

Bogna Opolska-Kokoszka

École Supérieure de Communication,
des Sciences Politiques et des Relations
Internationales de Varsovie

DE LA PHILOGIE NUMÉRIQUE AU SEIN D'UNE SÉMANTIQUE DES TEXTES

Nous avons renvoyé le problème de la compréhension à celui de l'interprétation.

L'interprétation part de la matérialité philologique des textes pour y revenir en leur assignant du sens ; pour la philosophie, c'est l'objectif de l'herméneutique. La sémantique interprétative n'est pas une herméneutique, car elle n'a pas de caractère philosophique.

(...) Les signes linguistiques sont le support de l'interprétation, non son objet.

Seuls des signifiants, sons ou caractères, sont transmis : tout le reste est à reconstruire.

En d'autres termes, l'interprétation ne s'appuie pas sur des signes déjà donnés, elle reconstitue les signes en identifiant leurs signifiants et en les associant à des signifiés.

L'identification des signes comme tels *résulte* donc de parcours interprétatifs (Rastier 1994b : 335)

0. INTRODUCTION

Nous nous proposons dans le présent article d'esquisser la conception de la philologie numérique proposée par François Rastier et d' expliciter son contexte, notamment le projet de la fédération des sciences traitant des langues et des cultures au sein de la sémiotique des textes faisant partie d'une sémiotique des cultures. Nous avons décrit ailleurs (Opolska-Kokoszka 2000, 2010) l'approche théorique et les instruments descriptifs, dans le cadre de sa méthode de l'analyse textuelle nommée originellement la sémantique descriptive unifiée (SDU), actuellement connue comme sémantique interprétative (SI).

Le projet, dont il est question, est la conséquence de la conception du texte de Rastier comme objet culturel (Rastier 1990, 1992a,b, 1994b, 2001, 2005a,b). L'évocation de sa description sera notre point de départ, pour passer ensuite au terme de la philologie numérique et à l'explication de ses origines.

Ensuite nous allons mentionner des méthodes de l'extraction du contenu de Web, auxquelles Rastier (2007, 2008, 2011) oppose une sémantique différentielle, et continuerons avec quelques exemples pratiques de cette méthode pour la détection des pages racistes sur Internet (Rastier 2006b).

1. TEXTE – OBJET PROPRE D'UNE SÉMANTIQUE DES TEXTES

Le postulat évoqué par Rastier depuis quelques années concerne la nécessité de créer une sémiotique des cultures. Dans son ouvrage récent (Rastier 2011, à paraître), il

propose de la considérer en tant qu'une méthodologie plutôt qu'une discipline. Il la décrit comme réflexion fédérative pour l'ensemble des sciences de la culture qui va à l'encontre de la décentralisation des disciplines traitant des langues et des cultures. Ce postulat est la conséquence de la critique du réalisme logique et ontologique qui se trouvent à la base de la théorie de la signification (conséquence de la conception *vox-conceptus-res*) qui ne peut pas être adaptée au sens textuel. L'extrait du chapitre qui suit consacré au texte et à la triade sémantique a été rédigé, en partie, d'après une publication antérieure (Opolska-Kokoszka 2001).

Comme un texte est pour Rastier avant tout un objet culturel, il le définit d'une manière négative :

1. Un texte n'est pas exclusivement une chaîne de caractères (ce qui est posé par l'informatique linguistique), 2. un texte n'est pas une suite d'instructions de tel ou tel type, 3. un texte ne peut pas être identifié à des schémas cognitifs compris comme modèles mentaux, scripts, plans. (Rastier 1996 : 20)

Surtout la première objection aura des conséquences pour sa vision d'une élaboration numérique à laquelle nous reviendrons plus tard.

Rastier (2001) rappelle que dans la tradition philologique la notion du texte a été limitée à l'écrit et opposée au discours comme l'écrit à l'oral. Il parle de « la dématérialisation numérique » que nous observons actuellement et il propose de dépasser l'opposition entre discours et texte dans cette définition positive grâce à la notion de support :

Un texte est une suite linguistique empirique attestée, produite dans une pratique sociale déterminée, et fixée sur un support quelconque. (Rastier 2001 : 21)

Un texte peut alors avoir une forme orale, ou écrite, peut se trouver en interaction avec d'autres objets sémiotiques. La définition met en valeur l'authenticité des textes qui ne sont pas coupés du contexte ni créés par les linguistes. Il est basé sur un support dont la nature ne détermine pas sa validité. Il appartient à une pratique sociale.

