

ALISE LEHMANN
FRANÇOISE MARTIN-BERTHET

Lexicologie

Sémantique, morphologie,
lexicographie

5^e édition

ARMAND COLIN

Illustration de couverture © Daniel Fung — shutterstock.com

Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du

Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Armand Colin, 2018

Armand Colin est une marque de
Dunod Éditeur, 11 rue Paul Bert, 92240 Malakoff

ISBN 978-2-200-62237-4

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2^o et 3^o a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

Sommaire

Avant-propos	5
1 Mot et lexique	9
PARTIE 1	
SÉMANTIQUE LEXICALE	
2 Le signe linguistique	27
3 Les analyses du sens lexical	33
4 Les relations sémantiques	73
5 La polysémie	95
PARTIE 2	
MORPHOLOGIE LEXICALE	
6 La formation des mots en diachronie	145
7 L'analyse des mots en synchronie	159
8 La dérivation	185
9 La composition	223
Annexe aux parties 1 et 2. Études de mots dans un texte	245

PARTIE 3
ÉLÉMENTS DE LEXICOGRAPHIE

Introduction. Lexicographie, métalexigraphie, dictionnairique	255
10 Les types de dictionnaires	257
11 La macrostructure	267
12 La microstructure	283
Bibliographie	329
Table des matières	343
Index	349

Avant-propos

Objet de la lexicologie

La *lexicologie* a pour tâche d'inventorier les unités qui constituent le lexique, et de décrire les relations entre ces unités.

Le lexique en effet n'est pas une simple liste, qu'on ne pourrait ordonner que par l'ordre alphabétique ; il s'organise sur les deux plans du sens et de la forme :

- La *sémantique lexicale* étudie l'organisation sémantique du lexique : elle analyse le sens des mots et les relations de sens qu'ils entretiennent entre eux.
- La *morphologie lexicale* étudie l'organisation formelle du lexique : elle analyse la structure des mots et les relations de forme et de sens qui existent entre eux.

Ces deux aspects concourent à construire une structure complexe, elle-même insérée dans l'ensemble du système de la langue.

Relations avec les autres domaines

Le lexique est lié à la *syntaxe*, puisque les mots sont employés dans des phrases.

Le sens des énoncés n'est pas seulement fait du sens des mots qui les composent : il dépend aussi de la syntaxe de la phrase et de la situation d'énonciation. La *sémantique lexicale* est donc une partie de la *sémantique*, étude du sens.

La forme des mots variables est en partie liée à leur emploi dans la phrase : cette variation, appelée *flexion* (conjugaison, déclinaison, variation en nombre et en genre) est l'objet de la *morphologie flexionnelle*, ou *morphologie grammaticale*. La *morphologie lexicale* est donc une partie de la *morphologie*, étude de la forme des mots.

La forme orale des mots est étudiée par la *morphophonologie*.

Domaines voisins

Le lexique est partiellement représenté et décrit dans les dictionnaires : on appelle *lexicographie* la fabrication et l'étude des dictionnaires. La *lexicologie* et la *lexicographie* se sont constamment inspirées mutuellement, mais l'objet « langue » et l'objet « dictionnaire » sont de nature différente.

On distingue le lexique général, ou commun, et les lexiques de spécialité (disciplines scientifiques, techniques, métiers, secteurs d'activité, etc.) : l'étude des lexiques de spécialité est la *terminologie*, dont les unités sont des *termes*.

Les mots enfin sont attestés dans des textes ou discours, et l'étude des textes a pu privilégier ce niveau d'analyse : la *lexicologie* a, par là, partie liée avec la *stylistique* et l'*analyse de discours*. L'un des outils de ces disciplines est la *lexicométrie*, ou *statistique lexicale*, qui mesure la fréquence des mots dans un texte ou un ensemble de textes (corpus). On oppose parfois *lexique* de la langue et *vocabulaire* d'un texte ou d'un discours.

Dans ce livre, on s'en tiendra au lexique général dans sa forme écrite.

Dans les deux premières parties, on s'efforce de dégager les fonctionnements *sémantiques* et *morphologiques* fondamentaux, en les illustrant par de nombreux exemples ; sans retracer l'histoire des idées sur le lexique, on indique autant que possible l'origine des notions issues des traditions philosophique, rhétorique et grammaticale, ou des développements plus récents de la linguistique.

Le point de vue est synchronique ; cependant, les faits diachroniques peuvent être évoqués quand ils apportent un argument à l'analyse ou qu'ils permettent de mieux comprendre la terminologie héritée de la tradition historique.

Des exercices d'application permettent de vérifier les notions acquises et de les affiner, à l'occasion des problèmes particuliers rencontrés.

À l'issue des deux premières parties, l'exercice consistant à étudier des mots dans un texte, comme le demandent notamment les épreuves de français moderne des concours de recrutement des professeurs de français, met en jeu l'ensemble des notions de sémantique et de morphologie.

Dans la troisième partie, on aborde les représentations du lexique données par les dictionnaires avec des éléments de *lexicographie*.

.....

La première partie, *Sémantique lexicale*, a été rédigée par A. Lehmann.
La deuxième partie, *Morphologie lexicale*, a été rédigée par F. Martin-Berthet.
Dans la troisième partie, *Éléments de lexicographie*, les chapitres 10 et 11 ont été rédigés par F. Martin-Berthet ; le chapitre 12 a été rédigé par A. Lehmann.

■ Chapitre 1

Mot et lexique

1. Le mot
2. Le lexique

1. Le mot

Chaque domaine de la linguistique doit définir son unité : les unités de la phonétique et de la phonologie sont le *son* et le *phonème* ; l'unité de la syntaxe est la *phrase* ; l'unité du lexique est communément appelée *mot*. L'identité d'un mot est constituée de trois éléments : une forme, un sens et une classe syntaxique.

