

## CHAPITRE 3

# Structures sémantiques et sémiométrie

La stabilité mise en évidence au chapitre précédent suscite une série de questions, relatives aussi bien à la nature qu'à l'origine même des traits structuraux.

Une des questions, fondamentale, concerne le rôle et l'importance des relations de type purement sémantique dans la structure sémiométrique. Plusieurs travaux ont été réalisés pour tenter de répondre à cette question. Les relations sémantiques ne sont certes pas faciles à définir ni à mesurer, et dans la discipline connue sous le nom de Traitement du Langage Naturel<sup>1</sup>, les questions qui touchent la sémantique restent parmi les plus ardues.

En admettant que l'on puisse vaincre ces difficultés, il subsiste l'interrogation suivante : *les proximités sémantiques entre mots ne sont-elles pas responsables de l'essentiel de la structure observée ?* Autrement dit, de façon un peu sommaire, *la structure que l'on observe est-elle une structure linguistique plutôt que psychologique ou psycho-sociologique ?*

Deux mots ayant des sens voisins seraient notés de façon similaire et donc corrélés, et le *pattern* observé ne serait autre que le réseau de ces liens sémantiques.

La stabilité de la structure en découlerait, puisque la langue est relativement stable dans le temps. Elle est aussi pratiquement la même pour

---

1. Discipline d'où sont dérivés les produits usuels que sont les correcteurs d'orthographe, les logiciels de traduction, les moteurs de recherche documentaire.

les différentes classes d'âge, le sexe, etc.. Un tel réseau sémantique doit aussi résister - plus ou moins - à une traduction du questionnaire, d'où la relative stabilité observée d'un pays à un autre<sup>1</sup>.

Pour apporter des éléments de réponse à ces interrogations, le présent chapitre relate trois séries de travaux, formant autant de parties.

La première partie, intitulée « Les champs sémantiques internes » montre comment les réponses au questionnaire sémiométrique et les corrélations entre mots calculées à partir de ces réponses décrivent effectivement, et assez finement, des liens de nature sémantique. Ce résultat n'était pas évident *a priori*, car, ne l'oublions pas, les questions posées ne portent que sur le caractère *agréable* ou *désagréable* des mots de la liste.

La deuxième partie intitulée « Voisinage sémantique du questionnaire sémiométrique », est dévolue à l'expérience suivante : les 210 mots de la sémiométrie vont être décrits par leurs *synonymes*, ou au moins par leurs *voisins sémantiques*, ceux-ci étant fournis par un dictionnaire de synonymes. Autrement dit, plus de questionnaire à faire remplir par un échantillon représentatif... c'est la langue elle-même qui répond, à travers certains des outils qui ont été forgés pour la décrire et la comprendre. Bien sûr, les dictionnaires de synonymes ne sont pas des instruments parfaits, et ils diffèrent notablement entre eux. L'expérience sera d'ailleurs l'occasion de voir l'étendue des difficultés qui nous attendent.

La troisième partie intitulée « Réseau sémantique des verbes français », relate une expérience beaucoup plus générale qui a consisté à prendre tous les verbes de la langue française (en fait les 829 verbes les plus usuels figurant dans le manuel classique de grammaire Bescherelle, parmi lesquels se trouvent déjà de nombreux verbes très peu usités, voire quasi-inconnus) et à étudier la structure sémantique de ces verbes à partir de leurs voisins.

Cette expérience confirmera qu'il existe des proximités locales sémantiques, mais qu'il n'existe pas d'axes principaux dominants et stables.

---

1. Disons d'emblée, en anticipant sur les résultats présentés dans ce chapitre, que la structure sémiométrique ne se réduit que très partiellement à une structure sémantique. Au niveau local, autour d'un point représentant un mot, on pourra trouver des voisins sémantiques, mais les grandes oppositions responsables d'axes stables ne sont pas observées spontanément à partir d'analyses de simples proximités sémantiques.

### 3.1 Les « champs sémantiques internes »

On appellera champ sémantique interne pour un mot donné l'ensemble des mots (pris dans une liste de mots fixée *a priori*, d'où le qualificatif d'*interne*) qui lui sont corrélés. La distance utilisée est d'autant plus petite que le coefficient de corrélation entre les deux mots est élevé<sup>1</sup>. Cette dénomination est justifiée par l'interprétation *a posteriori* des proximités observées. Il n'était pas évident que des notes fondées seulement sur l'agrément ou le désagrément engendrent des proximités sémantiques. Cette section montre la cohérence et la finesse des proximités observées.

Certains mots possèdent un champ sémantique interne riche et dense défini par des corrélations élevées avec de nombreux autres mots de la liste. C'est le cas pour les mots *Efficace* et *Courage*. Des mots comme *Montagne*, *Voluptueux*, *Mystère*, proches d'un petit nombre de mots, ont chacun un champ plus restreint.

Un même mot peut bien évidemment, dans certaines limites, être corrélé avec des mots non corrélés entre eux. C'est ainsi que les champs sémantiques de *Courage* et *Efficace* ne sont pas les mêmes malgré leur forte corrélation et l'appartenance de certains mots comme *Robuste* et *Dynamique* aux deux champs.

#### Champ sémantique du mot *Efficace*

*Précision* (.51)<sup>2</sup>, *Concret* (.40), *Robuste* (.40), *Courage* (.39), *Astucieux* (.38), *Maîtriser* (.38), *Dynamique* (.37), *Raison* (.37), *Confiance* (.35), *Produire* (.35), *Utilitaire* (.35), *Certitude* (.34), *Honnête* (.34), *Volontaire* (.34), *Construire* (.33), *Logique* (.32), *Récompense* (.32), *Réfléchir* (.32), *Respect* (.32), *Bâtisseur* (.31), *Fermeté* (.31), *Honneur* (.31), *Douceur* (.30).

