

Un dictionnaire électronique relationnel des noms propres liés à la géographie

Denis Maurel, Claude Belleil

Abstract

When we parse a text with Dictionary or Lexicon-Grammar, we inevitably meet Unknown Words and above Proper nouns. So, we have begun to built a French Proper Nouns Dictionary. To day, we confine to French Proper Nouns linked up to Geography. We think that these nouns are more easy to define and to collect.

Furthermore, French Proper Nouns linked up to Geography have to be recognized, but also to be combined with them : for instance, a town with the name of this inhabitants, or with a department or a region... That's why our project is divided in two-parts, a morphological processing and a relational processing.

Résumé

L'analyse automatique des textes commence généralement par une première phase, morphologique et lexicale, suite à la consultation de dictionnaires électroniques. Il est nécessaire d'envisager alors un traitement particulier pour les noms propres. Parmi ceux-ci, nous sommes en train de constituer un dictionnaire électronique relationnel des noms propres liés à la géographie.

Citer ce document / Cite this document :

Maurel Denis, Belleil Claude. Un dictionnaire électronique relationnel des noms propres liés à la géographie. In: Linx, n°34-35, 1996. Lexique, syntaxe...automatique. Hommage à Jean Dubois. pp. 77-88.

doi : 10.3406/linx.1996.1417

http://www.persee.fr/doc/linx_0246-8743_1996_num_34_1_1417

Document généré le 29/09/2015

Un dictionnaire électronique relationnel des noms propres liés à la géographie

Denis MAUREL

IRIN¹, Nantes – Laboratoire d'informatique/
E3i/Université de Tours

Claude BELLEIL

IRIN

0. Introduction

0.1. Les dictionnaires électroniques

L'analyse automatique des textes commence généralement par une première phase, morphologique et lexicale, suite à la consultation de dictionnaires électroniques.

L'évolution des recherches, tant en linguistique qu'en informatique, a eu une répercussion certaine sur la constitution de tels dictionnaires. Ainsi, par exemple, le dictionnaire *DELAS* [Courtois B. 1992], qui ne comportait à l'origine que des informations grammaticales et flexionnelles, contient aujourd'hui des indications sémantiques, des liens avec le lexique-grammaire [Gross M. 1975] [Leclère C. 1990] et une structuration en couches hiérarchiques (vocabulaire courant, occasionnel ou technique) [Garrigues M. 1993].

La théorie informatique des automates à nombre fini d'états permet l'utilisation directe de ces dictionnaires pour le traitement de corpus [Silberztein M. 1993] ou la préanalyse de formes relativement stables, comme certains adverbiaux [Maurel D. 1991]. Il est aussi

¹ Institut de Recherche en Informatique de Nantes.

possible de "prévoir" l'utilisation de mots nouveaux, par exemple en créant systématiquement tous les déverbaux à partir des tables du lexique-grammaire ; bien que cette entreprise semble *a priori* conduire à une explosion de la taille du dictionnaire, il n'en est rien si l'on considère celui-ci non comme une liste de mots, mais comme un automate. Par exemple, la dérivation systématique des 384 verbes de la table 32RA double le nombre de mots du *DELAS* mais n'ajoute qu'un kilo-octet à l'automate [Clémenceau D. 1993].

Il restera cependant toujours des "mots inconnus", c'est-à-dire non répertoriés par le dictionnaire... ne serait-ce que des fautes d'orthographe, des abréviations, des sigles ou des noms propres. Mais dans quelle proportion ? Une typologie de ces "mots inconnus" a été réalisée par X. Ren et F. Perrault [1992], à partir du dépouillement de deux corpus canadiens, mais sans prendre en compte les noms propres². C'est pourquoi nous sommes en train, à Nantes, de dépouiller un corpus constitué de numéros du journal *Ouest-France*, pour repérer les différentes sortes de mots inconnus et pour donner une classification statistique des noms propres liés à la géographie.

