La partie textuelle disponible sur support électronique de ces sources d'information forme un corpus électronique dont la taille justifie l'exploration avec des outils de traitement automatique du langage naturel. Un exemple d'exploitation d'une enquête réalisée auprès de randonneurs afin de définir les spécifications d'une carte adaptée à la randonnée est présenté. Les connaissances extraites du corpus électronique doivent être mises en forme et organisées afin de faciliter leur utilisation, d'assurer leur augmentation et leur évolution. Dans ce but, la construction d'une base de connaissances est proposée. Elle s'articule sur une ontologie, centrée sur l'utilisation de la couleur dans la construction de la légende, à laquelle est associée une base de règles de construction de la carte, elles aussi extraites du corpus. La mise en œuvre de cette base de connaissances est illustrée au travers de trois applications développées au laboratoire COGIT de l'IGN, qui permettent de construire des cartes adaptées au besoin de leurs utilisateurs et qui respectent les règles de cartographie.

1 Introduction

Le besoin de cartes sur mesure est manifeste : par exemple, un randonneur qui veut placer, sur des données topographiques, un itinéraire et des étapes personnalisés (restauration, hôtellerie, sites remarquables) ou un bureau d'études qui doit ajouter à des données topographiques les bassins versants et des informations sur les risques d'inondation. À ce besoin de cartes sur mesure répond l'explosion des outils cartographiques sur Internet¹ ainsi que la profusion des systèmes d'information géographique, qui offrent à tous des moyens techniques pour réaliser ces cartes. Cependant la qualité des cartes produites est rarement bonne : problèmes de lisibilité, piètre rendu esthétique, problèmes de compréhension du message censé être délivré par la carte sont les critiques majeures relevées par les cartographes profession-

nels. En effet, la majorité de ces outils superposent des symbolisations associées par défaut aux données de l'utilisateur sans vérifier la cohérence de ces superpositions, ou bien proposent des palettes de couleurs définies sans tenir compte du contexte d'utilisation de la carte.

La figure 1 montre un exemple de carte dans lequel le mode de représentation des données nuit à leur compréhension. L'information à représenter est quantitative et donc ordonnée : sept classes de zéro à plus de quinze saisines par région. Pour qu'il soit intuitivement compréhensible de percevoir cet ordre, il faut que la légende utilise une variable visuelle ordonnée. Or c'est la teinte (marron, gris, bleu) éventuellement combinée avec la valeur (différentes intensités de bleu) qui est utilisée ici pour cartographier les classes ordonnées. Il faut donc se

¹ Géoportail : <http://www.geoportail.fr/> Google Maps : <http://maps.google.fr/>

référer à la légende pour savoir que la quantité représentée (i.e. le nombre de saisines) est par exemple supérieure en région Rhône-Alpes à celle de la Bourgogne. Une symbolisation plus explicite et conforme aux règles de sémiologie graphique de (Bertin 1967) serait d'utiliser la variable visuelle valeur pour construire un dégradé d'une même teinte dont les variations d'intensité ordonnées traduiraient les variations du nombre de saisines.

La conception d'une carte reste donc un processus complexe à la fois technique - soutenu par une théorie cartographique particulièrement riche (Bertin 1967, Béguin et Pumain 2000, Cuenin 1972, Robinson 1952, Monmonnier 1991) - et créatif - requérant des aptitudes artistiques (Krygier 1995, Rekacewicz 2006).

Dans ce contexte, notre travail s'attache à proposer un modèle de conception cartographique en vue de créer des outils d'aide à la conception de cartes sur mesure. Ces outils visent un public novice ou expert. Ils prennent en compte le contexte et les goûts de l'utilisateur. Ils intègrent des connaissances appartenant à des domaines variés. L'objectif de cet article est de décrire la mise en place de ce modèle de conception cartographique. Notre démarche s'appuie sur la définition et l'exploitation d'un corpus textuel de référence afin d'en extraire des connaissances pertinentes qui sont ensuite formalisées dans une base de connaissances. Cette démarche est structurée selon trois axes principaux de travail qui sont présentés ici. Les sources d'information pertinentes utilisées pour la mise en place de la base de connaissances sont précisées dans la 2^e partie. L'extraction, à l'aide d'outils de traitement automatique du langage naturel, des connaissances contenues dans ces sources est exposée au travers d'un exemple dans la 3^e partie. Dans la 4^e partie, la construction d'une base de connaissances est proposée afin d'organiser ces connaissances : une ontologie associée à une base de règles déductives. L'utilisation de la base de connaissances est illustrée au travers de trois applications développées au laboratoire COGIT de l'IGN qui sont présentées dans la 5^e partie, avant de conclure.

2 Des connaissances pertinentes pour la construction de cartes sur mesure

La mise en place de connaissances pertinentes pour la construction semi-automatique de cartes sur mesure passe d'abord par l'identification de ces connaissances. Le processus simplifié de conception assistée, illustré par la figure 2, permet de recenser ces connaissances et leurs conditions d'exploitation sont présentées en fin de paragraphe.

2.1 Les connaissances pertinentes

Dans l'exemple de la figure 2, l'utilisateur de l'outil décrit son besoin de carte : il souhaite dessiner une *carte routière* possédant une propriété : *joyeuse*. Ces critères de description appartiennent au vocabulaire de l'utilisateur mais pas nécessairement à celui, normalisé, de l'application. Pour réaliser l'étape 1, i.e. traduire le besoin de carte exprimé par l'utilisateur dans le vocabulaire contrôlé de l'application, des connaissances linguistiques, cartographiques, des connaissances sur les propriétés esthétiques et symboliques des couleurs doivent être mises en œuvre. Des connaissances sur les perceptions des couleurs et des contrastes colorés, ainsi que sur le ressenti du lecteur à la lecture des assemblages de couleurs que propose la carte, sont aussi nécessaires. Grâce à ces connaissances, la propriété *joyeuse* est traduite en une propriété visuelle de la carte : *de teinte dominante jaune d'or*.

L'étape 2 a pour objectif d'interpréter le besoin, traduit dans le vocabulaire contrôlé de l'application, en spécifications de cartes. Pour cela, des connaissances en cartographie, sémiologie graphique et des connaissances sur les propriétés des couleurs doivent être mobilisées ; *carte lumineuse* est ainsi traduit en *carte dont les couleurs dominantes sont dans des teintes jaunes ou orange et d'intensité moyenne*. Une proposition de carte peut alors être dessinée par l'application, puis évaluée à la fin de l'étape 2 grâce à des connaissances concernant la manipulation des variables visuelles (Bertin 1967, Itten 1985), et en particulier l'évaluation des contrastes colorés (Chesneau 2006, Buard et Ruas 2009).

Les connaissances en cartographie, en sémiologie graphique, celles concernant la manipulation des variables visuelles sont exposées dans des ouvrages didactiques :

- manuels de cartographie : (Alhinac 1959, Cuenin 1972, Béguin et Pumain 2000) ;
- cours de cartographie topographique : (Weger 1998) ;
- traités de sémiologie graphique : (Bertin 1967, Denègre 2005) ;
- dictionnaires : (*Glossaire de cartographie* 1990, *Vocabulaire de la topographie* 1980) ;
- recueil de pratiques, méthodes de cartographie thématique : (Brunet 1987, Le Fur 2007, Zanin et Trémélo 2003).

