



Construction semi-automatique d'une ontologie de la perception des paysages

Julie Bourbeillon, Laëtitia Piel, Ramla El Ayeb, Damien Rousselière, Thomas Guyet

► To cite this version:

Julie Bourbeillon, Laëtitia Piel, Ramla El Ayeb, Damien Rousselière, Thomas Guyet. Construction semi-automatique d'une ontologie de la perception des paysages. IC - 24èmes Journées francophones d'Ingénierie des Connaissances, Mar 2013, Lille, France. 2013. <hal-00920722>

HAL Id: hal-00920722

<https://hal-agrocampus-ouest.archives-ouvertes.fr/hal-00920722>

Submitted on 19 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Construction semi-automatique d'une ontologie de la perception des paysages

Julie Bourbeillon^{1,2,3}, Laëtitia Piel⁴, Ramla El Ayebe⁵,
Damien Rousselière^{4,5}, Thomas Guyet⁶

¹ AGROCAMPUS OUEST, UMR1345, Institut de Recherche en Horticulture et Semences, SFR 4207 QUASAV, PRES LUNAM, F-49045 Angers, France prenom.nom@agrocampus-ouest.fr

² Université d'Angers, UMR1345, Institut de Recherche en Horticulture et Semences, F-49045 Angers, France

³ INRA, UMR1345, Institut de Recherche en Horticulture et Semences, F-49045 Angers, France

⁴ AGROCAMPUS OUEST, UMR-MA49, Groupe de Recherche ANgevin en Économie et Management, F-49045 Angers, France

⁵ Université d'Angers, UMR-MA49, Groupe de Recherche ANgevin en Économie et Management, F-49045 Angers, France

⁶ AGROCAMPUS-OUEST / IRISA - UMR 6074, F-35042 Rennes, France

Résumé : Selon la Convention Européenne du Paysage, le paysage est « une partie de territoire telle que perçue par les populations, dont le caractère résulte de l'action de facteurs naturels et/ou humains et de leurs interrelations ». Nous nous intéressons à la conception d'une ontologie axée « perception par les populations » en exploitant les aspects perceptifs disponibles dans les atlas de paysage. Pour palier les faibles fréquences de ce vocabulaire, nous présentons une approche semi-automatique qui nous a permis de construire une première version de l'ontologie.

Mots-clés : ontologies, extraction à partir de textes, domaine environnemental.

1 Introduction

Les systèmes de classification tels que le Land Cover Classification System de la FAO qualifient les paysages par les objets qui les composent et leur agencement. Mais, selon la Convention Européenne du Paysage, le paysage est « une partie de territoire telle que perçue par les populations » (Conseil de l'Europe, 2000) et la perception doit être incluse dans la caractérisation des paysages.

Notre objectif est de fournir un outil pour l'interprétation des paysages, axé sur le volet « perception par les populations ». La solution proposée est de concevoir une ontologie du domaine, qui contribuera à une description objective de la perception des paysages et devrait permettre les comparaisons entre les paysages décrits et leur classification. Si d'autres approches recourent à des questionnaires (Lelay *et al.*, 2005), nous exploitons ici les "Atlas du paysages" (Soufflet-Leclerc, 2007), documents standardisés au niveau national et contenant les informations recherchées. Le défi posé par ces documents réside dans la difficulté d'identification des termes spécifiques au champ lexical de la perception.

2 État de l'art

Des ontologies ont été développées en géographie (Lepczyk *et al.*, 2008), traitant principalement d'occupation des sols et topographie (pour une liste, voir : (Minard, 2008)), mais seuls quelques travaux se sont intéressés aux ontologies du paysage. Les travaux de (Lepczyk *et al.*, 2008) exposent les concepts clés du paysage en écologie. (Cataldo & Rinaldi, 2010) se sont penchés sur le concept de « paysage culturel ». Ces ontologies abordent des aspects consensuels de la notion de paysage, mais ne sont pas ancrées dans les territoires. Cet ancrage géoréférencé consiste généralement en classifications des différents types de paysages rencontrés, par exemple en Belgique (Van Eetvelde & Antrop, 2009). Regardant la perception, les ontologies existantes sont rares. Elles relèvent plutôt de domaines proches, comme les émotions (Mathieu, 2005). Nous visons donc à proposer une ontologie de la perception des paysages, premier pas vers une qualification formalisée du domaine sensible du paysage.

