Corela

Cognition, représentation, langage

HS-2 | 2005

Le traitement lexicographique des noms propres

Description linguistique en toponymie contrastive dans une base de données multilingue

Jonas Löfström et Bettina Schnabel-Le Corre

Édition électronique

URL : <http://journals.openedition.org/corela/1167>

DOI : 10.4000/corela.1167

ISSN : 1638-573X

Éditeur

Cercle linguistique du Centre et de l'Ouest - CerLICO

Référence électronique

Jonas Löfström et Bettina Schnabel-Le Corre, « Description linguistique en toponymie contrastive dans une base de données multilingue », *Corela* [En ligne], HS-2 | 2005, mis en ligne le 02 décembre 2005, consulté le 02 avril 2021. URL : <http://journals.openedition.org/corela/1167> ; DOI : <https://doi.org/10.4000/corela.1167>

Ce document a été généré automatiquement le 2 avril 2021.

Corela – cognition, représentation, langage est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Description linguistique en toponymie contrastive dans une base de données multilingue

Jonas Löfström et Bettina Schnabel-Le Corre

Introduction

- 1 Une raison évidente pour étudier les noms propres (et particulièrement les toponymes) dans une perspective lexicographique est le fait que le nom propre occupe une place à part dans le système d'une langue donnée mais également en raison du contact entre les langues. D'un côté, leur catégorie n'est pas clairement définie et ils sont moins dépourvus de sens qu'on ne le croit, d'un autre côté, par le biais des noms propres et spécialement des toponymes, un certain nombre des morphèmes de langue étrangère entrent comme emprunts dans une langue donnée. Deux constatations sont à la base de la réflexion présentée ici : (1) Les dictionnaires, qu'ils soient mono- ou plurilingues, contiennent rarement de noms propres, considérés comme des objets d'encyclopédie. Pourtant, les encyclopédies ne fournissent pas d'une façon cohérente les informations grammaticales nécessaires à l'utilisateur, ce qui est encore plus frappant dans les dictionnaires bi- ou plurilingues. (2) Les noms propres ont une fréquence textuelle non négligeable dans des documents très divers, et leur fréquence lexicale relative est encore plus importante.
- 2 Si, parmi les noms propres, nous avons choisi d'étudier les toponymes, c'est qu'ils semblent les plus proches du vocabulaire général. En effet, ils résultent souvent de la combinaison d'un élément d'un nom propre (proprial) et d'un appellatif (nom commun ou adjectif) ainsi que de morphèmes grammaticaux : pour ces raisons, ils ont beaucoup d'affinité avec le vocabulaire général. Cette place hybride pose d'autant plus de questions pour les dictionnaires bilingues. Nous proposons donc ici une approche comparative basée sur une analyse fonctionnelle de ces éléments constitutifs appliquée à notre base de données multilingue de toponymes (DiNoPro)

- 3 Le Dictionnaire des noms propres (DiNoPro) est un projet au sein de l'équipe LIDILE visant la publication d'une base de données multilingue de toponymes consultable en ligne. Cette base a été élaborée par une douzaine d'enseignants-chercheurs de l'Université Rennes 2 Haute Bretagne et inclut actuellement dix langues. Nous présentons ici les principes d'une analyse linguistique des toponymes ainsi que son application, principalement au français.

1. Objet topographique et toponymie contrastive

- 4 Le nom propre désigne des êtres vivants ou imaginaires et des objets. Il permet "de désigner directement une réalité en particulier" (Molinié, 1998 :268). Selon Grevisse, "contrairement au nom commun, il n'a pas de signification véritable, de définition, il se rattache à ce qu'il désigne par un lien qui n'est pas sémantique, mais par une convention qui lui est particulière" (Grevisse, 1975 :702). Il cite comme « véritables noms propres » d'abord les noms de lieux (villes, villages, rues, monuments, régions, pays, îles, montagnes, cours d'eau, mers, étoiles et astres) puis les personnes. D'où une motivation supplémentaire de commencer l'élaboration d'une base de données des noms propres par les toponymes.
- 5 Un toponyme est un nom propre qui désigne un lieu (du grec *topos* 'lieu' et *onoma* 'nom'). Il explicite parfois la nature de l'objet topographique (*mer Baltique*, *lac Léman*, *île de Pâques*) sous la forme d'un nom ou d'un adjectif (*mont Blanc*, *monts Métallifères*, *cap de Bonne Espérance*), ce qui permet de parler d'un statut hybride du toponyme. D'après la définition de l'Institut Géographique National, la toponymie est la science dont l'objet est « l'étude de la formation et de l'évolution des noms de lieux, ou toponymes. C'est également l'ensemble des noms de lieux d'un pays ou d'une région, d'une carte ou d'une nomenclature ».
- 6 Sans prendre en compte l'aspect historique ou étymologique des toponymes, puisque notre but est aussi de faire des descriptions structurelles et permettre des comparaisons entre les langues, dans une perspective de toponymie contrastive, nous nous inscrivons dans ce domaine de recherche. Nous ajoutons aux objectifs formulés par l'IGN la recherche de la forme particulière du toponyme appelée **exonyme**, nom donné à un objet topographique dans une langue autre que celle de la région où est situé l'objet (*Grèce* (fr), *Greece* (angl), *Griechenland* (all), *Grekland* (sué) pour *ελλάς* (grec)). L'**endonyme** est le toponyme dans la (ou les) langue(s) parlée(s) dans la zone linguistique dans laquelle se situe l'objet géographique. Chaque endonyme peut avoir un nombre inconnu d'exonymes correspondants dans d'autres langues (cf. Back, 2002, passim).

