

# **LA LEXICOLOGIE**

2<sup>e</sup> édition



# LA LEXICOLOGIE

AÏNO NIKLAS-SALMINEN

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<sup>er</sup> juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

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, Paris, 1997

ISBN : 978-2-200-26936-4

© Armand Colin, 2005, 2010

© Armand Colin, 2015 pour la présente édition

Internet : <http://www.armand-colin.com>

Armand Colin est une marque  
de Dunod Éditeur, 5 rue Laromiguière, 75005 Paris

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2<sup>o</sup> et 3<sup>o</sup> 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

<b>Introduction</b>	<b>7</b>
<b>1  NOTIONS FONDAMENTALES</b>	<b>8</b>
1. Le signe linguistique	8
2. Le mot et le morphème	16
3. Le lexique et la grammaire	25
4. Le lexique et ses usages	27
5. Le français fondamental et la statistique lexicale	34
6. Le lexique et l'univers	41
<b>2  ÉTUDE DE LA FORME DES MOTS</b>	<b>46</b>
1. La dérivation	47
2. La composition	65
3. L'abréviation et la siglaison	73
<b>3  ÉTUDE DES RELATIONS LEXICALES</b>	<b>78</b>
1. La synonymie	78
2. L'antonymie	83
3. L'hyponymie et l'holonymie	88
4. L'homonymie et la paronymie	90
5. La polysémie	93
6. Le sens propre et le sens figuré	97
7. L'autonymie	101
<b>4  ANALYSE DU SENS DES MOTS</b>	<b>102</b>
1. L'analyse sémique	102
2. La sémantique du prototype	121
3. Les rapports entre le lexique et la syntaxe	126
4. La sémantique par rapport à la pragmatique	133
<b>5  ÉVOLUTION DU LEXIQUE</b>	<b>136</b>
1. Les couches diachroniques du lexique français	136
2. La néologie et le néologisme	140
3. Les néologismes formels et sémantiques	142
4. L'emprunt	144
5. L'évolution du sens des mots	153

<b>6 LA LEXICOGRAPHIE</b>	<b>159</b>
1. Les différents types de dictionnaires	160
2. La macrostructure du dictionnaire de langue	162
3. Le mot-entrée	167
4. La définition et l'exemple	169
5. Autres informations données dans l'article de dictionnaire	176
6. Présentation de quelques dictionnaires actuels	177
<b>Exercices corrigés</b>	<b>185</b>
<b>Alphabet phonétique international</b>	<b>209</b>
<b>Bibliographie succincte</b>	<b>211</b>
<b>Glossaire des notions importantes</b>	<b>213</b>
<b>Index des principales notions utilisées</b>	<b>219</b>
<b>Table des encadrés</b>	<b>223</b>

# INTRODUCTION

La lexicologie, qui étudie les unités lexicales d'une langue, est une science relativement récente. Ses méthodes sont l'objet de constantes discussions et les résultats encore partiels. Dans un sens restreint, la lexicologie est considérée comme une branche de la sémantique qui a pour objet l'étude du sens des unités lexicales. Selon cette perspective, elle se confond, en grande partie, avec la sémantique structurale. Dans un sens plus large et plus généralement accepté, la lexicologie s'intéresse également à la forme des unités lexicales et même aux relations qui existent entre le lexique et la syntaxe. Menée dans une perspective historique, la lexicologie étudie l'apparition de nouvelles unités lexicales et l'évolution du sens des mots.

Le lexique est situé au carrefour des autres secteurs de la linguistique, la phonologie et la morphologie pour la forme des mots, la sémantique pour leur signification et la syntaxe pour leurs propriétés combinatoires. Le lexique, au lieu de constituer un système au sens strict, forme un ensemble ouvert et non autonome. C'est pour cela que l'on ne peut pas en donner une description systématique ou simple, mais seulement des descriptions complémentaires, selon le point de vue adopté.

Ce manuel, composé de six chapitres, s'adresse en priorité aux étudiants de lettres du premier cycle et aux candidats aux concours de recrutement de l'enseignement. Le premier chapitre met en lumière les notions fondamentales utilisées en lexicologie et leur propose la terminologie la plus généralement acceptée. Le deuxième chapitre est consacré à l'étude de la forme des différents types de mots qui caractérisent le lexique de la langue française. Les relations lexicales que les mots entretiennent entre eux, ainsi que le sens propre et le sens figuré des mots sont abordés dans le troisième chapitre. L'analyse sémantique des mots et l'étude des rapports qui existent entre le lexique et la syntaxe sont présentées dans le quatrième chapitre. Le cinquième chapitre porte sur l'évolution du lexique et le sixième sur l'étude des dictionnaires.