Les approches actuelles de construction d'ontologies oscillent entre l'interrogation d'experts et le traitement informatisé de corpus textuels du domaine. Les méthodes automatisées sont basées sur des principes statistiques, linguistiques ou associent les deux (Bourigault & Aussenac-Gilles, 2003). Par exemple, TermoStat (Drouin, 2003) compare corpus spécialisés et non-spécialisés en vue de l'identification des termes. De plus, certains outils implémentent une méthode globale, comme Terminae (Szulman *et al.*, 2002). IRAMUTEQ (Ratinaud, 2009) et Tropes (Molette, 2009) sont utilisés conjointement. La pertinence de cette combinaison est fondée sur les différences théoriques de ces deux approches :

- lexicométrie d'inspiration benzécriste pour IRAMUTEQ/ALCESTE, *i.e.* les thématiques émergent du texte

- sémiométrie pour Tropes, *i.e.* les thématiques du texte sont confrontées à des univers posés a priori

L'association de ces deux méthodes a montré son intérêt en sociologie (Jenny, 1996).

Méthodes et outils statistiques, linguistiques ou mixtes font l'hypothèse que les termes du domaine et leurs associations sont fréquents (idée de répétition d'ALCESTE (Reinert, 2003)). Or, même si le volet perception du paysage est présent dans les documents considérés, on peut supposer qu'il n'est pas le cœur du discours et risque d'être difficile à identifier, comme nous l'avons montré dans l'analyse de rapport d'activité d'une coopérative (Rousselière & Vézina, 2009). Il nous faut donc étudier plus précisément le corpus pour vérifier cette hypothèse et adapter la méthodologie mise en place.

3 Corpus documentaire

3.1 Documents contenant des informations de perception

Dans un premier temps, un inventaire des documents sources de vocabulaire sur la perception des paysages a été réalisé. Les documents d'urbanisme, de protection des espaces naturels, documents touristiques, documents recensant des conflits d'usage (sites internet militants, actes judiciaires, etc.) présentent des éléments quant à l'attachement des individus à des lieux emblématiques mais sont biaisés par l'intérêt des intervenants. Pour réduire l'introduction de ce vocabulaire biaisé, nous nous sommes limités aux Atlas de Paysage. Ces documents sont en effet destinés « à dresser l'état des lieux des paysages et des dynamiques qui les transforment » (Luginbühl, 1994), à l'échelle du département ou de la région. La Direction de l'Architecture et de l'Urbanisme a proposé en 1994 une méthode de rédaction, qui inclut une analyse de la dimension sensible, afin que ces atlas constituent « un état de référence partagé » (Brunet-Vinck, 2004). Ces atlas s'articulent en une partie diagnostic, qui identifie et caractérise des unités paysagères, zones géographiques homogènes, et une partie prospective, qui évalue la dynamique des paysages.

3.2 Pré-traitements des atlas

Parmi les départements français, nous avons pu constituer un corpus de 48 atlas en version numérique, qui ont été transformés en format texte

Thème	Nombre de mots	Pourcentage de mots
Autres concepts	69588	17%
Géographie	66724	16%
Caractéristiques	64154	15%
Agriculture et environnement	55988	13%
Politique et société	21402	5%
Comportements et sentiments	14662	3%
Forces et quantités	14118	3%
Animaux et végétaux	12955	3%
Autres (<3%)	101396	24%
Total	420987	100%

TABLE 1 – Analyse Tropes. Les « Autres concepts » correspondent à des concepts n’entrant dans aucune autre catégorie, les « Caractéristiques » correspondent à des termes définissant les propriétés d’autres concepts, la ligne « Autres » regroupe les thèmes peu fréquents.

utilisable par des outils informatiques. Les atlas présentent un contenu iconographique important (cartes, schémas et photographies). Cette première phase conduit donc à une perte d’information. Pour la limiter, les légendes des figures ont été conservées, car elles sont susceptibles de contenir du vocabulaire “perceptif” pertinent pour notre travail.

