



UNIVERSITÉ DU BURUNDI
FACULTE DES LETTRES ET SCIENCES HUMAINES
Département de Langues et Cultures Africaines

LINGUISTIQUE GENERALE

(LCA1102)

Notes de l'ECUE (60H)

1^{ère} année de Bachelier en : - Langues et Cultures Africaines et
- Langues et Littératures Française

Par

Dr Gélase NIMBONA

Bujumbura, janvier 2026

Fiche signalétique du cours

Processus	Paramètres	Description
Élaboration	Thème	Linguistique générale
	Objectif général	L'objectif du cours est, d'une part, de fournir aux étudiants les informations essentielles sur la nature particulière du langage humain, de son origine, de son extension, de sa diversité et, d'autre part, de donner un aperçu des théories et courants linguistiques modernes, afin d'éclairer comment notre conception actuelle du langage est le résultat d'éclairages différents mais complémentaires.
	Prérequis	Les étudiants doivent avoir suivi préalablement le cours d' <i>Introduction aux sciences du langage</i> .
	Objectif spécifique	Fournir les outils nécessaires à l'étude du langage et des langues
	Bref contenu	Une bonne partie du cours consistera en un « éclairage historique ». Il s'agit de replacer les théories linguistiques dans leur histoire, pour voir comment la linguistique en tant que science a fini par se séparer d'autres formes de métalangage, et ce que le passé éclaire dans les théories modernes de la linguistique. Chacun des courants de la linguistique sera considéré en regard des méthodes avec lesquelles il analyse le langage. Le plus souvent en effet, c'est cette méthode d'analyse qui donne au courant son nom: comparatisme, structuralisme, distributionnalisme, cognitivisme, etc. Les grandes approches théoriques abordées sont le comparatisme, le structuralisme saussurien et américain, le générativisme, le cognitivisme et le fonctionnalisme. Quelques chapitres qui visent la rattachement des concepts et de méthodes développés dans ces courants par les étudiants sont proposés en annexes. Ces chapitres sont structurés en fonction de différents niveaux d'analyse de la langue et préparent les étudiants aux autres cours de la linguistique (la morphologie, la syntaxe, la sémantique, la phonétique et la phonologie, etc.
	Références bibliographiques	Aucun ouvrage n'est imposé. Toutefois, les étudiants peuvent au début consulter les ouvrages suivants pour se familiariser avec les concepts clés de la linguistique : <ul style="list-style-type: none"> • Auchlin, A. & J. Moeschler (2014). Introduction à la linguistique contemporaine. Armand Colin, Paris. • DUCROT, O. & SCHAEFFER J.-M. (1995) Nouveau dictionnaire encyclopédique des sciences du langage. Paris : Éditions du Seuil. • SIOUFFI, G. & VAN RAEMDONCK D. (1999), 100 fiches pour comprendre la linguistique. Rosny: Bréal.
	Activités	- Lectures et travaux pratiques.
Intervention	Déroulement	- Exposé magistral du cours - Exercices de description et d'analyse linguistique (TD) en rapport avec les concepts fondamentaux dans les domaines et courants de la linguistique - Échange avec les étudiants sur les travaux réalisés.
	Interactions	Discussion avec les étudiants sur les notions acquises ainsi que leur application dans les travaux réalisés.
Appropriation	Évaluation	TP/TD : 40% Examen écrit : 60%

Table des matières

1	DEFINITION DES CONCEPTS ET NAISSANCE DE LA LINGUISTIQUE.....	7
1.1	Qu'est-ce que la linguistique (générale) ?.....	7
1.1.1	Linguistique vs grammaire.....	7
1.1.2	Linguistique et logique.....	8
1.1.3	Linguistique et philosophie du langage.....	8
1.1.4	Linguistique et rhétorique.....	9
1.1.5	Qu'est-ce que la linguistique.....	9
1.2	La linguistique, discipline scientifique.....	10
1.2.1	Qu'entend par une discipline scientifique ?.....	10
1.2.2	Implication du terme Science (en linguistique).....	10
1.2.3	Linguistique = discipline scientifique, depuis quand ?.....	10
1.3	L'éclatement de la linguistique.....	12
1.4	Question terminologique : langue vs langage.....	12
1.4.1	Qu'est-ce qu'un langage ?.....	13
1.4.2	Qu'est-ce que la langue ?.....	14
1.5	Origine(s) des langues et diversité linguistique.....	14
1.5.1	Quid de l'origine du langage et des langues ?.....	14
1.6	La diversité linguistique.....	18
1.6.1	Monogenèse et polygenèse.....	18
1.6.2	Les langues du monde.....	20
1.6.3	Classification des langues.....	22
2	LES APPROCHES PRETHEORIQUES DES LANGUES ET DU LANGAGE.....	30
2.1	L'Antiquité.....	30
2.1.1	L'Inde.....	30
2.1.2	La Grèce et la philosophie du langage.....	30
2.1.3	La grammaire en Grèce et à Rome.....	31
2.1.3.3	La grammaire de Pāṇini (VI ^e s. av. J-C).....	32
2.2	Le Moyen Age.....	32
2.2.1	Les théories du langage dans le monde arabe.....	32
2.2.2	La grammaire spéculative en Occident.....	33
2.4	L'Age classique (17 ^e – 18 ^e siècle).....	35
2.4.1	L'enregistrement des données.....	35
2.4.2	Le projet de « grammaire générale ».....	35
2.3	Le 19 ^e siècle (le romantisme allemand).....	36
2.3.1	De nouvelles manières de comprendre le langage.....	36
2.3.2	La découverte du sanscrit.....	36
2.3.3	La grammaire comparée.....	37
2.3.4	La grammaire historique : les néogrammairiens.....	44
3	Les grands courants de la linguistique moderne.....	46
3.1	Ferdinand de Saussure (1857-1913) ou le Structuralisme.....	46
3.1.1	Au tournant de la linguistique moderne.....	46
3.1.2	La conception saussurienne de la linguistique.....	46
3.1.3	Le caractère linéaire du langage.....	51
3.1.4	La théorie de Saussure en bref.....	52
3.2	Le distributionnalisme (Bloomfield [1887-1949]).....	55
3.2.1	Structuralisme européen vs structuralisme américain.....	56
3.2.2	L'approche behavioriste et son application à la linguistique.....	56
3.2.3	Les propriétés du langage : les principes de base de la théorie distributionnaliste.....	59

3.3	Le générativisme (Noam Chomsky)	63
3.3.1	Critique du behaviorisme	63
3.3.2	Le programme générativiste	64
3.3.3	La grammaire générative	65
3.3.4	La question de l'apprentissage/acquisition	68
3.3.5	Les universaux dans la GGT	68
3.4	Le cognitivisme	69
3.4.1	Aperçu comparatif des théories précédentes	69
3.4.2	L'avènement <i>du cognitivisme</i>	70
3.4.3	Les nouvelles questions théoriques soulevées par le cognitivisme	73
3.4.4	<i>Postulats théoriques</i>	74
3.4.5	<i>Sens grammatical vs sens lexical</i>	75
3.4.6	<i>Le sens lexical : le prototype et la métaphore, deux principes fondamentaux</i>	78
3.4.7	Le sens syntaxique	80
3.4.8	La construction dynamique du sens : les espaces mentaux	81
3.4.9	L'apprentissage d'une langue dans la perspective cognitive	83
3.5	Le fonctionnalisme	83
3.5.1	La grammaire fonctionnelle	83
3.5.2	La pragmatique	88
3.5.3	La sociolinguistique	90
4	Langage et Communication	96
4.1	La communication linguistique	96
4.1.1	Le système linguistique (Rappel)	96
4.1.2	Caractéristiques de la communication linguistique	96
4.1.3	Les éléments constitutifs de toute communication verbale	97
4.1.3.1.	La communication chez F. de Saussure	97
4.1.3.2.	La communication chez L. Bloomfield	98
4.2	La communication non-linguistique	102
4.2.1	Culture et communication	102
4.2.2	La communication, système de codes interdépendants	102
4.2.3	Les procédés de communication non linguistique	104
4.3	Complexité de la communication humaine et limites du Schéma de communication proposé par Roman Jakobson	104
5	Des « mots » aux unités morphologiques et lexicales	107
5.1	Problèmes posés par la notion de mot	107
5.2	Des critères formels pour définir le mot	107
5.3	Le morphème	108
5.3.1	Les catégories de morphèmes	108
5.3.2	Morphème, morphe, allomorphe	109
5.4	La formation des unités lexicales	110
5.4.1	Formation par dérivation	110
5.4.2	Formation par composition	113
5.4.3	Autres procédés de renouvellement du lexique	114
6	La sémantique lexicale	118
6.1	Du signifiant au signifié: sémasiologie	118
6.1.1	La polysémie	118
6.1.2	Polysémie vs homonymie	120
6.2	Des signifiés aux signifiants : onomasiologie	120
6.2.1	Synonymie et antonymie	121
6.2.2	Champs lexicaux et traits sémantiques	121

7	Unités syntaxiques.....	123
7.1	Une première définition	123
7.2	Analyse en constituants immédiats et structure hiérarchique de la phrase	123
7.2.1	L'analyse en constituants immédiats et la notion de constituant	123
7.2.2	La structure hiérarchique de la phrase.....	124
7.3	Les classes de mots (analyse distributionnelle).....	125
7.3.1	La notion de distribution	125
7.3.2	Catégorie, sous-catégorie et extension d'une catégorie	127
7.3.3	Autres critères de caractérisation d'une catégorie grammaticale.....	127
7.3.4	Catégories grammaticales couramment admises.....	128
7.3.5	Conclusion.....	128
7.4	Les règles syntagmatiques.....	128
7.4.1	La structure interne des groupes syntaxiques.....	128
7.4.1.1.	Le Groupe Nominal (GN)	129
7.4.2	L'ambiguïté, un problème classique en syntaxe	131
8	Des « sons » aux unités phonologiques.....	132
8.1	Les phonèmes	132
8.2	La phonétique	133
8.3	Des sons aux systèmes phonologiques.....	134
8.3.1	L'organisation des systèmes phonologiques	134
8.3.2	Les règles d'agencement des phonèmes : la syllabe	135
9	Références bibliographiques	136

1 DEFINITION DES CONCEPTS ET NAISSANCE DE LA LINGUISTIQUE

1.1 Qu'est-ce que la linguistique (générale) ?

Les réflexions sur le langage et les langues ne sont pas récentes. Depuis l'Antiquité, le langage et les langues ont en effet suscité diverses investigations. Toutefois, la linguistique en tant que discipline scientifique et autonome ne s'est constituée qu'au cours du XIXe siècle. Aujourd'hui, on a tendance à ne plus parler de linguistique, mais de la ranger dans le domaine plus vaste des sciences du langage. Cette conception actuelle est révélatrice de la vision que l'on a fini par avoir de la linguistique.

1.1.1 Linguistique vs grammaire

Le terme « grammaire » issu du grec *grammatikê*, de *gramma*, qui signifie « lettre », a longtemps été le seul employé pour désigner l'étude de la langue. Depuis presque deux siècles maintenant, est apparu un nouveau terme, celui de « linguistique », bâti sur la racine latine du mot « langue* ». Aujourd'hui, ces deux termes ne sont plus synonymes ; ils renvoient à deux démarches/approches différentes (Siouffi & Raemdonck, 1999 :8).

La première approche est celle de ce que l'on pourrait appeler « la grammaire ». Au sens traditionnel du terme, la grammaire réfère à « l'art d'exprimer ses pensées par la parole ou par l'écriture d'une manière conforme aux règles établies par le bon usage » (cf. le dictionnaire le Littré). La visée de cette grammaire est proprement normative : la grammaire prescrit la manière de bien écrire, de bien s'exprimer. Les grammairiens tentent de décrire une langue donnée et son fonctionnement en essayant d'y trouver des règles et des régularités plus ou moins abstraites. Ces règles ne sont pas nécessairement toutes appliquées, mais on postule que celui qui les connaît est un bon locuteur de la langue concernée. En d'autres termes, l'objectif de la grammaire, au sens traditionnel du terme, consiste à définir ce qu'il convient de dire et d'écrire pour une utilisation correcte de langue. Il s'agit donc d'une optique normative : par l'étude de la langue, on veut établir une *bonne* manière de s'exprimer. En français par exemple, on dira qu'un mot ou qu'une tournure n'appartient pas au *beau* français, ne correspond pas à ce qui est recommandé par l'Académie française. La grammaire traditionnelle illustre donc une approche de la langue, un objet différent, une démarche différente de celle de la linguistique.

A partir du XIXe Siècle, on observe une certaine révolution scientifique. On ne va plus considérer une langue de manière normative, mais privilégier une approche descriptive de la langue étudiée : la langue est décrite en elle-même et pour elle-même. La linguistique, contrairement à la grammaire traditionnelle, vise en effet à rendre compte du fonctionnement de la langue en dégageant, sur la base de l'observation et de la description des actes linguistiques, des règles destinées à expliquer la compétence des sujets parlants. Elle s'attache au traitement de la langue dans tous ses aspects (phonétique, phonologie, morphologique ou syntaxique, sémantique, etc.) et sous toutes ses formes en usage chez les locuteurs. Elle tente de découvrir des régularités sous-jacentes à la production et à l'interprétation des échanges langagiers. Elle envisage la langue telle qu'elle est actuellement parlée ou telle qu'elle a été parlée à un moment donné de son histoire. Le linguiste adopte donc un point de vue synchronique, descriptive.

Les premiers travaux de linguistique descriptive sont cependant encore fortement marqués par la composante normative. Peu à peu, la linguistique devient une discipline scientifique qui a pour objet la langue.

En effet, on constate que la discipline souffre bien vite d'un véritable problème de théorisation. Qu'est-ce que la langue ? Qu'est-ce que le langage ? Qu'est-ce qu'un mot ? Par ce fait, on éprouve des difficultés à savoir ce que l'on peut réellement dire et étudier à propos

de formes linguistiques qui filent entre les doigts des chercheurs, particulièrement au niveau de la linguistique générale (On parle de linguistique générale lorsque son objet d'étude dépasse celui d'une langue particulière afin de dégager des généralisations et des théories sur le langage humain en général).

Par ailleurs, la définition du terme « grammaire » a évolué. D'une part, le terme « grammaire » renvoie à l'ensemble des règles à suivre pour parler et écrire correctement une langue et d'autre part, ce terme réfère à une « étude systématique des éléments constitutifs d'une langue, sons, formes, mots, procédés [terme vague pour syntaxe, stylistique ou rhétorique]. Si la première acception renvoie sans trop de difficulté à la définition de Littré, la seconde change la perspective. La visée de l'étude systématique n'est plus prescriptive, mais descriptive et explicative ; le terme « grammaire » pouvant aujourd'hui donc être utilisé pour désigner la démarche normative ou la démarche descriptive.

1.1.2 Linguistique et logique

Le mot logique vient du mot grec logos, qui signifie « verbe », « langue ». Etymologiquement, la « science du langage », c'est donc la logique. L'ouvrage d'Aristote intitulé *Logique* étudiait aussi bien des types de raisonnement que la manière que nous avons de les formuler. Par la suite, les travaux portant sur la pensée et ceux portant sur le langage vont graduellement se séparer. C'est ainsi que la logique est souvent identifiée de nos jours à une étude formelle des processus de pensée et de raisonnement, alors que la linguistique est identifiée à une étude du langage. Un premier pas vers cette séparation fut accompli à l'époque du philosophe allemand Gottfried Wilhelm Leibniz. Celui-ci estima que, si elle devait traiter de la validité de certaines façons de raisonner, la logique ne devait pas se contenter d'employer les mots de tous les jours, autrement, les langues naturelles, car celles-ci sont chargées d'ambiguïtés. A la fin du 19^e S, on remarque que certains penseurs (cf. notamment Heimann Steinthal 1823-1899) s'élèvent contre cette assimilation entre la logique et le langage. Ils estiment que les catégories du langage et de la logique sont incompatibles.

Toutefois, la linguistique et la philosophie du langage continuent à avoir besoin de logique. On signalerait en passant que les recherches logiques du philosophe allemand Edmund Husserl (parues en 1900-1901) inspirent par exemple le structuralisme. Husserl remarque que, dans le langage courant, tous les signes n'ont pas le même degré de signification, certains n'en ont que très peu, les conjonctions par exemple. Tout le travail de la logique a consisté à essayer de séparer l'analyse des formes de l'analyse de leur contenu. L'étude structurale des langues va s'efforcer d'appliquer ce principe et, sur ce modèle, de nombreuses modèles théories linguistiques séparent la syntaxe et la sémantique, considérant notamment volontiers que la syntaxe a une structure logique, qui a besoin d'être formalisée, comme on pourrait le faire d'un langage logique.

1.1.3 Linguistique et philosophie du langage

Depuis ses origines grecques, la philosophie est très concernée par les problèmes du langage. On appelle traditionnellement philosophie du langage un domaine de la philosophie qui choisit de se pencher aux problèmes généraux posés par le langage. Mais, dans certains de ses développements, la philosophie du langage porte aussi sur une investigation interne « interne » du langage. Là, elle rejoint alors directement certaines questions que se pose la linguistique.

Le langage a posé un problème à la philosophie dès que celle-ci s'est pensée comme une réflexion sur le vrai et sur l'être. Platon, dans le *Sophiste*, s'aperçoit que le fait d'articuler, dans une phrase, un nom et un verbe permet de dire le faux, de parler ce qui n'est pas. Toute l'histoire sera dès lors marquée par des interrogations sur le rapport entre le langage et la réalité, entre le langage et la vérité. Cette question se retrouve au Moyen Age, à l'époque classique, et

au 20^e siècle aussi.

La philosophie du langage s'est beaucoup développée au 20^eS et son développement est curieusement parallèle à celui de la linguistique. Il faut remarquer notamment qu'à partir des années 30, les idées du philosophe Wittgenstein vont exercer une grande influence, autant en philosophie du langage que certaines branches naissantes de la linguistique, comme la pragmatique. En philosophie, on parle *d'école d'oxford* ou *de philosophie analytique*. Le fondateur de la pragmatique, John L. Austin appartient à ce courant. De l'idée de Wittgenstein selon laquelle les mots et les phrases n'ont pas de sens en dehors du contexte où ils sont prononcés, il va tirer la notion d'acte de langage. Cependant, la philosophie du langage et la linguistique restent des disciplines relativement séparées. La philosophie du langage estime ne pas devoir s'occuper de ce que les langues naturelles comportent des codes. La linguistique, de son côté, se propose d'être une discipline infiniment plus pratique et concrète ; elle se veut d'abord une étude des langues plutôt que du langage. Si elle utilise des concepts dont le sens est parfois mis au point par la philosophie du langage, l'essentiel pour elle, est d'en tirer des applications pratiques dans l'étude de telle configuration linguistique ou de telle situation de parole.

1.1.4 Linguistique et rhétorique

L'Antiquité traitait les problèmes de langage dans trois domaines : la grammaire, la logique et la rhétorique. Cet héritage s'est transmis dans la tradition occidentale grosso modo jusqu'au XIX^e S. Depuis, de nombreux concepts utilisés par la rhétorique ont été récupérés par la linguistique.

Quant à l'origine de la rhétorique, il faut remarquer que l'apparition de la rhétorique en Grèce au 5^e S avant J.C. est étroitement liée à la pratique des orateurs qui, dans les domaines juridique et politique, étaient devenus des virtuoses dans l'art d'organiser un discours et de convaincre les foules. Ainsi, la rhétorique est d'abord une pratique du langage. Aujourd'hui, on emploie encore l'adjectif rhétorique pour désigner un usage du langage qui répond à cet objectif utilitaire sans nécessairement refléter une quelconque sincérité.

Dès l'Antiquité, la pratique de la rhétorique s'est fortement théorisée (cf. la *Rhétorique* d'Aristote). A Rome également, sous l'influence de Cicéron, on continue de théoriser l'éloquence et de mettre en formes les conseils qui permettent d'accéder à la carrière d'orateur, carrière très prestigieuse. Toute l'époque moderne sera redevable de cet enseignement. Aujourd'hui, la rhétorique n'est plus enseignée en tant que telle et sa disparition, à la fin du 19^e S, est liée à l'apparition des premières disciplines techniques de la linguistique.

La rhétorique couvrait traditionnellement beaucoup de questions de langage. Lorsque l'enseignement de la rhétorique a disparu, la linguistique a continué d'approfondir chacune de ces questions séparément. On notera, ainsi, que de nombreux domaines de la linguistique actuelle ont une origine rhétorique, la pragmatique, par exemple ou la linguistique de l'énonciation. Les études portant sur l'implicite, la connotation, l'ironie, l'argumentation sont toutes d'inspiration rhétorique.

1.1.5 Qu'est-ce que la linguistique

Jusqu'ici nous avons dit tout ce que la linguistique n'était pas et il est question de se demander ce qu'elle est réellement. Si le terme linguistique date du début du 19^eS, on peut dire que c'est à la fin de ce siècle, après les grandes réalisations de la grammaire comparée, que la linguistique va chercher à se constituer en discipline scientifique, au moyen, notamment, d'un effort de théorisation et de conceptualisation des termes qu'elle utilise (cf. Chapitre 3 de ce cours).

1.2 La linguistique, discipline scientifique

1.2.1 Qu'entend par une discipline scientifique ?

Karl Popper (28/07/1902 – 17/09/1994) est un des philosophes des sciences les plus influents du XXe siècle. Il a écrit de nombreux ouvrages dans le champ de l'herméneutique (ou philosophie des sciences, qui est une branche de la logique). Il critique la théorie vérificationniste de la signification et met l'accent sur l'idée de la réfutabilité par l'expérimentation ou l'échange critique comme critère de démarcation entre *Science* et *Pseudosciences* (cf. son ouvrage *Logik der Forschung – La logique de la recherche*).

Pour Popper, toute science a pour but la construction d'un savoir théorique. Elle construit des théories générales, qui sont « *un ensemble cohérent de principes qui permettent de rendre compte des observations que l'on fait en rapport avec l'objet* ».

Cette vision des choses de Popper a fortement influencé la façon d'envisager une étude scientifique et après lui, on ne va plus jamais concevoir les théories scientifiques de la même manière. Une théorie qui ne définit pas en son sein ses conditions de falsification ne peut plus être considérée comme étant scientifique, même si les apparences semblent montrer le contraire. On doit toujours être en mesure de pouvoir falsifier toute théorie, et de la réfuter en prouvant cette fausseté par des faits objectifs.

La science n'est donc pas en quête de la Vérité absolue, d'un modèle qui expliquerait tout, mais est un ensemble de théories imaginées qui peuvent un jour être démenties.

1.2.2 Implication du terme Science (en linguistique)

Suivant les réflexions R.H. Robins (1973 : 21), « Dire que la linguistique est une science dans le sens plus strict, c'est dire qu'elle traite d'un objet spécifique (le langage parlé et écrit), et qu'elle utilise des procédures qui peuvent être communiquées et décrites d'une manière systématique et qui peuvent être justifiées par rapport à des principes énonçables et une théorie explicite. Son but, en l'occurrence, est d'analyser le matériau, et de rendre compte si possible en le reliant à des règles et des régularités, de la variété infinie des phénomènes (les énoncés oraux et écrits) qui sont de son ressort ».

1.2.3 Linguistique = discipline scientifique, depuis quand ?

Aux États-Unis, la linguistique a été reconnue comme une discipline scientifique à partir de 1940. Ailleurs, il faudra attendre les années 50. Cette reconnaissance est due à la conjonction de deux facteurs favorables :

- Des « **pères fondateurs** » qui ont défini des concepts et la base d'une méthodologie scientifique ;
- La reconnaissance des modèles élaborés par la linguistique par une application de ces modèles hors de son champ propre.

1.2.3.1 Les pères fondateurs

Vers cette époque, il y a eu quelques « pères fondateurs » de la linguistique qui ont donné les fondements de la linguistique d'un point de vue épistémologique. On peut donner ce rôle à des scientifiques du dernier XIX^e siècle : Whitney et Saussure.

William Dwight **Whitney** (en Amérique), pour avoir été le premier à rompre avec les concepts métaphysiques de la grammaire comparée.

Ferdinand de Saussure (en Europe) dont le cours de linguistique générale publié dans les années 20 (mais dont les idées n'ont fait boue de neige qu'à partir des années 30) offre les fondements de la linguistique moderne. Ses propositions de modèles ne seront approfondies que plus tard par d'autres, notamment :

- **N. Troubetzkoy** : le modèle phonologique (encore d'actualité).
- **E. Sapir** : la conception du langage en tant que produit culturel.
- **L. Bloomfield** : le premier modèle formalisant des énoncés.

Ces deux pères fondateurs ont défini (i) les critères définissant l'objet de la linguistique et par là-même sa validité. Ils ont également (ii) élaboré des concepts de base de la linguistique et (iii) mis en place une méthodologie efficace. Dans l'étude des langues et du langage, on ne passait plus par exemple par des reconstructions bancales (boiteuses, approximatives) de l'indo-européen, mais par des méthodes efficaces d'analyse de la langue comme objet empiriquement observable, qui permettront à la linguistique d'être reconnue comme science.

Mais la linguistique ne faisait alors que commencer. Avant 1950, *tout* le travail des linguistes fut un travail de construction théorique. Il s'agissait de démontrer quelle était la spécificité de la linguistique par rapport aux autres sciences qui prenaient le langage comme objet d'étude (acoustique, sémiologie, psychologie,...).

Les linguistes en sont progressivement venus à considérer le langage comme un instrument de communication, de manière relativement unanime.

Dès 1950, on peut dire que la linguistique dispose d'un outillage suffisamment unifié, dont le but principal et premier est la **description du fonctionnement du langage pour lui-même**, en tant qu'objet, et en tant que **faculté humaine permettant la communication**.

1.2.3.2 La reconnaissance des modèles

Plusieurs disciplines des sciences humaines ont emprunté à la linguistique des modèles qui tenaient la route. Dès le moment où une autre discipline se les est accaparés, et que s'est développée ce qu'on peut appeler la « linguistique appliquée », on a pu considérer la linguistique comme une science aux fondations bien assises.

➤ *En linguistique appliquée à l'apprentissage des langues*

Le premier domaine dans lequel la linguistique a été appliquée est l'enseignement des langues étrangères, particulièrement dans des aspects psycholinguistiques. Deux noms ont été très influents dans ce cadre : Robert Lado et Charles C.Fries.

➤ *Dans le traitement automatique du langage*

Un autre élément qui a joué en faveur d'une reconnaissance scientifique de la linguistique est son application pour le traitement automatique des langues (un exemple très actuel en est [Google translation®](#)). Les premiers balbutiements de la linguistique pure y ont pu être appliqués, et cela dès que l'apport de l'ordinateur a été possible.

➤ *Autres domaines d'application*

D'autres domaines d'application de la linguistique ont fini par se couper de la linguistique pure, et à vivre leur vie propre. A titre d'exemple, il importe de remarquer que les personnes qui

s'occupent de la linguistique appliquée (autrement 'didactique des langues') notamment dans l'enseignement des langues se coupent des enseignements de la linguistique et essaient de construire des modèles différents dans un domaine qui ne peut plus s'en défaire, tant la linguistique théorique a élaboré des modèles fondateurs.

Depuis 1950, on constate également une évolution de la linguistique, la **sociolinguistique**, menant sa vie propre et se détachant parfois des enseignements de la linguistique formelle. Ici, les modèles linguistiques sont mis de côté en connaissance de cause, à tort ou à raison, sous le prétexte qu'il s'agit ici d'étudier le langage en tant que comportement social et en tant que marque distinctive de l'être humain parmi les autres comportements sociaux

1.3 L'éclatement de la linguistique

Depuis 1955-1957, on constate un éclatement de l'approche unifiante. On assiste d'une part à une « nationalisation » de la linguistique : les linguistes français se referment sur eux-mêmes, tandis que les américains cultivent un certain isolationnisme scientifique dans le domaine. 1955 et 1957 sont également deux dates de « grand chambardement » pour la linguistique :

- En 1955, John Austin publie *How to do things with words* qui constitue l'acte de naissance de la pragmatique et de remise en cause de certains concepts du formalisme linguistique ;
- En 1957, Noam Chomsky publie un ouvrage transcendant : *Syntactic structures*, allant dans un sens très différent d'Austin.

Les deux ouvrages montrent l'image d'une linguistique où prolifèrent un nombre incroyable de théories, et sans terminologie fixe. Chacun y va de sa conception, critiquant à outrance ses prédécesseurs.

En effet, si la linguistique s'est autant développée au fil du 20^e siècle, c'est essentiellement en s'appuyant sur une ambition théorique nouvelle par rapport au 19^e siècle. Par la suite, elle s'est ramifiée en de nombreuses écoles, et de nombreux domaines nouveaux sont apparus. Mais, dans chacun de ces écoles et dans chacun de ces domaines, se retrouvent tout de même quelques constantes : le désir de créer des concepts, le souci d'appliquer à l'étude du langage des méthodes scientifiques et, surtout, la volonté de s'écarter absolument de toute idée de norme comme de tout parti esthétique, moral ou évaluatif.

1.4 Question terminologique : langue vs langage

Dans toute une série de langues, le mot qui désigne le *muscle qui sert à la déglutition* et celui qui désigne l'*activité de la communication humaine orale* est le même :

- Français ('*langue*') ;
- Italien ('*lingua*') ;
- Kirundi ('*Ururimi*')

Dans d'autres langues, en revanche, il ne s'agit pas de deux acceptions, mais de deux mots bien distincts :

- Anglais : *tongue* vs *language*
- Néerlandais : *tong* vs *taal*

- Allemand : *Zunge* vs. *Sprache*
- Espagnol : *lengua* vs. *idioma*
- Latin : *sermo* [langage, parole, discours]
vs. *lingua* [muscle; idiome d'une communauté donnée]

1.4.1 Qu'est-ce qu'un langage ?

Différents angles d'approche sont possibles pour définir le « langage ». Les théories de la communication estiment qu'il y a langage chaque fois qu'il y a système de signes destinés à transmettre une information.

Sous cette acception, l'informatique et les mathématiques disposent également de « langages », que l'on rangera plutôt dans la catégorie des langages artificiels. Par ailleurs, d'autres langages existent : le langage des fleurs, le langage des signes, etc.

Les langues du monde correspondent également à cette définition d'outils de communication. Mais elles ont en plus la caractéristique d'être des langues « naturelles », en ce qu'elles se différencient des autres par une sémantique et une syntaxe. Aussi, peut-on que les langues naturelles peuvent être porteuses d'un sens *ajouté*, d'ordre contextuel : on peut donner un sens sémantique qui n'existe pas comme un sens établi dans la langue, c'est cela la signification pragmatique.

Chez les philosophes, le langage couvre un sens plus large : il s'agit de l'outil qui nous permet d'exprimer le monde et la pensée, en général considéré comme le propre de l'homme.

Dans le sens commun, on oppose deux sens principaux au mot « langage » :

- Le langage comme compétence à produire un langage donné.
- Le langage comme la faculté d'expression propre à l'humain.

Ces deux acceptions du langage ne sont pas neuves. On peut revenir à l'origine mythique que l'on donne du langage dans la plupart des religions, et dans certains textes philosophiques (cf. infra).

En tant que système de signes, on peut également dire que le langage a quelque chose d'intentionnel : il est énoncé par un sujet pensant, conscient du sens qu'il véhicule. Cet aspect permet de le différencier des caractéristiques typiques des indices. Les nuages noirs, par exemple, annoncent la pluie, mais n'ont pas l'*intention* de communiquer cela. En effet, même s'il est possible de lire dans la forme des nuages l'annonce d'une évolution météorologique pour les jours à venir, on ne pourra pas dire qu'il y a langage. Il n'y a pas de communication entre les nuages et le météorologue. Le langage verbal humain, en revanche, résultera toujours d'une intention de communication.

On peut faire un parallèle, selon cette définition excluant toute absence d'intentionnalité, avec le **langage animal**. De nombreux animaux, utilisent bien un type de langage qui est bien intentionnel, mais celui-ci est beaucoup plus limité que le langage humain. L'animal, par exemple, est incapable d'innover à l'intérieur de son propre langage, tandis que l'homme peut être très créatif. Par ailleurs, l'homme peut apprendre d'autres langues tandis qu'un animal sera incapable d'apprendre un autre langage animal, ou encore le langage humain.

1.4.2 Qu'est-ce que la langue ?

Dans le dictionnaire de l'Académie Française, la définition de langage et de langue est *grosso modo* la même :

- **LANGUE** : « idiome, termes et façon de parler dont se sert une nation »
- **LANGAGE** : « idiome, langue que parle une nation »¹

Dès Saussure cependant, on verra que la distinction est bien installée et est devenue à sa suite incontournable et unanimement reconnue même dans la disparité des terminologies linguistiques.

Langue vs. Langage. Dans son sens courant, on peut dire que la langue est un langage (particulier), commun à un groupe social, à une communauté linguistique. Le langage proprement dit est la faculté humaine de pouvoir acquérir une langue et d'en apprendre d'autres². La distinction langue/langage semble cependant une particularité française. Dans la linguistique anglo-saxonne, un seul mot *language* recouvre les deux notions.

Langue naturelle vs. langage artificiel. La langue naturelle s'oppose aux langages artificiels par le fait qu'elle possède une intentionnalité, qu'elle possède une syntaxe et une sémantique, et qu'elle est actualisable en contexte. Il vaut mieux utiliser le terme de langue naturelle que de langage naturel : on a ici affaire à un code donné. On définit les langues naturelles comme des systèmes de signes vocaux doublement articulés – unités distinctives (les phonèmes) et les unités significatives (les morphèmes) ; cela afin de les opposer à d'autres systèmes de communication humains (comme la musique) ou animaux (le langage des abeilles).

Langue vs. système indiciare. Toute langue est un code, un système symbolique arbitraire, permettant de transmettre des informations de manière intentionnelle.

Chaque courant de la linguistique développera ensuite sa définition de ce qui, pour lui, est le véritable critère définitoire de la langue. Nous verrons pour chaque théorie les différents éléments définissant la langue et/ou la parole : toute théorie se construit en effet via la façon dont elle aborde son objet d'étude.

1.5 Origine(s) des langues et diversité linguistique

1.5.1 Quid de l'origine du langage et des langues ?

La question posée dans ce chapitre est de savoir si nous possédons de naissance la capacité de langage ou si notre capacité de langage résulte d'un apprentissage. Cette question a été traitée par les philosophes dans le passé avant d'être abordée par les scientifiques.

Aujourd'hui encore, aucune donnée ne permettrait de dire comment la langue est apparue. Par ce fait, la cause suite à laquelle l'homme aurait commencé à parler est plus de l'ordre de l'opinion, même pour les scientifiques. Certains postulent que la langue aurait une

¹La coïncidence présupposée entre un langage et une nation est une idée largement répandue. Cette idée absolument fautive peut être vue comme risible, mais également dangereuse.

² On parle d'acquisition pour la langue maternelle, et d'apprentissage lorsqu'il s'agit de langues étrangères. De fait, il ne s'agit pas du même processus cognitif.

origine divine, d'autres qu'elle est le **fruit de l'évolution** (on retrouve ici le débat entre créationnisme et évolutionnisme). D'autres, enfin, pensent qu'elle aurait été **inventée** –plus ou moins consciemment– par l'homme.

1.5.1.1 Une origine divine ?

L'Éternel Dieu forma de la terre tous les animaux des champs et tous les oiseaux du ciel, et il les fit venir vers l'homme, pour voir comment il les appellerait, et afin que tout être vivant portât le nom que lui donnerait l'homme. Et l'homme donna des noms à tout le bétail, aux oiseaux du ciel et à tous les animaux des champs. GENÈSE II, 19-20

Dans ce texte de la Genèse, la langue est conçue comme un ensemble de noms donnant des « étiquettes » à la création. Suivant le texte, on pourrait considérer que c'est par le biais de Dieu que l'homme a créé le langage pour nommer toutes les choses de la terre. Dans la mythologie judéo-chrétienne, on ne retrouve donc pas vraiment l'idée d'une origine purement divine du langage.

Dans d'autres mythologies, d'une manière très différente de la mythologie judéo-chrétienne, on retrouve des histoires à peu près similaires pour expliquer l'origine du langage :

- **Égyptiens**: création par le dieu **Thot** (inventeur de l'écriture et du langage) ;
- **Babyloniens**: création par le dieu **Nabû** (dieu mésopotamien du savoir et de l'écriture) ;
- **Hindous**: création par déesse **Sarasvatī** (créditée de l'invention du sanskrit).