0.2. Les noms propres

Une question surgit alors : faut-il considérer un dictionnaire électronique de noms propres comme une simple extension du dictionnaire *DELAS* ? Si la morphologie flexionnelle des noms propres est probablement identique à celle des noms communs, il n'en est sûrement pas de même, ni pour la morphologie dérivationnelle (cf. [Ganelko M. 1977]), ni pour la syntaxe (cf. [Gary-Prieur M. N. 1991]³). Il est possible aussi que les règles de phonémisations établies par E. Laporte [1988] soient à compléter par de nouvelles règles spécifiques.

Quant à la sémantique, assurément, "Les noms propres renvoient aux trois dimensions de la deixis, la personne, l'espace et le temps" [Molino J. 1982:19], ce qui explique qu'un nom propre ne se comprend que par référence à d'autres noms propres. C'est bien ainsi d'ailleurs que sont conçus les dictionnaires imprimés, où la partie "noms propres" ne comporte pas de définitions, au sens des noms communs, mais plutôt une description encyclopédique des relations que le mot étudié entretient avec d'autres noms propres.

Notre dictionnaire sera organisé en deux parties (un peu à l'image de la correspondance *DELAS-DELAF*) : d'une part le recensement exhaustif de toutes les formes morphologiques⁴ des noms propres qui y figurent (les flexions bien sûr, mais aussi les abréviations, très fréquentes) et, d'autre part, les mises en relation des différents noms

² Ils éliminent en effet systématiquement les mots commençant par une majuscule (c'est-à-dire, entre autres, les noms propres).

³ p. 16 : "Toute définition en termes de sous-catégorisation stricte est vouée à l'échec en français".

⁴ ou morphophonémiques lorsque nous aurons entrepris la phonémisation du dictionnaire.

propres entre eux (inclusions géographiques, nom de lieu-nom d'habitants, personnages ou événements célèbres évocateurs d'un lieu, etc.) (figure 1).