À la fin du livre, un certain nombre d'exercices commentés permettent de pratiquer la discipline de manière active et d'approfondir quelques points particuliers. Bibliographie et références sont, dans cet ouvrage, très succinctes. En effet, si celles-ci et des notes sont indispensables dans un ouvrage de recherche, elles se révéleraient encombrantes dans ce manuel qui a pour but de donner aux débutants une présentation accessible et précise de la lexicologie actuelle.

*Je tiens à remercier Joëlle Gardes Tamine pour ses remarques et conseils constructifs.*

# CHAPITRE 1

# NOTIONS FONDAMENTALES

## 1. LE SIGNE LINGUISTIQUE

## 2. LE MOT ET LE MORPHÈME

## 3. LE LEXIQUE ET LA GRAMMAIRE

## 4. LE LEXIQUE ET SES USAGES

## 5. LE FRANÇAIS FONDAMENTAL ET LA STATISTIQUE LEXICALE

## 6. LE LEXIQUE ET L'UNIVERS

Avant d'aborder l'étude morphologique et sémantique du lexique, il nous semble indispensable de définir les notions fondamentales utilisées en lexicologie. Il nous paraît également important de proposer notre terminologie. En effet, les discussions suscitées entre linguistes par la définition de ces notions ne vont pas sans une grande variété dans la terminologie. Nous essaierons de définir aussi clairement que possible celle que nous emploierons et, dans une perspective pédagogique, d'utiliser la plus simple possible.

Au cours de ce chapitre, on mettra d'abord en lumière la nature du signe linguistique dans l'univers des signes. On s'intéressera ensuite aux différents types de mots qui constituent le lexique. On verra que parmi les unités lexicales du français, il y a des mots simples et des mots construits qui entretiennent des relations paradigmatiques et syntagmatiques avec d'autres mots. On s'intéressera aussi aux différences qui existent entre le lexique et la grammaire, ainsi qu'aux usages du lexique et à la notion de variation. Le volet suivant s'attachera au français fondamental et à la statistique lexicale. À la fin de cette partie consacrée à la présentation des notions fondamentales, l'on examinera les rapports qui existent entre le lexique et l'univers.

## 1. LE SIGNE LINGUISTIQUE

### 1.1. L'UNIVERS DES SIGNES

Le signe linguistique appartient à l'univers des signes. Il est important de distinguer d'abord le signe linguistique des autres signes. Un **signe**, au sens le plus large, désigne un élément X qui représente un autre élément Y ou lui sert de substitut. Dans le vocabulaire technique de la **sémiologie** (= science qui étudie tous les procédés ou



systèmes de communication et de signification), un signe est une entité composée de deux éléments solidaires : une **forme** et un **sens**. La forme est un élément perceptible par les organes sensoriels, par exemple un tracé que l'on peut voir, ou un son, simple ou complexe, que l'on peut entendre. En fait, il existe de nombreuses tentatives de classification des signes. Ici, on a choisi d'adopter le classement qui propose les distinctions suivantes :




– 1) Certains signes sont produits **sans volonté de communication** et d'autres impliquent une **intention de communiquer**. Cette distinction permet déjà une première approche des signes. En effet, on peut opposer l'**indice** au **signal**. À ce propos, on donne souvent l'exemple du ciel d'orage (Georges Mounin, *Clefs pour la linguistique*, Paris, Seghers, 1987, p. 37) : le *ciel d'orage* noir et menaçant n'a pas l'intention de communiquer avec le météorologiste, mais il est cependant l'indice d'une pluie possible. La *fumée* est l'indice du feu, les *larmes* l'indice de la douleur, les *boutons sur la peau* l'indice de telle ou telle maladie, etc. L'indice peut être défini comme un fait immédiatement perceptible qui fait connaître quelque chose à propos d'un autre fait qui ne l'est pas.

Dans ces exemples, il y a un rapport physique ou d'appartenance entre l'objet représentant et la chose ou l'idée représentée. Le ciel noir est étroitement lié à la pluie. La fumée, les boutons et les larmes sont les conséquences naturelles du feu, de la maladie et de la douleur : la fumée et les boutons ne sont pas volontaires, les larmes ne sont jamais censées l'être.