Ces théories postulant d'une origine divine sont mises à mal par les données scientifiques, et ne sont évidemment pas réfutables. On retrouve manifestement un lien entre cette origine divine et des propriétés magiques de la langue. On retrouve cela dans le sanskrit, et également dans la tradition judéo-chrétienne. La Genèse, ou l'Évangile de Jean parlent de Dieu comme du Verbe :

Au commencement était le Verbe et le Verbe était avec Dieu et le Verbe était Dieu. Il était au commencement avec Dieu. Tout fut par lui, et sans lui rien ne fut. JEAN I, 1-3

Dieu serait, en un sens, assimilé au langage, ou du moins à la faculté de langage. On croyait dès lors pouvoir détenir une sorte de pouvoir magique sur un objet ou une personne en en connaissant son « nom véritable ». On part du principe que Dieu a créé, voire *est* la faculté de langage. C'est dans cet ordre d'idée qu'a également été (un jour) inventée l'élucubration « abracadabra », formule divine créatrice du langage en lui-même³ pour inventer les propriétés magiques du langage.

Le mythe de la langue originelle

On a longtemps considéré que la diversité des langues parlées dans le monde résulterait de l'éclatement d'une langue qui serait à l'origine de toutes les autres (à laquelle on impute une origine divine, à l'instar du mythe de Babel). De nombreux pseudo-scientifiques se sont ainsi positionnés pour dire que la langue originelle était le **phrygien** (Psamtek I, VII^e s. av. J-C), l'**hébreu** (James IV d'Écosse, XV^e s.), l'**allemand** (Becanus, XVI^e s.), l'**araméen** (Webster

³En araméen, « *abra-ka-dabra* » signifie « il a créé comme il a parlé ».

XIX^e s.), le **chinois** (Elkins XIX^e s.),... D'autres, dans le même ordre d'idées, tenteront de reconstituer une langue aujourd'hui éteinte. On appelle cette théorie la théorie de la *monogénèse* : il y aurait une langue racine, à partir de laquelle les autres se seraient toutes diversifiées. Bien sûr, encore une fois, il est à ce jour impossible de vérifier ou de réfuter toute hypothèse de ce type.

1.5.1.2 Une invention humaine ?

Différentes théories partent de l'idée que la langue serait une invention humaine. L'homme aurait eu l'initiative de la création du langage :

- **Platon**, dans le *Cratyle*, affirme par la bouche de Socrate que « les choses périssables portent des noms de convention, dus au hasard, mais que les choses éternelles portent des noms naturels, donnés par les dieux. »⁴. Les noms donnés par les dieux **imiteraient** l'idée de l'objet qu'elles désignent, tandis que les noms « accidentels » seraient totalement **arbitraires**.
- **Jean-Jacques Rousseau**, dans « *Discours sur l'origine et les fondements de l'inégalité parmi les hommes* » avance l'idée qu'à l'origine, l'homme ne produisait que des sons liés à ses besoins essentiels (le « **cri de la nature** ») et que la vie en société aurait rendu ses besoins de plus en plus complexes, ce qui aurait abouti *in fine* à la naissance du langage. C'est également l'idée du linguiste russe Nikolai (1865-1934) qui estimait que le langage était une sorte de phénomène de classe, une invention sociale, fruit du contact répété et du travail en commun ; une position qui a été soutenue en 1950 par G. Révész dans *Origine préhistoire du langage*.
- Le danois **Otto Jespersen** (XIX^e-XX^e s.) estimait quant à lui que le langage résultait d'un **besoin expressif** plutôt que communicatif. Il estimait que le chant, plaisir du son mélodieux, avait précédé le langage (cf. la « *la-la* »). Otto Jespersen était convaincu d'une liaison entre le son et le sens.

1.5.1.3 Fruit de l'évolution ?

Une autre hypothèse plus raisonnable, serait de considérer que le langage est le fruit de l'évolution. Le langage se serait constitué à partir du moment où certains éléments nécessaires à son apparition auraient été attestés chez l'homme par la suite de mutations successives.

Si cette théorie peut paraître la plus raisonnée, elle n'en est pas plus scientifique que les autres. En effet, on manque de preuves directes qui permettraient de l'affirmer. Les premières traces effectives de langage remontent à 3500 av. J-C : ce sont les premières traces d'écriture. Mais on est presque certain que l'apparition du langage a dû être orale et précéder l'apparition de l'écriture de plusieurs milliers d'années.

On peut induire de certains faits l'apparition du langage chez l'homme. Ces faits sont de trois types :

- Les faits d'ordre **génétique**.

⁴<http://bbouillon.free.fr/univ/hl/Fichiers/Cours/langage3.htm>,

- Les faits d'ordre **anatomique** (appareil phonateur, oreille, cerveau).
- Les attestations d'une pensée **symbolique**.

1.5.1.3.1 *Anatomie et génétique*

On constate une différence essentielle entre les primates et l'homme au niveau de l'appareil phonateur. C'est à cause de cette différence que les primates ne sont pas capables de produire de sons suffisamment diversifiés pour pouvoir communiquer verbalement.

Le pharynx

La principale différence anatomique se situe au niveau du pharynx, beaucoup plus grand chez l'homme que chez les primates. C'est chez les premiers *Homo sapiens* que, pour la première fois dans le règne des primates, les conditions sont réunies pour produire suffisamment de sons, et permettre ainsi une communication orale. Plusieurs éléments peuvent nous permettre de postuler une capacité à produire la parole chez nos ancêtres lointains. On sait donc que nos ancêtres avaient la possibilité de parler, mais on ne sait pas s'ils l'ont effectivement fait.

Suite à une étude réalisée sur un homme de Cro-Magnon dont la mâchoire a été conservée relativement intacte, on a découvert que les premiers *Homo sapiens* n'étaient en effet sans doute pas capables de produire les sons [i], [u] et [a]. Or, ces sons apparaissent aujourd'hui dans toutes les langues de la terre, car ce sont les trois voyelles cardinales. Par conséquent, l'*Homo sapiens*, notre ancêtre, ne disposait pas de toutes les possibilités articulatoires dont nous jouissons en tant qu'*Homo « sapiens sapiens »*, et s'il a pu produire une parole articulée, elle devait être sans rapport avec les langues que nous parlons aujourd'hui.

La cavité laryngée

Un autre élément confirmant l'apparition de la faculté de langage chez l'*Homo sapiens* est l'apparition de la cavité laryngée. Tout homme capable de parole a en effet la capacité de produire des sons laryngaux.

Le cerveau

i) Taille du cerveau

Un autre élément qui a été nécessaire à l'apparition du langage est un volume cérébral suffisant. L'australopithèque disposait d'un cerveau dont le volume était de 410 cm³. Chez l'*Homo sapiens*, on en arrive à des proportions bien plus imposantes : 1700 cm³ pour l'homme de Neandertal, et 1400cm³ pour l'*Homo sapiens sapiens*. On peut partir du principe que si le cerveau a augmenté en taille, il a en principe augmenté sa faculté cognitive et donc sa faculté de produire un langage.

ii) Encéphalisation

Mais la taille n'est finalement pas tout à fait révélatrice de cette capacité cognitive. Une

meilleure mesure pourrait en être le **degré d'encéphalisation**. Si le cerveau humain est grand, certains mammifères font en effet mieux que lui (l'éléphant, la baleine,...). Mais là où le cerveau humain demeure imbattable, c'est dans la croissance qu'il poursuit après la naissance. Une croissance du cerveau à rythme fœtal se poursuit en effet encore durant les premiers mois, grâce à la *frontanelle*, partie du cerveau qui se développe progressivement et permet un apprentissage plus tardif et très différent des autres mammifères.

iii) *Latéralisation*

Un troisième élément propice à l'apparition du langage est la latéralisation de la structure cérébrale. On constate que le cerveau est « divisible » :

- en un hémisphère gauche et hémisphère droit,
- en aire de Broca et en aire de Wernicke

Cela va permettre au cerveau de pouvoir optimiser son fonctionnement. Toutefois, on sait que la latéralisation existait déjà chez le chimpanzé et certains oiseaux. Cette latéralisation ne correspond en fait pas directement à la fonction du langage, mais c'est plus que certainement un élément qui a joué en la faveur de son apparition. On n'en est pas convaincu, et ce n'est certainement qu'une condition suffisante, et non nécessaire, à l'apparition du langage.

1.5.1.3.2 *Apparition de la pensée symbolique*

L'anatomie et la génétique sont une catégorie de faits scientifiques qui nous permettent d'induire la *possibilité* de l'apparition du langage. Un autre élément que l'on considère comme ayant pu intervenir, en amont ou en aval des premières formes de langage, est l'apparition d'une pensée symbolique. Certains animaux possèdent une capacité symbolique (oiseaux, abeilles,...), et on peut y entraîner les chimpanzés. Mais chez l'homme cette capacité peut également intervenir dans des domaines où sa survie n'est pas en jeu.

Une première preuve de l'apparition de cette pensée symbolique se retrouve dans l'art pariétal (du latin *parietalis* « relatif au mur, paroi », l'art pariétal désigne l'ensemble des œuvres d'art au sens large réalisées par l'homme sur des parois des grottes). Cet élément a pu jouer un rôle en ce que cet art est une représentation de la réalité au moyen d'images, et donc l'expression d'un symbolisme. On peut dire, pour simplifier, que dès que l'homme est capable de « dessiner », il est capable d'avoir l'idée conceptualisée de certains éléments du réel, et de les représenter, en leur absence : il génère des symboles. Dès lors, le langage n'est donc plus très loin : du point de vue du cerveau de l'homme, on peut faire le rapprochement entre ce type d'art pariétal et l'utilisation de sons pour générer des symboles du réel.

1.6 La diversité linguistique

1.6.1 Monogenèse et polygenèse

Deux théories différentes et contradictoires prétendent expliquer la diversité des langues parlées dans le monde : une monogenèse ou une polygenèse des langues.

- Monogénèse. Le langage humain s'est développé une seule fois à un endroit précis avant les premières migrations humaines et s'est répandu et diversifié grâce à ces migrations.
- Polygénèse. Le langage humain s'est développé plusieurs fois à des endroits différents après ou pendant les grandes migrations.

Actuellement, on privilégie la thèse de la monogénèse, en partant du principe que le langage se serait développé une fois, à partir de la région du Kenya, avant les grandes migrations et les premières dispersions des différents peuples. On part du principe que la monogénèse *tient mieux la route*, par rapport aux modèles explicatifs de la préhistoire dans d'autres sciences.

Un récit comme celui de la tour de Babel soutient également une thèse de la monogénèse, où Dieu serait venu confondre le langage des hommes :

Tout le monde se servait d'une même langue et des mêmes mots. Comme les hommes se déplaçaient à l'Orient, ils trouvèrent une vallée au pays de Shinéar et ils s'y établirent. Ils se dirent l'un à l'autre : « Allons ! Faisons des briques et cuisons-les au feu ! » La brique leur servit de pierre et le bitume leur servit de mortier. Ils dirent : « Allons ! Bâtissons-nous une ville et une tour dont le sommet pénètre les cieux ! Faisons-nous un nom et ne soyons pas dispersés sur toute la terre ! »

Dieu descendit pour voir la ville et la tour que les hommes avaient bâties. Et Dieu dit : « Voici que tous font un seul peuple et parlent une seule langue, et tel est le début de leurs entreprises ! Maintenant, aucun dessein ne sera irréalisable pour eux. Allons ! Descendons ! Et là, confondons leur langage pour qu'ils ne s'entendent plus les uns les autres. » Dieu les dispersa de là sur toute la face de la terre et ils cessèrent de bâtir la ville. Aussi la nomma-t-on Babel, car c'est là que Dieu confondit le langage de tous les habitants de la terre et c'est là qu'Il les dispersa sur toute la face de la terre.

GENÈSE XI, 1-9

On pourrait cependant avancer des arguments allant précisément à l'encontre de cette théorie de la monogénèse. En effet, plus il y a de langues, plus la thèse de la polygénèse deviendrait possible.

Pour qu'il y ait monogénèse, il faudrait que toutes les langues soient les descendantes de cette langue primitive ; mais vu la diversité des langues, cet élément pourrait être mis en doute, et rien ne montre que les universaux linguistiques sont les « résidus » d'une évolution : ils pourraient tout aussi bien n'être que l'expression de contraintes intrinsèques au langage. On pourrait également dire pour défendre la théorie d'une polygénèse qu'un stade « pré-linguistique » très élémentaire aurait tout d'abord été traversé, où le langage aurait été lié aux besoins directs de l'homme, puis se serait développée en des foyers multiples et de façons différentes.

1.6.2 Les langues du monde

1.6.2.1 Le recensement

Plus il y a de langues, plus la thèse d'une polygenèse devient plausible.

Les sources divergent quant au nombre de langues parlées dans le monde :

- **1929** : Académie française: **2 796**
- **2000** : Summer Institute of Linguistics (centre de recherche d'Ethnologue.com) : **6 809**

Area	Living languages		Number of speakers	
	Count	%	Count	%
Africa	2092	30.3	675 887 158	11.8
Americas	1002	14.5	47 559 381	0.8
Asia	2269	32.8	3 489 897 147	61.0
Europe	239	3.5	1 504 393 183	26.3
Pacific	1310	19	6 124 341	0.1
Total	6912	100	5 723 861 210	100.0

Les thèses les plus pessimistes déclarent que des 6000 langues parlées dans le monde, il n'en restera plus en 2100 qu'environ 3000.

Si on fait le décompte des langues vivantes et mortes en Europe (par exemple), on constate qu'il y a environ 275 langues usitées. On peut constater ce nombre en regardant la diversité linguistique présente en France, malgré l'apparente uniformité que celle-ci affiche. Toutefois, la grande majorité des langues d'Europe sont parlées en ex-URSS. Et contrairement à ce que l'image de la France peut montrer (celle d'une diversité linguistique), l'Europe n'est pas vraiment « riche » en termes de diversité linguistique : elle ne totalise que 3,5% des langues du monde (contre 15% en Amérique par exemple, et 30% en Afrique, 30% en Asie, et 20% dans le pacifique).

Malgré ce faible pourcentage, l'Europe compte 5 des 10 premières langues parlées dans le monde. Les 10 langues les plus parlées dans le monde sont aujourd'hui la langue maternelle d'environ la moitié de la population du globe, mais elles ne représentent que 0,10 à 0,15% des langues parlées dans le monde. Parmi ces 10 langues, 5 d'entre elles sont parlées en Europe, mais la majorité des locuteurs de l'espagnol, de l'anglais, du français ou du portugais vivent dans d'autres régions du monde.

À côté de ces « super-langues » parlées par des centaines de millions de personnes, plus de 95% des langues du monde ont moins d'un million de locuteurs. Le nombre moyen de locuteurs d'une langue est probablement de 5000 à 6000, et 95% des langues parlées ont moins d'un million de locuteurs natifs. Il faut aussi remarquer que 83 à 84% des langues parlées dans le monde sont endémiques, i.e. on ne les parle que dans un seul pays.

1.6.2.2 Critères de différenciation et de classement des langues

Au niveau des critères de différenciation des langues, on constate que des critères politiques prennent le plus souvent le dessus par rapport à des critères purement linguistiques. Certains linguistes, pour différencier une langue d'un dialecte, vont même jusqu'à dire qu'« *une langue est un dialecte composé d'une armée et d'une marine.* » (Max Weinreich). En effet, de nombreuses langues ne sont définies comme telles que parce qu'elles sont validées par les autorités en place.

Ainsi, le serbo-croate est une langue séparée en deux dialectes, le serbe d'une part et le croate d'autre part, sur une distinction politico-ethnique et non véritablement linguistique. Le chinois, de la même manière, est le fait d'une écriture répandue dans un pays entier, mais les langues orales parlées en Chine sont en réalité très diversifiées et ne sont pas toutes intelligibles entre elles (cantonais, hakka, mandarin, wu,...)⁵ ; ces langues ont elles-mêmes leurs dialectes propres.

Deux chercheurs français (Alain Calvet et Louis-Jean Calvet 2022) établissent un classement basé sur plusieurs critères. Ils ont compilé une liste de 6014 langues, chacune avec un code standardisé et univoque. Mais, toutes ces langues ne sont pas comparables entre elles. Si l'on retient les langues parlées par plus de 300.000 personnes dans des pays ayant un certain niveau de développement, de fécondité et d'accès à internet, on aboutit à une liste de 634 langues comparables entre elles. Un certain nombre de critères se sont révélés pertinents :

- 1) Le nombre de locuteurs ;
- 2) La diffusion d'une langue et le nombre de pays où elle est parlée. Par exemple, le chinois mandarin est parlé par 921 millions de personnes et peu diffusé, alors que l'anglais est parlé par 362 millions de personnes et extrêmement diffusé.
- 3) Le statut politique d'une langue lui donne aussi du poids plus ou moins important, surtout lorsqu'elle est une langue officielle ;
- 4) Des facteurs culturels tels que :
 - 5.1. Le nombre de traductions à partir d'une langue ou vers celle-ci ;
 - 5.2. Le nombre de prix littéraires internationaux,
 - 5.3. Le nombre de pages publiées sur Wikipédia ;
 - 5.4. L'enseignement d'une langue dans les universités
 - 5.5. L'alphabet utilisé pour écrire une langue peut aussi lui donner du poids ;
- 5) Le poids du (ou des) pays où une langue est pratiquée influence sa valeur.
- 6) L'indice de développement (qui tient compte du PIB, du niveau d'éducation et de l'espérance de vie) ;
- 7) l'indice de fécondité ;
- 8) le nombre d'utilisateurs d'internet pèsent lourd dans le classement.

Sur la base de tous ces critères, on peut classer les 634 langues retenues. Tous critères confondus, le top 5 est : anglais, français, espagnol, allemand et russe. Le mandarin (11^e) arrive le japonais (8^e), à cause des critères de développement. L'ourdou (33^e) et l'hindi (34^e) se classent assez bas, malgré un nombre élevé de locuteurs. Si l'on tient compte uniquement

⁵ Voir : http://fr.wikipedia.org/wiki/Langues_chinoises.

du nombre de locuteurs, le top 5 est : mandarin, hindi, espagnol, anglais, portugais.

Le top 5 et le top 50 se modifient selon le poids qu'on donne à chacun des critères. Si le critère démographique prime, alors on trouve aux 5 premières places : anglais, ourdou, français, yué (langue chinoise parlée à Canton, Hong Kong ou Macao) et tagalog (Philippines).

Ce critère démographique décline les langues de la vieille Europe. Il place 24 langues africaines et 13 langues asiatiques dans le top 50. Au contraire, le critère de prestige permet de surclasser les langues à forte reconnaissance et diffusion culturelles, qui sont enseignées dans les universités et pratiquées dans des pays développés où elles jouissent d'un statut officiel. Avec ce critère, anglais, français, allemand, espagnol et russe restent dans le top 5. De sorte que 30 langues européennes figurent dans le top 50 (lire la chronique *C'est du français*, Le Soir, 2-3/12/2023).

1.6.2.3 Cohabitation des systèmes linguistiques

Le terme de **diglossie** fait référence à cette situation où, dans un même état, deux systèmes linguistiques différents cohabitent et que l'un de ces deux systèmes occupe, souvent pour des raisons historiques, un **statut sociopolitique inférieur** ; cette situation pouvant très souvent à l'origine des conflits sociopolitiques. On parle de **bilinguisme (politique)** lorsque cette situation conflictuelle entre deux systèmes linguistiques a pu être résolue, et que les deux langues vivent statutairement **sur un pied d'égalité**.

1.6.3 Classification des langues

Généralement, on considère qu'il y a deux grandes approches possibles :

- La classification typologique. Les langues sont regroupées selon des **caractéristiques** structurelles qu'elles partagent ; c'est le point de vue **synchronique** qui est privilégié.
- La classification généalogique. Les langues sont regroupées sur principe de leurs **liens de parenté** (qu'elles aient un ancêtre attesté ou non) ; c'est le point de vue **diachronique** qui est mis en avant.

1.6.3.1 Classification typologique

La classification typologique des langues peut se faire sur un grand nombre de critères :

- Phonologie : analyser les systèmes phonologiques globaux, le système vocalique et consonantique (inventaire des phonèmes, voyelles longues, voyelles nasales, etc.), prosodique (différencier les langues à isochronie accentuelle et les langues à isochronie syllabique, différencier les langues à accent des langues à tons, etc.).
- Morphologie : analyser les stratégies de combinaison des morphèmes, étudier la profondeur du système flexionnel et dérivationnel, etc.
- Syntaxe : c'est ici l'ordre des mots qu'on cherche à analyser, particulièrement dans la phrase déclarative simple.
- Etc.

1.6.3.1.1 *Typologie phonologique*

Inventaire des phonèmes

On peut dresser un inventaire vocalique et consonantique des différentes langues du monde. On constate qu'il y a maximum 1000 phonèmes qui existent dans toutes les langues réunies. Chaque langue a une moyenne de 32 phonèmes, et la majorité d'entre elles en a entre 26 et 40. Mais il existe certains extrêmes ; ainsi, l'hawaïen ne comporte que 12 phonèmes, tandis que le !kung en comporte non moins de 141.

Nombre de voyelles

On constate que le système vocalique de la majorité des langues du monde se compose de en général de 3 à 9 voyelles, 5 étant le nombre moyen (nombre qui se retrouve dans 50% des langues, dont l'espagnol).

i	u	i	i	u
		e	o	e
				ə
	a	a		ʌ
				a
Three-vowel system Gudanji (Australia)		Four-vowel system Navajo (Arizona)		Seven-vowel system Ge'ez (Ethiopia)

Voyelles longues

Il est possible de différencier les langues sur un critère phonologique en prenant encore en considération l'existence ou non de voyelles longues. On peut constater, dans ces langues à voyelles longues (45% des langues du monde) que les brèves sont plus souvent internes que les longues (bien qu'on y ait jamais trouvé d'explication).

Nasales

Sauf dans des cas *excessivement* rares, le nombre de voyelles nasales est, sensiblement inférieur au nombre de voyelles orales, dans toutes les langues du monde: en effet, la confusion entre les nasales est potentiellement plus importantes (d'où le problème du [œ̃] précité en français). Un très grand nombre de langues ne contiennent pas du tout de voyelles nasales.

Nombre de consonnes

Le spectre de variation dans les langues du monde s'étend de 8 à plus de 90 consonnes, avec une moyenne arithmétique de 22 consonnes. L'écart de la moyenne varie plus que pour les voyelles.

1.6.3.1.2 Typologie morphologique

Sur un critère morphologique, on peut différencier les langues selon les rapports qu'elles privilégient entre les différents morphèmes. On différencie de cette manière deux extrêmes :

- **Les langues isolantes**, dans lesquelles les morphèmes sont séparés les uns des autres. Dans le chinois mandarin, chaque mot constitue un seul morphème. Il n'existe pas d'affixes pour marquer le temps et le nombre : tout sens se prend par des mots différents de la phrase.

Tā chī fàn le (il {manger} viande {passé} = il a mangé de la viande)

Tā chī le fàn (il {manger} {passé} viande = il mange de la vieille viande)

- **Les langues polysynthétiques**, dans lesquelles les mots sont constitués de racines et d'affixes. Dans les cas les plus significatifs, une expression signifiée par un seul mot composé équivaldrait ainsi à une phrase entière dans d'autres langues.

Qasuiirsarvigssarsingitluinarnarpuq

Qasuiir sar vigssar si ngitluinarnarpuq

(Quelqu'un n'a pas trouvé un endroit totalement propice au repos)

Peu de langues sont totalement isolantes ou totalement polysynthétiques. Entre ces deux extrêmes, on trouve deux types de langues répandus :

- **Les langues agglutinantes** : les mots peuvent contenir plusieurs morphèmes, mais on peut en identifier relativement facilement les composantes. Une autre caractéristique des langues agglutinantes est que chaque affixe correspond à une seule catégorie grammaticale.

En turc : köy köy-ler köy-ler-in
 village village-pl village-pl-génitif

- **Langues flexionnelles** : dans ces langues (d'entre lesquelles deux exemples marquants sont le latin et le russe) les mots peuvent être constitués de plusieurs morphèmes comme pour les langues agglutinantes, mais les affixes marquent ici non pas une, mais souvent plusieurs catégories grammaticales simultanément. Un affixe renferme donc une importante quantité d'information.

Pater	libros	fratri	meo	praebuit
nom	acc		dat	parf
sing	pl		sing	3p sing

Un grand nombre de langues constituent un enchevêtrement complexe de ces différentes stratégies. Dans l'anglais, on peut voir certaines stratégies typiques d'une langue analytique, mais également de nombreuses flexions et agglutinations :

I will leave (isolant)
 Reenactment (agglutinant)
 Him (flexionnel)

On s'est donc vite aperçu que la plupart des langues n'étaient pas « blanches ou noires », mais qu'elles correspondaient souvent à des catégories mixtes, ou que du moins, il était nécessaire de nuancer leur classement dans l'une ou l'autre catégorie. Certains ont alors suggéré que les termes d'isolante, agglutinante ou flexionnelle ne devraient plus être utilisés pour catégoriser une langue, mais pour qualifier certaines structures au sein d'une langue donnée. Winfred Lehmann (1995) montre, à travers son cycle de grammaticalisation, comment les différentes stratégies morphologiques des langues sont nécessairement instables.

1.6.3.1.3 1.6.3.1.3. Typologie syntaxique

➤ *Ordre des principaux constituants*

Joseph H. Greenberg (1963) est parvenu à dégager plusieurs types d'ordres syntaxiques dans les différentes langues du monde, en fonction de la place des constituants principaux dans leurs « phrases canoniques ». Il distingue trois types de constituants :

- O est le groupe objet-complément du verbe.
- S est le groupe du sujet ou de l'agent.
- V est le groupe verbal ou le pivot de la phrase.

Réalisant une typologie sur base de ces critères, il s'aperçoit que les trois schémas SOV, SVO et VSO regroupent 90% des langues du monde :

- *Le type VSO, unidirectionnel, représente 15% des langues connues: sémitiques et celtiques entre autres.*
- *Le type SOV, également unidirectionnel mais de sens inverse, représente 39% des langues du globe telles le turc, le japonais, l'hindi et de nombreuses langues amérindiennes et océaniques.*
- *Le type OSV n'est attesté que dans 10% des langues, de même que les types OVS et VOS (cas du malgache, des langues de Polynésie et de Mélanésie notamment).*
- *Le type SVO représente lui 36% des langues connues (romanes, slaves, môn-khmères entre autres).⁶*

Le français est par exemple une langue purement SVO ; les langues germaniques sont *grosso modo* des langues SVO, parfois SOV. Ces langues privilégiant la relation sujet-verbe tendent à employer de nombreuses prépositions dont le sens équivaut à la fonction syntaxique qu'elles remplissent (introduction de complément indirect, ...). En outre, de nombreuses langues SVO (le néerlandais et l'allemand par exemple) « changent de sens » dans les propositions subordonnées.

Les langues SOV ont tendance à utiliser des modificateurs avant ce qui est modifié, et utilisent souvent des postpositions. Là où les langues SVO utilisent des prépositions, c'est dans

⁶ Source : Wikipédia, http://fr.wikipedia.org/wiki/Typologie_linguistique.

ce type de langues que l'on retrouve le plus de marques casuelles (en partant du nominatif et de l'accusatif, nécessaires pour différencier le sujet de l'objet, contigus dans l'ordre SVO).

Latin (SOV): *Pater librum fratri meo praebet.* → **Emploi privilégié de formes casuelles**

Français (SVO): *Mon père prête un livre à mon frère.* → **Emploi privilégié de prépositions**

➤ ***Langues à satellites vs Langues à cadrage verbal***

On peut également distinguer, en syntaxe, les **langues à cadrage verbal** (verb-framed) et les **langues à satellites** (satellite-framed).

Si on prend par exemple la phrase « Marie traversa la rivière à la nage », l'élément « à la nage » n'est pas incorporé au verbe mais est un syntagme prépositionnel, contenant un nom, une idée qui pourrait être également un sujet. Ici ce n'est pas le cas car il s'agit d'exprimer un mode (moyen ou manière) de faire l'action de traverser. Par contre, le déplacement (« path ») est exprimé par le verbe (« traversa »). La phrase correspondante en anglais sera « Mary swam across the river ». La manière (ou « mode ») n'est plus exprimée par un syntagme prépositionnel mais par le verbe en lui-même. C'est la préposition, en revanche, qui exprime le *path* (« accros »), à l'exact inverse du français.

Cette différence de configuration a diverses implications : dans ce genre de phrase, la manière doit en anglais être obligatoirement incorporée, alors qu'on l'omet souvent en français.

<i>Langues à satellites</i> <i>(satellite-framed)</i> (= anglais, langues germaniques en général, russe, mandarin, serbe, ...)	<i>Langues à cadrage verbal</i> <i>(verb-framed)</i> (= français, espagnol, hébreu, turc, ...)
<ul style="list-style-type: none"> • <i>Le verbe exprime la manière / la cause.</i> • <i>Le mouvement est exprimé par le « satellite » (préposition, particule, adverbe).</i> 	<ul style="list-style-type: none"> • <i>Le mouvement est exprimé par le verbe.</i> • <i>La manière n'est pas incorporée par le verbe mais sera exprimée de façon « périphérique » (syntagme prépositionnel, gérondif, ...).</i> • <i>La manière est souvent omise.</i>
Mary swam across the river. The car drove through the tunnel. The bird flew into the room.	Marie traversa la rivière (à la nage). La voiture traversa le tunnel. L'oiseau entra dans la pièce (en volant).

1.6.3.1.4. Typologie et universaux

Cette approche typologique a permis de dégager les différents universaux du langage qui peuvent coexister. Chomsky distingue deux types d'universaux : les universaux de substance, liés à des contraintes matérielles, et les universaux formels, plus intéressants car ils disent quelque chose du fonctionnement cognitif du langage, et de la grammaire universelle. Greenberg, quant à lui, distingue les universaux absolus, qui ne souffrent d'aucune exception, et les universaux statistiques, qui sont ceux qui souffrent de très peu d'exceptions (par exemple, le fait que, si une langue possède des fricatives, elle possèdera aussi des occlusives). Parmi les principaux universaux, on peut noter notamment :

- **Phonologie.** Si une langue possède des occlusives sonores, elle possèdera aussi des occlusives sourdes. Si une langue possède des fricatives, elle possèdera aussi des occlusives.
- **Morphologie.** Si une langue possède des affixes flexionnels, elle possèdera aussi des affixes dérivationnels. Si une langue possède des affixes dérivationnels et flexionnels, l'affixe dérivationnel est le plus proche de la racine.
- **Syntaxe.** La proposition principale précède toujours le verbe dans les langues OV, mais le suit généralement dans les langues VO.

1.6.3.2 Classification génétique des langues

La classification génétique, contrairement à la classification typologique, suggère l'existence d'un ancêtre⁷ commun. Pour trouver cet ancêtre commun, elle ne s'intéresse pas principalement à la forme structurelle des langues (phonologie, morphologie, syntaxe), comme le fait la classification typologique : elle se base davantage sur des parentés phonétiques et sémantiques entre certaines unités lexicales appartenant à des langues différentes.

La classification génétique des langues fut l'objet d'étude principal de la grammaire comparée. Il s'agissait d'établir une méthodologie rigoureuse (Critères ? Quantité de données à récolter ?), afin de déceler les liens de parenté entre les langues. Lors de la classification typologique, nous pouvons constater que l'anglais et le russe n'ont presque rien en commun. En revanche, d'un point de vue génétique, une étude sémantique et lexicale permet de dégager un lien de parenté entre ces deux langues, qui sont toutes deux issues de la souche indo-européenne. Toutefois, il faut toujours rester prudent en induisant un tel lien : il n'existe en effet plus de traces matérielles, ni de preuves directes de l'indo-européen.

On distingue grosso modo entre 15 et 20 macro-familles de langues. Cette classification est cependant très controversée, notamment parce qu'elle contient des « inclassables » comme le basque par exemple. En effet, le basque a la structure très différente des autres langues européennes, alors qu'elle n'est parlée que dans une région enclavée au cœur de l'Europe. Il faut noter que même dans une Europe presque entièrement indo-européenne, quelques langues font malgré cavalier seul : l'estonien, le hongrois et le finnois.

⁷ Lors de l'emploi de tels termes, il faut garder à l'esprit qu'une expression telle que « lien de parenté » ou encore « ancêtre commun » reste une métaphore conceptuelle. Utile pour la représentation mentale, elle ne doit pas être prise pour argent comptant, car poussée dans ses retranchements elle inviterait à des interprétations erronées de ce qu'est en vérité une langue, et de la manière dont elle évolue.

Europe : indo-européen, altaïque, ouralien (ou finno-ougrien).

Afrique : afro-asiatique (chamito-sémitique), nilosaharien, nigéro-kordofanien, khoisan.

Asie : dravidien, austro-asiatique, austronésien, sino-tibétain (auxquelles s'ajoutent les familles partagées par l'Europe et l'Afrique).

Océanie : australien.

Amérique : eskimo-aléoute, amérindien

D'autres chercheurs ont tenté des classifications plus poussées au sein des différents groupes linguistiques. Les langues « filles » de l'indo-européen ont ainsi été catégorisées en différentes classes et sous-classes⁸.

⁸Image tirée de : <http://lewebpedagogique.com/blocagesenfrançais/etymologie/>

2 LES APPROCHES PRETHEORIQUES DES LANGUES ET DU LANGAGE

2.1 L'Antiquité

L'Antiquité est la première époque où l'on trouvera des théories du langage. En effet, dans de nombreuses civilisations archaïques, le langage humain faisait l'objet de mythes fondateurs. A un dieu particulier était dévolue par exemple la mission de créer le langage. En Grèce, la philosophie a eu ceci de particulier qu'elle a cherché à interpréter les mythes et à formuler certaines questions en termes abstraits. Au cours de l'antiquité ont été écrites notamment une série de grammaires, de glossaires et de traités de rhétorique, à finalité uniquement pratique, qui ne faisaient pas de réflexion sur le langage en tant que tel. Ces ouvrages étaient principalement normatifs et prescriptivistes, souvent doublées d'une prétention esthétique.

2.1.1. L'Inde

C'est pour des raisons religieuses que les anciens Indiens ont tout d'abord été amenés à étudier leur langue. Il importait que les textes sacrés, réunis dans le Veda, ne subissent aucune altération au moment où ils devaient être chantés ou récités pendant les sacrifices. D'où l'effort fait pour les conserver dans leur pureté première. Dans la suite, les grammairiens indiens – dont le plus célèbre est Pāṇini (vers le IV^e S av. J.-C.) firent porter leurs observations sur les valeurs et l'emploi des mots et dressèrent de leur langue, avec une précision et une minutie admirables, des descriptions phonétiques et grammaticales qui sont des modèles du genre. Longtemps oubliées, elles furent découvertes par les savants occidentaux à la fin du XVIII^e siècle et fournirent le point de départ indispensable à la création de la grammaire comparée. Mais ces études étaient statiques portant sur le seul sanskrit et étaient réalisées par des hommes qui, conformément au génie propre de l'Inde, manquaient totalement de sens historique, elles se bornaient à classer les faits sans en chercher l'explication.

2.1.2. La Grèce et la philosophie du langage

2.1.2.1. Les mots et les choses

Les philosophes les plus archaïques de la Grèce considéraient que le langage n'était qu'une autre expression du cosmos, autrement dit, du monde dans son sens le plus large. Le point de vue philosophique (adéquation du langage à la pensée) pris par ces philosophes s'avère intéressant puisque leurs spéculations sont pour une bonne part à l'origine de la pensée linguistique moderne dans ses errements comme dans les voies de sa réussite. Le problème essentiel qui se posait à des philosophes soucieux d'élaborer une théorie de la connaissance consistait à définir les rapports entre la notion et le mot qui la désigne. La grande question qui était débattue chez les sophistes et les philosophes antiques (et qui restera d'actualité jusqu'à la scholastique médiévale) était de savoir si le langage a été créé par la nature ou à la suite d'une convention ; en d'autres termes s'il y a un rapport nécessaire entre les mots et leur signification, entre le signifiant et le signifié.

Dans le dialogue de Platon, le *Cratyle*, deux personnages discutent de la nature du langage. *Cratyle*, disciple d'Héraclite, (540-480 av. J.C) estime que la substance profonde des choses se retrouve dans l'aspect des mots qui les représentent. Les mots aident à connaître les choses. Hermogène, au contraire, développe l'idée du caractère conventionnel, arbitraire, du langage. Platon fait dialoguer deux théories et essaie de les concilier : le langage est une création humaine (et donc conventionnelle), mais il découle de l'essence des choses et n'est donc pas libre.

2.1.2.2. Le langage et la pensée

Les Grecs n'avaient qu'un terme pour désigner le langage et la raison : le terme *logos*. Cela explique que dans l'Antiquité, une grande théorie du langage se retrouve exprimée dans la discipline de la logique. La logique essaie de décrire rationnellement l'organisation de la pensée et la façon dont celle-ci est exprimée dans le langage. Les Latins, en traduisant **logos** par deux termes, *ratio* (qui est devenu « raison ») et *oratio* « oraison », ancien mot utilisé pour désigner à la fois la pensée et le langage, ont commencé à dissocier les deux perspectives. Cependant, dans l'Antiquité, le langage est très souvent pensé en liaison étroite avec la pensée.

