

Analyse lexicale du récit fantastique
Le système UNITEX

Séminaire d'Analyse du texte fantastique:
Littérature et informatique

Kalliopi Ploumistaki
Université Aristote de Thessalonique

Sommaire

1. Nouvelles technologies et texte. Informatique et littérature
2. Présentation du système INTEX

1. Nouvelles technologies et texte. Informatique et littérature

L'usage des nouvelles technologies s'avère important et, finalement, obligatoire dans plusieurs domaines de la vie quotidienne. Elles se présentent comme un outil et un support dans toute manifestation humaine. Il est vrai que le discours oral et écrit ne pourrait pas rester intact des logiciels et des programmes informatiques. Une des questions qui pourrait surgir est la qualité et le profondeur des recherches. D'un côté, l'informatique devient le support du chercheur pour que ce dernier puisse analyser et étudier les structures du texte. De l'autre, l'informatique – les services de l'Internet inclus – constitue le moyen de l'écrivain (comme source d'inspiration ou défi) pour écrire, s'expérimenter, créer de nouveautés (littérature combinatoire, littérature numérique). Après les années '60, un grand intérêt est marqué à propos du traitement automatique des textes techniques au début, et littéraires récemment. Tout discours dispose de ses particularités ; ainsi, le discours politique ou de caractère technique s'adapte plus facilement aux exigences des systèmes informatiques, puisque les résultats de ces traitements textuels fournissent un produit plus cohérent que celui d'un texte littéraire.

Le traitement automatique des textes littéraires se déroule à trois niveaux : morphologique, syntaxique et sémantique. L'un dépend de l'autre, car la langue s'étend simultanément à ses trois niveaux. L'état de la recherche et l'élaboration de programmes et de logiciels semble embryonnaire. Les conclusions des recherches sur le discours littéraire n'ont pas apporté l'ouverture de nouveaux horizons à part les objectifs de la science de statistique. Un logiciel peut donner des informations sur les types grammaticaux (articles, adjectifs, pronoms, verbes avec leurs temps et leurs modes) et les types lexicaux sous la forme des pourcentages et des chiffres qui demanderaient beaucoup de temps si le chercheur ne l'utilisait pas. Autrement dit, les nombres, les chiffres ou les tableaux récapitulatifs touchaient peu la spécificité du texte littéraire à n'importe quelle langue, et, par conséquent, sa difficulté : le sens, la signification. Un des problèmes d'application est que l'analyse automatique demeure superficielle parce que selon les logiciels créés, le champ sémantique est rarement abordé puisqu'il emporte la majorité d'accrocs : les ambiguïtés notionnelles qui peuvent conduire à des absurdités ou à des notions équivoques et incompréhensibles.

2. Présentation du système INTEX

Le système INTEX vise à analyser tout texte des langues naturelles d'une manière précise. Max Silberztein est son créateur¹. A présent, ce système est appliqué à plusieurs langues (anglais, français, italien, espagnol). Le but de ce système est de décrire les langues, et d'analyser les textes volumineux à temps réel. Mais, les logiciels présentent une analyse moins profonde laissant en marge le champ sémantique. Le traitement des textes est soutenu par l'élaboration de dictionnaires et de grammaires électroniques.

Inspirée des recherches et du système INTEX de M. Silberztein, Tita Kiriakopoulou, professeur de l'Université Aristote, et avec les collaborateurs du Laboratoire de l'Informatique de l'Université ont appliqué l'UNITEX, leur propre système, aux textes littéraires grecs. Ainsi, ils ont créé :

- un dictionnaire de mots simples (126 000) et les formes déclinables respectives
- un dictionnaire de mots composés (28 000) et les formes déclinables respectives.
- un ensemble de programmes d'Informatique.

De plus, ils ont adapté pour les particularités de la langue grecque les codes qui marquent :

- la forme déclinable
- le lemme
- la catégorie grammaticale
- la détermination du nom (le cas, le genre, le nombre) et du verbe (l'inclinaison, le temps, la personne, le nombre).