Contrairement aux indices non intentionnels, il y a des signes qui impliquent une volonté de communication. Dans ce cas, on parle de **signaux**. On a vu que le *ciel d'orage* n'a pas l'intention d'annoncer le mauvais temps, mais cet indice va conduire le responsable de la sécurité d'une plage à hisser un *drapeau rouge*. Ce drapeau, qui indique que la baignade est dangereuse, est donc un fait qui a été produit artificiellement pour servir d'indice. Alors que la forme du drapeau s'offre immédiatement à la vue, la signification (= le danger) associée à cette forme doit être comprise, ce qui suppose un apprentissage préalable. Selon cette perspective, la *canne blanche* est le signal de la cécité, la *croix verte* le signal des pharmacies, le *feu vert* le signal du passage libre, le *clin d'œil* le signal de la complicité, etc.

– 2) L'intention de communiquer permet donc de faire la différence entre l'indice et le signal. L'observation des rapports qui existent entre l'objet perçu et ce qu'il représente permet une deuxième distinction : celle entre le **symbole** et le **signe**. Un Z sur un panneau routier annonce un virage, une *tête de cheval* indique une boucherie chevaline, un *dessin de cuiller et fourchette entrecroisées* un restaurant. Dans ces trois exemples, il y a un rapport de ressemblance formelle entre la forme de l'objet représentant et celui de l'objet représenté. Le Z, la *tête de cheval* et le *dessin de cuiller et fourchette entrecroisées* sont des **symboles**. Le symbole est un signal qui marque un rapport analogique, constant dans une culture donnée, avec l'élément qu'il signifie.

Cependant, il convient de souligner que la plupart du temps il n'existe aucun lien naturel entre la forme de l'objet représentant et celui de l'objet représenté. Il n'y a pas de rapport d'analogie entre un *drapeau rouge* et une baignade dangereuse, ou entre une *canne blanche* et la cécité, ou encore entre une *croix verte* et la pharmacie. Le *drapeau rouge*, la *canne blanche* et la *croix verte* sont donc des signes. Il n'y a pas, non plus, de lien d'analogie entre le signe linguistique *cheval* [ʃ(ə)val], par exemple, et l'animal désigné par ce signe. Les indices relèvent des sciences d'observation, et les signes non linguistiques et les symboles, de la sémiologie. Pour schématiser et clarifier ces notions, on peut présenter le tableau suivant :

Aucune intention	Intention de communication	
INDICE	SIGNAL	
	A ↔ B	A...B
	 SYMBOLE	 SIGNE NON LINGUISTIQUE

(Source : in Christian Baylon et Paul Fabre, *Initiation à la linguistique*, Paris, Nathan, 1990, p. 5.)

Le **signe linguistique** est un signe particulier dans cet univers des signes, car le langage humain est un langage incomparablement plus riche, plus souple et plus efficace que n'importe quel autre langage. Comme tout système signifiant utilisé à des fins communicatives, les langues sont organisées sur deux plans :

- celui des **formes** (ou **signifiants**)
- et celui des **contenus** (ou **signifiés**).

On vient de constater que parallèlement au langage des hommes, il existe de nombreux autres systèmes de communication non linguistiques. Les systèmes de symboles ou les systèmes de signes arbitraires en font partie. La carte routière est un bon exemple des systèmes de symboles : chaque élément a sa représentation symbolique, les petits avions symbolisent des aérodromes, les petites touffes d'herbe des marais, les croix des cimetières, etc. Le code de la route, dans lequel les panneaux circulaires signifient une injonction, les panneaux rectangulaires une information, les panneaux triangulaires un danger, et ainsi de suite, forme également un système de communication non linguistique. En effet, la notion de système implique la présence de signes stables d'un message à l'autre qui se définissent fonctionnellement par leur opposition les uns aux autres.

Les langues naturelles se différencient de la plupart des autres systèmes par la propriété d'être doublement articulées. Chaque langue naturelle possède un petit nombre de **phonèmes** (= une unité de la chaîne parlée qui a une fonction différentielle, mais qui n'a pas de signification). Phonétiquement, le français possède 19 consonnes (auxquelles viennent s'ajouter deux consonnes dues à des mots empruntés à l'anglais et à l'espagnol), 3 semi-consonnes et 16 voyelles. Avec ces 38 unités sonores (cf. alphabet phonétique international, p. 209), on peut construire une infinité d'unités lexicales et morphologiques.

Ce type de combinaison s'appelle donc la « double articulation du langage ». On considère que les unités significatives constituent la première articulation, parce que c'est la couche du langage que l'on appréhende en premier. C'est elle qui véhicule le sens. Ainsi la suite phonique ou graphique : *Un enfant joue dans le jardin* se découpe en six de ces unités : *un, enfant, joue, dans, le et jardin*. Ces unités de première articulation sont généralement appelées **morphèmes** (= la plus petite unité ayant une signification dans la langue) pour les distinguer des **mots**, qui sont souvent constitués d'un seul morphème (ex. : *enfant, jardin, masque, juste*), mais qui peuvent aussi être formés de deux ou de plusieurs morphèmes (*enfant-in, jardin-age, dé-masqu-er, in-juste-ment* et *anti-constitution(n)-elle-ment*).