2.1.3. La grammaire en Grèce et à Rome

Une grande partie de la réflexion sur le langage et de la recherche grammaticale qui aura lieu en Occident repose sur des concepts élaborés dans l'Antiquité grecque. Il est donc important de se remémorer dans quel esprit ils ont été forgés, de manière à comprendre comment a pu naître la linguistique.

2.1.3.1. La Grèce et la naissance de l'esprit grammatical

Le terme *grammatikê*, en grec, est issu du terme *gramma*, qui signifie « lettre ». La grammaire doit donc d'abord être comprise comme une étude pratique des lettres. Cela est dû au fait que, en empruntant aux Phéniciens le modèle d'une écriture alphabétique, les grecs ont été les premiers à se représenter le langage sous la forme d'un assemblage d'éléments pouvant être analysés séparément. Pendant longtemps, la grammaire restera associée à la pratique d'une lecture détaillée et de la copie des textes. Mais les recherches d'Aristote allaient entraîner les réflexions sur le langage dans une autre voie : la constitution de la grammaire. Son mérite, essentiel, est celui d'avoir, le premier, tenté une analyse précise de la structure linguistique sans doute la grammaire générale n'est-elle pour lui qu'une partie ou qu'un aspect de la logique mais son souci d'élaborer une théorie de la phrase, la distinction qu'il a faite entre les parties du discours, l'énumération qu'il a proposée des catégories grammaticales placent Aristote et les aristotéliens au début d'une longue tradition : leur aperception des données linguistiques comme le vocabulaire qu'ils ont mis au point à ce sujet se sont imposés jusqu'à une date récente et pèsent encore aujourd'hui sur la pédagogie et la méthodologie de nos études.

Chez Denys le Thrace (IIe siècle av. J.-C), qui est l'auteur du premier traité grammatical grec connu, on trouve une première liste de parties de discours (nom, verbe, etc.) ainsi que de catégories (le genre, le nombre, etc.). Toutefois, ce traité n'est pas, à l'échelle mondiale, le premier ouvrage de grammaire connu : il existe au IVe siècle av. J.-C., en Inde, un ouvrage d'un dénommé Pāṇini qui comporte une description très minutieuse de l'ancienne langue de l'Inde, *le sanscrit*.

2.1.3.2. Rome : les premières grammaires constituées

Les penseurs latins sont moins intéressés par la philosophie du langage que leurs prédécesseurs grecs. Ils sont d'abord concernés par des disciplines pratiques telles que la rhétorique. A titre d'exemple, Quintilien propose, au Ier siècle après J.-C., une somme de tout ce qu'il est nécessaire de savoir pour prononcer un discours éloquent. De leur côté, les grammairiens latins affinent la description linguistique esquissée par les grecs. Certains, comme Varron (Ier siècle av. J.-C), tentent d'ordonner davantage la description des mots en utilisant leur morphologie (mots variables vs mots invariables) et leur étymologie. En effet, au moment où les grammairiens latins se sont efforcés plier servilement l'étude de leur langue aux « règles » formulées par les théoriciens grecs, ne faisant que reprendre et propager les idées de ceux-ci, on remarque que, Varron a fait un gros effort pour définir la grammaire à la fois comme une

science et comme art et a aperçu, avec plus de lucidité que les Grecs, la valeur de l'opposition aspectuelle dans le système. Malheureusement, du point de vue moderne, ces recherches manquent souvent de la documentation et de la rigueur nécessaire.

Mais c'est chez le grammairien latin Priscien (Ve siècle ap. J.-C.) que l'on trouve le premier travail d'ampleur ressemblant à ce que nous appelons aujourd'hui une « grammaire ». Disciple du grammairien grec Apollonios Dyscole, Priscien approfondit la notion de syntaxe. Il estime que dans tout énoncé complet, certains mots appellent nécessairement d'autres. Par exemple, si je dis « il accuse », il est clair qu'il manque un autre mot qu'on appellerait aujourd'hui « le complément ». Chaque fois qu'il manquera un mot de la sorte, Priscien dira que le discours est imparfait. La syntaxe pourra ainsi être définie comme l'étude de ce qui est nécessaire à la construction d'un discours parfait.

2.1.3.3. La grammaire de Pāṇini (VI^e s. av. J-C)

Dans ce contexte d'une antiquité qui n'avait pas encore éprouvé le besoin d'ériger la linguistique en tant que discipline scientifique, Pāṇini fait figure de linguiste *avant la lettre*. Il a écrit la première grammaire connue (« अष्टाध्यायी », *Aṣṭādhyāyī*, qui signifie « en quatre parties »). Cette grammaire résume en 4000 règles environ le fonctionnement du sanskrit. Ces règles étaient composées de manière brève, afin de pouvoir être apprises par cœur.

À l'époque où fut écrit *Aṣṭādhyāyī*, le sanskrit traversait un mouvement d'épuration linguistique. Les règles de l'ouvrage sont qualifiées soit de SAMŚKRITA (épuré, parfait), soit de PRAKRIT (naturel, peu soigné). Ces données indiquaient donc si la règle décrite était ou non reconnue par la norme. Les règles comportent également des symboles, qui réfèrent à des classifications de phonèmes et de lexèmes. On peut dire de la grammaire de Pāṇini qu'il s'agit de la première grammaire générative avant la lettre, car ses règles avaient pour but de décrire la langue, dans la perspective de pouvoir en former de nouveaux énoncés. Il s'agit donc de la première description linguistique du langage. Cet ouvrage précurseur a eu par ce fait une importance capitale pour le développement de la linguistique contemporaine. Il se distingue en effet par la finesse des concepts linguistiques fondamentaux, l'exactitude et la précision de l'analyse qu'il fait du sanskrit. Pour la première fois, on peut parler d'un véritable métalangage, utilisant déjà toute une série de règles, un système de symboles et une terminologie rigoureuse, et interprétant objectivement les faits de langue. Outre cet ouvrage, Pāṇini a été à l'origine d'un modèle intellectuel et littéraire encore admiré aujourd'hui pour sa grande finesse. Au moment où Pāṇini a rédigé sa grammaire, le sanskrit était en passe de devenir une langue morte. On peut donc considérer que son *Aṣṭādhyāyī* en reste donc l'ouvrage de référence. Le sanskrit, à cette époque, n'était plus la langue du peuple, mais alors une langue de culture utilisée par une certaine élite sociale. On continue encore aujourd'hui à l'utiliser un peu comme le latin dans les textes religieux européens : la langue est devenue un véhicule culturel.

2.2 Le Moyen Age

Du point de vue philosophique, le Moyen Age marque le moment où l'on prend conscience de la force du langage et du mystère de la signification qui s'y accomplit, indépendamment des lois générales qu'étudie la logique. A cette époque, la grammaire commence insensiblement à se détacher de la philosophie et à prendre un envol autonome.

2.2.1 Les théories du langage dans le monde arabe

Le monde arabe s'est intéressé très tôt au langage. Grâce aux richesses de la bibliothèque d'Alexandrie, les philosophes arabes avaient pu se familiariser avec l'enseignement d'Aristote. La première grammaire complète de l'arabe date du VIII^e siècle. Dans l'islam, l'étude du langage est étroitement liée à la lecture du Coran et à l'effort qui doit être fait pour le comprendre. C'est pourquoi les penseurs arabes ont privilégié la dimension pragmatique des

énoncés, plutôt que leur sens immédiat. Ils s'intéressent à la valeur que peut avoir un énoncé que peut avoir un énoncé pour celui qui le reçoit, s'il doit être interprété comme une simple déclaration ou un acte véritable. Leurs théories se rapprochent beaucoup de la théorie actuelle des actes de langage.

Aussi, dans le monde arabe comme dans le monde chrétien, le Moyen Age est une période de grande activité dans le domaine de la logique et beaucoup de travaux de cette époque approfondissent l'enseignement d'Aristote.

2.2.2 La grammaire spéculative en Occident

Les penseurs chrétiens du Moyen Age considèrent volontiers que le monde ne nous est pas accessible directement, mais que nous pouvons avoir des aperçus de son essence en étudiant le langage, qui en est le miroir (*speculum*, en latin, d'où le nom de « grammaire spéculative », discipline qui naît vers 1255 à la Faculté des arts de l'Université de Paris).

Ainsi, pour les grammairiens médiévaux, le langage est un système complexe de significations. Il existe plusieurs « modes de signification » par lesquels les choses manifestent leur existence à l'intelligence humaine. L'étude du langage est d'une certaine façon l'étude de ces modes de signification. Cette théorie a valu à ses inventeurs le nom de grammairiens « modistes ». Les modistes considéraient qu'il existe une structure grammaticale une et universelle inhérente à toutes les langues et que, par conséquent, les règles de grammaire sont, en tant que règles, parfaitement indépendantes des langages particuliers dans lesquels elles trouvent leur réalisation.

Il importe également de remarquer qu'une grande partie de la production grammaticale européenne du Moyen Age concerne le latin. Les grammairiens approfondissent donc les théories de Priscien et considèrent que seul le latin est une langue véritablement grammaticale. Pendant la période qui précède le Xe siècle s'élaborent des systèmes de notation des diverses langues « barbares » qui remplacent le latin en Europe : *l'alphabet ogamique* (IVe s) pour les langues celtes, *l'alphabet gothique* (VIe s) pour les langues germaniques, *l'alphabet cyrillique* (Xe s) pour les langues slaves. Cependant, pendant plusieurs siècles, les langues vivantes de l'Europe continueront d'être vues comme des dialectes grossiers et impossibles à décrire. Dante (1265-1321) a été le premier à défendre une langue vulgaire contre le latin. Pour lui, c'est dans le modèle de la langue maternelle, apprise sans règles et par imitation, et non dans le modèle grammatical, que se trouve la véritable « noblesse » du langage. On voit alors qu'à partir de la Renaissance, on va s'intéresser de plus en plus aux langues modernes, principalement le français, l'italien, l'espagnol, l'allemand, l'anglais, et qu'on va les considérer comme des langues à part entière.

2.3. De la renaissance à la fin du 18^e s : comparaison de langues, langue originelle

Il faudra attendre le XVI^e S pour voir, dans l'intense fermentation intellectuelle qui caractérise cette époque, se créer un climat plus favorable à une étude minutieuse et ce, pour plusieurs raisons :

- (i) Une meilleure connaissance du grec et des traditions antiques se répand en Italie à la suite de l'arrivée de nombreux savants byzantins chassés de Constantinople après la prise de la ville.
- (ii) D'autre part, les controverses théologiques rendent indispensable le recours à l'hébreu, langue sémitique d'une structure différente de celle des langues européennes, ce qui amène forcément des comparaisons d'ordre linguistique.
- (iii) De plus, la ferveur religieuse des réformés provoque la traduction des livres sacrés en de nombreux dialectes et, si le latin reste toujours la langue universelle, le mépris longtemps

témoigné à l'égard des langues vulgaires s'atténue et disparaît devant l'épanouissement de riches et vigoureuses littératures nationales.

- (iv) Enfin des voyageurs, des commerçants, des diplomates rapportent de leurs expériences à l'étranger quelques notions sur des idiomes jusqu'alors totalement inconnus.

L'abondance de matériaux ainsi recueillis suscite l'étonnement et éveille l'intérêt ; on sent le besoin de classer les faits et de les offrir à l'observation attentive des lettrés ou à la curiosité amusée des dilettantes. C'est le type de classement géographique qu'on trouve appliqué dès le 15^e S soit dans des dictionnaires en plusieurs langues, soit dans des recueils présentant un texte, généralement le *Pater Noster*, traduit en différentes langues. Le plus ancien (1502) dictionnaire polyglotte est celui de l'Italien Ambrosio Calepino : il connut un succès tel qu'il fut à plusieurs reprises remanié et enrichi et qu'il en sortit le nom commun « Calepin ». Mais c'est en 1555 que Conrad Gesner publie à Zurich le premier recueil important du *Pater Noster* en 22 langues intitulé **Mithridates** ; ce titre se référant au roi du Pont qui, au premier siècle de notre ère, s'était attiré l'admiration de ses contemporains pour ses facultés, peu communes à l'Antiquité, de polyglotte. On rapporte que ce roi était capable de s'entretenir sans interprète et avec aisance avec tous les sujets de son royaume. On peut aussi parler du *Thesaurus Polyglottus* de Jérôme Megiser (1603) où sont recensés des dialectes du continent américain découvert 100 ans plus tôt. Et dans les siècles qui suivirent, de nombreux ouvrages s'ingénierent à faire le point de la connaissance des langues du globe, rassemblant des spécimens d'écritures, collationnant de petits textes, établissant des vocabulaires comparés ; le couronnement de ce genre d'études étant l'édition, en 1806, par J. Chr. Adelung (et poursuivie après sa mort par J.S. Vater) d'un imposant recueil, intitulé lui aussi **Mithridates** et dont les 4 volumes inventoriaient, selon un plan purement géographique, quelque cinq cents langues.

Cependant, dès le début de la recherche au 16^e s, un autre principe de méthode se faisait jour qui aurait dû rationaliser l'étude du rapport entre les dialectes : c'est celui de la communauté d'origine permettant de classer les langues en familles. Mais ces premiers essais furent malheureux : les érudits qui s'attachaient ainsi à classer les langues en transcendant le point de vue géographique partaient non pas de « examen des documents mais d'une conception a priori, à savoir la précellence de l'hébreu. Pour des raisons religieuses, ils considéraient l'hébreu, langue de l'Ancien Testament, comme la langue primitive à partir de laquelle il convenait d'expliquer toutes les autres, suivant en cela une tradition chrétienne qui semble remonter aux spéculations du judaïsme alexandrin. Une telle conception conduisait à une impasse car ce n'est qu'en torturant la forme et le sens des mots qu'on parvient à trouver certaines similitudes entre les langues européennes et l'hébreu, dialecte sémitique d'un type tout différent.

Un certain Joseph-Juste Scaliger se montrait tout de même inspiré. Dans sa *Diatriba de Europaeorum*, il se bornait à présenter le tableau des langues d'Europe, y distinguant 11 langues matrices dont étaient issues des propagines, ces dernières étant parentes entre elles. Les 4 principales « langues-souches » étaient désignées d'après la forme du nom de « Dieu » : *Boge*, *Godt*, *Deus*, *Θεός*, c'est-à-dire de façon fort exacte les langues slaves, germaniques, romanes et le grec. Certes, il subsistait des erreurs de présentation et des lacunes dans son travail mais dans l'ensemble il y avait une vue lucide des problèmes, de méthode correcte et les rapprochements proposés étaient justifiés par des exemples. Malheureusement, il est important de remarquer que les théoriciens s'accrochaient pendant plus de deux siècles contre l'évidence à l'explication par la monogénèse, un dialecte – l'hébreu le plus souvent – étant considéré comme langue-mère ayant donné naissance à toutes les autres.

Au début du 18^e s, Leibniz – qui d’ailleurs avait, comme Platon, reconnu le caractère arbitraire du signe linguistique – combattit vigoureusement et avec bon sens l’hypothèse de l’origine hébraïque mais sa protestation ne rencontra guère d’écho. Il recommandait, pour étudier le langage humain, de se reporter aux monuments écrits des époques anciennes afin de les comparer aux langues modernes dans leur état le plus récent. Mais pour déforer les principes méthodologiques qu’il venait d’énoncer, il bâtissait en même temps sans base sérieuse un système généalogique selon lequel les langues d’Europe et d’Asie – voire celles d’Afrique et d’Amérique – dériveraient d’un prototype commun.

Un autre chercheur original qui retient notre attention dans ce cours est l’Italien Giambattista Vico. Dans sa *Scienza nuova* (1725), il propose une définition d’une *storia ideal eterna* conçue comme une histoire cyclique du genre humain (car arrivée à son apogée, l’humanité se corrompt et retourne à la barbarie puisque son évolution se marque par une série de flux et de reflux) en 3 étapes :

- (1) théocratique (crainte religieuse devant les phénomènes naturels)
- (2) héroïque (société aristocratique) et ;
- (3) démocratique (stade rationnel où fleurissent au maximum la justice et la civilisation).

Vico tire les conséquences en ce qui concerne le langage. Pour lui, la langue du premier âge fut muette, les hommes communiquant entre eux aux moyens des signes, mais le premier langage articulé fut symbolique, c’est-à-dire poétique et les hommes se sont exprimés tout naturellement en vers. Le troisième état de langue est la langue humaine composée de vocables dont les peuples peuvent fixer le sens à leur gré.

2.4. L’Age classique (17^e – 18^e siècle)

2.4.1. L’enregistrement des données

Comme l’on peut le remarquer au point précédent, les premières grammaires des langues modernes datent de la Renaissance, époque à laquelle on s’intéresse à ce qui fait peut-être leur spécificité par rapport aux langues anciennes. Il faut remarquer toutefois que la plupart des travaux réalisés à cette époque continuent à être rédigés en latin, ce qui provoque un décalage entre les concepts utilisés (concepts empruntés à la grammaire latine) et les langues ainsi décrites. Les efforts s’accompagnent aussi parfois des tentatives de modernisation du système orthographique et d’études sur la prononciation. Aussi, faut-il remarquer qu’à cette époque il est encore difficile de parler de langues. Beaucoup d’habitants ne parlaient que leur patois et la langue de leur métier ; beaucoup d’érudits n’écrivaient qu’en latin. C’est pourquoi en effet qu’un peu partout en Europe apparaissent des méthodes d’apprentissage des langues dans lesquelles nous trouvons, par ailleurs, les premiers concepts grammaticaux modernes.

Certains grammairiens de cette époque n’étaient intéressés que par la récolte de mots, de façons de parler, de tournures, etc. On les a appelés les « grammairiens d’usage » ; c’est grâce à eux que la grammaire savante a pu évoluer (cf. les *Remarques sur la langue française* de Vaugelas en 1647). Il faut noter également que c’est aussi l’époque de la création, en Italie, puis en France d’académies destinées à améliorer la qualité des langues. Sous leur influence, les grammairiens d’usage, qui se voulaient au départ de simples enregistreurs, ont glissé insensiblement vers la grammaire normative.

2.4.2. Le projet de « grammaire générale »

A la fin du 17^e s, plusieurs penseurs forment le projet de décrire les langues, non plus dans ce qui les différencie, mais dans ce qui les rapproche. C’est l’époque de la « grammaire générale ». De semblables grammaires voient le jour dans beaucoup de pays européens. En France, on peut citer la Grammaire de Port-Royal (1660), la grammaire la plus connue de l’époque.

Les tenants de cette méthode estiment que, si les langues sont différentes dans leur détail, leur forme profonde, en revanche, est identique, car l'esprit humain n'a foncièrement qu'une forme.

Ces grammaires s'appuient donc très fermement sur la logique et essaient de démontrer comment plusieurs énoncés peuvent se ramener à un seul.

Au 18^e s, beaucoup de grammairiens se posent également des problèmes de terminologie. Il leur paraît clair que la terminologie latine est inadéquate pour décrire les langues modernes. Ils essaient d'en inventer de nouvelles, capables de refléter fidèlement chacune des langues étudiées. C'est ainsi que les grammairiens français, par exemple, Gabriel Girard, ou les grammairiens de l'Encyclopédie, César Chesneau Du Marsais et Nicolas Beauzée, vont essayer de distinguer les classes de mots et les emplois de ces mots dans la phrase, que Girard appelle « fonctions ». C'est de cette époque, que datent, par exemple, certaines classes, comme l'article ou l'adjectif.

2.3 Le 19^e siècle (le romantisme allemand)

2.3.1 De nouvelles manières de comprendre le langage

Deux penseurs allemands révolutionnent la manière de concevoir le langage : il s'agit de Johann Gottfried Herder (1744-1803), et de Wilhelm von Humboldt (1767-1835). Le premier se repose la question de l'origine du langage et formule l'hypothèse selon laquelle « le langage n'a pas comme origine le besoin de communiquer ou d'exprimer ses passions, mais une prise de conscience de l'homme par lui-même ». C'est en prenant conscience de lui-même que l'homme accède à la fois à la pensée et au langage. Chaque langue illustre une manière différente selon laquelle ce mouvement se réalise. C'est pourquoi chaque langue est en étroite liaison avec « le génie » du peuple qui la parle. Il fut en cela le précurseur de l'idée du rôle actif de la langue maternelle dans la formation de notre vision du monde grâce à l'union organique de la pensée et du langage. Selon Herder, la langue est non seulement un instrument [de communication], mais aussi un dépôt et une forme de la pensée⁹. En effet, pour Herder, la langue est le moule des sciences dans lequel et conformément auquel les pensées se configurent.

Pour le philosophe Humboldt, le langage n'est pas un tout achevé, figé, il est une énergie en activité constante. Chaque langue, dans son histoire, peut être considérée comme une illustration de ce qu'est le langage en général : une constante marche en avant. Humboldt adapte ici à l'analyse des langues un modèle de type « organique » que Goethe avait développé à propos des plantes : quelle que soit la différence apparente de structure qu'il y a entre une graine et une fleur, il existe une forme interne qui les relie et qui permet d'expliquer pourquoi, de la graine, est née la fleur. Pour Humboldt, chaque langue possède sa forme intérieure, qui se développe d'elle-même, et qui contient toutes ses transformations grammaticales à venir.

2.3.2 La découverte du sanscrit

A la fin du 18^e s, une découverte importante a modifié la façon de considérer les langues européennes. Quelques traducteurs de textes anciens de l'Inde, notamment William Jones¹⁰,

⁹ Un « dépôt », car l'expérience et le savoir des générations passées s'accumulent précisément dans la langue grâce à laquelle ils sont transmis aux générations suivantes. Une « forme de la pensée », car nous pensons non seulement dans une langue, mais aussi par l'intermédiaire d'une langue.

¹⁰ Dans la redécouverte du sanskrit qui s'opérait alors, Sir William Jones (1746-1794), érudit anglais vivant dans la colonie des Indes, fait figure de pionnier (il est par ailleurs le fondateur de la Société Asiatique du Bengale). Grâce à lui, l'étude du sanskrit est remise au goût du jour. Il rend la langue et la littérature de l'antiquité accessibles aux européens, et est à la source d'une série de travaux d'étude de la langue qu'il va redécouvrir dans une approche comparative. Par ses traductions d'écrits indiens antiques, on considère également qu'il a été indirectement responsable de l'éclosion du

avaient bien remarqué qu'entre le sanscrit – la langue sacrée de ce pays –, le grec et le latin, existaient des ressemblances qu'il paraissait d'attribuer au hasard. Mais c'est avec la publication, en 1816, par Franz Bopp (1791-1867), d'un mémoire sur leurs systèmes de conjugaison respectifs qu'on a la preuve d'une réelle parenté entre ces trois langues. On fait l'hypothèse qu'il a existé à un moment donné une langue mère, « l'indoeuropéen », dont sont issus aussi bien le grec que le sanscrit, le latin ou les langues germaniques. A partir de là, Bopp va pouvoir proposer sa *Grammaire comparée des langues indo-européennes* (1833). L'orientation que l'on donne à la discipline change radicalement de nature.

2.3.3 La grammaire comparée

La grammaire comparée, inscrite dans le comparatisme qui touche également d'autres sciences humaines, voit le jour à partir du XIXe siècle, principalement entre 1810 et 1875. La grammaire comparée avait pour but d'établir les liens de parenté entre deux ou plusieurs langues éloignées dans le temps, et le plus souvent dans l'espace.

Dès 1870, on considère qu'on est arrivé à un point de rupture par rapport au comparatisme : la grammaire comparée va alors se muer en une linguistique historique. *Cette nouvelle discipline se distingue de la grammaire comparée en ce qu'elle a pour but de reconstituer dans le détail l'intervalle et les lignes d'évolution d'une langue B depuis une langue A, chronologiquement antérieure* (et souvent éteinte). On ne tente plus d'induire une parenté hypothétique seulement par la comparaison, mais on élabore une véritable théorie basée sur des « lois » d'évolution, par le biais d'un véritable modèle théorique précédant l'observation.

Cette évolution est à mettre en relation avec la redécouverte du sanskrit en occident (c'est à cette époque qu'on redécouvre le manuscrit de Pāṇini). Les milieux lettrés commencent à s'intéresser à cette langue ancienne, peut-être bien antérieure au grec et au latin et présentant avec ceux-ci d'étonnantes similitudes.

2.3.3.1 Les précurseurs (17^e -18^e s)

La linguistique comparée a également connu plusieurs précurseurs, dont les travaux ont annoncé ce qui par après s'édifierait comme une véritable discipline de recherche.

HIOB W. LUDOLF (1624-1704) a montré de manière concluante la proximité entre l'éthiopien et l'amharique. Il s'agit de linguistique contrastive.

GOTTFRIED W. LEIBNITZ (1646-1716), mathématicien allemand, encourage la constitution d'une encyclopédie des langues parlées au sein de l'empire russe de l'époque.

Ces deux travaux ont débouché sur des méthodes d'analyse rigoureuses, et novatrices pour l'époque. Mais dans tous les cas, les liens linguistiques subodorés par les chercheurs étaient basés sur une intuition et des sentiments induits, bien plus que sur une attestation linguistique affirmée et vérifiable. Mais certains points de vue intéressants sont formulés, et il existe des éléments tangibles qui permettent d'ouvrir un horizon de recherches.

En 1756, **TURGOT** rédige pour l'Encyclopédie l'article sur l'*étymologie*. Cet article, contenant des idées nouvelles, va donner matière à réflexion aux premiers comparatistes, et en particulier au danois Raske.

J. C. ADELUNG (1732-1806), pour sa part, s'est illustré dans divers domaines. Il a contribué, par ses dictionnaires et ses travaux de grammaires, à une normalisation de la langue allemande. Il a également eu la volonté de rédiger un traité prétendument complet des langues connues à son époque. Il le nomme *Mithridates*, nom d'un souverain grec qui dit-on maîtrisait

romantisme britannique (Byron,...), leur fournissant de nouvelles sources d'inspiration et véhiculant une nouvelle esthétique. Jones remarque que le sanskrit, le latin et le grec partagent des ressemblances, notamment au niveau des racines verbales et des formes verbales qui ne peuvent être dues au hasard et qui seraient probablement d'une langue commune.

les 22 langues de ses sujets. Il n'a publié qu'une infime partie de cet ouvrage, puisqu'il est mort avant de le finir ; il n'a pu considérer que les langues asiatiques. À l'époque, il ne disposait pas d'outils philologiques performants. Il s'est donc fondé sur le critère empirique de la proximité géographique des langues, qui permet de formuler l'hypothèse de leur probable parenté. Plus tard, on s'apercevra que cette présupposition n'est pas nécessairement vraie.

Le russe **P. S. PALLAS** (1714-1811) publie de son côté un **VOCABULAIRE COMPARE DES LANGUES DU MONDE** (1786-1789).

2.3.3.2 Orientation de la grammaire comparée¹¹

On peut distinguer, au niveau du comparatisme linguistique, deux orientations générales :

- 1) L'analyse des langues classiques, réalisées par certains, comparant presque exclusivement, mais de manière approfondie, le sanskrit, le latin et le grec.
- 2) L'analyse des langues européennes et de leurs traits de ressemblance. Il faut noter que les linguistes qui s'employaient à cette comparaison étaient fréquemment motivés par une idéologie nationaliste sous-jacente.

Ces deux orientations ont mené à un double constat, qui amena les comparatistes à formuler l'hypothèse selon laquelle les langues du monde s'organisent en grandes familles qui sont chacune cohérente et qui chacune développe des systèmes grammaticaux, qu'il devrait exister différents types d'affinités entre les langues, que l'on pourrait appeler des « LIENS DE PARENTE ». L'hypothèse formulée était double :

- D'une part, on postulait que les langues, de la Scandinavie à l'Inde, soient nées de la transformation d'une même langue souche. On nomma alors cette langue l'**INDO-EURO-PÉEN**, à laquelle on parvenait par le biais de la reconstruction. De l'indo-européen se ramifient une série d'héritages, circulant dans une généalogie complexe.
- D'autre part, la comparaison des éléments grammaticaux de ces langues a permis d'établir des correspondances formelles entre les langues, et de modéliser leur évolution. Cette comparaison se fait soit dans le détail, soit dans les grandes lignes : langues mères/grandes langues/familles de langues.

La méthode est d'abord de comparer, tant en grammaire qu'en phonétique, les différentes langues européennes actuelles et de montrer comment tel élément dans l'une joue le même rôle que tel élément dans l'autre (par exemple, une voyelle ou un élément de morphologie). Dans un mot, ils distinguent les éléments lexicaux (radicaux) et les éléments grammaticaux.

2.3.3.3 Principaux partisans du domaine indoeuropéen

Au départ, ce sont essentiellement des recherches d'auteurs allemands qui développeront le concept de l'indoeuropéen (initialement nommé « *indogermanisch* »). Les recherches qui appuient cette théorie coïncident fortement avec l'émergence du romantisme en Allemagne (à noter donc que s'il y a des linguistes, il y a encore forcément des littéraires derrière).

¹¹ Il est important de distinguer la grammaire comparée de la linguistique contrastive. La grammaire comparée est essentiellement diachronique, tandis que la linguistique contrastive s'inscrit dans une approche synchronique ; elle ne se préoccupe pas de reconstituer des familles de langues, ni de découvrir des liens de parenté.

2.3.3.3.1 *Les frères Schlegel*

Les frères Schlegel (Karl Wilhelm Friedrich von Schlegel [1772-1829] et August Wilhelm von Schlegel [1767-1845]) furent les premiers à utiliser le terme de *Vergleichende Grammatik*. Friedrich est influencé par l'*Essai sur l'origine des langues* de Rousseau, dans le climat intellectuel de l'*Aufklärung* et du *Sturm und Drang*; il porte un intérêt sur la langue en tant qu'expression des cultures. Il sera le premier à mettre en valeur l'analyse morphologique, et y montrer l'intérêt pour démêler les relations *génético-linguistiques* qui existent entre les langues. Son frère, August Schlegel, écrit notamment en 1808 *De la langue et du savoir des Indiens*. Il proposera plus tard la typologie des langues sur base de leurs structures morphologiques (la différence entre les langues isolantes, affixantes et flexionnelles). Cette typologie sera reprise plus tard par Humboldt.

2.3.3.3.2 *2.4.3.3.2. Les frères Grimm*

Les frères Grimm ont notamment rédigé une *Histoire de la langue allemande* et un *dictionnaire allemand* ainsi qu'une série d'ouvrages qui portent sur le folklore allemand.

Jakob Grimm (1785-1863) a le premier défini des lois phonétiques, parmi lesquelles la célèbre loi de la première mutation consonantique. Il publie en 1822 un traité des grammaires germaniques, dans lequel il va systématiser les observations faites par Rask. Jakob Grimm est à la base de différentes publications importantes, parmi lesquelles **Histoire de la linguistique et de la philologie orientale en Allemagne** (1869), et **Recherches étymologiques dans le domaine des langues indo-européennes** (1833-1836). Grimm montre entre autres dans cette dernière que les langues germaniques comportent régulièrement un /f/ là où le grec comporte un /p/. Il énonce qu'il a dû y avoir une mutation, survenue à la période préhistorique du germanique. Il énonce les lois suivantes :

- Les occlusives sourdes deviennent des fricatives sourdes : par exemple, la souche indo-européenne *PH₂tèr ('père') qui a donné le latin 'Pater', donne en vieux germanique '*Fadar' : (/p/ devient /θ/).
- Les occlusives sonores deviennent des occlusives sourdes : par exemple, la souche indo-européenne *De₂kemt ('dix') qui a donné le latin 'Decem', donne en vieux germanique '*Tehan' : (/d/ devient /t/ ou /ð/).

2.3.3.3.3 *Rasmus Rask (1787-1832)*

Contrairement aux frères Schlegel et Grimm, Rasmus Rask est héritier du rationalisme empiriste¹², et également du naturalisme. Cette double filiation va lui permettre de reculer les limites du comparatisme linguistique, en prenant davantage de recul par rapport à la philologie, à la littérature et aux aspects culturels de la linguistique. Rask est **le premier à chercher à donner des explications naturelles au changement linguistique, précédant Grimm dans la formulation de la loi des mutations consonantiques.**

Rask a effectué des voyages en Islande, en Russie et en Perse. Suite à cela, il écrit les ouvrages *Grammaire de l'islandais* (1811) et *Recherches sur l'origine de l'ancienne langue nordique ou islandaise* (1818), dans lesquels **il est le premier à rapprocher l'islandais avec le gothique ancien, mais aussi avec le vieux slave, le grec, les langues baltes et les langues latines.** Lorsqu'il publie ses travaux, on n'avait pas encore redécouvert le sanskrit, ce qui l'a empêché de pouvoir étendre la portée de son analyse. Il va prendre appui sur la réflexion de

¹² Le rationalisme empiriste était un courant de pensée proposant une méthodologie scientifique de classification de données réelles, attestées, avant toute analyse.

Turgot, en estimant que ce sont **les relations étymologiques** qui prévalent dans l'évolution des langues. Il essaie de montrer qu'il y a entre certaines langues d'Europe, et plus particulièrement les langues nordiques, un lien de parenté évident. Il analyse l'évolution des langues sous le regard qui est celui du naturalisme pré-darwinien. L'idée en est que les « espèces » étaient relativement stables avant de subir des mutations. Il applique ce postulat au rapport étymologique, par lequel il va pouvoir démontrer la parenté d'un **son** d'une langue avec le **son** d'une autre langue.

Il faut noter que parmi les comparatistes (qui travaillent sur des documents écrits), Rask a été le premier à discerner le son de la graphie pour en dégager la parenté entre les différentes langues. Il est l'un de ceux, et **l'un des premiers, qui ont établi et reconnu la parenté indo-européenne**. Il est une grande figure de la philologie nordique, et on le considère comme un des précurseurs du structuralisme. Son rayonnement a hélas été moindre que celui de Franz Bopp.

2.3.3.3.4 *Franz Bopp (1791-1867)*

Bopp est principalement influencé par les travaux des frères Schlegel et Grimm. Dans l'idée, déjà avancée par Rask, de comparer les langues à des organismes vivants, il est à la recherche d'une langue mère qui précéderait les autres. Parallèlement à Rask, il va orienter la grammaire comparée à un examen de l'organisation morphologique des mots, principalement du nom et du verbe.

Bopp ne va pas utiliser la perspective naturaliste, mais évolutionniste, en suivant notamment les travaux de Darwin. Selon Bopp, le sanskrit serait la souche dont toutes les langues indoeuropéennes dérivent (on nomme cette thèse « la thèse de la primogéniture du sanskrit »). Il va postuler que les principes en vertu desquels les langues dérivent de la langue mère sont le fruit de transformations, qu'il va essayer d'établir.

Le philologue rédige plusieurs ouvrages sur le sanskrit, parmi lesquels *Système de conjugaison du sanskrit* (1816), qui lui servira de point de départ pour ses travaux ultérieurs. Ces travaux, d'une étendue immense, vont se consacrer à comparer les grammaires du latin, du grec et des langues slaves. Il va tenter d'expliquer d'une part l'existence des régularités formelles que l'on retrouve dans ces langues, et d'autre part l'origine des différentes formes flexionnelles qu'elles utilisent.

2.3.3.3.5 *Wilhelm von Humboldt (1767-1835)*

Alexander von Humboldt est surtout connu pour avoir développé une nouvelle philosophie du langage. Il a repris au littéraire **HERDER** une idée de la philosophie de l'histoire, selon laquelle une langue est l'expression du génie populaire. Pour Humboldt, « la langue d'un peuple est son esprit et son esprit est sa langue »¹³.

Contrairement à ses prédécesseurs, le langage comme la langue sont considérés chez Humboldt comme des activités distinctives de l'humain. Humboldt est le premier à proposer une définition de la compétence linguistique, et à avoir une conception de la langue comme quelque chose de dynamique constituant un système. Cette position particulière explique qu'il occupe dans le champ de la grammaire comparée une place à part : il conçoit l'évolution des langues comme un changement hérité du système, dans une conception structurale du langage.

Humboldt est également influencé par les idées de grands penseurs tels que **KANT** et **GOETHE**. À ce dernier, il reprend l'idée que chaque langue possède en elle toutes ses formations ultérieures. Par rapport à Kant, on retrouve chez lui la notion de forme interne de la langue.