Les connaissances pertinentes pour la construction de carte et concernant la couleur appartiennent donc à des champs de connaissance variés. La couleur *y* est décrite selon différents points de vue qui

sont tous sagaces et utiles pour proposer des gammes colorées adaptées à la construction d'une légende de carte sur mesure. Ces connaissances sont exposées dans :

- des ouvrages exposant différents points de vue sur la description de la couleur et des phénomènes colorés : (Dérivé 1964, Sève et al. 2007) ;
- des ouvrages décrivant les propriétés physiques des couleurs : (Fairchild 2005) ;
- des dictionnaires des couleurs et de leurs propriétés esthétiques et symboliques (Sève et al. 2007, Mollard-Desfour 1998, Pastoureau et Simonnet 2005) ;
- des manuels d'utilisation de la couleur dans la publicité, le design, l'ameublement (Sawahata 2001) ;
- des ouvrages ou des sites médicaux décrivant les propriétés physiologiques de l'œil mobilisées pour la perception des couleurs, ses variations et les anomalies associées (SNOF 2009) ;
- des ouvrages décrivant la manipulation des couleurs dans le contexte d'une pratique artistique (Itten 1985, Carron 2009) ;
- des ouvrages littéraires qui donnent une interprétation esthétique ou lyrique de la perception colorée comme (Anouilh 1946) : « Tout était gris. Maintenant, tu ne peux pas savoir, tout est déjà rose, jaune, vert. C'est devenu une carte postale. Il faut te lever plus tôt, nourrice, si tu veux voir un monde sans couleurs » ou (Colette 1922) qui décrit des correspondances entre la perception de la couleur et d'autres sens : « Mais peut-être ne retrouvera-t-elle pas sa subtilité d'enfant, et la supériorité de ses sens qui savent goûter un parfum sur la langue, palper une couleur et voir - «fine comme un cheveu, fine comme une herbe» - la ligne d'un chant imaginaire » ou (Baudelaire 1857) qui expose une autre pratique de la synesthésie :

« Les parfums, les couleurs et les sons se répondent.
Il est des parfums frais comme des chairs d'enfants,
Doux comme les hautbois, verts comme les prairies,
Et d'autres, corrompus, riches et triomphants » ;

- des classiques de la littérature enfantine (Lionni 1959) qui mettent en place des bases sur lesquelles l'individu va construire la perception et l'interprétation des couleurs qui lui sont propres.

À l'abondance des ouvrages traitant d'un ou plusieurs aspects de la couleur s'oppose la pénurie des sources d'information décrivant le ressenti des utilisateurs lors de l'utilisation des cartes. De telles investi-

gations sont rares et leurs résultats souvent difficiles à formaliser. Pour pallier ce manque, plusieurs enquêtes ont été réalisées à l'IGN, qui visent à recueillir les commentaires des lecteurs de cartes et à formaliser leur ressenti :

- sur les propriétés visuelles de la carte (Dominguès et Bucher 2006) ;
- sur les associations d'idées mises en place à la lecture de cartes (Dominguès 2008) ;
- sur les spécifications des cartes de randonnée (Baldit-Schneller et Dominguès 2010).

L'ensemble de ces sources d'information totalise plusieurs milliers de pages.

2.2 Exploiter les sources d'information pertinentes

Ces sources d'information sont donc foisonnantes, mais aussi évolutives. En particulier, il est légitime de vouloir étendre le corpus didactique à d'autres ouvrages ou à d'autres champs de connaissances, ou bien d'ajouter de nouvelles enquêtes utilisateurs au corpus. Pour exploiter ces sources d'information, il paraît opportun d'utiliser des méthodes et des outils de traitement automatique du langage naturel (TALN) parce qu'ils permettent d'exploiter des ensembles textuels de grande taille et facilitent la capitalisation des connaissances extraites de ces textes pour les modifier et les enrichir. Ces outils nécessitent de disposer d'une version électronique des textes. Nous avons donc constitué un corpus cartographique électronique composé des textes pertinents (cf. §2.1) et disponibles sur support électronique. Il compte environ 1 400 pages.

Des outils exploitent le lexique des textes. Par exemple, la fréquence d'utilisation des mots permet d'extraire les termes significatifs d'un texte ou d'un groupe de textes ; les collocations de mots permettent d'identifier des termes liés sémantiquement. Ces outils tirent aussi parti de l'organisation des phrases dans les textes pour repérer, par exemple, la manière d'exprimer un conseil, une critique, un ordre, une règle.

3 Extraction des connaissances

Nous souhaitons extraire des connaissances des sources d'information décrites dans la 2^e partie. Cette extraction s'appuie sur des outils de TALN. Le mode opératoire mis en place pour exploiter les documents électroniques est décrit dans ce paragraphe au travers d'un exemple : la définition des spécifications des cartes de randonnée à partir d'une enquête réalisée auprès de randonneurs.

3.1 Objectifs de l'enquête

Une enquête a été réalisée par l'IGN auprès de différents groupes de randonneurs afin d'exprimer leurs critiques vis-à-vis des supports existants, dans le but de préciser les spécifications d'une carte adaptée à la randonnée. (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 par les utilisateurs. Notre démarche consiste à mettre en évidence :

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

Il s'agit de combiner une analyse quantitative du corpus de l'enquête - fondée sur les données lexicales (Lebart et Salem 1994) : vocabulaire employé, richesse lexicale, occurrences, etc. - avec une analyse qualitative, en principe dépourvue d'a priori quant à la détection des besoins (Roux et Lefèvre 2004). La figure 3 précise les différentes étapes de cette démarche.

3.2 Description de l'enquête

Dans l'enquête réalisée, 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 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 un 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 guidés par l'animateur ; celui-ci dispose d'un guide d'animation qui récapitule l'ensemble des questions auxquelles l'entretien est censé apporter des réponses. Le format des réponses est libre : les participants peuvent adopter le vocabulaire et la syntaxe qui leur conviennent ; ils peuvent aussi aborder des thèmes qui ne figurent pas explicitement dans les questions. Cette phase correspond à l'étape 1 de la figure 3.

3.3 Constitution du corpus

Les entretiens sont enregistrés. Questions et réponses sont ensuite retranscrites par écrit par l'animateur du groupe. La retranscription de ces entretiens forme le corpus d'étude et constitue une enquête qualitative à partir de questions ouvertes (Brugidou et al. 2004 ; Gélinas-Chebat et al. 2004).