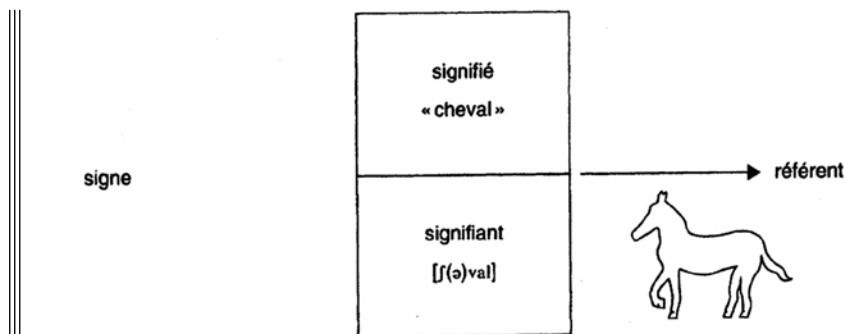
À un second niveau, les morphèmes s'articulent en segments distinctifs minimaux appelés « phonèmes ». Dépourvues en elles-mêmes de signification, ces unités de deuxième articulation ont pour unique fonction de distinguer entre elles les unités significatives de première articulation. Le mot *raison* [ʁɛzɔ̃], par exemple, est une combinaison de quatre phonèmes qui, comme telle, distingue ce mot des autres mots français : cette unité lexicale s'oppose en tous points à celle qui articule le mot *jardin*, mais ne se distingue que par son premier élément, r [ʀ], de celle qui articule le mot *saison* [sɛzɔ̃]. Toutes les langues naturelles sont orales avant d'être écrites, beaucoup de langues ne possèdent pas de forme écrite. Les alphabets dans les écritures alphabétiques font correspondre, bien que d'une manière souvent approximative, un nombre à peu près équivalent de lettres. Tous les énoncés d'une langue ayant adopté ce type d'écriture peuvent donc être retranscrits à l'aide d'un petit nombre de lettres. La langue française possède vingt-six lettres, quelques accents et quelques signes de ponctuation.

La double articulation donne au langage humain la créativité qui lui est propre, cette capacité d'exprimer par des combinaisons perpétuellement nouvelles des pensées constamment nouvelles.

## 1.2. LA NATURE DU SIGNE LINGUISTIQUE

On peut dire, en s'appuyant en partie sur des idées de Ferdinand de Saussure (*Cours de linguistique générale*, Paris, Payot, 1972), que le signe linguistique se caractérise par les traits suivants :

– a) Il est formé par l'association d'une « image acoustique », appelée « **signifiant** » et d'un concept ou « image mentale », appelé « **signifié** ». Ces deux faces du signe linguistique sont indissociables, puisque quand on prononce les sons [ʃ(ə)val], on évoque aussitôt le concept « cheval » et, inversement, ce concept ne peut exister indépendamment du matériel phonique. Ces deux faces sont donc solidaires comme le sont le recto et le verso d'une feuille de papier. Le signe linguistique réfère à un objet du monde, appelé « **réfèrent** ».



On doit prendre garde à ne pas confondre le réfèrent et le signifié. Le réfèrent est un fragment de réalité et le signifié est une représentation de cette réalité. Le signifié est donc une abstraction, une espèce de réalité psychologique. Il est plus pauvre et mieux organisé que la réalité. Il simplifie la complexité du réel et met en évidence l'essentiel en donnant un premier classement des éléments du monde. Par exemple, le signifié du signe *cheval* ne tient pas compte de la diversité des chevaux qui existent, mais ne retient que ce qui est commun à tous (crinière, sabots, queue...).

– b) Le lien entre signifiant et signifié est **arbitraire** : il n'existe aucun rapport interne entre le concept, celui de « cheval » par exemple, et la suite de sons qui le représente : [ʃ] + [(ə)] + [v] + [a] + [l]. On en veut pour preuve la variété des dénominations de langue à langue pour une même réalité signifiée : français *cheval*, anglais *horse*, finnois *hevonnen*, suédois *häst*. Chaque langue a sa propre façon de nommer le réel.

