

Sciences humaines combinées

ISSN : 1961-9936

: Université de Bourgogne, Université de Franche-Comté, COMUE Université Bourgogne Franche-Comté

4 | 2009

Actes du colloque interdoctoral 2009

Modèle de dictionnaire électronique de recherche par critères de termes dans la langue française

Article publié le 01 septembre 2009.

Thierry Lecolinet

DOI : 10.58335/shc.140

✉ <http://preo.ube.fr/shc/index.php?id=140>

Thierry Lecolinet, « Modèle de dictionnaire électronique de recherche par critères de termes dans la langue française », *Sciences humaines combinées* [], 4 | 2009, publié le 01 septembre 2009 et consulté le 29 janvier 2026. DOI : 10.58335/shc.140. URL : <http://preo.ube.fr/shc/index.php?id=140>

La revue *Sciences humaines combinées* autorise et encourage le dépôt de ce pdf dans des archives ouvertes.

PREO

PREO est une plateforme de diffusion voie diamant.

Modèle de dictionnaire électronique de recherche par critères de termes dans la langue française

Sciences humaines combinées

Article publié le 01 septembre 2009.

4 | 2009

Actes du colloque interdoctoral 2009

Thierry Lecolinet

DOI : 10.58335/shc.140

✉ <http://preo.ube.fr/shc/index.php?id=140>

Introduction

1. Qu'est-ce qu'une recherche sur la base de critères ?

1.1. Définition :

1.2. Que peuvent être des critères de recherche ?

1.3. Les objectifs de ce type de recherche :

2. Le corpus de travail :

2.1. La partie de la langue sélectionnée pour tester ce type de recherche :

2.2. Comment définir ce type de termes :

3. L'analyse du corpus

3.1. Forme et origine des termes :

3.2. Sens des termes :

3.3. Le cas des expressions :

4. Les résultats de l'analyse des termes de notre corpus :

5. De la théorie à l'application :

Conclusion

Introduction

¹ A l'heure actuelle, les différents modèles de dictionnaires existants qu'ils prennent la forme classique d'ouvrage ou celle de systèmes électroniques ne proposent généralement pour seule et unique option de recherches de termes, que des recherches basées sur la gra-

phie des termes qui les composent ou tout du moins une approximation de cette dernière. D'autres permettent avec des réussites qui peuvent paraître assez discutables, des recherches thématiques ou ne se concentrent que sur une terminologie spécifique de la langue, autrement dit ne traitent qu'une partie de la langue : ce sont par exemple, des dictionnaires de noms propres ou des dictionnaires qui ne vont aborder que les termes du domaine de la botanique. Parmi tous ces modèles aucun ne permet réellement de répondre à un des problèmes rencontrés actuellement dans le domaine de la linguistique : comment rechercher des termes dans une langue sans en connaître la graphie ou encore le sens ? Une option pour tenter de répondre à ce genre de questionnement est de déterminer différents moyens qui pourraient nous permettre de définir les termes d'une langue ou, du moins, d'une partie de celle-ci sans prendre en compte la manière dont ceux-ci s'écrivent et en considérant que ces termes pourraient être rechercher par une personne qui n'a même pas conscience, à la base, de leur existence ou qui, au mieux, ne dispose que d'informations très partielles à leur propos.

- 2 Ce que nous proposons au travers de cet article est une tentative de réponse à cette question à propos de la possibilité de recherche de termes dans une langue sans pour autant passer par la graphie de ces derniers.

1. Qu'est-ce qu'une recherche sur la base de critères ?

1.1. Définition :

- 3 Nous pouvons définir, dans le cas présent, le mot « critère » ainsi : tout élément, autre que la graphie, permettant de différencier un terme d'un autre dans la langue.
- 4 Une recherche sur une base de critères est, par conséquent, un type particulier de recherche qui permet de retrouver des termes dans une langue par des moyens autres que la seule référence à la manière dont ces termes vont pouvoir s'orthographier.

1.2. Que peuvent être des critères de recherche ?

- 5 A un premier niveau, certains de ces critères peuvent s'avérer être les suivants : la classe grammaticale d'un terme (nom, adjetif, adverbe, etc.), l'origine de ce terme (provient-il du latin, de l'anglais, de l'espagnol), le domaine d'usage de ce terme : est-il utilisé plus particulièrement dans un domaine spécifique de la langue (domaine linguistique, juridique, etc.), la manière dont il a été formé (troncation, mot valise, affixation), sa date d'apparition dans la langue, cette liste n'est pas exhaustive.
- 6 Toutefois, il est évident que ce premier niveau de détermination n'est pas suffisant pour une recherche en profondeur dans la langue ou même dans une partie de cette dernière. Trop de termes pouvant par exemple correspondre à la combinaison des critères suivants : « substantif venant de l'anglais, appartenant au domaine de l'informatique et ayant été recensé au XX^e siècle ». Il va donc se révéler nécessaire de prendre en considération d'autres éléments de distinction plus précis, un second niveau de critères de que nous allons définir en analysant, dans l'optique de les différencier de quelque manière que ce soit, chacun des termes qui vont entrer dans notre champ d'investigation.