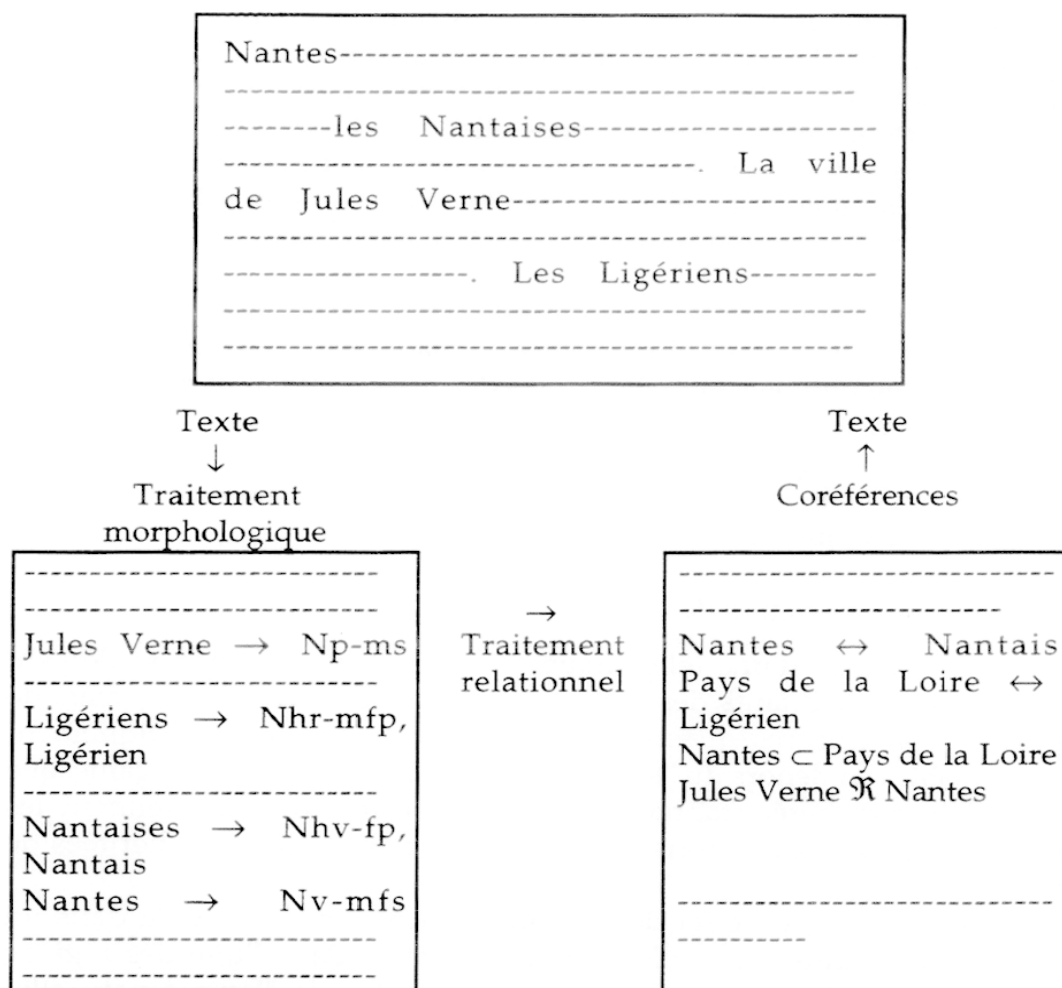


Figure 1 : Organisation bipartite du dictionnaire

Nous allons tout d'abord présenter le travail de collection des données que nous avons entrepris (§1), puis viendront les traitements morphologique (§2) et relationnel (§3) du dictionnaire. Ce dernier aspect, qui constitue peut-être le point le plus original de notre traitement lexical, ne sera qu'une introduction à des développements en cours⁵.

1. La collection des données

1.1. A partir des dictionnaires existants

Ce travail de collection des données est moins trivial qu'il n'y paraît de prime abord. Car, d'une part, celles-ci ne sont pas accessibles

⁵ Dans le cadre d'une thèse de doctorat qui sera soutenue par Claude Belleil.

directement et, d'autre part, le choix des entrées de notre dictionnaire déterminera son usage futur.

Prenons un exemple simple, celui des noms de ville (française) et des noms d'habitants correspondants. Contrairement à ce que l'on pourrait croire, il n'existe pas, à notre connaissance, de liste exhaustive de ce type. Bien sûr, des listes d'exemples sont fournies par des dictionnaires ou encyclopédies, mais elles sont toujours incomplètes et souvent imprécises sur l'aspect flexionnel : le *Quid* pose par exemple la question "Comment s'appellent les habitants de ?" et répond uniquement par des formes masculin-pluriel. Il en est de même pour les *Guides Verts Michelin*.

Notre première liste (cf. [Maurel D. et al. 1992]) était donc sujette à caution et nous avons préféré mettre en place une collection directe des données. Une première enquête auprès des préfectures nous a appris que, contrairement à nos espérances, celles-ci n'étaient pas mieux renseignées que nous sur les gentils de leur région ! Aussi, nous avons décidé une enquête nationale auprès des habitants eux-mêmes, à travers les mairies de quelque six mille communes.

1.2. Une enquête nationale

L'avantage d'une telle enquête était de connaître non seulement les noms des habitants des différentes communes concernées, mais de poser aussi d'autres questions dont les réponses sont pratiquement introuvables en dehors du "terrain", à moins de se livrer à un gigantesque travail historico-géographique : à savoir, le lien qu'entretiennent les communes avec les régions géographiques que nous avons héritées de l'histoire.

Il est facile en effet de connaître l'inclusion d'une ville dans un département, ainsi que celle d'un département dans l'une des régions administratives qui découpent notre pays. Mais la "langue naturelle" n'est pas codifiée par l'administration (heureusement d'ailleurs) et les références régionales que l'on trouve, par exemple dans la presse, sont le plus souvent des références historiques et non administratives. Ces régions sont parfois très petites, comme la *Saintonge* (partie de la région Poitou-Charentes) ou le *Pays Bigouden* (région autour de la ville de Pont-l'Abbé en Bretagne).

Et, pire encore, les régions historiques et les régions administratives portent parfois le même nom sans se référer au même lieu géographique ! Ainsi, la ville de *Vervins* est située dans la *Thiérache*, région elle-même située à la fois dans la Picardie historique et dans la Picardie administrative, alors que la ville de *Marseille-en-Beauvaisis*, qui est situé, comme son nom l'indique, dans le *Beauvaisis*, appartient à la Picardie administrative, sans pour autant appartenir à la Picardie historique.