Pour bien préciser en quoi consiste la textualité il faudrait distinguer, selon Rastier, « le texte » et « les textes ». Le premier appartient au langage et le deuxième est l'objet empirique de la linguistique. Les textes sont des produits culturels de l'homme exprimés dans les langues.

1.1. SIGNE LINGUISTIQUE VERSUS CONCEPT

L'idée de la représentation et du signe linguistique est le point central de l'argumentation de Rastier (1990, 1991, 1994b, 2001), en ce qui concerne l'autonomie du niveau sémiotique (ou symbolique) du signifié appartenant à un monde textuel. Rastier oppose le signe linguistique à la conception aristotélicienne de la signification, qui est la conséquence de la triade mot/concept/chose. Il est distinct du concept philosophique qui instaure un lien direct entre les mots et les choses. Sa spécificité consiste dans la dualité signifiant/signifié. La relation signifiant-signifié est proprement linguistique tandis que la relation concept-référent suppose une réalité extra-linguistique et un aspect ontologique. Rastier critique donc la conception du réalisme logique et ontologique, selon lesquels les signes représentent des concepts qui, à leur tour, représentent des êtres.

Rastier considère la culture comme un système symbolique et les signes (plus exactement les signifiés) n'imitent pas la réalité mais ils la créent. Le rôle des signifiés est ainsi décrit comme la médiation entre le physique et le représentationnel. Ces précisions sont encore complétées par la description de l'impression référentielle qui devrait remplacer la notion de référence. Les signifiés doivent être distingués des représentations (images mentales en sciences cognitives) qu'ils entraînent. L'impression référentielle opposée à la référence est entendue comme référence philologique (monde textuel des signifiés) et non pas ontologique.

Le sens d'un texte est ainsi défini par Rastier (1992b : 90) comme :

Un ensemble de contraintes linguistiques sur la formation des représentations psychiques (...) il incombe ainsi à la sémantique de décrire les conditions différenciées des impressions référentielles.

Chaque texte présente des contraintes sur la formation des images mentales par ses structures sémantiques. Elles ne peuvent pas obéir aux règles de la linguistique restreinte, analogues à celles de la théorie des langages formels. L'élaboration des typologies des textes semble plus adéquate que la construction d'un système universel d'analyse textuelle. Cette idée de variété de textes et de leurs typologies nous introduit dans la problématique philologique.

2. PHILOGIE NUMÉRIQUE

La notion de support de texte a été introduite à l'occasion de la description du texte que nous venons de présenter. Rastier en évoquant cette idée rappelle son histoire qui a commencé par l'écriture, ensuite l'imprimerie pour arriver à la numérisation d'où le terme « philologie numérique ». Rastier (2001) en l'introduisant se réfère à Roberto Busa comme premier qui, à partir des années 1950, codait Thomas d'Aquin sur des cartes perforées et dont le projet s'appelait « philologie électronique ».

C'est seulement à partir de la découverte du scanneur et de l'informatisation de l'édition dans les années 1980 que le corpus informatisé devient important (la taille du British National Corpus augmente de un à dix millions de mots) et permet le progrès des traitements automatiques du langage (TAL) et des linguistiques de corpus. Rastier décrit la tâche des TAL (issus de la tradition logico-formelle) de la manière suivante :

TAL ont servi avec le chomskysme de carrière aux principales conceptions du cognitivisme, et ont été utilisés pour « prouver » des théories plutôt que de décrire des objets. À l'objectif de la formalisation, les cognitivistes ont ajouté la tâche de simuler la compréhension. (Rastier 2001 : 76)

Dans les années 1990, suite à l'évolution des traitements automatiques du langage et des nouvelles linguistiques du corpus, ainsi qu'à la quantité de textes disponibles, nous observons un progrès dans leur codage et leur structuration. Néanmoins, les unités supérieures au syntagme, les unités lexicales, les structures textuelles, les paramètres de genre échappent aux analyses. Dictionnaires, thésaurus, réseaux sémantiques sont tous favorisés, ce dont témoigne le succès de Wordnet, le réseau le plus développé aujourd'hui.

