



Mots. Les langages du politique

80 | 2006

La politique mise au net

Sémiotique des sites racistes

François Rastier



Édition électronique

URL : <http://journals.openedition.org/mots/17113>

DOI : 10.4000/mots.17113

ISSN : 1960-6001

Éditeur

ENS Éditions

Édition imprimée

Date de publication : 1 mars 2006

Pagination : 73-85

ISBN : 2-84788-095-X

ISSN : 0243-6450

Référence électronique

François Rastier, « Sémiotique des sites racistes », *Mots. Les langages du politique* [En ligne], 80 | 2006, mis en ligne le 01 mars 2008, consulté le 17 mai 2019. URL : <http://journals.openedition.org/mots/17113> ; DOI : 10.4000/mots.17113

Méthodes

François Rastier

Sémiotique des sites racistes

En marge du projet européen Princip consacré à la détection automatique de sites racistes¹, les réflexions et analyses qui suivent ont trouvé leur source dans un travail collectif dont j'ai plaisir à remercier ici les participants, en gardant toutefois la responsabilité de mes propos.

On estimait début 2004 à 10 000 le nombre, toujours croissant, des sites racistes ou xénophobes – contre 4 000 début 2002. Moyen de propagande, le web est devenu aussi un lieu de rencontre fédérateur. Cela ne concerne évidemment pas que les extrémistes: selon une estimation raisonnable, un quart des collégiens a visité en toute bonne foi au moins un site raciste ou négationniste, puisqu'il suffit de taper *holocauste* pour lire du Garaudy ou du Faurisson.

Éviter le « politiquement correct »

L'indignation miséricordieuse a pu donner naissance à des formes de conformisme, certes bénignes et bien intentionnées, mais somme toute peu efficaces. La condamnation reste utile dès lors qu'elle ne s'arrête pas à la bonne conscience et ne dispense pas de comprendre: cependant, le pathos et l'appel aux sentiments restent des ressorts éprouvés du racisme.

Le recueil d'un corpus requiert une précompréhension: l'analyse suppose

1. PRINCIP est un système multilingue (anglais, allemand, français) de détection des documents racistes sur Internet. Fondé sur une approche sémantique du document web qui intègre tous les paliers de la complexité textuelle, il vise à pallier les insuffisances des systèmes de filtrage ordinaires qui reposent, le plus souvent, sur une simple liste de mots clés (voir <http://www.princip.net>). Par ailleurs, et en complément, le projet PRAX (2004-2005) vise à réaliser un matériel pédagogique et un outil d'enseignement assisté par ordinateur (<http://crim.inalco.fr/prax/>).

Crim, Inalco, Lpez@ext.jussieu.fr

de reconnaître ce qui fait système, d'entrer dans le propos du racisme. Ainsi, pour constituer le corpus d'apprentissage initial du projet Princip, nous avons d'abord recueilli les « plus beaux » sites racistes: concentrant les caractéristiques les plus discriminantes, ils semblent à leur façon exemplaires.

Si le politiquement correct préjuge volontiers des victimes et des agresseurs, il convient toutefois d'éviter tout a priori car la source et la cible du racisme demeurent éminemment variables:

– la *cible* semble une variable non pertinente, car le langage du racisme peut être utilisé contre n'importe quel groupe: juifs, Palestiniens, handicapés, homosexuels, etc., peuvent être l'objet du même discours déshumanisant et animalisant. Ainsi, on trouve à présent des sites où l'on explique comment nourrir son chien avec de la viande de Palestinien, et il n'est plus rare que des stéréotypes antijuifs soient appliqués aux musulmans: « Les fils d'Allah se multiplient comme des rats »². Dans les faits, on remarque la même variabilité des cibles: tel groupe de skinheads de Reims, parti « à la chasse aux Arabes », a naguère assassiné un homosexuel;

– la *source* est également une variable non pertinente. Ainsi, le racisme « de gauche » a les mêmes fondements que le racisme de droite. Il nous paraît a priori peu pertinent qu'un site se réclame d'une religion ou d'une autre, d'un camp ou d'un autre: si, par exemple, l'auteur d'un site nomme un kamikaze palestinien un « déchet explosif », peu importe qu'il se réfère ou non au judaïsme, il emploie le langage caractéristique du racisme et sera décelé comme tel par le système.