1.3. Les objectifs de ce type de recherche :

- 7 L'objectif primordial de cette recherche est de définir s'il existe des éléments dans le langage qui nous permettraient de définir un terme par d'autres moyens que celui de sa simple graphie dans la perspective de trouver des palliatifs aux lacunes que ce type de recherche induit.
- 8 L'objectif secondaire, qui dépend du précédent est de concevoir différents procédés de recherche d'un terme dans une langue que ce soit de manière individuelle ou encore dans le contexte de recherches thématiques : les termes du domaine politique par exemple, voire des recherches plus techniques du point de vue linguistique : tous les

substantifs de la langue française dont l'étymon provient du grec et qui sont homographes de leur étymon.

- 9 Enfin, ce type de recherche pourrait nous permettre d'éviter des confusions entre les différents sens d'un terme, dans un contexte particulier ou une terminologie spécifique un terme n'a pas forcément le sens que nous pensons qu'il peut avoir : par exemple, dans le domaine médical ou vétérinaire, un mot tel que « dragon » va désigner une tâche dans l'œil d'un être humain ou d'un cheval et non plus un saurien légendaire ou encore un soldat de cavalerie ou d'une unité motorisée.

2. Le corpus de travail :

2.1. La partie de la langue sélectionnée pour tester ce type de recherche :

- 10 Il n'est pas question dans le cas présent de concevoir cet outil dans l'optique de traiter le langage dans son intégralité. Nous allons plus simplement envisager de traiter une partie spécifique de ce dernier, en l'occurrence les termes du français qui sont issus d'un étymon¹ mythologique. Les raisons de ce choix particulier peuvent se définir comme les suivantes : tout d'abord ce domaine nous est connu au travers de précédents travaux que nous avons réalisé. Ensuite, ce domaine est cloisonné : les termes qui le composent ont été, pour une majorité d'entre eux, bien définis. Enfin, il n'est pas producteur d'une grande quantité de néologismes, autrement dit, peu de termes de ce domaine sont producteurs de mots nouveaux à l'heure actuelle, ce qui limite les oubli potentiels, comparé à d'autres domaines actuels tels que celui de l'informatique par exemple.

2.2. Comment définir ce type de termes :

- 11 Le terme retenu peut aussi bien être un nom commun qu'un nom propre, un homographe ou un dérivé d'un terme du domaine mythologique dans le sens où il répond bien aux caractéristiques de sélection : son étymon doit être reconnu comme faisant partie d'une des

sphères de la mythologie antique (mythes européens, orientaux, asiatiques, des deux Amériques). De ce fait, il sera fait abstraction de la cryptozoologie², et des « légendes urbaines³ ». Ce même étymon doit avoir un sens spécifique dans ce domaine, il se doit ainsi de désigner : une divinité, un héros, une créature fantastique, un lieu fabuleux, etc. Il doit, de surcroît, être attesté dans la langue française : un terme tel que « drake » qui désigne bien une entité mythologique, en l'occurrence une sorte de dragon des mythes scandinaves, qui est à l'origine du terme « drakkar » mais qui n'est pas reconnu dans la langue française ne peut ainsi être pris en compte.

12 Le terme doit donc avoir des référents multiples, désigner plusieurs concepts, un premier dans le domaine mythologique et un ou différents autres dans d'autres domaines. Il doit nous être possible d'établir un lien d'analogie quel qu'il soit entre le terme mythique et celui sélectionné, ceci nous permet d'éviter de sélectionner des termes qui ne seraient, en définitive, que de simple homographes.

13 Enfin, pour ce qui est des expressions, dans la mesure où certaines d'entre elles sont éligibles dans le cadre de notre corpus, elles doivent se rapporter au domaine des mythes et des légendes en illustrant explicitement⁴ ou non ce dernier : l'expression « se retirer sous sa tente », par exemple, se rapporte bien à la mythologie sans préciser qu'elle met en scène le personnage d'Achille durant le siège de Troie.

3. L'analyse du corpus

14 Notre analyse s'est faite à différents niveaux : la forme et l'origine des termes, le sens de ces mêmes termes, une étude spécifique des expressions, avec pour objectif principal de déterminer le maximum d'informations à propos de chacun des termes qui en faisaient partie, pour pouvoir au mieux les différencier les uns des autres.