« Un mot résulte de l'association d'un sens donné à un ensemble de sons donnés susceptible d'un emploi grammatical donné. » (A. Meillet, *Linguistique historique et linguistique générale*, Paris-Genève, Champion-Slatkine, 1982 [1913], p. 30)

À ces trois aspects correspondent trois types de difficultés pour définir le mot et établir la liste des unités du lexique.

1.1 La classe syntaxique

Dans la définition du mot par Antoine Meillet citée ci-dessus, l'expression « emploi grammatical » désigne la classe syntaxique, par laquelle le mot s'intègre à la phrase. En français, la tradition grammaticale reconnaît huit *classes de mots*, appelées aussi *parties du discours* : nom, verbe, adjectif, déterminant, pronom, adverbe, préposition, conjonction.

L'expression traditionnelle **parties du discours** est le calque du latin *partes orationis*, où *partes* signifie « parties » par rapport au « tout » de l'énoncé (*oratio*) ; ce sont les unités constituantes de la phrase.

On peut parler aussi de *catégories morphosyntaxiques* (puisque des traits morphologiques interviennent dans leur définition, à savoir la variation en genre, nombre, personne) ou de *catégories grammaticales* ; mais ce dernier terme a aussi un autre sens : il désigne les catégories notionnelles qui ont une expression grammaticale, comme le temps, la personne, le genre, le nombre.

Dans la tradition de la grammaire scolaire, la classe du mot est sa **nature** (opposé à **fonction**).

Les mots comme *allo !*, *ouf !*, *chut !* sont recueillis dans une neuvième classe, celle des **interjections** ; mais il ne s'agit pas d'une classe de mots semblable aux autres : ce sont des mots qui sont des phrases (ou *mots-phrases*) et non des constituants de la phrase.

Certaines formes correspondent à plusieurs classes et ont des sens tout à fait distincts, comme *boucher* Nom et *boucher* Verbe. (Du point de vue étymologique, ils ont des origines différentes.) Ce sont donc deux mots différents, qui sont *homonymes* (cf. chap. 5, 1.3).

L'un des mots peut être dérivé de l'autre : seule la classe change, le sens étant globalement conservé ; par exemple, *déjeuner* Nom dérive de *déjeuner* Verbe. C'est une *conversion* (cf. chap. 8, 2).

La forme *faux* peut être a) un adjectif qualificatif (*une fausse note, ce passeport est faux*) ; b) un adverbe (*il chante faux* ; on parle aussi d'adjectif employé comme adverbe) ; c) un nom (*ce document est un faux*), qui vient de l'adjectif par conversion ; d) un autre nom homonyme (*couper l'herbe à la faux*, en relation avec le verbe *faucher*), qui n'a aucun rapport de sens avec le précédent (et qui a une autre origine). Selon les analyses, on distinguera deux mots (a/b/c + d), ou quatre (a + b + c + d).

De plus, certains mots ont plusieurs **emplois** pour une même classe, c'est-à-dire qu'ils entrent dans plusieurs constructions et contextes sans changement de classe. Par exemple, le verbe *casser* a un emploi transitif (*casser une vitre*) et un emploi intransitif (*la branche casse*) ; le *café* que l'on boit n'est pas celui où l'on va ; un *pauvre* homme n'est pas un homme *pauvre* ; le dérivé *pauvreté* est lié à ce second emploi. La différence de sens liée à la différence d'emploi est

plus ou moins importante (cf. ci-dessous 2.1) ; mais chaque emploi est isolable comme une unité du lexique.

1.2 Le sens

1.2.1 Mots homonymes et mot polysémique

À certaines formes de même classe correspondent plusieurs sens. Ainsi, *bise* signifie « vent du nord » ou « baiser » ; *perle* signifie « petite boule de nacre », « personne remarquable (dans sa partie : *une perle rare*) », « erreur ridicule (dans un texte) ». On devrait alors, pour s'en tenir à la définition donnée, considérer qu'il y a autant de mots différents que de sens différents (*homonymie*) ; dans le deuxième cas pourtant (*perle*), la notion d'un lien métaphorique entre les différents sens pourra suggérer que c'est un mot unique qui a plusieurs acceptions (*polysémie*). Il est souvent difficile de trancher avec la seule intuition sémantique (cf. chap. 5). Il y a des contextes différents (*cette personne est une perle, un collier de perles, perles relevées dans une copie ; cf. ci-dessus emplois*) mais aussi des contextes communs, non discriminants (*trouver une perle*).

1.2.2 Mots pleins et mots vides

Certains mots dits *vides* n'ont pas de sens par eux-mêmes : il s'agit surtout des **mots grammaticaux** (prépositions, conjonctions, déterminants, pronoms), opposés aux **mots lexicaux** ou *mots pleins* (noms, adjectifs, verbes, adverbes ; on parle aussi de *catégories majeures*). Ceux-ci ont un sens plus autonome et une référence : ils renvoient à un objet ou une notion définissable (cf. chap. 2).

On appelle **grammaticalisation** le « passage d'un mot autonome au rôle d'élément grammatical¹ » c'est-à-dire le fait qu'un mot lexical devienne grammatical ou acquière un emploi grammatical, avec une perte de sens. Par exemple, les formes verbales de participe présent *pendant* (verbe *pendre*) et *durant* (verbe *durer*) donnent des prépositions : *pendant ce temps, durant cette période* ; le verbe de mouvement *aller* devient un auxiliaire de futur proche : *je vais partir dans cinq minutes* ; les noms *pas, point* deviennent éléments d'adverbe négatif avec *ne*, à

1. A. Meillet, *Linguistique historique et linguistique générale*, Paris-Genève, Champion-Slatkine, 1982 [1912], p. 130-133.

partir de contextes comme : *je ne marche pas, je ne vois point*¹. Le nom *question* « a donné naissance à une nouvelle préposition : *Question chômage, c'est catastrophique*. » : la grammaticalisation conjugue à des degrés divers recatégorisation (morphologique), réanalyse (syntaxique) et réinterprétation² (sémantique).