¹³ « Ce que l'on a appelé l'hypothèse humboldtienne se rejoint avec l'[hypothèse Sapir-Whorf](#), laquelle veut que les catégories de la langue parlée prédéterminent nos catégories de pensée. Chaque langue renfermerait une vision du monde irréductible. »

Humboldt décline ce concept en trois principes avancés comme vrais pour toutes les langues :

- (1) En deçà de leur forme structurelle, les langues articulent des représentations sémantiques : elles sont un MODE D'ACTIVITÉ DE L'ESPRIT HUMAIN, structurant le sens, et sont révélatrices de la pensée et de la perception des phénomènes, dans un modèle proche de celui de la *Critique de la Raison pure* de Kant.
- (2) Chaque langue, grâce à sa forme interne, comporterait un MÉCANISME ACTIF DE SYNTHÈSE DES PERCEPTIONS. On peut en dégager que le langage, ou la langue, est un mode d'action de l'esprit humain : il en est à la fois la production et le produit. La langue a donc une dimension « spirituelle », nécessairement associée à une matière sonore, sensible. C'est la première fois que l'on voit apparaître la notion de spécificité du signe linguistique, et dans une perspective qui est plus dynamique (En d'autres termes, Saussure n'a rien inventé, en ce sens qu'Humboldt avait déjà annoncé cette spécificité du signe).
- (3) Enfin, pour Humboldt, chaque langue dans sa forme interne DÉTERMINE UN IDIOME PARTICULIER (la LANGUE détermine la PAROLE). La phrase la plus simple est représentative de l'unité de tout le système, et cette langue est considérée alors comme un organisme vivant. Humboldt est précurseur de ce que Saussure va mettre *noir sur blanc*, également en la matière de la distinction entre diachronie et synchronie.

Humboldt dépassera les limites de l'Allemagne pour inspirer l'école linguistique américaine, notamment Sapir et Whorf. Un siècle plus tard, en effet, Edward Sapir sera le premier à fonder une analyse conceptuelle, conçue comme structurante de l'esprit, opérée par le langage. Benjamin Lee Whorf reprend cette idée, en l'appliquant à une anthropologie linguistique des parlars amérindiens.

▪ La pensée de Sapir

Pour Sapir, « la langue d'une société humaine donnée qui pense et parle dans cette langue, est l'organisatrice de son expérience et, de ce fait façonne son « monde » et « sa réalité sociale », i.e. chaque langue contient une vision spécifique du monde. Cette idée n'est pas nouvelle, mais elle se caractérise par les faits suivants :

- (1) Sapir défend un point de vue plus modéré que celui des Humboldistes. Par ex, il énonce simplement que la langue socialement formée influence à son tour la manière dont la société conçoit la réalité.
- (2) Sapir met en valeur l'ordre social de la langue (Sociolinguistique). Il dit ceci : « *Le langage est un guide dans la 'réalité sociale'. Bien que le langage ne soit pas d'ordinaire considéré comme un objet d'étude des sciences sociales, il conditionne cependant considérablement toutes nos pensées sur les problèmes et les processus sociaux* ». (cf. The status of linguistics as a science, 1969, p.162 cité par Haddad 2002 : 16).
- (3) Sapir précise l'influence de la langue dans la perception de la réalité : « *Il est illusoire, dit-il, de s'imaginer que les hommes s'adaptent essentiellement à la réalité sans se servir du langage, et que le langage est seulement un moyen accidentel de résoudre les problèmes scientifiques de la communication et de la réflexion. (...) Nous voyons, nous entendons et nous expérimentons comme nous le faisons, parce que les habitudes linguistiques de notre communauté prédisposent à certains choix d'interprétation* ». (idem, p.99 cité par Hadda 2002 : 16).

▪ La pensée de Benjamin Lee Whorf

Whorf raconte qu'avant d'avoir rencontré Sapir et connu de ses conceptions, la pratique le

poussa à méditer sur sujet de l'influence qu'exercent les mots, acceptées par un groupe social, sur le comportement des individus. Il profita de sa connaissance de la langue et de la culture des Hopi pour vérifier – sur des matériaux concrets – ses thèses générales empruntées à Sapir. Il affirme que la langue et l'image du monde Hopi, non seulement diffèrent des langues et des images du monde des peuples européens, mais leur sont même opposées. Whorf parvint ainsi à concrétiser les thèses de Sapir et à considérablement les radicaliser, formulant en définitive le principe de *la relativité linguistique*. Il le formule comme suit :

« Nous morcelons la nature conformément aux directives esquissées par notre langue maternelle. Les catégories et les types que nous isolons du monde des phénomènes, nous ne les y trouvons pas... Au contraire, le monde se présente à nous comme un kaléidoscopique d'impressions qui doit être organisés par nos esprits – et ce dans une grande mesure, par les systèmes linguistiques que nos esprits contiennent. Nous disséquons la nature, nous l'organisons en concepts auxquels nous attribuons des significations d'une manière définie, principalement parce que nous sommes les parties contractantes d'une convention aux termes de laquelle nous devons l'organiser ainsi ; d'une convention qui est de rigueur pour notre communauté linguistique et est codifiée dans les modèles de notre langue. Cette convention, évidemment, est implicite et n'est pas spécifiée, mais ces directs sont en vigueur d'une manière absolue ; nous ne pouvons parler que si nous souscrivons à l'organisation et à la classification des données que la convention décrète.

Ce fait est très important pour la science moderne ; parce qu'il signifie que l'individu ne peut pas décrire la nature avec une impartialité absolue, mais qu'il est contraint à certains modes d'interprétation même quand il est convaincu d'être libre. Relativement, le plus libre à ce point de vue serait un linguiste familiarisé avec de nombreux systèmes linguistiques qui différeraient considérablement entre eux. Jusqu'à présent, aucun linguiste ne se trouve dans cette position. Nous avons donc abouti à un nouveau principe de relativité qui dit que la même réalité physique n'amène pas tous les observateurs à la même image de l'univers, à moins que leurs fonds linguistiques ne soient semblables ou puissent être d'une manière quelconque ramenés aux mêmes termes ». (cf. Haddad 2002 : 17-18).

Actuellement, la thèse Sapir-Whorf est surtout connue pour avoir énoncé l'idée que la langue est le reflet d'une vision du monde. En effet, contrairement à la tradition romantique selon laquelle les correspondances entre les phénomènes de langue et le comportement humain s'expliquent par le fait que le génie particulier de chaque peuple s'exprime à travers sa langue, Whorf et Sapir posent l'hypothèse selon laquelle la langue d'une communauté donnée organise sa culture, c'est-à-dire l'appréhension que ce peuple a de la réalité et la représentation qu'il se forme du monde. Pour Whorf et Sapir, la différence de langues a pour conséquence une structuration intellectuelle et affective différente ; il s'agit de deux mondes différents et non pas du même monde sous deux séries d'étiquettes différentes. C'est là qu'apparaît l'analyse opérée par le langage, avancée par Humboldt. Sapir a inspiré un certain Bloomfield, qui est le père du structuralisme américain.

2.3.3.3.6 AUGUST Schleicher (1821-1868)

Si Humboldt occupait une place à part dans le champ de la linguistique comparée, on peut considérer que Schleicher pose un tournant radical. Grâce à l'influence de l'évolutionnisme linguistique, mais également du romantisme philosophique, il propose un dépassement de l'œuvre de Rask et de Bopp. Dans deux ouvrages qui auront une importance capitale dans l'histoire de la linguistique historique : **ABREGE DE LA GRAMMAIRE COMPAREE DES LANGUES INDOEUROPEENNES** (1861-1862) et **THEORIE DARWINIENNE ET L'ESPRIT DES LANGUES** (1865).

On peut le considérer comme l'aboutissement du comparatisme traditionnel, dans la visée historique qui y était présente en germe. Il sera le premier à récuser le postulat du sanskrit

comme étant la langue mère, et propose à la place de celui-ci l'indoeuropéen comme *Ursprache*. À partir de Schleicher, on constate que la grammaire comparée se redéfinit en linguistique historique, qui va chercher à reconstituer l'évolution des langues. Il s'agit désormais, à partir d'outils développés, de reconstituer les stades d'évolution manquants ainsi que les filiations directes ou transversales entre les langues.

Schleicher va proposer deux nouveaux outils méthodologiques qui vont marquer les théories suivantes : L'utilisation d'un **ARBRE PHYLOGENETIQUE** des langues, classant chacune des langues selon leur ancêtre direct à la manière d'une généalogie.

Schleicher sera également le premier à reconstituer l'**INDOEUROPEEN PRIMITIF**¹⁴, par ce qu'il va appeler la *Stammbaumtheorie* (« théorie du tronc commun »). Cette théorie tente d'illustrer le processus de différenciation à partir de la figuration explicite de liens, et de la représentation d'une langue mère et de langues filles.

La *LECTURE ASCENDANTE DE L'ARBRE* permet de suivre l'ordre possible de la reconstruction des liens de parenté, en vérifiant le schéma suivant : les langues communes, les subdivisions en sous-familles, la langue parente commune et la langue archaïque (*Ursprache*).

La *LECTURE DESCENDANTE* de l'arbre, elle, fait référence à l'ordre historico-génétique, dans lequel les langues se différencient progressivement ; une langue archaïque (*Ursprache*) se situant à l'origine. Pour Schleicher, cet arbre se serait ramifié en trois branches : les langues isolantes (le chinois), les langues agglutinantes (le hongrois) et les langues flexionnelles (le sanscrit), avant d'entrer dans une phase de déclin. Influencé par les idées de Darwin, Schleicher estime qu'à l'intérieur de cet arbre les langues naissent, se développent et meurent, comme des organismes vivants.

2.3.3.4 Les limites de la grammaire comparée

Le défaut majeur des différentes théories qui ont jalonné le courant de la grammaire comparée est de ne pas reposer sur une base empirique suffisante. Son objet d'études, une langue mère plus ou moins reconstituée (le sanskrit ou l'indoeuropéen) n'est connu que par le biais de filiations linguistiques, perdu dans les frimas d'un passé insondable. Il est donc bien évidemment impossible de¹⁵ falsifier des thèses comme celle de l'indoeuropéen.

Un autre élément par lequel on peut considérer que le comparatisme est une théorie risquée, est qu'on ne pourra jamais exclure, dans l'analyse de prétendues racines communes, d'une part la possibilité d'une coïncidence, et d'autre part la possibilité d'un emprunt, qui est quasiment le corollaire de tout contact culturel. Seuls quelques champs lexicaux d'une langue sont *presque* en mesure d'échapper à l'emprunt : les parties du corps, les liens familiaux, ...

Certains mots échappent également aux grammairiens, partageant une ressemblance formelle avec d'autres mots d'une autre langue, alors que leur sens aura été changé au cours du temps et de l'évolution culturelle (par préférence d'un sens second, etc.). Pour d'autres racines, la situation inverse se produira : alors que le sens est demeuré similaire, une évolution phonologique très importante a rendu le mot méconnaissable. De là, on voit comment

¹⁴Voir à ce propos l'article de Wikipédia « La fable de Schleicher ». Cette fable a été imaginée par Schleicher pour constituer un texte entier en indoeuropéen. Elle a ensuite été réécrite au fil des théories : [http://fr.wikipedia.org/wiki/Fable_de_Schleicher_\(Le_mouton_et_les_chevaux\)](http://fr.wikipedia.org/wiki/Fable_de_Schleicher_(Le_mouton_et_les_chevaux))

¹⁵Il faut cependant noter qu'une thèse avancée par Saussure lors de son mémoire, en 1878, a pu par la suite être démontrée. Elle a « procuré à la grammaire comparée l'une de ses découvertes les plus marquantes, dont la communauté scientifique de l'époque a néanmoins eu beaucoup de peine à reconnaître le bien-fondé. [...] Saussure démontrait prophétiquement l'existence, en indo-européen, de deux phonèmes supplémentaires notés *A* et *O* : il se souciait essentiellement de spécifier leur fonctionnement dans les alternances grammaticales, mettant délibérément entre parenthèses le problème de leur contenu phonétique. Longtemps contestée, la réalité des "coefficients sonantiques" (plus tard appelés "schwa" et enfin "laryngales") fut prouvée de manière externe, près de cinquante ans plus tard et bien après la mort de Saussure, par le témoignage du hittite nouvellement déchiffré. » (<http://www2.unine.ch/files/content/sites/linguistique.francaise/files/shared/documents/MethodeBader.pdf>)

considérer des aspects culturels ou sémantiques en relations avec des données formelles est une démarche risquée, à la scientificité contestable.

2.3.4 La grammaire historique : les néogrammairiens

Dans la vague positiviste de seconde moitié du XIX^e siècle, bercée par les idées d'Auguste Comte, cette génération réclame une base empirique plus forte, et est hostile aux conceptions romantiques des langues comme nécessairement attachées à la culture et à un « génie du peuple ». Les grammairiens prennent leur distance avec les ambitions philosophiques et globalisantes de la génération de Schleicher, et essaient d'appliquer à la grammaire des principes similaires à ceux appliqués dans les sciences. Cette école a été nommée les « néogrammairiens ».

L'objectif de cette école était d'expliquer le changement linguistique. On fait l'hypothèse qu'il existe des « lois phonétiques » entièrement mécaniques, qui ont des causes physiologiques. En français par exemple, le *c* suivi du *a* est devenu systématiquement *ch* : *campus* donne *champ* (Siouffi & Van Raemdonck, 1999 : 27). En France, la seconde moitié est l'époque des premières études sur l'ancien français, mais aussi celle de grands dictionnaires (Littré, Larousse), d'atlas linguistiques comme l'Atlas linguistique de la France, qui date du début XX^e siècle ou de grandes synthèses historiques sur la langue et la littérature.

Les néogrammairiens vont critiquer les visions *a priori* de la linguistique historique, et en particulier Schleicher et la question de l'origine des langues, en taxant son hypothèse de non scientifique, et en désignant le postulat d'une langue mère comme relevant du mythe, et donc non falsifiable et non vérifiable. L'idée défendue par Humboldt dressant un parallélisme nécessaire entre pensée et langage est également largement critiquée par les néogrammairiens. Enfin, la nouvelle école ne cautionne pas cette conception du changement comme étant une dégradation progressive conduisant à la mort des langues. Pour les néogrammairiens, c'est l'idée des lois phonétiques qui est primordiale :

« Les changements phonétiques que nous pouvons constater dans les documents de l'histoire linguistique procèdent selon des lois immuables qui n'admettent de variation qu'en conformité avec d'autres lois » (Scherer 1875; cité par Paveau & Sarfati 2003: 25).

Par l'examen des transformations phonétiques, il s'agit donc d'élaborer un modèle théorique en amont, dans une démarche à la fois inductive et déductive (on identifie les *causes* du changement). Plusieurs figures importantes de la linguistique intégreront le cercle des néogrammairiens. Parmi eux, on peut citer le slaviste August Leskien (1840-1916) qui concentrera les néogrammairiens autour de l'université de Leipzig, accompagné de Karl Brugmann (1849-1919), Hermann Paul (1846-1921) Karl Verner (1846-1896).

Karl Verner va corriger et compléter la loi avancée par Grimm. Ce faisant, il illustre bien quel va être l'apport des néogrammairiens par rapport à la grammaire comparée ; en fait, plus que la corriger, il va proposer une explication à certaines de ses irrégularités. Pour les néogrammairiens en effet, « **il doit exister une règle aux exceptions à une règle; le seul problème est de la découvrir** » (August Leskien).

Pour Verner, il faut donc prendre en compte, en plus de la règle avancée par Grimm, d'autres régularités. En linguistique, une exception n'est jamais une exception à la règle : il y a toujours une règle de l'exception.

Pour la première fois dans l'histoire de la linguistique, les néogrammairiens ont fait primer l'oral sur l'écrit. On voyait déjà poindre cette idée chez Rask, mais pas encore comme principe de base. Le changement linguistique obéit donc à des lois qui sont universelles et générales. Plutôt qu'une volonté humaine, il s'agit d'une tendance inconsciente, répondant tantôt à une nécessité articulatoire (explication physiologique ou « loi du moindre effort »), tantôt à une propension à l'analogie (explication psychologique). On voit comment les

néogrammairiens ont déjà établi, un siècle auparavant, le programme des linguistes typologistes « à la Greenberg ».

Deux autres figures importantes graviteront en périphérie de ce cercle de Leipzig : l'italien Graziado Ascoli, fondateur de la dialectologie, mais également le suisse Ferdinand de Saussure, qui rejoindra les néogrammairiens, avant de leur adresser de vives critiques. Elles lui permettront alors de fonder la linguistique moderne ; avant cela toutefois, il avait apporté une importante contribution théorique à l'approche néogrammairienne, dans son *Mémoire sur le système primitif des voyelles en indoeuropéen* (1879).

3 Les grands courants de la linguistique moderne

Après la grammaire comparée, la linguistique historique et l'approche néogrammairienne, les théories peu à peu s'enchaînent en une sorte de réaction par rapport aux courants précédents. Ainsi, à la fin du XIXe siècle, les recherches en linguistique atteignent un développement évident. L'optique est nettement historique, mais une nouvelle sensibilité apparaît, qui privilégie la théorisation de l'organisation des phénomènes. Ces réactions seront l'œuvre de plusieurs grandes figures de la linguistique.

3.1 Ferdinand de Saussure (1857-1913) ou le Structuralisme

Ferdinand de Saussure est mort en 1913, mais son *Cours de linguistique générale* n'a été publié qu'en 1916, sur base de notes prises par deux de ses assistants, Bally et Séchehaye. Si Saussure n'avait de son vivant jamais écrit d'ouvrage sur la théorie du langage qu'il développait, les notes diffusées sur son cours après sa mort sont devenues la pierre angulaire de la linguistique moderne.

3.1.1 Au tournant de la linguistique moderne

Saussure envisageait l'étude de la langue comme un *système* : il s'agit d'un terme primordial dans l'appréhension de Saussure de la langue. On considère par ce fait que le *Cours de linguistique générale* marque une rupture épistémologique, étant donné que les faits de langue sont désormais considérés de manière complètement différente ; d'une part, parce qu'il y a une rupture complète avec la linguistique comparatiste de son époque, et d'autre part, parce qu'il propose une approche non historique, descriptive et systématique de la langue.

Saussure parle pour la première fois de *système*, mais ce n'est qu'à partir des années 1920 que l'on va muer ce concept clé en une notion plus générale de « structure » qui sera à la base de la pensée structuraliste. C'est ainsi qu'on considère Saussure comme le père du structuralisme, quoiqu'il n'eût jamais lui-même employé ce terme.

Ce sont en effet ses idées sur le système de la langue qui ont été reprises, principalement par la phonologie telle qu'elle a été mise en place par le cercle de Prague, et notamment par Troubetzkoy, inventeur du terme de « phonème ». Mais, outre la phonologie pragoise, on constate également que les concepts mis en place par Saussure ont été récupérés par d'autres disciplines :

L'*ANTHROPOLOGIE* avec Lévi-Strauss

La *SEMILOGIE* avec Barthes

La *PSYCHOLOGIE* (du développement) avec Piaget

Etc.

On peut dire que Saussure a fait des émules, bien au-delà de la linguistique. Or en histoire des sciences, on considère souvent une science comme véritablement autonome à partir du moment où, dans ses concepts, elle se trouve récupérée par d'autres.

3.1.2 La conception saussurienne de la linguistique

3.1.2.1 Les bases d'une linguistique générale

Saussure édifie la linguistique générale en science, en jetant les *BASES* et en définissant les *TÂCHES* et l'*OBJET*. Il distingue des dichotomies fondamentales, notamment entre langue et parole. Saussure propose trois tâches à la linguistique :

(1) faire la description et l'histoire de toutes les langues qu'elle pourra atteindre. Ce qui revient à faire l'histoire des familles de langues et à reconstituer dans la mesure du possible les langues mères de chaque famille. Le terme de description est donc ici fondamental. C'est ce qui la distingue de la grammaire, elle normative ;

- (2) rechercher les forces en jeu d'une manière permanente et universelle dans toutes les langues, et de dégager les lois générales auxquelles on peut ramener tous les phénomènes particuliers de l'histoire ;
- (3) Se délimiter et se définir elle-même (cet aspect était déjà en germe chez Whitney). La linguistique a des rapports étroits avec d'autres sciences qui tantôt lui empruntent des données, tantôt lui en fournissent (ethnologie, ethnographie, préhistoire, anthropologie, physiologie, etc.).

La motivation de Saussure pour définir ces trois tâches est la recherche de la généralité.

3.1.2.2 L'objet de la linguistique

Pour Saussure, l'objet de la linguistique n'est pas le même que celui des sciences exactes. Il n'est pas donné d'avance : il faut construire l'objet de la linguistique de sorte qu'on puisse l'étudier et adopter une attitude résolument scientifique. Il nourrit alors l'objectif de définir l'objet intégral et concret de la linguistique afin de délimiter une discipline scientifique et autonome. Étant aux prises avec le caractère hétérogène, confus et duel des faits abordés, il conclut que l'unique solution consiste à « se placer de prime abord sur le terrain de la langue et la prendre pour norme de toutes les manifestations du langage ». Selon donc Saussure ce qui devrait être l'objet de la linguistique c'est la langue et non le langage.

Saussure pose comme postulat que le langage est une faculté humaine qui est beaucoup plus vaste, et moins spécifique que la langue. Pour lui, le langage est une faculté que nous tenons de la nature, tandis que la langue est un produit social, qui permet à l'individu de pouvoir exercer sa faculté de langage. Il l'explique comme suit :

« Mais qu'est-ce que la langue ? Pour nous, elle ne se confond pas avec le langage ; elle n'en est qu'une partie déterminée essentiel il est vrai. C'est à la fois un et un ensemble de conventions nécessaires, adoptées par le corps social pour permettre l'exercice de cette faculté chez les individus produit social de la faculté du langage. Pris dans son tout, le langage est multiforme et hétéroclite ; à cheval sur plusieurs domaines, à la fois physiques, physiologiques et psychiques, il appartient encore au domaine de l'individuel et au domaine du social ; il ne se laisse dans aucune catégorie des faits humains, parce qu'on ne sait comment dégager son unité. La langue, au contraire, est un tout en soi et un principe de classification. Dès que nous lui donnons la première place parmi les faits de langage, nous introduisons un ordre naturel dans un ensemble qui ne se prête à aucune autre classification. [...] En séparant la langue et la parole, on sépare du même coup : 1° ce qui est social de ce qui est individuel ; 2° ce qui est essentiel de ce qui est accessoire et plus ou moins accidentel. La langue n'est pas fonction du sujet parlant, elle est produit que l'individu enregistre passivement [...]. La parole est au contraire un acte individuel de volonté et d'intelligence, dans lequel il convient de distinguer : 1° les combinaisons par lesquelles le sujet parlant utilise le code de la langue en vue d'exprimer sa pensée personnelle ; 2° le mécanisme psychophysique qui lui permet d'extérioriser ces combinaisons. »

Cette idée se reflète encore maintenant lorsque l'on parlera de l'*acquisition* de la langue maternelle (du langage), tandis que l'on parlera de l'*apprentissage* des langues étrangères. Le langage est une « boîte noire » que Saussure préfère laisser de côté, à la fois parce qu'il est plus difficile d'en décrypter le système (et c'est cela qui intéressera Chomsky. Voir infra), et parce que le langage peut tout autant être l'objet de la psychologie. En effet, Saussure range la linguistique dans une discipline plus générale la « Sémiologie », considérée elle aussi comme un sous-ensemble de la psychologie sociale :

« On peut donc concevoir une science qui étudie la vie des signes au sein de la vie

*sociale; elle formerait partie de la psychologie sociale, et par conséquent de la psychologie générale; nous la nommerons sémiologie (du grec sēmeion, "signe"). [...] C'est au psychologue à déterminer la place exacte de la sémiologie; la tâche du linguiste est de définir ce qui fait de la langue un système spécial dans l'ensemble des faits sémiologiques. »*¹⁶

La dichotomie langue/parole

Alors qu'il cherche à délimiter l'objet de la linguistique, Saussure veut distinguer deux éléments fondamentaux : la langue et la parole.

« L'étude du langage comporte deux parties : l'une, essentielle, a pour objet la langue, qui est sociale dans son essence et indépendante de l'individu [...]; l'autre, secondaire, a pour objet la partie individuelle du langage, c'est-à-dire la parole. [...] Le langage a un côté individuel et un côté social, et l'on ne peut concevoir l'un sans l'autre. »⁽¹⁷⁾
 « [L]e côté exécutif reste hors de cause, car l'exécution n'est jamais faite par la masse; elle est toujours individuelle, et l'individu en est toujours le maître; nous l'appellerons la parole. »¹⁸

On peut donc comparer la langue et la parole par le biais du tableau suivant :

<i>LANGUE</i>	<i>PAROLE</i>
Fait social. Cette vision est centrale dans la perspective de Saussure. Qui dit « fait social » dit « fait extérieur à l'individu »	Fait individuel.
Essentielle. La langue, en tant que fait social est acquise, et puisqu'elle est acquise, elle s'impose à l'individu.	Accessoire et accidentelle. Cela découle du fait que la parole est individuelle.
Enregistrée passivement. La langue, en tant que fait social à caractère permanent, est enregistrée passivement ; elle est acquise.	Acte de volonté et d'intelligence. La parole, de son côté, est éphémère. Elle dépend du locuteur et de sa compétence ; elle est apprise.
Psychique. On aurait dit aujourd'hui « cognitive » : la langue nous dit en effet quelque chose de l'esprit humain en général.	Psychophysique. La parole est liée à un contexte d'apparition particulier.
Somme des empreintes verbales dans chaque cerveau.	Somme de ce que les gens disent.
Modèle collectif.	Élément non collectif.

Saussure estime que c'est la langue, et non la parole, qui doit être l'objet de la linguistique. En plus de ces éléments, Saussure s'oppose à la prédominance de la représentation écrite de la langue, et

¹⁶*Cours de linguistique générale*, éd. Tullio de Mauro, p. 33

¹⁷*Ibidem*, p. 24

¹⁸*Ibidem*, p. 30

érige comme véritablement analysables les manifestations **orales** de la langue :

« Si nous pouvions embrasser la somme des **images verbales** emmagasinées chez tous les individus, nous toucherions le **lien social** qui constitue la langue. C'est un trésor déposé par la pratique de la parole dans les sujets appartenant à une même communauté, un **système** grammatical existant virtuellement dans chaque cerveau, ou plus exactement dans les cerveaux d'un ensemble d'individus; car la langue n'est complète dans aucun, elle n'existe parfaitement que dans la masse. »¹⁹

D'une certaine manière, on peut considérer que Saussure jette déjà les bases de la phonologie. En effet, la linguistique doit s'intéresser à un domaine abstrait, et cet abstrait doit être oral : il faut donc formaliser, abstraire les sons de la langue.

3.1.2.3 Le signe (linguistique)

Le signe est considéré chez Saussure comme étant une image verbale. L'ingéniosité de Saussure est de faire apparaître que ce signe est une association, ou une unité indissociable d'un élément perçu (l'image auditive, acoustique) et d'un élément non perçu (le concept). Le premier correspond à une manifestation scientifique visuelle, auditive ou odorante et le second à une idée, à un concept. Tous les deux sont d'ordre purement mental :

« 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; elle est sensorielle, et s'il nous arrive de l'appeler matérielle, c'est seulement dans ce sens par opposition à l'autre terme de l'association, le concept, généralement plus abstrait »²⁰

Le concept est ce qu'il appellera plus tard le signifié, et l'image auditive le signifiant. L'idée directrice est d'insister sur l'aspect conventionnel du signe. En effet, tout signe n'est qu'un lien entre ces deux entités, lequel lien est purement arbitraire : il est défini par la communauté sociale (conventionnel) et imposé à l'individu (arbitraire).

Le système

Saussure pose également l'existence d'une association dynamique entre ces différents signes. Les signes linguistiques entretiennent des rapports et constituent en ceci un système, une structure²¹ (cette notion sera à la base du structuralisme) :

« Tout [...] revient à dire que dans la langue il n'y a que des différences. Bien plus : une différence suppose en général des termes positifs entre lesquels elle s'établit; mais dans

¹⁹Ibidem, p. 30

²⁰Ibidem, p. 98

²¹ À distinguer *Structurel* de *Structural* :

Structurel : Par ce terme, on qualifie toute forme concrète d'organisation, directement perceptible dans la réalité. Tout ce qui relève de la pratique effective, par exemple les réalités sociales, économiques d'une région (les infrastructures). Mais aussi tout arrangement réel qui résulte de phénomènes naturels, par exemple la cristallisation de certains minéraux, les solidarités biologiques...

Structural : par ce terme on qualifie tout arrangement qui dans les langages et les signes humains, produit de la signification, par exemple l'opposition rouge/vert dans la signalisation routière. À la différence du structurel, le structural ne peut directement être repéré et expliqué. Il faut pour en rendre compte plusieurs épreuves artificielles : par exemple remplacer le rouge par le violet, le vert par le bleu, dans le signal routier, se rendre compte que bleu/violet ne veut plus rien dire. Autrement dit, pour étudier le structural, il faut le reconstruire. Dans ce cours, nous parlons de structures au sens structural et de structuralisme tant que jouera le modèle linguistique (voir J.-B. Fages, *comprendre le structuralisme*, 1968).

la langue il n'y a que des différences sans termes positifs. Qu'on prenne le signifié ou le signifiant, la langue ne comporte ni des idées ni des sons qui préexisteraient au système linguistique, mais seulement des différences conceptuelles et des différences phoniques issues de ce système. »²²

Les pensées (concepts) et les sons (phonèmes) n'existent qu'en vertu de la structure que la langue impose à notre réflexion et à notre phonation. Chaque langue amenant une autre structure, Saussure pensera à l'époque qu'il y aurait peu de chances que des traits universels existent ; au vu des découvertes sur les universaux linguistiques qui ont été réalisées par la suite, on peut sur ce point le considérer comme archaïque.

3.1.2.3.1 *Les propriétés du signe linguistique*

Les signes linguistiques se caractérisent par un ensemble de propriétés plus ou moins spécifiques :

- (1) **Ils sont arbitraires** : rien, exceptée la convention sociale, ne justifie la relation établie entre le signifiant et le signifié ;
- (2) **Ils sont linéaires** : ils se déploient successivement dans le temps et ne peuvent pas apparaître simultanément ;
- (3) **Ils sont discrets** : ils s'opposent les uns aux autres sans gradations intermédiaires ;
- (4) **Ils sont doublement articulés** : si on entreprend de segmenter la chaîne parlée, *il va à la poste*, on identifie d'abord des unités significatives, dites de première articulation, *il/va/à/la/poste*, ensuite des unités distinctives, dites de seconde articulation, *i//v/a/*, etc. L'étude des différents modes de communication non linguistiques démontre que cette dernière propriété est propre au langage humain.

3.1.2.3.2 *Le fonctionnement du signe linguistique*

Un signe ne dispose pas d'autonomie, il appartient à un tout appelé « Système ». Dans cette structure organisée, chaque élément obtient sa valeur d'un réseau de relations de réciprocité. Par exemple, le feu vert de la circulation routière ne véhicule aucune valeur en soi, mais seulement par opposition aux deux autres dans le système des feux tricolores. De la même façon, le signe linguistique n'existe qu'en opposition aux autres unités du système de la langue.

Ce système ne consiste pas en un ensemble d'étiquettes prédéfinies et disponibles pour désigner la réalité extralinguistique. Tout au contraire, c'est la langue elle-même qui structure le réel, elle donne la forme à la réalité substantielle. En effet, la langue correspond à un ensemble de conventions sociales adopté par tous les membres d'une communauté linguistique. Lorsque nous formulons des énoncés, nous nous approprions, en dépit de variations personnelles, un système de règles et de relations institutionnel (et c'est précisément ce système, la langue, qui constitue l'objet de la linguistique).

Parler de système, de grammaire, ne consiste pas à définir la langue comme un code figé et immuable. Elle est, tout au contraire, conçue comme une structure vivante en continuelle évolution. Les notions d'immutabilité et de mutabilité du signe linguistique témoignent de cette propriété. La première souligne que le lien entre le signifiant et le signifié, imposé conventionnellement, est nécessaire. La seconde reconnaît que le signe linguistique peut cependant subir des altérations liées à différents phénomènes d'évolution de la langue.

²²*Ibidem*, p. 166

3.1.3 Le caractère linéaire du langage

Dans toutes les langues, cette structure est toutefois linéaire, puisqu'une approche linéaire des choses est liée à un flux de parole nécessairement physiquement contraint dans la ligne temporelle :

« *D'une part, dans le discours, les mots contractent entre eux, en vertu de leur enchaînement, des rapports fondés sur le caractère linéaire de la langue, qui exclut la possibilité de prononcer deux éléments à la fois. [...] Ces combinaisons qui ont pour support l'étendue peuvent être appelées syntagmes.* »²³

Le syntagme²⁴ se compose toujours de deux ou plusieurs unités consécutives (p.ex. *re-lire, contre tous, la vie humaine ; Dieu est bon ; s'il fait beau, nous sortirons, etc.*). Placé dans un syntagme, un terme n'acquiert sa valeur que parce qu'il a ce qui précède ou ce qui suit, à tous les deux.

« *D'autre part, en dehors du discours, 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* ».

Ainsi le mot *enseignement* fera surgir inconsciemment devant l'esprit une foule d'autres mots (*enseigner, renseigner, etc.* ou bien *armement, changement, etc.* ou bien *éducation, apprentissage, etc.*) ; par un côté ou un autre, tous ont quelque chose de commun entre eux.

« [Ces coordinations] sont d'une toute autre espèce que les premières. Elles n'ont pas pour support l'étendue; leur siège est dans le cerveau; elles font partie de ce trésor intérieur qui constitue la langue chez chaque individu. Nous les appellerons rapports associatifs. »²⁵

Le rapport syntagmatique est *in praesentia* : il repose sur deux ou plusieurs termes également présents dans une série effective. Au contraire le rapport associatif unit des termes *in absentia* dans une série mnémorique virtuelle.

Considérant les propriétés et caractéristiques de cette structure linéaire de la langue, Saussure identifie plusieurs rapports possibles entre les signes :

- 1) Les **RAPPORTS SYNTAGMATIQUES** sont donc ceux des signes combinés sur l'étendue de la chaîne langagière, *in praesentia*. Ces rapports se constituent selon un ORDRE DÉFINI, et font toujours intervenir un NOMBRE LIMITÉ D'UNITÉS. Le rapport syntagmatique limite l'arbitraire, dans la mesure où son effet est sémantiquement plus ou moins motivé. Cependant, il n'existe que dans la mesure où les éléments qui en sont les termes ont chacun des rapports associatifs avec d'autres éléments.
- 2) Les **rapports paradigmatiques (ou associatifs)**. Ce type de rapport est le produit d'une évocation. Il s'agit de liens réalisés *in absentia*²⁶, imposés par la communauté ou en tous

²³*Ibidem*, p. 170

²⁴ La notion de syntagme s'applique non seulement aux mots, mais aux groupes de mots, aux unités complexes de toute dimension et de toute espèce (mots composés, dérivés, membres de phrase, phrases entières).

²⁵*Ibidem*, pp. 170-171

²⁶ Les relations paradigmatiques sont dites *in absentia* car elles regroupent des éléments qui ne sont pas présents simultanément dans la chaîne parlée, mais qui sont associés par l'esprit.

cas par la société dans laquelle on se trouve. Contrairement aux rapports syntagmatiques, leur ORDRE est INDÉFINI et leur NOMBRE POTENTIELLEMENT ILLIMITÉ. En effet, comme l'on peut le remarquer à partir de l'exemple suivant pour le concept « enseignement », le signe entretient de rapports paradigmatiques avec d'autres du point de vue de son signifié et/ou de son signifiant.

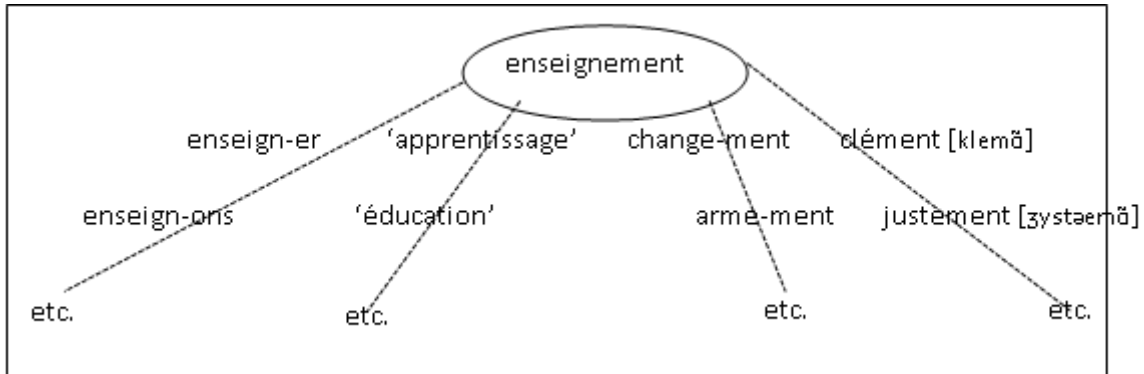


Illustration des rapports paradigmatiques que peut entretenir le signe « enseignement » du point de vue de son signifié et/ou de son signifiant.

Cette distinction entre deux types de rapports, édifiant une conception du langage à la fois linéaire et, par sa linéarité même bidimensionnel, va grandement intéresser les structuralistes à la suite de Saussure.