Pour que ce corpus puisse être manipulé avec des outils de TALN, il doit être prétraité. Ce prétraitement recouvre différentes tâches, en particulier :

- **La correction des interprétations et des fautes de frappe et d'orthographe** engendrées lors de la transcription des entretiens oraux ;
par exemple :

	doit être corrigé en
<i>coordonnées lombaires</i>	<i>coordonnées Lambert</i>
<i>point côté</i>	<i>point coté</i>
<i>déclinité</i>	<i>déclivité</i>
<i>bonen</i>	<i>bonne.</i>

- **La normalisation de la graphie**

Pour que les outils de TALN puissent opérer, il est nécessaire d'unifier les différentes désignations d'un même objet. Par exemple, les randonneurs font souvent référence à l'échelle de la carte mais la désignent sous différentes formes comme : *la 25*, *la carte au 1/25* qui sont alors remplacées dans le corpus par *la carte au 1 : 25 000*.

Ces tâches sont réalisées manuellement mais guidées par des outils lexicaux. Par exemple, les mots contenant des fautes de frappe ressortent, le plus souvent, comme inconnus dans les dictionnaires électroniques utilisés par les outils de TALN. C'est de cette manière que peuvent être repérées les erreurs de frappe comme *déclinité* ou *bonen*. Les erreurs qui reviennent à mettre un mot à la place d'un autre, *lombaire* pour *Lambert* ou *point côté* pour *point coté*, sont plus difficiles à détecter automatiquement parce qu'elles sont liées au sens des mots dans un contexte, notion plus difficilement accessible à des outils automatiques. Cette phase correspond à l'étape 2 de la figure 3.

Finalement, le corpus prétraité est formé de quatre sous-corpus (un par entretien et donc par type de randonnée) et compte plus de 46 000 mots, soit plus d'une centaine de pages.

3.4 Objets, notions et concepts importants pour les randonneurs

L'enquête fournit des commentaires de randonneurs différenciés selon leur pratique de la randon-

née. Il est donc possible, et pertinent, d'identifier, selon chaque pratique de la randonnée, les objets, notions et concepts importants pour construire des cartes de randonnée adaptées à chaque pratique. Pour cela, une analyse contrastive du vocabulaire employé par chaque type de randonneur a été mise en place (étape 3 de la figure 3). Les mots «significativement fréquents» par rapport à un corpus de référence ont été recherchés dans chaque sous-corpus. Cette méthode relève de l'analyse lexicale descriptive (Labbé 2000 ; Beaudoin 2000). Elle s'appuie ici sur l'outil Lexico3². Il s'agit de quantifier les fréquences des unités lexicales³. Une forme est dite caractéristique d'une partie si sa fréquence est «anormalement élevée» dans cette partie. Cette fréquence s'observe par rapport à un seuil de spécificité et à un corpus de référence. Ici, les parties étudiées sont successivement chacun des sous-corpus, et le corpus de référence est formé de l'ensemble des quatre sous-corpus. La structure grammaticale des sous-corpus et celle du corpus de référence étant comparables, les mots-outils, tels que les prépositions, déterminants, auxiliaires grammaticaux, conjonctions de coordination ou de subordination, ne peuvent apparaître comme des mots significativement fréquents. Le corpus de référence ayant pour thème la randonnée, les mots appartenant au domaine de la randonnée mais non spécifiques d'une certaine pratique sont aussi éliminés. Les résultats de cette analyse contrastive sont présentés en figure 4.

3.5 Critiques des cartes de randonnée existantes et attentes des randonneurs

Des cartes existent déjà pour aider les randonneurs à préparer et réaliser leur randonnée. Un des objectifs de l'enquête est d'évaluer ces outils : il s'agit donc aussi de recueillir et d'analyser, par type de pratique et donc dans chaque sous-corpus, les critiques des randonneurs et leurs attentes quant à un support de randonnée mieux adapté à leur pratique de la randonnée. La démarche développée ici relève de l'analyse lexicale interprétative. Les modalités élocutives «précisent la manière dont le locuteur révèle sa position vis-à-vis du propos qu'il énonce» (Charaudeau 1992). La méthode (étape 4 de la figure 3) 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 les plus fréquents dans l'ensemble du corpus sont les suivants :

- pour marquer une opinion : *je pense que, je trouve que, pour moi ;*
- pour indiquer une préférence : *je préfère ;*
- pour indiquer un souhait : *j'aimerais bien que, si ... ce serait ... ;*
- pour indiquer une critique : *<être> trop, pas assez ;*
- pour indiquer une évaluation positive : *j'aime bien, c'est bien ;* une satisfaction moyenne : *c'est pas mal ;* une évaluation négative : *il n'y a pas, (il) manque, j'aime pas/je n'aime pas.*

Ils ont permis d'extraire 337 occurrences, par exemple :

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

3.6 Synthèse des résultats : critiques et attentes des randonneurs concernant les notions importantes représentées sur la carte

Les critiques et les attentes des randonneurs comme les notions importantes que doit représenter la carte sont synthétisées dans le tableau de la figure 5. Elles 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). Les résultats sont analysés à la suite du tableau.

3.6.1 Infrastructure utilisée et balisage

Les termes identifiés sur ce thème ne renvoient pas tous au même objet mais sont sémantiquement assez proches. La notion de balisage s'ajoute à celle d'infrastructure utilisée pour la randonnée ; elle est inhérente à des termes comme *circuit, GR, circuit cyclotouristique, piste* ou ajoutée sous la forme du modifieur *balisé* à des termes comme *sentier* ou *chemin* pour former les composés : *sentier balisé* ou *chemin balisé* qui figurent dans le corpus. L'ensemble des randonneurs insiste sur la nécessaire mise en valeur des circuits et sentiers de randonnée pour mieux adapter les cartes à la randonnée.

3.6.2 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 lec-

2 <http://www.cavi.univ-paris3.fr/llpga/ilpga/tal/lexicoWWW/>

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

ture des cartes. Ils cherchent 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 de la randonnée que pendant son déroulement.

3.6.3 Déclivité

La déclivité d'un parcours constitue un des éléments les plus importants pour les cyclistes. Elle est identifiée par la pente qui n'est pas indiquée sur les cartes IGN. Ce terme apparaît dans le corpus des cyclistes, justement parce que ceux-ci déplorent cette absence. En parallèle, les cyclistes utilisent avec une fréquence significativement élevée le terme *Michelin* ; en effet, sur ses cartes, Michelin utilise les chevrons pour figurer le sens et la raideur de la pente.

3.6.4 Échelle 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 (la référence fréquente à l'échelle 1 : 50 000 est probablement due au fait que l'enquête était orientée principalement sur les usages de cette carte). Ce type de randonneur a l'habitude d'utiliser les cartes : il y a trois fois plus d'occurrences concernant l'échelle dans leur sous-corpus par rapport à celui des randonneurs occasionnels.

3.6.5 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. Cette information ne figure pas pour le moment sur les cartes IGN.

3.6.6 Quadrillage

Les quatre catégories de randonneurs pensent que le quadrillage est inutile.

3.6.7 Lisibilité

Cette propriété de la carte est évoquée par tous les groupes. D'après les travaux de (Bertin 1967), la lisibilité d'une carte est améliorée par le choix des couleurs, des contrastes de couleurs et d'un lettrage 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 les objets cartographiques importants pour un randonneur.