B. Pottier oppose *lexèmes* et *grammèmes* et appelle *grammémisation* la transformation d'un lexème en grammème³.

Cependant, les mots grammaticaux sont plus ou moins vides : par exemple, certaines prépositions ont un sens spatial particulier (*sur, devant*). D'autre part, le statut grammatical implique moins l'absence de sens qu'un mode de signification : le mot grammatical a un sens très abstrait, il concourt à l'interprétation de la phrase en mettant en relation les mots lexicaux. On opposera alors *sens grammatical* et *sens lexical*.

Dans une terminologie plus récente, on oppose *sens instructionnel* (consistant plutôt en « instructions » guidant la mise au point du sens des mots et de la phrase) et *sens référentiel* ou *descriptif* (consistant à décrire les propriétés d'un référent ; cf. chap. 2, 1 ; chap. 3, 4).

1.3 La forme

1.3.1 Mot fléchi

Les mots variables ont plusieurs formes : ils ont une **flexion** (conjugaison, déclinaison, variation en genre et en nombre). Les différentes formes du mot variable sont aussi des mots : on peut parler du mot *grand* (adjectif), en tant qu'unité lexicale abstraite transcendant ses formes observables dans les phrases (*grand, grande, grands, grandes*), et du mot *grands* dans la phrase : *Les enfants sont grands* ; ou du mot *partir* (verbe) et du mot *partirons* dans la phrase : *Nous partirons à midi*.

On utilise conventionnellement l'une des formes du mot variable pour le nommer en tant qu'unité lexicale : le masculin singulier des mots variables en

1. G. Gougenheim, *Les Mots français dans l'histoire et dans la vie*, Omnibus, 2018 [Picard, 1966], p. 636, *Mots à sens plein devenus mots grammaticaux*.

2. Ch. Marchello-Nizia, *Grammaticalisation et changement linguistique*, p. 15, p. 42 sq.

3. *Linguistique générale, Théorie et description*, Klincksieck, 1974, p. 272.

genre et en nombre, le singulier des mots variables en nombre, l'infinifit présent des verbes, le cas sujet des mots à déclinaison.

Cette forme est celle qui sert d'**adresse** (ou **entrée**) dans les dictionnaires (sauf indication de la variation en genre et en nombre par l'entrée : *blanc, blanche* ; *œil, yeux* ; cf. chap. 12, 1.3), ou de **lemme** dans le domaine du traitement automatique du langage. La **lemmatisation** consiste à indexer les différentes formes fléchies sous cette forme unique.

Si une des formes ou si chaque forme a un sens particulier, elle acquiert le statut d'unité lexicale différente. Cela arrive souvent avec la variation en genre : par exemple, *cuisinière* « appareil » n'est pas la forme féminine de *cuisinier* « humain » ; *loup* « demi-masque noir » n'est pas la forme masculine de *louve* « animal » ; *couturier* et *couturière* désignent des métiers différents. Le nombre peut aussi être concerné : on distinguera *la vacance* (d'un poste) et *les vacances* (les congés).

On utilise souvent le terme **lexème** (par analogie avec *phonème* et *morphème*) pour référer à l'unité lexicale abstraite. Ce terme est opposé à *mot-forme* ou *lexe* (cf. l'opposition *morphème / morphe*, voir chap. 7, 1) chez I. Mel'čuk¹.

1.3.2 Mot graphique

Fur, guingois, guise, plupart sont des mots graphiques : ils sont délimités à l'écrit par les blancs². Ce sont d'anciens noms. Aujourd'hui, ce ne sont plus des unités lexicales : les unités intégrant de la phrase sont les expressions complexes dont ils font partie : locutions adverbiales (*de guingois, au fur et à mesure, à ma, ta, sa... guise*) ; locution prépositionnelle (*en guise de*) ; locution conjonctive (*au fur et à mesure que*) ; déterminant ou pronom indéfinis (*la plupart de Article N, la plupart*).

Pomme, de et terre sont trois unités lexicales du français ; *pomme de terre* en constitue une quatrième, composée des précédentes.

L'orthographe ne délimite pas toujours les mots (unités lexicales) ; l'identification et la délimitation de ceux qui sont composés de plusieurs mots

1. *Cours de morphologie générale*, Presses de l'Université de Montréal-CNRS Éditions, 1993, volume 1.

2. Le terme de typographie est *espace*, nom féminin.

graphiques, ou mots **polylexicaux**, doit se faire d'après des critères linguistiques (cf. chap. 9, 1).

2. Le lexique

L'ensemble des mots d'une langue constitue son lexique.

Cet ensemble est structuré par des relations entre ses unités ; il se diversifie selon un certain nombre de variables ; il n'est pas clos, et ses contours ne sont pas fixés de manière absolue.

2.1 Les relations

2.1.1 Relations paradigmatiques et syntagmatiques

Les unités de la langue sont reliées entre elles par deux grands types de relations :

- Les *relations* (ou *rappports*) **syntagmatiques** sont des relations d'enchaînement (ou concaténation) et de contiguïté : *le + livre, lire + un + livre* ; un *syntagme* est un groupement d'unités (*le livre, lire un livre, le livre de ton ami*).
- Les *relations* (ou *rappports*) **paradigmatiques** sont des relations d'équivalence ; un *paradigme* est l'ensemble des unités substituables dans un même enchaînement syntagmatique : par exemple, *le / un / ce / mon / chaque /etc. (livre doit être lu) ; lire un(e) livre / texte / roman / note / affiche / etc.*

Chaque point d'une séquence d'unités est le croisement d'un *axe vertical* (*paradigmatique*), où est sélectionnée l'unité, et d'un *axe horizontal* (*syntagmatique*), où les unités se placent les unes à la suite des autres, en raison de la *linéarité* du langage.