1.1. Types d'objets topographiques

- 7 Les toponymes peuvent être classés en catégories, en fonction de l'objet topographique qu'ils désignent. Le type d'objet topographique est indiqué pour chacun des termes de la base de données et permet d'interroger efficacement la base. Pour l'analyse détaillée de chaque langue cette information est essentielle.
- 8 Le type de l'objet topographique peut en effet influencer la structure linguistique du toponyme. Les toponymes désignant des villes n'ont pas le même comportement

syntactique que ceux désignant des fleuves. En français par exemple, le genre des villes est difficile à déterminer et l'usage peut varier d'un contexte à l'autre : *le Paris du XVIII^e siècle* ≠ *Paris libérée*. Dans les langues germaniques, par contre, le genre des noms de villes est stable. En allemand et en suédois, les villes sont du genre neutre, en anglais, puisqu'il n'y a pas de genre grammatical, la question ne se pose pas. Pour d'autres types d'objets topographiques, comme les fleuves par exemple, certaines langues utilisent un article défini *le Danube* (fr. masc), *die Donau* (all. fém.) dans tous les contextes (sauf énumération), contrairement au suédois où l'exonyme *Donau* (sué., genre commun) est utilisé sans article, mais d'autres noms de fleuve peuvent être suivis d'article du genre commun. Dans l'optique de notre projet d'une base de données de toponymes multilingue nous avons donc retenu 12 objets topographiques : les agglomérations, les capitales, les constructions, les continents, les cours d'eau, les étendues d'eau et côtes, les îles, les pays, les régions, les reliefs, les sites, les subdivisions administratives de premier ordre et une rubrique 'autre'.

- 9 Notre base de données comporte dix langues (allemand, anglais, arabe, breton, chinois, espagnol, français, italien, portugais, suédois). Le fait d'avoir inclus des langues à écriture non-latine dans la base ajoute une autre problématique connue dans la toponymie, celle de la translittération et de la transcription.

1.2. Relation endonyme - exonyme

- 10 L'introduction d'un toponyme étranger se fait de façons différentes, souvent tel quel, mais aussi avec des adaptations plus ou moins importantes.
- 11 **Variations phonétiques** : Un exonyme par rapport à l'endonyme peut garder exactement la même forme à l'écrit et la variation n'apparaît qu'au niveau phonétique :
Endonyme *Paris* (fr) [pa'ʁi] ≠ *Paris* (all) [pa'ʁi :s] ≠ *Paris* (ang) ['pæris]
- 12 **Variations phonétique et orthographique** : Il peut s'agir d'une adaptation à la langue au niveau de l'orthographe et de la phonétique, soit d'une langue à écriture latine à une autre :
transcription : Endonyme *Firenze* (ita) ≠ *Florence* (fr) ≠ *Florenz* (all)
Strasbourg (fr) [stras'buʁ] ≠ *Straßburg* (all) ['ʃtʁa :sbuʁk]
Ou d'une langue à écriture non-latine vers une écriture latine :
translittération : Exonyme *Moscou* (fr) ≠ *Moskau* (all) ≠ *Moskva* (sué)
- 13 **Traduction** : Il peut également s'agir d'une traduction proprement dite d'une partie de l'endonyme si celui-ci contient des composants appellatifs :
Traduction partielle :
Endonyme : *lago di Como* (it) ≠ *lac de Come* (fr) ≠ *Comer See* (all)
Ou s'agir d'une traduction complète de l'endonyme :
Traduction complète :
Endonyme *Rocky Mountains* (angl) ≠ *Montagnes Rocheuses* (fr)
Endonyme *Zweibrücken* (all) ≠ *Deux Ponts* (fr)
- 14 **Exonyme opaque** : Un exonyme peut par ailleurs prendre une forme différente de celle de l'endonyme à tous les niveaux, phonétique, orthographique, syntaxique et également au niveau sémantique, si sémantique il y a :
Endonyme *Deutschland* (all) ≠ *Allemagne* (fr) ≠ *Germany* (ang)
Endonyme *Aachen* (all) ≠ *Aix-la-Chapelle* (fr)

- 15 Dans le toponyme allemand aucun élément ne correspond à un nom commun, contrairement au toponyme (désignation) français 'chapelle'.
- 16 À l'intérieur d'une langue, aucune différence n'est faite entre endonymes et exonymes dans l'analyse et la présentation des toponymes.
- 17 La fréquence non négligeable de l'exonyme à l'intérieur d'une langue combinée avec sa forme parfois atypique a donné lieu - surtout depuis quelques années - à de multiples réflexions par rapport à sa classification linguistique. Ce statut de l'exonyme lui attribue également un rôle particulièrement important dans le contact entre les langues (et par conséquent entre les nations) et son utilisation dans les relations internationales nécessitant une normalisation proposée par les Nations Unies (Nations Unies 1977 : 253-260) et soulevant parfois des remarques critiques (Utenlandske namn i Norden 1997).