#### Champ sémantique du mot *Courage*

*Honnête* (.41), *Dynamique* (.40), *Respect* (.40), *Efficace* (.39), *Confiance* (.38), *Honneur* (.38), *Elégance* (.36), *Politesse* (.36), *Amitié* (.35), *Guérir* (.35), *Robuste* (.35), *Charitable* (.34), *Précision* (.33), *Volontaire* (.33), *Douceur* (.32), *Protéger* (.32), *Pureté* (.32), *Raison* (.32), *Gaieté* (.31), *Récompense* (.31), *Certitude* (.30).

---

1. Le carré de la distance entre deux mots *a* et *b* dont le coefficient de corrélation vaut  $r(a,b)$  s'écrit  $d^2(a,b) = (1 - r(a,b))$ .

2. Les quantités entre parenthèses sont les coefficients de corrélation avec le mot en caractères gras qui précède. Notons que tous les coefficients de corrélation cités sont hautement significatifs statistiquement compte tenu de la taille de l'échantillon (11 055).

### Champ sémantique du mot *Voluptueux*

*Sensuel (.47), Charnel (.45), Désir (.34), Moelleux (.34), Nudité (.32), Séduire (.31), Souplesse (.31), Sublime (.30), Féminin (.25), Rêver (.24), Astucieux (.23), Audace (.23), Douceur (.23), Intime (.23), Peau (.23), Caresse (.22), Légèreté (.22), Original (.22), Art (.20), Élégance (.20), Viril (.20).*

### Champ sémantique du mot *Montagne*

*Sommet (.47), Escalader (.34), Campagne (.33), Fleuve (.32), Arbre (.31), Océan (.31), Ile (.26), Lune (.26), Fleur (.24), Musique (.23), Eau (.22), Animal (.21), Bleu (.21), Construire (.21), Nid (.21), Douceur (.20).*

### Champ sémantique du mot *Mystère*

*Orage (.26), Magie (.25), Aventurier (.24), Emotion (.24), Original (.24), Secret (.24), Sauvage (.23), Infini (.22), Feu (.21), Océan (.20).*

### Champ sémantique du mot *Dieu*

*Foi (.74), Prêtre (.65), Recueillement (.60), Sacré (.55), Ame (.48), Créateur (.33), Patrie (.32), Pardon (.29), Charitable (.27), Morale (.27), Héros (.26), Honneur (.26), Noble (.26), Cérémonie (.25), Eternel (.25), Tradition (.25), Mariage (.23), Souverain (.22), Famille (.21), Pureté (.21), Discipline (.20), Elite (.20).*

Ce premier cadrage d'un espace des mots, par la constitution de quelques champs sémantiques internes, peut se généraliser à l'ensemble des mots du questionnaire. On utilise pour cela les méthodes de classification qui consistent à effectuer une partition de l'ensemble des éléments à classer c'est-à-dire les mots. Cependant, dans ce cas, un mot ne peut appartenir à plusieurs champs.

### **3.1.1 Représentation hiérarchique des proximités sémantique**

Les mots qui se ressemblent le plus quant à leur manière d'être notés et donc ressentis vont être regroupés maintenant de façon automatique. La méthode d'analyse en composantes principales présentée au premier chapitre réalise une telle représentation globale des corrélations, et donc des proximités. Mais les deux méthodes de classification appliquées ici, *classification hiérarchique*<sup>1</sup> et *cartes auto-organisées* de Kohonen<sup>2</sup>, décrivent peut-être avec plus de finesse les proximités locales qui sont, nous le verrons, les seules à être interprétables en termes sémantiques.

Pour ce faire, on réalise dans un premier temps une *classification hiérarchique* sur les 210 mots munis des distances définies précédemment.

---

1. Le principe des méthodes de classification hiérarchique est brièvement présenté en annexe A1.7.

2. Le principe des cartes auto-organisées est brièvement présenté en annexe A1.8.

Le principe de ce traitement (l'algorithme, en fait) est de regrouper les mots deux par deux par agglomération progressive fournissant ainsi une hiérarchie de *partitions* des mots. Nous retenons ici les partitions emboîtées en 12, 24 et 36 classes représentées par le tableau 3.1 qui schématise le *dendrogramme* de la hiérarchie.

Les résultats parlent d'eux-mêmes. Les classes obtenues regroupent des mots qui ne sont pas de vrais synonymes<sup>1</sup>, mais appartiennent au même *halo sémantique*.

Examinons, par exemple, la partition en 12 classes (la moins fine des trois partitions) dont les classes sont repérées dans la première colonne du tableau 3.1 par les symboles 1/12, 2/12, ..., 12/12.

Sont regroupés au niveau de cette partition, les mots ayant trait au concept de « Sublimation » (classe 1/12, *Absolu, Immense, Infini, Admirer, Adorer, ..., Secret, Sublime*). D'autres classes correspondent à des thèmes comme l'art de l'écriture (classe 3/12 : *Art, Ecrire, livre, Poésie, Théâtre*), la nature (classe 4/12 ; *Escalader, Île, .., Sommet, Animal, ..., Vert*), l'émotion (classe 5/12 ; *Emotion, ..., Rêver, Humour, ..., Rouge, Bohème, .., Révolte*).

Notons, en passant, que l'on retrouve dans la classe 28/36 le champ sémantique du mot *Efficace* ; le mot *Courage* n'en fait pas partie et se regroupe avec les mots *Respect, Politesse, Honnête*.

### 3.1.2 Représentation plane des proximités sémantiques

La représentation des classes emboîtées du tableau 3.1 ne permet pas d'apprécier les distances relatives entre les groupements de mots obtenus par la classification hiérarchique. La carte de Kohonen (figure 3.1) est un mode de visualisation de données qui regroupe les mots en classes en conservant autant que faire se peut la topologie initiale de l'espace des mots. Elle rend compte ainsi des proximités entre classes qui se ressemblent (*cf.* annexe A1.8).