4 Organisation des connaissances extraites

L'exploitation des sources d'information fournit des connaissances. Celles-ci doivent être organisées afin d'être utilisables par des applications informatiques d'aide à la conception de cartes sur mesure. En outre, ces connaissances pertinentes sont susceptibles d'augmenter, d'être modifiées et élargies. Leur organisation doit garantir cette évolution. Nous présentons ici l'organisation que nous avons retenue pour répondre à cette problématique : une base de connaissances formée d'une ontologie de concepts⁴, baptisée Ontocarto, associée à un ensemble de règles. Le consensus préalable à la construction de l'ontologie et la formalisation induite par sa construction garantissent la cohérence de l'ensemble des concepts et leur évolution. En particulier, la formalisation des concepts liés à la description des différents points de vue sur la couleur est un point d'intégration crucial des différents champs de connaissance que nous avons décrits au paragraphe 2. La formalisation dans l'ontologie des outils graphiques existants et utilisés pour la conception manuelle de légendes, comme les cercles chromatiques par exemple, repose sur cette notion fondamentale de couleur. Ces concepts sont présentés dans les paragraphes suivants.

4.1 Différents points de vue sur la couleur

L'acception courante du mot *couleur* est trop polysémique et imprécise pour être utilisable dans l'ontologie. Cependant, les connaissances esthétiques et symboliques concernant les couleurs utilisent ce terme en ignorant son imprécision. Par exemple, la littérature utilise des noms simples comme *vert* ou composés comme *couleur jaune* pour définir des propriétés : (Pastoureau et Simonnet 2005) relèvent que *la couleur jaune est associée au mensonge ou le vert représente la chance, et aussi la malchance* sans préciser l'intensité concernée. Le concept *CouleurGénérique* introduit dans l'ontologie rend

4 Une ontologie est un ensemble structuré de termes (et des concepts associés à ces termes) qui modélise un domaine de connaissances. Les termes, et les concepts, sont liés par différents types de relation. Une ontologie constitue un support au raisonnement sur les objets du domaine et les relations existant entre objets.

compte de cette acception courante du terme *couleur* et permet de traduire ces propriétés. Les sous-classes de *CouleurGénérique* sont les couleurs désignées par (Mollard-Desfour 1998) : *Bleu, Blanc, Brun [synonyme : Marron], Gris, Jaune, Noir, Orange, Rose, Rouge, Vert, Violet*.

En sémiologie graphique, la notion de couleur renvoie à deux variables visuelles dont les concepts sont retenus dans l'ontologie : *Teinte* et *Valeur* qui sont subsumés⁵ par *VariableVisuelle*. Selon les sources d'information étudiées, le terme *couleur* renvoie à la teinte ou à sa valeur. Par sa conception, un cercle chromatique fournit de fait un codage et permet de désigner sans ambiguïté une couleur par l'intermédiaire de sa teinte et de sa valeur. D'autres codages de couleurs sont aussi référencés dans l'ontologie : *CodageRGB*⁶, *CodageTSL*⁷, *CodageCIELab*⁸ ainsi que les algorithmes de passage d'un codage à l'autre pour les couleurs descriptibles dans plusieurs codages.

Le concept *CouleurPhysique* renvoie à la définition physique d'une couleur, c'est-à-dire une longueur d'onde, un intervalle de longueurs d'ondes (pour une couleur monochromatique) ou une union d'intervalles de longueurs d'ondes (pour les couleurs usuelles non monochromatiques).

4.2 Cercle chromatique

Le cercle chromatique de (Chesneau 2006) est modélisé dans Ontocarto et représenté dans la figure 10. Il est composé de douze teintes (concept *CercleChromatiqueTeinte* équivalent au concept *CercleChromatiqueQuartier*) dont la valeur varie en sept paliers (concept *CercleChromatiqueValeur* équivalent au concept *CercleChromatiqueSecteur*). À ce cercle principal sont ajoutés deux cercles de couleurs grisées et de gris colorés adaptés à la cartographie de données d'importance secondaire.

Le concept *CercleChromatique* est relié à *CercleChromatiqueQuartier* par une relation *EstComposéDe* ; une même relation *EstComposéDe* relie *CercleChromatiqueQuartier* et *CercleChromatiqueSecteur*.

Afin de pouvoir leur associer les mêmes propriétés qu'à une *CouleurGénérique*, *CercleChromatiqueQuartier* et *CercleChromatiqueCouleur* sont subsumés par *CouleurGénérique*. La figure 6 montre l'organisation permettant d'articuler les concepts liés à la couleur et ceux du cercle chromatique. Les relations représen-

tées sont des relations de subsomption, à l'exception des relations de composition entre *CercleChromatique* et *CercleChromatiqueQuartier* (un *CercleChromatique* est composé de *CercleChromatiqueQuartier*) ainsi qu'entre *CercleChromatiqueQuartier* et *CercleChromatiqueSecteur* (un *CercleChromatiqueQuartier* est composé de *CercleChromatiqueSecteur*).

Une famille de couleurs est une construction fondée sur l'acception du langage courant, par exemple : la *famille des bleus*. Elle permet de regrouper sous le même terme différentes intensités d'une même teinte ou de teintes visuellement proches dans le cercle chromatique. Dans l'ontologie, le concept *FamilleDeCouleurs* est relié au concept *CercleChromatiqueTeinte* par une relation *EstComposéDe*.

4.3 Les règles associées à l'ontologie Ontocarto

À l'instar de (Iosifescu-Enescu et Hurni 2007), nous proposons de distinguer les concepts pertinents pour la représentation cartographique des règles d'utilisation de ces concepts. Pour construire des cartes qui respectent les règles de sémiologie graphique, (Dao 2004) a proposé un système automatique qui vise à interdire les représentations non valides. Notre proposition a pour objectif de construire des légendes de carte cartographiquement correctes en simulant, à l'aide de règles qui manipulent les concepts de l'ontologie, les règles et recommandations telles qu'elles sont définies par les auteurs cités au paragraphe 2.

Dans un premier temps, nous avons choisi de travailler sur les trois catégories de règles suivantes :

- **Règles traduisant l'utilisation conventionnelle des couleurs.** Par exemple, les objets géographiques appartenant au thème *hydrographie* sont cartographiés avec un bleu («couleur» définie par une teinte et une valeur et appartenant à la *FamilleDeCouleursDesBleus*). De même, les objets du thème végétation sont cartographiés avec un vert de la *FamilleDeCouleursDesVerts*.

- **Règles traduisant les relations entre thèmes à l'aide des teintes.** Par exemple : les objets appartenant au même thème sont cartographiés avec des couleurs de même teinte ou formant un *ContrasteDeTeinte* faible. Les objets appartenant à des thèmes différents sont cartographiés avec des couleurs formant un *ContrasteDeTeinte* fort.

5 La subsomption est une relation hiérarchique entre concepts de l'ontologie : si B est subsumé par A alors B est une spécialisation de A. Dans Ontocarto, *Teinte* et *Valeur* sont des spécialisations du concept *VariableVisuelle*.

6 PASCALE D. A review of RGB color spaces ... from xyY to R'G'B'.

7 MUNSELL Color Co (1943). Munsell book of color, Baltimore. Edité par Luneau Industrie. Paris.

8 <http://www.fho-empden.de/~hoffmann/cielab03022003.pdf>.