« Le signifiant, étant de nature auditive, se déroule dans le temps seul et a les caractères qu'il emprunte au temps : a) il représente une étendue, et b) cette étendue est mesurable dans une seule dimension : c'est une ligne. »
L'écriture « substitue la ligne spatiale des signes graphiques à leur succession dans le temps. » Ferdinand de Saussure, *Cours de linguistique générale*, Payot, 1995 [1916], Première partie, Chapitre 1, p. 103.

L'opposition entre *paradigme* et *syntagme* est issue d'une « dichotomie saussurienne » voisine mais différente, où les rapports « verticaux » sont dits **associatifs** et sont indépendants des rapports horizontaux syntagmatiques : « les mots offrant quelque chose de commun s'associent dans la mémoire, et il se forme ainsi des groupes au sein desquels règnent des rapports très divers ». Par exemple, *enseignement* entre dans les *séries associatives* a) *enseignement, enseigner, enseignants*, etc. b) *enseignement, armement, changement*, etc. c) *enseignement, instruction, apprentissage, éducation*, etc. Le *rapport associatif* unit des termes *in absentia* dans une série mnémorique virtuelle ». « Le *rapport syntagmatique* est *in praesentia* : il repose sur deux ou plusieurs termes également présents dans une série effective » ; son support est l'étendue. Ces deux ordres « correspondent à deux formes de notre activité mentale, toutes deux indispensables à la vie de la langue. » *Cours de linguistique générale*, Deuxième partie, Chapitre V, p. 170-175)

Le terme *syntagme* désigne chez Saussure toute combinaison d'unités consécutives : *re-lire, contre tous, la vie humaine, Dieu est bon, s'il fait beau, nous sortirons*, etc. (p. 170). Il désigne plutôt aujourd'hui un groupe de mots syntaxique à l'intérieur de la phrase.

Le terme **paradigme**, emprunté au grec ancien, vient de la grammaire, où il signifie :

- 1) « exemple », « modèle » (c'est le sens du mot en grec) ; c'est le mot type, celui qui sert de modèle pour présenter une conjugaison ou une déclinaison : par exemple, *chanter* est le paradigme des verbes du premier groupe ; en latin *rosa* est le paradigme de la première déclinaison des noms.
- 2) « liste close de formes fléchies » (conjugaison, déclinaison), par exemple les formes du verbe *chanter* au présent de l'indicatif ; on passe de l'exemple à ce qu'il exemplifie, par métonymie.

Le point commun entre ces sens en grammaire et le sens linguistique évoqué ci-dessus est la notion de rapports « verticaux » et « *in absentia* » par opposition aux rapports syntagmatiques, « horizontaux » et « *in praesentia* ».

2.1.2 Relations sémantiques, morphologiques, syntaxiques

Le terme **relations lexicales** (relations entre les mots) désigne communément les **relations de sens** entre mots de même classe syntaxique et sans lien morphologique : synonymie, antonymie, hyperonymie et hyponymie, méronymie et holonymie (voir chap. 4). Une relation sémantique comme l'antonymie est lexicale (*grand / petit*) ou morphologique (*juste / injuste*). Une certaine synonymie peut exister aussi entre mots de classe syntaxique différente, comme

entre un verbe et un nom d'action ; elle est lexicale (*tomber* sujet humain et *chute*) ou morphologique (*chuter* et *chute*) (voir chap. 8, 1.2.2).

Les **relations morphologiques** sont des relations de forme et de sens : elles ont lieu entre un mot simple et les mots construits qu'il produit (dérivés, composés) et entre les mots construits de même structure (voir Deuxième partie, « Morphologie lexicale »).

Les **relations syntaxiques** sont des relations combinatoires (syntagmatiques) de construction (comme entre un verbe et ses objets : par exemple, écrire + lettre, livre, message, etc.) et de collocation (une peur intense est une peur bleue ; voir chap. 4, 2.1.2). Ce sont aussi les relations paradigmatiques de classe et sous-classe (nom, nom d'animé, verbe, verbe intransitif, etc.).

Ces trois types de relations sont intimement liés. La combinatoire syntaxique est associée à des compatibilités sémantiques. Une relation de forme sans relation de sens est purement phonétique (et graphique) et relève de la paronymie ou de l'homonymie (voir chap. 5, 1). Les relations morphologiques portent sur des emplois définis syntaxiquement (voir *supra* 1.1 ; chap. 8, 1.2.1). Une relation comme celle entre un verbe et le nom d'action synonyme concerne à la fois la morphologie, le sens et la syntaxe.

Fonctions lexicales. Le modèle Sens-Texte (qui cherche à représenter comment on va du sens au texte qui l'exprime) regroupe les liens entre les mots en deux grands types : les *dérivations sémantiques* et les *collocations*.

Les *dérivations sémantiques* comprennent non seulement les relations de sens classiquement répertoriées (telles que synonymie, antonymie) mais aussi le changement de classe grammaticale (*marchander / marchandage, vendre / vente, frapper / coup ; terre / terrestre, ville / urbain*), et la relation situation / élément de situation, participant (actant) ou circonstant (*nager / nageur, dormir / lit*). Elles sont ou non morphologiques. (Cette notion est reprise à J. Picoche, voir chap. 8, 1.2.2).

Les *collocations* associent une base à ses *collocatifs* : ex. *dormir : comme une souche*.