2.1. ORDRE HERMÉNEUTIQUE

Rastier propose donc de passer d'une problématique de la représentation à une problématique de l'interprétation, avec recours à l'herméneutique, pour pouvoir traiter les structures textuelles sémantiques.

A côté des ordres : syntagmatique, paradigmatique et référentiel, Rastier introduit l'ordre herméneutique propre aux conditions de la production et de l'interprétation des textes dans lequel il inclue aussi les facteurs pragmatiques.

Le projet intellectuel de Rastier vise à « réunifier la philologie et l'herméneutique au sein d'une sémantique des textes à vocation descriptive » (Rastier 2001 : 2).¹ Quand on interprète des signes dans un contexte on ne peut pas procéder comme dans le cas de l'analyse morpho-syntaxique .

Au niveau textuel les unités sémantiques ne se prêtent pas à l'analyse à travers les domaines comme l'analyse en mots-clés car, par exemple, une isotopie sémantique spécifique peut être constituée par les sémèmes appartenant aux différents domaines sémantiques.

L'aspect philologique fait que les textes ne sont pas considérés comme des données – chaînes de caractères (ce qui est le cas de TAL), mais comme conséquences de procédures différentes comme : inscription, établissement, système graphique et variabilité de sens.

L'inscription concerne la transposition du texte sur un support et son changement modifie chaque fois le texte. L'établissement dépend de l'interprétation, ce qui est conforme à l'herméneutique philologique qui mène à l'emendatio. L'établissement constitue l'illustration d'une forme du cercle philologique car il est lui-même issu d'une interprétation et permet la création des autres. Le système graphique d'un texte entraîne plusieurs sémiotiques (au palier graphémique, par exemple, il s'agit de lettres et diacritiques, de paragraphes, chapitres, au palier séquentiel). Le balisage (tagging) est considéré par Rastier comme une dénomination nouvelle mais un procédé connu car il rappelle les letrines médiévales qui peuvent être identifiées comme une sorte de balisage.

On peut distinguer quatre types de balises qui sont des points importants dans le processus de l'interprétation : articulation (p. ex. une position dans la section), étiquette (une glose minimale – p.ex. prosodique), indice (un point local hétérogène – p.ex. soulignement), ancre (p.ex. appel de note). Les systèmes de codage sont polysémiotiques.

L'interprétation n'a pas de caractère stable, change avec les objectifs et les instruments de la description et varie avec la pratique sociale. Tout le texte ne peut pas être interprété que dans un corpus des textes du même genre et du même discours. Les textes doivent être intégraux et non pas les échantillons de textes car ils sont des objets culturels à analyser dans leur histoire et dans leurs contextes inter-textuel et inter-sémiotique. Un corpus est donc un ensemble de documents et non pas de données. Rastier souligne cette différence quand il analyse le système Web.

¹ Rastier parle d'une herméneutique matérielle ou philologique en l'opposant à l'herméneutique philosophique. Elle a pourtant sa méthodologie qui unit la critique philologique et le comparatisme, elle admet en plus une relativité historique.

2.2. WEB DES DONNÉES VERSUS WEB DES DOCUMENTS

La prolongation du programme classique de la représentation des connaissances (dont nous venons de parler), est assurée par le Web sémantique (Web des données) qui est opposé par Rastier (2008) à la sémantique du Web (Web des documents).

Le Web sémantique est basé sur une sémantique dénotationnelle et ceux qui mettent en ligne des textes doivent suivre les règles d'une infrastructure appelée « cake ». Cela entraîne un système normalisé de métadonnées qui représentent le contenu comme un thésaurus (emploi des termes décontextualisés) où la donnée se présente comme une chaîne de caractères.

Ce système de consultation que constitue le Web sémantique est un programme politico-technique² qui réduit la langue à une nomenclature, qui ne rend pas compte des structures textuelles tandis que la sémantique du Web est un projet méthodologique fondé sur une sémiotique des corpus. Ce dernier est issu de la linguistique de corpus pour laquelle les connaissances se trouvent dans les textes et leurs extractions sans tenir compte de valeurs contextuelles (ce qui est le cas du Web sémantique), les privent de pertinence.

Le chapitre qui suit décrit un exemple de la mise en œuvre d'une approche sémantique du document Web. Cette approche se veut combler les lacunes d'un filtrage ordinaire par mots-clés.