La méthode différentielle

En étendant aux documents multimédias les principes de la sémantique différentielle de tradition saussurienne, nous nous sommes éloigné de la pratique ordinaire dans les milieux du traitement automatique du langage. Au lieu de procéder à l'extraction de connaissances pour construire une ontologie (qui serait ici l'ontologie du racisme), nous avons pris le parti de contraster un corpus de sites racistes et antiracistes. C'est en effet leur contraste qui permet d'accuser les spécificités du discours raciste, à tous les niveaux, jusqu'à celui des étiquettes html. Si nous avons cherché à contraster d'emblée les sites racistes avec un problématique corpus neutre, supposé représenter un échantillon du web général, nous n'aurions pas pu déceler leurs traits caractéristiques. Un discours peut se caractériser par rapport à d'autres discours, mais il serait improductif de le contraster avec une « langue générale », qui reste un artefact sans doute illusoire des linguistes.

2. O. Fallaci, 2002, *La rage et l'orgueil*, Paris, Plon, p. 145.

La question des positions et l'impossibilité d'une ontologie

Les domaines sémantiques instaurent des « fonds » sémantiques communs correspondant à des pratiques reconnues et normées. Cependant, ils ne sont partagés qu'en partie: par exemple, certaines unités lexicales ne se rencontrent que dans certaines « positions de discours ». Même les éléments « référentiels » communs véhiculent des valorisations différentes: ainsi, dans le discours raciste, les anti-négationnistes sont-ils nommés « exterminationnistes » par leurs adversaires.

Les positions sont en nombre a priori indéterminé, mais en tout domaine on en trouve plusieurs. Chacune se traduit par: 1 / un *choix de genres* de prédilection – satiriques, par exemple; 2 / un *usage* particulier de ces genres, engageant une thématique, une dialectique (formes narratives ou argumentatives), une dialogique qui se traduit par la fréquence de grammèmes particuliers (ex.: *notre* dans le discours sectaire), des expressions comme [*sic*] (caractéristique de la polémique de droite comme de gauche); 3 / à chaque *position* dialogique correspond un *ethos* (au sens rhétorique), c'est-à-dire une forme de modalisation du propos qui s'exprime par la représentation de l'énonciateur et du destinataire, comme par toutes sortes d'évaluations.

Si l'image idéale d'une classification ontologique appartient au discours scientifique ou technique, le discours mythique manie des valeurs et, dans sa version politisée qui nous occupe ici, les répartit selon des *positions* correspondant sociologiquement ou politiquement à des *camps* qui ne partagent ni les mêmes objets, ni a fortiori la même vision sur ces objets. Ainsi, chaque position crée en quelque sorte son *axiologie* ou système de normes sémantiques. Les axiologies propres à des positions sont des systèmes de valorisations doxales qui s'affrontent et interdisent la constitution d'ontologies neutres par définition car purement référentielles. Même si chaque position contient une image de l'autre, cette image lui appartient en propre.

L'insuffisance des mots clés et les critères lexicaux

Dans l'état de l'art des étiqueteurs morphosyntaxiques, la syntaxe des textes reste peu caractérisante³; leur lexique l'est davantage, mais reste à décrire en contrastant des corpus, sans s'appuyer sur des mots clés répertoriés dans des

3. L'analyse morphosyntaxique en temps réel reste difficile – du moins pour un logiciel de détection qui doit pouvoir être annexé aux navigateurs du commerce. En outre, les règles de différenciation des positions qu'elle permet n'ont rien d'absolu. Par exemple, dans les textes racistes, *blanc* et *national* sont fréquemment des noms, mais uniquement des adjectifs dans les textes antiracistes; toutefois, il reste difficile de tirer de ce type d'exemple une règle de substantivation « raciste »: a contrario, si *juiif* est fréquemment un nom dans les sites racistes, il en va souvent de même dans les sites antiracistes.

ontologies. Il diffère en effet selon les positions: par exemple, le mot *homme*, évidemment fort rare dans les sites racistes, peut devenir un indice alors que personne ne penserait à en faire un mot clé. *Homme* appartient au vocabulaire anti-raciste⁴, alors que dans toutes les ontologies *africain* ou *européen*, *Pierre* ou *Paul* seraient sous la dépendance du terme superordonné *homme*.

Les syntagmes sont évidemment plus discriminants que les mots isolés : ainsi, l'expression *égalité des races* ne se trouve que sur les sites racistes; tous parlent de *génocide*, mais l'expression *génocide blanc* reste caractéristique de ces sites.