Ces liens sont représentés comme des *fonctions* (au sens mathématique) appliquées à des *arguments* : « instrument » (*couper*) > *couteau, hache*, etc. ; « intensificateur » (*méchant*) > *comme la gale, comme une teigne*¹.

1. I. Mel'čuk, A. Polguère, *Lexique actif du français*, Bruxelles, 2007, p. 18-23.

2.2 Les variables

2.2.1 Les domaines

Le lexique général est commun à tous les locuteurs ; les *lexiques de spécialité* sont liés à un *domaine* : science (chimie, astronomie), science et technique (informatique), métier (menuiserie), activité (jardinage), métier et activité (politique), etc. L'étude des lexiques de spécialité est la *terminologie*.

2.2.2 Les usages

Une langue vit longtemps et peut être parlée dans un espace très étendu : l'usage du lexique varie selon deux facteurs principaux, qui sont le temps et le lieu. Il se diversifie aussi selon les groupes sociaux et la situation de communication.

La variation **diachronique** a lieu dans le temps. Si un fonds stable demeure depuis les débuts du français, le lexique du français contemporain n'est pas identique à celui des périodes précédentes : certains mots anciens ne sont plus utilisés ni compris (ex. *malandrin*) ; des mots nouveaux, ou *néologismes*, apparaissent (ex. *zapper*). Cette variation est sensible entre les générations.

La variation **diatopique** a lieu dans l'espace. Les régions de France et les pays de la francophonie présentent des mots qui leur sont propres : ainsi *cheni* « saleté, désordre », *panosse* « serpillère » sont employés en Savoie et en Suisse ; *magasiner* est un québécoisisme qui signifie « aller faire des achats dans les magasins ».

La variation **diastratique** est liée aux groupes sociaux (représentés comme des *strates*) : certaines façons de parler sont considérées comme « populaires », d'autres au contraire comme marquant l'appartenance à un milieu social favorisé et cultivé. Certains groupes ont des argots spécifiques (« argot des malfaiteurs », argot scolaire, verlan des « jeunes des cités »). Cependant, la détermination sociale impliquée par cette notion est moins facile à appliquer au lexique qu'à d'autres aspects du langage, comme la prononciation ou la syntaxe.

La variation **diaphasique** (*-phas-* « parler », cf. *aphasie*) est liée à la situation de communication, qui sera plus ou moins familière ou formelle, induisant un vocabulaire plus ou moins familier ou recherché chez un même locuteur ou entre des locuteurs par ailleurs non différenciés par les autres facteurs.

Les dimensions *diastratique* et *diaphasique* ont des expressions voisines ; elles sont souvent confondues et exprimées en termes de *registres* ou de *niveaux de langue*. La *diastratie* est aussi *diatopie* dès lors que les groupes sociaux sont séparés dans l'espace : c'est ce que montre une expression comme « langage des cités ». Les classes d'âge (« langage des jeunes ») concernent la *diachronie* et la *diaphasie*. Les notions de *strates* et de *niveaux* peuvent véhiculer une dimension hiérarchique et une attitude normative, au détriment de leur vocation descriptive.

La description des usages est complexe, comme le montrent les marques lexicographiques courantes, qui se fondent sur des critères divers : situation de communication (*familier*), attitude de l'énonciateur (*péjoratif*), jugement moral ou esthétique (*vulgaire*), etc. La marque *littéraire* est tantôt une marque de domaine (elle réfère à un type de texte), tantôt une marque de registre (signifiant « soutenu »). Un terme comme *populaire* signifie plutôt « très familier » que « utilisé par les classes populaires ». (Voir chap. 11, 1.2 ; chap. 12, 2)

Le terme *diachronique* est, chez Saussure, opposé à *synchronique*. Les autres termes ont été forgés pour rendre compte de la variation synchronique. Les termes *diatopique* et *diastratique* sont dus au norvégien Leiv Flydal. Eugenio Coseriu les reprend et introduit le terme *diaphasique* pour traiter plus spécifiquement des registres liés aux situations de communication, tels que « usuel », « solennel », « familier ». Toutefois, les aspects sociaux et les niveaux de langue étant très mêlés, on utilise couramment aujourd'hui *diastratique* de façon extensive¹.

Les mots attachés à l'usage d'un temps, d'un lieu ou d'un registre sont dits *marqués*, par opposition à un fonds commun *non marqué*.

2.2.3 La fréquence

Un petit nombre de mots très fréquents constituent un noyau autour duquel se superposent des couches de fréquence moindre, jusqu'aux mots très rares et aux *hapax*, qui sont les mots dont on a une seule attestation. La fréquence est évaluée intuitivement par les locuteurs d'après leur expérience : par ex. chacun sent que *quand* est plus fréquent que *lorsque*. La fréquence est une donnée

1. E. Coseriu, « Structure lexicale et enseignement du vocabulaire », *Les théories linguistiques et leurs applications*, Conseil de la coopération culturelle, AIDELA, Nancy, 1967, p. 33. Pour un développement de cette nomenclature par F.-J. Haussman, voir l'*Encyclopédie internationale de lexicographie*, Berlin / New-York, Walter de Gruyter, 1989-1991, t. I, p. 649-657.

objective quand elle est tirée de l'observation des occurrences du mot dans un *corpus*, ou ensemble de textes ; mais elle est alors dépendante de ce corpus. Les moyens informatiques permettent aujourd'hui de réunir des corpus suffisamment importants pour fournir une bonne image de la fréquence des mots. Pour le français, on dispose notamment, grâce au laboratoire ATILF (Analyse et traitement informatique de la langue française, CNRS – Université Nancy 2, www.atilf.fr), de la base FRANTEXT, qui rassemble un très grand nombre de textes à dominante littéraire.