– c) Le rapport constitutif du signe linguistique peut être considéré comme **conventionnel**, puisqu'une fois établi, il s'impose aux usagers qui sont obligés d'accepter tels quels les signes en usage dans leur communauté. Toute infraction à la règle admise est sanctionnée socialement.

– d) Le signe linguistique donne la possibilité de parler d'objets ou de choses absents ou imaginaires. Quand on parle de chats, de démons ou de fées, on n'a pas besoin de les voir. On peut même n'en avoir jamais vu.

– e) Le signe linguistique est **typiquement humain**. Le pouvoir d'abstraction du signe linguistique fait partie des propriétés qui distinguent le langage humain du langage des animaux. D'après certains chercheurs, le gorille possède 22 cris différents, mais chacun d'eux est étroitement associé à une situation particulière. Aucun animal n'est capable de raconter ou d'évoquer un événement du passé, ou encore d'exprimer par la voix des sentiments différents en l'absence du stimulus qui aurait pu les provoquer.

La double articulation permet aussi de distinguer la communication linguistique humaine et la communication non linguistique animale. Chez certains animaux, on rencontre un système de combinatoire qui peut présenter quelques ressemblances avec la double articulation humaine. Un oiseau qui possède 5 notes de musique peut former, par exemple, 10 messages différents. Mais là s'arrête le processus. Cet oiseau ne composera jamais un 11<sup>e</sup> message.

– f) Les signes linguistiques sont la plupart du temps **arbitraires** (ou non naturels) puisque l'implication réciproque entre leurs deux faces n'est pas fondée sur une correspondance naturelle entre la forme du signifiant et les traits définitoires du signifié. D'une langue à l'autre, la même réalité notionnelle est souvent exprimée par des formes lexicales totalement différentes.

Cependant, dans chaque langue, malgré le grand nombre de signes arbitraires, on rencontre des signes qui entretiennent avec la réalité des relations moins arbitraires. Dans ce cas, on a affaire à des signes **motivés**. Un exemple de motivation du signe linguistique est offert par les **onomatopées** qui constituent une frange très marginale du lexique : *coucou*, *cocorico*, *meuh* imitent respectivement le chant d'un oiseau, du coq et le meuglement d'une vache. *Crac*, *boum*, *patatras*, *tic tac*, *froufrou*, *piqueter*, *miauler*, *picoter*, *murmurer*, *chuchoter*, *ronronner*, etc. reproduisent des bruits.

On peut pourtant constater que ces productions imitatives apparaissent elles-mêmes fortement conventionnalisées. Si la langue française a *cocorico* pour imiter le chant du coq, l'anglais a *cock-a-doodle-do*, l'italien *chichirichi* ([kiki-riki]), le japonais *kokekokko*, le hollandais *kukeleku* et le finnois *kukkokiekuu*. Cela montre que chaque langue interprète le réel selon ses propres habitudes phonologiques. Le signe linguistique, même motivé, reste donc conventionnel à l'intérieur d'une même communauté. Les onomatopées sont rarement compréhensibles aux membres d'une autre collectivité linguistique.

La motivation du signe linguistique s'observe aussi dans les formes complexes ou construites. Les signes du lexique français peuvent parfois apparaître motivés du point de vue morphologique : le *pommier* est l'arbre dont le fruit est la pomme, l'*abricotier* produit des abricots, le *cerisier* des cerises, le *citronnier* des citrons, le *châtaignier* des châtaignes, etc. De même, l'*ourson* est le petit de l'ours et le *chaton* le petit du chat. Dans ces exemples, la motivation résulte de l'emploi des procédés de dérivation.

Les mots dérivés, contrairement à leurs bases totalement arbitraires, sont relativement motivés. *Pommier* a été formé à partir de *pomme* à l'aide du suffixe *-ier*. *Abricotier*, *cerisier*, *citronnier*, *châtaignier*, ainsi qu'un grand nombre d'autres noms d'arbres fruitiers ont été créés de la même façon. Un signe à motivation relative est donc nécessairement complexe.

La motivation relative se rencontre aussi dans les mots composés combinant des signes élémentaires immotivés. Contrairement à *cent*, la forme composée *quatre-vingt-dix-neufs* s'interprète analytiquement comme le résultat de la multiplication de *vingt* par *quatre* auquel s'additionne la somme de *dix* et de *neuf*.

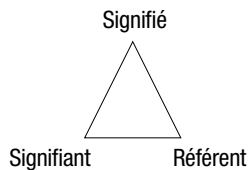
La poésie exploite les rapports qui existent entre le niveau phonique et le niveau sémantique du langage. Autrement dit, elle cherche à mettre les récurrences phoniques non signifiantes en relation avec le sens et ainsi à rendre le lien qui existe entre le signifiant et le signifié du signe linguistique moins arbitraire, plus naturel. Tout le monde sait que les poètes favorisent les onomatopées, les créations lexicales, les changements de sens, etc.

