

18 Une base de données aux services de la création terminologique

Daphné Deliyianni

RÉSUMÉ

Dans cette communication je vais présenter le travail qui a été effectué pour examiner comment et dans quelle mesure la diachronie peut servir la synchronie en fournissant de termes ou de formants pour remplir les lacunes correspondant à des besoins actuels, et ceci pour une langue technique donnée, celle du domaine de la vigne et du vin. Ayant comme point de départ l'extraordinaire continuité du grec, je vais exposer la méthodologie que j'ai utilisé pour former une base de données de termes du domaine, qui couvre une période historique de vingt siècles dans le but de repérer des termes éventuellement réutilisables en grec moderne ou même de formants entrant dans la création terminologique. Je vais aussi examiner les limites d'une telle démarche, ainsi que les résultats que le chercheur peut espérer d'obtenir selon les différentes analyses de son corpus.

Μια βάση δεδομένων στις υπηρεσίες της ορολογικής δημιουργίας

Δάφνη Δεληγιάννη

ΠΕΡΙΛΗΨΗ

Σε αυτήν την ανακοίνωση θα παρουσιαστεί η εργασία που περατώθηκε για να εξεταστεί πώς και κατά πόσον η διαχρονία μπορεί να συνδράμει τη συγχρονία δίνοντας όρους ή μέρη όρων, για να καλύψει κενά στις σημερινές ορολογικές ανάγκες, για το ειδικό λεξιλόγιο της οιολογίας. Έχοντας ως αφετηρία την αδιάσπαστη συνέχεια της ελληνικής, θα περιγραφεί η μεθοδολογία που ακολουθήθηκε για τη δημιουργία βάσης ορολογικών δεδομένων, η οποία καλύπτει μια χρονική περίοδο είκοσι αιώνων και που σκοπό έχει τον εντοπισμό εν δυνάμει όρων ή μέρους τους. Θα εξεταστούν έτσι τα όρια αυτής της μεθόδου, οι δυσκολίες της αλλά και τα αποτελέσματα που μπορεί να αναμένει ο ερευνητής ανάλογα με τον τρόπο ανάλυσης των δεδομένων του.

0 Introduction

Dans ce papier je vais présenter le travail que j'ai effectué pour la constitution d'une base de données terminologiques dans le but de repérer dans l'histoire du grec des unités terminologiques ou de formants aptes à être réintroduits en grec moderne pour remplir des lacunes dans une langue technique donnée, à savoir celle de l'œnologie. Autrement dit, j'ai cherché à savoir comment et dans quelle mesure je pouvais compter sur la possibilité qu'a

la langue à se ressourcer elle-même. L'idée principale vient du fait que les frontières entre synchronie et diachronie surtout pour une langue comme le grec qui n'a jamais cessé d'être parlée et écrite depuis des siècles, sont floues. Une approche diachronique pourrait non seulement fournir des termes tout prêts, mais aussi démontrer la manière avec laquelle la langue a évolué pour faire face aux besoins terminologiques du domaine.

Ces besoins se sont faits ressentis quelques années auparavant, au moment où, dans le cadre de mon parcours académique, j'ai essayé de mettre en grec une partie du *Lexique de la Vigne et du Vin*, édité par l'Office International de la Vigne et du Vin¹ en 1963. Il s'agit d'un dictionnaire thématique et multilingue, qui ne comprend pas le grec mais qui est, faute de dictionnaire grec, largement utilisé par les œnologues grecs. Ce groupe linguistique présente d'ailleurs un bilinguisme particulier, étant donné que la plupart d'entre eux ont effectué une partie ou même la totalité de leurs études en France. De ce fait, l'introduction de termes français dans leur langage est massive mais devient problématique du moment où la nature de leur travail fait que la communication avec d'autres groupes linguistiques, les viticulteurs, les producteurs de vin, le public, est très étendue.