2. Analyse fonctionnelle des composants des toponymes

- 18 D'un point de vue purement morphologique, on distingue les toponymes simples (*Paris*, *Londres*, *Malte*), et les toponymes complexes (composés : *Châteaugiron*, *Albertville* ou groupes nominaux : *Mur de Bretagne*, *Le cap de Bonne Espérance*).
- 19 Pour pouvoir comparer la structure morphologique des toponymes dans les différentes langues une analyse fonctionnelle des éléments s'est avérée indispensable. Elle permet, d'une part, d'effectuer des comparaisons entre les différentes langues et, par ailleurs, d'élaborer séparément une description détaillée de la structure interne des toponymes de chacune des langues.

2.1. Le toponyme complexe

- 20 Les exonymes simples comme, en français, *Londres* (pour *London*) ou *Rome* (*Roma*) ne posent pas de problèmes de description. Des toponymes simples correspondant à un appellatif (voir ci-dessous) sont assimilés par définition à un proprial simple : *Centre* (fr), *Centro* (port) ; *Le Cap* (fr), *El Cabo* (esp). Dans la morphologie interne de ces deux langues, ces deux derniers toponymes sont constitués en effet d'un segment identique à un article défini antéposé (*Le*, *El*) et d'un nom commun.
- 21 Les toponymes complexes, par contre, demandent une approche systématique. Ainsi les exonymes *Kapstadt* (all), *Cape Town* (ang), *Ker ar C'hab* (bre) et *Kapstaden* (sué) désignant un même objet toponymique peuvent être analysés de la façon suivante.
 - Sont des toponymes composés : *Kapstadt* (all), *Cape Town* (ang) et *Kapstaden* (sué) - malgré la différence entre une composition soudée et une juxtaposition - consistant en un appellatif spécifique *Kap* ou *Cape* et un appellatif *stadt*, *town* et *stad*, respectivement. *Kapstaden* (sué) se comporte comme les autres langues germaniques au niveau de la composition, et comprend en plus l'article défini sous forme suffixale, conformément à la structure interne du suédois.

- Est un groupe nominal l'exemple du breton Ker ar C'hab (bre) qui est composé d'un appellatif ker, d'un morphème grammatical ar et de l'appellatif spécifique C'hab.
- Comme point de départ pour l'analyse nous nous sommes basés sur une étude menée par des cartographes autrichiens et publiée à Vienne (Stani-Fertl, R., Exonyme und Kartographie, 2001.).

2.2. La dichotomie terminologique

- 22 Nous avons adopté une description fonctionnelle des toponymes complexes à l'aide d'une double dichotomie :
- 23 D'une part, nous faisons la différence entre des éléments **appellatifs** et **propriaux** ; ainsi, *lac Victoria* ou *détroit de Behring* sont composés d'un élément appellatif (*lac* et *détroit*) et d'un élément proprial (*Victoria* et *Behring*). Les prépositions, marques casuelles ou autres des toponymes, différentes d'une langue à l'autre, ne sont pas prises en compte ici, mais jouent un rôle important dans la description interne des toponymes d'une langue donnée. Par extension, appellatif désigne aussi bien un nom commun qu'un adjectif (cf. Back 2002 : 10). Ainsi, dans le cadre d'une approche contrastive, les constituants appartenant à différentes classes de mots d'une langue à l'autre - dans la partie qui n'est pas proprial - peuvent facilement être comparés : *Alpes maritimes* (fr) (appellatif = adjectif) ≠ *Seealpen* (all) (appellatif = nom commun).
- 24 D'autre part, il faut distinguer entre éléments **génériques** et **spécifiques**. Ainsi, dans *le canal de Suez*, *canal* est un élément générique, qui renvoie à la catégorie d'objet topographique à laquelle le lieu qu'il désigne appartient. *Suez*, l'élément proprial, est spécifique.
- 25 Cependant, un nom propre (ou son adjectif dérivé) peut fonctionner comme générique, constituer 'l'objet de base' que l'on spécifie à l'aide d'un autre terme, dit spécifique : *la Guyane française* (fr), où *Guyane* est générique et *française* spécifique. De quelle *Guyane* (générique) s'agit-il ? De *la (Guyane) française* (spécifique).
- élément **proprial** : élément d'un toponyme se basant sur un nom propre déjà existant, par exemple : *lac Victoria* (fr), *lago di Como* (it), *Scottish Highlands* (ang).
 - élément **appellatif** : élément d'un toponyme composé qui, lui-même, n'est pas un nom propre ou une dérivation d'un nom propre : *lac/lago* dans l'exemple le *lac Victoria*, *lago di Como*.
 - élément **spécifique** : élément d'un toponyme composé qui s'ajoute à l'élément générique (ou type d'objets) pour restreindre sa portée à l'objet en question, par exemple *rouge* dans *la mer Rouge* (fr), *française* dans *la Guyane française* (fr), *Báltico* dans *Mare Báltico* (port), *hei* dans *hei/hai* (chinois) 'mer Noire' ou *Nord* dans *mer du Nord* (fr), *Tartaruga* dans *Ilha da Tartaruga* (port) ('île aux tortues').
 - élément **générique** : élément d'un toponyme composé qui indique de quel type d'objet toponymique l'objet en question fait partie, par exemple : *lac* dans le *lac Victoria* (fr) et *mer* dans *la mer Rouge* (fr), où *mer* désigne bien une mer. Par contre, *la Mer morte* (fr) ne correspond pas à la même réalité géographique, car il s'agit d'un lac. Ici, le mot *mer* n'est pas un générique dans le sens restreint du terme.
- 26 C'est justement ce type de constatation qui nous amène à faire également quelques distinctions supplémentaires pour permettre, au besoin, de peaufiner l'analyse.
- élément générique **"faux"** : élément d'un toponyme composé qui indique une 'fausse' catégorie (dans le même champ de catégories), comme *mer* pour *lac* (dans *la Mer morte* ci-