Comme, au demeurant, les lexies intégrées de longue date, les néologismes sont des formations de discours qui ne peuvent figurer dans les ontologies: où placer *crouillophile* ou *licrasseux*? La fonction des néologismes, singulièrement fréquents sur les sites racistes, mérite l'attention. Le mot rare distingue l'orateur⁵ et, à travers lui, l'auditeur; il introduit du nouveau, dans un discours qui prône un changement radical. À cela s'ajoute, dans le cas des néologismes, un aspect satirique qui met les rieurs de son côté et discrédite l'adversaire sur le registre de l'injure. Plus profondément, l'invention néologique revêt une fonction mimétique et crée, par un effet de réel, les répugnants hybrides que dénonce le discours: ainsi, les deux morphèmes de *crouillophile* s'opposent tant par leur niveau de langue (vulgaire vs savant) que par leur contenu évaluatif (péjoratif vs mélioratif). De même, les *fédérastes* (Jean-Marie Le Pen) conjoignent-ils l'abomination politique du fédéralisme et la contre-nature de la pédérastie. Polémiques, les modes compositionnels des néologismes se retrouvent dans les syntagmes descriptifs: ainsi de *charcutière casher* (J.-M. Le Pen) pour désigner la journaliste Anne Sinclair.

Qu'il s'agisse de dérivations ou de mots-valises, de lexèmes ou de syntagmes inédits (comme le syntagme-valise *Durafour crématore*), les néologismes sont un moyen éprouvé de corrélérer des dimensions, domaines ou champs sémantiques ordinairement séparés, voire opposés. Ainsi de la politique et de la sexualité: on se souvient du général Vanuxem, revenant de Dien Bien Phu pour dénoncer les Français « étendus sur les plages, prêts à toutes les sodomies ». Quand on crée *fédéraste* ou *saddamiser*, on procède de même, non plus au palier de la proposition mais à celui du mot.

En passant du politique au sexuel, le langage de la polémique naturalise son propos: il s'agit de rabattre le politique sur le biologique, ce qui est comme on sait l'objectif définitoire du racisme. Ainsi, au palier du mot, de la phrase, du texte entier, le même projet politique s'exprime avec des moyens différenciés. Il en va de même, nous le verrons, sur le plan de l'expression: vraisemblablement,

4. En revanche, le mot *dieu* est rarissime dans le corpus antiraciste.

5. On sait que Joseph Goebbels était réputé pour ses mots rares, et Jean-Marie Le Pen use d'un lexique remarquablement étendu.

le substrat mythique du racisme – appuyé même sur les mythes scientistes – lui permet une élaboration « littéraire » absente des sites antiracistes.

Bref, la mise en contraste des corpus confirme qu'aux deux positions raciste et antiraciste correspondent des systèmes lexicaux différents, indices sans doute de systèmes de valorisations différents. Si le vocabulaire des deux positions se recoupe en partie, les différences restent significatives: le nom *Hitler* est par exemple beaucoup plus fréquent sur les sites antiracistes ; et le mot *bougnoles* leur appartient en propre, car il correspond à l'image quelque peu désuète que les antiracistes se font des racistes. Il semble en somme que chaque camp emploie les mots qu'il croit que l'autre a en tête mais évite d'employer. Du moins, les antiracistes mentionnent-ils fréquemment les dirigeants adverses (J.-M. Le Pen, Bruno Mégret, Martine Lehideux...), des leaders étrangers (Jörg Haider) et des idéologues divers (Martin Heidegger...). Toutefois, comme les racistes sont sentimentaux à leur manière, ils restent les seuls à mentionner leurs martyrs, Robert Brasillach ou Rudolf Hess.

On retrouve ici le problème de la paradiastole, dont on peut donner pour exemple *Vos héros sont des assassins*. Quand des termes se répondent d'une position à l'autre, c'est sans se correspondre, voire en s'opposant: ainsi de *national vs xénophobe*, ou d'*invasion vs flux migratoires*. Même les repères spatiaux et temporels diffèrent: les antiracistes citent des sigles, lieux et dates, car ils s'inspirent du discours historique, alors que les racistes se meuvent dans le mythique, qui appelle une sorte d'intemporalité et d'exemplarité: les prénoms étrangers, comme *Ahmed*, *Mohammed*, *Aziz*, souvent employés sans nom de famille, campent des types et non des individus. Le projet de construire une arborescence unique serait d'autant plus absurde que les renvois entre domaines sémantiques sont constants. Il n'y a pas d'arborescence du discours raciste, car tout est transposable en tout domaine: les métaphores constantes, entre la vie sociale et la vie animale notamment, revêtent une fonction structurelle dans la construction d'un champ « biopolitique ».

Une conception du monde social

Pour trouver des mots clés et appeler les premiers sites du corpus, les insultes ne suffisent évidemment pas et suscitent beaucoup de bruit; c'est plutôt la description de quoi « Nous » (les Occidentaux, euphémisme pour « blancs ») serions victimes qui est révélatrice. D'où cette hypothèse: il y a racisme quand la « race » illustrée (généralement blanche)⁶ est présentée comme *victime*; sinon, on a affaire à une « simple » xénophobie.