– g) Qu'il s'agisse de leur structure interne ou de leurs combinaisons, les signes linguistiques sont **linéaires**. Ce caractère linéaire du signe est dû à la nature orale du langage. Il est tout à fait impossible de prononcer simultanément deux sons, deux syllabes ou deux mots. La manifestation orale du langage se déroule donc dans le temps. Cette linéarité se répercute sur la transcription alphabétique qui se déroule dans l'espace : on ne peut pas écrire les unes sur les autres les différentes unités graphiques de la langue. Les lettres et les mots se succèdent sur la dimension de la ligne.

Le langage exploite doublement cette dimension unique. D'une part, quand on emploie les trois phonèmes /p/, /a/ et /l/, leur ordre dans /pal/ « pal », /alp/ « alpe » et /pla/ « plat » est extrêmement important pour la signification du message. D'autre part, quand on dit : *Pierre bat Paul*, cela ne veut pas dire la même chose que *Paul bat Pierre*. Cette caractéristique différencie les langues naturelles humaines de tous les systèmes de communication qui construisent leurs messages sur la trame de l'espace et non sur celle du temps.

### 1.3. LE SIGNE LINGUISTIQUE ET LE RÉFÉRENT

Le signe linguistique est souvent présenté sous forme d'un triangle.



Les signes linguistiques ont la propriété de pouvoir renvoyer aux objets du monde, extérieurs à la langue ; ces objets sont des **référents** qui peuvent être des entités matérielles ou conceptuelles (objets, êtres, processus, lieux, événements, etc.). Ils relèvent de l'univers extralinguistique réel ou fictif (par exemple, la licorne ou le dragon). Il est important de souligner que le processus référentiel de mise en relation d'unités linguistiques et d'unités extralinguistiques n'est pas unitaire. Selon Catherine Kerbrat-Orecchioni (*L'Énonciation de la subjectivité dans le langage*, Paris, A. Colin, coll. « Linguistique », 1997, p. 35), on peut faire la différence entre une référence absolue, une référence relative au contexte linguistique et une référence relative à une situation d'énonciation.

Dans le cas de la **référence absolue**, l'opération de référenciation se caractérise par son autonomie. Cela veut dire que le repérage du référent dépend uniquement des informations contenues dans l'expression référentielle. Par exemple, la construction référentielle de l'énoncé *la rose est une fleur*, ne dépend que de la connaissance des mots et des savoirs extralinguistiques des interlocuteurs.

Dans le cas de la **référence relative**, les expressions référentielles ne disposent pas d'un contenu descriptif suffisant pour permettre un acte de référenciation autonome. La référence doit se réaliser grâce à des informations différentes de celles qui sont disponibles dans les unités linguistiques. Si les données convoquées font partie de l'environnement linguistique de l'unité référentielle, la référence est dite « **relative au contexte linguistique** » ou « **anaphorique** ». Par exemple, dans l'énoncé *Marine est une fille intelligente, mais elle ne travaille pas assez*, l'interprétation référentielle du pronom *elle* dépend de la prise en compte d'un élément du contexte linguistique, le nom propre *Marine*. Cette procédure référentielle anaphorique est indirecte dans la mesure où elle effectue un détour par le contexte linguistique pour désigner le référent visé.

Si les données convoquées coïncident avec l'un des éléments constitutifs de la situation d'énonciation, la référence est « **relative à la situation d'énonciation** » ou « **déictique** ». Par exemple, si l'on dit *Est-ce que tu aimes ça ?*, l'objet désigné par le pronom démonstratif *ça* est repérable dans la situation de communication. Cette procédure référentielle est directe, elle n'est pas médiatisée par une autre forme linguistique présente dans l'énoncé.

#### 1.4. LES RELATIONS SYNTAGMATIQUES ET PARADIGMATIQUES

Les mots entretiennent entre eux deux types de relations fondamentales : les relations syntagmatiques qui s'observent entre les termes d'une même construction et les relations paradigmatiques qu'on peut établir entre une unité et toutes celles qui pourraient la remplacer dans un environnement donné.