dessus et la *Mer Aral*), et qui constitue un phénomène marginal, mais existant dans d'autres langues. Ex : *al-bahr al-Abyad* (arabe) 'la mer blanche' désigne le *Nil blanc* (cf. German Glossary of Toponymic Terminology 2002 : 'falsches generisches Element').

- élément générique "**transféré**" : élément d'un toponyme composé qui indique un type d'objet topographique transféré à un autre : l'agglomération *Pont Réan* (fr) porte dans son nom le mot *pont* désignant un autre type d'objet topographique que ville ou village, cas courant également dans d'autres langues (*Cambridge* (angl), *Nybro* (sué), *Saarbrücken* (all)).
 - élément générique "**différent**" : élément d'un toponyme composé qui est un élément (éventuellement récurrent dans la langue), non-toponymique, parfois figuré, *Corps nuds* (fr), *Trollhättan* (sué) qui signifie littéralement 'le bonnet du troll'.
- 27 Chaque élément d'un toponyme est analysé en fonction de ces deux dichotomies. Les combinaisons suivantes sont possibles : Ps pour proprial spécifique (*lac Victoria*), Ag pour appellatif générique (*mer Rouge*), As pour appellatif spécifique (*mer Rouge*), et Pg pour proprial générique (*Guyane française*).
- 28 Ces catégories permettent l'analyse en synchronie de la structure des toponymes complexes d'une langue. Cette approche conceptuelle évite d'avoir recours aux concepts propres à l'analyse diachronique ou étymologique des toponymes, et présente l'avantage d'être applicable aussi bien à l'analyse des endonymes que des exonymes d'une même langue.
- 29 Elle se distingue également d'une autre approche synchronique développée dans le projet Prolex (Maurel, Piton, Eggert 2000:623-641). Ce projet de traitement automatique des toponymes et de gentils s'inscrit dans le cadre plus large de la linguistique computationnelle et vise globalement l'identification formelle des noms propres en vue de la traduction (Grass 2003).

3. Outils pour une description des toponymes en français

- 30 Seule l'analyse fonctionnelle des composants permet de comparer les toponymes des différentes langues représentées dans notre base de données. Elle permet aussi de fournir une description de la structure fonctionnelle à des toponymes de chaque langue, y compris les règles d'orthographe qui varient d'une langue à l'autre. Grâce à ces descriptions, l'utilisateur dispose d'une présentation globale concernant l'usage du terme en contexte et ne doit plus se contenter des informations sommaires fournies avec chaque entrée.

3.1. Structure fonctionnelle des toponymes en français

- 31 Nous présentons ici la version expérimentale de la description de la structure fonctionnelle du français.

3.1.1. Toponymes simples

- 32 La structure fonctionnelle la plus simple et fréquente d'un toponyme est un élément proprial unique (*Paris, le Rhin, la France, Le Caire, Les Arcs*). Ces exemples sont marqués Ps. La mention spécifique 's' pour *Paris* etc. est superflue puisque un proprial isolé est

forcément spécifique, mais cette précision l'oppose clairement à des cas de *proprial* générique comme *Guyane française* (Pg + Ps).

- 33 *Le Caire* et *Le Mans* sont des *propriaux*, dont *Le* se comporte en fait comme un article défini puisqu'on ne dit pas : **je reviens de Le Caire*, mais *du Caire*. Mais le test : *Dans le Le Caire du Moyen Age* montre que *Le* est un article lexicalisé et de ce fait une partie intégrante du *proprial*.

3.1.2. Toponymes complexes

- 34 Parmi les toponymes complexes on distingue les toponymes composés (soudés ou juxtaposés) et les toponymes constitués d'un groupe nominal. Les toponymes à plus d'un élément ont majoritairement, dans la langue romane qu'est le français, l'élément spécifique en deuxième position

- Le toponyme composé et soudé n'est pas rare en français (*Montparnasse* (Ag + Ps)) mais ne semble plus productif dans la constitution actuelle des exonymes. En effet, le plus souvent un toponyme composé soudé étranger correspond à un exonyme français juxtaposé ou complexe : *Deux ponts* (fr) ≠ *Zweibrücken* (all) ; *barrage d'Assouan* (fr) ≠ *Assuandammen* (sué).
- Le toponyme composé et juxtaposé est une structure fréquente qui se présente le plus souvent comme une combinaison analysée Ag+Ps à savoir un appellatif générique suivi d'un élément *proprial* (*mer Baltique*) ou bien un appellatif générique suivi d'un appellatif spécifique Ag+As (*mer Rouge*).
- Le toponyme complexe avec une marque grammaticale ' % ' (en français la préposition) est la combinaison d'un appellatif et d'un *proprial*, marquée Ag %Ps (*lac de Constance*, *col de Brenner*, *cordillère des Andes*). Parfois, l'appellatif va être combiné avec un autre appellatif qui 'spécifie' le toponyme : Ag %As (*îles du Vent*). Le cas du toponyme coordonné constitue une sous-catégorie du toponyme complexe, car aucun élément ne spécifie l'autre (*Ille-et-Vilaine* (fr), *Saint-Pierre-et-Miquelon* (fr)) et doit être analysé comme un seul *proprial* (Ps).