3. « PRINCIP »³ – UN EXEMPLE DE L'APPROCHE SÉMANTIQUE DU DOCUMENT WEB

François Rastier (2006b) présente un projet européen de création d'un système de détection automatique des sites racistes, nommé « PRINCIP » comme exemple de la sémantique du Web.

L'établissement du corpus admet une précompréhension pour discerner les caractéristiques les plus discriminants et exclue tout à priori.⁴ Le langage raciste peut être utilisé contre tout un groupe (ethnique mais aussi concernant les handicapés, les homosexuels, etc.) dans un même discours insultant. La source du racisme peut être « de gauche » comme « de droite ». Ainsi ni la source, ni la cible du racisme ne semble pertinente pour l'analyse.

La méthode employée est différentielle et met en œuvre le contraste qui se crée entre les sites racistes et non racistes. Cette méthode est jugée plus efficace qu'un simple contraste avec une « langue générale » car elle permet de détecter un système de normes sémantiques (axiologies) propres à ces deux discours qui s'affrontent. L'établissement de ces systèmes de valorisation, constitue cette précompréhension et cette

² Rastier soulève le problème économique du soutien du Web sémantique par le Département de Commerce des États-Unis.

³ C'est un système multilingue : anglais, allemand, français (<http://www.principle.net>).

⁴ On aura tendance à croire que les antiracistes se présentent comme victimes de l'agression des racistes et il s'avère que c'est le contraire.

interprétation préalable qui permet de distinguer les éléments discriminants qui peuvent être considérés comme équivalents des mots-clés. Ce procédé est illustration du cercle philologique, dont nous avons parlé plus haut, on part du texte pour revenir après à d'autres textes du même corpus en testant la pertinence du choix du système de description préétablie.

L'analyse montre que les mots isolés sont moins discriminants que les syntagmes (différence entre les expressions « génocide » et « génocide blanc », ce deuxième utilisé par les racistes). Les néologismes sont un exemple intéressant plus fréquemment utilisés sur les sites racistes. Ils prennent un aspect satirique et introduisent le registre de l'injure, comme des syntagmes descriptifs, souvent visant une personne ou un groupe précis (exemple de Jean-Marie Le Pen désignant la journaliste Anne Sinclair de « charcutière casher »).

Les néologismes sont un moyen de corrélation entre des dimensions et des domaines sémantiques séparés ou même opposés (p.ex. fédérastes, saddamiser).

En établissant les systèmes de valorisations Rastier pose l'hypothèse suivante :

« Il y a le racisme quand la 'race' illustrée est présentée comme 'victime' ; sinon on a affaire à une 'simple' xénophobie » (Rastier 2006). Cette hypothèse s'appuie sur un schéma binaire qui oppose « Nous » et les « Autres ». Ces derniers se divisent en deux groupes : les « allogènes » et les « cosmopolites ». Chez ces deux groupes, il n'y a pas de désignation de l'ennemi. En se posant en rôle de victime les racistes ont l'impression de faire appel à de justes violences car « Nous » sommes opprimés.

Comme le décrit l'auteur, la péjoration de l'Autre se fait sur tous les codes sensoriels (odorat, goût, etc.) car le domaine politique n'est pas totalement privilégié. Ce qui est donc caractéristique pour la pensée raciste (ou mythique) c'est le passage d'un domaine sémantique à un autre.

L'opposition de Nous aux Autres fait appel à d'autres oppositions des domaines comme le nombre, l'identité, la pureté, l'évaluation qui organise l'univers sémantique.⁵

Ces catégories sémantiques que nous venons de citer ne sont pas spécifiques pour le discours raciste. Elles peuvent caractériser chaque cadre spatio-temporel et mode d'action. Parmi les champs sémantiques les plus caractéristiques, nous avons : « l'invasion », « l'animalité » et « l'impureté ».

Il y a aussi d'autres difficultés qui font la détection automatique impossible. Il s'agit des citations, qui sont le plus souvent utilisées sur les sites antiracistes, ensuite la parodie qui aura pour but d'imiter l'ennemi et la prosopopée qui donne la parole à l'ennemi.

Il existe donc plusieurs stratégies de masquage. Ils vont dans leur discours du politique au scientifique, du scientifique au judiciaire etc. Une autre stratégie de masquage assez fréquente est d'écrire des articles de polémique sous forme d'articles scientifiques.