6. Mais pas obligatoirement. Dans un journal ivoirien pro-gouvernemental, *Le National*, le leader de l'opposition Alassane Ouattara, marié à une Française, est appelé le « négro américain et

Le propos s'appuie sur un schéma binaire⁷ qui oppose *Nous* et les *Autres*. Les *Autres* se divisent en deux groupes: les *allogènes* qu'on doit renvoyer chez eux et les *cosmopolites* qui ne sont nulle part (une solution s'imposera); les premiers forment une « lie sociale parasitaire », les seconds une « ploutocratie naturellement internationale ». On s'adresse aux classes moyennes victimes de ces basses et « hautes » classes. Juifs, francs-maçons, marxistes, capitalistes, fédéralistes, la désignation de l'ennemi semble incohérente, mais peu importe car son identité affreusement mêlée est réputée masquée par le complot; d'où, par exemple, le rôle de premier plan donné au B'nai Brith américain, groupe maçonnique juif des plus marginaux. L'indistinction catégorielle de l'ennemi fait elle-même partie de son impureté métissée et peut être rapportée au complot qu'il ourdit.

L'histoire connaît trois phases: indépendance nationale de naguère, oppression actuelle et justice future. Ce schéma est celui du récit héroïque indo-européen où le héros a pour charge de restaurer l'âge d'or. Ainsi la Révolution nationale a-t-elle pour but la restauration de la Nation. Comme « Nous » sommes opprimés, il faut résister: on emprunte le vocabulaire de la Résistance, en appelant par exemple les juifs des *collabos*, pour s'opposer au génocide (des Blancs). En se posant en victime, on appelle à de justes violences.

Cet extrémisme tient le langage de la modération: c'est pour rétablir des limites que l'on prépare des exterminations. L'évaluation, et surtout les dépassements de seuils évaluatifs sont donc fondamentaux pour le racisme: d'où la fréquence du vocabulaire de la restriction ou de l'excès, jusque dans les grammèmes (comme *tellement, assez, trop, ne... que*, etc.; voir Mathieu Valette, 2004⁸), le thème du seuil de tolérance, etc.

Les arguments topiques sont récurrents et, semble-t-il, internationaux: *Les X nous pillent, Les X nous submergent*, etc. Privilégiant l'isotopie biologique, la thèse biopolitique remotive la banale animalisation insultante qu'illustrent la vermine, les rats, les poux, les insectes divers, innombrables et néfastes. Elle appuie un discours sur la pureté ethnique, le nettoyage salvateur – le Zyklon B employé dans les chambres à gaz, d'abord créé pour la lutte contre les insectes qui dévastaient les forêts allemandes, était utilisé aussi pour la dératisation.

Contrairement aux apparences, le domaine politique n'est donc pas toujours privilégié, même si l'on peut penser qu'il est déterminant. Par exemple,

sa juive blanche » (cité par *Le Monde*, 8 février 2003, p. 3). On retrouve par hypallage de riches associations « négro + blanc, américain + juif » qui rappellent Louis-Ferdinand Céline sur les « judéo-nègres ».

7. D'ascendance religieuse, il est hérité, par divers intermédiaires que nous ne pouvons détailler ici, du manichéisme antique.
8. M. Valette, 2004, « Détection et interprétation automatique de contenus illicites et préjudiciables sur Internet », P. Enjalbert et M. Gaio éd., *Approches sémantiques du document numérique* (Actes du 7^e Colloque international sur le document électronique, 22-25 juin 2004).

on retrouve une péjoration de l'Autre sur tous les codes sensoriels; ainsi chez J.-M. Le Pen, se répondent l'odorat (*empuantit*), le goût (*fruit amer*) et l'ouïe (*mélodées moroses, tintamarre des exotismes de pacotille vs péan haut et clair des combats antiques, chant de guerre et de victoire*).

La biopolitique s'exprime aussi, bien entendu, sur une isotopie politique qui, en vertu d'une conception policière de l'histoire, se résume au complot d'un lobby. Comme toute forme sémantique, ce scénario est éminemment transposable: *Le lobby juif s'est infiltré dans nos organes d'information, L'immigration épuise les budgets sociaux, Les poux sucent le sang* deviennent ainsi des propositions homologues. L'idéologie, précisément, joue de ces transposabilités, alors que la science interdit la totalisation mythique; en quoi elle inquiète, et l'idéologie rassure. Le passage constant d'un domaine à l'autre reste caractéristique de la pensée mythique; or les massacres sont souvent causés par l'irruption du mythe dans l'histoire.