3.3 Premières analyses

Afin de cerner la place du vocabulaire de la perception dans ce corpus, nous l’avons traité avec le logiciel Tropes. La classification du corpus selon les concepts définis dans le « scénario standard », qui correspond à une taxonomie générale, est présentée Table 1. Y prédominent les thèmes qui sont au cœur des atlas de paysage : géographie, agriculture et environnement, ainsi que leur caractérisation. Par contre, les éléments relevant de la perception sont rares : ils représentent une fraction des thèmes « Caractéristiques », « Comportements et sentiments » ou « Forces et quantités ».

Toujours avec Tropes, nous avons ensuite estimé la connexion entre termes de perception, en étudiant les co-occurrences pour quelques termes emblématiques du domaine. Ainsi, le « Paysage » est impliqué dans près de 12000 relations. Mais ces relations n’impliquent des termes de perception que pour une dizaine de pourcents en volume et seuls 5 autres termes de perception sont concernés. La « vision » n’implique qu’environ 2000 relations ou l’« ambiance » uniquement quelques centaines, et la moitié des co-occurrences ne concerne pas d’autres termes de perception.

Nous avons enfin testé des outils d'acquisition automatique de termes tels que TermoStat. Les candidats de regroupement fréquents relèvent de la géographie (vallée, coteau, bourg, etc.) et les graphes de relations syntaxiques incluent peu de termes de perception.

Ces analyses préliminaires mettent en évidence la faible fréquence, le peu de connexion et le masquage par le vocabulaire géographique des termes de perception, conduisant à un difficile traitement automatique. Nous avons donc mis en place une approche semi-automatique que nous avons appliquée sur le territoire d'étude du Maine-et-Loire.

4 Construction de l'ontologie

La méthode de construction mise en place, inspirée de (Bourigault & Aussenac-Gilles, 2003), présente quatre étapes : identification des termes pertinents, choix des critères de regroupement des termes, établissement des concepts et relations par une approche experte et enrichissement par étendue du corpus analysé.

4.1 Identification des termes pertinents

Nous avons extrait statistiquement les termes de l'atlas du Maine-et-Loire à l'aide d'IRAMUTEQ. La rareté des termes d'intérêt nous a conduit à ne pas réaliser d'élagage lié à la fréquence. Par contre, nous avons procédé à une lemmatisation et à une mise de côté des mots outils, avec la base Lexique3 (New *et al.*, 2001). Cette première extraction a conduit à une liste de 3687 termes. Nous avons ensuite éliminé automatiquement les termes non pertinents à l'aide de vocabulaires tels Eurovoc¹ et GEMET². Nous avons alors obtenu une liste de 2721 termes. Un raffinement manuel a conduit à 384 termes du domaine de la perception des paysages.

4.2 Choix de critères de regroupement

Les regroupements automatiques (clustering d'IRAMUTEQ ou candidats-termes de TermoStat) étant peu pertinents, l'ensemble des termes ont été regroupés par thématique en exploitant des rapports de synonymie, d'antonymie, de méronymie, d'hypo- ou d'hyponymie par un expert.

1. Thésaurus de l'Union Européenne, <http://eurovoc.europa.eu/>

2. GEneral Multilingual Environmental Thesaurus de l'Agence Européenne pour l'Environnement. <http://www.eionet.europa.eu/gemet>

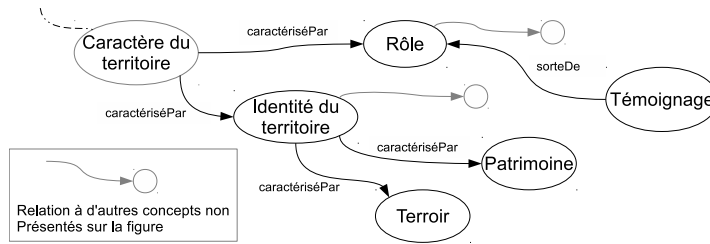


FIGURE 1 – Représentation d'une partie du concept 'Caractère'.