- 35 D'après cette analyse une répartition parallèle de ces combinaisons peut être établie : le groupe de structures fonctionnelles comprenant Ag comme premier élément (Ag+Ps, Ag %Ps, Ag+As, Ag %As) trouve son parallèle dans les structures avec un Pg initial : Pg+Ps (*Guyane française*), Pg %Ps (*Nogent-sur-Marne*), Pg+As (*Alpes maritimes*), Pg %As (*Rocheport-sur-la-côte*).

- 36 Plus marginales sont les combinaisons avec un appellatif spécifique (As) en première position, comme As+Pg (*Basse Silésie*) ou As+Ag (*Basse-Terre*). Cette structure correspond, entre autres, au groupe nominal avec un adjectif antéposé. La combinaison As %Pg (*Le Haut du Faing*) est rare et la combinaison As %Ag paraît impossible à cause de l'incompatibilité avec la structure interne du groupe nominal en français.

Ps Londres

Ag+Ps mer Baltique Pg+Ps Guyane française As+Pg Basse Silésie

Ag %Ps île de Pâques Pg %Ps Nogent-sur-Marne As %Pg Le Haut du Faing

Ag+As mer Morte Pg+As Alpes maritimes As+Ag Basse-Terre

Ag %As îles du Vent Pg %As Rocheport-sur-la-côte As %Ag Ø

- 37 Il est par ailleurs possible de trouver des variantes des combinaisons d'analyses fonctionnelles citées selon le principe de récursivité, à savoir qu'un élément dans le toponyme peut être représenté par une structure subordonnée : *Grand Lac des Esclaves* peut être analysé comme As+(Ag %As) ou sans l'analyse au niveau inférieur : As+Pg. Une autre possibilité serait même (As+Ag) %As. C'est le contexte toponymique qui fera la

différence. Est-ce l'appellatif *Grand* qui spécifie *Lac des Esclaves*, As+(Ag %As) comme dans le premier exemple ou *Esclaves* spécifie-t-il *Grand lac*, (As+Ag) %As ?

3.2 .Marquage morphologique des toponymes

- 38 Après avoir proposé une analyse de la structure fonctionnelle des toponymes en français, le problème du marquage quant au genre et au nombre des toponymes doit être résolu.

3.2.1. Genre

- 39 “Tout nom est pourvu d’un **genre** inhérent, **masculin** ou **féminin** (...), une caractéristique qui lui reste attachée même hors emploi (stockée dans notre mémoire lexicale et enregistrée comme telle dans les dictionnaires), mais qu’il transmet, à l’intérieur du GN, au déterminant et à l’adjectif épithète (...)” (RPR 2001 :150).
- 40 Le toponyme est un groupe nominal minimal ou étendu. Chaque terme de la base de données a un genre. Il est explicite pour les toponymes construits à partir d’un appellatif (*mer de Barents* (f), *désert de Libye* (m), *chutes du Niagara* (fpl), et également *Les Arcs* (mpl)), et aussi souvent bien établi dans l’usage des cas comme *Loire* (f), *Rhin*, (m). Mais le genre peut varier ou être perçu comme incertain dans d’autres cas, sans que cela veuille dire que ces toponymes n’auraient pas de genre. Pour les toponymes simples, l’usage est variable et est parfois sensible au contexte (Földes 1999 :303-315).
- 41 Le genre d’un toponyme donné est probablement l’aboutissement d’un long usage et donc arbitraire (le *Rhin* mais la *Loire*). Dans d’autres cas, il est parfois possible de dresser la liste des principes qui déterminent le genre (en langue française, toutes les provinces de la Chine ont le genre masculin ; c’est bien entendu une convention purement française, sans influence de la morphologie d’une autre langue étant donné que le chinois ne connaît pas de genre).

3.2.2. Nombre

- 42 Le « **nombre**, **singulier** ou **pluriel** (...), n’est attribué au nom que dans l’instance du discours. Il résulte d’un choix du locuteur guidé par les besoins d’ajustement référentiel (unicité ou pluralité du référent désigné par le GN). Aussi porte-t-il sémantiquement sur l’ensemble du GN (...) » (RPR 2001 : 151). Cette règle est sans objet concernant le nom propre puisqu’il est par nature mono-référentiel. En effet, les noms propres, même quand ils contiennent un élément appellatif, sont des termes fixés à un moment donné et dont la morphologie interne n’est pas vivante. *Le col de Brenner*, *la plaine du Pô*, *les chutes du Niagara* et *Les Arcs* ne changent pas de nombre “par les besoins d’ajustement référentiel”. En revanche, au niveau syntaxique, le mécanisme d’accord dans le groupe nominal est respecté et les adjectifs et les pronoms s’accordent au toponyme ainsi que le genre et le nombre des formes verbales conjuguées. Ainsi, le toponyme *Pays Bas* est au niveau grammatical un pluriel, mais réfère à un seul objet dans le monde.