- Règles exploitant les contrastes de valeurs.

Les contrastes de valeur permettent de traduire des relations d'ordre entre données du même thème. Les superficies relatives des objets influent également sur les choix des valeurs. Par exemple, le fond cartographique, qui occupe en général la plus grande superficie de la carte, doit être contrasté avec les objets qui s'y superposent. Pour cela, la couleur du fond doit être de faible valeur (par exemple les deux premiers niveaux de valeur des quartiers du cercle chromatique) et les objets de valeur plus forte (les intensités 5, 6 et 7 du cercle chromatique).

5 Exploitation visée de l'ontologie

La base de connaissances formalise des concepts et une organisation de ces concepts consensuelle (au moins dans l'équipe concernée du laboratoire COGIT). Plusieurs applications s'appuyant sur cette base de connaissances ont été construites. La conceptualisation qui y est proposée est reproduite dans les applications sous forme de code (écrit en l'occurrence en JAVA⁹). La prochaine étape consiste à utiliser directement la base de connaissances (ontologie plus règles) dans des raisonnements déployés par un moteur de raisonnement de type Corese (Corby et al. 2004). Il serait alors possible d'activer directement dans l'application les règles ou les groupes de règles pertinents. L'objectif est de pouvoir construire plus rapidement des applications mieux ciblées et plus flexibles en capitalisant dans la base de connaissances les concepts et les règles que ces applications pourront sélectionner et manipuler selon leurs spécifications et leurs objectifs.

Trois applications ont été conçues qui mettent en oeuvre les concepts et les règles de la base de connaissances :

(Buard et Ruas 2009) proposent de faire progresser la lisibilité d'une carte en améliorant ses contrastes colorés. Tous les contrastes sont sélectionnés, évalués et classés automatiquement. Des propositions de symbolisation sont ensuite faites à l'utilisateur afin d'améliorer les contrastes évalués comme les plus mauvais. La démarche est itérative. La figure 7 montre une carte de risque dans sa version initiale et après le traitement des contrastes. Les teintes, valeurs et surfaces respectives de tous les objets cartographiques sont prises en compte pour l'évaluation et l'amélioration des contrastes de la carte.

(Jolivet 2009) propose à l'utilisateur de créer, de manière automatique, interactive et rapide, une carte sur mesure par l'intermédiaire d'un service Web. Ce

service s'appuie sur une description formalisée de la carte qui guide l'organisation de la légende en thèmes ainsi que les choix de symbolisation. L'application propose des légendes qui respectent les règles de la base de connaissances et que l'utilisateur peut éventuellement modifier pour mieux mettre en valeur des objets qu'il juge importants. La figure 9 montre la carte au 1 : 25 000 obtenue en appliquant la symbolisation, générée automatiquement, à des données topographiques de l'Isère. Cette symbolisation prend en compte une préférence exprimée par l'utilisateur sur la visualisation prioritaire de la zone arborée. La symbolisation est donc adaptée (ici une valeur forte choisie pour le thème *forêt*) tout en respectant les règles cartographiques de la base de connaissances (ici, les règles d'utilisation conventionnelle des couleurs).

(Christophe et Ruas 2009) proposent un modèle de conception coopérative de légende qui gère des contraintes cartographiques formalisées dans la base de connaissances et les préférences que l'utilisateur exprime au cours du dialogue établi par l'application. Une stratégie consiste à proposer à l'utilisateur des palettes de couleurs extraites de toiles de maîtres. La figure 10 montre celle tirée du tableau *Montagnes à Collioure* de Derain (1905). Les couleurs de la peinture (rapprochées de celles du cercle chromatique de Chesneau par la mise en place d'une distance colorimétrique) sont associées aux thèmes de la légende en tenant compte des règles de la base de connaissances (la traduction des relations entre thèmes à l'aide des teintes) et des préférences exprimées par l'utilisateur, pour construire des propositions de légende (cf. fig. 11).

6 Conclusions et perspectives

Une carte adaptée à son contexte d'utilisation et conforme aux règles de cartographie permet que le message qu'elle est censée délivrer soit compréhensible par le lecteur. La conception automatisée, ou quasi-automatique, d'une telle carte pose de nombreux problèmes. Afin d'aider un utilisateur dans cette tâche, nous avons proposé un modèle de conception fondé sur l'utilisation d'une base de connaissances (ontologie plus règles). Pour construire cette base de connaissances, nous avons sélectionné des sources pertinentes qui ont été exploitées avec des outils de TALN. De ces sources ont été extraites des connaissances qui ont été formalisées dans la base de connaissances. Celle-ci constitue un réservoir de concepts et de règles pour la construction d'applications logicielles qui mettent ainsi à disposition des utilisateurs, novices ou experts, une expertise en cartographie. Ces applications permet-

⁹ www.java.com

tront de réaliser une carte sur mesure parce qu'elles donnent la possibilité à l'utilisateur de personnaliser la représentation de ses données (ce que n'autorisent pas les logiciels existants) et prennent en compte ses préférences et sa satisfaction. Enfin, parce qu'elles intègrent à la fois une expertise en cartographie et les préférences des utilisateurs, ces outils peuvent être proposés à un large éventail de profils : à des novices, ils apporteront d'abord l'expertise en cartographie qui leur fait défaut pour concevoir des légendes correctes ; aux experts, ils permettront d'affiner leur utilisation de la couleur en proposant des palettes colorées, riches et éprouvées.

Des travaux complémentaires s'intéressant à la traduction de l'harmonie et du contraste dans la carte toute entière sont en cours de développement. Un travail a été entamé sur l'expression du ressenti des lecteurs de carte et sa traduction à l'aide des concepts de la cartographie ; des connaissances complémentaires sur ce sujet pourront être ajoutés à l'ontologie. Celle-ci devra aussi être élargie à l'utilisation des variables visuelles autres que la teinte et la valeur pour proposer des légendes innovantes plus variées. Enfin, la définition et l'intégration dans la base de connaissances de règles décrivant plus spécifiquement le savoir-faire et l'expertise des professionnels pourraient aussi augmenter la variété des propositions de légende.