Les expressions *mot courant* ou *sens courant* sont ambiguës, parce qu'elles peuvent renvoyer aux trois variables évoquées ci-dessus : *courant* peut s'opposer 1) à *spécialisé* 2) à *marqué* 3) à *rare*. La rareté est d'ailleurs souvent liée au caractère spécialisé ou marqué.

Le terme *registre* est parfois utilisé de manière extensive, renvoyant à tous les aspects indiqués ici (domaine, usage, fréquence).

La *fréquence absolue* est le nombre d'occurrences dans un corpus. La *fréquence relative* est rapportée à la longueur du corpus.

Le TLF, *Trésor de la langue française, Dictionnaire de la langue du XIX^e et du XX^e siècles* (www.atilf.fr/tlfi), donne à la fin de certains articles la fréquence absolue de l'entrée dans les textes littéraires et les fréquences relatives dans les quatre demi-siècles, d'après le *Dictionnaire des fréquences, Vocabulaire littéraire des XIX^e et XX^e siècles* d'É. Brunet¹.

Dans *Le vocabulaire français de 1789 à nos jours, d'après les données du Trésor de la langue française*², É. Brunet donne les fréquences des 6 700 mots de fréquence absolue supérieure à 500 dans FRANTEXT, par genre et par période.

J. Picoche et J.-C. Rolland, pour établir la nomenclature de leur *Dictionnaire du français usuel*³, sont partis des 907 mots de fréquence supérieure à 7 000 dans la liste d'É. Brunet, pour sélectionner 442 entrées consistant en mots lexicaux hyperfréquents et polysémiques qui conduisent, par leurs relations sémantiques et morphologiques, à un lexique de 15 000 mots.

Il faut noter que des mots usuels dans les échanges quotidiens apparaissent peu dans les listes statistiques, parce que leur apparition est liée à des thèmes

1. Paris, Didier, 1971.

2. Tome II, Slatkine, 1981.

3. De Boeck-Duculot, 2002.

ou à des situations : c'est le cas notamment des noms d'objet comme *livre*, *chaise*. Aussi, lorsqu'on a cherché dans les années 1950 à établir un « français fondamental » pour l'enseignement, a-t-on corrigé la fréquence par la *disponibilité*, évaluée par des enquêtes auprès d'un échantillon de locuteurs¹.

L'étude des phénomènes de fréquence lexicale dans les corpus est la *lexicométrie*, ou *statistique lexicale* (encore appelée *statistique linguistique* ou *linguistique quantitative*, dans la mesure où les comptages peuvent porter aussi sur des aspects morphosyntaxiques, comme les temps des verbes). Depuis P. Guiraud², elle a porté notamment en France sur des œuvres littéraires³ et sur le vocabulaire politique⁴. Elle entre aujourd'hui dans le cadre plus général de la *linguistique de corpus*.

2.3 Les marges

Le lexique comporte à ses marges un ensemble flou de mots plus ou moins intégrables dont le statut est incertain.

2.3.1 Noms propres

On considère que, à la différence des noms communs et des autres mots de la langue, ils n'ont pas véritablement de sens, mais seulement un référent (*cf.* chap. 2), qui est une entité individuelle : *Paris*, *la Loire*, *Zola*.

Ce point a été débattu par la philosophie : pour certains, il y a un sens des noms propres, qui se constitue des caractéristiques de leur référent et se dégage de leurs emplois dans les discours. De plus, il existe des classes sémantiques de noms propres : noms de pays, prénoms, noms d'animaux domestiques, etc.

1. G. Gougenheim, R. Michéa, P. Rivenc, A. Sauvageot, *L'élaboration du français fondamental (1^{er} degré)*, Didier, 1964.

2. *Les caractères statistiques du vocabulaire*, PUF, 1954, *Problèmes et méthodes de la statistique linguistique*, PUF, 1960.

3. Ch. Müller, *Étude de statistique lexicale, Le vocabulaire du théâtre de Pierre Corneille*, Slatkine, 1993 [Larousse, 1967] ; É. Brunet, *Le vocabulaire de Jean Giraudoux, Structure et évolution*, Slatkine, 1978 ; D. Mayaffre éd., *Etienne Brunet - Écrits choisis Volume 1 : Comptes d'auteurs. Études statistiques*. De Rabelais à Gracq, Champion, 2009.

4. M. Tournier et « groupe de Saint-Cloud », *La parole syndicale, Étude du vocabulaire confédéral des centrales ouvrières françaises, 1971-1976*, PUF, 1982 ; D. Mayaffre, *Le poids des mots, Le discours de droite et de gauche dans l'entre-deux-guerres*, Champion, 2000.

Certains noms propres entrent dans le lexique par leurs dérivés, qu'il s'agisse de noms de lieux (*français, parisien, savoyard*) ou de noms d'auteurs (*cornélien, rabelaisien, marxisme*) et de personnages (*gargantuesque, ubuesque*).

D'autres passent dans la langue en devenant des noms communs. C'est le cas des noms de personnages emblématiques désignant un type d'individus : un *don Juan* ou *don juan* est un séducteur, un *tartufe* ou *tartuffe* est un hypocrite, une *Lolita* ou *lolita* (du nom de l'héroïne éponyme du roman de Nabokov) est une « nymphette », un *Machiavel* est un « homme d'État sans scrupule » (définitions du *Petit Robert*) ; Mécène, Mentor (*un mécène, un mentor*) sont des personnages de l'histoire et de la littérature antiques. La rhétorique classique a répertorié ce fait dans les « figures de mots », sous le terme *antonomase*. Ils peuvent prendre la marque du pluriel (*les don Juan* ou *les dons juans, les Lolita* ou *les lolitas, les tartuffes, les Machiavels, les mécènes*) et donner des dérivés : *donjuanesque* et *donjuanisme, tartuf(f)erie, machiavélique* et *machiavélisme, mécénat*. Le passage à l'initiale minuscule ne se fait pas toujours (*Machiavel*) : la lexicalisation comporte des degrés ; l'allusion culturelle est plus ou moins présente, le registre est souvent soutenu.