Ainsi, les mots *Théâtre* et *Poésie* sont proches à la fois des regroupements « *Nid, Livre, Fleuve, Ecrire, Art* », « *Vert, Sublime, Nager, Musique, Montagne, Eau, Bleu* », « *Sommet, Secret Lune, Immense* », « *Justice* », « *Recueillement, Créateur, Ame* » (*cf.* figure 3.1).

---

1. La liste de mots excluait *a priori* la présence de synonymes qui auraient introduit des redondances inutiles.

**Tableau 3.1 : Classification hiérarchique en 12, 24 et 36 classes emboîtées (cf. annexe 1.7)**

1/12	1/24	1/36	absolu, immense, infini
		2/36	admirer, adorer, éternel, précieux, secret, sublime
2/12	2/24	3/36	douceur, féminin, intime, moelleux, peau, souplesse
		4/36	caresse, désir, séduire
	3/24	5/36	charnel, nudité, sensuel, voluptueux
3/12	4/24	6/36	art, écrire, livre, poésie, théâtre
4/12	5/24	7/36	escalader, île, lune, montagne, océan, sommet
	6/24	8/36	animal, arbre, bleu, campagne, eau, fleur, fleuve, nager, nid, vert
5/12	7/24	9/36	émotion, évasion, légèreté, magie, original, rêver
		10/36	humour, jeu, musique, rouge
		11/36	bohème, changement, différent, étranger, inconnu
	8/24	12/36	acharnement, aventurier, défi, ruse, vitesse
	9/24	13/36	désert, feu, mystère, noir, orage, sauvage
		14/36	critiquer, désordre, ironie, révolte
6/12	10/24	15/36	guerre, trahir
		16/36	angoisse, cri, danger, détachement, doute, faute, labyrinthe, méfiance, mort, rompre, vide, vieillir
	11/24	17/36	armure, attaquer, chasse, fusil
	12/24	18/36	immobile, masque, muraille, nœud
		19/36	frontière, interdire, métallique, punir, rigide, sacrifice

**Tableau 3.1 (suite) : Classification hiérarchique en 12, 24 et 36 classes emboîtées**

7/12	13/24	20/36	bijou, élégance, minceur, mode, parfum
		21/36	acheter, cadeau, confort, gratuit, maison, récompense
	14/24	22/36	ambition, conquérir, gloire, puissance, victoire
		23/36	argent, hériter, or, propriété, richesse
8/12	15/24	24/36	noble, perfection, raffiné
		25/36	commander, élite, héros, honneur, patrie, soldat, souverain, tradition, viril
	16/24	26/36	commerce, industrie, matériel, produire, utilitaire
	17/24	27/36	discipline, économiser, fermeté, loi, morale, obéir, règle
9/12	18/24	28/36	astucieux, audace, certitude, concret, dynamique, efficace, logique, maîtriser, précision, raison, robuste, volontaire
	19/24	29/36	effort, interroger, question, réfléchir
	20/24	30/36	bâtitteur, chercheur, construire, école, enseigner, inventeur, science, travail
10/12	21/24:	31/36	âme, créateur, dieu, foi, prêtre, recueillement, sacré
11/12	22/24	32/36	enfance, ensemble, féconder, maternel
		33/36	cérémonie, famille, mariage, naissance
12/12	23/24	34/36	courage, fidélité, honnête, politesse, respect
		35/36	amitié, confiance, gaieté, guérir, paix, rire, tendresse
	24/24	36/36	attachement, charitable, consoler, humble, justice, modération, modestie, pardon, patience, protéger, prudence, pureté, soigner

RICHESSE OR HERITER GLOIRE ARGENT	RECOMPENSE PRECIEUX MODE MINCEUR GRATUIT ELEGANCE CONFORT CADEAU BIJOU ADORER ACHETER	MAISON FECONDER DOUCEUR	TENDRESSE PROTEGER PAIX NAISSANCE MATERNEL GUERIR GAITE FLEUR FIDELITE FAMILLE ENFANCE CONFIANCE CAMPAGNE ATTACHEMENT AMITIE	SOIGNER RESPECT PUISSANCE POLITESSE PATIENCE HUMBLE HONNETE CONSOLER	TRAVAIL REGLE MORALE MODERATION LOI EFFORT ECOLE
VIRIL VICTOIRE PRUDENCE PROPRIETE CONQUERIR COMMERCE AMBITION	CERTITUDE	RAFFINE MARIAGE CEREMONIE ADMIRER	PURETE PARDON CHARITABLE	MODESTIE ENSEIGNER COURAGE	UTILITAIRE RAISON PRODUIRE PRECISION LOGIQUE FERMETE
SOLDAT MATERIEL INDUSTRIE ECONOMISER DISCIPLINE COMMANDER	TRADITION SOUVERAIN PERFECTION PATRIE HONNEUR HEROS ELITE	SACRE PRETRE NOBLE FOI ETERNEL DIEU	RECUEILLEMEN CREATEUR AME	JUSTICE	VOLONTAIRE SCIENCE ROBUSTE REFLECHIR MAITRISE INVENTEUR EFFICACE CONSTRUIRE CONCRET CHERCHEUR BATISSEUR
SACRIFICE RIGIDE OBEIR METALLIQUE INTERDIRE FRONTIERE CHASSE	INTERROGER	ABSOLU	THEATRE POESIE	NID LIVRE FLEUVE ECRIRE ART	DYNAMIQUE AUDACE ASTUCIEUX ARBRE
ROMPRE PUNIR NOEUD MURAILLE MEFIANCE IMMOBILE FUSIL CRI ARMURE	QUESTION MASQUE DETACHEMENT ACHARNement	ROUGE INFINI INCONNU ETRANGER ESCALADER	SOMMET SECRET LUNE IMMENSE	VERT SUBLIME NAGER MUSIQUE MONTAGNE EAU BLEU	SOUPLESSE PEAU JEU INTIME HUMOUR FEMININ ENSEMBLE ANIMAL
VIEILLIR VIDE TRAHIR MORT LABYRINTHE GUERRE FAUTE DOUTE DANGER ATTAQUER ANGOISSE	RUSE REVOLTE ORAGE NOIR IRONIE DESORDRE DESERT CRITIQUER	VITESSE SAUVAGE MYSTERE FEU DIFFERENT DEFI CHANGEMENT	ORIGINAL NUDITE MAGIE LEGERETE EVASION EMOTION BOHEME AVENTURIER	VOLUPTÉ SENSUEL OCEAN ILE CHARNEL	SEDUIRE RIRE REVER PARFUM MOELLEUX DESIR CARESSE