Voilà quelques codes à titre indicatif :

| | |
|--------------------|--------|
| /Adjectif | (:A) |
| /Adverbe | (:ADV) |
| /Nom | (:N) |
| /Forme canonique | (:W) |
| /Participe présent | (:G) |

¹ Max Silberztein, *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*, Masson, Paris, 1993.

| | |
|------------------------|---------------------------------|
| /Indicatif présent | (:P1s,:P2s,:P3s,:P1p,:P2p,:P3p) |
| / imparfait | (:I1s,:I2s,:I3s,:I1p,:I2p,:I3p) |
| / aoriste | (:J1s,:J2s,:J3s,:J1p,:J2p,:J3p) |
| / futur instantané | (:F1s,:F2s,:F3s,:F1p,:F2p,:F3p) |
| / futur continu | (:D1s,:D2s,:D3s,:D1p,:D2p,:D3p) |
| /Subjonctif instantané | (:S1s,:S2s,:S3s,:S1p,:S2p,:S3p) |
| /Subjonctif continu | (:T1s,:T2s,:T3s,:T1p,:T2p,:T3p) |
| /Impératif continu | (:Y2s,:Y2p) |
| /Impératif instantané | (:Z2s,:Z2p) |
| /Conditionnel présent | (:C1s,:C2s,:C3s,:C1p,:C2p,:C3p) |

Et aussi les codes qui qualifient le genre : M (Masculin), F (Féminin) et N (Neutre) ; le nombre : S (singulier) et P (Pluriel) ; et le cas N (nominatif), G (Génitif), A (Accusatif) et V (Vocatif). Ainsi, le mot *mort* en grec a la codification suivante dans le dictionnaire électronique du Laboratoire:

| | |
|---------------------------------|----------------|
| θάνατος, N:Nms:Ams:Vms | (= la mort) |
| θανάτου, θάνατος.N:Gms | (= de la mort) |
| θάνατοι, θάνατος. N:Nmp:Amp:Vmp | (= les morts) |
| θανάτων, θάνατος. N:Gmp | (= des morts) |

En plus, nous notons les premiers mots encodés du conte de Karagatsis :

Άγγλο,Άγγλος.N:Ams
 Άγγλος,.N:Nms
 Άκου,.N:Nms:Gms:Ams:Vms
 Άντριους,.N:Nms:Gms:Ams:Vms
 Άρχοντα,Άρχοντας.N:Gms:Ams:Vms:Nfs:Gfs:Afs:Vfs
 Άρχοντα,Άρχοντος.N:Gms:Ams:Vms:Nfs:Gfs:Afs:Vfs
 Άρχων,.N:Nms:Gms:Ams:Vms:Nmp:Gmp:Amp:Vmp:Nfs:Gfs:Afs:Vfs:Nfp:Gfp
 :Afp:Vfp

Ἄρχων, Ἄρχος.N:Gmp

Ἄτλας,.N:Nms

Ἐλα,.N:Nfs:Gfs:Afs:Vfs

Ἐλληνα, Ἐλλην.N:Ams

Ἐλληνα, Ἐλληνας.N:Gms:Ams:Vms

Ἐλληνες, Ἐλλην.N:Nmp:Amp:Vmp

Ἐλληνες, Ἐλληνας.N:Nmp:Amp:Vmp

Ἐμπα,.N:Nfs:Gfs:Afs:Vfs

Ἐνα,.N:Nfs:Gfs:Afs:Vfs

Ὀλα,.N:Nfs:Gfs:Afs:Vfs

Ἀγαπητέ, Ἀγαπητός.N:Vms

Ἀνδρέα, Ἀνδρέας.N:Gms:Ams:Vms:Nfs:Gfs:Afs:Vfs

Mais des outils supplémentaires pour l'analyse du texte littéraire sont approuvés nécessaires :

- un programme d'élaboration automatique des tableaux de mots
- un programme de tableaux contextuels pour trouver des structures morphosyntaxiques dans les textes. Ces structures apparaissent avec leur contexte afin de prendre en considération le sens qui est moins abordé.

Le traitement automatique d'un texte littéraire en grec dispose de quelques restrictions. Le texte doit être écrit sous la forme d'ASCII et sans typographies particulières ; pour le grec un problème important s'impose : la fidélité sur la forme du texte original parce que, même si aujourd'hui il n'y a qu'un accent, auparavant il y en avait plusieurs. Exemple, le conte de Manolis Karagatsis "Ένας χαμένος κόσμος"² dont la typographie contient plusieurs accents mais qui ne peuvent pas être reproduits pour être traité par le système UNITEX ; respectant donc cette restriction, nous avons réservé un seul accent. Nous citons le début du conte par la suite, la forme en ASCII et l'original :

² 'Un monde perdu' in Μάκης Πανώριος, *Το ελληνικό φανταστικό διήγημα*, Αίολος, Αθήνα, 1987, pp. 97 - 111 [Makis Panorios, *To eliniko fantastiko diigima*, Athènes, Eolos, 1987, p. 97 - 111].