Catégories

En-deçà des divers domaines sémantiques (politique, biologique, etc.) articulés par le discours raciste, des catégories binaires d'une grande généralité structurent l'univers sémantique. En première approximation, on parvient à cette homologation de catégories productives qui sous-tendent de multiples oppositions, NOUS VS AUTRES:

- nombre: *unique vs multiple*;
- identité: *homogène vs hétérogène*;
- pureté: *pur vs impur*;
- évaluation: *positif vs négatif*;
- statut véridictoire: *véridique vs mensonger*;
- position: *centré vs périphérique, supérieur vs inférieur, frontal vs dorsal*;
- aspects: *perfectif vs imperfectif, ponctuel vs itératif*;
- processus: *menacé vs menaçant*.

Ces catégories sémantiques restent « invisibles » dans la mesure où elles structurent les conditions générales de l'expérience représentée, le cadre spatio-temporel et les modes élémentaires de l'action. Par exemple, la multiplicité se traduit dans les sites racistes par la fréquence des expressions de grands nombres (*centaines, milliers, millions, milliards*). Elles peuvent tout aussi bien être véhiculées par des grammèmes: par exemple, la catégorie du nombre sous-tend une opposition comme *masse vs élite*; la catégorie aspectuelle, la perfectivité combinée à celle de la pureté, une opposition comme *autochtone vs allogène*. Leur combinatoire est productive: ainsi, l'opposition *pur vs impur*, sur l'axe sémantique de la luminosité, soutient une opposition comme *sombre vs brillant*; combinée à la couleur, elle se traduit par *noir vs blanc*; sur l'axe de l'action, elle se traduit par *souillure vs purification*. De même, les oppositions *nord*

vs *sud* et *oriental* vs *occidental* se corrént à l'opposition *central* vs *périphérique*, importante pour le propos identitaire du racisme.

Les champs sémantiques de la discrimination

L'homologation de ces catégories permet une grande variété de métaphores. Par exemple, au baby-foot, dans certaines banlieues, *tirer en feuj* signifie « tirer par derrière »⁹. En effet, ces catégories, par leur généralité même, peuvent s'articuler sur diverses isotopies: /par derrière/ sera manifesté indifféremment sur une isotopie guerrière (attaquer par derrière) ou sexuelle; la menace dissimulée se traduira par une aspiration itérative, instanciée aussi bien dans l'exploitation économique (capitalistes) que la prédation animale (buveurs de sang). Parmi les champs sémantiques les plus fréquemment manifestés, on relève:

– *invasion*: « Les musulmans ont envahi les prisons françaises » (un responsable du Front national); « *Stop Arab occupation* » (slogan d'une manifestation à New York);

– *animalité*: insectes (poux, fourmis, etc.); animaux sauvages (sangliers, etc.); espèces inférieures: singes (le *macaque au keffieh*, désignation de Yasser Arafat; adj. *simiesque*); canidés divers (voir l'écrêteau que portaient des restaurants pendant l'Occupation: « Interdit aux juifs et aux chiens »);

– *impureté*: l'impureté active (sida, peste) comme l'impureté passive (détchet) appellent une action générale de nettoyage: opérations « de nettoyage » en Algérie, « purification ethnique » en Serbie, *zatchiskas* en Tchétchénie.

Les combinaisons entre champs sémantiques et les passages constants d'un domaine à l'autre témoignent de ce que Claude Lévi-Strauss appelait jadis la « structure feuilletée du mythe ». Par exemple, une confusion de l'ordre de la syllepse permettra de juxtaposer le politique et le sexuel, la dénonciation des « clandestins qui *violent* nos règles et les valeurs d'hospitalité » et la prosopopée raciste: « Je *viole* les femmes, je pille les banques et les petits commerces ».

Les combinaisons entre champs sémantiques ne font que redoubler l'effet des homologations entre domaines. Par exemple, « violées par une horde de sangliers » (O. Fallaci, *supra*) combine: invasion (sexuelle) + multiplicité + animalité; ou encore « J'espère que tu seras violé par cinq ou six bougnoules et qu'ils te colleront le sida! » (Une lepéniste âgée s'adressant à un jeune mégrétiste, le 1^{er} mai 1999) combine: invasion + multiplicité + impureté (homosexualité).

Une première difficulté: la représentation réciproque

Même sous la simple forme de la citation, elle soulève des problèmes considérables, du moins pour la détection automatique. Les citations de textes racis-

9. La balle ricoche sur la planche en fond de « terrain ».

tes sont fréquentes sur les sites antiracistes, qui privilégient l'analyse commentée. Parfois, la citation tourne au jeu de miroirs: par exemple, le journal égyptien *al-Ahram* reproduisait un long extrait du *Protocole des Sages de Sion*, faux antisémite de la police tsariste; un site israélien reproduit à son tour la page d'*al-Ahram*, et cette reproduction en elle-même polémique peut évidemment induire en erreur un système de détection.