Figure 3.1 : Proximités entre mots décrites par une carte de Kohonen

La classe « *Recueillement, Créateur, Âme* » est elle-même proche de « *Modestie, Enseigner, Courage* » et « *Pureté, Pardon, Charitable* », « *Raffiné, Mariage, Cérémonie, Admirer* », « *Sacré, Prêtre, Foi, ...* ».

Les classes sont sans doute moins homogènes et parfois moins cohérentes que les partitions hiérarchiques mais la représentation par la carte auto-organisée de Kohonen insiste davantage sur la cohérence entre les classes.

Les cartes de Kohonen ont cependant l'inconvénient d'être plus instables que les classifications ascendantes hiérarchiques et s'adressent davantage à des ensembles de données non structurées. De fait, nous verrons au chapitre suivant qu'elles sont mieux adaptées aux questionnaires ouverts qu'à l'analyse de corpus de mots déjà fortement structuré.

### 3.2 Voisinage sémantique du questionnaire sémiométrique

Dans cette première expérience, chaque mot de la sémiométrie est décrit par ses « synonymes », collationnés à partir d'une source externe<sup>1</sup>. Le tableau 3.2 nous donne un court extrait du recueil ainsi obtenu pour les premiers mots de *Absolu* à *Angoisse*. Comme on le voit, le nombre de voisins sémantiques est très variable.

Ce nombre de voisins peut être très grand et peut atteindre 22 lignes de texte pour le mot *Changement* qui ne figure pas dans l'extrait du tableau 3.2, ou très court (voire inexistant, ainsi, l'adjectif *Maternel* – également extérieur au tableau 3.2 – n'a pas de voisin sémantique dans ce dictionnaire).

A partir de ce nouveau recueil de textes est construit un *tableau croisé* contenant en lignes les 210 mots de base, et en colonnes tous les mots rencontrés dans le recueil, c'est-à-dire tous les voisins sémantiques des 210 mots.

Le tableau, à l'intersection de la ligne *i* et de la colonne *j* contient la valeur « 1 » si le mot *j* figure parmi les voisins sémantiques du mot *i*, et contient la valeur « 0 » sinon.

Si certains mots ont des voisins sémantiques communs, un calcul de distance approprié les désignera comme proches.

---

1. Synonymes et voisins sémantiques ont été fournis par le *Dictionnaire de synonymes et contraires* de Henri Bertaud du Chazaud, publié par Robert, 1994.

**Tableau 3.2 :**  
**Exemples de voisins sémantiques des neuf premiers mots**

– **Absolu**

catégorique, complet, dirimant, discrétionnaire, dogmatique, entier, exclusif, foncier, formel, impératif, impérieux, inconditionnel, indispensable, infini, parfait, plein, radical, rédhitoire, total, unilatéral, autocratique, autoritaire, arbitraire, cassant, césarien, despotique, dictatorial, dogmatique, exclusif, impérieux, intransigeant, jupitérien, omnipotent, souverain, totalitaire, tyrannique, idéal, infini, intégrité, intransigeance, perfection, plénitude

– **Acharnement**

ardeur, effort, énergie, persévérance, ténacité, cruauté, entêtement, fureur, furie, obstination, opiniâtreté, rage, sadisme

– **Acheter**

acquérir, corrompre, soudoyer

– **Admirer**

apprécier, s'émerveiller, s'enthousiasmer, s'extasier, louer, louer

– **Adorer**

aimer, bénir, glorifier, idolâtrer, admirer, honorer, respecter, révéler, vénérer, aduler, courtoiser, encenser, flagorner, flatter

– **Ambition**

appétit, ardeur, aspiration, brigue, convoitise, désir, faim, fringale, idéal, passion, prétention, quête, recherche, soif, arrivisme, mégalomanie, but, dessein, fin, mobile, objet, rêve, visée, vue

– **Ame**

cœur, conscience, dedans, esprit, fond, intérieur, mystère, pensée, principe, secret, spiritualité, transcendance, vie, air, ectoplasme, émanation, essence, éther, étincelle, feu, flamme, souffle, vapeur, émotion, expression, intelligence, sensibilité, sentiment, agent, animateur, centre, cerveau, chef, instigateur, maître, moteur, nœud, organisateur, patron, responsable

– **Amitié**

accord, cordialité, entente, sympathie, amabilité, compliment, hommages, caresse, flagornerie, flatterie, grimace, affection, affinité, amour, attachement, bonté, complaisance, dévotion, dévouement, dilection, douceur, inclination, inclination, intérêt, lien, penchant, piété, respect, sollicitude, tendresse, union, vénération, goût, prédilection, vocation, bienveillance, affabilité, altruisme, compréhension, convivialité, cordialité, gentillesse, indulgence, mansuétude, obligeance, prévenance, bonté, charité, clémence, compassion, générosité, indulgence, mansuétude, miséricorde, indulgence, pitié, serviabilité, socialité,

– **Angoisse**

inquiétude, anxiété, appréhension, crainte, émotion, ennui, malaise, peine, préoccupation, scrupule, souci, stress, supplice, tension, trac, transe, trouble, alarme, alerte, émoi, affolement, agitation, désarroi, détresse, effarement, effroi, épouvante, panique, peur, terreur

Selon une méthode d'application courante pour les analyses statistiques de texte, on peut donc construire un tableau des distances entre les mots, et représenter graphiquement ces distances par des proximités entre points.