1 Méthodologie du travail

C'était dans ce stade du travail que j'ai découvert une série d'articles écrits par un chimiste, Dimitris Kissopoulos, qui a voulu recueillir tous les termes utilisés par « les grecs anciens, les byzantins et les romains », dans un but, dit-il, purement terminologique. Kissopoulos donne aussi une liste exhaustive de tous les auteurs qui ont fait référence au monde du vin, que ces œuvres soient scientifiques ou littéraires. La mauvaise structure des articles et les incohérences repérées dans ces articles² m'ont conduite à reprendre tous les termes qu'il avait recueillis pour les vérifier dans leur source première. Ce sont au total deux cent quatre vingt dix-sept auteurs grecs ou hellénophones et soixante quatorze romains ou latinistes, couvrant une période de la langue grecque de vingt siècles, du VIII^e siècle av. J. C. jusqu'au XII^e siècle de notre ère. Inspirée par les possibilités de la technologie, j'ai entamé plus qu'un simple reclassement ou réorganisation des articles, la création d'une véritable base de données terminologique, en repérant les termes directement dans le texte. Si j'arrivais à vérifier l'existence de chaque terme dans ces listes, tout en les replaçant dans leur contexte et leur cotexte, je pourrais fabriquer une base terminologique proprement dite. Au niveau

¹ Aujourd'hui appelé Organisme International de la Vigne et du Vin.

² Voir Daphné Deliyianni, *La langue technique de la vigne et du vin ; une étude terminologique*, Thèse de doctorat, Université Montpellier III – Paul Valéry, février 2006, p. 123 et suiv.

des unités repérées, une base de données offre la possibilité d'une étude terminologique très complète. Étudier des unités terminologiques tant dans leur environnement textuel étendu que dans leur cotexte immédiat, permettrait par exemple de voir leur comportement lexical. Dans quelle mesure ces unités sont-elles libres, est-ce qu'elles apparaissent en collocations, et avec quelles autres unités, est-ce qu'elles forment de syntagmes terminologiques ou se présentent-elles plutôt comme des termes lexématiques, etc.

Faire un tel dépouillement terminologique serait une enquête à mener au cours de toute une vie. Si cette recherche est devenue possible à accomplir en quelques mois, c'est grâce au *Thesaurus Linguae Graecae*. Dans ce Cd-rom est collectée et mis sous forme numérique la totalité de la littérature grecque, et présente comme avantage l'assurance pour le chercheur de travailler sur des textes de très grande qualité³.

Le logiciel d'exploitation du TLG que j'ai utilisé pour faire la recherche terminologique est le MUSAIOS qui est spécialement conçu pour fonctionner en environnement Windows. Le menu principal du programme permet à l'utilisateur d'effectuer la configuration du programme et d'introduire ses préférences. Un menu assez riche en outils donne la possibilité d'effectuer des recherches ciblées, sur un ou plusieurs auteurs, ou même sur la totalité de la base. L'outil qui m'a permis de réaliser la recherche, c'est le *Multi Document Search*. En effet, il permet d'effectuer une recherche dans les œuvres des auteurs que l'on choisit. La liste d'auteurs que l'utilisateur constitue ainsi peut même être préfigurée en tant que Liste par Défaut. Ceci permet de la constituer une seule fois, au début de la recherche. Le bouton *Search* de l'outil, permet d'introduire l'unité, le terme ou la phrase que l'on soumet à la recherche. En introduisant un seul mot, l'utilisateur peut effectuer une recherche concentrée sur l'unité exacte telle qu'il la soumet, sur le préfixe ou le suffixe de l'unité, ou même, en choisissant l'option *Any*, d'avoir tous les trois types de résultats.

³ Maria Pantelia, *The Thesaurus Linguae Graecae Project: Looking Towards the 21st century*, in Ιωάννης Ν. Καζάζης, (υπο την εποπτεία του), *Η λεξικογραφία της αρχαίας, μεσαιωνικής και νέας ελληνικής γραμματικής*, Υπουργείο Εθνικής Παιδείας και Θρησκευμάτων, Κέντρο Ελληνικής Γλώσσας, Θεσσαλονίκη, 2003, p. 32.