Cette enquête a été parrainée par trois partenaires, la Direction régionale de la Poste, le Conservatoire National des Arts et Métiers de

Nantes et le journal Ouest-France. Nous avons pu, grâce à leur appui logistique, atteindre notre objectif, à savoir interroger :

- 1) les huit cent une villes de plus de dix mille habitants (suivant le recensement de l'INSEE en 1975) ;
- 2) les mille neuf cent soixante-cinq villes connues pour leur intérêt historique ou touristique (et citées de ce fait dans les *Guides Verts Michelin*) ;
- 3) les quatre mille trois cent soixante et une villes de l'Ouest de la France (d'après la couverture régionale du journal Ouest-France).

Soit un total de cinq mille neuf cent quatre-vingts communes (après élimination des doublons entre ces trois listes).

1.3. Le questionnaire

Outre le nom de la ville et de ses habitants (avec les quatre formes fléchies), le questionnaire envoyé demandait à quelle(s) région(s) historique(s) appartenait la commune (avec, aussi, le nom des habitants de cette région). Puis, d'une manière moins contrainte, nous cherchions à connaître l'ensemble des noms propres liés à la ville. Quelles remarques pouvons-nous faire au vu des premiers résultats⁶ ? quatre points ont, pour l'instant, retenu notre attention :

- * le nombre très important de petites régions historiques ou géographiques auxquelles les habitants d'une ville se sentent rattachés ;
- * l'existence fréquente⁷ de plusieurs noms d'habitants pour une même ville, parfois à l'insu des dictionnaires ; par exemple, les habitants de la ville de *Gap* sont appelés les *Gappençois* d'après le *Robert*, alors que le questionnaire nous rapporte la forme *Gappençais* ;
- * la présence aussi d'alias, parfois très locaux ; ainsi, si tout le monde identifie les *Phocéens* avec les *Marseillais*, on nous signale que les habitants de *Contes* (Alpes-Maritimes), les *Contois* ou *Countes*, sont aussi surnommés les *Lu tremp'oli*⁸ ;
- * le lien qu'entretiennent certains personnages célèbres avec certaines villes ; on parlera de la *ville de Jules Verne* pour Nantes, mais qui sait que Racine est né à La Ferté-Millon (dans l'Aisne) ? Il y a là comme une sorte de lexicalisation comparable peut-être à celle des mots composés.

⁶ À la date de rédaction de cet article, l'ensemble des réponses n'a pas encore pu être traité ; bien que le questionnaire ait été envoyé en novembre 1994, nous recevions encore des réponses quatre mois plus tard. Il ne s'agit pas, bien sûr, d'une simple saisie, chaque renseignement étant comparé à l'article correspondant du *Dictionnaire Robert* des Noms Propres et à celui des *Guides Verts Michelin*.

⁷ Dans l'état actuel du dépouillement, 15 % des communes examinées possèdent plus d'un nom d'habitants ; trois cas se présentent : des suffixations différentes, comme ici, pour la ville de *Gap* ; des dérivations à partir de formes supplétives, comme ci-dessous pour la ville de *Contes* ; et des dérivations à partir des différents composés d'un nom, comme, par exemple, *Amélien* et *Palaldéen* pour la ville d'*Amélie-les-Bains-Palalda* (Pyrénées-Orientales).

⁸ À cause de leur réputation de "trempeurs d'huile" (dans l'eau).

2. Le traitement morphologique

2.1. Les noms propres de lieu

Les noms propres de lieu n'ont évidemment pas de flexion ; cependant, il sera nécessaire de leur attribuer un genre et un nombre. Il est possible de les regrouper en deux catégories, ceux qui admettent (ou exigent) la présence d'un déterminant et les autres. Mis à part les noms de ville et de personnages célèbres ou imaginaires, presque tous les noms propres liés à la géographie se retrouvent dans ce premier groupe. Le nombre est alors le même que celui du déterminant ; de même pour le genre des noms singuliers précédés d'un déterminant non éliminé. Pour les déterminants éliminés ou pluriels, une recherche du genre est indispensable, par exemple par la recherche de modificateurs attestés, éventuellement dans un autre nom propre (comme *Les Hautes-Alpes*) ; ou dans la décomposition des noms composés (comme *Les Pays de la Loire*) (figure 2).