Le premier résultat obtenu, décevant, mais prévisible, est une séparation entre les noms et les verbes... qui ne peuvent pas être synonymes, et assez rarement voisins sémantiques dans les dictionnaires usuels, alors que noms et verbes figurent parmi les mots de la sémiométrie.

On procédera donc à l'analyse des seuls noms et adjectifs, qui, au nombre de 177, sont donc très majoritaires dans la liste originale des 210 mots du questionnaire.

On n'observe plus des nuages de points réguliers et équilibrés comme ceux des graphiques sémiométriques présentés dans les sections précédentes, mais des grappes de mots qui s'opposent à tous les autres.

Ce phénomène semble dû à la *non-transitivité* des similitudes sémantiques, et à l'absence de pertinence de la notion d'éloignement sémantique. La non-transitivité peut s'exprimer de la façon suivante : si le mot A a un sens voisin de celui du mot B, et si le mot B a un sens voisin de celui du mot C, alors, A n'a pas forcément un sens voisin de celui du mot C. On peut même assez rapidement aboutir à un antonyme de A<sup>1</sup>.

L'observation des liens sémantiques issus d'un dictionnaire avec ceux que l'on peut dériver des notes sémiométriques aide à comprendre la nature du fait statistique et de la structure que l'on observe.

Tout d'abord, il existe, parmi les 210 mots de la sémiométrie, des mots qui n'ont pas de voisins sémantiques communs avec les autres (en fait avec plus de deux autres) comme par exemple les mots : *Lune, Peau, Arbre, Fusil, Ile, Théâtre, Soldat*. On comprend bien que pour *Lune*, par exemple, ses synonymes (*Artémis, Diane, Hécate, Phoebé, Séléné*<sup>2</sup>), trop particuliers, ont peu de chance de se retrouver parmi les voisins, même éloignés, des autres mots. De même pour *Fusil* (*carabine, lebel, mousqueton, arquebuse, mousquet, tromblon*).

---

1. Ainsi, *abandonner* est un voisin sémantique de *donner*, voisin de *partager*, qui est lui-même un voisin de *participer*, à son tour voisin de *se joindre*. On rencontre en général des chaînes plus longues, qui conduisent plus simplement à une indifférence sémantique. Comme la séquence *littérature – poésie – musique – vibration*. Le premier et le dernier terme de la chaîne n'ont pas – ou peu – de rapports, sans pour autant avoir des sens contraires.

2. Bien que n'appartenant pas au questionnaire sémiométrique, ces mots, étant des noms propres, figurent avec une majuscule.

Les corrélations observées entre ces mots et les autres ne sont donc pas des artefacts sémantiques, elles relèvent d'un contexte perceptif plus général.

Il existe, en revanche, des mots ou groupes de mots qui ont entre eux des liens sémantiques incontestables, lesquels vont se traduire simultanément par des proximités sur les cartes sémiométriques et sur les graphes sémantiques<sup>1</sup>.

Ainsi, les mots *Charnel*, *Sensuel*, *Voluptueux* qui figurent dans le questionnaire sémiométrique sont des voisins sémantiques les uns des autres, et ont plusieurs voisins sémantiques en commun dans le dictionnaire de synonymes que nous avons utilisé (comme par exemple : *animal*, *bestial*, *érotique*, *lascif*, *libidineux*, *luxurieux*). Il s'agit cependant d'une exception dans le questionnaire. Il n'existe pas d'autres groupes de trois mots aussi proches sémantiquement.

On pourrait alors objecter que cette redondance partielle (partielle car ces mots ont malgré tout des sens clairement distincts) donne un poids excessif aux notions sous-jacentes. Ce groupe de trois mots serait alors responsable de certains des traits de structure observés et décrits au chapitre précédent. Mais ce n'est pas le cas : comme on le vérifie aisément, les axes de base dont nous avons décrit la stabilité ne sont pas modifiés si l'on supprime, dans le fichier des notes, deux sur trois des mots incriminés.

Fait encore plus marquant, les six premiers axes sont conservés si l'on supprime de la liste l'ensemble des trois mots *Charnel*, *Sensuel* et *Voluptueux*.

Le second axe, le plus corrélé aux trois mots supprimés, et donc le plus susceptible d'être affecté par cette suppression, est étonnamment stable, comme l'atteste le tableau 3.3, qui nous montre les mots les plus extrêmes sur l'axe. Dans les deux cas, ces listes s'opposent à *Patrie*, *Argent*, *Propriété*, *Soldat*, *Rigide*, *Economiser*, *Obéir*, *Discipline*.

Ainsi, les proximités sémantiques, particulièrement importantes dans le cas de ces trois mots, semblent jouer un rôle secondaire dans l'établissement de la structure stable sous-jacente.

Il existe aussi des couples de mots, comme *Ironie* et *Humour*, qui ont un grand nombre de voisins sémantiques communs (*esprit*, *causticité*, *malice*, *verve*, *spirituel*, *comique*, *drôle*, *facétieux*, *malicieux*, *piquant*, *satirique*, *moquerie*, *saillie*, *satire*) et qui sont donc très proches sémantiquement, que

---

1. L'outil retenu pour donner une appréciation visuelle agréable des graphes sémantiques est de nouveau la carte auto-organisée de Kohonen. Cf. annexe A1.8 pour plus de détails sur cette méthode neuronale et la justification de son emploi dans ce contexte.

la sémiométrie sépare de façon significative. Ainsi, sur le troisième axe « Attachement / Détachement », *Ironie* se situe du côté du « Détachement » et « Humour », de celui de l' « Attachement » .