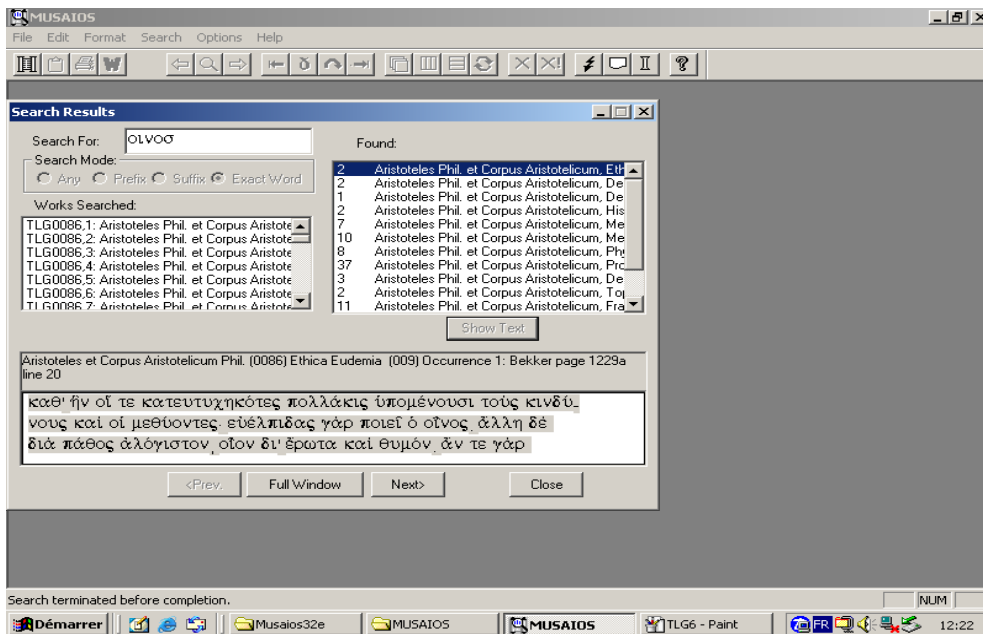


Figure 1 : La fenêtre de l'outil Multi Document Search

Dans un premier temps, j'ai pu dresser la liste des auteurs qui, selon Kissopoulos, ont utilisé des termes de la vitiviniculture. Ensuite, chaque terme recherché a été introduit dans la machine qui, six minutes plus tard, donnait toutes les occurrences trouvées sur la liste des ouvrages choisis. Une fois le nombre d'occurrences trouvé, le bouton *Results* permet de visualiser les résultats, qui sont affichés par œuvre. Il suffit alors de les choisir un par un pour les visualiser. Un bouton nommé *Next*, permet de visionner l'occurrence suivante mais ne peut pas passer automatiquement d'une œuvre à l'autre. Enfin, le bouton *Full Window* permet de visualiser le texte dans lequel a été puisée l'occurrence en entier.

La mise en place de la base s'est faite en deux temps. J'ai commencé par un premier remplissage des fiches : les premières cases que nous avons remplies sont celles qui correspondent aux résultats de la première recherche, c'est-à-dire celle de l'unité terminologique, du nombre des occurrences repérées, et du code de chaque occurrence⁴.

⁴ En effet, j'ai dû imaginer un système de codification des occurrences pour que la base soit lisible et facilement exploitable. Voir Daphné Deliyianni, *La langue technique de la vigne et du vin ; une étude terminologique*, Thèse de doctorat, Université Montpellier III – Paul Valéry, février 2006, p. 146.

J'ai ensuite vérifié les termes non repérés dans des documents secondaires, en l'occurrence des dictionnaires. Celle-ci concernait les unités ayant donné zéro résultats et donna lieu au deuxième remplissage de fiches terminologiques.

Le dépouillement direct dans un corpus, qui plus est, aussi vaste que celui que j'ai traité, ne se limite pas aux seules données terminologiques. D'autres informations peuvent être relevées pour nous aider à la meilleure exploitation de nos données terminologiques. Le champ des renseignements que nous pourrions ainsi inclure dans les fiches de la base de données est assez étendu. Dans la partie qui suit, je décris de quelle manière j'ai travaillé pour constituer la base de données et en décrire le contenu.