Ainsi dans la phrase : *La petite fille joue avec son chat*, l'adjectif *petite* est en **relation syntagmatique** avec l'article défini *la* qui le précède et avec le substantif *fille* qui le suit. Ce groupe nominal est en relation syntagmatique avec le verbe *joue*, et ainsi de suite. Les relations syntagmatiques affectent donc des éléments qui sont présents dans l'énoncé. On leur donne souvent le nom de relations *in praesentia* (lat. : « en présence »).

Dans la même phrase, l'article défini *la* est en **relation paradigmatique** avec d'autres déterminants : *cette, ma, une, notre*, etc. ; l'adjectif *petite* avec d'autres adjectifs tels que *grande, jolie*, etc. ; le substantif *fille* avec des substantifs comme *voisine, cousine, dame*, etc. ; le verbe *joue* avec d'autres verbes tels que *court, danse*, etc. Les éléments en relation paradigmatique sont mutuellement substituables dans un environnement donné, s'y excluent les uns les autres et forment ensemble un paradigme. Les déterminants *la, cette, ma, une, notre*, etc. font partie du même paradigme, tout comme les adjectifs *petite, grande, jolie*, etc. Le paradigme est défini comme un ensemble d'unités virtuellement substituables dans un contexte donné. Les relations paradigmatiques existent entre des termes qui ne sont pas présents dans l'énoncé. On les appelle souvent relations *in absentia* (lat. : « en absence »). Elles s'opposent par là aux relations syntagmatiques.

## 2. LE MOT ET LE MORPHÈME

### 2.1. LE MOT

Le lexique désigne conventionnellement l'ensemble des mots au moyen desquels les membres d'une communauté linguistique communiquent entre eux. Cette définition du lexique oblige à donner une définition précise de l'**unité lexicale**, du **mot** en tant qu'élément de base de l'ensemble. La notion traditionnelle de mot est l'une de celles qui ont tendance à solliciter le plus constamment l'attention des linguistes. Cette notion, qui semble être familière et évidente pour le grand public, constitue pour le linguiste une source de difficultés théoriques considérable. Dans une langue telle que le français, on arrive à donner une définition simple et rigoureuse du mot uniquement au niveau de la manifestation graphique, où le mot est un groupement de lettres, séparé, à gauche et à droite, par un blanc des autres éléments du texte. Mais on constate très vite que ces segments ne correspondent pas à une réalité linguistique bien déterminée. Les blancs du texte ne s'harmonisent que rarement avec les limites du code oral. On peut malgré tout chercher à mettre en lumière les **critères phoniques** qui aident à identifier et à délimiter les unités lexicales dans la chaîne parlée.

#### 2.1.1. POINT DE VUE PHONIQUE

L'accent, les contraintes phonologiques et le coup de glotte démarcatif peuvent contribuer à l'identification des unités lexicales du point de vue sonore. Dans un grand



nombre de langues, l'**accent** joue un rôle démarcatif important. C'est le cas des langues à **accent lexical**. Dans certaines langues (ex. : tchèque, finnois, hongrois), l'accent tombe sur la première syllabe du mot ; dans d'autres, il peut occuper diverses places, mais toujours la même pour un mot donné (ex. : anglais, italien, russe). Le français est une langue à **accent syntaxique**. Si le mot est isolé, l'accent se place toujours sur la dernière voyelle du mot à l'exception du *e* caduc. Mais le mot ne fonctionne que rarement seul : la plupart du temps, il est inséré à l'intérieur d'un syntagme, d'un groupe et dans ce cas, l'unité d'accentuation est beaucoup moins le mot que le syntagme.

maison  
une maison rouge

En outre, en français, en dehors de cet accent dit « interne », non emphatique, qui est purement linguistique, il existe un accent dit « externe » ou emphatique qui est lié à des facteurs expressifs et intellectuels. En effet, l'accent peut être déplacé sous l'effet d'une volonté d'expressivité : l'accent du mot *formidable!*, qui frappe normalement la voyelle /a/ de la dernière syllabe peut très bien passer sur le /o/ initial. En français, le rôle démarcatif de l'accent dans l'identification des mots est donc très faible.

Les **contraintes phonologiques** peuvent également aider l'auditeur à identifier et à délimiter les unités lexicales dans la chaîne sonore. Il y a des phonèmes et des groupes de phonèmes qui sont possibles à toutes les places du mot. Dans ce cas, leur valeur démarcative est nulle. En revanche, certains phonèmes ou combinaisons de phonèmes sont impossibles à telle ou telle place du mot et peuvent ainsi donner des indications sur les limites initiales ou finales de l'unité lexicale. Alors que la présence d'un /ø/, possible à toutes les places (ex. *heureux, heureusement* : [øRø], [øRøzmã]) n'enseigne rien, celle d'un /œ/, impossible à la finale absolue, signale qu'on n'est pas arrivé à la fin d'un mot (cf. Jacqueline Picoche, *Précis de lexicologie française*, Paris, Nathan, 1992, p. 14).