3.1.4 La théorie de Saussure en bref

Pour Saussure, la langue est donc avant tout un système de signes. L'objet de la linguistique est d'étudier ces signes, en tant que termes des rapports constitutifs de la structure du langage. En tant que système, elle doit être acquise comme une convention sociale. Elle n'est pas vue comme universelle, mais comme modulaire en langue d'une culture à l'autre, et modulable en parole d'un individu à l'autre.

3.1.4.1 Positionnement de la linguistique

Saussure positionne la linguistique de trois manières :

UNE LINGUISTIQUE DESCRIPTIVE. Pour Saussure, la linguistique générale doit être une linguistique descriptive. Elle doit se démarquer de la stigmatisation des erreurs et de la prescription pour faire de la description pure, sans jugement de valeur.

PRIMAUTE DE L'ORAL SUR L'ECRIT. Saussure affirme la primauté de l'oral par rapport à l'écrit. En philologie, on étudie la langue, en effet, mais on ne l'étudie pas pour elle-même et en elle-même. La philologie a pour intérêt le contenu et le commentaire de textes. Pour Saussure, étant donné que les traditions orales précèdent les traditions écrites, il va de soi de dire que l'oral prime, et que l'écrit n'est qu'une tentative de le retranscrire.

UNE BRANCHE DE LA SEMIOLOGIE. Avec Saussure, la langue sera étudiée en tant que système, et en tant que système de signes plus spécifiquement.

3.1.4.2 Langue, langage et parole

Saussure a cru nécessaire de différencier la LANGUE, objet d'étude de la linguistique, d'autres phénomènes que sont le langage et la parole. La PAROLE est la réalisation concrète de la faculté de LANGAGE. À côté du langage, on a le concept de « langue » qui est le système sous-jacent qui permet de générer la parole.

« Le langage a un côté individuel et un côté social, et l'on ne peut concevoir l'un sans l'autre. »²⁷

En effet, langue et parole ne s'opposent pas comme deux aspects irréductibles du langage, elles entretiennent des rapports de compréhension réciproque. Il s'agit de deux objets qui sont étroitement liés et se supposent l'un l'autre :

- 1) La langue est nécessaire pour que la parole soit intelligible et produise des effets ; mais celle-ci est nécessaire pour que la langue s'établisse. En effet, les actes de parole précèdent la langue et la bâtissent peu à peu, mais ces mêmes actes de parole vont puiser dans la réserve commune et obéissent aux règles collectives, (cette réserve et ces règles, formant la langue) si bien que la langue précède, d'un autre point de vue, la parole.
- 2) C'est en entendant les autres que nous apprenons notre langue maternelle ; elle n'arrive à se déposer dans notre cerveau qu'à la suite d'innombrables expériences.
- 3) C'est la parole qui fait évoluer la langue : ce sont les impressions reçues en entendant les autres qui modifient nos habitudes linguistiques.

La langue représente l'aspect social et codifié du langage. Elle est institution sociale résultant d'un vaste contrat, d'un ensemble de conventions entre les hommes. L'individu qui parle et veut se faire comprendre ne peut modifier cette convention collective, il doit au contraire s'y conformer. La langue est commune à tous les membres d'une même communauté. En termes simples, on peut dire que la langue existe dans la collectivité sous la forme d'empreintes déposées dans chaque cerveau, à peu près comme un dictionnaire dont tous les exemplaires, identiques, seraient répartis entre les individus. C'est donc quelque chose qui est dans chacun d'eux, tout en étant commun à tous et placé en dehors de la volonté des dépositaires.

Un système de langue ne peut fonctionner qu'à l'intérieur d'une communauté de personnes qui le partagent.

La parole est un acte individuel, elle est le fait concret d'un homme qui parle à son semblable. Elle dispose d'une certaine liberté d'expression ou de variation, car chaque individu a sa manière spéciale de manier les mots et les phrases. Mais cette liberté reste étroitement surveillée : elle combine des signes, bien qu'elle ne les crée, elle les combine sur la base du code commun, sur la base de la langue.

Saussure décide donc que ce qui doit constituer l'objet de la linguistique n'est pas la parole mais la langue. On considère souvent l'approche Saussurienne comme dénigrant la parole, mais cette idée n'est pas l'exact reflet de la pensée de Saussure. En fait, la parole joue pour Saussure un autre rôle. On remarque ici que Saussure donne à la parole un rôle très important, et ce en vertu de trois raisons énoncées ci-avant et reprises ci-après :

- i) La langue, premièrement, ne peut être observée directement.
- ii) De plus, on se contente de dire que la parole est un objet d'étude secondaire, mais c'est elle qui conditionne le changement linguistique.

²⁷*Ibidem*, p. 24

- iii) Enfin, un enfant n'acquiert sa langue qu'en étant confronté à la parole, à l'usage de sa langue maternelle.

3.1.4.3 La langue comme système de signes

Le système de la langue est conçu par Saussure comme étant un système de signes, défini par une entité abstraite ayant deux faces : l'une signifiante (représentation mentale d'une suite de sons) et l'autre signifiée (représentation mentale d'une idée ou d'une chose). Un grand débat se déroule chez les théoriciens pour savoir quel est exactement ce signifié. Le signifiant, s'il est plus « palpable » n'est pas pour autant simple à appréhender. Néanmoins, il est clair que l'union des deux forme le signe linguistique.

Le signe linguistique est de nature linéaire et arbitraire. La linéarité du langage, d'une part, signifie qu'on produit les sons les uns à la suite des autres, sans pouvoir signifier deux choses à la fois. Son aspect arbitraire, d'autre part, réside dans le fait qu'aucune raison objective ne peut expliquer qu'un signifiant renvoie à un concept. Le lien d'un signifiant et d'un signifié n'est pas naturel, et échappe à toute explication logique : on aura toujours que des signes arbitraires fonctionnant à l'intérieur d'une communauté bien précise.

Saussure insiste également sur le caractère systémique du signe :

« C'est une grande illusion de considérer un terme simplement comme l'union d'un certain son avec un certain concept. Le définir ainsi, ce serait l'isoler du système dont il fait partie; ce serait croire qu'on peut commencer par les termes et construire le système en en faisant la somme, alors qu'au contraire c'est du tout solidaire qu'il faut partir pour obtenir par analyse les éléments qu'il renferme. »²⁸

Tout signe ne reçoit donc son sens qu'à l'intérieur d'un système. Pris individuellement, il n'a pas de sens. Ainsi, le mot « table » ne prend son sens que par rapport à « bureau », « guéridon », etc. Les rapports existant entre ces signes à l'intérieur du système sont les fameux rapports syntagmatiques et paradigmatiques : on étudie les rapports qui existent entre les différents éléments d'une suite linéaire, et l'opposition d'un élément par rapport à un autre.

3.1.4.4 Synchronie et diachronie

Les rapports entre les signes peuvent être étudiés en considérant d'autres dimensions que les aspects associatifs et syntagmatiques. On peut également considérer la dimension temporelle. Tout comme la linéarité du signifiant conditionne une bidimensionnalité des rapports entre les signes, la linéarité du temps conditionne une bidimensionnalité du système :

La **SYNCHRONIE** étudie un système dans ses rapports à lui-même à un moment donné du temps.

La **DIACHRONIE** étudie un système dans ses rapports à sa propre évolution au cours du temps.

Saussure déclare que la seule manière d'étudier véritablement le système en tant que tel est de l'observer en synchronie. Il utilise pour cela la métaphore de la partie d'échecs. Admettons que des personnes soient en train d'observer une partie d'échecs ; si une tierce personne arrive au

²⁸*Ibidem*, pp. 157.

milieu de la partie et demande ce qui se passe, les observateurs vont décrire la situation actuelle, et non retracer l'histoire des coups précédents. L'important est de dire quelque chose sur la partie au moment où elle en est : la description synchronique est la plus importante. Évidemment, même s'il n'est pas capital, cela n'empêche pas de considérer le point de vue diachronique, et d'expliquer les stades qui ont pu amener d'une situation passée à la situation présente.

Saussure veut se démarquer de la diachronie qui primait chez les néogrammairiens. Ce qui doit primer, c'est la synchronie, parce qu'expliquer ce qui s'est passé auparavant n'est pas réellement un *besoin* scientifique. Pour Saussure, par ailleurs, la diachronie n'est rien de plus qu'une succession de synchronies. L'évolution de la langue, et en particulier l'évolution phonétique qui était chère aux néogrammairiens est accessoire. Les phonèmes eux-mêmes n'existent qu'en fonction d'une structure que la langue impose à notre réflexion et à notre phonation. Elle va être leur grille de lecture.

3.1.4.5 Qu'entendre du système linguistique?

La linguistique est généralement définie comme l'étude scientifique du langage. Mais cette discipline, qui part donc de l'étude du langage, en arrive surtout à s'occuper des langues. Elle a comme objet spécifique (visé en dernière analyse) la langue, et a comme objet étudié (qu'il faut connaître au départ) le langage. La linguistique étudie principalement et, marginalement, le langage, instrument de communication par manifestation vocale. Il importe de remarquer dès lors que deux conceptions de la linguistique coexistent :

La linguistique comme description des langues : elle s'appuie sur l'observation objective du comportement linguistique des sujets parlants ; elle décrit tout ce qui caractérise en propre une langue reconnue comme telle, considérant que l'étude d'un état de langue peut avoir une valeur explicative, et pas seulement descriptive. Se proposant comme idéal les méthodes des sciences de la nature et l'objectivité du physicien, elle considère une langue comme un système de signes linguistiques.

La linguistique comme l'étude du fonctionnement du langage : elle s'appuie sur une conception unitaire du langage humain, considérant que les langues particulières ne sont que des cas particuliers du langage. Les langues diffèrent les unes des autres, non dans leurs structures profondes, mais du fait des variations tout à fait superficielles ; une analyse approfondie ferait apparaître des universaux linguistiques qui témoignent d'une faculté naturelle du langage. Il y aurait ainsi un répertoire universel d'éléments sémantiques ou phonétiques dans lequel chaque langue choisirait les éléments de base de sa combinatoire. À la notion de langue comme système de signes est substituée la notion de langue comme système de règles. D'où l'on pourra définir la linguistique comme l'étude scientifique du langage humain, vocal, tel qu'il se réalise par une langue, système de signes linguistiques et/ou système de règles (voir Christian BAYLON & Paul FABRE, 1975, p17).

3.2 Le distributionnalisme (Bloomfield [1887-1949])

Le distributionnalisme peut être vu comme marquant une succession au structuralisme. Bloomfield était germaniste ; il étudie l'allemand à Harvard et entre ainsi en contact avec les langues européennes et les théories structuralistes. Il réalise une thèse en linguistique historique : *A Semasiologic differentiation in Germanic secondary Ablaut*. Il va ensuite se consacrer à l'application de ces idées de la linguistique historique à la description des langues

amérindiennes, en particulier à la famille algonquienne. Bloomfield est surtout le premier à décrire les comportements linguistiques à travers les dernières avancées de la psychologie (*An introduction to the study of language* [1914] *Language* [1933]). En appliquant cette démarche, il va faire évoluer la théorie dans une série de directions qui vont marquer la linguistique.

3.2.1 Structuralisme européen vs structuralisme américain

La linguistique structurale européenne s'est développée, on le sait, à partir du *Cours de Linguistique générale*, de F. de Saussure, professé de 1906 à 1911 et publié, par les soins de Ch. Bally et A. Séchehaye, en 1916, et à partir des Travaux du *Cercle linguistique de Prague*, fondé en 1926 et dont les thèses, œuvre collective de V. Mathesius, de R. Jakobson et N. Troubetskoï, paraissent en 1929. Elle s'est diversifiée alors en plusieurs courants, comme *l'École de Copenhague* (V. Brendal et L. Hjelmslev) ou *le fonctionnalisme* (R. Jakobson, N. Troubetskoï, A. Martinet et A. K. Halliday); ceux-ci, tout en se distinguant par l'accent mis sur tel ou tel aspect de la description (phonologie ou morphologie), par l'analyse des rapports entre les unités des divers rangs (phonème, morphème, syntagme, phrase) ou des relations entre les plans formel et sémantique (isomorphisme ou non des deux plans, identité ou non des méthodes d'analyse), par la prévalence d'une méthode inductive empirique ou d'une méthode déductive axiomatique, impliquent en réalité tous le même modèle linguistique fondamental : le langage naturel est constitué d'ensembles d'objets arbitraires; ces ensembles sont organisés en des systèmes qui sont définis par les seules relations que ces objets entretiennent entre eux; *la construction de ce modèle se fait à partir d'un ensemble de données effectives, collectées et réunies sous le terme de corpus.*

Aux États-Unis, si E. Sapir (*Language*, 1921), dont les hypothèses sont proches des Européens, reste isolé, l'ouvrage décisif est celui de L. Bloomfield, *Language*, en 1933. Suivant la fondation en 1924 de la *Société de Linguistique américaine* et la création de la revue *Language* en 1929, il oriente la linguistique structurale dans une direction différente, marquée par un double développement. D'un côté s'élaborent :

- (1) **la méthode « distributionnelle »**, qui consiste à définir avec rigueur une méthode formelle de segmentation de la chaîne parlée en unités distinctives, définies par les seules relations qu'elles entretiennent dans cette chaîne, c'est-à-dire par leur environnement, et
- (2) **la méthode des « constituants immédiats »** qui définit une analyse combinatoire de ces unités en allant de l'unité minimale aux unités de rang supérieur (du morphème à l'énoncé, du phonème au morphème).

D'un autre côté, se précise l'hypothèse fondamentale que la ou les structures découvertes par ces méthodes sont dans les objets « physiques » eux-mêmes, dans les phénomènes linguistiques ; d'où l'accent mis sur le rapport étroit entre les structures de la langue et les structures physiques, par exemple entre l'unité phonique et sa réalisation physique.

3.2.2 L'approche behavioriste et son application à la linguistique

Le point de départ que Bloomfield va prendre sera de considérer des idées en vogue à l'époque avec le behaviorisme de *Skinner* : selon lui, le comportement humain peut être schématisé en une suite de stimulus-réponse. Il considère ainsi que la langue fonctionne selon les mêmes principes de stimulus et de réponse que le comportement. On peut de cette manière considérer la langue comme une simple réponse à un stimulus reçu. Dans cette perspective, la tâche du linguiste sera la description des rapports entre stimulus linguistique et réponse linguistique.

Prenons par exemple le contexte de la cueillette d'une pomme. La perception et l'envie de la pomme va produire sur le sujet un désir de la cueillir ; c'est là qu'intervient le langage, si par

exemple A (ici Jill) désire la pomme sans pouvoir la cueillir, tandis que B (Jack), lui, le peut et peut coopérer avec Jill par la voie du langage :

I. Behavior

L'organisme Jill	stimulus	Faim ; perception de la pomme.
	réaction	cueillir la pomme.

enchaînement: $S \rightarrow R$

II. Linguistic Behavior

L'organisme Jill	stimulus	Faim ; perception de la pomme.	} <i>Rôle du langage</i>
	réaction	Vibration des cordes vocales ; Configurations successives de la bouche.	
L'organisme Jack	stimulus	Vibration des tympan.	
	réaction	Cueillir la pomme ; Donner la pomme à Jill.	

enchaînement: $S \rightarrow r \rightarrow s \rightarrow R$

Ce rapport stimulus-réponse des objets aux individus peut être modélisé, mais on ne peut en avoir une approche purement théorique. On va analyser les données dans leur milieu, dans une démarche inductive, constituer des documents, et en tirer des conclusions. Il s'agit de recueillir des données, autrement dit un échantillon représentatif d'« individus » qui sont susceptibles de représenter la communauté dans son ensemble.

En bref, on constitue un corpus pour décrire et classer les données. De là, on tire des catégories pour enfin former ce que l'on va appeler une taxonomie. On s'attache particulièrement à l'environnement dans lequel les différents faits linguistiques ont pu être observés. On cherche à décrire les contextes plutôt que le sens : l'approche behavioriste ne permet pas d'accéder à la sémantique des énoncés.

3.2.2.1 Démarche du linguiste

Dans une perspective behavioriste, l'analyse de la capacité linguistique (« linguistic behavior ») doit donc être constituée de quatre étapes :

- (i) L'observation
- (ii) La description des données observées
- (iii) La catégorisation de ces données
- (iv) L'établissement de la distribution, consistant en une liste complète des contextes.

3.2.2.2 L'analyse en constituants immédiats

De cette approche behavioriste, Bloomfield va instituer une nouvelle manière de regarder le langage : *l'analyse en constituants immédiats*, encore très porteuse aujourd'hui. Cette analyse permet de considérer n'importe quelle langue sur le terrain ; sans connaissance de la langue, elle demeure toujours le meilleur outil dont le linguiste puisse disposer. Les énoncés vont être disséqués en des éléments de plus en plus petits, pour en arriver aux constituants minimaux.

Pour étudier la distribution, on procède à une étude du contexte phrastique. Chaque mot est vu comme la réponse au stimulus produit par les autres mots de la phrase. On vérifie par conséquent ces stimuli au moyen de divers tests, parmi lesquels deux tests peuvent véritablement être universels :

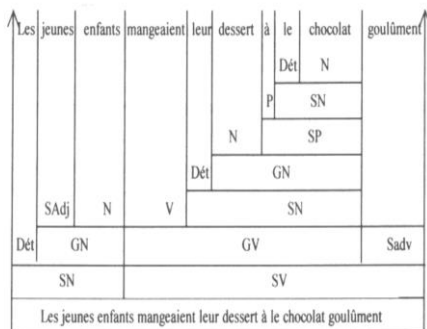
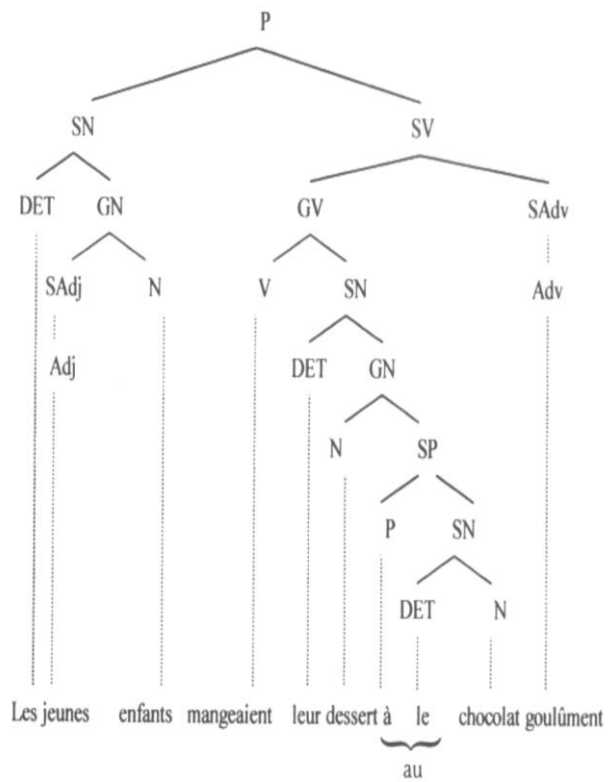
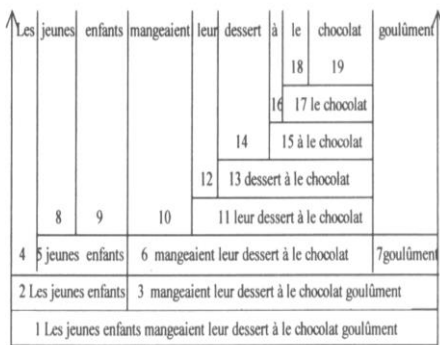
- Le test de commutation *Un chat très intelligent* → *un chat assez intelligent*

On sait grâce à ce test que « très » et « assez » font partie de la même classe grammaticale. Mais si on prend le mot « tôt », on constate que cela ne fonctionne absolument pas : cela doit faire partie d’une autre catégorie. Il s’agit d’un test de substitution : on remplace un élément par un autre pour savoir si la phrase est correcte.

- Le test de combinaison *Un joli chapeau* / *Le joli chapeau* → **Un le joli chapeau*

La combinaison nous donne une indication sur ce qui se passe à l’intérieur d’une classe distributionnelle donnée, comme ici celle des déterminants ; un seul déterminant peut exister à un endroit donné.

Ces types de tests sont également utilisés dans d’autres disciplines de la linguistique. Ils sont par exemple utilisés en phonologie, si par exemple on cherche à savoir si deux sons sont les réalisations d’un même phonème. Ces tests, réalisés sur une langue dans un grand nombre de contextes, permettent de formaliser les énoncés d’une langue. En aval de ces simples tests, on pourra donc en arriver à des modélisations complexes (ci-dessous, la représentation en tiroirs et en arbre syntaxique).



3.2.3 Les propriétés du langage : les principes de base de la théorie distributionnaliste

La formalisation des langues naturelles implique la reconnaissance d'un certain nombre de propriétés du langage, qui constituent les principes de base de la théorie distributionnelle.

3.2.3.1 Le caractère discret des éléments

Les éléments qui entrent dans la structure linguistique ont pour propriété essentielle d'être des segments discrets. Dans le continuum physique que constitue le flux verbal, l'analyse linguistique reconnaît des éléments discrets.

Ce caractère discret des éléments, qui est la condition de la segmentation de la chaîne parlée et de l'établissement de la structure de la langue, écarte donc de l'analyse linguistique un ensemble d'informations acoustiques, liées au flux verbal (hésitations, pauses, variations de timbre ou d'intensité, gestes, etc.) qui se modifient selon l'âge, le sexe, la fatigue ou la personnalité, etc. : ces informations, qui ne sont pas représentées par un système de segments discrets, n'entretiennent pas de relations grammaticales avec les éléments de la structure linguistique.

En revanche, la théorie distributionnelle intègre à son modèle grammatical les traits prosodiques que sont les accents, les hauteurs, les contours d'intonation (interrogation, emphatique, impératif, etc.) : ceux-ci sont analysés comme des segments discrets.

3.2.3.2 L'antériorité et la stabilité de la structure grammaticale

Les éléments discrets ainsi définis sont communs à l'émetteur et au récepteur; ils leur sont antérieurs et, en quelque sorte, préétablis. Le locuteur et l'auditeur se réfèrent tous deux dans la communication à un ensemble d'objets linguistiques qui leur sont communs.

Cette propriété rend compte, par exemple, dans le fonctionnement du langage, du fait que le sujet parlant, lorsqu'il a à répéter un énoncé, le « reproduit », c'est-à-dire qu'il l'interprète par rapport aux éléments grammaticaux qu'il a intériorisés et qu'il le reconstitue selon les particularités de son propre système ; il ne l'imité pas, c'est-à-dire qu'il ne le redonne pas comme une copie conforme de ce qu'il a entendu.

Cette intériorisation des éléments grammaticaux se fait par l'apprentissage : les objets linguistiques agissent comme un ensemble de stimuli permettant chez le sujet la reconnaissance de certains contrastes et la constitution d'une structure grammaticale. Cette structure étant propre à l'objet que forme la langue est donc d'abord extérieure au sujet. Ce qui est seulement donné à la naissance, c'est la spécificité des structures neuronales des hémisphères cérébraux, propres à l'espèce humaine et capables de sous-tendre les fonctionnements complexes de structures composées d'éléments discrets. Ce qui est appris, ce sont les éléments grammaticaux, les structures linguistiques elles-mêmes.

Ce caractère antérieur au sujet et préétabli rend compte, avec l'arbitraire des symboles verbaux, de la stabilité des éléments discrets; ceux-ci sont relativement indépendants des distorsions que peuvent leur faire subir les locuteurs dans la mesure où ils doivent rester le lieu commun de référence qui, seul, permet la transmissibilité des messages.

3.2.3.3 Le caractère arbitraire des éléments grammaticaux.

Le caractère arbitraire des éléments discrets est impliqué par la nécessité de la transmission sans erreur. Les sons qui composent les morphèmes ne suggèrent pas le sens de ces morphèmes; il existe un lien arbitraire et conventionnel entre les symboles verbaux et les objets.

Dans le cas où un lien naturel existerait entre le son et le sens des mots, l'instabilité du

système serait grande : la transmission sans erreur ne pourrait se faire puisque le locuteur et l'auditeur ne seraient pas nécessairement d'accord sur ce lien ; les sens et les sons varieraient en fonction des expériences individuelles et le système ne pourrait pas être intériorisé chez des locuteurs différents, ni rester le même pendant un temps assez long.

3.2.3.4 . Le caractère linéaire des combinaisons d'éléments.

Les énoncés sont des suites d'éléments discrets ordonnés de façon linéaire. Chaque morphème est une suite de phonèmes, chaque syntagme une suite de morphèmes, chaque phrase une suite de syntagmes, chaque discours une suite de phrases. Lorsque les éléments paraissent se chevaucher ou que l'étendue de certains items paraît plus grande que d'autres, on peut toujours obtenir une représentation linéaire, conforme à la succession linéaire des événements linguistiques. Ainsi les phonèmes et les morphèmes suprasegmentaux d'intonation, d'accent et de hauteur sont simultanés aux phonèmes et aux morphèmes segmentaux.

3.2.3.5 Le caractère fini des combinaisons.

Les éléments grammaticaux d'une langue, les classifications et les opérations grammaticales forment un ensemble fini. Certes l'application des règles peut être récursive, mais les opérateurs utilisés dans les règles de combinaison sont en nombre fini.

Les phrases sont donc toujours des suites finies d'éléments discrets. Il est en effet impossible de dire qu'une phrase est grammaticale et acceptable, c'est-à-dire qu'une suite est une phrase de la langue, avant qu'elle ne soit finie, car il peut toujours intervenir avant son achèvement un élément qui pourrait violer telle ou telle régularité.

Si chaque phrase est finie, l'ensemble des phrases est infini, mais dénombrable ; infini puisque l'on peut toujours ajouter un élément dans une phrase pour en former une autre (ainsi un adjectif peut être adjoint à un syntagme nominal), de la même manière que l'on peut former un nouveau nombre en ajoutant 1 ; mais cet ensemble est dénombrable, puisque cette phrase est alors une occurrence d'un type de phrase.

3.2.3.6 Les restrictions de combinaisons

Toutes les combinaisons des éléments linguistiques ne sont pas possibles dans une langue donnée. Ceci veut dire que toutes les suites d'éléments ne constituent pas un discours.

En effet, si toute suite de phonèmes donnait un morphème et si toute suite de morphèmes (ou toute suite de syntagmes) donnait une phrase, etc., il n'y aurait aucun moyen de déterminer les relations entretenues par les éléments à l'intérieur du morphème ou de la phrase, ni de fixer leurs frontières (et, par conséquent, leur signification). Cette détermination ne peut pas se faire seulement par la dénotation (désignation des objets dénotés) ou par la mimique (ceci étant vrai en particulier pour les morphèmes grammaticaux) ni par la référence à une autre langue.

Toutes les combinaisons n'étant pas possibles, on peut donc définir des éléments grammaticaux, comme les morphèmes, par les restrictions qui sont imposées aux éléments qui les composent (les phonèmes) et on peut définir les phrases par les restrictions qui sont imposées aux éléments (morphèmes ou suites de morphèmes) qui les composent.

On peut donc réunir les morphèmes en des classes définies par le seul critère de leurs combinaisons dans des suites de morphèmes (syntagmes). De la même manière, on définira des classes de syntagmes en constatant que toute suite de morphèmes n'entre pas dans une phrase. Ce processus de classification permet de définir des classes de phonèmes, de morphèmes, de

syntagmes et de phrases par les restrictions aux combinaisons possibles. La question la plus importante de la linguistique structurale est de savoir en effet :

- (i) comment on distinguera les suites de phonèmes, de syntagmes ou de phrases acceptables et celles qui ne le sont pas ;
- (ii) comment on trouvera des régularités dans les suites d'éléments discrets constituant le discours ;
- (iii) comment on pourra déterminer dans l'ensemble infini et dénombrable des phrases les combinaisons et les opérations grammaticales récurrentes et finies qui forment la structure grammaticale.

3.2.3.7 Les structures de phrases.

Ces régularités ne peuvent pas être déterminées sur la seule base de la transition entre un élément A et un élément B. On ne pourra pas définir des classes à partir des transitions des éléments successifs d'une phrase, car on constate souvent que les phrases peuvent avoir plusieurs sens, c'est-à-dire qu'elles peuvent être ambiguës.

Ainsi une phrase comme *J'ai rapporté un vase du Japon* a deux sens au moins : j'ai rapporté un vase, celui-ci vient du Japon (et moi-même je peux l'avoir rapporté d'un autre pays), ou bien j'ai rapporté du Japon ce vase, qui peut être un vase de Chine.

Cette ambiguïté, dont on ne saurait rendre compte par la seule transition entre des morphèmes qui ont toujours le même sens, est liée à certains types de groupements de morphèmes.

On partira de la constatation empirique qu'il y a dans les suites de syntagmes ou de morphèmes des régularités qui permettent de définir des structures de phrase. Dans un discours quelconque, les parties que l'on découvre alors sont des occurrences particulières d'un certain type de suite bien formée de morphèmes ou de syntagmes. Toute langue comporte ainsi un nombre restreint de structures de phrase.

Il s'agit donc pour le linguiste distributionnaliste de définir les suites caractéristiques des classes d'éléments qui constituent les structures de phrase de la langue par opposition à celles qui ne la constituent pas.

Ainsi on peut définir une classe N des noms, V des verbes, P des prépositions, etc. Toute suite NVN constitue, par exemple, une structure de phrase, mais une suite VPP ne constitue pas en français une suite acceptable. Toutefois toute suite VN, par exemple, n'entre pas dans une phrase acceptable; il faudra donc constituer des sous-classes de V, la classe Vt (transitif), la combinaison VtN étant acceptable, et la classe Vi (intransitif), la combinaison ViN étant non acceptable.

3.2.3.8 La relation de contiguïté

La seule relation existant entre deux morphèmes ou deux suites de morphèmes est la contiguïté, c'est-à-dire la proximité immédiate de deux morphèmes ou suites de morphèmes (NVN par exemple).

Il n'est pas en effet possible de définir et de mesurer un espace en linguistique, comme on peut le définir en musique. Toute suite bien formée de morphèmes est constituée de mor-

phèmes reliés entre eux par la seule relation de contiguïté. Ainsi SN est formé de D + N (déterminant + nom). Si nous avons D + A -f N (déterminant -f- adjectif -f nom), nous avons une autre suite.

La seule propriété qui fait que cette suite est un objet de la grammaire est celle d'être elle-même constituée de mots appartenant à certaines classes spécifiques. Il n'y a pas d'espace mesurable entre les éléments constitutifs d'une suite.

3.3.3.9. Le changement linguistique comme propriété du langage.

Le changement est une propriété de toute langue naturelle vivante. Toutes les langues se modifient avec le temps. Et le problème qui se pose à la linguistique distributionnelle est de savoir de quelle manière elle va traiter de ce changement.

Or, les modifications que présentent une langue au cours du temps sont continues, en ce sens que le changement est une sorte de continuum sur lequel la description d'un état de langue va définir un segment.

On peut en effet décrire la langue à un moment t_1 de son fonctionnement, en établir les classes et les opérations, et on peut déterminer un temps t_2 , antérieur au temps t_1 et procéder à la même description sur des textes qui seront des échantillons de la langue considérée (cf. la bi-dimensionnalité temporelle de F. de Saussure). On procède ainsi à une série de segmentations arbitraires sur la ligne continue du changement.

Supposons que nous ayons ainsi deux grammaires G_1 de t_1 et G_2 de t_2 , tel que t_2 soit confondu avec le moment présent et que t_1 lui soit antérieur (par exemple le français du 17^e siècle). En ce cas on pourra comparer les grammaires G_1 et G_2 au regard d'un certain nombre d'items, ces items étant des classes de morphèmes, des classes de suites ou des opérations affectant les types de phrases.

Le changement pourra alors être décrit de la manière suivante : la grammaire G_2 antérieure à G_1 est valable au temps t_2 ; au temps t_1 (moment présent) le changement peut être défini soit en établissant la grammaire G_2 dont tous les traits rendent compte des caractéristiques de la langue en t_2 , soit en considérant dans la grammaire G_1 les traits de G_2 qui sont encore valides pour décrire la langue. En ce sens, on ne considère pas seulement la différence entre les deux grammaires G_1 et G_2 , mais on rend compte aussi de la continuité du changement tout en procédant à des coupes synchroniques dans une langue qui a une histoire.

3.3 Le générativisme (Noam Chomsky)

Chez Saussure, on essayait de comprendre la langue en tant que système, en désirant garder ses distances par rapport à la parole et la production individuelle. Chez Bloomfield, on trouve également cette notion de système, mais cette fois complétée par une dimension psychologique imprégnée de behaviorisme. Faisant progresser la linguistique vers davantage de cognition, Noam Chomsky va se placer dans la continuation de Saussure et de Bloomfield en cherchant encore davantage à percer les secrets du siège de la langue : le cerveau humain.

Après avoir été éduqué dans une école à pédagogie alternative, Chomsky entre à l'université de Pennsylvanie où il étudie parallèlement la philosophie et la linguistique auprès du distributionnaliste Zellig Harris, disciple de Bloomfield. Il rédige un mémoire sur la langue hébraïque *Morphophonemics of modern Hebrew* (1951). Imprégné par la pensée distributionnaliste ; il commençait toutefois à développer des idées en marge de la théorie de Bloomfield. Doctorant, il est engagé au M.I.T et travaille avec d'autres chercheurs sur la conception d'une « machine à traduire », avant la révolution informatique.

Dans sa thèse, *Transformational analysis*, il initiera une approche logico-mathématique du langage, hautement formelle, inaugurant les analyses transformationnelles qui vont ouvrir la voie à une nouvelle compréhension de la syntaxe. Il approfondira par la suite ses théories dans deux ouvrages qui vont profondément marquer la linguistique : *Syntactic Structures* (1957) et *Aspects of the Theory of Syntax* (1965). Par après, son approche éclatera en une multitude de variantes, et Chomsky lui-même sera maintes fois amené à revoir certaines conjectures trop ambitieuses ; nous ne présenterons ici par conséquent que les grandes lignes de sa théorie.

3.3.1 Critique du behaviorisme

Déjà en tant que distributionnaliste, Chomsky réagit face à ce que faisait Bloomfield. Celui-ci désirait fournir une classification (taxonomie) des unités de langue, basée sur la seule observation d'un corpus, en excluant l'intuition linguistique.

Chomsky va critiquer cet empirisme de l'approche behavioriste, soulignant qu'il serait impossible de rassembler un corpus contenant toutes les expressions d'une langue donnée. Mais Chomsky a d'autres ambitions : ne plus s'intéresser à la langue en tant que telle, mais au système sous-jacent qui a permis de produire des énoncés. Il veut produire une « grammaire » dite **générative et transformationnelle** (GGT), qui serait une machine à produire les énoncés. Cette grammaire serait faite d'un ensemble de règles abstraites, à l'instar d'un code informatique.

Par ce fait, Chomsky ne part pas de l'observation pour baser sa théorie : il va d'abord formuler une grammaire pour parvenir à la tester sur les faits. On quitte alors la démarche inductive de Bloomfield pour se trouver face à une approche hypothético-déductive dans laquelle on part de la théorie pour en arriver aux faits.

La première publication scientifique de Chomsky sera ainsi le compte-rendu critique d'un livre, *Verbal behavior* de Skinner, ouvrage fondateur du behaviorisme. Il y attaque durement les conceptions behavioristes en les dénonçant comme « simplistes ». Plusieurs arguments étayaient cette affirmation :

- Une théorie qui considérerait l'acquisition du langage selon le simple processus stimulus-réponse ne peut pas raisonnablement expliquer le langage. En s'inscrivant dans la

logique du behaviorisme, un enfant, pour acquérir sa langue maternelle, devrait produire uniquement ce à quoi il a été confronté, ne faisant que répéter des stimuli issus de son expérience. Or, il s'avère bel et bien que tout être humain est en mesure de produire et de comprendre un nombre illimité de structures même s'il ne les a jamais entendues comme telles par le passé.

- Dans une perspective behavioriste, les parents devraient stimuler la faculté linguistique d'un enfant en l'exposant à des énoncés les mieux formés possible. Or, en analysant le comportement des parents envers leur enfant, on s'est aperçu de l'exact contraire : l'enfant recevait des énoncés exprimés dans une langue appauvrie. Si on applique la théorie de Skinner, Proust aurait d'abord nécessairement *entendu* les phrases complexes qu'il livre dans *À la recherche du temps perdu*. Or, les stimuli envoyés par les parents à l'enfant qu'a été Proust ont certainement été bien plus pauvres.

3.3.2 Le programme générativiste

3.3.2.1 Distinction compétence/performance

Chomsky établit une distinction entre la **compétence** d'un système sous-jacent, et la **performance**, qui est la pratique des locuteurs. Cette distinction n'est pas novatrice, car elle se rapproche grandement de la distinction langue/parole déjà opérée par F. de Saussure. Toutefois, cette distinction est ici remise dans la vision mentaliste prônée par Chomsky.