3 La base de données

Les fiches de la base de données terminologiques que j'ai constituée comportent chacune quatorze cases : je les présente ici dans l'ordre de leur apparition dans la base, avec leur intitulé et une petite explication de leur contenu :

1. **terme** : il s'agit du terme tel qu'il est donné par Kissopoulos. Dans le document initial, les substantifs se présentent de manière générale au nominatif singulier, les verbes à la première personne du singulier de l'indicatif présent avec les deux désinences dans le cas des verbes contractes, et les adjectifs au cas et à la personne du substantif qu'ils déterminent.

2. **catégorie grammaticale** : la terminologie étant par définition limitée aux catégories grammaticales des substantifs et syntagmes nominaux, aux adjectifs et syntagmes adjectivaux et aux verbes et aux syntagmes verbaux, j'ai utilisé les initiales *n.m*, *n.f.*, *n.n*, *adj.*, et *v.* pour chaque catégorie respectivement. Seule exception à cette règle, le cas où les substantifs se présentaient au pluriel. Dans ce cas, la lettre *p.* venait s'ajouter à côté des autres initiales.

3. **unité d'origine** : j'ai introduit dans cette case l'unité à partir de laquelle le terme a été créé avec l'ajout d'un formant préposé ou postposé. C'était par exemple le cas d'adjectifs déterminant différents types de vins fabriqués à partir de plantes autres que le raisin. L'adjectif était alors formé à partir du nom de la plante en question, plus le formant postposé – *ίτης* ; ex. φοινικίτης = φοίνιξ + ίτης.

4. **1^{er} composant** : j'y ai introduit la première des unités qui se sont juxtaposées pour former un terme composé. Nous avons par exemple rencontré en constituant notre base plusieurs breuvages dont le nom était composé ; ex. ιπέμελι = ιπέα [+ μέλι].

5. **2^e composant** : j'y ai introduit la deuxième des unités qui se sont juxtaposées pour former un terme composé. Malgré la grande capacité du grec à former des unités composées, nous n'avons pas rencontré de termes ayant été formés avec plus de deux éléments de composition.

6. **préfixe** : nous avons recueilli sous cette catégorie tous les formants préposés que nous avons repérés dans la dérivation de termes, comme par exemple le – *προ* dans *προλήνιον*.

7. **suffixe** : nous avons recueilli dans cette case tous les formants postposés entrant dans la formation de termes, mais aussi les désinences de termes composés dont l'unité d'origine n'appartenait pas au même genre.

8. **synonymes** : nous y avons recueilli tous les synonymes que nous avons repérés pour un certain nombre de termes. La question de la synonymie en langue technique et scientifique pose certainement moins de problèmes qu'en langue commune, puisque l'on accepte la présence de synonymes parfaits en terminologie. Nous n'avons traité dans cette case que des synonymes pour lesquels nous avons eu assez d'informations pour les considérer comme synonymes parfaits.

9. **variante orthographique** : la coexistence de plusieurs dialectes grecs avant l'apparition de la *koiné* nous a mis en face de l'existence d'une ou plusieurs variantes orthographiques de la même unité terminologique. Nous avons introduit ces variantes dans une case séparée, sans cependant oublier que la recherche s'effectue sur des textes extrêmement anciens pour la plupart, dont une grande partie nous est parvenue à travers des copies postérieures à l'antiquité, ce qui n'écarte pas la possibilité d'une transcription erronée.

10. **type du document** : j'ai voulu marquer dans cette case les unités terminologiques que j'ai repérées dans des documents secondaires et, en premier lieu dans le Liddell & Scott.

11. **nombre d'occurrences** : il s'agissait de marquer le nombre des occurrences repérées par le TLG dans le corpus des auteurs que j'ai soumis à l'examen.

12. **occurrence** : j'ai inscrit dans cette case, codifié, le cotexte dans lequel nous avons repéré chaque terme, dans le cas où la recherche avait été fructueuse.

13. **remarques** : j'ai librement utilisé cette case pour introduire des remarques d'ordre linguistique ou méthodologique, dans les cas où j'avais jugé que cela serait nécessaire pour la meilleure exploitation des données.