Masculin singulier	Féminin singulier	Masculin pluriel	Féminin pluriel
Le Lot	La Loire		
Le Jura			Les Alpes
Le Limousin	La Bretagne	Les Pays de la Loire	Les Côtes d'Armor

Figure 2 : Genre et nombre des noms propres avec déterminant

Dans le second groupe, les quelques noms de région qui y figurent correspondent à des compositions hétérogènes du point de vue du genre ou du nombre, comme *Midi-Pyrénées* ou *Provence-Alpes-Côte d'Azur*. Les noms de ville, quant à eux, admettent les deux genres, d'après Grevisse, qui note à ce sujet que *l'usage est tout à fait flottant* (§549).

Quelques noms de ville semblent précédés d'un déterminant, mais il s'agit en fait d'une forme figée qui n'influe en rien sur le genre ou le nombre. *Les Sables d'Olonnes* n'est évidemment pas un pluriel et *La Baule* ou *Le Croisic* admettent les deux genres, tout comme les autres noms de ville.

2.2. Les noms d'habitants

Les noms d'habitants appartiennent aux mêmes catégories flexionnelles que celles décrites dans le DELAF. La figure 3 présente un exemple de chacune des classes rencontrées actuellement. Les trois premières (N61, N41 et N32) représentent plus de 95 % du total.

Code du DELAF	Exemple
N61 (0, e, 0, es)	Nantais
N41 (n, nne, ns, nnes)	Parisien
N32 (0, e, s, es)	Toulousain

N31 (0, 0, s, s)	Roscovite ⁹
N39 (0, sse, s, sses)	Druides ¹⁰
N42 (er, ère, ers, ères)	Berruyer ¹¹
N43 (et, ète, ets, ètes)	Bussenet ¹²
N53 (u, lle, us, lles)	Millassou ¹³
N72 (au, lle, aux, lles)	Tourangeau
N80 (0, 0, 0, 0)	Martigaux ¹⁴

Figure 3 : Les classes flexionnelles rencontrées

2.3. Les abréviations

Certains noms composés désignant un lieu admettent des abréviations très fréquentes et d'ailleurs tout à fait officielles. La figure 4 en donne quelques exemples.

Abréviation		Exemple
St	Saint	St-Étienne
Ste	Sainte	Ste-Geneviève-des-Bois ¹⁵
Pt	Pont	Pt-l'Évêque ¹⁶
Mt	Mont	Mt-Saint-Michel
s	sous	Aulnay-s-Bois ¹⁷
/	sur	Neuilly/Seine

Figure 4 : Quelques abréviations

2.4. La représentation informatique

Pour le traitement morphosyntaxique, nous envisageons la construction d'un transducteur qui reconnaîtra l'ensemble des mots du dictionnaire (y compris les formes fléchies et les abréviations) et leur associera, d'une part, une forme lemmatisée, et, d'autre part, un genre et un nombre. Prenons l'exemple de la ville de *Saint-Cyr-l'École* (Yvelines) et de ses habitants, les *Saint-Cyriens* : en supposant que *Saint* peut être abrégé en *St* et que le *É* peut ne pas être accentué, nous obtenons douze formes pour deux lemmes (figure 5¹⁸).

⁹ habitant de Roscoff (Finistère).

¹⁰ habitant de Druy-sur-Parigny (Nièvre).

¹¹ habitant de Bourges (Cher).

¹² habitant de Bussang (Vosges).

¹³ habitant de Millas (Pyrénées-Orientales).

¹⁴ habitant de Martigues (Bouches-du-Rhône).

¹⁵ ville de l'Essonne.

¹⁶ ville du Calvados.

¹⁷ ville de la Seine-Saint-Denis.