Dépasant la reprise, la parodie imite l'ennemi; par exemple, le site SOS Racaille, devenu SOS Rats après son interdiction, imite SOS Racisme¹⁰; le site islamophobe MRAY (Mouvement contre le racisme, l'antisémitisme et la violence) imite celui du MRAP, etc. Les renversements sont légion: il existe par exemple un site antiraciste intitulé Pro-bougnoûle.

La prosopopée, genre de prédilection des racistes, radicalise la même stratégie: elle donne en effet la parole à l'ennemi et à ses affidés; par exemple, tel témoignage commence par: « Bonjour, je suis une salope: je couche avec un Arabe. Oui je sais, c'est dégueulasse, ils sentent mauvais (et je ne vous parle même pas des Blacks, que j'ai aussi essayé), sont violents, méprisants, faînénants, terriblement machos, et leur Q. I. leur permet à peine de distinguer une BMW d'une Peugeot (pour savoir laquelle voler) » (site Racisme antiblanc).

Les stratégies de masquage

Elles participent de la diversité du discours raciste en multipliant les effets de surimpression d'autres genres et d'autres discours; ils vont par exemple:

- du politique au scientifique dans le discours négationniste, qui mime à s'y méprendre le positivisme ordinaire des sciences humaines;
- du scientifique au judiciaire quand, dans un article « savant », Robert Faurisson accuse Primo Levi d'être un « faux témoin » par syllepse sur l'acception historique et l'acception judiciaire;
- du politique au religieux: telle secte (Église du Créateur) plaide par exemple « Pour un monde plus blanc et plus brillant »;
- du politique au publicitaire: *Bravo Ariel!* (site France-Toujours) fait allusion à Ariel Sharon et au slogan d'une marque de lessive. Le programme de la « grande lessive » demeure, comme celui de la purification;
- du politique au pornographique: les sites sado-masochistes utilisent souvent les symboles nazis et partagent les mêmes polices typographiques que les sites néo-nazis, où l'on peut les télécharger.

On trouve aussi des *genres* qui se masquent, notamment des articles polémiques qui prennent la forme d'articles scientifiques, comme c'est le cas pour

10. Jusque dans son usage, devenu péjoratif, du mot *potes*: « Je vois de jour en jour cette invasion musulmane qui s'emplifie, les pots qui terrorisent les gens honnettes et foutent le bordel ds NOTRE BEAU PAYS doivent être éradiquer sur le champ » ([sic], cité dans M. Valette, 2004, *op. cit.*, p. 215-230).

le négationnisme. Tous ces masquages sont des indices de manipulation entendue linguistiquement comme travestissement de discours ou de genre.

Diversions

Alors que le mythe se donne sur le mode de l'évidence, le discours raciste procède ainsi par masquages divers. Voici quelques-unes des stratégies discursives qui rendent difficile sa description.

– *L'euphémisation*: elle est classique (ex.: *Endlösung*, solution finale). Elle permet d'éviter des poursuites et de créer de la connivence. Par exemple, là où l'on attendrait *arabe* ou *noir*, on a *ethnique*, voire simplement *18^e arrondissement*. L'euphémisme peut signaler aussi – à bon entendeur, salut – que l'on n'est pas libre de s'exprimer; mais, passé un certain seuil de banalisation, il suffit pour les racistes d'employer le langage du politiquement correct, comme si la vérité s'imposait d'elle-même (ainsi, l'expression *les jeunes des quartiers* peut parfaitement référer à des « délinquants allogènes »). Même les désignations allusives par un simple pronom valent pour un antécédent sous-entendu: *Ils se reproduisent*¹¹.

– *L'allusion*: puissant moyen de propagande, elle induit une participation (au sens rhétorique et parfois politique) présupposant des valeurs, ou du moins des savoirs partagés. Comme l'euphémisation, elle témoigne que nous sommes sous la coupe de la censure, mais pouvons cependant la contourner. Elle prend volontiers une forme satirique; ainsi, dans « le petit rat qui deale sa came et tourne des Gauloises en compagnie d'autres colons exotiques » (site Tribune-Libre), *tourne des Gauloises* est une reprise de l'expression bien de chez nous *roule des gauloises*, mais légèrement transformée pour signifier *viole des Françaises*.