Certains auteurs parlent de *nom propre métaphorique*, par opposition aux *noms propres métonymiques* tels que : un Picasso = « un tableau de Picasso ».

Les noms de personnages emblématiques entrent aussi à titre de parangon dans des comparaisons figées à valeur intensive, appartenant là encore à un registre plutôt soutenu ; ils gardent alors la majuscule et la référence particulière : *fier comme Artaban, riche comme Crésus, vieux comme Hérode*.

Des noms d'inventeurs ou de responsables, généralement oubliés, deviennent par métonymie le nom de leur invention ou de l'objet qu'ils ont promu : *godillot* (Alexis Godillot, fournisseur de l'armée), *poubelle* (Eugène Poubelle, préfet de la Seine), *guillotine* (le docteur Guillotin).

Les noms de marque (noms déposés) s'intègrent au lexique ou aux phrases de deux manières différentes :

- ils servent à désigner un type de produit : *frigidaire, goretex, sopalin* ; cet emploi est critiqué par les puristes ;
- ils fonctionnent comme noms de sous-classes d'objets (*hyponymes*, cf. chap. 4, 1.1) : *rouler en Renault* (sorte de voiture), *mettre ses Nike* (sorte de chaussures), *être en Chanel* (sorte de vêtements) : dans ce cas, ils gardent

leur majuscule, restent invariables, et les dictionnaires de langue ne les répertorient pas.

Inversement, des noms communs deviennent des noms propres : c'est le cas des noms d'événements ou de périodes (*la Renaissance*), ou encore d'institutions (*l'Assemblée nationale*).

2.3.2 Mots étrangers

Toute langue emprunte à d'autres une partie de son lexique. Les mots étrangers sont « hors système » : ils ont leurs particularités phonétiques et morphologiques. Les *emprunts* sont assimilés quand ils sont conformes aux structures du français (*sentimental*, par exemple, est un anglicisme daté de 1769), ou quand ils se coulent dans ses moules phonétiques, orthographiques et morphologiques : *beefsteak* devient *bifteck* ; *blog*, *look* donnent *blogueur*, *relooker*. En revanche, *apartheid* (afrikaans) ou *tchador* (persan) comportent des sons qui ne sont pas dans le phonétisme français. Nombre de mots étrangers peu utilisés restent aux marges du lexique.

Emprunt, calque, xénisme. Le terme *emprunt* désigne tout élément provenant d'une autre langue, il a une valeur générique. Un *calque* est un emprunt qui résulte de la traduction littérale d'une expression (*col blanc* : anglais *white-collar* ; *gratte-ciel* : *skyscraper*), d'un dérivé (*surpêche*, *surréagir* : anglais *overfishing*, *overreact*) ou d'une acception figurée (*souris* au sens de « boîtier connecté à un ordinateur » : anglais *mouse*). Dans ce dernier cas, on parle aussi d'*emprunt sémantique*. Le terme *xénisme* est réservé à l'emprunt qui correspond à une réalité étrangère (afrikaans *apartheid*, russe *toundra*).

L'emprunt aux langues anciennes (grec et latin). Un mot peut être emprunté au grec ancien ou au latin de la même façon qu'aux langues modernes : *humus* est latin ; *agora*, *chaos*, *misanthrope* sont grecs. Deux autres types d'emprunt aux langues anciennes ont un statut différent :

- le latin étant à la fois langue mère et langue d'emprunt, le français a des *mots savants* empruntés au latin, qui parfois doublent un *mot populaire* de même étymon : le latin *auscultare* donne en français *écouter* et par emprunt *auscul-ter* (voir chap. 6, 1) ;
- le français emprunte au latin et au grec des *éléments* pour construire des *composés savants* (*omnivore* ; *xénophobe*), des *dérivés savants* dont le radical a la forme d'un mot latin (*épiler* de *pilus* « poil », *carcéral* de *carcer* « prison ») ou pour servir de préfixe ou de suffixe (*hypersensible*, *insecticide*).

Il s'agit alors de modes de formation internes au français. Voir chap. 6, 2.1 ; chap. 8, 1.2.2 ; chap. 9, 3.

2.3.3 Mots nouveaux

L'apparition d'un mot nouveau, ou *néologisme*, est un fait historique (qu'il n'est pas toujours facile de dater d'ailleurs) qui a lieu dans un discours individuel, par nécessité (nommer une réalité ou une notion nouvelle), par recherche d'expressivité, par jeu, etc. Il s'intègre ou non dans la langue comme bien commun de la collectivité¹.

La nouveauté est dans la forme, le sens ou la syntaxe :

- Les néologismes de forme créent une nouvelle forme ; pour la plupart, ils sont produits par les procédés de formation que sont la dérivation (cf. chap. 8) et la composition (cf. chap. 9) : par ex. *démondialisation*, *éco-quartier*, *monte-escalier* ; ce sont aussi des emprunts : *selfie*, *sudoku* ;
- Les néologismes de sens créent un nouveau sens pour une forme donnée, selon certains mécanismes sémantiques, tels que les métaphores, les métonymies, les extensions ou restrictions (cf. chap. 5, 3) : par ex., on peut parler de *polluer l'atmosphère* dans un sens moral et psychologique, par métaphore.
- Les néologismes syntaxiques créent un nouvel emploi, soit par changement de classe grammaticale (ou *conversion*, cf. chap. 8, 2), soit par changement de construction : par ex. à *angoisser* transitif à objet humain (*cet endroit m'angoisse*) s'ajoute depuis peu une construction intransitive à sujet humain : *j'angoisse* « j'éprouve de l'angoisse ».