4.3 Établissement des concepts et de leurs relations

L'analyse des termes a abouti à 7 thèmes principaux. Prenons l'exemple du thème du *Caractère du territoire* :

- Adjectifs : spécifique, typique, particulier, local, identitaire
- Noms : caractère, témoin, trace, caractéristique, témoignage, vocation, rôle, identité, visage, terroir, faciès, patrimoine, profil
- Verbes : spécialiser, caractériser, identifier, réputer

Les sous-thèmes suivants ont été identifiés par un expert du paysage : *Caractère du territoire* (caractère, caractéristique, spécifique, typique, particulier), *Rôle* (rôle, vocation), *Témoignage* (témoin, trace, témoignage), *Identité du territoire* (identité, visage, faciès, profil), *Terroir* (terroir) et *Patrimoine* (patrimoine) où les termes en italique ont été définis comme concepts et les autres comme termes associés.

Le *Caractère du territoire* peut être caractérisé par son *Identité* et son *Rôle*. *Témoignage* est une sorte de rôle. L'*Identité du territoire* est décrite par son *Terroir* et son *Patrimoine*. Le concept de *Caractère du territoire* est donc relié aux autres comme présenté Figure 1.

Selon cette approche chaque concept a été désigné par un terme présent dans un thème ou sous-thème et associé aux autres termes du thème. Ces concepts ont ensuite été reliés par des relations sémantiques telles que « sorteDe », « caractériséPar », « entraîne ». Finalement, cette première version de l'ontologie comporte 77 concepts au total.

4.4 Enrichissement de l'ontologie

Nous avons complété l'ontologie en étendant l'analyse aux départements limitrophes (Indre-et-Loire, Loire-Atlantique et Sarthe), où se prolongent certaines unités paysagères du Maine-et-Loire. Leur description

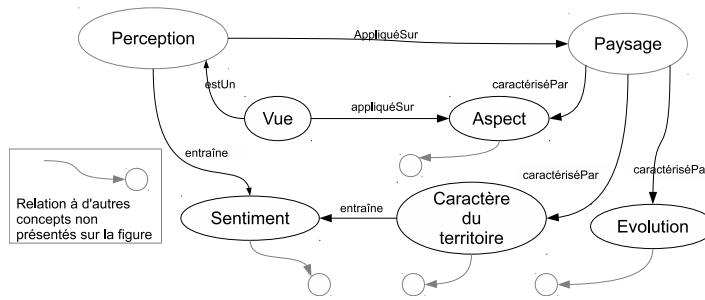


FIGURE 2 – Partie supérieure de l'ontologie de perception du paysage.

dans ces atlas a été traitée selon la méthode précédente (cf. section 4.1 à 4.3), conduisant à une liste de 316 termes déjà identifiés et de 247 termes supplémentaires, qui ont pu être intégrés aux concepts existants. Un travail identique a été effectué sur l'Atlas du Morbihan, conduisant à 296 termes déjà rencontrés et 1120 termes supplémentaires. Les nouveaux termes ont enrichi les concepts existants mais aussi entraîné l'émergence de nouveaux concepts.

4.5 L'ontologie « perception du paysage »

L'ontologie finale³ de 144 concepts présente la structure illustrée par la Figure 2. La *Perception* appliquée au *paysage* est déclinée entre autres en la *Vue* des différents *Aspects* du *Paysage*. Celle-ci entraîne des *Sentiments* causés par le *Caractère du territoire*. De plus, le *Paysage* est caractérisé par une *Evolution* temporelle. Plus précisément, l'*Aspect* du paysage englobe *Texture*, *Lumière*, *Couleur* ainsi que le niveau de *Visibilité* ou d'*Occultation* du paysage et sa *Composition* (*Formes*, *Variété* des constituants, ainsi que leur *Organisation* et leur *Echelle*), etc.