I. Texte sous la forme d'ASCII.

ΜΙΧΑΛΗΣ ΚΑΡΑΓΑΤΣΗΣ

Ένας χαμένος κόσμος

Το καλοκαίρι του 1942, το ελληνικό υποβρύχιο «Νηρηίς» βρισκόταν σε αποστολή στον Ατλαντικό Ωκεανό, όχι μακριά από το Γιβραλτάρ. Την ημέρα περιπολούσε εν καταδύσει. Και τη νύχτα ανέβαινε στην επιφάνεια να φορτώσει τους συσσωρευτάς του. Μια νύχτα, λοιπόν, καθώς η «Νηρηίς» έπλεε στην επιφάνεια, ακούστηκε μακριά ο βόμβος ενός αεροπλάνου. Αμέσως ο κυβερνήτης υποπλοίαρχος Γιάννης Ρεΐζης που βρισκόταν στη γέφυρα, χτύπησε το κλάξον του κινδύνου και φώναξε:

- Ταχεία κατάδυσις!

Όσοι ναύτες ήσαν στη γέφυρα μπήκαν βιαστικά στο εσωτερικό του υποβρυχίου. Τελευταίος κατέβηκα ο κυβερνήτης αφού έκλεισε μόνος του την πόρτα της καθόδου. Κάτω, στο θάλαμο, ο υποπλοίαρχος Μανώλης Τρέκκας είχε αρχίσει κιόλας τους χειρισμούς καταδύσεως:

- Ηλεκτροκινητήρες πρόσω ολοταχώς! «Εξαεριστικά υδροδοτών» ένα και τρία! Πρωραία πηδάλια βάθους συν δέκα!

Το υποβρύχιο άρχισε να καταδύεται. Η σωτηρία των γενναίων ναυτικών και η διάσωση του σκάφους εξαρτιόταν από το βάθος που θα έπαιρνε το υποβρύχιο. Έξαφνα μια δυνατή έκρηξη συνεκλόνησε το σκάφος. Τα ηλεκτρικά λαμπιόνια έσπασαν και γίνηκε σκοτάδι. «Γρήγορα! Τις λάμπες του πετρελαίου!» φώναξε ο Ρεΐζης. Άναψαν μια λάμπα.

- Μας πέτυχε ο Γερμανός! είπε ο Τρέκκας.

- Θα του ξεφύγουμε, αποκρίνεται ο Ρεΐζης.

Εκείνη τη στιγμή, δεύτερη έκρηξη ταράζει το υποβρύχιο, ακόμα πιο δυνατή. Από την πόρτα του θαλάμου μηχανών μπαίνει ο δεύτερος μηχανικός κατακίτρινος. «Διαρροή στις μηχανές!» φωνάζει.

Την ίδια στιγμή, ακούγονται από την πλώρη φωνές τρομαγμένες: «Διαρροή στο πρωαίο διαμέρισμα! Πνιγόμαστε!» Τότε, μπροστά τον κίνδυνο, ο κυβερνήτης παίρνει τη σκληρή μα αναγκαία απόφαση: «Κλείσατε φράκτας και πόματα ασφαλείας!» διατάζει. Οι ναύτες εκτελούν τη διαταγή και κλείνουν τις υδατοστεγείς πόρτες που χωρίζουν το κεντρικό διαμέρισμα του υποβρυχίου από την πλώρη και την πρύμνη. Όσο για τους άλλους. Γι' αυτούς δεν υπήρχε πια καμιά ελπίδα... Ήταν καταδικασμένοι σε θάνατο.

Στο μεταξύ, τα διαμερίσματα της πλώρης και της πρύμνης γεμίζουν νερά από τα ρήγματα που προκάλεσαν οι γερμανικές βόμβες. Όσοι βρίσκονται εκεί, καταλαβαίνουν πως ήρθε η ώρα να πεθάνουν σαν Έλληνες! Στέκονται όρθιοι σε στάση προσοχής. Και με φωνή σταθερή τραγουδούν τον Εθνικό μας Ύμνο. Κι όταν πια το νερό έφτασε κοντά στο στόμα τους, όταν δηλαδή ήρθε η στιγμή του θανάτου, εφώναξαν όλοι μαζί: «Ζήτω η Ελλάδα!» και πνίγηκαν. Πέθαναν σαν πραγματικά ελληνόπουλα για το μεγαλείο της γλυκειάς Πατρίδας.