« Le problème, pour le linguiste aussi bien que pour l'enfant qui apprend la langue, consiste en ceci : déterminer à partir des données de la performance, le système sous-jacent de règles qui a été maîtrisé par le locuteur-auditeur et qu'il met en usage dans sa performance effective. De ce fait, la théorie linguistique est mentaliste, au sens technique de ce mot, puisqu'elle s'attache à découvrir une réalité mentale sous-jacente au comportement effectif. » (Chomsky, 1965 : 13).

En un sens, on retrouve le caractère de la langue que Saussure désignait comme « psychique » (par rapport à une parole « psychophysique »), qui est ici approfondi et trouve un nouvel écho. Pour Chomsky, à l'instar de Saussure, la linguistique ne doit pas s'intéresser à la performance. Si elle est l'illustration concrète du langage, la performance demeure en effet imparfaite, jamais produite dans des conditions idéales où l'on aurait directement accès à l'activité cérébrale « nue » du locuteur/auditeur.

3.3.2.2 Une linguistique de la compétence

On cherche donc à mettre à nu cette compétence linguistique, à voir clair dans les rouages de cette machine à produire les énoncés qu'est la grammaire. Cette « **grammaire** », compétence du locuteur, est internalisée par l'individu sous la forme d'un **ensemble fini de règles**, permettant de produire un **nombre virtuellement infini de phrases**.

Mais la compétence linguistique que cette grammaire représente n'est cependant pas observable directement. Le linguiste générativiste se trouve face à une « boîte noire », devant laquelle s'étaient arrêtés les behavioristes. Chomsky parvient toutefois à franchir cette barrière en rejetant la démarche inductive, et en mettant en place une démarche introspective,

hypothético-déductive. Le linguiste est reconnu dans sa faculté de produire lui-même des énoncés et d'en tester la grammaticalité, sans recourir à un assemblage de données collectées, qu'il ne pourrait maîtriser qu'extérieurement.

3.3.2.3 L'élaboration d'une grammaire générative

Pour Chomsky, « il est évident que l'ensemble des phrases grammaticales ne peut être assimilé à aucun corpus particulier d'énoncés, recueilli par le linguiste dans son travail d'enquête. La grammaire d'une langue doit **projeter** le corpus fini et toujours plus ou moins accidentel des énoncés observés. » (1957 :17).

La démarche générativiste va être de rechercher et d'appliquer des règles construites, en produisant une grande quantité de données (« output »). L'important n'est pas, dans un premier temps, de savoir si l'énoncé produit fait sens. On va par la suite présenter ces données à des évaluateurs, eux-mêmes locuteurs natifs de la langue, et leur demander de porter un jugement sur la grammaticalité des énoncés.

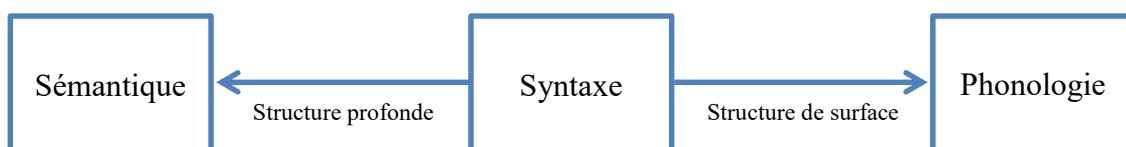
Chomsky veut donc élaborer une linguistique libérée de l'analyse d'un contexte, touchant directement au but : l'analyse de l'aspect formel des énoncés. Pour dégager ses règles d'énoncés, Chomsky prend donc le pari de travailler avec la représentation d'un locuteur natif idéal, possédant un « *mental organ* » parfaitement fonctionnel, et vivant dans une communauté linguistique virtuelle qui serait parfaitement homogène. Pour les commodités de l'analyse formelle, il n'est pas distrait, termine ses phrases et ne rencontre pas de problèmes de mémoire :

« L'objet premier de la théorie linguistique est un **locuteur-auditeur idéal**, appartenant à une communauté linguistique complètement homogène, qui connaît parfaitement sa langue et qui, lorsqu'il applique en une performance effective sa connaissance de la langue, n'est pas affecté par des conditions grammaticalement non pertinentes, telles que limitation de mémoire, distractions, déplacements d'intérêt ou d'attention, erreurs (fortuites ou caractéristiques). » (Chomsky 1965 :12).

3.3.3 La grammaire générative

3.3.3.1 La structure

Selon Chomsky, « *une grammaire générative doit être un système de règles qui peuvent être itérées pour engendrer un nombre indéfiniment grand de structures. Ce système de règles peut être analysé en trois parties, qui sont les trois composants principaux d'une grammaire générative : les composants syntaxique, phonologique et sémantique.* » (Chomsky, 1965 : 31).



La GGT que décrit Chomsky est composée de trois modules agissant de concert pour produire des énoncés : la sémantique, la syntaxe et la phonologie.

Dans la conception générativiste, la grammaire d'une langue n'est pas un dispositif tout d'une pièce qui engendre directement l'ensemble des descriptions des paires sons-sens bien formées. C'est l'agencement de trois dispositifs ou composantes spécialisées dans une tâche particulière. Chacune de ces composantes est un système de règles.

La composante syntaxique engendre un ensemble de structures syntaxiques dont chacune contient entre autres une séquence de morphèmes bien formée. La composante sémantique attribue un sens à chacune de ces structures syntaxiques, et la composante phonologique leur attribue une prononciation.

Aussi, la composante syntaxique est-elle l'épine dorsale du système. C'est elle qui rend compte de la créativité, de cette aptitude à faire toujours des combinaisons nouvelles à partir d'un nombre fini de morphèmes. Les deux autres composantes se contentent d'interpréter les objets engendrés par la composante syntaxique en les munissant d'une face audible (prononciation) et d'une face intelligible (sens). On appelle structure superficielle (de surface) la partie d'une structure qui contient toutes les informations nécessaires pour spécifier la prononciation de la phrase correspondante. En effet, pour spécifier la prononciation d'une phrase, il faut connaître les morphèmes qui la composent et l'ordre dans lequel ils sont rangés (Dell, 1973 : 31). Le lexique est l'équivalent d'un dictionnaire ; c'est la liste des éléments lexicaux, en quelque sorte de pièces détachées où la syntaxe puise les matériaux nécessaires à la construction des structures syntaxiques ; les éléments lexicaux étant des morphèmes isolés (*ami, cheval*) ou des assemblages fixes de morphèmes (*entre+prendre, mé+connaiss+able, porte+plume, ...*).

Le sens de chaque morphème contribue pour une part qui lui est propre à déterminer le sens de cette phrase. La grammaire d'une langue doit contenir des principes généraux qui permettent, étant donné la structure syntaxique, de lui assigner un sens global en partant des sens individuels des morphèmes qui la composent et de la façon dont ces morphèmes sont combinés. C'est la tâche des règles sémantiques, dont l'ensemble constitue la composante sémantique (Dell 1973 : 32). A partir de l'ensemble des structures syntaxiques engendrées par la composante syntaxique, la composante sémantique construit un ensemble infini de représentations sémantiques qui correspond terme à terme à cet ensemble.

Étant donné n'importe quelle structure syntaxique, la composante phonologique lui assigne une prononciation en partant de la prononciation de chaque morphème pris individuellement et de la manière dont ces morphèmes sont combinés. Prédire la prononciation d'une séquence de morphèmes n'est donc pas un problème trivial. Il ne suffit pas malheureusement de mettre bout à bout les prononciations individuelles des morphèmes dont elle est composée. Car la prononciation d'un morphème donné varie le plus souvent d'une combinaison à l'autre.

Sans la mettre de côté, Chomsky ne considère pas la pragmatique en tant que telle. Il l'insère dans le module sémantique, mais reconnaît que son fonctionnement n'est pas véritablement grammatical. Chomsky concède à la théorie behavioriste qu'elle est la seule capable de donner les clés du fonctionnement pragmatique du langage.

La phonologie et la syntaxe sont deux modules pour lesquels on pourrait arriver à une formalisation complète, tandis que la sémantique comme la syntaxe peuvent dire quelque chose du fonctionnement interne du cerveau humain. C'est la raison pour laquelle Chomsky met la syntaxe au cœur de sa théorie : elle est à la fois formalisable et révélatrice d'un dispositif mental ;

elle est vue comme le module le plus proprement linguistique.

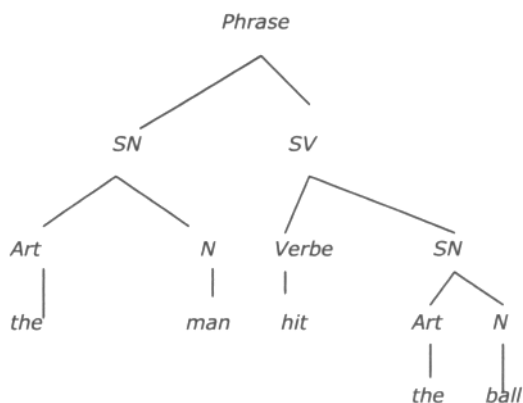
3.3.3.2 La genèse des énoncés (déroulement général)

On appelle l'énoncé tel qu'il est avant d'être produit par le locuteur la « **structure profonde** » (voir ci-après). Cette structure profonde résulte de ce que les générativistes reconnaissent encore comme la véritable « boîte noire » du cerveau : la construction du sens.

De la structure profonde, la syntaxe fournit une forme de squelette, un énoncé général que l'on peut formaliser sous la forme d'un arbre (voir figure ci-dessous). Ce squelette est renvoyé à nouveau vers le module sémantique, qui fournit un lexique mental, remplissant les trous créés par l'arbre syntaxique à l'aide de mots. Le module phonologique va ensuite se charger de la mise en œuvre finale de la phrase. L'idée de la GGT n'est pas de décrire la formation d'un énoncé qui ait un sens, car ce sens est reconnu comme inaccessible à une formalisation.

Extrait de *Structures syntaxiques* p. 29-30 (édition française, 1969)

- (13) (I) *Phrase* → *SN* + *SV*
 (II) *SN* → *Art* + *N*
 (III) *SV* → *Verbe* + *SN*
 (IV) *Art* → *The*
 (V) *N* → *man, ball, etc.*
 (VI) *Verbe* → *hit, took, etc.*
- (14) *Phrase* (I)
SN + *SV* (II)
Art + *N* + *SV* (III)
Art + *N* + *Verbe* + *SN* (IV)
The + *N* + *Verbe* + *SN* (V)
The + *man* + *Verbe* + *SN* (VI)
The + *man* + *hit* + *SN* (VII)
The + *man* + *hit* + *Art* + *N* (VIII)
The + *man* + *hit* + *the* + *N* (IX)
The + *man* + *hit* + *the* + *ball*



Le concept de **structure profonde** a été élaboré par Chomsky lorsqu'il s'est aperçu que la compréhension d'un énoncé ne pouvait découler directement des phrases telles qu'elles apparaissent une fois produites. Il postulera alors que la phrase déclarative et la phrase interrogative proviennent d'une même structure profonde. Celle-ci découle directement du module sémantique, même si elle est pleinement syntaxique, et en elle-même vide de sens. Un

certain nombre de règles, décrivent la structure interne des constituants. Ces constituants vont être déplacés suite à des transformations, qui vont permettre d'avoir plusieurs réalisations formelles différentes en surface. Cette formulation de surface est liée au contexte, et arrive assez tard dans la formulation de l'énoncé.

3.3.4 La question de l'apprentissage/acquisition

Chomsky est le premier linguiste arrivant avec une théorie assez perfectionnée de l'apprentissage, permettant de faire des liens entre les différentes langues :

« Pour apprendre une langue, l'enfant doit dès lors être en possession d'une méthode propre à élaborer une grammaire, à partir des données linguistiques primaires » (Chomsky 1968 : 43).

Pour lui, comprendre comment se déroule l'apprentissage constitue un enjeu majeur de la linguistique. Il tentera de rendre compte de la compétence du locuteur adulte. C'est ainsi que les théories de l'apprentissage des langues ne « tombent pas du ciel », mais découlent d'un modèle linguistique fort. De la même manière que Chomsky s'intéresse à l'apprentissage, il va considérer l'acquisition de la langue maternelle, et ce dès les premiers développements de sa théorie. L'idée en est que malgré la pauvreté des stimuli, les enfants sont en mesure de produire une grammaire. Ils doivent donc avoir un élément inné qui leur permet de produire des énoncés linguistiques. Pour cette raison, les théories générativistes développent le concept de LAD (« *Language Acquisition Device* ») :

« *L'objet d'apprentissage, la structure cognitive acquise, aura les propriétés de Grammaire Universelle (GU) [= paramètres], tout en possédant aussi d'autres propriétés, des propriétés contingentes [= settings]. Toutes les langues humaines seront conformes à GU; leurs différences tiendront à ces propriétés contingentes.* » (Chomsky 1968 :43).

« *Quant à ce qu'on appelle parfois "grammaire universelle" [...]; c'est une théorie de l'état initial de la faculté de langage avant qu'elle ne soit affectée par une expérience externe quelconque.* » (Chomsky, 1968 : XIV).

« *L'acquisition d'une langue tiendrait du miracle si ses propriétés fondamentales n'étaient pas prédéterminées par l'état initial de la faculté de langage.* » (Chomsky, 1965 : XVI).

Dès sa venue au monde, l'enfant est donc doté d'une « boîte noire » qui va lui permettre d'acquérir la langue maternelle, et qui contient ces universaux du langage. Ces universaux, paramètres généraux, vont être recalibrés pour en arriver au figement dans la langue maternelle.

3.3.5 Les universaux dans la GGT

« *Définissons la "grammaire universelle" (GU) comme le système des principes, des conditions et des règles qui sont des éléments ou des propriétés de toutes les langues humaines, pas simplement par accident, mais par nécessité – nécessité biologique et*

non logique, évidemment. Ainsi on peut considérer que GU exprime l'essence du langage humain" » (Chomsk1968 : 40-41)

*« L'idée directrice est que la faculté de langage peut être comparée à un réseau électrique fixe connecté à une boîte de commutateurs. Le réseau correspond aux **principes invariants du langage**, les commutateurs aux choix paramétriques qu'impose l'expérience. [...] Chacune des langues-1 possibles correspond à un ensemble particulier de positions des commutateurs. Il devrait donc être possible de déduire les propriétés d'une langue d'un choix particulier de valeurs pour ces paramètres. » (Chomsky, 1968 : XVIII).*

La linguistique générative apporte ici une contribution importante à la compréhension des universaux linguistiques. Que toutes les langues aient des caractéristiques communes nous permet de comprendre des choses sur le fonctionnement de l'esprit humain, de façon générale. Les universaux, à l'instar de la langue, sont un miroir de l'esprit humain. Étudier la langue permet donc de découvrir les principes généraux qui gouvernent le fonctionnement de l'esprit.

3.4 Le cognitivisme

Nous avons parlé du structuralisme, du distributionnalisme et du générativisme. Le cognitivisme va nous replonger à nouveau dans cette boîte noire qu'est le cerveau que Chomsky cherchait à mettre à nu.

3.4.1 Aperçu comparatif des théories précédentes

STRUCTURALISME	GENERATIVISME
<p>Opposition langue – parole Le structuralisme opère une distinction entre langue et parole. La parole n'est que la diversification individuelle d'un système commun.</p>	<p>Opposition compétence – performance L'angle d'approche du générativisme est l'opposition entre compétence et performance. Pour Chomsky, la performance est « sale » : elle est peuplée d'hésitations, de retours en arrière, etc.</p>
<p>La langue est sociale (externe) Pour Saussure, la langue est un phénomène social. Chaque individu possède une partie de ce système.</p>	<p>La langue est individuelle (interne) Pour Chomsky, la langue est avant tout interne et individuelle.</p>
<p>La langue est acquise Chez Saussure, l'enfant, en apprenant la parole, voit le système se déposer en lui par le biais de l'environnement.</p>	<p>La langue est innée et universelle Chomsky a une perception innéiste et universaliste de l'apprentissage : l'enfant a ce qu'on appelle une grammaire universelle, qui va être recalibrée au fur et à mesure des acquisitions. Mais l'enfant n'est pas une <i>tabula rasa</i> : il vient avec quelque chose.</p>
<p>Système de signes (rapport entre les signes) Le structuralisme étudie « la vie sociale des signes ».</p>	<p>Module(s) génératif(s) (opérations formelles) Ce qui intéresse Chomsky est de comprendre la machinerie qui permet de produire l'ensemble des phrases d'une langue. On voit clairement une</p>

	approche formelle proche de ce qu'on peut trouver dans les ordinateurs.
<p>Induction (cf. corpus) Dans le structuralisme, on part de données concrètes pour établir une conclusion.</p>	<p>Introspection et jugement d'un locuteur natif idéalisé</p> <p>Dans le générativisme, le linguiste forme des énoncés linguistiques sur la seule base de son intuition, et les soumet éventuellement à des locuteurs pour évaluer leur grammaticalité. De plus, un individu, même pendant toute sa vie, ne produira qu'un corpus limité : on ne parviendra jamais à quelque chose de complet. La langue qui est donc visée pour la formalisation est la langue d'un locuteur natif, un locuteur parfait qui maîtrise toutes les possibilités de la langue, qui n'hésite pas, ne reformule pas. On parle d'une compétence internalisée.</p>

Le cognitivisme sera une réaction à la fois au courant structuraliste et au courant générativiste. En effet, le cognitivisme comme la linguistique générale se définissent comme une « science du langage appréhendé à travers la diversité des langues ». Il s'agit, dans les deux cas, de s'intéresser aux universaux linguistiques et à la structure des langues. Mais ces découvertes trouvent dans le cognitivisme un écho nouveau : elles doivent dire quelque chose sur le fonctionnement de la cognition humaine, du fonctionnement de l'esprit et de la pensée humaine, à travers les langues :

« C'est de l'intérieur même du système des langues et à partir de l'étude de l'organisation structurale et signifiante de ce système que la linguistique cognitive propose d'appréhender les liens entre langage, esprit et cerveau » (Fuchs 2004 : 3).

Il s'agit de découvrir les liens entre le langage, l'esprit humain et le cerveau. Cette propriété peut sembler rapprocher la linguistique cognitive de la linguistique générative qui l'a précédée. Cependant, pour les générativistes, l'essentiel est la syntaxe. Si on n'a pas ce squelette, on ne peut rien faire : le sens ne peut qu'être étudié de manière subordonnée à la syntaxe. Avec les cognitivistes en revanche, on s'intéresse à la structure signifiante de l'énoncé de manière bien plus fondamentale : le sens joue un rôle désormais crucial.

3.4.2 L'avènement du cognitivisme

L'historiographie officielle fait remonter le tournant cognitif aux deux conférences qui, en 1956 aux Etats-Unis (l'une à Cambridge, l'autre à Dartmouth), ont réuni des représentants de plusieurs disciplines scientifiques différentes autour d'un projet épistémologique commun, connu sous le nom de « *programme cognitiviste* ».

Le linguiste Noam Chomsky côtoyait le psychologue Herbert Simon et le spécialiste d'intelligence artificielle Marvin Minsky dans cette entreprise pluridisciplinaire. L'objectif était de caractériser le fonctionnement de l'esprit à travers les facultés qu'il développe, et notamment à travers la faculté de langage.

L'hypothèse fondatrice de ce projet était que, de façon générale, la cognition humaine

pourrait être définie, à la manière d'une machine, en termes de *calculs* (« computations »), correspondant au traitement des divers types d'informations reçues par l'humain. C'est ainsi que la linguistique s'est trouvée (par l'intermédiaire des tenants d'une approche formelle de la langue) partie prenante de l'entreprise « cognitive » dès ses débuts, participant à ce que d'aucuns qualifient parfois de « révolution des sciences cognitives ». *C'est à la même époque que le générativisme arrive au goût du jour.*

Chomsky va donc nourrir avec les psychologues des discussions intéressantes. Pour mémoire, avec sa critique du behaviorisme, Chomsky avait pourtant adressé une lourde critique au courant de la psychologie qui dominait alors : il les avait chassés du domaine de la linguistique. Mais contrairement au behaviorisme, le cognitivisme entend en effet plonger dans le fonctionnement interne de l'esprit humain. Les cognitivistes rencontrent Chomsky en arrivant avec une théorie très différente, selon laquelle la cognition humaine peut être définie à la manière d'une machine réalisant des calculs.

On voit là de grandes similitudes avec les théories nourries par le générativisme et au départ, le cognitivisme demeurera très proche du générativisme.

Il convient de signaler que l'étape de 1956, réputée marquer la naissance de la linguistique cognitive, avait été précédée, dès le tournant des années 1940, par la période de la « *cybernétique* », dont les pères fondateurs ont pour noms von Neumann, Wiener, Turing, McCulloch. La cybernétique visait à instaurer une nouvelle « science de l'esprit » en s'appuyant notamment sur *la logique mathématique* (pour décrire le fonctionnement du raisonnement), sur *la théorie des systèmes* (pour formuler les principes généraux gouvernant tout système complexe) et sur *la théorie de l'information* (comme la théorie statistique du signal et des canaux de communication).

L'idée même d'une pensée fonctionnant comme un calcul, à l'instar d'une machine, et d'un cerveau dont les constituants incarneraient des principes logiques, est constitutive de l'approche cybernétique (c'est de là que procédera l'invention de l'ordinateur, selon les principes de von Neumann). Or c'est à cette tradition que se rattachent les tout débuts de la *linguistique dite « computationnelle »*, c'est-à-dire du courant s'inspirant de l'étude des langages formels pour élaborer des traitements automatiques des langues.

A l'approche cybernétique avait donc succédé le « *cognitivisme* » des années 1950. Si celui-ci reprenait l'idée de calcul, en revanche la nature de ce calcul avait changé : il s'agissait désormais d'un calcul sur des représentations symboliques. D'où le nom de « *paradigme computo-représentationnel symbolique* » donné au paradigme épistémologique constitutif du courant cognitiviste classique. Les calculs y sont définis en termes d'opérations sur des symboles. Les symboles (les éléments linguistiques) sont considérés comme ayant une réalité à la fois physique (ils seraient « inscrits » dans le cerveau) et sémantique (ils « représenteraient » le monde objectif)²⁹.

²⁹ Chomsky (1965 : 27-30) distingue deux types d'universaux linguistiques : les universaux *substantiels* sont les pierres angulaires de la structure linguistique, il s'agit des traits distinctifs phonologiques et de la notion de syllabe, des parties du discours (qui sont organisées ultérieurement en fonction de deux traits [+/-N, +/-V], les noms recevant la double caractérisation [+N, -V], les verbes [-N, +V], les adjectifs [+N, +V] et les adpositions • [-N,-V]), et le principe de hiérarchisation arborescente des syntagmes. Le second type est celui des universaux *formels* qui constituent le répertoire des types de règles que l'enfant doit assimiler. Dans cette conception, l'enfant dispose d'un répertoire de règles (par exemple de formation des syntagmes et de dérivation) qui se développeront en règles de la grammaire interne de l'enfant une fois alimentées par la perception des régularités de la langue maternelle.

Dans cette perspective, la cognition se laisse caractériser comme un traitement d'informations mettant en jeu (au niveau symbolique) des règles de manipulation de symboles caractérisés comme des éléments physiques (au niveau neurobiologique) et représentant adéquatement le monde réel (au niveau sémantique). Le cognitivisme des années 1950 se fondait donc largement sur la métaphore de « l'esprit-machine » (partagée par la psychologie cognitive, la philosophie cognitive, et l'intelligence artificielle).

Le paradigme computo-représentionnel symbolique est le paradigme théorique auquel se rattachent, en linguistique, la *grammaire chomskienne* (Chomsky [1965], [1981], [1995]) et, à sa suite, un certain nombre d'autres modèles formels en grammaire qui, tous, accordent une place centrale à la notion de « système formel ».

La linguistique chomskienne a donc été la première à se revendiquer explicitement comme une linguistique cognitive. Les caractéristiques de cette approche peuvent être résumées comme suit :

- (1) La démarche adoptée est de nature hypothético-déductive, et les exemples étudiés sont forgés par le linguiste sur la base de sa compétence de la langue, en tant que sujet parlant.
- (2) L'approche est « modulariste » : il est fait l'hypothèse que le module « langue » est spécifique par rapport aux autres modules cognitifs, et une place centrale est accordée à la syntaxe, conçue comme devant assurer l'interface entre une représentation sémantique et une représentation phonétique.
- (3) La conception du langage est représentationnelle : le langage est caractérisé comme instrument d'expression de la pensée, permettant la transmission d'informations à propos du monde.
- (4) Enfin, la modélisation a, de façon classique, recours à des calculs d'ordre logico-algébrique.

Cette approche linguistique s'inscrit dans le cadre plus général du développement des sciences cognitives et des neurosciences (en lien avec la biologie moléculaire et la génétique) ; la « grammaire universelle » proposée par Chomsky fait écho, en matière de langage, à l'idée d'un code génétique se transmettant de génération en génération : c'est « l'instinct du langage », selon les termes de Pinker [1994].

Il va falloir cependant un certain temps pour que les cognitivistes quittent cette approche purement formelle. Après un bout de chemin commun, on va assister à une autonomisation progressive du cognitivisme par rapport au générativisme. Cette autonomisation va prendre une trentaine d'années et va principalement être l'œuvre de :

- *George Lakoff (1941-) : Women, fire, and other dangerous things, 1987*
- *Ronald Langacker (1942-) : Foundations of the cognitive grammar, 1987*
- *Gilles Fauconnier (1944-) : Espaces mentaux, 1984*
- *Leonard Talmy (1943-) : The relation of grammar to cognition, 1988*

Ces auteurs ont élaboré diverses formes de grammaires cognitives en réaction contre l'option modulariste et la prééminence de la syntaxe prônée par les chomskiens. Nous parlons donc ici de la linguistique telle que formée par ces figures. On pourrait ne parler que de Langacker,

mais faire l'impasse sur les autres serait regrettable pour la diversité des théories développées.

3.4.3 Les nouvelles questions théoriques soulevées par le cognitivisme

Ces grammaires *cognitives* contestent le postulat selon lequel les grammaires formelles constitueraient des modèles adéquats de la cognition linguistique et cherchent à relier les phénomènes langagiers aux processus généraux de la cognition (comme par exemple la perception).

Par différence avec les grammaires formelles, cet autre courant de linguistique cognitive se caractérise par une démarche inductive et une approche interactionniste : une place est accordée à la sémantique, réputée informer la syntaxe et le lexique avec lesquels elle interagit. La conception du langage est davantage émergentiste que représentationnelle : le langage est envisagé comme instrument de conceptualisation active du monde et/ou comme instrument de communication. Pour ces théories, le noyau dur de la langue ne réside pas dans les règles de la syntaxe mais dans les opérations de construction de la signification.

En effet, pour les tenants des théories cognitivistes, plusieurs questions se posent avec le cognitivisme classique. Pour eux, le cognitivisme en linguistique conduit à s'interroger sur l'ensemble des connaissances spécifiques que maîtrise l'esprit humain au travers de la faculté de langage (elle-même appréhendée à partir du système des langues) mais aussi à se demander comment ces connaissances sont organisées pour pouvoir être acquises et mises en œuvre dans l'activité du langage.

D'où toute une série de questions qui concernent, par-delà l'architecture structurale des connaissances linguistiques, leur architecture fonctionnelle :

- (1) Sous quelle forme ces connaissances sont-elles organisées dans l'esprit et le cerveau humain pour permettre l'apprentissage et le fonctionnement effectif du langage ainsi que ses dysfonctionnements : s'agit-il des modules (indépendants ou en interaction) ?
- (2) Quelle place accordée aux variations entre les langues face à l'universalité de la faculté de langage ? *L'étude des langues, a permis de s'apercevoir qu'elles présentaient des différences, et que même à l'intérieur d'une langue les faits n'étaient pas uniformes. La question qu'on se pose dès lors en linguistique cognitive est de savoir quels pourront en être les impacts du point de vue de la cognition.*
- (3) *La cognition humaine a en effet dû s'adapter à des particularités intrinsèques au langage, qu'il s'agisse de la variation de la prononciation, de l'évolution permanente et active des langues, ou du fait qu'il puisse exister deux mots pour renvoyer à un même objet. Comment, dès lors, la communication humaine est-elle encore possible sans poser de problèmes (majeurs) de compréhension ? Comment donc fonctionne le langage pour parvenir à réaliser de manière concordante la synthèse de subjectivités différentes ?*
- (4) Une linguistique à visée cognitive ne doit pas se désintéresser de la question des liens entre le langage et d'autres facultés humaines, notamment des liens entre le langage et la pensée. Quelle est la nature du processus de représentation qui sous-tend l'activité de langage ? est-ce un processus spécifique, ou bien présente-t-il des analogies avec

d'autres types de processus cognitifs (par exemple la vision) ? comment la langue met-elle en œuvre de façon particulière ce processus ?

3.4.4 Postulats théoriques

Le cognitivisme a donc quatre pères fondateurs, avec chacun sa conception de ce que devrait être la linguistique cognitive. Malgré tout, ils partagent quelques postulats théoriques communs. Ils partent en effet tous du principe que tout en ayant ses spécificités, la faculté de langage est régie par des mécanismes cognitifs généraux à l'œuvre dans toutes les activités cognitives.

➤ **L'opposition « figure-ground »**

Pour les cognitivistes, le langage est une faculté qui est tout à fait liée à d'autres facultés du cerveau. En réalité, cette division langage-cerveau est créée par notre vision du monde : c'est déjà une division qui est le fait de notre propre cognition. Le phénomène en est le suivant : quand on regarde quelque chose dans la vie courante, on conçoit toujours un élément d'avant-plan et des idées d'arrière-plan, le décor. Tout le monde perçoit les choses de cette manière, et cette particularité est en réalité centrale. On a toujours un point de départ, avec un élément d'avant-plan et un élément d'arrière-plan : à un élément connu (topique) s'ajoute un élément nouveau (focus). On retrouve l'opposition « figure – ground » déjà opérée en psychologie cognitive par le mouvement gestaltiste.

➤ **Rejet de la primauté de la syntaxe pour la sémantique**

Pour Chomsky, l'élément indispensable de la langue était la syntaxe. Prenons par exemple l'énoncé « He sent an e-mail to Susan » par rapport à « He sent Susan an e-mail ».

Pour les générativistes, il s'agit d'une transformation très intéressante. Il s'agira d'avancer une théorie permettant de démontrer pourquoi il y a eu « transformation », et par exemple pourquoi il y a insertion/suppression de la préposition « to ».

Les cognitivistes constatent la différence de structures entre ces deux énoncés, mais pour eux, les deux phrases renvoient en réalité au même résultat. La différence que l'on va tenter d'approfondir est que les éléments mis en avant ne sont pas les mêmes : on met en avant la trajectoire par rapport au résultat, ou le résultat par rapport à la trajectoire. La plus petite différence est en réalité, pour les cognitivistes, porteuse de sens. Quand on fait du cognitivisme, on fait avant tout de la sémantique.

➤ **Pas une relation langage-cognition à sens unique**

En fait, l'apport du cognitivisme se fait nécessairement dans deux sens : l'étude de la cognition permet d'expliquer certaines théories linguistiques, mais l'étude du langage permet également de comprendre certains mécanismes mentaux. Le sens a dans le cognitivisme une primauté dans le langage, qu'on retrouve à tous les niveaux :

« Du caractère intrinsèquement symbolique du langage se dégage la primauté du sens, dans pratiquement tous les problèmes linguistiques. [...] Je prends ceci comme un

principe directeur : les descriptions linguistiques doivent tenir compte du sens non seulement d'une manière explicite, mais également intégrée. » (Langacker 1983 : 6).

On peut concevoir que le sens est capital à plusieurs niveaux discursifs : au niveau grammatical, au niveau du lexique et au niveau de la syntaxe. Le cognitivisme a même quelque chose à dire sur la construction du sens au niveau du discours ; il s'agit de montrer comment dans le discours différentes idées s'enchaînent.

3.4.5 *Sens grammatical vs sens lexical*

S'appuyant sur la distinction classique pour les unités linguistiques entre classes ouvertes (noms, verbes, etc.)³⁰ et classes fermées (prépositions, déterminants, etc.)³¹, Talmy (2000, vol. 1, p. 21) définit deux « sous-systèmes » au sein des langues qui ont des fonctions sémantiques nettement différenciées :

- (i) **le sous-système grammatical**, qui détermine la structure de la représentation cognitive évoquée par un énoncé, et
- (ii) **le sous-système lexical**, qui en détermine le contenu.

Le sous-système grammatical contient donc les morphèmes grammaticaux (libres ou liés), tels que les prépositions, les conjonctions, les flexions nominales et verbales, etc.

Mais dans la conception de Talmy, il ne se limite pas à ces morphèmes. Il comporte aussi des éléments plus abstraits, tels que les catégories grammaticales (parties du discours), les sous-catégories (le fait pour un nom d'être massif ou comptable, par exemple), les fonctions syntaxiques (sujet, objet, etc.) et les constructions syntaxiques (l'ordre des mots, notamment). D'où la distinction est plus subtile que celle qui est opérée traditionnellement.

Un mot comme *chien*, par exemple, présente à la fois des aspects grammaticaux et lexicaux.

En tant que nom comptable (caractéristique qu'il partage avec des mots comme *chat*, *table*, *arbre*, etc.), il relève du sous-système grammatical. Et c'est uniquement en tant que lexème possédant un contenu sémantique qui l'oppose aux autres noms comptables qu'il fait partie du sous-système lexical.

Le point clé qui fait tout l'intérêt de cette approche, c'est de considérer que les aspects grammaticaux sont aussi porteurs de sens en eux-mêmes, indépendamment des aspects lexicaux. La même remarque s'applique aux autres classes ouvertes (verbes, adjectifs, etc.). Notamment, pour un bon nombre de verbes de mouvement, c'est pratiquement l'intégralité de leur sens qui relève de la sémantique grammaticale.

Cette sémantique grammaticale est essentiellement configurationnelle : le rôle du sous-système grammatical est d'organiser les différents éléments évoqués par un énoncé les uns par rapport aux autres en une scène complexe cohérente. *La personne, dans la perception d'un langage, va reconstruire un fil. Il va reconstituer l'action, la voir se dérouler et, par ces*

³⁰ *Chaque jour, on invente de nouveaux noms, de nouveaux verbes qui entrent dans une langue et sont en perpétuelle évolution.*

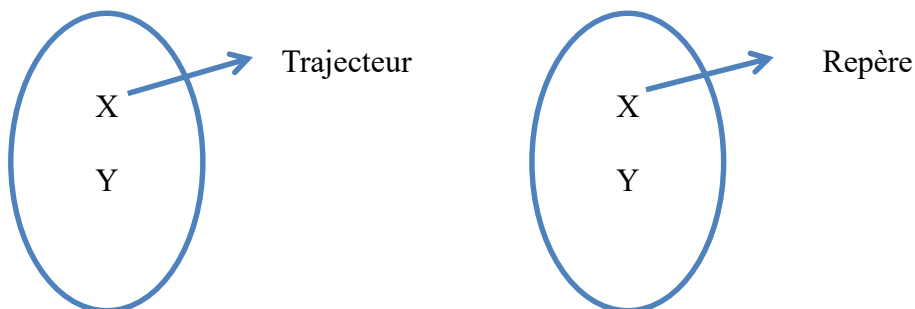
³¹ *Ici, on a un système plus ou moins fixe, ou du moins qui évolue beaucoup moins : prépositions, articles, etc.*

différents éléments, permettre aux représentations mentales de se faire le « film » de ce qui s'est passé.

Le sens grammatical est d'abord conçu en termes de propriétés et de relations topologiques et cinématiques.

Pour représenter ce lien de sens, les cognitivistes utilisent une forme de diagramme ; les énoncés ne sont plus représentés par des arbres, comme chez les générativistes, mais par des dessins, des diagrammes, qui ont la propriété d'être multidimensionnels et d'ainsi pouvoir porter des informations plus complexes. Ainsi l'entité évoquée par un nom est représentée par une région dans un espace multidimensionnel que l'on appelle son domaine (et qui dépend bien sûr du contenu sémantique du nom). Les unités linguistiques relationnelles, comme les prépositions, sont représentées dans ces diagrammes par des relations géométriques statiques entre les régions représentant des entités. Ces relations sont en général asymétriques : l'une des entités, appelée « trajecteur » (*trajector*), joue un rôle particulier par rapport aux autres, appelées « repères » (*landmarks*), qui servent de points de référence pour localiser le trajecteur.

Pour prendre un exemple simple, les deux prépositions *above* et *below* correspondent à un même diagramme représentant deux entités X et Y disposées le long d'un axe vertical dans le domaine constitué par l'espace physique orienté. La différence entre les énoncés *X is above Y* et *Y is below X* consiste en une inversion du trajecteur et du repère : dans le premier énoncé, X est le trajecteur, alors qu'il est le repère dans le second.