3.1. Forme et origine des termes :

15 Travailler sur la forme et l'origine des termes nous permet de définir une première série de critères d'analyse : leur nature, sont-ils des noms propres ou des noms communs ? La seconde option nous permet de mettre en avant leur classe grammaticale. Sont-ils des homographes ou des dérivés de l'étymon mythologique sur la base duquel

ils ont été créés et s'ils sont des dérivés, sur la base de quel processus de dérivation ont-ils été formés ? Toujours à propos de leur étymon, de quel type d'entité sont-ils issus (le nom d'héros, d'un dieu, d'un monstre) ? De quelle sphère mythologique proviennent-ils ?

3.2. Sens des termes :

- 16 Le sens des termes va nous permettre de mettre en lumière les éléments suivants : tout d'abord, le domaine d'usage de ceux-ci. Ce qu'ils vont désigner dans la langue (des animés, des inanimés). La manière dont ils vont le désigner par rapport au sens de leur étymon : de manière littérale ou de manière ironique⁵. Le lien qui va exister entre le sens actuel et le sens originel de l'étymon.

3.3. Le cas des expressions :

- 17 Pour les expressions, il a été nécessaire d'appliquer un traitement plus particulier d'analyse, dans le sens où nous n'avions plus ici affaire à des termes isolés mais à des parties de la langue. Il a été toutefois possible de définir quelques éléments de différenciations identiques à ceux que nous avions mis à jour pour les termes isolés, telles leur mythologie d'origine ou encore le type d'entité mythologique qu'elles mettaient en scène. Toutefois nous avons du mettre en évidence un autre type de critère pour mieux les différencier : leur motivation, l'élément particulier qui leur donne leur sens, en d'autres termes de quelle spécificité de l'entité mythologique qu'elles mettent en scène tirent-elles leur inspiration : un événement particulier ou une aventure qui est arrivé au personnage, un attribut physique ou mental de ce dernier, un comportement remarquable qui peut lui être attribué ou encore une accréditation de ce même personnage ?

4. Les résultats de l'analyse des termes de notre corpus :

- 18 Il nous a été possible de classifier quasiment l'intégralité des éléments de notre corpus suivant différents critères.
- 19 La combinatoire des critères que nous avons établis durant notre analyse ajoutée à ce que nous avons défini comme le premier niveau

de critères de classification, ou critères généraux, nous permet donc de déterminer des termes par d'autres moyens que celui de leur unique graphie et d'établir des cheminements dans le langage pour atteindre ces termes sans en connaître ni leur sens ni même, à l'origine, leur existence.

- 20 Ces critères ont été rassemblés et organisés de manière hiérarchisée en fonction de quatre principales catégories de recherche de départ : « terme », « étymon », « domaine d'usage », « lien entre le terme et son étymon ». Chacune de ces catégories se subdivise ensuite en sous-catégories de plus en plus précises et donne la possibilité de tracer des chemins déterminés dans la langue qui vont nous guider vers un ou plusieurs termes qui répondront aux choix qui auront été définis lors de la recherche parmi les sous-catégories. Cette organisation particulière se présente aussi comme un guide, si d'une certaine manière, l'agencement des catégories fait que nous pourrions nous sentir contraints dans nos choix de parcours, il implique aussi qu'aucune erreur de cheminement ne sera possible et que donc, une réponse sera toujours donnée à une requête quelle qu'elle soit. Ce choix d'ordination a été fait selon les possibilités que nous permettaient les termes retenus dans notre corpus, un ajout de termes impliquerait indubitablement des options de cheminement supplémentaires.
- 21 Cette classification établie, nous devions désormais aller plus loin et concevoir un outil qui allait nous permettre d'exploiter les résultats que nous avions mis en lumière.

5. De la théorie à l'application :

- 22 Nous avons donc pu établir de manière théorique qu'il était possible de différencier et de définir individuellement d'une façon assez précise la quasi totalité des éléments de notre corpus suivant les deux niveaux de critères que nous avons mis à jour.
- 23 Il était en effet nécessaire dans le cadre de cette recherche de, non seulement, présenter les résultats de cette analyse mais avant tout de proposer un exemple d'outil qui soit en mesure d'illustrer et d'exploiter ces mêmes résultats. Cet outil prend la forme d'un dictionnaire

électronique, support nettement plus adapté à ce type de recherche qu'un dictionnaire de facture classique de type ouvrage écrit.

- 24 Notre modèle fonctionne de deux manières différentes, tout d'abord comme les modèles classiques de dictionnaires électroniques sur la base d'une recherche se rapportant à la graphie d'un terme, en effet si notre outil se voulait innovant dans ces options de recherche de termes, il ne pouvait toutefois pas omettre les possibilités habituelles, cela aurait réduit son potentiel d'utilisation à des personnes simplement intéressées par des recherches par critères, mais aussi suivant plusieurs choix d'options de recherches thématiques réalisées sur les bases de l'analyse de notre corpus ainsi que sur les critères mis à jour par ce biais.
- 25 Toutefois cet outil reste, pour l'heure, un programme de démonstration servant à illustrer nos résultats, il n'est pas encore finalisé pour une réelle exploitation commerciale. Malgré tout, il donne un exemple substantiel de ce que pourrait être une future version terminale d'un dictionnaire de recherches par critères de termes dans une langue.