D'autres masquages, moins évidentes, rendent la description encore plus difficile. Voici quelques exemples :

⁵ Quelques exemples de ces oppositions : « Nombre » : Unique vs Multiple ; « Identité » : Homogène vs Hétérogène ; « Pureté » : Pur vs Impur ; « Évaluation » : Positif vs Négatif ; « Position » : Centré vs Périphérique, Supérieur vs Inférieur, Frontal vs Dorsal ; « Aspects » : Perfectif vs Imperfectif, Ponctuel vs Itératif.

« L'euphémisation »⁶ où il suffit d'employer le langage du politiquement correct et qui permet de créer des connivences, « l'allusion », qui a toujours été un moyen puissant de propagande prenant même la forme satirique et contournant la censure et pour finir le « cryptage » qui est une sorte de forme sectaire de l'allusion et qui emploie des signes de reconnaissance. Les stratégies de masquage employées dépendent le plus souvent du cadre juridique du pays concerné.

Pour finir, François Rastier nous propose le résumé illustrant les choix méthodologiques de nouveaux observables et les enjeux théoriques. L'exploitation du corpus se divise en étapes suivantes :

1. Recrutement du corpus suivi d'une catégorisation qui répartit les documents dans deux classes (raciste et antiraciste).
2. L'étude qui engage une caractérisation qui décrit les positions axiologiques. L'application doit catégoriser les sites en deux classes grâce à la discrimination.
3. Cycle de validation : schéma incluant le recrutement du corpus par les mots-clés, sélection des variables sensibles, association de chaque variable comme règle de discrimination et son indexation, test sur le corpus, extension à des nouveaux corpus et finalement mise à jour des règles de discrimination pour être au courant de l'évolution des sites racistes.

Les indices du système PRINCIP appartiennent le plus souvent à des unités linguistiques peu décrites en analyse du discours, à des non-mots ou à des niveaux négligés par les études textuelles comme la police de caractères, couleurs, capitales, etc. Ces indices échappent à la conception classique d'un discours argumentatif. Les structures des sites racistes et des sites antiracistes sont différentes. Sur les sites racistes, il faut souvent cliquer plusieurs fois pour accéder aux documents que nous désirons, dépasser les mises en garde, tandis que les sites antiracistes mettent chaque analyse à la ligne. Pour faire un système de recherche de sites racistes pertinent, il ne suffit donc pas de procéder à une détection par mots-clés et de ne pas réduire les phrases en suite de caractères.

Rastier nous présente trois directions vers lesquelles les recherches devraient se poursuivre :

1. Codage des corpus.
2. Linguistique de corpus permettant à présent une observation des normes.
3. Extension de l'analyse sémantique à la corrélation contenu/expression.

CONCLUSION

Ce propos, vu l'ampleur intellectuelle du projet de Rastier, n'a pu être qu'une sorte d'introduction à la problématique de l'analyse sémantique de corpus. Sémantique de corpus (à vocation différentielle) fait partie d'une linguistique de corpus, elle-même se trouvant dans le projet fédérateur d'une sémiotique des cultures. Rastier appelle cette approche sémantique « le saussurisme renouvelé »⁷ et souligne que sa pertinence relève

⁶ Au lieu d'utiliser « arabe » ou « noir » on a « ethnique ».

⁷ Il s'agit avant tout de la dualité du signe linguistique et des signifiés qui sont à reconstruire dans le processus de l'interprétation.

de ses liens avec la philologie (responsable pour le recueil critique et l'indexation des documents).

Pour conclure, nous rappelons que cette sémantique de corpus est constituée par la philologie (pour établir et documenter les textes) et une herméneutique matérielle (pour interpréter des textes en partant du principe que : « Les textes ne signifient pas moins par leurs passages jugés obscurs que par les passages censés clairs » (Rastier 2005b).

Rastier (2011) distingue quatre régimes de complexité par rapport au corpus : sémiotique (unissant contenu et leur expression), intersémiotique (relation du texte aux sémiotiques non verbales), textuelle (relations entre le niveau global et local du texte) et inter-textuelle (rapports entre textes du même corpus) d'où la nécessité d'une approche intégrant aussi bien facteurs sémiotiques que sémantiques dans l'analyse des textes.