**Tableau 3.3 :**

<b>Liste des mots les plus corrélés positivement avec le deuxième axe</b>	
Avec 210 mots	Avec 207 mots
<i>Art</i>	<i>Art</i>
<i>Bohème</i>	<i>Bohème</i>
<i>Livre</i>	<i>Livre</i>
<i>Original</i>	<i>Original</i>
<u><i>Sensuel</i></u>	<i>Poésie</i>
<i>Poésie</i>	<i>Orage</i>
<i>Révolte</i>	<i>Théâtre</i>
<u><i>Voluptueux</i></u>	<i>Musique</i>
<i>Orage</i>	<i>Révolte</i>
<i>Sauvage</i>	<i>Océan</i>
<i>Océan</i>	<i>Sauvage</i>

Autre exemple de proximité sémantique non traduite par des proximités sémiométriques : le mot *Animal*, dont on a vu qu'il figurait parmi les voisins sémantiques des mots *Charnel*<sup>1</sup>, *Sensuel*, *Voluptueux* ne leur est pas voisin dans les plans sémiométriques. Il s'en éloigne notamment le long du second axe (« Devoir /Plaisir »). Il faut cependant préciser qu'il y a ici ambiguïté entre le substantif *animal* et l'adjectif *animal* qui peuvent incontestablement avoir des voisins sémantiques différents. Le questionnaire sémiométrique spécifie bien « Un animal ».

En conclusion de ce paragraphe, on retiendra que :

- la structure sémantique de la liste sémiométrique telle qu'elle est décrite (grossièrement, il est vrai) par un dictionnaire de synonymes ne donne pas lieu à des axes (ou directions principales) stables.
- localement, on retrouve des associations sémantiques sur les cartes sémiométriques, mais il existe aussi des exceptions notables.
- les grandes oppositions stables observées dans le champ sémiométrique ont un caractère psycho-sociologique, voire socio-démographique marqué (liens de certains axes avec le sexe, l'âge, la catégorie socio-

<sup>1</sup> Le mot *Charnel* n'apparaît pas dans le tableau 3.3 car il n'occupe que le treizième rang sur le deuxième axe.

professionnelle), et ne relèvent pas ou peu du registre synonymie-antonymie.

### 3.3 Un réseau sémantique des verbes français

Cette seconde expérience, contrairement à la précédente, est indépendante du corpus de 210 mots. C'est l'ensemble des verbes français usuels (les 829 verbes les plus fréquents figurant dans le manuel classique de grammaire « *Bescherelle* »<sup>1</sup>) qui est maintenant pris en considération.

La question posée est aussi beaucoup plus générale et ambitieuse : se dégage-t-il une structure analogue à celle observée lors des analyses sémiométriques lorsque l'on décrit l'ensemble des distances sémantiques entre les verbes de la langue française ?

Chaque verbe sera encore décrit par l'ensemble de ses synonymes, ce qui va permettre de calculer des distances entre verbes, et donc d'observer la structure induite par ces distances.

Pourquoi avoir choisi les verbes ? Parce que ceux-ci constituent un ensemble beaucoup plus restreint et homogène que celui des noms ou des adjectifs, et qu'il est donc possible d'être (quasi) exhaustif.

Les résultats confirmeront ceux observés au paragraphe précédent : la structure sémantique est essentiellement locale, et les « grandes distances sémantiques » ne donnent pas lieu à des axes stables, ni à des oppositions interprétables.

Le corpus formé par les verbes et leurs synonymes (analogue au texte du tableau 3.2 ci-dessus, mais relatif cette fois à 829 verbes) comporte 17 446 occurrences de mots (qui sont évidemment tous des verbes).

Ce même texte comporte 3 839 verbes distincts. Ce nombre est beaucoup plus grand que les 829 verbes de départ, car des verbes moins usités peuvent figurer parmi les synonymes des verbes du *Bescherelle*. Le tableau 3.4 donne, à titre d'exemple, les synonymes retenus du verbe *couper*.

Deux verbes seront d'autant plus proches qu'ils auront beaucoup de synonymes communs.

Evidemment, beaucoup de paires de verbes n'auront aucun synonyme commun, et les distances entre ces paires seront équivalentes.

---

1. *Bescherelle, La conjugaison, Dictionnaire de 12 000 verbes*, Hatier, Paris, 1990.

### Tableau 3.4 Synonymes du verbe *couper*

*amputer, champlever, cisailier, coupailier, débillarder, découper, diviser, ébarber, élaguer, entrecouper, hacher, inciser, massicoter, réséquer, scarifier, sectionner, taillader, tailler, trancher, tronçonner, essoriller, émonder, séparer, supprimer, soustraire, censurer, exclure, châtrer, castrer, chaponner, émasculer, hongrer, traverser, franchir, mélanger, abattre, scier, faucher.*

Comme cela est habituel lors des analyses statistiques de textes, il est nécessaire de faire intervenir un seuil de fréquence minimale des mots.

Ne seront retenus ici dans un premier temps que les verbes apparaissant au moins quatre fois dans le corpus formé de l'ensemble des verbes et de leurs synonymes. Ceci laisse 1 619 mots distincts, et réduit, dans une proportion moindre, la taille du corpus global qui comporte maintenant 13 660 occurrences.