Conclusion

- 26 Nous avons montré au travers de cette recherche et de cette application qu'il est possible de définir une partie de la langue selon certains critères précis et, de la même manière, de rechercher différents termes d'une langue sans pour autant se servir de leur graphie. Si ce modèle de dictionnaire s'avère viable pour la partie de la langue que nous avions sélectionnée, il sera forcément caduc pour le même type de démarche sur un autre domaine. Il n'existe pas de modèle type pour ce genre de recherche, même si, toutefois, les bases de conception et de programmation restent identiques, les critères de recherche, hormis ceux que nous avons appelé « critères de premier niveau » devront forcément être modifiés. A chaque domaine étudié correspondra donc un modèle de dictionnaire spécifique.

- Langages, *La terminologie : nature et enjeux*, numéro 157, Paris, mars 2005, pp.25-35.
- F. De Saussure, *Cours de linguistique générale*, éditions Payot, Paris, 1985.
- L. Depecker, *L'invention de la langue. Le choix des mots nouveaux*, Armand Colin / Larousse, Paris, 2001.
- L. Depecker, *Entre signe et concept. Eléments de terminologie générale*, éditions Presse Sorbonne nouvelle, Paris, 2002.
- L. Depecker, "Contribution de la terminologie à la linguistique", *Langages, La terminologie, nature et enjeux*, numéro 157, Paris, mars 2005, pp.6-13.
- R. Eluard Robert, *La lexicologie*, P.U.F, Paris, 2000.
- F. Gaudin and L. Guespin, *Initiation à la lexicologie française*, éditions Duculot-De Boeck, Paris, 2000.
- P. Guiraud, *Structure étymologique du lexique français*, éditions Payot, Paris, 1967.
- A. Lehmann and F. Martin-Berhet, *Introduction à la lexicologie : sémantique et morphologie*, éditions Armand Colin, Paris, 2005.
- R. Martin, *Sémantique et automate*, éditions P.U.F, Paris, 2001.
- M.F. Mortureux, *La lexicologie entre langue et discours*, éditions Armand Colin, Paris, 1997.
- A. Niklas-Salmien, *La lexicologie*, éditions Armand Colin, Paris, 1997.
- J. Picoche, *Précis de lexicologie française : l'étude et l'enseignement du vocabulaire*, éditions Nathan Université, Paris, 1992.
- A. Rey, *La terminologie. Noms et notions*, P.U.F, Paris, 1992.
- S. Tillier, "Terminologie et nomenclatures scientifiques : l'exemple de la taxonomie zoologique", *Langages, La terminologie : nature et enjeux*, numéro 157, Paris, mars 2005, pp.103-116.
- I. Tomba, *La sémantique*, P.U.F, Paris, 1988.

-
- 1 Mot plus ancien à l'origine d'un mot nouveau.
- 2 Etude des créatures imaginaires telles que le Yéti ou encore le monstre du Loch Ness.
- 3 Appelées aussi « légende contemporaines », les légendes urbaines sont des récits contemporains ou adaptés à l'époque contemporaine qui mettent parfois en scène des personnages ou des créatures fantastiques.
- 4 En employant un terme mythologique parmi ceux qui composent l'expression : dormir dans les bras de Morphée.
- 5 Un terme comme « apollon » peut, entre autre, aussi bien désigner un bel homme ou, au contraire, un homme laid.

Français

L'enjeu de cet article est de présenter ce que pourrait être un dictionnaire de recherche par critères dans un corpus donné de termes de la langue française. Sur la base de ce corpus, nous allons présenter les moyens que nous allons mettre en œuvre pour classifier les différents éléments qui en font partie par d'autres moyens que celui de leur simple graphie. Nous présenterons, par la suite, un modèle d'outil électronique qui nous permettra d'exploiter nos résultats et ainsi de rechercher des termes dans notre corpus grâce aux critères que nous aurons définis.

English

The main idea of this paper is to present a model of what a dictionary of research of terms based on criteria in the French language in a selected corpus could be. On the basis of this corpus we will present the means brought into play to classify the elements of this corpus with other means than resorting to their written forms. Afterwards, we will present a model of an electronic program which allows us to research terms inside our corpus with the help of the criteria we have established.

Thierry Lecolinet

Docteur en Sciences du Language, Centre Lucien Tesnière - EA 2283 - UFC

IDREF : <https://www.idref.fr/113724918>

ISNI : <http://www.isni.org/0000000358643976>