14. **bruit** : aussi performant qu'il soit, le TLG exploité par MUSAIOS reste un logiciel qui traite ses données de manière aveugle. Si l'on tient compte du fait que les textes sont introduits en tant qu'images sur lesquels nous ne pouvons pas intervenir, il est arrivé que les résultats soient altérés par ce que la machine croyait être un résultat. En effet, le dépouillement s'effectue sans référence au contenu conceptuel des textes. De ce fait, ils présentent un taux considérable de bruit, c'est-à-dire des découpages qui ne délimitent pas des termes entiers ou qui regroupent ensemble des éléments fortuits du discours sans désigner forcément une unité de sens.

4 Les possibilités de recherche dans la base

Les possibilités de recherche que présente une telle base de données sont nombreuses. Dans un premier temps, on peut dresser à nouveau la liste des auteurs qui ont utilisé des termes de la vitiviniculture. La liste qui figure au début du premier des articles de Kissopoulos semble, certes, exhaustive mais elle n'en reste pas moins vague. Les résultats de la recherche permettent de revoir cette liste sur une base plus réelle, dans la mesure où l'on s'appuie sur les occurrences trouvées. On peut ainsi la reconstituer, en y mettant les auteurs qui se sont réellement occupés du sujet.

Dans le même esprit, on peut avancer dans la démarche en établissant une liste avec les œuvres de chaque auteur dans lesquels on a trouvé des occurrences, où sera mise en filigrane l'importance qu'avait prit la vitiviniculture dans leur œuvre. On pourrait établir aussi, en effectuant une catégorisation des œuvres, une liste avec les types d'ouvrages dans lesquels la thématique du vin revient le plus souvent.

Au niveau des unités repérées, une base de données offre la possibilité d'une étude terminologique très complète. Étudier des unités terminologiques tant dans leur environnement textuel étendu que dans leur cotexte immédiat, permettrait par exemple de voir leur comportement lexical. Dans quelle mesure ces unités sont-elles libres, est-ce qu'elles apparaissent en collocations, et avec quelles autres unités, est-ce qu'elles forment

de syntagmes terminologiques ou se présentent-elles plutôt comme des termes lexématiques⁵ etc.

Les possibilités offertes par MUSAIOS, et plus particulièrement par la recherche d'unités lexicales précises, pourraient aider le chercheur à établir une table de fréquence d'apparition des termes dans la littérature, même si elle risquait d'être artificielle dans la mesure où l'on pourrait pas tirer des résultats sur leur fréquence réelle dans le système en général. L'étude de la fréquence, ainsi que leur répartition dans les sources étudiées pourrait donner des informations précieuses sur leur aptitude à faire partie de termes candidats à la réintroduction dans le langage vitivinicole d'aujourd'hui. La démarche étant diachronique, la fréquence d'apparition des unités serait alors utilisée en sens inverse d'une étude terminologique à visée synchronique. Si la fréquence est la mesure de lexicalisation de l'unité au cours d'une étude synchronique, l'idée consiste à l'utiliser pour mesurer ses possibilités d'adaptabilité dans le système actuel. En effet, on peut savoir si une unité de discours accidentelle, un hapax legomenon, est plus apte à être réutilisée. Une unité du lexique plus stable risquerait d'avoir plus de chances de survie au cours des siècles, son signifiant pourrait être encore utilisé aujourd'hui avec un autre signifié, et pourrait donc avoir un contenu sémantique encore assez stable pour que l'on puisse lui en attribuer un autre. En revanche, les unités de lexique, grâce justement à leur fréquence pourraient donner des formants ayant survécu et qui pourraient alors être utilisés dans le même but.

Le dépouillement direct dans un corpus, qui plus est, aussi vaste que celui que j'ai traité, ne se limite pas aux seules données terminologiques. D'autres informations peuvent être relevées pour aider à la meilleure exploitation des données terminologiques. Le champ des renseignements que l'on pourrait ainsi inclure dans les fiches de la base de données est assez étendu. L'analyse terminologique des unités peut alors être morphologique, sémantique, textuelle, de fréquence lexicale etc., selon une procédure répondant à chaque fois à une problématique différente.