Un nouveau sens est souvent aussi un nouvel emploi : dans le « langage des cités », l'adjectif *moelleux* s'applique à une personne (nouvel emploi) avec le sens « mou, sans énergie² » ; *être (un) décroissant* (Adj et N) quand il s'agit d'une personne, c'est remettre en cause la croissance économique.

On appelle parfois mots *théoriques* ou *possibles* les formes produites par le système virtuel de la langue et non (encore) attestées, dans lesquelles puisent les néologismes de forme.

1. J.-F. Sablayrolles, *La néologie en français contemporain, Examen du concept et analyses de productions néologiques récentes* ; J. Pruvost, J.-F. Sablayrolles, *Les néologismes*.

2. *Lexik des cités*, Éditions Fleuve, 2007.

« Un mot que j'improvise, comme *in-décor-able*, existe déjà en puissance dans la langue ; on retrouve tous ses éléments dans les syntagmes tels que *décor-er*, *décor-ation* : *pardonn-able*, *mani-able* : *in-connu*, *in-sensé*, etc., et sa réalisation dans la parole est un fait insignifiant en comparaison de la possibilité de le former. » F. de Saussure, *Cours de linguistique générale*, Payot, 1996 [1967], Troisième partie, Chapitre IV, p. 227

Le lexique attesté s'augmente ainsi de la masse des mots « en puissance » dans la langue. Cependant, on ne peut pas dire quels sont les mots « théoriques » sans avoir élucidé tous les mécanismes de la formation des mots ; or, ces mécanismes ne sont pas totalement décrits. Il est encore plus difficile de prédire les sens ou les emplois possibles. Il faut noter aussi que l'usage n'est pas indifférent : savoir que *irrécupérable* est lexicalisé, mais non *indécorable*, ou *imbattable* et moins *battable* (cf. chap. 6, 2.3 ; chap. 7, 2.3.3), fait partie de la compétence lexicale.

■ **Partie 1**

Sémantique lexicale

■ Chapitre 2

Le signe linguistique

1. Le signe linguistique et le référent
2. Extension vs intension
3. Dénotation vs connotation(s)

La sémantique lexicale a pour objet l'étude du sens des unités lexicales. Elle se sert des concepts liés au signe, hérités souvent de la philosophie. On définira brièvement ici les concepts fondamentaux.

1. Le signe linguistique et le référent

1.1 Le signe

Ferdinand de Saussure (1857-1913) définit le signe linguistique comme une entité à double face, l'une sensible qui est son **signifiant** (l'image acoustique), l'autre, abstraite, qui est son **signifié** (le concept).

Signifiant et signifié s'impliquent réciproquement. Au signifiant oral [flœR] ou au signifiant graphique (*fleur*) est associé le signifié « production de certains végétaux, colorée, souvent parfumée ».

« Le signe linguistique unit non une chose et un nom, mais un concept et une image acoustique. Cette dernière n'est pas le son matériel, chose purement physique, mais l'empreinte psychique de ce son, la représentation que nous en donne le témoignage de nos sens [...] »

F. de Saussure, *Cours de linguistique générale*, première partie, chap. 1, p. 98

Le lien unissant le signifiant au signifié est arbitraire.

« Ainsi l'idée de "sœur" n'est liée par aucun rapport intérieur avec la suite de sons *s-ø-r* qui lui sert de signifiant ; il pourrait aussi bien être représenté par n'importe quelle autre... » (*Ibid.*, p. 100)

1.2 Le référent

Les signes linguistiques permettent au locuteur de parler de la réalité qui l'entoure. Ils ont, en effet, la propriété de pouvoir renvoyer aux objets du monde, extérieurs à la langue ; ces objets sont les référents. Signes linguistiques et référents ne doivent pas être confondus. Ainsi c'est bien la fleur (la plante concrète) qui embaume et non pas le mot *fleur* ni le signifié de *fleur*. La confusion entre le signe et le référent désigné par le signe s'observe dans l'apprentissage du langage. Dans sa thèse, I. Berthoud-Papandropoulou rapporte différents tests qu'elle présente à des enfants âgés de 4 à 6 ans¹. L'un des tests consiste à leur soumettre des mots de longueur différente en demandant de juger de leur dimension. À la question : « Pourquoi *armoire* est un mot long ? », l'enfant répond : « parce qu'il y a beaucoup d'affaires dedans ». Les enfants répondent à la question **métalinguistique** – c'est-à-dire portant sur la langue – par des réponses liées aux caractéristiques du référent. Ces réponses traduisent le refus de l'arbitraire du signe et le besoin de motivation (*cf.* chap. 6, 3).

Les référents sont des entités matérielles ou conceptuelles (êtres, objets, lieux, processus, propriétés, événements, etc.). Ils relèvent de l'univers extralinguistique réel ou fictif (par exemple, la sirène). Les référents ne sont pas tous des données immédiates du réel. Ils sont médiatisés par la langue. Les rapports entre l'ordre de la langue et l'ordre du monde sont complexes ; ils ont nourri les débats philosophiques dès l'Antiquité.

On distingue l'emploi référentiel du signe de son emploi autonymique. Lorsque le signe ne renvoie pas à un référent (externe) mais à lui-même, il est dit **autonyme** (il se désigne lui-même comme signe). Dans l'exemple : *Chat rime avec rat*, *chat* est autonome (on dit que le locuteur fait mention du signe). En revanche, dans l'exemple : *Le chat miaule sur le balcon*, *le chat* renvoie au référent (on dit que le locuteur fait usage du signe).

1. I. Berthoud-Papandropoulou, *La Réflexion métalinguistique chez l'enfant*, Université de Genève, 1980.