Il est intéressant de noter que le mot français isolé peut commencer par n'importe quelle consonne. Toutefois, les mots commençant par les semi-consonnes [j, w, ɥ], ainsi que ceux qui ont pour initiale un *h* dit « aspiré » sont très rares :

ouaille, ouate, ouest, ouïe, ouistiti, etc.  
hachis, haïr, haine, hardiesse, etc.

Ce que l'on appelle aujourd'hui un *h aspiré* en français, n'est en fait ni un [h], ni aspiré. Il s'agit d'un phénomène d'absence de liaison (*les hachis* : on ne dit pas [lezafi]\*, mais [le'afi]) et d'absence d'éliision (on ne dit pas *l'hachis*\* [lafi], mais *le hachis* [lə'afi]). Le *h* dit « muet » permet l'éliision et la liaison, comme dans *l'homme* [lɔm] et *les hommes* [lezɔm].

Les trois consonnes nasales [m], [n] et [ɲ] peuvent toutes apparaître à l'intermédiaire : *hameau, anneau, agneau*, [amo], [ano], [aɲo] et à la finale des mots : *came*,

*canne*, *khâgne* [kam], [kan], [kaŋ]. Mais, à l'initiale, alors que les mots commençant par [m] ou [n] sont très nombreux, les mots commençant par [ɲ], comme *gnocchi* [ɲɔki], *gnôle* [ɲol] ou *gnognote* [ɲɔɲɔt], se comptent sur les doigts d'une main. En revanche, le *h* aspiré ne s'entend jamais dans une autre position qu'au début du mot. D'une façon générale, les consonnes (occlusives et constrictives) ont du mal à se maintenir à la finale du mot en français : par exemple, la consonne qui apparaît dans la flexion [simũt+e] (*cimenter*) ou tombe quand le radical est nu [simũ] (*ciment*).

Certaines langues (ex. l'allemand) connaissent le **coup de glotte démarcatif** qui a pour but de signaler la frontière des mots composant un syntagme. Ce phénomène est très rare en français (*le un*, *le onze*, *le harnais*, *un haricot* : [lɔ'œ̃], [lɔ'ɔz], [lɔ'arɛ], [œ̃'ariko]) qui a tendance à favoriser toutes sortes de liaisons et d'élisions.

Si l'on essaie de trouver des critères phoniques de délimitation du mot, on risque d'être déçu. Les blancs du texte écrit ne correspondent que rarement aux limites du code oral. On vient de constater qu'en français le rôle démarcatif de l'accent dans l'identification des mots est très faible par rapport aux accents des langues à accent lexical. Les locuteurs peuvent souvent faire entendre des accents ou effectuer des pauses en des points différents d'une même séquence sonore. On a également vu que les contraintes phonologiques ne peuvent aider l'auditeur que partiellement et que les liaisons et les élisions de toutes sortes ont tendance à contribuer à effacer les frontières entre les différents mots composant un syntagme.

Les critères sonores ne permettent donc pas aux linguistes d'appréhender le fonctionnement réel de l'élément mot. En effet, les traits démarcatifs phoniques ne semblent jouer, dans le découpage en mots de la chaîne parlée, qu'un rôle secondaire, les critères principaux étant d'ordre syntaxique et sémantique.

### 2.1.2. POINT DE VUE SYNTACTICO-SÉMANTIQUE

Dans les paragraphes suivants, on cherchera à mettre en relief les critères qui aident à identifier et à délimiter les unités lexicales dans la chaîne parlée en se plaçant sur le plan syntactico-sémantique.

Les unités lexicales sont inscrites comme unités dans le code de la mémoire du sujet parlant qui doit les reproduire telles quelles, en bloc, dans le discours. Le locuteur n'a pas la liberté de changer, à sens égal, l'ordre des éléments dans ces unités (*maisonnette*, *\*ettemaison*, *pomme de terre*, *\*de terre pomme*), ni de les séparer par des insertions (*\*pomme jaune de terre*).

En effet, sur le plan syntaxique on peut isoler des unités de fonctionnement en utilisant divers critères dont les principaux sont les procédés d'**inséparabilité** et de **commutation**. Ces manipulations ont pour but de segmenter les unités de la chaîne parlée et de les analyser (cf. Joëlle Gardes, *La Grammaire. Méthodes et notions*, « Cours », Paris, Armand Colin, 2012).