3.3. Contexte syntaxique

- 43 Si les toponymes simples sont toujours marqués par la majuscule, l’usage de majuscule initiale pour les toponymes avec un élément appellatif diffère selon les contextes

d'emploi. En texte courant, l'appellatif initial d'un toponyme s'écrit avec une minuscule et l'usage de l'article défini correspond à celui d'un nom commun dans un contexte semblable. Par contre, isolé sur une carte ou dans une énumération, le toponyme figure sans article et le statut de toponyme est marqué par une majuscule sur l'appellatif initial également.

3.4. Entrée lexicale du toponyme

- 44 La structure syntaxique des différentes catégories de toponymes dans l'usage en texte courant ne fait pas l'objet de l'entrée lexicale.
- 45 Le traitement lexicographique du toponyme dans la base de données sera donc, d'une part, la présentation du terme toponymique avec le marquage de la ou les majuscule(s) (ainsi que d'éventuels traits d'union) qui constitue la marque typographique du nom propre et, d'autre part, une indication grammaticale du genre et du nombre. Figurera également une autre dénomination éventuelle.
- 46 L'article défini ne va pas être marqué dans l'entrée lexicale. Le segment *Le, La, Les* qui fait partie intégrante du toponyme *Le Mans, Le Caire, La Rochelle* n'est pas concerné par cette règle. Dans les autres cas, le genre et le nombre sont marqués pour indiquer quel est l'accord à respecter, l'usage d'article défini, l'accord d'adjectif et de pronom.
- 47 Le toponyme peut donc figurer dans la base comme mot simple (*Paris, Le Caire*), comme juxtaposition de deux ou plusieurs éléments (*lac Léman, Alpes maritimes, Région Rhône-Alpes*), comme mot composé (*Châteaugiron*), comme mot composé avec trait d'union (*Alpes-Maritimes*) et comme syntagme complexe (*mer d'Oman, île de Pâques, Grand Lac des Esclaves*) ou comme deux éléments coordonnés (*Ille-et-Vilaine*).
- 48 Les règles d'orthographe proposées ici sont simples et reflètent les principes observés dans l'usage de la majuscule et du trait d'union. Nous ne prenons pas en compte les variations, mais résumons des usages fréquents, systématisés par domaine ou installés par habitudes (mais sans principes évidents) dans un souci de cohérence. Autrement dit, nous visons une présentation descriptive et systématique, sans ambition prescriptive.

3.4.1. Majuscule

- 49 La règle générale, la "majuscule s'emploie pour marquer un nom propre (...), par opposition à un nom commun" (RPR 2001 :74), n'est pas suffisante pour rendre compte des variations orthographiques des toponymes. Les règles d'utilisation des majuscules peuvent, pour l'essentiel, être exprimées en termes d'analyse fonctionnelle :
 - Le mot simple (analysé Ps) prend une majuscule initiale (*Paris, Londres*). Dans le toponyme simple constitué d'un segment *Le, La, Les* et d'un mot simple (l'ensemble analysé Ps), les deux mots prennent une majuscule (*Le Caire, La Rochelle*).
 - Pour les formations plus complexes, le principe de base est que chaque élément propre qu'il soit spécifique ou générique (Ps, Pg) prend la majuscule (*mer Baltique, terre Adélie (Ag+Ps), col de Brenner, trouée de Belfort, mer du Nord (Ag %Ps) ; Basse Silésie (As+Pg)*). Les mots grammaticaux (conjonctions, prépositions) ne prennent pas de majuscule.
 - L'élément As prend également la majuscule, comme dans *Basse Silésie (As+Pg)*, puisqu'il est spécifique et établit l'entité en question comme nom propre, que ce soit en combinaison

avec un proprial générique (Basse Silésie (As+Pg)) ou un appellatif générique (mer Rouge (Ag+As)). L'élément Ag prend la minuscule.

- Dans Basse Silésie (As+Pg), l'appellatif spécifique (Basse) établit le nom propre et prend avec sa position initiale la majuscule. Par contre, il y a une nette tendance à utiliser la minuscule pour des As ainsi que des Ps en forme adjectivale en position non-initiale et en combinaison avec un proprial, comme Alpes maritimes (Pg+As), Guyane française (Pg+Ps). L'usage varie pourtant et on peut lire Guyane Française et Alpes-Maritimes (voir aussi 3.4.2).
- Dans le cas de la variation entre Asie du sud-est (Pg %As) et Asie du Sud-Est (Pg %As), on peut faire l'analyse suivante : les points cardinaux peuvent être vus comme des propriaux (Asie du Sud-Est (Pg %Ps) et prennent donc la majuscule.