– *Le cryptage*: forme sectaire de l'allusion, il réserve la compréhension aux *happyfew*, leur prodiguant des signes de reconnaissance. Beaucoup plus caractéristiques sont les signes de connivence (comme les abréviations): *niak*, *gnoul*, *caille* (pour *racaille*); ainsi le message de Maxime Brunerie avant son attentat contre Jacques Chirac s'achevait-il par *Death to ZOG, 88!* Cela signifie *Mort à Israël, Heil Hitler!* triplement codé par la langue étrangère, l'acronyme (de *Zionist Organization Government*) et la valeur numérique de la huitième lettre de l'alphabet.

Les stratégies employées dépendent évidemment du cadre juridique: les sites racistes français basés en France emploient massivement l'euphémisme par crainte de tomber sous le coup de la loi Gayssot. En revanche, les sites basés à l'étranger, la plupart aux États-Unis, ne semblent connaître aucune inhibition.

11. Le pronom *ils* est caractéristique des sites racistes – par contraste avec *nous* et *vous* sur les sites antiracistes.

Choix méthodologiques, nouveaux observables et enjeux théoriques

Cette étude reste centrée sur la sémantique de corpus et non sur l'implantation informatique. Outre le manque de place, notre relative discrétion peut se justifier par diverses raisons externes: méthodologie sous licence, curiosité flatteuse mais déplacée de certains services ordinairement plus discrets, enfin, refus de voir une application servir à des fins contraires (le filtrage de sites anti-racistes, par exemple).

La méthode suivie applique à des corpus numériques les principes de la sémantique différentielle de tradition saussurienne, étendue au palier textuel et intertextuel conformément aux propositions de la sémantique interprétative. Résumons les étapes suivies dans l'exploitation des corpus:

- le recrutement du corpus d'apprentissage s'accompagne d'une *catégorisation* (« manuelle », c'est-à-dire intellectuelle) qui place les documents dans deux classes (racistes et antiracistes);

- l'étude scientifique du corpus d'apprentissage isole des critères descriptifs qui contrastent ces deux classes à tous les niveaux d'analyse accessibles. Elle engage une *caractérisation*: on décrit les positions, pour justifier par des critères explicites la catégorisation des documents. Au besoin, elle peut aller jusqu'à l'*individuation*, de manière à distinguer les textes les uns des autres au sein de leur classe, dans le même camp. Elle utilise les méthodes de la linguistique de corpus, aménagées pour une étude contrastive;

- l'application se contente d'une *discrimination*: comme elle opère à la volée, dans le temps réel de la navigation, elle doit catégoriser les sites à partir d'un minimum d'indices locaux ou globaux, pour les placer dans deux classes: racistes, et non racistes quels qu'ils soient;

- le cycle de validation se laisse schématiser ainsi: recrutement du corpus de travail, par mots clés puis par élimination des documents non pertinents; sélection des variables sensibles; encapsulation de chaque variable retenue comme règle de discrimination dans un agent (petit automate élémentaire); test sur le corpus du travail; extension à des corpus nouveaux recrutés automatiquement; mise à jour périodique par accroissement de la population d'agents, pour tenir compte de l'évolution des sites racistes.

La méthodologie retenue ne peut cependant se confondre avec des choix d'implantation informatique qui dépendent d'éléments contingents comme l'état de l'art, voire simplement les collaborations au projet; elle nous retient ici par ses enjeux théoriques et l'empan des observables qu'elle met en évidence.

Préoccupée par les contenus politiques, utilisant des méthodes lexicologiques et lexicométriques, l'École française d'analyse du discours a naturellement privilégié le lexique, puis l'argumentation. Le discours était conçu comme

un *logos* qui aurait à répondre à l'inévitable question platonicienne: *Qui parle?*, les positions de parole étant par ailleurs assimilées à des positions de classe (sur les attendus théoriques de cette perspective énonciative, voir François Rastier, 2001¹²). Or les indices retenus dans le système Princip échappent pour beaucoup à la conception classique d'un discours argumentatif articulant des concepts exprimés par un vocabulaire politique. Beaucoup appartiennent en effet:

- à des unités linguistiques peu décrites en analyse du discours (grammèmes, morphèmes, ponctèmes)¹³;
- à des non-mots (chaîne de caractères *qu-* plus fréquente dans le discours raciste français, *wh-* plus fréquente dans le discours raciste anglais);
- enfin, à des niveaux ordinairement négligés par les études textuelles car censés appartenir à des documents et non aux textes: polices de caractères, graisse, capitales, couleurs. À cela s'ajoutent les spécificités des documents web, étiquettes, balises, bannières, etc. Ces traits sont caractérisants, car les sites racistes diffèrent grandement par leur structure des sites antiracistes: alors que les antiracistes se contentent de mettre en ligne leurs analyses, sur un site raciste il faut multiplier les clics, dépasser les mises en garde (*warnings*) de la page d'accueil, s'enfoncer dans le site et cliquer enfin, par exemple, sur *Mort aux juifs* pour pouvoir sortir. Aussi, alors que pour discriminer un corpus mêlé de textes racistes et antiracistes au format texte, une classification automatique produit 80 % de bonnes réponses, sur le format html elle en produira 94 % avec le même algorithme (voir Romain Vinot *et al.*, 2003¹⁴). Que l'on puisse caractériser des éléments globaux comme les positions raciste et antiraciste avec des éléments locaux comme des chaînes de caractères ou des balises html, cela suppose une *solidarité d'échelle* entre niveaux de complexité: c'est l'indice que le global détermine le local.

De plus, dans l'état de l'art, les éléments de l'expression restent évidemment beaucoup plus faciles à identifier et répertorier que les éléments sémantiques. On peut donc s'appuyer sur une autre solidarité, celle qui lie le contenu et l'expression et qui détermine la *sémiosis textuelle*. Elle se concrétise notamment, au palier du texte, par les genres. Quand on parvient à lier des variables locales de l'expression aux variables globales du contenu, on peut caractériser un type de texte à la volée et automatiser la détection.

Pourtant fort discriminantes, les variables locales de l'expression n'appartiennent pas au champ de réflexion ordinaire de la linguistique textuelle préoccupée par l'énonciation dont la prise en compte distinguerait le discours et le texte. Pour rendre compte de l'expression, il ne faut évidemment pas réduire

12. F. Rastier, 2001, *Arts et sciences du texte*, Paris, Presses universitaires de France, ch. 8.

13. Voir notamment le remarquable travail de M. Valette (*op. cit.*) sur les morphèmes.

14. R. Vinot, N. Grabar et M. Valette, 2003, « Application d'algorithmes de classification automatique pour la détection des contenus racistes sur Internet », *Actes TALN 2003*, Batz sur Mer.

le texte à une chaîne de caractères (comme en informatique) ni à une suite de mots ou de phrases (comme en linguistique), mais il convient de restituer toute sa complexité philologique.

Pour progresser sur cette question, il semble que la recherche puisse s'avancer dans trois directions complémentaires:

- le codage des corpus, et notamment des en-têtes de textes, reste une condition pour le développement de la linguistique de corpus. Cela demande un travail approfondi sur les discours, champs génériques et genres, ainsi que sur les « séquences » ou configurations textuelles;

- la linguistique de corpus permet à présent une observation des normes (notamment des normes sémantiques de la doxa) qui sont le « chaînon manquant » entre la linguistique de la langue et la linguistique de la parole¹⁵;

- l'extension de l'analyse sémantique à la corrélation contenu/expression permet de corréler à des variables sémantiques globales des variables d'expression de « bas niveau », y compris les ponctèmes, la casse ou la graisse typographiques, voire les balises html.

Mais comment interpréter les nouveaux observables? Ils deviennent tels par le détour de méthodes quantitatives qui attirent l'attention sur des phénomènes autrement imperceptibles; mais aussi par une hypothèse interprétative qui les retient et les qualifie dans la masse des « données ». Par exemple, pourquoi les sites racistes privilégient-ils le singulier, les antiracistes le pluriel (à *la civilisation* par exemple, répondent *les civilisations*, etc.)? Dans leur stratégie identitaire, les racistes créent une ontologie, constituent un monde d'essences typiques (l'Arabe, le Juif, la Civilisation – nécessairement unique). Pourquoi encore trouve-t-on sur les sites racistes des syntagmes et des phrases écrits en majuscules? C'est sans doute l'équivalent typographique du coup de gueule viril, comme en témoigne la fréquence des ponctuations fortes, qui plus est répétées. Comme toutes les interprétations, ce ne sont là que des conjectures; mais coalisées, elles emportent la décision du système, sinon la certitude du linguiste. Aucun indice n'est certes par lui-même nécessaire et suffisant; mais plus les indices sont hétérogènes, plus la conjecture que stabilise leur coalition devient vraisemblable. Le choix d'implantation retenu dans le projet Princip concrétise le raisonnement conjectural par un système multiagents: chaque indice est intégré dans une règle, et ces règles, par dizaines, se coalisent pour emporter la décision.

15. Pour un développement, voir F. Rastier, 2005, « Lexique et doxa en corpus », J. Pauchard et F. Canon-Roger éd., *Recherches en linguistique et psychologie cognitive*, n° 21, Presses universitaires de Reims, p. 55-104. Réédition : <http://www.revue-texto.net>.