5 Conclusions

Dépouiller un corpus qui comporte des centaines de textes, couvrant une période de vingt siècles de littérature grecque en donnant en plus la garantie au chercheur de travailler sur

⁵ Nous avons rencontré dans nos lectures linguistiques beaucoup de termes concurrents pour désigner « les syntagmes lexicaux pourvus d'une définition spécialisée ». R. Kocourec affirme en avoir repéré vingt-sept expressions différentes. Nous utilisons ici la terminologie de R. Kocourec. Voir *La langue française de la technique et de la science : vers une linguistique de la langue savante*, Brandstetter Verlag, Wiesbaden, p. 135-137.

des textes de bonne qualité, serait une tâche qui demanderait des années de travail. Le travail avec le TLG a offert un gain de temps considérable, tout comme l'assurance d'une qualité extraordinaire dans ses conditions de travail. Pour ce qui est de mon cas, le fait de pouvoir disposer d'un tel outil m'a permis d'essayer de mettre en place une procédure terminologique fondée sur une problématique de travail qui, sinon, serait prohibitive parce que jugée trop prenante ou peu productive.

Je considère le travail effectué jusque là constitue la première étape de la procédure que j'ai imaginé et qu'elle ouvre de très riches perspectives. Le résultat est une base de données contenant 2520 termes de la vitiviniculture puisés dans la culture œnologique de la Grèce telle qu'elle se reflète dans son fonds lexical, qui continue d'avancer et qui admet de multiples analyses permettant de connaître ses limites mais surtout son potentiel. Le projet final est la création d'une base de données décrivant le domaine actuel de manière plus exhaustive, comprenant les étapes suivantes du grec, au fur et à mesure que celles-ci seront comprises dans le TLG. Ainsi, les différents problèmes vont être mieux cernés, mieux décrits, mais le plus important, la mise en parallèle des termes des différents états de la langue va être mieux décrite et le travail purement terminologique, dans son aspect onomasiologique, facilité.

Bibliographie

- [1] AUGER Pierre – ROUSSEAU Louis-Jean, *Méthodologie de la recherche terminologique*, Office de la langue française, Service des travaux terminologiques, L'Éditeur officiel du Québec, 1978, 80 p.
- [2] BÉJOINT Henri - THOIRON Phillipe (sous la dir. de), *Les dictionnaires bilingues*, Champs Linguistiques, Upef – Uref - Duculot, Louvain-la-Neuve, 1996, 256 p.
- [3] CABRÉ Maria-Teresa, « Terminologie et dictionnaires », in revue *Meta*, vol. XXXIX, numéro 4, 1994.
- [4] CABRÉ Maria-Teresa, *La terminologie, théorie, méthode et applications*, Ottawa, Les Presses de l'Université d'Ottawa - A. Colin, 1998, 322 p.
- [5] CANDEL Danielle, « Néologie et Terminologie : activités et réflexions », in *Terminologies Nouvelles*, no 20, Rint, Bruxelles, 1999, p.44-53.
- [6] COUTIER Martine, « Tropes et termes : le vocabulaire de la dégustation du vin », in revue *Meta*, vol. XXXIX, numéro 4, 1994.
- [7] DELIYIANNI Daphné, *La langue technique de la vigne et du vin ; une étude terminologique*, Thèse de doctorat, Université Montpellier III – Paul Valéry, février 2006, 295 p.
- [8] DEPECKER Loïc, *L'invention de la langue ; le choix des mots nouveaux*, Armand Colin – Larousse, 2001, 719 p.

- [9] GAMBIER Yves, « Travail et vocabulaire spécialisés : prolégomènes à une socioterminologie », in revue *Meta*, vol. XXXVI, numéro 1, 1991.
- [10] GAUDIN François, *Socioterminologie ; une approche sociolinguistique de la terminologie*, Champs Linguistiques, Duculot, Bruxelles, 2003.
- [11] GUILBERT Louis, « La néologie scientifique et technique », in revue *La banque des mots*, numéro 1, Presses Universitaires de France, 1971, p. 45-54.
- [12] KOCOUREC Rotislav, *La langue française de la technique et de la science : vers une linguistique de la langue savante*, Brandstetter Verlag, Wiesbaden, 1991²

Daphné Deliyianni

Docteur en Etudes Néo-helléniques

19 B, rue Jacques Brives

34090 Montpellier

France

Courriel : daphne.deliyianni@gmail.com