Cette fréquence minimale de quatre laisse quand même quelques verbes relativement peu courants dans le corpus, tels que : *contrister, lanciner, transmuier, gourmander, congutiner, épisser, déconfire, carguer, excaver, ravauder, désencombrer.*

Le tableau 3.5 donne les verbes les plus fréquents dans le corpus, en se limitant maintenant, pour des questions d'encombrement, aux verbes apparaissant plus de 16 fois<sup>1</sup>.

Afin d'obtenir rapidement une vue d'ensemble du graphe sémantique des verbes de la langue française, on est finalement conduit à procéder à une analyse des correspondances (*cf.* annexe A1.4) d'un tableau binaire ayant 1 619 lignes et 829 colonnes<sup>2</sup>.

---

1. Le tableau dans son intégralité est disponible auprès des auteurs.

2. Le terme situé à l'intersection de la ligne *i* et de la colonne *j* de ce tableau vaut 1 si le verbe *i* est un des synonymes du verbe *j*, et vaut 0 sinon.

**Tableau 3.5 : Extrait de la liste des verbes (classés ici par fréquences décroissantes (de 50 à 17) du nombre de synonymes)**

<i>verbes</i>	<i>frequence</i>	<i>verbes</i>	<i>frequence</i>	<i>verbes</i>	<i>frequence</i>
enlever	50	régler	22	organiser	18
fixer	39	comprendre	22	incorporer	18
prendre	36	conduire	22	aider	18
examiner	33	encourager	22	entamer	18
joindre	33	assurer	22	accabler	18
arrêter	33	lier	22	apprendre	18
entraîner	32	écarter	22	étendre	18
assembler	32	céder	22	tirer	18
pousser	31	supprimer	22	condamner	18
soutenir	31	diriger	22	juger	18
abandonner	30	trancher	21	signaler	18
augmenter	30	garantir	21	remplir	18
atteindre	29	déterminer	21	transporter	17
montrer	28	attirer	21	communiquer	17
protéger	28	approuver	21	rassembler	17
maintenir	28	marquer	21	emporter	17
admettre	28	élever	21	percer	17
disposer	28	déclarer	20	attraper	17
couper	28	attaquer	20	considérer	17
couvrir	28	retourner	20	garder	17
exposer	27	former	20	ravir	17
arranger	27	développer	20	éloigner	17
toucher	26	renverser	20	renvoyer	17
relever	26	unir	20	discerner	17
gagner	26	imaginer	20	assujettir	17
appuyer	26	apprécier	20	dire	17
découvrir	26	conclure	20	émouvoir	17
agiter	25	estimer	20	diminuer	17
détruire	25	tenir	20	extraire	17
dégager	25	combiner	20	introduire	17
attacher	24	accorder	20	bouleverser	17
abattre	24	instruire	20	corriger	17
indiquer	24	frapper	19	contraindre	17
saisir	24	représenter	19	exalter	17
observer	24	concevoir	19	évaluer	17
retenir	24	reprendre	19	ordonner	17
soustraire	24	avancer	19	étudier	17
écraser	24	recueillir	19	repousser	17
exciter	24	animer	19	rapprocher	17
percevoir	23	anéantir	19	placer	17
établir	23	exprimer	19	remuer	17
associer	22	donner	18	scruter	17
distinguer	22	mener	18	désigner	17
réunir	22	rejeter	18	tourmenter	17
troubler	22	rapporter	18		

Avec le seuil de fréquence choisi de 4, le même groupe de mots (*trépider, s'agiter, trembler, grelotter, trembloter, frissonner, frémir...*) s'oppose au groupe (*crier, gémir, clamer, mander, confirmer, hurler,*

*promettre, notifier, corroborer, certifier, témoigner, narrer, relater, proclamer, réciter...*) sur le premier axe, et au groupe (*démoraliser, exténué, ternir, lasser, massacrer, assombrir, abîmer, désespérer, contrister, navrer, noircir, désoler, déshonorer, corrompre, mâchurer, contaminer...*) sur le second axe.

En prenant un nouveau seuil de fréquence minimale de 5 (on élimine les verbes ayant moins de cinq occurrences dans le corpus), il reste 1 265 verbes distincts dans un corpus formé maintenant de 12 244 occurrences.

Avec ce nouveau seuil de cinq, c'est le groupe de mots (*désirer, rêver, vouloir, espérer, ambitionner, souhaiter, songer, prier, se représenter, supplier, implorer, demander...*) qui s'oppose à deux nouveaux groupes sur les deux premiers axes : le groupe – moins homogène – (*manger, désespérer, mâchurer, trembler, refroidir, décourager, lasser, déprimer, démanteler, abîmer, raser...*) sur le premier axe.

Le même groupe de mots s'oppose au groupe (*embrancher, agglutiner, corrélér, agréger, conjuguer, conjoindre, souder, jumeler, juxtaposer, fusionner, raccorder, connecter, moiser, annexer, marier, braser...*) sur le deuxième axe.

On pourrait encore incrémenter le seuil de fréquence pour montrer que l'on continue à obtenir des structures triangulaires (un groupe s'oppose à deux autres dans le premier plan factoriel engendré par les deux premiers axes), et, de surcroît, des structures instables, puisque les groupes caractérisant les axes principaux changent lorsque le seuil de fréquence minimale change.

En fait, l'analyse géométrique du nuage multidimensionnel de points-verbes montre que ce nuage est presque sphérique<sup>1</sup>. Cette quasi-sphère comporte à la périphérie des « grumeaux » qui sont des amas de verbes sémantiquement voisins. Ces « grumeaux » créent les axes principaux au gré de leur taille, qui dépend d'ailleurs du seuil de fréquence minimale choisi au départ. On est donc très loin de la structure elliptique du nuage de mots de la sémiométrie<sup>2</sup>, avec ses axes principaux stables et significativement distincts.