Les procès, évoqués par des verbes, sont représentés par des schémas qui comportent une dimension supplémentaire : le temps. Autrement dit, un procès sera représenté par la succession le long de l'axe du temps d'une série de diagrammes statiques, le premier correspondant à la situation au début du procès, et le dernier à la situation résultante. L'opposition entre procès perfectifs et imperfectifs est rendue par l'existence ou non de bornes, cette fois sur l'axe temporel.

La **notion de profilage** est très importante pour comprendre la manière dont les éléments grammaticaux font intervenir les sens. La notion de Profilage s'inspire de l'opposition figure/fond de la théorie de la Gestalt et renvoie à la notion de **saillance relative** d'un énoncé par rapport à un autre. Un élément est considéré comme **saillant**, tandis que l'autre est davantage un élément **d'arrière-plan** : il constitue le décor. Il s'agit de l'opposition **profil/base** explicitée plus haut.

Dans un schéma, certains éléments sont plus saillants que les autres parce qu'ils sont plus directement évoqués par l'expression étudiée : ils constituent le « profil » (*profile*), qui se

détache de la « base » (*base*).

Pour expliquer cette notion, Langacker prend, entre autres, l'exemple du nom relationnel *uncle*. Le domaine impliqué est ici celui des relations de parenté que Langacker symbolise par un arbre généalogique. La représentation du sens de *oncle* demande la prise en considération de toute une région d'un tel arbre, permettant d'illustrer la relation entre deux nœuds : l'oncle et le neveu. C'est l'ensemble de cette région qui compose la base ; le profil étant constitué par le seul nœud oncle. Ainsi la représentation du nom *neveu* possèdera la même base, mais c'est cette fois le nœud neveu qui sera profilé.

Loin de se limiter à la représentation des noms relationnels, le profilage se révèle un outil très puissant, qui permet effectivement de différencier des expressions linguistiques de toute sorte se rapportant à un même champ conceptuel. Notamment, cela permet de distinguer un verbe évoquant un procès, un participe passé associé à l'état résultant de ce procès, une nominalisation de ce même procès, etc. Prenons l'exemple de la situation « *le professeur arrive* ». On voit ici la description sémantique de l'énoncé « le professeur arrive ». Il s'agit de montrer comment on va représenter le processus « arriver », mis en relation avec une perception de la temporalité. Pour représenter cela, on va utiliser une série de diagrammes qui représentent chacun un état.

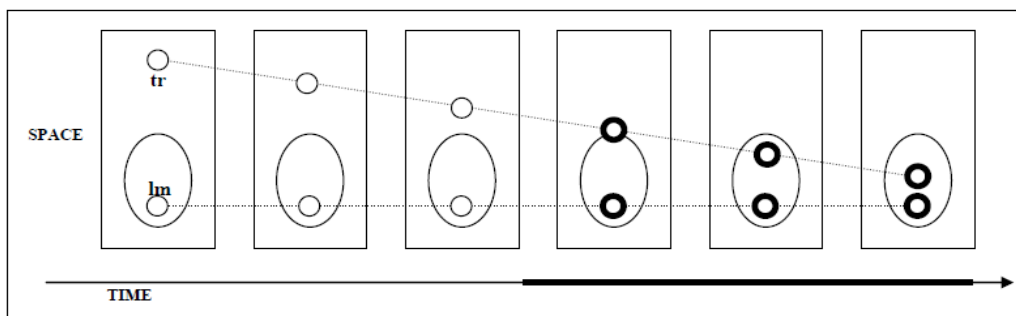


Figure 1 : Représentation de *arrive*

Le premier diagramme représente le point de départ, le dernier représente l'arrivée en tant que processus fini. Un point de référence qui est celui de la position du locuteur (lm) est confiné dans une zone qui est la zone de perception du locuteur. C'est la zone proximale, le champ de vision du point de repère. Un trajecteur va évoluer par rapport à ce point de repère. Plus on avance dans le processus, plus on s'approche du point de repère, pour enfin se situer dans son point de vue immédiat, au moment où le trajecteur est arrivé au-devant de la zone proximale : le processus duratif de l'arrivée est alors parvenu à son achèvement. Ce processus est entretemps ce que dénote le verbe « arriver ». Les éléments en gras matérialisent le profil dans ce schéma. Ce profilage indique que le sens du verbe est limité à une partie du processus représenté : l'entrée du trajecteur dans la zone proximale du repère, ce qui constitue effectivement une bonne définition de *arrive*.

Bref, le rôle grammatical va nous permettre de construire des images mentales et de les véhiculer par le langage. Les représentations ne sont pas simplement décrites, elles sont présentées selon une vision du monde, un système de référence. Elles sont soit mises en évidence, soit en arrière-plan, à partir d'une certaine perspective. Certains éléments sont plus saillants que d'autres, et

ces différents rôles vont évoluer. Les entités et événements évoqués par les énoncés ne sont pas simplement décrits, ils sont mis en scène, présentés selon un certain point de vue, avec des éléments saillants et un arrière-plan dans des configurations dynamiques qui évoluent dans le temps en fonction des forces qui animent les entités représentées. (Langacker 1987 : 81).

3.4.6 *Le sens lexical : le prototype et la métaphore, deux principes fondamentaux*

3.4.6.1 Le prototype

Il s'agit d'une réaction par rapport à une conception ancienne de la sémantique, qui voulait qu'un mot fasse partie ou non d'une catégorie de façon blanche ou noire. Les cognitivistes vont réagir à cette conception en disant que, dans une catégorie lexicale donnée, des éléments présentent totalement les caractéristiques d'une catégorie, tandis que d'autres sont plus périphériques. L'appartenance à ces catégories n'est pas du genre "tout ou rien", mais plutôt du genre "plus ou moins". *Certains objets sont alors plus exemplaires d'une catégorie que d'autres ; on parle d'une **approche prototypique de la signification des concepts**.*

Cela conduit à refuser de définir les sens d'un mot comme on le fait en logique, par un ensemble de conditions, nécessaires et suffisantes, qu'une entité doit satisfaire pour pouvoir être désignée par le mot en question. Il faut plutôt le définir comme **un ensemble de traits, plus ou moins caractéristiques : plus une entité possède de traits associés au mot, plus elle a vocation à être désignée par ce mot**. Par exemple, les éléments « avocat », « banane » et « ananas » (au Burundi surtout) sont ainsi les plus exemplatifs de la catégorie des fruits : ils en sont les éléments les plus prototypiques. La tomate, en revanche, est beaucoup moins proche du prototype. On aboutit donc, là aussi, à une représentation géométrique du sens.

On peut en effet définir une distance entre entités désignées par un même mot, distance plus ou moins grande suivant le nombre de traits qu'ils partagent. Le sens du mot est alors représenté par une région, au centre de laquelle se trouvent les meilleurs exemplaires (ceux qui possèdent le maximum de traits) : plus on s'éloigne de ce centre, moins le mot est pertinent. Les frontières de cette région sont donc floues, ce qui correspond aux hésitations des locuteurs (et aux désaccords entre locuteurs) sur la désignation d'exemplaires particulièrement atypiques. **C'est le centre de la région associée à un mot que l'on appelle le prototype.**

Il s'agit d'une sorte de réseau radial : ces conceptions sont toutes reliées entre elles, pour former un réseau. De même qu'il expliquera la hiérarchie interne des catégories, le concept de prototype permet donc également de montrer la hiérarchie interne des significations pour un mot, dans le cas des mots dits polysèmes. Cela veut dire qu'on peut catégoriser la polysémie d'un mot donné par une appartenance graduelle des significations, par signification prototypique.

3.4.6.2 La métaphore

Jusqu'à récemment, la métaphore a été le terrain de chasse gardé des rhétoriciens et des stylisticiens. Ce n'est que depuis peu les linguistes s'y sont intéressés. Une explication de cet état de chose est que la linguistique moderne (structurale et générativiste) considère les faits rhétoriques en général et la métaphore en particulier comme des faits de parole (ou de discours)

et non de langue, ou externes au domaine de la théorie linguistique (limitée à la grammaire dans le paradigme chomskien). Depuis les travaux des philosophes du langage (notamment ceux de Grice 1979 et Searle 1982), cette séparation n'est plus de mise. Le clivage entre tradition linguistique et tradition rhétorique s'est d'autant plus estompée que des disciplines connexes (psycholinguistique, psychologie cognitive, intelligence artificielle) se sont intéressées de plein droit à la métaphore. La métaphore et les tropes ont reçu un statut central. L'on notera ici notamment quelques travaux :

- (i) La caractérisation des tropes (métaphore, ironie, litote, hyperbole) comme des implications conversationnelles chez Grice (1975) ;
- (ii) La tentative d'un traitement homogène des faits d'indirection comme les actes de langages indirects, l'ironie et la métaphore (Searle 1982) ;
- (iii) Le rôle central dans la communication assigné aux usages non littéraires du langage comme les énoncés vagues, les implications et les tropes (Sperber & Wilson 1986 et 1989) ;
- (iv) Le développement de recherches linguistiques systématiques sur l'organisation conceptuelle et sémantique des métaphores dans les langues naturelles (Lakoff et Johnson 1985).

Ainsi, si la théorie du prototype étendu est utilisée pour traiter une partie des phénomènes de polysémie, d'autres mécanismes viennent compléter le dispositif, et, en premier lieu, la métaphore, qui joue un rôle capital en sémantique lexicale pour les tenants des grammaires cognitives.

En effet, à la suite de Lakoff et Johnson (1980), **la métaphore est considérée comme un mécanisme général de la pensée, qui permet d'utiliser le domaine de l'expérience sensori-motrice pour appréhender des concepts plus abstraits**. Précisément, nous ne sommes généralement pas conscients des métaphores dans lesquelles nous pensons, mais elles peuvent être mises en évidence en étudiant les traces qu'elles ont inscrites dans les langues. En effet, **l'étude des métaphores permet donc de remplir un double objectif** : d'une part, au plan linguistique, rendre compte d'une grande partie de la polysémie lexicale, et d'autre part, au plan cognitif, mieux comprendre un mécanisme central de la pensée.

Prenons l'exemple de la métaphore, très répandue dans les langues, qui projette sur l'axe vertical le domaine de variation de n'importe quelle variable quantitative. On dit ainsi en français que les **prix « montent », qu'ils « atteignent des sommets », qu'ils « dégringolent »**, etc.

Tous ces concepts sont calqués sur des processus qu'on peut observer dans la réalité, dans la vie de tous les jours. Cette métaphore, ancrée dans l'expérience perceptive quotidienne (le niveau d'eau monte dans un récipient quand sa quantité augmente), explique la lexicalisation de sens quantitatifs d'un grand nombre d'unités linguistiques exprimant le mouvement vertical et l'altitude (*monter, descendre, niveau, sommet, élevé, bas, au-dessus de*, etc.). Mais au-delà du langage, c'est le raisonnement sur le **concept même de variation quantitative** qui s'appuie sur cette métaphore.

La métaphore s'exerce sur tous les champs de l'expérience. Tout nouveau domaine, notamment, est immédiatement investi par le lexique du domaine le plus apte à développer une correspondance analogique opératoire : ainsi a-t-on pu assister récemment à l'introduction d'un vocabulaire biologique et médical en informatique, avec l'apparition des « vers » et des « virus » informatiques, plus ou moins « bénins » ou « virulents », qui « contaminent » les

ordinateurs insuffisamment « immunisés » contre ces « infections ». C'est ici encore un vocabulaire de la vie réelle, propre de la biologie, qui est transformé par le biais du langage en une métaphore appliquée à un autre domaine de la vie réelle.

Parmi les métaphores les plus structurelles, Lakoff insiste sur l'universalité de la conception du temps comme un double déplacement spatial : nous avançons vers l'avenir et l'avenir vient vers nous. Les événements passés sont derrière nous, ils s'éloignent, alors d'autres arrivent à grands pas (*l'observateur est fixe et le temps « passe »*). Mais nous pouvons aussi dire que nous entrons dans une nouvelle ère, et que nous nous rapprochons d'une échéance (*le temps est immobile et l'observateur se déplace*).

D'une manière générale, les métaphores ne sont pas limitées, là encore, à la sémantique lexicale. Elles sont aussi invoquées pour rendre compte de la polysémie des unités grammaticales.

3.4.7 Le sens syntaxique

On a vu que chaque unité grammaticale ou lexicale est porteuse de sens. Il en est de même des constructions syntaxiques qui contribuent elles aussi à l'élaboration du sens global d'un énoncé, de manière autonome par rapport aux unités qu'elles agencent. Comme l'écrit Langacker :

I contend that grammar itself, i.e. patterns for grouping morphemes into progressively larger configurations, is inherently symbolic and hence meaningful. Thus it makes no more sense to posit separate grammatical and semantic components than it does to divide a dictionary into two components, one listing lexical forms and the other listing lexical meanings. (1987, p. 12).

On doit donc attribuer à chaque construction syntaxique un sens qui lui soit propre, et qui permette de rendre compte des effets différentiels obtenus par des permutations de construction sur un même matériau sémantique. Car tout choix d'une structure syntaxique porte également son propre impact au niveau de la sémantique : la différence de structure permet de mettre en scène différemment une même réalité. **Il s'agit donc, pour les cognitivistes, de rendre compte des différences de sens provoquées par des différences de construction sur un même matériau sémantique.**

Exemples :

He sent an e-mail to Susan – He sent Susan an e-mail
*He threw the ball to the fence – *He threw the fence a ball*

Le second énoncé n'a pas de sens pour les cognitivistes. Une règle sémantique impose en effet qu'on ne puisse pas exprimer la possession pour un objet inanimé tel que « a fence » (une clôture) : on ne peut donc mettre l'accent que sur la trajectoire de l'action.

Marie paraît malade – Il paraît que Marie est malade

Ici, on voit que le choix d'une structure syntaxique n'est pas innocent. Chaque structure syntaxique, on le voit, est porteuse de sens. Le sens grammatical, résultant d'une unification ; pour avoir le sens global d'un énoncé, il s'agit de mettre ensemble ses éléments, les fusionner pour arriver à une représentation globale de la réalité représentée par le locuteur. Les éléments constitutifs de l'énoncé (unités grammaticales et lexicales ainsi que les tours syntaxiques) apportent chacun leur contribution au sens global qui est obtenu, en fusionnant les représentations schématiques associées à chacun d'eux. Quand on additionne le sens de tous les éléments d'un énoncé, on en arrive donc au sens global d'un énoncé, même si celui-ci peut être

parfois plus intéressant que la somme des parties. Par exemple, on pourrait voir dans la phrase « il fait chaud ici » une pure situation factuelle. Mais imaginons un contexte où quelqu'un est dans une salle, d'autres personnes près des fenêtres ouvertes : ici, il s'agit d'une requête implicite sur le plan pragmatique. Les cognitivistes sont obligés de reconnaître que les énoncés arrivent accompagnés d'un sens ajouté, lié au contexte et non généralisable à partir des parties de l'énoncé.

3.4.8 La construction dynamique du sens : les espaces mentaux

Les éléments qu'on va rencontrer en cours de discours apportent chacun leur pierre à l'édifice de la construction du sens ; ils vont construire les dispositions sémantiques qui permettront d'arriver à une représentation globale. En effet, pour Fauconnier, les formes linguistiques sont des instructions de construction de configurations cognitives, structurées et interconnectées, qui se mettent en place au fur et à mesure du déroulement de la parole :

A language expression does not have a meaning in itself; rather, it has a *meaning potential*, and it is only within a complete discourse and in context that meaning will be actually be produced. The unfolding of discourse brings into play complex cognitive constructions. They include the setting up of internally structured domains linked to each other by connectors; this is effected on the basis of linguistic, contextual and situational clues. Grammatical clues, although crucial to the building process, are in themselves insufficient to determine it. (Fauconnier 1997, p. 37-38).

Ce sont ces configurations cognitives que Fauconnier appelle des espaces mentaux. Ces espaces forment un réseau qui s'enrichit tout au long du discours. L'espace de départ s'appelle *la base du système*. A tout moment, on peut distinguer deux autres espaces particuliers : *le point de vue*, sorte de repère à partir duquel sont spécifiés de nouveaux éléments, et *le focus*, l'espace qui reçoit ces nouveaux éléments. **Base**, **point de vue** et **focus** ne sont pas forcément distincts (notamment au début du discours). Certaines expressions linguistiques, les « constructeurs d'espace » (*space builders*), servent à construire de nouveaux espaces à partir du point de vue, ou de la base, à laquelle il est toujours possible de revenir. D'autres expressions servent à changer de point de vue ou de focus. D'autres encore servent à spécifier les relations entre espaces, notamment les relations d'accessibilité (le statut énonciatif d'un espace par rapport au point de vue dont il est issu). D'autres enfin servent, naturellement, à introduire du contenu, en évoquant de nouvelles entités, relations et événements qui viennent enrichir la structure du focus.

Une notion fondamentale est celle de correspondance (*mapping*) : les relations entre espaces mettent en place des liens entre certains éléments de ces espaces, permettant ainsi un transfert d'une partie de la structure d'un espace vers un autre. Prenons un exemple très simple (Fauconnier, 1997, p. 42-43) de la phrase '*Maybe Romeo is in love with Juliet*'. A lui seul, cet énoncé construit deux espaces (figure 4). En effet, *maybe* est un constructeur d'espace qui, en l'occurrence, spécifie une relation modale épistémique : le nouvel espace est présenté comme une situation « possible », non pleinement validée par le locuteur. Le réseau comporte donc une base (l'espace B de la figure) dans laquelle sont représentés les éléments a et b, associés aux

noms *Romeo* et *Juliet*. Toutes les connaissances partagées entre les interlocuteurs sur les deux personnes ont vocation à être représentées dans cet espace. En revanche la relation amoureuse supposée est représentée dans un autre espace mental M, qui est le focus actuel, accessible à partir du premier (qui est donc aussi le point de vue) par la relation modale de possibilité (indiquée sur la figure par la ligne en pointillés).

Deux éléments a' et b' sont arguments de la relation amoureuse (notée LOVE) dans l'espace M. Ils sont aussi mis en correspondance avec les éléments a et b de l'espace B (relation d'identité notée *Id* sur la figure), ce qui permet de transférer dans ce nouvel espace tout ou partie des connaissances présentes dans l'espace B, à commencer par les noms *Romeo* et *Juliet*, que l'on peut donc attribuer respectivement à a' et b' .

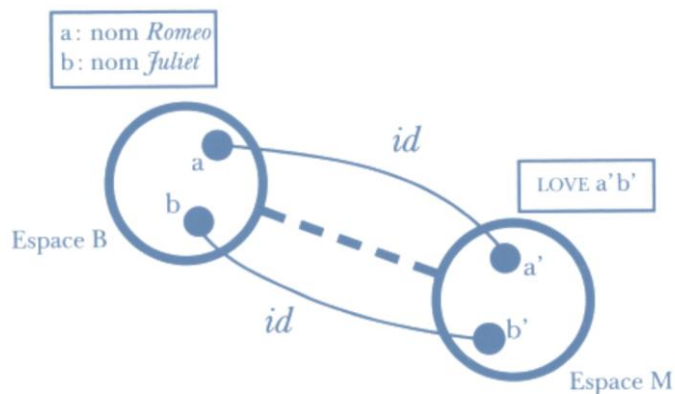


Fig. 4 – Représentation d'espaces mentaux.

N.B. Notions importantes

Représentation cognitive : espace mental aménagé entre les deux locuteurs par la construction énonciative.

Base : point de départ sur lequel s'entendent les deux locuteurs avant la construction du sens.

Point de vue : situation dans laquelle l'énoncé est produit

Focus : « trajecteur » de l'énoncé : information qu'il vise à atteindre.

Dans l'énoncé "Maybe Romeo is in love with Juliet", on peut comprendre ces notions de la façon suivante :

Base : dans un tel énoncé, les locuteurs partent avec des connaissances préétablies sur Roméo et Juliette (parce qu'ils suivent le déroulement de la pièce, parce qu'ils ont des connaissances générales qui leur permettent de savoir quels personnages ils désignent, etc.).

Espace mental : le propos tenu va activer une certaine zone, ouvrir un espace mental dans lequel deux protagonistes interagissent.

Focus : nos connaissances préalables se voient, avec ce nouvel énoncé, ajouter une nouvelle donnée : un second espace mental va être occupé par une nouvelle information.

Point de vue : on voit sur le schéma que visiblement, les deux espaces sont interconnectés : il y a une relation entre eux, qui relie les deux espaces mentaux. On voit cependant que la ligne est en pointillés ; il n'y a, en fait, pas d'affirmation (« maybe ») : l'énoncé n'est qu'une hypothèse.

3.4.9 L'apprentissage d'une langue dans la perspective cognitive

Pour les cognitivistes, la langue est purement acquise, et cette acquisition se fait par l'entremise de l'usage. On se rend pourtant compte que certains processus sont universels : le processus de la métaphore par exemple, est ainsi observable partout, même si toutes les cultures n'en ont pas la même réalité. Ici, l'idée est que le principe de la métaphore est une caractéristique générale de la cognition humaine : c'est un trait universel, mais il ne produira pas le même résultat selon son usage. C'est donc ici le résultat du processus qui est acquis.

3.5 Le fonctionnalisme

3.5.1 La grammaire fonctionnelle

Le fonctionnalisme linguistique contemporain (qu'on peut dire de *troisième* génération) n'est pas la simple continuation du fonctionnalisme de *première* génération, représenté par N.S. Troubetzkoy, V. Mathesius ou A. Martinet, consacré en priorité aux structures phonologiques et à l'identification des fonctions du langage (K. Bühler, R. Jakobson), ni de celui de *seconde* génération orienté vers la perspective fonctionnelle de la phrase et l'articulation entre phrase et discours (J. Firbas, P. Sgall, M.A.K. Halliday)

Les grammaires fonctionnelles se sont développées très tôt dans l'émergence du fonctionnalisme, raison pour laquelle elles seront considérées avant d'autres sous-branches plus tardives du courant.

La première grammaire fonctionnelle trouve ses origines à Prague dans les années 1950. Après la seconde guerre mondiale, des structuralistes héritiers du cercle de Prague se centrent sur un nouvel objet d'étude pour concevoir leurs grammaires formelles : le contexte. Le courant formaliste qui s'est développé à Prague a produit des résultats impressionnants, qui n'ont hélas jamais pu se faire connaître du monde occidental, en raison de la Guerre Froide. Des auteurs comme **Jan Firbas** sont ainsi, encore aujourd'hui, totalement inconnus malgré le caractère inédit de leurs travaux. C'est que par ailleurs, d'autres auteurs leur ont quelques années plus tard « volé la vedette » du côté de l'Occident, en publiant des recherches similaires. **Michael Halliday**, en 1973, développe un modèle de *linguistique systémique fonctionnelle* comblant les mêmes préoccupations heuristiques que Firbas. Encore aujourd'hui, ce sont ses recherches qui font l'objet du plus grand nombre de références dans le monde scientifique, nonobstant leur superficialité : elles ne sont basées que sur l'anglais, et ne peuvent s'appliquer qu'à ce dernier.

Le néerlandais Simon Dik développera un peu plus tard une autre théorie de grammaire fonctionnelle, plus large dans ses applications.

Les locuteurs disposent d'une compétence communicationnelle impliquant deux composantes :

- la **compétence grammaticale**, c'est-à-dire l'aptitude à construire et à interpréter des énoncés, et
- la **compétence pragmatique**, c'est-à-dire l'aptitude à mettre en œuvre ces expressions

linguistique de manière à atteindre l'effet communicationnel recherché.

Les règles d'expression déterminent la forme, l'ordre et le contour prosodique des constituants en fonction de leur statut structural et fonctionnel à l'intérieur de la structure sous-jacente de la phrase élémentaire (Dik 1989 : 289).

3.5.1.1 Axiomes

3.5.1.1.1 *La langue est avant tout interactive*

Le point de départ du fonctionnalisme et de la grammaire fonctionnelle est de dire que la langue est *en premier lieu* un instrument d'interaction sociale. Cela ne signifie pas qu'ils ont le monopole de cette conception ; de fait, leurs prédécesseurs ne nient pas cette interaction. Ainsi, on retrouve même chez Chomsky la notion de **performance**, chez Saussure la notion de parole : ces concepts indiquent une prise en compte du facteur social et interactif de la langue, mais ils ne constituent pas, pour eux, l'objet central de la linguistique. L'usage est vu comme quelque chose de fuyant, d'extrêmement « variable », pour reprendre les arguments de Chomsky. C'est effectivement le cas lorsqu'on désire dans la langue mettre au jour une systématique.

Mais considérer la langue de cette manière c'est laisser de côté un point important, qui est de dire que le langage sert avant tout à la communication. C'est cet aspect que la grammaire fonctionnelle, la première, va tenter de mettre à jour : ils vont considérer que même des « intouchables » d'une approche formaliste comme la grammaire, vont être avant tout régies par une intention des locuteurs d'utiliser la langue pour créer l'interaction et optimiser la communication.

3.5.1.1.2 *La compétence communicationnelle*

Qui dit utilisation de la langue, dit compétence : les grammaires fonctionnelles ne mettent donc pas de côté une notion de compétence innée, que les locuteurs devraient maîtriser. Pour Dik cependant, il ne s'agit pas de mécanismes mentaux abstraits comme chez Chomsky. Il parle plutôt de « compétence communicationnelle ». Les locuteurs disposent d'une compétence communicationnelle qui implique deux composantes :

- la **composante grammaticale**, c'est-à-dire l'aptitude à **construire** et à **interpréter** des expressions linguistiques et ;
- la **composante pragmatique** c'est-à-dire la capacité d'utiliser les expressions linguistiques d'une façon appropriée et efficace **conformément aux conventions** de l'interaction verbale en vigueur dans une communauté (on ne parle pas de règles grammaticales mais de règles d'emploi) pour atteindre l'effet communicationnel recherché (François & Cornish 1995 :13).

Dik définit la grammaire fonctionnelle comme une théorie de la composante grammaticale de la compétence communicationnelle. Le modèle néerlandais ne prévoit ni syntaxe ni sémantique autonomes. Il insiste sur la dimension interactionnelle de l'acte de communication. Dans sa conception, la fonction de la syntaxe consiste à permettre aux locuteurs de construire des énoncés complexes à l'aide desquels des significations complexes sont exprimables, et la sémantique a pour fonction de livrer aux locuteurs les contenus complexes à partir desquels ils peuvent poursuivre différents objectifs communicationnels (François & Cornish 1995 : 12). Il

présente ainsi une métalangue unique dans laquelle sont exprimées les représentations sous-jacentes qui sont conçues de manière à fournir l'entrée des règles d'expression.

La métalangue dans laquelle les représentations sous-jacentes (abrégé en « RSJ ») sont exprimées est polyvalente. Elle comportant des données d'ordre :

(i) sémantique.

Le modèle distingue dans un cadre des prédicats :

- (1) des grandeurs sémantiques (les termes, prédications, propositions et clauses [c'est-à-dire des propositions dotées d'une force illocutoire] ;
- (2) des rôles, i.e. fonctions sémantiques (agent, patient, destinataire, locatif, directionnel, origine, manière, instrument, etc.) et ;
- (3) des opérateurs et des satellites.

(ii) pragmatique

La métalangue proposée dans modèle est également pragmatique dans la mesure où les constituants des représentations sous-jacentes sont spécifiables en termes de Topique et de Focus, de Thème et de Rhème (cf.3.6.1.1.3)

(iii) syntaxique

La métalangue proposée dans le modèle s'avère aussi syntaxique par la mention des fonction sujet et objet en complément des rôles.

(iv) morphologique

Mais cette métalangue est aussi morphologique par la spécification des prédicats comme prédicat verbal, nominal ou adjectival ainsi par les d'expression qui confèrent un statut morphologique aux constituants des représentations sous-jacentes.

Les principaux symboles employés dans les représentations sous-jacentes s'interprètent comme suit :

X_j = quelque chose, quelqu'un (quelque entité) ;

E_j = quelque procès (état, événement, action, sus) ;

X_j = quelque proposition (procès doté d'une valeur de vérité) ;

E_j = quelque « clause » (proposition dotée d'une force illocutoire) ;

$\Omega_{1,2}$ = opérateurs applicables aux termes (nombre, définitude) ;

$\pi_{1,2,3,4}$ = opérateurs applicables respectivement aux prédicats, prédications, propositions et « clauses » ;

$\sigma_{1,2,3,4}$ = satellites respectivement de prédicat, prédication, proposition et « clause ».

Exemple : Paris ParisNp (X_i) \emptyset

Np symbolise « nom propre » et \emptyset le fait que la propriété d'« être-Paris » assignée à l'entité X_j ne nécessite pas l'introduction d'une fonction sémantique particulière (si l'on substituait à Paris le prédicat relationnel capitale $N[X_i\emptyset, (X_j)_{Loc}]$, X_i conserverait la fonction sémantique \emptyset , mais X_j (= la France) recevrait la fonction de localisateur.

On voit comment une transition s'opère entre une linguistique formelle et une linguistique de l'usage et de la performance, sans qu'on puisse parler d'une rupture avec certains principes fondamentaux de la recherche linguistique : on s'intéresse toujours à une forme, qui est de l'ordre de la langue et non de la parole (au sens saussurien), mais cette forme est désormais perçue différemment. On ne pense pas, chez les fonctionnalistes, que l'ordre des mots sera fortuit : il est régi par des règles formelles. Mais ces règles ont pour origine une optique communicative, qui n'exclut pas l'existence de contraintes d'ordre cognitif. Il s'agit de prendre conscience que tout énoncé est produit dans un contexte donné : celui-ci a un rôle prégnant dans la manière de construire les énoncés.

3.5.1.1.3 *Thème et rhème*

Dans un message, on a toujours des éléments déjà connus, qui constituent un point de départ pour l'énoncé : on le nomme **thème**, ou encore topique. Ensuite, on a des éléments nouveaux, importants pour le locuteur, qui constituent son apport d'information : c'est ce que l'on qualifie de **rhème**, ou focus. Les énoncés sont donc décrits à la mesure de leur valeur informative. Le rhème, par opposition au thème, a une valeur informative élevée ; on a comme une bipolarité dans les énoncés.

Quand on observe les langues, on se rend compte qu'il y a une tendance à ordonner les mots dans un agencement thème-rhème. On le considère donc comme l'ordre non marqué : il y aurait une tendance à adopter dans l'énoncé une sorte de « crescendo » canonique de l'information. Cette idée, empruntée aux cognitivistes (cf. *Opposition figure-ground*), témoigne d'une manière d'appréhender le monde : on part du connu pour aller vers l'inconnu.

3.5.1.2 Une nouvelle hiérarchie des contraintes

Avec la grammaire fonctionnelle, l'ordre « classique » des contraintes entre les différents systèmes de la linguistique est bouleversé. La pragmatique sera désormais considérée comme la donnée la plus importante pour rendre compte du langage, tandis que la sémantique et la syntaxe sont laissées de côté : la sémantique, avant tout au service de l'interaction, est subordonnée à la pragmatique ; la syntaxe est encore plus bas dans l'échelle, étant subordonnée à la sémantique pour « mettre en forme » son contenu propositionnel. Cela ne signifie pas que de tels niveaux ne doivent plus être pris en compte, mais qu'ils doivent toujours être considérés en relation avec les données qui sont d'ordre supérieur. Dik va par exemple étudier les contraintes qui permettent d'expliquer l'ordre des mots dans un énoncé (considération syntaxique), en mettant l'accent sur des aspects pragmatiques qui régissent cette syntaxe.

[L]a relation entre les différents constituants de l'organisation linguistique est considérée de telle sorte que la pragmatique est envisagée comme le cadre général englobant l'étude de la syntaxe et de la sémantique. La sémantique est considérée comme un outil de la pragmatique, et la syntaxe comme un outil de la sémantique. En ce sens, il n'y a pas de place pour une syntaxe qui serait « autonome ». (Dik, 1989 :7)

Le rôle des différents ordres de contraintes (syntaxe, sémantique) peut être capturé par les règles

de l'expression. Grâce à elle, la pragmatique est en mesure d'expliquer et de régir à la fois le **placement** (au niveau de l'énoncé) et la **formation** (au niveau du groupe de mot, dans la formation morphologique des adverbes, etc.) des constituants.

On parle de **règles de placement** et de **règles de formation**. Dans le cadre de ces règles, une série de principes interagissent, parmi lesquels certains sont plus forts que d'autres : l'idée sera que ceux qui sont au-devant de la scène sont d'ordre pragmatique.

On peut voir un exemple de cette régence pragmatique, dans la succession d'éléments dans l'énoncé (qui constitue une préoccupation syntaxique). La raison pour laquelle, dans une phrase non marquée, on a plus une tendance à voir des syntagmes très longs à la fin qu'au début de la phrase, s'explique par la pragmatique : elle préférera pour une communication optimisée une succession du thème au rhème. Le thème étant une information connue d'avance, le début de la phrase est donc, en principe, réservé à des énoncés relativement simples ; et inversement pour le rhème. Quelle que soit la langue que l'on étudie, il y a de cette manière des places dans l'énoncé qui vont soit accueillir une information thématique, soit rhématique : de fait, les positions « à droite » sont généralement rhématiques.

3.5.1.3 Typologie pragmatique des langues

La typologie des langues joue un rôle important dans la théorie de Dik. En effet, quelle que soit la langue que l'on étudie, l'objectif demeure de découvrir le principe de communication. Mais chaque langue développe en vérité sa propre solution au problème de la communication. Les langues varient sur le plan pragmatique, notamment au niveau des éléments rhématiques et thématiques.

Le marquage linguistique du focus (rhème) a été extrêmement étudié dans les grammaires fonctionnelles. Suite à de telles études, les linguistes fonctionnels se sont aperçus qu'il existait **cinq stratégies possibles** pour le marquage du focus, et que toutes les langues du monde utilisaient au minimum une seule d'entre elles (Dik 1997 :327) :

- **Ordre des mots.** Typiquement, le focus se signale alors en étant le plus « à droite » possible dans l'énoncé.
- **Syntaxe.** On utilise alors une construction où on met le focus en évidence. En français, cette stratégie est très courante dans ce que l'on nomme la construction clivée : « c'est *x* qui *p* » ; « c'est *y* que *q* ».
- **Morphologie.** La morphologie peut également être utilisée pour signaler le focus : **on va ajouter un suffixe spécial qui se met sur l'élément mis en évidence** (-ka 'か' et 'wa' は en japonais ; mais aussi en wolof, en inuktitut,...).
- **Lexique.** On utilise des mots pour marquer le focus, que l'on nomme des « particules de focalisation » En néerlandais par exemple, le mot « *maar* » remplit cette fonction : « *het is de wind maar* » ou « *het is maar de wind* ».
- **Prosodie et intonation.** A l'oral, la prosodie et en particulier l'intonation peuvent marquer le focus dans presque toutes les langues du monde.

À ce jour, on n'a trouvé rien d'autre que ces cinq stratégies de focalisation, dans toutes les langues du monde qui ont été observées. Quand on va se pencher plus spécifiquement sur une

langue ou une autre, on va se rendre compte que ces stratégies peuvent parfois être utilisées en combinaison. En français, ce procédé de surenchère d'accentuation est plus difficile à produire que dans d'autres langues européennes, du fait de la moins grande souplesse de sa syntaxe.

En comparant le français et le néerlandais, on peut s'apercevoir que les deux langues utilisent le même inventaire de stratégies : seule est absente celle de la morphologie. Mais leur fréquence dans les deux langues est différente. Les constructions clivées seront ainsi, en néerlandais, presque absentes de l'oral, alors que dans des corpus oraux francophones elles se rencontrent davantage encore qu'à l'écrit.

3.5.1.4 La question de l'acquisition

Les chercheurs ayant approché le problème de l'acquisition du langage d'un point de vue contextuel ont montré que le processus d'acquisition du langage est fortement co-déterminé par un input très structuré des données linguistiques, présentées à l'enfant dans des conditions naturelles, et adaptées au développement progressif de son niveau de compétence communicative (Dik 1989 : 6).

Chomsky nous disait que, du point de vue de l'apprentissage, on est obligé d'admettre que l'enfant a une capacité innée avec laquelle il vient au monde et sans laquelle il ne serait pas en mesure d'apprendre la langue. **De fait, l'enfant ne pourrait apprendre la langue sans le moindre appareil inné** : il est mathématiquement impossible d'apprendre toutes les phrases de la langue par cœur. Cela est encore moins défendable en tenant compte de la pauvreté de l'input reçu par l'enfant de ses parents. Les fonctionnalistes reconnaîtront à la pauvreté de cet input une fonction pour l'apprentissage de la langue : il permet de le lancer, en permettant une compréhension du fonctionnement de la communication avant d'en comprendre directement le contenu.