3.4.2 Trait d'union

- 50 La présence du trait d'union entre éléments constitutifs d'un toponyme ne modifie pas son analyse fonctionnelle, mais elle permet de distinguer entre dénomination géographique et dénomination administrative :
- 51 *Alpes maritimes* est une région géographique alors que *Alpes-Maritimes* désigne une subdivision administrative.

3.4.3. Genre et nombre

- 52 Les indications sur les catégories grammaticales sont très succinctes mais se veulent sans ambiguïté. Ainsi, msg et mpl, fsg et fpl sont indiqués quand les rédacteurs sont sûrs du genre du terme. Autrement l'indication -sg ou -pl signifie que le choix de genre varie dans l'usage, soit que le genre n'a pas pu être déterminé.

3.4.4. Article défini

- 53 Une information, distincte de celle du genre et du nombre, dont l'utilisateur a besoin, est de savoir si l'usage de l'article défini est de règle pour un toponyme donné ou non.
- 54 Ainsi, *Malte* ne prend jamais d'article, même dans un texte courant, mais *Iran* en contexte est précédé de l'article. Ex. : *Il est représentant de Malte*, mais *Il est représentant de l'Iran*.
- 55 Pour chaque entrée est indiqué, dans le champ 'Catégorie grammaticale', si l'article défini est d'usage dans les contextes où la syntaxe de la langue l'exige. Ex : *mer Baltique* (Afsg), où A veut dire présence d'article défini, pour un toponyme féminin singulier, comparé à *Londres* (Ø-sg), où Ø veut dire absence d'article défini pour un toponyme de genre indéterminé (codé par le tiret).
- 56 La description de la spécificité de la préposition 'en' qui remplace l'article qui précède les toponymes de genre féminin en français relève d'une grammaire générale de la langue française et n'est pas l'objet d'un dictionnaire (*Il est représentant de la France* mais *Il habite en France*).

3.4.5. Les informations données pour chaque entrée

- 57 En consultant la base de données l'utilisateur trouvera pour chaque entrée des informations supplémentaires sur le toponyme recherché, et présentées dans les

champs ci-dessous. Une éventuelle dénomination secondaire figure sur le même plan que l'entrée principale et fournit les mêmes types d'informations :

la langue et le terme : Français : **lac d'Averne**
catégorie grammaticale : Présence d'article **Amsg**
zone géographique : **Italie**
type d'objet topographique : **Etendue d'eau et côtes**
source : **Divers**
note :
la langue et le terme : Français **Londres**
catégorie grammaticale : Absence d'article **Ø-- sg**
zone géographique : **Royaume-Uni**
domaine : **Capitale**
source : **divers**
note :

Conclusion et perspectives

- 58 Le travail de mise en place de la version actuelle de la base de données a permis, en partant d'un objectif et d'un format relativement modestes, d'élaborer les exigences théoriques et pratiques pour bien rendre compte d'un vocabulaire précis, les toponymes, du point de vue linguistique tout en prenant en compte les différentes catégories topographiques que ce vocabulaire reflète.
- 59 Ainsi, nous pensons avoir trouvé un moyen de rendre compte des informations nécessaires à l'utilisateur sur la morphologie et l'orthographe des toponymes en français par une analyse fonctionnelle de leur structure en laissant les régularités syntaxiques de côté, celles-ci appartenant à la syntaxe générale. Dans la version actuelle de la base, les formes variantes (orthographe), voire des termes alternatifs, n'ont pas pu être intégrés d'une façon systématique et efficacement exploitable. Le même raisonnement est valable pour les informations topographiques. Une étude approfondie permettra raisonnablement d'améliorer cet aspect du projet. Nous comptons également sur les commentaires des utilisateurs.
- 60 Le développement souhaitable concerne donc aussi bien la forme que le contenu. Parmi les informations souhaitables figurent les adjectifs et les noms des gentilés (cf. Eggert 2005) correspondant aux toponymes, ainsi que la prononciation des entrées.
- 61 Quant au contenu de la base de données, nous envisageons un développement qualitatif qui intégrerait, d'une part, une analyse détaillée et, d'autre part, une information sur l'usage effectif des toponymes.
- 62 Dans le premier cas, il s'agirait de faire une analyse morphologique détaillée, une segmentation des termes qui permettrait d'identifier, dans les différents toponymes les allomorphes d'un même morphème proprial (ou appellatif). Cette analyse permettrait d'établir les régularités dans la formation des mots dans ce domaine spécifique qu'est la toponymie. Ainsi, les allomorphes *France*, *français*, et *franco-* pourraient être traités et présentés ensemble.
- 63 Dans le deuxième cas, l'établissement de liens entre les (termes des) entrées de la base de données et le vocabulaire d'un grand corpus de référence (par langue) permettrait de donner, sous forme de concordance, à l'utilisateur des contextes représentatifs de l'usage effectif des toponymes qu'il cherche. Les structures syntaxiques et la variation seraient ainsi accessibles en contexte.

BIBLIOGRAPHIE

Back O. (2002), Übersetzbare Eigennamen. Eine synchronische Untersuchung von interlingualer Allonymie und Exonymie. Edition Praesens, Vienne.

Eggert, E. (2005), La dérivation toponymes-gentils en français : mise en évidence des régularités utilisables dans le cadre d'un traitement automatique. Bisontins ou Besançonnais. A la recherche des règles pour la formation des gentils et application au traitement automatique, Narr, Tübingen.