5 Application à l'annotation des unités paysagères du Maine-et-Loire

Une application de l'ontologie est l'annotation de cartes des unités paysagères, dans un objectif de classification. Il s'agit de rechercher les termes de l'ontologie dans la description de chaque unité pour y associer les concepts pertinents. Cette méthode appliquée manuellement sur l'Atlas du Maine-et-Loire a pour résultat la Figure 3. Cette carte fait apparaître des anno-

3. Disponible sur le site du projet PAYTAL. <http://www.paytal.fr>

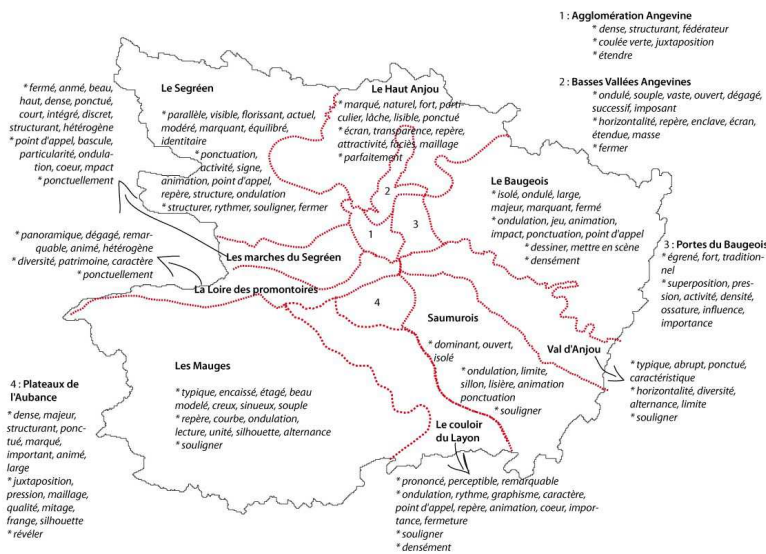


FIGURE 3 – Carte des unités paysagères du Maine-et-Loire annotée avec les concepts de l’ontologie.

tations différentes selon l’unité paysagère considérée, ce qui suggère des différences de perception du paysage selon les territoires.

6 Conclusion et perspectives

Dans cet article, nous avons présenté une méthode pour construire une première ontologie de la perception des paysages à partir d’atlas de paysages. Nous avons pallié la faible fréquence des termes pertinents dans le corpus par une approche semi-automatique et en limitant l’espace d’étude.

Une première limite est liée à cette rareté des éléments de perception. Les extractions automatiques présentent des performances limitées (précision de 12,1% et rappel de 85.9%), conduisant à une importante phase manuelle et empêchant le passage à l’échelle. Nous envisageons donc d’améliorer l’automatisation, en complétant l’élagage avec d’autres vocabulaires,

tel le WOLF⁴, et incluant les aspects morpho-syntaxiques en lien avec l'ontologie existante. Nous avons aussi en projet de constituer un corpus centré sur la perception par des enquêtes auprès des acteurs territoriaux.

Une seconde limite est le biais de subjectivité introduit par l'auteur de l'atlas, malgré les recommandations énoncées pour sa réalisation. Ce document est en effet chargé de médiatiser la/les perception(s) d'un territoire. Cette médiatisation est nécessairement une interprétation – ce qu'encourage l'approche sensible du paysage – et à ce titre susceptible de « dérives interprétatives » (Lahire, 2007). Nous explorons donc deux voies : travailler sur l'ensemble des atlas afin de savoir si la nature de l'auteur est susceptible d'influencer le type de vocabulaire utilisé et exploiter d'autres documents sur un même lieu pour identifier la nature et l'ampleur du biais spécifique à l'atlas de paysage.