¹⁸ Une remarque s'impose sur cette figure : il peut sembler étrange d'émettre *Saint-Cyr* dès la lecture du *S* ; cela tient au fait que les douze mots reconnus commencent tous par ces mêmes lettres. Le principe de ce type de transducteur est en effet

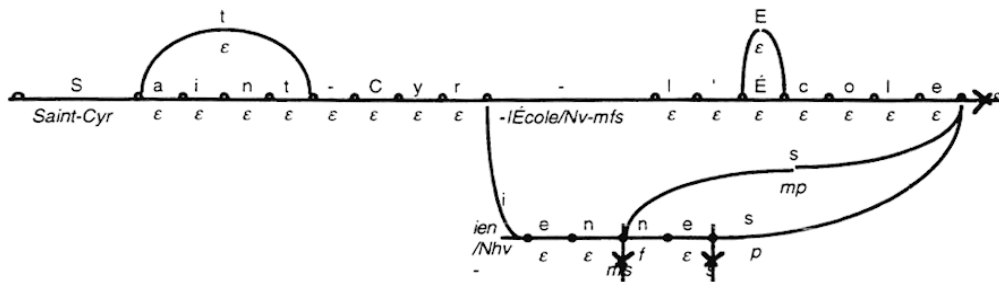


Figure 5 : Un transducteur pour reconnaître Saint-Cyr-l'École et les Saint-Cyriens

3. Le traitement relationnel

3.1. Un exemple

Comme cela a été dit en introduction, un nom propre est porteur de coréférences vis-à-vis d'autres noms propres. Ainsi, il est fréquent de voir dans un même texte (surtout dans les articles journalistiques) des noms propres différents, mais désignant un même référent sémantique, par exemple une équipe sportive. Il s'agit alors d'une métonymie où le nom de la ville remplace celui de ses habitants ; il arrive aussi assez fréquemment que soit utilisé, non pas le nom de la ville, mais celui de la région où elle est située ou celui des habitants de cette région. Prenons un exemple :

La décontraction tourangelle¹⁹ va sérieusement trancher avec l'anxiété mayennaise. Laval entend jouer les barrages pour la seconde année consécutive, un objectif qui passe par une performance au stade de la Vallée du Cher. En tout cas, Manceaux et Rouennais qui s'affrontent, Guingamp à Bourges et Cuiseaux à Ancenis espèrent que Tours fera mordre la poussière aux Mayennais.²⁰

où les *Lavallois* ne sont même pas cités, mais désignés successivement par *Laval* et par *Mayennais*²¹.

3.2. Les relations "géographiques"

Comme le montre l'exemple précédent, il sera donc nécessaire de connaître un certain nombre de liens géographique, comme "ville située

d'émettre dès que possible ce qui peut l'être, sans présumer d'ailleurs de la reconnaissance. Sur un transducteur plus grand, les émissions seront bien sûr beaucoup plus réparties sur l'ensemble des transitions.

¹⁹ Bien sûr, *tourangelle* et *mayennaise* ne sont pas des noms propres, mais les adjectifs à partir desquels les noms d'habitants sont dérivés (par conversion). Ces adjectifs seront aussi intégrés à notre dictionnaire.

²⁰ *La Nouvelle République du Centre-Ouest* du 23 mars 1992.

²¹ Laval est la préfecture de la Mayenne.

dans la région de", ou administratif, comme "préfecture de", etc. Mais peut-être aussi de savoir où coule telle ou telle rivière, où se situe telle ou telle montagne... Il y a peu de temps encore, on s'intéressait beaucoup à ce qui se passait dans la vallée de la Vologne...

Que dire aussi des relations privilégiées d'une ville avec un personnage célèbre ou un événement historique (Nantes n'est-elle pas la *ville de l'Édit* ?). Des œuvres littéraires ou des personnages imaginaires sont aussi parfois évocateurs d'un lieu... Une question majeure à laquelle il nous faudra répondre sera celle de la limitation des données ; mais comment faire notre choix ? et sur quels critères ?