Bibliographie

- Alinhac G.**, 1959, *Cartographie théorique et technique ; Premier fascicule Eléments constitutifs de la carte et moyens d'expression cartographiques.*, Ministère des travaux publics et des transports, IGN, ENSG.
- Anouilh J.**, 1946, *Antigone*, Paris, Editions de la table ronde.
- Baldit-Schneller P., Dominguès C.**, 2010, « La carte de randonnée vue par les randonneurs », *Actes de JADT'10 (10es Journées internationales d'Analyse statistique des Données Textuelles)*, Rome (à paraître).
- Baudelaire Ch.**, 1857, *Les fleurs du mal*, Paris.
- Beaudouin V.**, 2000 « Statistique textuelle : une approche empirique du sens à base d'analyse distributionnelle », *Texto !* [En ligne], URL : <http://www.revue-texto.net/index.php?id=642>
- Beguïn M., Pumain D.**, 2000, *La représentation des données géographiques*, Paris, Armand Colin, 2^e édition.
- 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. et Salomon A-C.**, 2004, « Méthode et stratégie d'analyse des questions ouvertes du panel électoral français », *Bulletin de méthodologie sociologique*, 82 | 2004, [En ligne] Mis en ligne le 09 juillet 2008, URL : <http://bms.revues.org/index1068.html>, Consulté le 24 septembre 2009.
- Brunet R.**, 1987, *La carte, mode d'emploi*, Paris-Montpellier, Fayard-Reclus.
- Buard E., Ruas A.**, 2009, « Processes for improving the colours of topographic maps in the context of Map-on-Demand », *Actes de 24th International Cartographic Conference (ICC'09)*, Santiago (Chile)
- Carron Ph.**, 2009, *L'enseignement de la couleur*, [En ligne] URL : <http://www.creatic.fr/cic/Psm6Theo.htm>, Consulté le 10 mai 2010.
- Charaudeau P.**, 1992, *Grammaire du sens et de l'expression*, Paris, Hachette Education.
- Chesneau E.**, 2006, Modèle d'amélioration automatique des contrastes de couleurs en cartographie. Thèse de doctorat, Université de Paris Est-Marne La Vallée.
- Christophe S., Ruas A.**, 2009. « Conception de légendes créatives sur mesure », numéro spécial de la revue *Le monde des cartes* du Comité français de cartographie sur les publications de la conférence ICC'09, n°202, décembre 2009.
- Colette**, 1921, *La maison de Claudine*, Paris.
- Corby, O., Dieng-Kuntz, R., Faron-Zucker, C.**, 2004, « Querying the Semantic Web with the Corese Search Engine », *Proc. of ECAI*, IOS Press, p.705-709.
- Cuenin R.**, 1972, *Cartographie générale (tome 1) : Notions générales et principes d'élaboration*. p. 109-179. Paris. Editions Eyrolles.
- Dao H.**, 2004, « Les principes de la représentation cartographique de données géographiques. : une approche ontologique et sémiologique », *Revue internationale de géomatique*. Les ontologies spatiales. 14/2004 p. 259-283.
- Denègre J.**, 2005, *Sémiologie et conception cartographique*, Paris, Hermes.

- Déribéré M.**, 1964, *La couleur*. Rééditions, 11^{ème} édition en 2008, Paris, . Presses universitaires de France.
- Dominguès C.**, 2008, « Description de cartes géographiques », *8es journées francophones Extraction et gestion des connaissances (EGC2008) – Atelier sur la modélisation utilisateur et la personnalisation d'interfaces web*, Sophia-Antipolis (France).
- Dominguès C., Bucher B.**, 2006, « Application d'aide à la conception de légende », *Actes du colloque SAGEO*, Strasbourg.
- Fairchild M.D.**, 2005, *Color appearance models*, Chichester, England, John Wiley & sons, Ltd.
- Gélinas-Chebat C., Daoust F., Dufresne M., Gallopel K. et Lebel M-E.**, 2004, « Analyse exploratoire d'entrevues de groupe : les jeunes Français et le tabac », *Actes de JADT'04 (7es Journées internationales d'Analyse statistique des Données Textuelles)*, p.479-487.
- Iosifescu-Enescu I., Hurni L.**, 2007, « Towards cartographic ontologies or «how computers learn cartography », *23rd ICA conference*, Moscow, Russia.
- Itten J.**, 1985, *Art de la couleur*, Réédition en 2004, Dessain et Tolra.
- Jolivet L.**, 2009, « Characterizing maps to improve on-demand cartography - the example of European topographic maps ». *Actes de la conférence GISRUUK*, Royaume-Uni, Durham.
- Krygier J.**, 1995, « Cartography as an art and a science? », *The Cartographic Journal* 32, 6. p. 3-10.
- Labbé D.**, 2000, « Analyse des données textuelles et statistique lexicale », *Actes de JADT'00 (5es Journées internationales d'Analyse statistique des Données Textuelles)*, Lausanne.
- Le Fur, A.**, 2007, *Pratiques de la cartographie* (2e édition), Paris, Armand Colin.
- Lebart L., Salem A.**, 1994, *Statistique textuelle*, Paris, Dunod.
- Lionni L.**, 1959, *Little blue and little yellow*. Mulberry Paperback. (traduction française : *Petit-bleu et petit-jaune*, 1970, Paris, L'école des loisirs).
- Mollard-Desfour A.**, 1998, *Le dictionnaire des mots et expressions de couleur du XX^e siècle, Le bleu*, Rééd. 2004, Paris, CNRS Editions.
- Monmonier M.**, 1991, *How to lie with maps*. University of Chicago Press.
- Pastoreau M., Simonnet D.**, 2005, *Le petit livre des couleurs*, Paris, Ed. du Panama.
- Reinert M.**, 1993, *Les «mondes lexicaux» et leur logique : Langage et société*, Paris, Maison des sciences de l'Homme, n°66,p.5-39.
- Rekacewicz Ph.**, 2006, « La cartographie, entre science, art et manipulation ». A propos de L'Atlas 2006 du Monde diplomatique, [En ligne] <http://www.monde-diplomatique.fr/2006/02/REKACEWICZ/13169>.
- Robinson A.H.**, 1952, *The Looks of Maps*, Madison, University of Wisconsin Press.
- Roux C., Lefèvre A.**, 2004, « Identification des besoins en information géographique », *Actes de JADT'04 (7èmes Journées internationales d'Analyse statistique des Données Textuelles)*, pp.969-976, [En ligne] www.cavi.univ-paris3.fr/lexicometrical/jadt/jadt2004/pdf/JADT_094.pdf
- Sawahata L.**, 2001, *Color harmony workbook*, Rockport Publishers.
- Sève R., Indergand M., Lanthony Ph.**, 2007, *Dictionnaire des termes de la couleur*, Terra rossa, Avallon
- SNOF, site du Syndicat national des ophtalmologistes de France, [En ligne] <http://www.snof.org/>
- Weger G.**, 1998, *Cartographie*, cours de l'Ecole nationale des sciences géographiques, Paris, Institut géographique national.
- Zanin C., Trémélo M-L**, 2003, *Savoir faire une carte*, Paris, Belin.

Dictionnaires

- Comité français de cartographie, 1990, *Glossaire de cartographie, Bulletin* n°123-124, mars-juin 1990, Paris.
- Conseil international de la langue française, 1980, *Vocabulaire de la topographie*, Paris, La maison du dictionnaire, Hachette.