Földes C. (1999), Ortsnamen im Spannungsfeld von Öffentlichkeitssprachen und sprachlichem Wandel. Germanistisch-linguistische Bemerkungen anhand der Kosovo-Krise. (*Muttersprache* 109, p. 303-315). Wiesbaden.

German glossary of toponymic Terminology version 2.0 (2002), Bundesamt für Kartographie und Geodäsie (BKG) Frankfurt.

Glossaire de termes pour la normalisation des noms géographiques (2002). United Nations, New York. unstats.un.org/unsd/geoinfo/N0244957.pdf

Grass T., Maurel D., Rossi N., Savary A. (2003), Identification formelle des noms propres et traduction (Journées d'études à l'Université François Rabelais, Tours).

Grevisse M. (1975), *Le Bon Usage*. Duculot, Gembloux (Belgique), 10e édition revue.

Helleland B., Nilsson L. (éds) (1997), Utenlandske namn i Norden. Rapport frå NORNAs tjuesjette symposium i Oslo 28 -30 mai 1997. Uppsala.

IGN Institut Géographique National (2004), Publications de la commission de toponymie. "Charte de toponymie". (www.ign.fr)

Löfström J. (2005), "Namn på främmande länders orter i det svenska språket". (*Svenskans beskrivning* 27), Växjö universitet, Växjö.

Maurel D., Piton O., Eggert E. (2000), Les relations entre noms propres : lieux et habitants dans le projet Prolex. *Traitement automatique des langues*, vol. 41-1, p. 623-641.

Molinié G. (1998), *Grammaire et communication*. Magnard Paris.

Nations Unies (1977), Essai de classification des exonymes, Troisième Conférence des Nations Unies sur la normalisation des noms géographiques, volume 2, 253-260.

Riegel, M., Pellat, J.Ch. et Rioul, R. (2001), *Grammaire méthodique du français*. Quadriga/PUF, Paris. (= RPR)

Schneider Alain J., Règles typographiques 1 : Les majuscules / minuscules. <http://alain.j.schneider.free.fr/majuscules-minuscules.htm>

Stani-Fertl, R. (2001) : Exonyme und Kartographie. Weltweites Register deutscher geographischer Namen, klassifiziert nach Gebräuchlichkeit, und ihrer ortsüblichen Entsprechungen. (*Wiener Schriften zur Geographie und Kartographie*, Band 14), Wien.

Vaxelaire J.L. (2001), *Pour une lexicographie du nom propre*. Université Denis Diderot (Paris VII).

RÉSUMÉS

Le projet de constitution d'une base de données multilingue de toponymes et l'objectif associé à ce projet à savoir la comparaison des exonymes à travers dix langues, a montré la nécessité d'analyser les toponymes (et avant tout les toponymes complexes) des différentes langues par une méthode globale. Inspirés d'une étude menée à Vienne (Stani-Fertl, 2001) nous avons déterminé les éléments des toponymes à l'aide d'une double dichotomie, respectivement en « proprial » versus « appellatif » et « générique » versus « spécifique ». Par cette méthode quatre types d'éléments et un grand nombre de combinaisons ont été déterminés, à l'aide desquels les toponymes peuvent être triés, classés et comparés entre eux dans une langue ou entre les différentes langues. Qui plus est, par la même méthode, une description concernant les structures des toponymes français et les mécanismes au niveau de l'orthographe, du genre et du nombre et de l'usage de l'article qu'elle engendre a pu être dégagée. Cette description, jointe aux informations fournies dans la base de données pour chaque entrée, servira aux futurs utilisateurs lors d'une consultation en ligne, comme repère pour une application correcte du toponyme français en contexte. Les descriptions des neuf autres langues, selon les mêmes principes, sont en cours et figureront à côté de celle du français dans le site Internet du dictionnaire des noms propres (DiNoPro) et devront assurer les mêmes compétences concernant les autres langues à l'utilisateur et à long terme aux utilisateurs non français.

The project to establish a multi-lingual database of toponyms and its associated purpose to compare exonyms through 10 languages has stressed the necessity to analyze the toponyms, in particular the analysis of complex toponyms of the different languages by means of a global method. We determined the elements of the toponyms by means of a double dichotomy: proprial against appellative and generic against specific. Four different types of elements and an important number of possible combinations have been determined by this method, by which the toponyms can be sorted, classified and compared in one language as well as between different languages. The same method made it possible to describe French toponyms and mechanisms with regards to spelling as well as gender, number and use of articles. This description combined with the information given for each entry in the data base will serve as a guideline for the correct application of French toponyms in their context. The descriptions of the nine other languages are being established along the same principles. They will appear next to the French descriptions on the web-site of the dictionary for proper names. (Dictionnaire des noms propres: DiNoPro). This should enable users and in the long run non-French users, to have access to the same information for the other languages.

INDEX

Mots-clés : appellatif, base de données multilingue, endonyme, exonyme, générique, proprial, spécifique, toponyme

AUTEURS

JONAS LÖFSTRÖM

Université Rennes 2 Haute Bretagne, LIDILE (EA 3874)

BETTINA SCHNABEL-LE CORRE

Université Rennes 2 Haute Bretagne, LIDILE (EA 3874)