Il nous faudra aussi traiter les ambiguïtés du lexique. Par exemple, *Mayenne* désigne à la fois :

- * une ville, dont les habitants sont les *Mayennais*, située dans le département de *La Mayenne*, traversée par la rivière nommée *La Mayenne* et associée à la *Guerre de Vendée* et, de ce fait, à la région historique dénommée *Vendée militaire* ;
- * un département, de la région administrative des *Pays de Loire*, dont les habitants sont les *Mayennais*, et qui a pour préfecture *Laval* ;
- * une rivière, qui est un affluent de *La Sarthe*, et qui traverse les villes de *Mayenne*, *Laval* et *Château-Gontier* ;
- * un personnage célèbre, le *Duc de Mayenne*, né à *Alençon*, mort à *Soissons* et associé à la *Sainte Ligue* et aux *Guerres de Religions*.

C'est sans doute l'utilisation de *grammaires locales*²² qui permettra de distinguer entre la *ville de Mayenne*, le *département de La Mayenne*, la *rivière de La Mayenne* et le *Duc de Mayenne*.

3.3. Les relations dérivationnelles

Peut-on déduire d'un nom de lieu le nom des habitants correspondant par une dérivation suffixale ([Dubois J. 1962 et 1969]) ? Nous avons montré qu'un nombre assez élevé d'exceptions rend plus complexe un système de construction par règles qu'une association sous la forme d'un transducteur [Maurel D. 1994], rejoignant en cela [Dressler W. U. 1985:49]. Cela tient au fait que les gentils ne sont pas toujours le résultat d'une dérivation morphologique à partir du nom de ville correspondant²³. Il arrive souvent que la formation du nom de la ville et que celle du nom des habitants de cette ville soient deux processus dérivationnels parallèles, à partir de deux formes existant dans une langue ancienne ou régionale, ou à partir d'une forme supplétive ([Adouani A. 1993:89]).

Cependant, E. Eggert [1994] propose un certain nombre de règles que nous comptons, avec son aide, formaliser dans une sorte de filtre des "mots inconnus" destiné à conjecturer des relations géographiques non

²² comme, par exemple, pour les particules préverbaux [Gross M. 1990].

²³ ou, plus exactement, comme cela a été signalé plus haut, d'une double dérivation :

$N_{ville} \rightarrow Adj_{habitants} \rightarrow N_{habitant}$ ([Corbin D. 1991]).

répertoriées dans notre dictionnaire. De plus, il remarque l'importance de la localisation géographique sur le choix de la suffixation d'un nom d'habitants et propose une carte des affixes les plus probables, carte qui serait aussi exploitable à partir de nos données sur les régions.

3.4. La représentation informatique

Nous envisageons actuellement de créer une base de données relationnelle dont l'accès serait possible à partir des différents lemmes obtenus, dès la lecture du texte lui-même, grâce à notre premier transducteur (cf. 2.4).

Cet accès se ferait par un second transducteur, un transducteur de hachage ([Revuz D. 1991:66]), qui associerait à chaque lemme un numéro unique (dans l'ordre alphabétique), ce qui permettrait la consultation d'une table contenant les adresses des données. Les ambiguïtés peuvent alors conduire à plusieurs entrées pour un même lemme. Voici, par exemple, à la figure 6, un transducteur de hachage pour la reconnaissance des cinq villes suivantes : Lille (0), Lyon (1), Nanterre (2), Nantes (3) et Paris (4). Sur ce transducteur, c'est la somme des émissions qui donne la place du mot dans la liste.