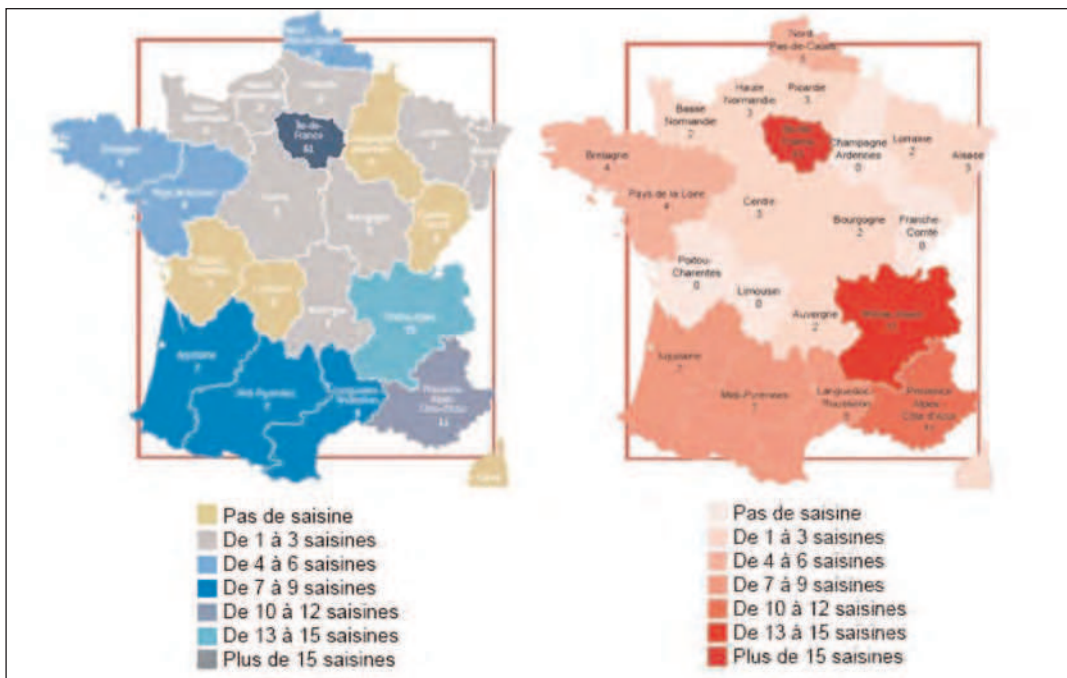


Figure 1 : Extrait de la carte illustrant l'origine des saisines traitées en 2008 par la Commission nationale de déontologie de la sécurité (http://www.la-cnds.eu/rapports/ra_pdf/Bilan_activite_2008.pdf)

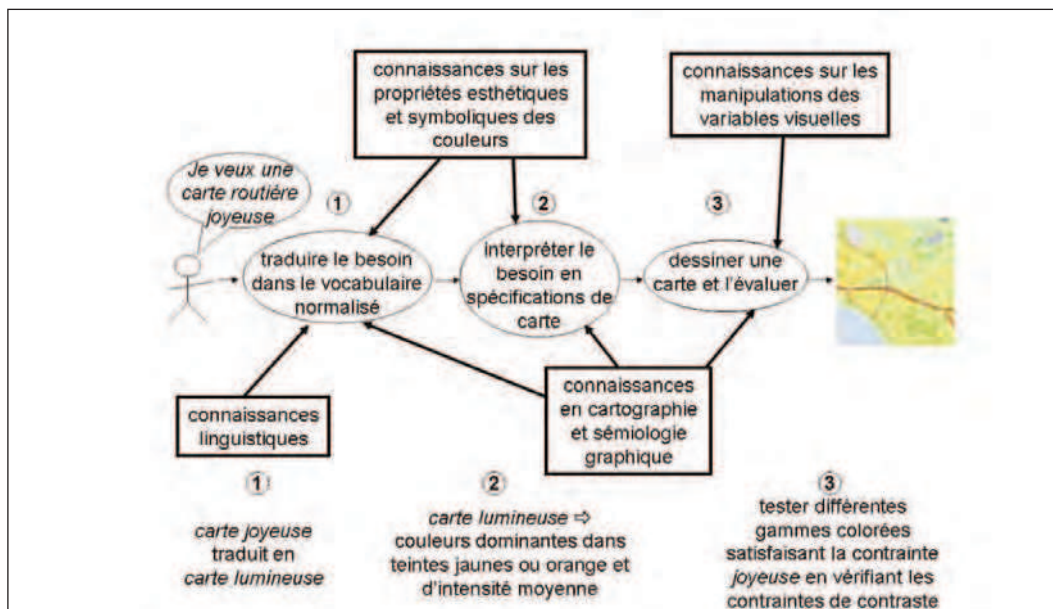


Figure 2 : Processus de conception assistée de cartes sur mesure

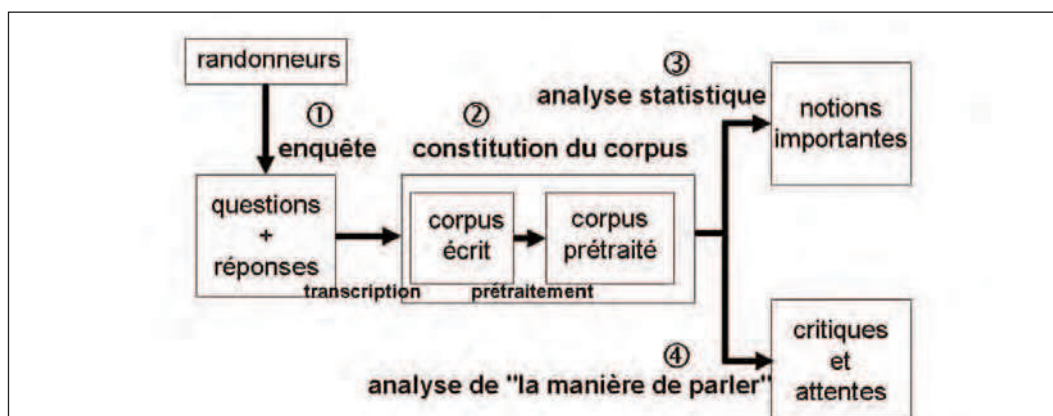


Figure 3 : Les différentes étapes de l'exploitation de l'enquête réalisée auprès des randonneurs

Pédestres	Vététistes	Cyclistes	Occasionnels
[carte au] 1:50 000	[carte au] 1:25 000	chevron/chevrons	balade/balades
chargée	cartes	équipements	compliqué <e s es>
chemin	CD-rom	forêt/forêts	facile
dénivelé/dénivelés	clarté	Internet	fouillis
kilométrage	confiance	kilomètres	itinéraire/itinéraire
[une vue] macro	courbe/courbes [de niveau]	logiciel	jeune
pédestre	GPS	Michelin	légende
randonne	GR	pente	massif
repérage	information	route/routes	nous
sentier/sentiers	Libris	vélo	points [de repère]
terrain	végétation	ville/villes	sympa
vous	VTT	visible	trop

Figure 4 : Termes spécifiques du vocabulaire de chaque sous-corpus (GPS est l'abréviation normalisée de Global Positioning System, GR celle de sentier de grande randonnée). Les termes spécifiques du vocabulaire de chaque catégorie de randonneurs ont été regroupés thématiquement (en utilisant le guide d'animation ou des connaissances en cartographie, externes à l'enquête) ; par exemple : carte, logiciel, internet sont des termes relatifs au support de la randonnée ; dénivelé, chevron, courbe de niveau concernent la déclivité du parcours. Les cases relevant d'un même thème sont représentées avec la même texture.

notion :	sous-corpus : pédestres	vététistes	cyclistes	occasionnels
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	
échelle utilisée	1:25 000 1:50 000	1:25 000	1:25 000 1:50 000 1:100 000	1:25 000 1:50 000
distance	+	++	++	
quadrillage	-	-	-	-
lisibilité	++	++	++	+

Figure 5 : Synthèse des informations importantes pour les randonneurs, leurs critiques et leurs attentes. Le tableau a été structuré par thèmes (les mêmes que ceux de la figure 4 pour les thèmes communs) ; les termes spécifiques à chaque groupe sont indiqués en regard du thème correspondant. 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. Les codes de texture sont identiques à ceux de la figure 4.