Finalement, l'utilisation possible de l'ontologie pour créer des indicateurs paysagers viendra enrichir des données sur les sols, le relief, l'hydrologie, les sites classés et les sites protégés, les activités agricoles, les bassins de vie de l'INSEE... et pourra permettre ainsi une meilleure compréhension des phénomènes qui agissent sur les territoires.

Références

- BOURIGAULT D. & AUSSENAC-GILLES N. (2003). Construction d'ontologies à partir de textes. In *Actes des 10ème conférence sur le Traitement Automatique des Langues Naturelles (TALN)*, p. 27–47.
- BRUNET-VINCK V. (2004). *Méthode pour les Atlas de paysages – Enseignements méthodologiques de 10 ans de travaux*. Rapport interne, Ministère de l'Ecologie et du Développement Durable.
- CATALDO A. & RINALDI A. M. (2010). An ontological approach to represent knowledge in territorial planning science. *Computers, Environment and Urban Systems*, **34**(2), 117–132.
- CONSEIL DE L'EUROPE (2000). *Convention Européenne du Paysage - Florence, 20 octobre 2000*. Bruxelles : Division de l'aménagement du territoire et du paysage.
- DROUIN P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, **9**(1), 99–117.
- JENNY J. (1996). Analyses de contenu et de discours dans la recherche sociologique française : Pratiques micro-informatiques actuelles et potentielles. *Current Sociology*, **44**(3), 279–290.

4. Wordnet Libre du Français, <http://alpage.inria.fr/~sagot/wolf.html>

- LAHIRE B. (2007). *L'esprit sociologique*. Paris : La Découverte / Poche.
- LELAY Y.-F., PIEGAY H. & COSSIN M. (2005). Les enquêtes de perception paysagère à l'aide de photographies. In *Actes des Rencontres de Théo Quant*.
- LEPCZYK C. A., LORTIE C. J. & ANDERSON L. J. (2008). An ontology for landscapes. *Ecological Complexity*, **5**(3), 272–279.
- LUGINBÜHL Y. (1994). *Méthode pour des atlas de paysages. Identification et qualification*. Rapport interne, Strates/CNRS - SEGESA, Direction de l'Architecture et de l'Urbanisme.
- MATHIEU Y. Y. (2005). Annotation of emotions and feelings in texts. In *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, volume 3784 of *Lecture Notes in Computer Science*, p. 350–357. Springer Berlin Heidelberg.
- MINARD A.-L. (2008). *Etat de l'art des ontologies d'objets géographiques*. Rapport interne, Laboratoire COGIT, Institut National de l'Information Géographique (IGN).
- MOLETTE P. (2009). De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste. In *Actes du colloque Psychologie sociale de la communication*.
- NEW B., PALLER C., FERRAND L. & MATOS R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUETM//A lexical database for contemporary french : LEXIQUETM. *L'année psychologique*, **101**(3), 447–462.
- RATINAUD P. (2009). IRaMuTeQ : Interface de R pour les analyses multidimensionnelles de textes et de questionnaires. <http://www.iramuteq.org/>.
- REINERT M. (2003). Le rôle de la répétition dans la représentation du sens et son approche statistique par la méthode "Alceste". *Semiotica*, **147**(1/4), 389–420.
- ROUSSELIÈRE D. & VÉZINA M. (2009). Constructing the legitimacy of a financial cooperative in the cultural sector : a case study using textual analysis. *International Review of Sociology—Revue Internationale de Sociologie*, **19**(2), 241–261.
- SOUFFLET-LECLERC E. (2007). L'identification et la qualification des paysages en france : les atlas de paysages. *Culture et Recherche*, **111**, 27–28.
- SZULMAN S., BIÉBOW B. & AUSSÉNAC-GILLES N. (2002). Structuration de Terminologies à l'aide d'outils d'analyse de textes avec TERMINAE. *Revue Traitement Automatique des Langues*, **43**(1), 103–128.
- VAN EETVELDE V. & ANTROP M. (2009). A stepwise multi-scaled landscape typology and characterisation for trans-regional integration, applied on the federal state of belgium. *Landscape and Urban Planning*, **91**(3), 160–170.