Le système INTEX, comme T. Kiriakopoulou explique dans son article « Η ανάλυση των λογοτεχνικών κειμένων : πλεονεκτήματα και περιορισμοί » ('L'analyse automatique des textes littéraires : avantages et restrictions')³ suit quelques étapes pour qu'un texte soit analysé. Le système, en premier lieu, repère tout caractère graphique dans le texte ainsi que sa fonction dans le contexte ; le point, par exemple, constitue le signe de ponctuation ou le signe des nombres décimaux. En deuxième lieu, après la détermination des signes graphiques, le type des mots doit être éclairci. Cette tâche présente de nombreuses difficultés en grec, et exige l'existence des dictionnaires électroniques contenant toutes les formes déclinables. Les problèmes qui sont créés, viennent souvent de mots suivants :

- les articles contractés (στο (= au), σε + το (= à + le))
- les mots composés
- les expressions figées

Nous ajoutons certains problèmes lexicaux qui sont apparus par le texte de Karagatsis, et qui exigent le renouvellement du dictionnaire électronique :

³ Communication en grec inédite.

➤ les mots qui constituent des restes du grec ancien ou de katharevousa, ainsi que quelques types des formes déclinales qui appartiennent soit à la conjugaison du grec ancien soit à la conjugaison du grec moderne (ήσαν, εκοίταξαν, εμίλησε, ενεργήσωμεν, ερρύθμιζε, εσήμαινε, εφώναξαν, εύρωμεν, τελειότης, σημείωσις).

➤ l'orthographe plus vieille de quelques mots la quelle est plus simple au grec moderne (βγήτε, γλυκειά, κατεβήτε, καμμία, λειβάδι, μακρυά).

Ces deux étapes alors, le repérage des signes graphiques et des types des mots constituent l'analyse morphologique dont les problèmes qui découlent vont être éclaircis par l'analyse syntaxique et, puis, sémantique.

Etudier les structures de tout type d'un texte ou d'une œuvre littéraire n'est pas toujours une procédure qui conduit à des conclusions fructueuses. Par contre, le choix d'une certaine structure doit être significative, et avoir un objectif concret. G. Mounin avait écrit à propos des structures dans un roman : « Il ne suffit pas de trouver dans une œuvre des structures, encore faut-il savoir quelle est la fonction de ces structures »⁴. Ainsi, par toutes ces structures qui constituent un conte fantastique, nous choisissons l'étude du champ lexical de la *mort* de quelques récits de la langue grecque parce que nous considérons que le thème de la mort est une composante de plusieurs contes fantastiques sinon le stimulus qui attire un écrivain.

Nous avons appliqué le système UNITEX au conte de Karagatsis. Il s'agit d'un conte fantastique qui distingue, du point de vue stylistique, de sa vivacité. La narration est enrichie de dialogues, fait qui est renforcé par l'utilisation des signes de ponctuation :

| | |
|-----|-----|
| 536 | {S} |
| 515 | . |
| 406 | , |
| 141 | - |
| 96 | ' |
| 81 | ! |
| 79 | " |

⁴ MOUNIN, G., *Travaux pratiques de Sémiologie générale*, Textes réunis et publiés par Alain Baudot et Claude Tatilon, Préface de Claude Tatilon, Toronto, Editions du GREF, Collection Theoria n. 3, 1994, p. 31.

60 ;

Pour passer à la *signification*, nous avons trouvé à propos du vocable *mort* les résultats ci-dessous :

S} Αλλά ξεχνάς το "Φράγμα του [Θανάτου](#)". - Δεν το ξεχνώ. {S}
 Ε
 ό ως ύποπτη. {S} Το φράγμα του [Θανάτου](#) έχει διακοπεί. {S} Έμπα
 ίο πως η Πύλη των ακτίνων του [Θανάτου](#) άνοιξε. {S} Και το όχημ
 θούν να περάσουν το "θόλο του [Θανάτου](#)", είπε ο Άτλας. {S} Κάθ
 ρίσκονταν κοντά στο "θόλο του [Θανάτου](#)". {S} Διέσχισαν ένα δάσ
 μόνο το ζήτημα του θόλου του [Θανάτου](#). {S} Πως θα περνούσαν τα
 α ιδεατά σύνορα του θόλου του [Θανάτου](#), θα χρειάζονταν περίπο
 άσουμε την Πύλη του θόλου του [Θανάτου](#). {S} Η Πύλη ήταν μια αφ
 λπίδα... Ήταν καταδικασμένοι σε [θάνατο](#). {S} Στο μεταξύ, τα διαμ
 αταδικάζει την Εάλη σε βέβαιο [θάνατο](#), μια και θ' ανακαλυφθεί
 πάρχει τρόπος να γλιτώσουν το [θάνατο](#). {S} Τότε, ο Ανδρέας βλέ
 διαμέρισμα, παρακολουθούν το [θάνατο](#) των συντρόφων τους με σ
 ς τους περίμενε ο φοβερώτερος [θάνατος](#) που υπάρχει: η ασφυξία
 κάνουν παρά να περιμένουν το [θάνατό](#)... - Για πόσες ώρες έχουμ
 όταν δηλαδή ήρθε η στιγμή του [Θανάτου](#), εφώναξαν όλοι μαζί: "