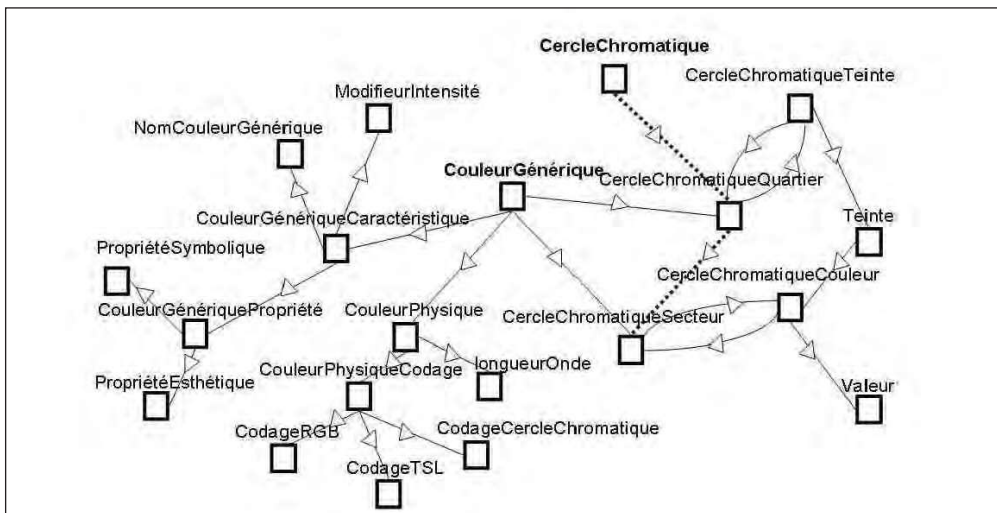


Figure 6 : Organisation des concepts liés à la couleur et au cercle chromatique. Les flèches en pointillés correspondent à une relation EstComposéDe.

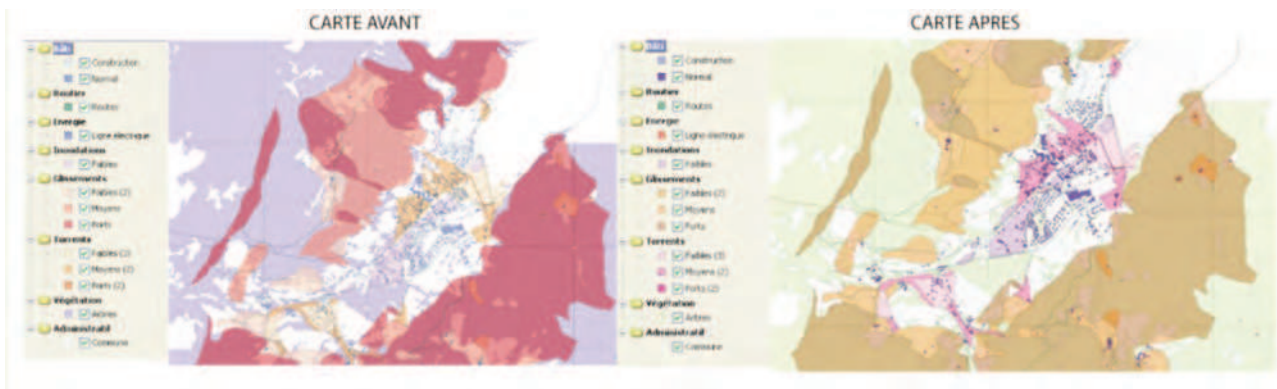


Figure 7 : Une carte de risque avant et après l'amélioration de certains contrastes colorés

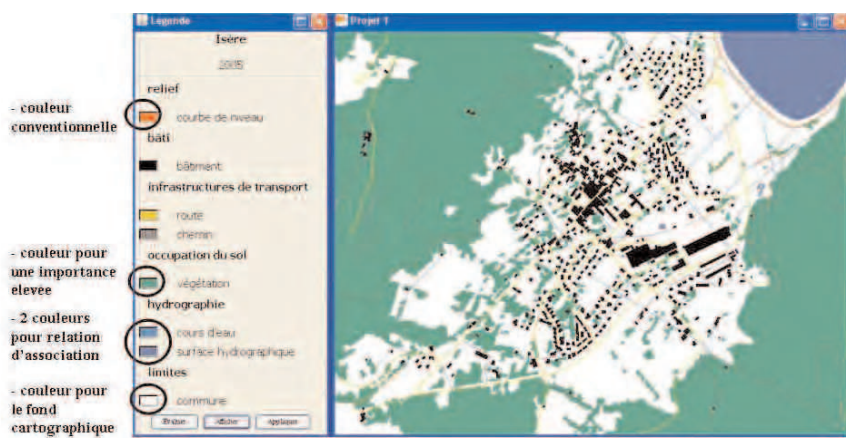


Figure 8 : Exemple d'une conception de légende de carte topographique à partir des choix d'un utilisateur

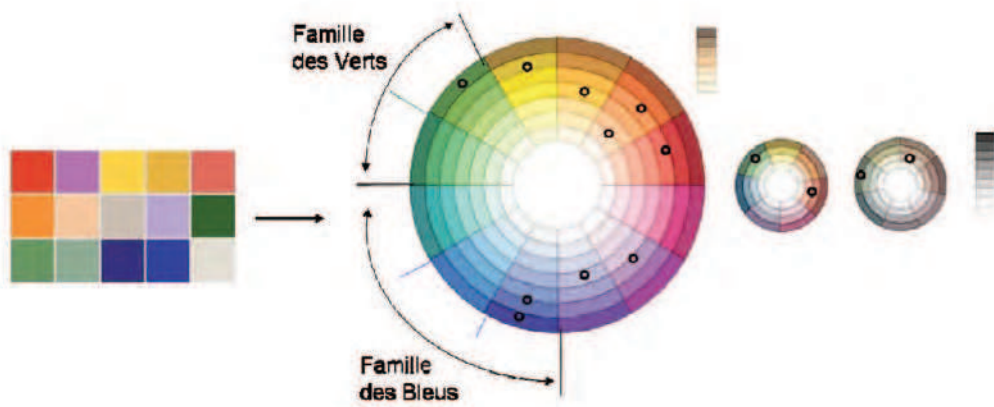


Figure 9 : Rapprochement de la palette Derain et des couleurs du cercle chromatique



Fig. 10 : Propositions de légende à partir de la palette Derain