Pour des seuils de l'ordre de 12 à 18, c'est à dire pour des verbes apparaissant de douze à dix huit fois comme synonymes d'autres verbes de

---

1. Cette sphéricité se vérifie en particulier par la similitude des première *valeurs propres* dont les pourcentages de variance valent, pour les cinq premières : 0.77, 0.73, 0.69, 0.68, 0.66 (cas de l'analyse avec le seuil 4).

2. Rappelons que la signification statistique des axes sémiométriques a été éprouvée par *les intervalles de confiance d'Anderson*, des répétitions d'échantillons indépendants dans le temps et dans l'espace, et les méthodes de *bootstrap*, cf. chapitre 2.

la langue française<sup>1</sup>, on voit bien apparaître les groupements précités, mais les grandes oppositions observées sur les axes principaux sont décevantes, inattendues.

Ainsi, pour plusieurs de ces seuils, on voit apparaître, à l'extrémité d'un des axes, le groupe :

*[conjoindre, annexer, marier, accoupler, allier, relier, rattacher, rapprocher, rassembler, réunir, associer, assembler, rallier, lier, mélanger, joindre, enchaîner, mêler].*

Mais ce groupe de verbes relativement homogène et assez lié au pôle « Attachement » du troisième axe de la sémiométrie, s'oppose souvent au regroupement :

*[réclamer, demander, souhaiter, requérir, aspirer, solliciter, explorer, inspecter, sonder, scruter, fouiller, appeler, penser, estimer, considérer].*

Il ne se retrouve jamais opposé au regroupement fréquent pourtant plus antonymique que le précédent :

*[dégager, séparer, distinguer, discerner, abandonner, écarter, renvoyer, rejeter, éloigner, repousser].*

En fait, la notion de synonymie telle qu'elle est traitée dans les dictionnaires ou les thésaurus nous éloigne de l'esprit du questionnaire sémiométrique où il est demandé si les mots présentés évoquent des sensations agréables ou désagréables, et où le contexte même du questionnaire (liste de mots volontairement hétéroclite) et le temps imparti à sa complétion impliquent de prendre les mots suggérés dans un de leurs sens les plus courants. Au contraire, un dictionnaire de synonymes s'efforce d'être exhaustif et de ne pas omettre les emplois secondaires, métaphoriques, familiers ou argotiques<sup>2</sup>.

Illustrons, en marge de ces analyses statistiques, ces difficultés de la notion d'éloignement sémantique par une des réflexions de Gaston Bachelard, reprise par Roland Barthes dans ses *Mythologies*, pour qui, d'une certaine façon, « le vin est le contraire de l'eau... »<sup>3</sup>. La distance est ici plus axiologique que sémantique, et l'eau pourra être le contraire de la

1. Avec des seuils aussi élevés, il n'est pas question de voir apparaître les verbes rares absents de la liste des 829 verbes de base de notre corpus.

2. Le lecteur aura peut-être été surpris, lors de l'analyse précédente avec le seuil 4, de voir le verbe *manger* en compagnie de *démanteler*, de *raser*... c'est que *manger* comporte aussi comme synonyme *dilapider*, *dissiper*, *gaspiller*.

3. « Bachelard a déjà donné la psychanalyse substantielle de ce liquide (*le vin*) à la fin de son essai sur les rêveries de la volonté, montrant que le vin est suc de soleil et de terre, que son état de base est, non pas l'humide, mais le sec, et qu'à ce titre, la substance mythique qui lui est le plus contraire, c'est l'eau. » (Roland Barthes, *Mythologies*, Seuil, 1957).

sécheresse pour un agriculteur, du froid pour un esquimau, du solide pour un physicien, de la terre pour un marin, ou du vin pour le Français moyen de 1950 dont Barthes décrypte avec humour les valeurs et les mythes qui leurs sont associés.

Dans la même veine, un simple proverbe comme « le mieux est l'ennemi du bien » doit sa force lapidaire à la remise en question – légèrement paradoxale – de certaines proximités sémantiques avérées<sup>1</sup>.

A propos de l'expérience sur l'ensemble des verbes français et de sa grande ambition : découvrir dans le réseau sémantique de ces verbes une structure stable qui aurait un rapport avec celle observée en sémiométrie, on doit aussi constater, comme cela pouvait être pressenti, que les verbes porteurs de valeurs sont très peu fréquents dans l'ensemble des 829 verbes retenus.

Par nature, le verbe est lié à l'action, au mouvement, au temps, au changement. Les verbes techniques ou descriptifs abondent et en quelque sorte étouffent, dans le cas de notre corpus de synonymes, les verbes plus abstraits, affectifs ou relationnels (*aimer, juger, connaître, soigner, admirer, interdire...*).

Enfin, nous concluons par une remarque qui concerne à la fois les deux expériences précédentes : qu'il s'agisse de noms, d'adjectifs ou de verbes, le réseau sémantique n'est pas homogène<sup>2</sup>.

Certains mots ont beaucoup de synonymes, d'autres en ont peu, ou pas du tout. Un mot qui n'a pas de synonyme n'est pas moins important qu'un autre. C'est le cas par exemple des verbes *menacer* ou *voter*, mais c'est le cas aussi du mot *maternel*. Il faudra, dans une phase de recherche ultérieure, étendre les expériences précédentes en utilisant des corpus faisant appel, au delà de la synonymie, à des analogies, voire à des définitions de dictionnaire et des recueils de citations.

---

1. Beaucoup de clauses de style exploitent cette complexité topologique de l'espace sémantique, qui, pour les statisticiens, plus prosaïquement, est un espace que l'on ne peut pas munir facilement d'une distance euclidienne.

2. En termes plus techniques, mais plus précis, le graphe sémantique est loin d'être un graphe régulier, ses sommets (les verbes) ont des degrés très variables, certains sommets sont même isolés.