σουμε το θόλο των ακτίνων του θανάτου.{S} Άμα καταφέρουμε αυ
 άξαμε με ένα θόλο κυμάτων του θανάτου.{S} Κανένας ζωντανός ο
 δυσκολία δεν ήταν ο θόλος του θανάτου.{S} Αν πήγαιναν στον Α

Le système a repéré 18 emplois du terme, soit 0.117% de l'ensemble lexical du conte. Parmi ces emplois, nous pouvons discerner deux catégories principales qui concernent le champ notionnel de la *mort*. Le premier emploi fait allusion à la voûte (Θόλος του θανάτου, c'est-à-dire la voûte de la mort) et au barrage qui sépare l'univers de l'Atlantide avec le reste du monde humain ; c'est le seuil que les visiteurs ne peuvent pas franchir sans la permission des gardiens de l'Atlantide. Cette appellation renforce le doute du lecteur, et favorise l'équivoque du récit, car le narrateur même s'il nous laisse croire à la possibilité de l'existence du monde mythique de l'Atlantide, il nous rappelle, en même temps, qu'il s'agit de la porte de la mort ; il semble se contredire.

Le deuxième emploi renvoie à la mort physique – l'asphyxie – qui menace l'équipage du sous-marin. Il est renforcé par le verbe *mourir* (πεθαίνω ; 9 emplois, soit 0.058%) qui apparaît au discours de l'équipage devant la menace de la mort :

ω η Ελλάδα!" και πνίγηκαν.{S} Πέθαναν σαν πραγματικά
 ελληνόπ
 του οξυγόνου από το στόμα. - Πεθαίνω!{S} Συλλογιέται ο Άντρ
 , τα μάτια κι' ετοιμάσθηκε να πεθάνει γιατί ήταν βέβαιος πως
 ειναν τους "φράκτες" θα είχαν πεθάνει κι αυτοί.{S} Μακάρι να
 κι αυτοί.{S} Μακάρι να είχαν πεθάνει.{S} Έτσι, θα γλύτωναν
 ετε να βγήτε χωρίς κίνδυνο να πεθάνετε από την πίεση του νερ
 ς, πρέπει να ετοιμαστούμε, θα πεθάνουμε κι εμείς σαν Έλληνες

ταλαβαίνουν πως ήρθε η ώρα να πεθάνουν σαν Έλληνες! {S} Στέκο
να φύγω από την Ατλαντίδα θα πεθάνω. {S} Πρέπει να με βοηθήσ

Le champ lexical de la mort, en général, est assez riche aux synonymes. Le système UNITEX dans ce cas permet le passage à l'approche sémantique du texte littéraire. L'analyse d'un texte littéraire à l'aide du système peut porter des informations intéressantes aux études linguistiques et du style de l'écrivain. A part les études ayant comme but les objectifs statistiques l'UNITEX donna la possibilité d'étudier l'usage d'une unité lexicale dans le contexte, dans l'œuvre ou dans l'ensemble d'une œuvre d'un écrivain ou de plusieurs écrivains. Enfin, tout texte littéraire impose ses règles et assigne le renouvellement des dictionnaires électroniques.

Bibliographie

- ΚΥΡΙΑΚΟΠΟΥΛΟΥ, Τίτα, « Η ανάλυση των λογοτεχνικών κειμένων : πλεονεκτήματα και περιορισμοί » ('L'analyse automatique des textes littéraires : avantages et restrictions'), communication en grec inédite.
- MOUNIN, Georges, *Travaux pratiques de Sémiologie générale*, Textes réunis et publiés par Alain Baudot et Claude Tatilon, Préface de Claude Tatilon, Toronto, Editions du GREF, 1994.
- SILBERZTEIN, Max, *Dictionnaires électroniques et analyse automatique de textes. Le système INTEX*, Paris, Masson, 1993
- VUILLEMIN, Alain, *Informatique et Littérature (1950-1990)*, Genève-Paris, Slatkine-Champion, 1990.

- hypermedia.univ-paris8.fr/jean/infolit/infolit.htm
- hypermedia.univ-paris8.fr/jean/infolit/cours/infolit8.html
- www.epi.asso.fr/revue/81/b81p171.htm
- www.epi.asso.fr/revue/94/b94p051.htm
- www.revue-texto.net/Inedits/Mezaille_Etudier/Mezaille_Etudier4.html