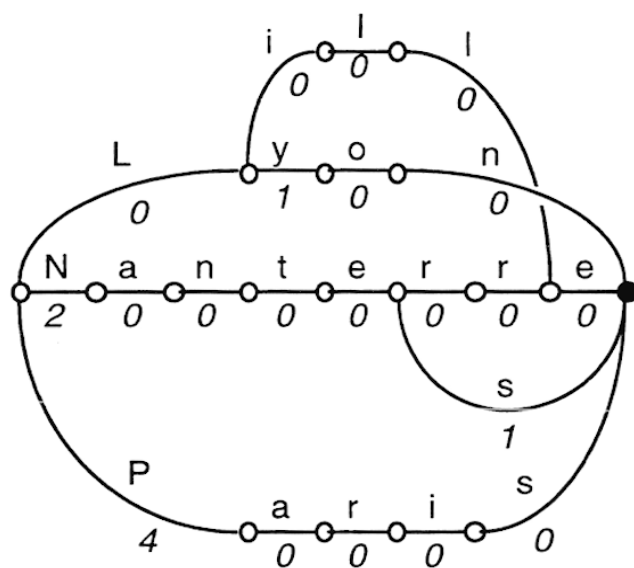


Figure 6 : Un transducteur de hachage

Denis MAUREL

16 rue du Cluzel

37100 NANTES

e-mail : maurel@univ-tours.fr

Claude BELLEIL

IRIN

2 rue de la Houssinière

44072 NANTES Cedex 03

e-mail : belleil@irin.univ-nantes.fr

Bibliographie

- Adouani A. (1993) : "Traitement dérivationnel des supplétismes lexicaux", *Cahiers de lexicologie*, vol. 63, 87-98, Paris, Didier.
- Clemenceau D. (1993) : *Structuration du lexique et reconnaissance de mots dérivés*, Paris, Thèse de doctorat (Université Paris VII).
- Corbin D. (1991) : "La formation des mots : structures et interprétations", *Lexique*, n°10, 7-30, Presses Universitaires de Lille.
- Courtois B. (1992) : *Dictionnaire électronique des mots simples du français DELAS V07-E1*, Rapport de recherche n°33 du LADL, Université Paris VII.
- Dressler W. U. (1985) : "Sur le statut de la suppléance dans la morphologie naturelle", *Langages*, n°78, 41-56.
- Dubois J. (1962) : *Étude sur la dérivation suffixale en français moderne et contemporain*, Paris, Larousse.
- Dubois J. (1969) : *Grammaire structurale du français : la phrase et les transformations*, Paris, Larousse.
- Eggert E. (1994) : *Étude dérivationnelle des dérivés de toponymes*, Mémoire de maîtrise, Université Lille III.
- Ganelko M. (1977) : "Sur la concurrence des types Pascal/Pascalien", *Revue Romane*, 123-126.
- Garrigues M. (1993) : *Méthode de paramétrage des dictionnaires et grammaires électroniques : Application à des systèmes interactifs en langue naturelle*, Thèse de doctorat en Sciences du langage, Université Paris VII.
- Gary-Prieur M. N. (1991) : "Le nom propre constitue-t-il une catégorie linguistique ?", *Langue française*, n°92, 4-25, Paris, Larousse.
- Grevisse M. (1982) : *Le français correct - Guide pratique*, Paris, Duculot.
- Gross M. (1975) : *Méthodes en syntaxe*, Hermann, Paris.
- Gross M. (1990) : "Automates pour les particules préverbaux", *Convention de Recherches ASSTRIL-UAP*, Rapport final pour l'année 1989, 1-9, Paris.
- Laporte E. (1988) : *Méthodes algorithmiques et lexicales de phonétisation de textes : applications au français*, Thèse de doctorat (Université Paris VII), Paris.
- Leclère C. (1990) : "Organisation du lexique-grammaire des verbes du français", *Langue française*, n°87, Paris, Larousse.
- Maurel D. (1991) : "Préanalyse des adverbes de date du français", *TA information*, volume 32, n°2, 5-17, Paris.
- Maurel D., Leduc B., Courtois B. (1995) : "Vers la constitution d'un dictionnaire électronique des noms propres", *Linguisticae Investigationes*, volume 19, 159-171.
- Maurel D. (1994) : "Le traitement informatique de la dérivation des noms de ville", *TA informations*, volume 35, n°2, Paris.
- Molino J. (1982) : "Le nom propre dans la langue", *Langages*, n°66, 5-20, Paris, Larousse.

- Ren X., Perrault F. (1992): *The typology of Unknown Words: An Experimental Study of Two Corpora*, Proceedings of COLING 92, Nantes.
- Revuz D. (1991): *Dictionnaires et lexiques - Méthodes et algorithmes*, Thèse de Doctorat en Informatique de l'Université Paris VII.
- Silberztein M. (1993): *Dictionnaires électroniques et analyse automatique de textes - Le système INTEX*, Paris, Masson.