BIBLIOGRAPHIE

- OPOLSKA-KOKOSZKA BOGNA, 2000, *Le discours philosophique d'Emmanuel Levinas selon la sémantique interprétative*, thèse de doctorat soutenue à la Faculté de Philologie de l'Université Jagellonne de Cracovie.
- OPOLSKA-KOKOSZKA Bogna, 2001, *Sémantique interprétative pour linguistes et littéraires*, Opera Romanica, Édition Universitatis Bohemiae Meridionalis, České Budějovice, 59–65.
- OPOLSKA-KOKOSZKA Bogna, 2010, François Rastier – sens odnaleziony, (in :) *En quête de sens. Etudes dédiées à Marcela Świątkowska. W poszukiwaniu znaczeń. Studia dedykowane Marceli Świątkowskiej*, J. Górniewicz, H. Grzmil-Tylutki et I. Piechnik (eds), Kraków : Wyd. UJ, 382–388.
- RASTIER François, 1989, *Sens et textualité*, Paris : Hachette.
- RASTIER FRANÇOIS, 1990, La triade sémiotique, le trivium, et la sémantique linguistique, *Nouveaux actes sémiotiques*, 9, 5–39.
- RASTIER François, 1991, *Sémantique et recherches cognitives*, Paris : PUF.
- RASTIER François, 1992a, La généalogie d'Aphrodite, *Littérature* 87, 105–121.
- RASTIER François, 1992b, Réalisme sémantique et réalisme esthétique, *TLE* 10, 81–91.
- RASTIER François, 1994a, *Sémantique pour l'analyse*, Paris : Masson.
- RASTIER François, 1994b, Sur l'immanentisme en sémantique, *Cahiers de Linguistique Française* 15, 325–335.
- RASTIER François, 1996, La sémantique des textes-concepts et applications, *Hermes. Journal of Linguistics*, 15–37.
- RASTIER François, 2001, Arts et sciences du texte, Paris : PUF.
- RASTIER François, 2005a, Enjeux épistémologiques de la linguistique de corpus, (in :) Williams (éd.), *La Linguistique de corpus*, Rennes, Presses Universitaires de Rennes, 31–46.
- RASTIER François, 2005b, Herméneutique et linguistique : dépasser la méconnaissance, *Texto!* (en ligne) X, 4.
- RASTIER François, 2006a, Formes sémantiques et textualité, *Langages* 163, 99–114.
- RASTIER François, 2006b, Sémiotique des sites racistes, *Mots* 80, 73–85.
- RASTIER François, 2007a, Du réalisme au postulat référentiel, *Texto!* (en ligne).
- RASTIER François, 2007b, Passages, *Corpus* 6, 127–162.
- RASTIER François, 2008, Sémantique du Web vs semantic Web ? Le problème de la pertinence, *Texto!* (en ligne) XVIII, 3.
- RASTIER François, 2011, *Sémiotique et linguistique de corpus*, *Signata* I(1), (à paraître), 9–34.

Summary

About numerical philology as a part of the text semantics

The aim of the article is to outline the concept of numerical philology, as proposed by François Rastier the French linguist, in the context of a project to federate disciplines engaged in languages and cultures as part of the semiotics of cultures. This project is the consequence of the concept of text as a cultural object. The starting point for the article will be the presentation of that position followed by a discussion of the issue of numerical philology and an explanation of its roots. The next part will present the methods for obtaining text from the Web which Rastier compares with the text semantic method, a constituent of which is numerical philology. The final chapter of the work will show practical examples of the use of this method in identifying racist websites.

Streszczenie

Filologia numeryczna jako część semantyki tekstów

Celem artykułu jest zarysowanie koncepcji filologii numerycznej, zaproponowanej przez François Rastiera, współczesnego francuskiego językoznawcę, w kontekście projektu federacji dyscyplin zajmujących się językami i kulturami w ramach semiotyki kultur. Projekt, o którym mowa, stanowi konsekwencję koncepcji tekstu jako obiektu kulturowego. Punktem wyjścia artykułu jest więc przedstawienie tego stanowiska oraz omówienie problematyki filologii numerycznej i wyjaśnienie jej źródeł. W następnej części zostały zaprezentowane metody pozyskiwania treści z systemu informacyjnego Web. Rastier przeciwstawia im metody semantyki tekstowej, której częścią jest filologia numeryczna. Praktyczne przykłady wykorzystania tej metody do wyszukiwania rasistowskich stron internetowych stanowią końcowy rozdział pracy.