3.5.2 La pragmatique

La pragmatique, au sens américain, vise essentiellement à dégager les conditions de la communication (linguistique), telles, par exemple, que la manière, pour les deux interlocuteurs, d'agir l'un sur l'autre. Contrairement à la sémantique, la pragmatique part du principe que le langage ne fait pas que décrire la réalité, mais qu'il agit aussi sur elle. Celui qui parle accomplit une action (un acte de langage : promesse, ordre, déclaration, ...), qui ne peut être déclarée vraie ou fausse, mais plus ou moins réussie selon que le destinataire comprend l'intention du locuteur. En fait, le locuteur a l'intention de transmettre un certain contenu à l'aide du code qu'est la langue. Le destinataire devra décoder pour comprendre le message. Cependant, même si le code est parfaitement partagé, la communication peut échouer, parce qu'elle contient une part de non-dit, d'implicite. Pour pouvoir en rendre compte, la pragmatique élabore un modèle qui explique comment, à partir d'informations contenues dans l'énoncé, et d'autres fournies par le contexte, le destinataire émet des hypothèses sur l'intention du locuteur.

Cette pragmatique du langage a trait aux caractéristiques de son utilisation ; elle concerne la compétence cognitive des sujets communicants telle qu'on peut la reconnaître à l'intérieur des discours-énoncés : le faire persuasif et le faire interprétatif ne constituent pas des paramètres « extralinguistiques » comme pourrait le laisser entendre une certaine conception mécaniste de la communication (voir le modèle de Shanon & Weaver), mais entrent à part entière dans le processus de la communication – tel qu'il est envisagé par la sémiotique – où le

destinateur et le destinataire ne sont pas des instances vides (tels l'émetteur, le récepteur), mais des sujets compétents.

La pragmatique pourra également être développée dans d'autres cours, nous revenons ici sur les principes fondateurs qui continuent d'exercer leur influence aujourd'hui, notamment le **principe de coopération** qui insiste sur les règles du discours.

En effet, selon Herbert Paul Grice, le discours est une activité réglée. L'idée principale de Grice est que les locuteurs, pour communiquer et se comprendre dans la conversation, respectent un principe général rationnel non spécifique à l'activité verbale, « le principe de coopération », et un certain nombre de règles, ou maximes conversationnelles. Il formule ce principe de la manière suivante :

« Que votre contribution à la conversation soit, au moment où elle intervient, telle que le requiert l'objectif ou la direction acceptés de l'échange verbal dans lequel vous êtes engagé » (traduction de Wilson & Sperber 1979 : 93). Le propos du locuteur doit être conforme à ce qu'on attend de lui au moment où il intervient, à l'intérieur d'un échange discursif dont le but ou la direction sont partagés par les interlocuteurs.

Ce principe est complété par quatre catégories de règles ou *maximes* (cf. tableau suivant) dont le respect devrait donner les résultats en accord avec ce principe. Ces *maximes* sont pour la communication autant de « règles de conduites ».

Principe de coopération

Maxime de quantité	Maxime de modalité / de manières
« Que votre contribution soit aussi informative que nécessaire ; ne donnez pas plus d'information que ce qui est requis. »	« Ne dites pas ce que vous croyez être faux, ni des choses pour lesquelles vous manquez de preuves. »
Si on demande à quelqu'un « tu n'as pas 1000BIF » et que la personne répond « oui j'en ai 5000 » elle a donné une information superflue : elle n'a pas respecté la maxime de quantité.	Un principe de base pour une communication coopérative est qu'on se doit de ne pas mentir : on doit dire des informations qui sont véridiques.
Maxime de qualité (ou de véridicité)	Maxime de relation (ou de pertinence)
« Ne dites pas les choses pour lesquelles vous manquez de preuves. Évitez l'ambiguïté. Soyez bref. Soyez ordonné. »	« Soyez pertinent. » (donner la contribution à propos)
Ici, on demande d'être le plus clair possible. Si on a par exemple l'énoncé « <i>Lucky Luke partit et sauta sur son cheval</i> » cet énoncé n'est pas clair, vu que le bon sens aurait voulu qu'on s'exprimât de la manière inverse.	La réponse que nous donnons doit être en relation avec ce que nous avons dit ou ce que nous allons dire. On peut considérer là qu'il s'agit d'une sous-catégorie de la maxime de qualité.

Le respect de ces maximes garantit la réussite de l'échange d'informations. Toute personne qui entreprend un échange discursif est présumée respecter les maximes ou au moins le principe de coopération dont elles dérivent.

Cependant, il arrive que les maximes soient transgressées : mensonges ou conversations apparemment hors propos, etc. Dans certains cas, la tâche du destinataire consiste à extraire une information supplémentaire afin de réconcilier les propos du locuteur avec la présomption du respect du principe de coopération. Ainsi, si Pierre dit à Marie : « Tu es le soleil de ma vie », en première analyse, cet énoncé paraîtra enfreindre la maxime de qualité (Marie n'est pas un soleil). Cependant, Marie peut réconcilier propos de Pierre avec le principe de coopération en extrayant une information du type : « Pierre veut me faire un compliment ; il utilise une image positive, le soleil ». Grice appelle implicature conversationnelle cette information supplémentaire, ce sous-entendu. Selon lui, le principe, les maximes et les implicatures constituent ce qui permet l'interprétation du propos du locuteur ainsi que de ses intentions. L'idée est que si toutefois le locuteur déroge à l'une de ces maximes, son interlocuteur va interpréter son énoncé en postulant que le principe de coopération est malgré tout maintenu : **la maxime violée l'a été intentionnellement pour introduire une plus-value de sens, liée au contexte**. En somme, l'interlocuteur postule que par défaut, le locuteur se conforme à ce qu'on attend de lui dans un contexte donné.

La linguistique du XXe siècle a opposé une linguistique de la langue à une linguistique de la parole. La pragmatique est une linguistique du discours, qui intègre les situations et contextes de communication. Ce faisant, elle ouvre la porte à l'étude de ce qui n'est pas dit, de l'implicite.

3.5.3 La sociolinguistique

Le contexte revêt dans le cadre de la sociolinguistique une acception plus large encore que dans le cadre de la grammaire fonctionnelle ou de la pragmatique de première génération. Comme dans les courants précédents, c'est à l'influence du contexte sur l'usage de la langue que l'on va s'intéresser, mais on va bien vite s'apercevoir que cette influence va bien au-delà d'une influence situationnelle, comme dans le cas de la pragmatique, ou d'une influence liée à la communication.

D'une influence centrée sur la parole, on va passer à une influence reconnue du contexte sur la langue en tant qu'objet psycho-social.

William Labov, linguiste américain, est considéré comme le père fondateur de la sociolinguistique. Il a été le premier à développer une étude de la variation dotée d'un arsenal descriptif véritablement performant (description quantitative, modèles théoriques,...). On lui doit de nombreuses études dans lesquelles il insiste sur la variation *diastratique*³² et *diaphasique*, des « axes » de la variation qui jusque-là n'avaient pas fait l'objet d'un intérêt suffisant. Par après, il donnera également un éclairage social à la variation diachronique en expliquant la relation entre variation et changement.

³² La **variation diastratique** relève les différences d'usage d'une même langue au sein des couches sociales. À titre d'exemple, le français des classes sociales élevées de Paris représente en France, la norme, le modèle à suivre, contrairement à certaines variétés dépréciées, voire stigmatisées (l'accent dit de « banlieue » par exemple). La **variation diaphasique** rend compte des usages différant d'une situation de discours à une autre. La production langagière est déterminée par le caractère « formel » ou « informel » de la situation de communication.

Labov part du principe qu'on ne peut isoler plus longtemps les systèmes linguistiques de leurs utilisateurs. De fait, la « linguistique de bureau » mettant en avant l'approche introspective hypothético-déductive est critiquée dans le sens où elle ne rend pas compte des variations et des phénomènes langagiers qui prennent place dans la société ; d'où l'intérêt de la « linguistique de terrain » qui remplit bien cette tâche.

La condition de cette nouvelle pratique de la linguistique est une méthodologie d'observation qui lui permette de lire avec précision l'incidence des interactions sociales sur la structure de la langue. L'objectif n'est pas tant d'apporter à la linguistique une nouvelle théorie mais une méthode. C'est pourquoi les données recueillies doivent constituer autant d'arguments empiriques susceptibles de valider ou d'invalider l'élaboration théorique. (Boyer et Prieur, 1996 : 36).

3.5.3.1 Les précurseurs

3.5.3.1.1 *La dialectologie*

La sociolinguistique trouve ses origines dans les premiers développements de la dialectologie³³. Le principe de base de la dialectologie, qui sera également celui de la sociolinguistique, est celui de la **variation**. Une langue n'est pas un système uniforme dans le chef de tous les locuteurs qui la pratiquent. La dialectologie permet de montrer qu'à partir d'une même aire linguistique, on peut diviser les prononciations ; que la prononciation canonique n'existe pas si on exclut tout positionnement normatif.

3.5.3.1.2 *La phonétique*

La phonétique a également précédé la sociolinguistique dans l'étude d'un domaine spécifique : **la variation phonétique**. Deux figures se doivent d'être citées en la matière : l'abbé français **Jean-Pierre Rousselot** et le suisse **Louis Gauchat**.

Rousselot est considéré comme le fondateur de la phonétique expérimentale tant en recherche fondamentale qu'en recherche appliquée. Il a proposé d'étudier les modifications phonétiques des parlers par une méthode expérimentale et d'en induire les lois phonétiques des changements en cours et plus généralement les processus analysés jusque-là par la linguistique historique.

Louis Gauchat, quant à lui, a créé le « Glossaire des patois de la Suisse romande », un institut fondé en vue de la publication d'un glossaire complet des patois romands.³⁴

À cette époque, l'enregistreur devient à la linguistique ce que l'appareil photo était devenu aux enquêtes policières : pour la première fois, la variation peut être empiriquement étudiée, et on est en mesure d'en conserver des traces et d'établir des bases de données chiffrées par la collecte d'enregistrements.

3.5.3.2 Missions de la sociolinguistique

Le langage n'est pas une pratique individuelle : c'est surtout une pratique sociale. En parlant, nous ne mettons pas seulement en jeu notre individualité : nous montrons notre attachement à

³³ Cf. point 2.1.3. sur les **néogrammairiens**, parmi lesquels l'italien **Graziado Ascoli** est, pour rappel, le fondateur de la dialectologie.

³⁴ Source : Wikipédia (fr.wikipedia.org), [Jean-Pierre Rousselot](#) ; [Louis Gauchat](#).

un groupe, à une communauté. Par rapport à l'idée de norme que peut se faire la grammaire, la sociolinguistique s'intéresse donc essentiellement à l'usage, ainsi qu'à la pratique orale de la langue. La tâche de la sociolinguistique est donc d'étudier l'interaction entre notre pratique du langage et les phénomènes sociaux qui nous entourent. Il s'agit d'expliquer les phénomènes linguistiques à partir de données extralinguistiques, autrement dit de facteurs politiques et sociaux. Plus particulièrement, il sera question de déterminer à quelle structure sociale répond une structure linguistique donnée et comment, d'une manière générale, les changements de structure sociale se traduisent par des changements linguistiques (Sioufi & Raemdonck 1999 :35 citant Meillet 1906).

Selon Labov, en effet, la sociolinguistique doit être en mesure d'expliquer et de décrire les variations dans l'usage de la langue, tant à l'échelle microsociale (au niveau de l'individu et des relations interindividuelles) qu'à l'échelle macrosociale (au niveau d'une communauté entière). La tâche consistera à décrire et à expliquer les variations tant chez des individus pris séparément que dans un groupe plus large ; le langage reflétant les relations sociales et étant marqueur identitaire.

La sociolinguistique pourra d'abord s'intéresser aux cas de variation à l'intérieur d'une même langue. La manière spécifique qu'a un individu de parler sera appelée *idiolecte*. Il arrive que le langage parlé par une communauté de locuteurs présente des différences significatives d'avec celui parlé par une autre communauté de la même langue. Si cette différence est d'origine régionale, on parlera de *dialecte* (objet qu'étudie la dialectologie). Si cette différence est d'origine sociale, on parlera de *sociolecte* ; et l'étude des sociolectes fonde une partie de la sociolinguistique.

Une autre direction de la sociolinguistique, qu'on appelle parfois ethnographie de la communication, s'intéresse à la manière dont les éléments linguistiques s'organisent dans le cadre d'un échange de manière à créer un contexte de communication. L'emploi de tel élément de lexique, par exemple, de telle intonation, de tel segment de syntaxe, peut avoir une signification particulière à l'intérieur des pratiques de langage d'un groupe donné. Dans une conversation réunissant les deux sexes, par exemple, l'emploi de telle ou telle tournure par l'un des participants pourra avoir un effet de définir à l'intérieur de cette conversation un micro-auditoire formé par les personnes d'un sexe ou de l'autre.

La sociolinguistique constitue ainsi une (énime) réaction à la position qui avait été tenue par Chomsky, selon lequel tout linguiste devait se placer au niveau de la compétence et non de la performance. La sociolinguistique a pour mission de montrer que contrairement à la vision chomskyenne, la performance n'est pas aléatoire. La variation qui la conditionne en effet n'est pas libre, mais structurée par des règles sous-jacentes.

3.5.3.3 Études fondatrices

3.5.3.3.1 *Martha's Vineyard (1963)*

William Labov, lors de vacances touristiques, se rend sur une petite île au large du Massachusetts, Martha's Vineyard. Durant son séjour dans cette île où le tourisme se développe depuis peu, il constate que certains autochtones sont ravis de recevoir les touristes fortunés, tandis que d'autres sont beaucoup plus refroidis à cette idée. En outre, il remarque une variation

caractéristique de l'accent de l'île mais qui n'est pas pratiquée dans les mêmes proportions chez tous les habitants.

Il fera l'objet de sa thèse de doctorat d'étudier une corrélation existant *a priori* entre ces deux phénomènes. Les habitants favorables aux touristes semblent prononcer de la même manière que ceux-ci, dans un américain plus standard, tandis que ceux qui leur sont plus défavorables, parmi lesquels les derniers marins de l'île, pratiquant encore la pêche au gros, auront tendance à conserver, voire à exagérer l'accent local. Son travail a bien montré que les variantes en cause (formes centralisées et non centralisées des deux diphtongues) définissent les identités conflictuelles, insulaire et continentale, opposant les habitants de l'île et se répartissant dans les différents sous-groupes (Boyer, Prieur, 1996 : 42).

3.5.3.3.2 *Prononciation du /r/ dans les grands magasins new-yorkais (1966)*

Labov s'intéresse également à la variation liée à la classe sociale dans la prononciation. Une étude fondatrice en la matière est celle qu'il a réalisée à propos de la prononciation du /r/ à New York, *The Social Stratification of English in New York City* (1966). Il s'est rendu, pour la réaliser, dans trois chaînes de magasins de vêtements dont l'offre était catégorisée comme de luxueuse (*Saks Fifth Avenue*), adressée aux classes moyennes (*Macy's*) ou aux classes populaires (*S. Klein*). Ayant choisi les magasins de telle sorte qu'à chaque fois, les bas en nylon se trouvaient au quatrième étage (« fourthfloor »), il demandait aux clients la localisation de ceux-ci, muni d'un enregistreur.

Son objectif était d'étudier la variation de prononciation, du plus luxueux au moins luxueux des magasins.

Il constate que la chaîne la plus luxueuse, Saks, est également celle dans laquelle le plus de /r/ sont prononcés. On désire avoir une prononciation plus « châtiée » : on prononce mieux, on prononce tout. En revanche, chez S. Klein, l'insistance sur la prononciation des /r/ est presque inexistante.

3.5.3.3.3 *La langue des ghettos*

On voit bien, avec *The Social Stratification of English in New York City*, l'ancrage de la question sociale dans la pratique du langage. Toutefois ce constat avait déjà été réalisé, et des descriptions avaient été tentées bien avant l'avènement de la sociolinguistique. C'est ainsi qu'on s'est beaucoup intéressé à l'emploi de l'anglais, par exemple, au sein des populations afro-américaines.

Le gouvernement américain avait constaté que les populations noires des ghettos avaient souvent du mal à s'adapter et étaient gênés dans leur ascension sociale par certains traits culturels qui les « excluaient » de la population américaine, et notamment au niveau de leur langue. On a donc cherché à remettre à niveau cette population des ghettos. L'idée initiale était que les afro-américains utilisaient un code restreint « restrictive code » : ils produisaient des énoncés très courts souvent incomplets, bien souvent de forme paratactique. Leurs emplois lexicaux étaient peu variés, leur vocabulaire relativement vague.

Leur langue pourrait donc, selon l'hypothèse de départ, être une variante diminuée par rapport au code élaboré « elaborated code » de l'anglais américain standard, et cela à cause de facteurs d'une part psychologiques (les noirs des ghettos, à cause de leur apprentissage de la

langue, auraient développé une capacité moindre à développer leur discours) ; mais aussi sociologiques (cette utilisation de la langue serait le produit d'une pratique limitée à une communauté trop refermée sur elle-même).

Le sociologue et linguiste Basil Bernstein a plus spécialement mis ces codes en relation avec les classes sociales, en mettant en évidence que les classes sociales inférieures utilisaient une forme de code restreint, tandis que les classes plus élevées auraient, en plus d'un *code restreint*, un *code* élaboré, efficient. Ce qu'il tient pour un état de fait aurait évidemment un impact sur le parcours scolaire, et d'une manière plus générale sur la réussite dans la vie : les personnes issues des milieux sociaux inférieurs, qui ne possèdent pas les registres élaborés typiques du système supérieur, n'ont pas aussi facilement accès à l'apprentissage du code élaboré. Il suggère alors au gouvernement américain le développement de programmes scolaires adaptés, baignant davantage les élèves dans le code élaboré.

Labov : mise en évidence du paradoxe de l'observateur

Toutefois, on se rend bien vite compte que ces programmes scolaires ne fonctionnent pas, ou très peu. La question que pose alors Labov est de savoir si on peut réellement qualifier l'emploi linguistique de ces personnes de déficient, du point de vue cognitif.

Il souligne que ces personnes qui sont enregistrées n'avaient pas l'habitude de se retrouver dans un petit local, à côté d'une lampe,... pour voir un étrange individu allumer un enregistreur et les interroger sur leur façon de s'exprimer. Il remarque que l'expérience avait en fait toutes les chances d'être biaisée par les conditions de l'observation.

Labov va alors réaliser la même expérience, en en changeant les conditions. Il se présente aux jeunes noirs de manière informelle, en leur présentant un paquet de chips ; la personne en face de lui peut parler de ce dont elle veut et aborder des sujets de conversation qui l'intéressent ; en somme, l'observateur essaie de se fondre dans la communauté linguistique et de s'adapter à la personne qui passe le test.

L'enquête ainsi réalisée par Labov va produire des résultats différents. On se rend compte que la compétence linguistique des personnes interrogées était élaborée, mais que leur problème est principalement **une question de registre et de cadrage de l'interaction**. Leur code est bien fonctionnel, et adapté à l'intérieur de la communauté utilisant une pratique uniforme, avec ses propres fonctionnements, même si cette variante est « non standard ».

En réalisant cette expérience, Labov a mis en évidence le problème méthodologique du paradoxe de l'observateur : les données doivent refléter la manière dont la personne interagit dans un contexte réel de production. Avant lui, on négligeait le contexte en pensant qu'il suffisait de prendre un tiers informateur et de l'enregistrer. La contestation faite par Labov de cette approche a conduit à une reconsidération méthodologique et à un regard nouveau sur les variétés non standard.

*

*

*

On peut voir donc, à travers ces études, comment **le contexte social détermine l'usage**

de la langue, que l'usage de la langue reflète en réalité les rapports entre les locuteurs. Ainsi, le cours de sociolinguistique aura à revenir en détails sur les facteurs de variation et les formes de l'usage de la langue en société. La sociolinguistique montre que les formes de l'usage de la langue sont multiples et qu'on peut les relier à de nombreux paramètres. En effet, on peut tout d'abord mentionner les registres de langue, les différentes formes d'argot, de jargon ou de technolecte (le langage particulier à un métier ou à un milieu professionnel, la médecine, par exemple). Mais il faut faire également leur part aux questions d'âge, de sexe et de localisation géographique.

Bref, comme on le voit, le champ d'études de la sociolinguistique est très vaste. Qu'elles soient réalisées sur de grands ensembles de locuteurs ou à petite échelle, ces études démontrent à quel point le langage est une fabrication sociale, et à quel point il échappe souvent à la volonté individuelle de ses locuteurs. L'usage de la langue répond à des contraintes sociales ; chaque locuteur dispose de registres de langue parmi lesquels il va opérer un choix. L'analyse sociolinguistique de la langue se doit donc de rendre compte de la contrainte sociale à laquelle le locuteur est soumis. Ces contraintes ne sont pas à voir comme des lois immuables, mais comme des éléments qui vont motiver le choix du locuteur d'un certain registre de langue dans un répertoire de « paradigmes interactionnels » possibles. Ce choix va être motivé notamment par la personne à qui on parle, du sujet ou du propos sur lequel on cherche à s'entretenir, de l'évocation que l'on fait du contexte, et du but recherché par le locuteur.

4 LANGAGE ET COMMUNICATION

4.1 La communication linguistique

4.1.1 Le système linguistique (Rappel)

La linguistique est généralement définie comme l'étude scientifique du langage. Mais cette discipline, qui part donc de l'étude du langage, en arrive surtout à s'occuper des langues. Elle a comme objet spécifique (visé en dernière analyse) la langue, et a comme objet étudié (qu'il faut connaître au départ) le langage. La linguistique étudie principalement et, marginalement, le langage, instrument de communication par manifestation vocale. Il importe de remarquer dès lors que deux conceptions de la linguistique coexistent :

- (1) *La linguistique comme description des langues* : elle s'appuie sur l'observation objective du comportement linguistique des sujets parlants ; elle décrit tout ce qui caractérise en propre une langue reconnue comme telle, considérant que l'étude d'un état de langue peut avoir une valeur explicative, et pas seulement descriptive. Se proposant comme idéal les méthodes des sciences de la nature et l'objectivité du physicien, **elle considère une langue comme un système de signes linguistiques.**
- (2) *La linguistique comme l'étude du fonctionnement du langage* : elle s'appuie sur une conception unitaire du langage humain, considérant que les langues particulières ne sont que des cas particuliers du langage. Les langues diffèrent les unes des autres, non dans leurs structures profondes, mais du fait des variations tout à fait superficielles ; une analyse approfondie ferait apparaître des universaux linguistiques qui témoignent d'une faculté naturelle du langage. Il y aurait ainsi un répertoire universel d'éléments sémantiques ou phonétiques dans lequel chaque langue choisirait les éléments de base de sa combinatoire. À la notion de langue comme système de signes est substituée la notion de langue comme système de règles. D'où l'on pourra définir la linguistique comme l'étude scientifique du langage humain, vocal, tel qu'il se réalise par une langue, système de signes linguistiques et/ou système de règles (voir Christian BAYLON & Paul FABRE, 1975 :17).

4.1.2 Caractéristiques de la communication linguistique

La communication linguistique implique l'utilisation du langage articulé, système de signes discrets, phoniques, oraux, vocaux, ou celle du langage écrit, code de signes substitutifs du langage parlé.

Un code présuppose en effet des unités préexistantes. Il existe un code graphique, l'écriture, qui offre, pour la forme phonique des unités linguistiques, des équivalents visuels mieux adaptés aux besoins de la conservation des messages. En face de ce code, il n'y a pas un autre code, mais la langue, « organisation sui generis des données de l'expérience », dont les unités (morphèmes, phonèmes) ne préexistent pas à la langue « parce qu'une langue ne consiste

pas en des étiquettes accrochées à des réalités données une fois pour toutes, de toute éternité, et identiques d'une communauté humaine à une autre ». Le code part toujours d'un message déjà formé pour aboutir à un autre message exprimé par des symboles différents ; la langue peut servir à renseigner directement sur la réalité extralinguistique.

Un message de la langue naturelle est construit avec des segments minimaux de deux espèces : la 1^{ère} articulation et la 2^e articulation du langage

(a) La première articulation du langage. C'est celle qui construit l'énoncé en unités significatives successives minima ou morphèmes :

Ex. [1] *Le vent souffle* : 3 morphèmes

[2] *Em-barqu-ons sur le bateau* : 6 morphèmes

Le morphème est la plus petite unité ayant à la fois une forme (Sa) et un sens (Sé).

(b) La deuxième articulation du langage. C'est celle qui construit l'unité significative elle-même à partir d'unités successives distinctives, les phonèmes.

[3] /vã/ : 2 phonèmes

[4] /bato/ : 4 phonèmes

Le phonème est un élément du Sa, du morphème. Il possède une forme phonique mais pas de signifié. Il permet de distinguer les morphèmes : chacun des phonèmes de /vã/ suffit à le distinguer d'autres morphèmes comme /bã/, /mã/, etc., et comme /võ/, /vi/, etc.... Bref, le phonème est la plus petite unité de la chaîne parlée qui ait une valeur pertinente d'opposition. Selon A. Martinet, « le langage humain peut être décrit comme doublement articulé : articulé d'abord sur les deux plans du contenu et de l'expression ; articulé ensuite sur le plan d'expression seulement ».

Il est difficile d'opposer la communication animale au langage humain par des caractères spécifiques : absence d'appareil vocal chez les insectes ; absence de réponse autre que comportementale ; absence de communication relayée ; caractère mimique de la communication ; pauvreté des messages sémantiques, etc. Pour George Mounin, « l'absence d'unités discrètes et l'absence de la double articulation (qui seule rend compte de l'infinie variété des messages possibles avec l'arbitraire du signe) suffit, seule, à nous empêcher pour l'instant de parler de langage des abeilles. Dans le système de A ou communication animale, on peut penser que l'unité la plus petite est le message global, l'énoncé complet, et que chaque message s'oppose à tous les autres, globalement.

4.1.3 Les éléments constitutifs de toute communication verbale

4.1.3.1. La communication chez F. de Saussure

Saussure présente la communication comme un événement social. On peut l'observer dans l'acte de parole. Il faut au moins deux personnes pour que s'établisse ce que Saussure appelle le circuit de la parole qu'on peut résumer comme suit :

Soit deux individus A et B qui s'entretiennent. Le circuit peut prendre naissance dans le cerveau de A où « *les faits de conscience que nous appellerons concepts se trouvent associés aux représentations des signes linguistiques ou images acoustiques servant à leurs expressions* ».

Il se passe donc quelque chose dans le cerveau de A : une impulsion y est donnée aux organes de la parole pour qu'ils produisent les sons appropriés. Ces sons se trouvent transmis sur les

ondes sonores de la bouche de A à l'oreille de B, puis de à son cerveau. Si B répond, un deuxième acte de parole se produit, la transmission se faisant cette fois du cerveau de B à sa bouche puis à l'oreille et enfin au cerveau de A, et ainsi de suite tant que dure la conversation.

4.1.3.2. La communication chez L. Bloomfield

Nous trouvons, chez Bloomfield, une autre présentation de la communication linguistique dans l'apologue bien connu de Jack et Jill. « Jack et Jill se promènent dans un sentier. Jill a faim. Elle voit une pomme sur un arbre. Elle produit des bruits au moyen de son larynx, de sa langue et de ses lèvres. Jack saute la barrière, grimpe à l'arbre, prend la pomme et la pose dans la main de Jill. Jill mange la pomme.

Bloomfield pose le problème en termes behavioristes : il décrit ce qui peut être observé de l'extérieur. Il distingue entre practical events – c'est-à-dire les faits et gestes et, de façon générale, la situation telle qu'elle se présente immédiatement avant l'acte de parole – et l'acte de parole lui-même (speech), et il analyse le tout en trois moments :

- A. La situation précédant l'acte de parole ;
- B. La parole ;
- C. La situation succédant l'acte de parole.

De A, il tire le stimulus du locuteur, et de C la réaction, ou « réponse » de l'auditeur, et insiste tout particulièrement sur l'intérêt et l'importance de cette réponse pour Jill. En effet, note-il, Jill, livrée à ses seules ressources, devrait, ou cueillir la pomme elle-même, ou rester sur sa faim, et se trouverait donc dans la même position que l'animal privé de parole. La faim et la vue (ou l'odeur) de la nourriture sont le stimulus S, et les mouvements pour atteindre cette nourriture la réaction R. ce procès, le même pour Jill solitaire et pour l'animal sans parole, est présenté ainsi :

S > > —————> R

En accomplissant l'acte de parole, Jill a pu faire réagir Jack à sa place. Le langage, donc, permet de susciter une réaction (R) chez une personne, quand une autre personne a éprouvé le stimulus (S). Bloomfield procède alors à l'analyse de l'acte de parole comme suit :

1. Réaction substitutive (R) correspondant à l'énoncé de Jill et
2. Stimulus substitutif (s) correspondant à l'audition de Jack.

La totalité du procès est représentée comme suit :

S> r s> R

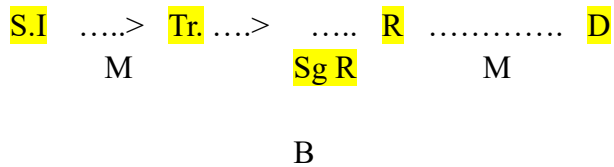
S étant le stimulus effectif (la faim de Jill et la vue de la pomme), la réaction active – négociée par la parole – (Jack saute la barrière, grimpe à l'arbre) ; r, la réaction substitutive (Jill produit un énoncé) ; les pointillés, les ondes sonores ; s, le stimulus substitutif (l'énoncé de Jill entendu par Jack). De tout ceci, finalement, Bloomfield ne retient que la section /r...s/, à savoir l'énoncé parlé, et y voit le seul objet qui devrait retenir l'attention du linguiste. Tout ce qui lie l'énoncé parlé B et ce que nous avons résumé comme les situations A et C, constitue le sens, dont l'importance est capitale pour l'homme, mais qui est antérieur au champ de la linguistique.

Bloomfield comme Saussure, met en évidence le caractère social de la langue et le fait qu'il exige des communautés linguistiques, c'est-à-dire des groupes humains qui utilisent des

systèmes spécifiques de signaux parlés.

4.1.3.3. La communication chez Shannon et Weaver

Dans le schéma de Shannon et Weaver, illustrant un système général de communication, ce qui prend le pas, c'est la transmission de l'information :

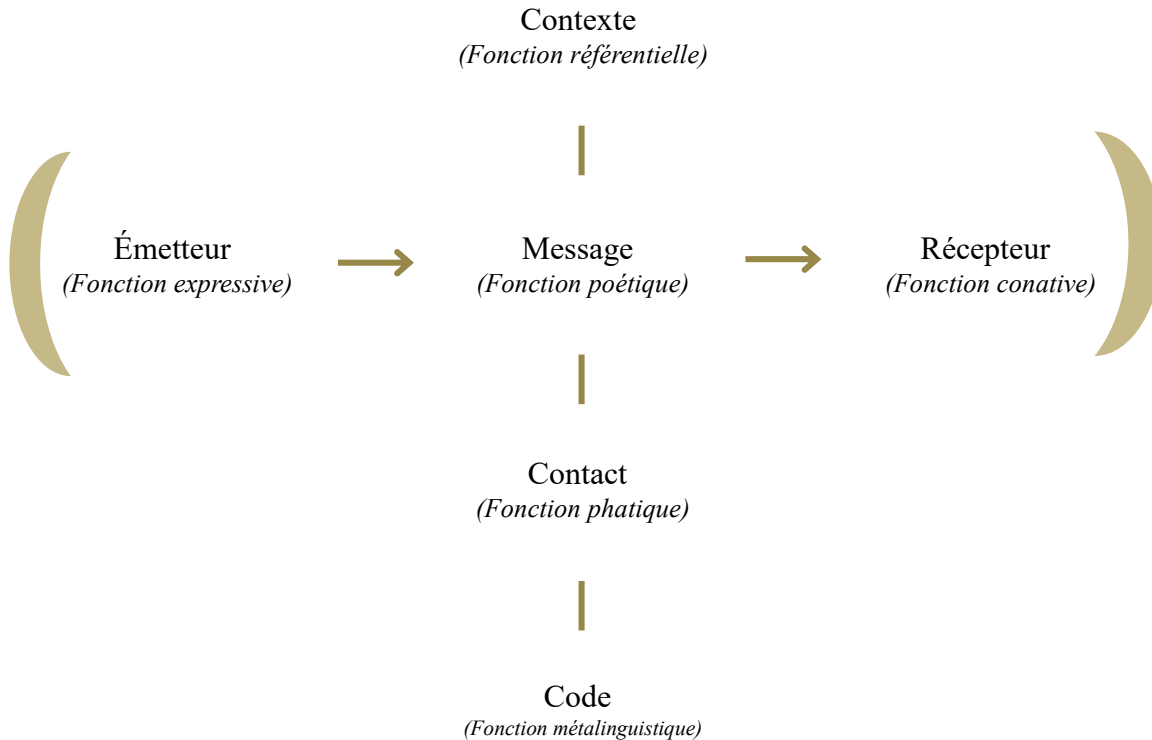


Selon cette présentation, une source d'information (S.I) produit un message ou une suite de messages (M.) qui doit être transmis à un récepteur (R). Le transmetteur (Tr.) traite le message afin de produire un signal susceptible d'être transmis par un canal (par exemple le codage d'un télégramme donnant une suite de points, de tirets et d'espaces). Le signal reçu (Sg R) est alors décodé par le récepteur, afin de restituer le message, qui atteint finalement sa destination (D). Les opérations d'encodage et de décodage sont exécutées au moyen d'un code. La transmission peut être entravée par des bruits, et l'élimination du bruit ou des bruits va constituer une préoccupation majeure. On se rappellera que Shannon et Weaver sont des ingénieurs, et que pour eux « *le problème fondamental de la communication consiste à reproduire, en un point, exactement ou approximativement, un message prélevé à un autre point. Le plus souvent, ces messages ont un sens, c.à.d. qu'ils se réfèrent à certaines entités physiques ou conceptuelles, ou se trouvent en relation avec celles-ci. Ces aspects sémantiques de la communication n'ont pas de pertinence du point de vue des problèmes techniques* ».

Ce qu'on retient du schéma de Shannon et Weaver c'est que des informations ou messages sont transmis, au moyen de signaux d'un émetteur à un récepteur, selon un canal. On peut considérer que l'information est une substance qui prend la forme d'un message par application d'un code. L'appareil constitué par le transmetteur, le canal et le récepteur peut être désigné comme le médium.

4.1.3.4. La communication chez R. Jakobson

Tout processus de communication met en relation, par un canal (support physique du message) minimalement deux interlocuteurs disposant d'un code commun dans un contexte particulier. Se fondant sur l'analyse de l'acte sémiologique global (le langage devant être étudié dans la variété de ses fonctions), Roman Jakobson propose une série de termes constitutifs de tout procès linguistique, de toute communication verbale :



Chacun de ces six éléments donne naissance à une fonction linguistique différente.

Toutefois, si nous distinguons six aspects fondamentaux dans le langage, il serait difficile de trouver des messages qui rempliraient seulement une et une seule fonction. La diversité des messages réside non dans le monopole de l'une ou l'autre fonction, mais dans les différences de hiérarchie entre celles-ci. La structure verbale d'un message dépend avant tout de la fonction prédominante. Mais, même si la visée du référent, l'orientation vers le contexte, bref **la fonction dite « dénotative », « cognitive », référentielle** - est la tâche dominante de nombreux messages, la participation secondaire des autres fonctions à de tels messages doit être prise en considération par un linguiste attentif.

La **fonction dite « expressive »** ou émotive, centrée sur le destinataire, vise à une expression directe de l'attitude du sujet à l'égard de ce dont il parle. Elle tend à donner l'impression d'une certaine émotion, vraie ou feinte; c'est pourquoi la dénomination de fonction « émotive », proposée par Marty, s'est révélée préférable à celle de « fonction émotionnelle ». La couche purement émotive, dans la langue, est présentée par les interjections. On n'a pas le droit de restreindre la notion d'information à l'aspect cognitif du langage. Un sujet, utilisant des éléments expressifs pour indiquer l'ironie ou le courroux, transmet visiblement une information, et il est certain que ce comportement verbal ne peut être assimilé à des activités non sémiotiques. La différence, en français, entre [si] et [si:], avec allongement emphatique de la voyelle, est un élément linguistique conventionnel, codé, tout autant que, en tchèque, la

différence entre voyelles brèves et longues, dans des paires telles que [vi] « vous » et [vi:] « sait » ; mais, dans le cas de cette paire-ci, l'information différentielle est phonématique, tandis que dans la première paire elle est d'ordre émotif.

L'orientation vers le destinataire, la **fonction conative**, trouve son expression grammaticale la plus pure dans le vocatif et l'impératif, qui, du point de vue syntaxique, morphologique, et souvent même phonologique, s'écartent des autres catégories nominales et verbales. Les phrases impératives diffèrent sur un point fondamental des phrases déclaratives: celles-ci peuvent et celles-là ne peuvent pas être soumises à une épreuve de vérité.

Il y a des messages qui servent essentiellement à établir, prolonger ou interrompre la communication, à vérifier si le circuit fonctionne (« Allô, vous m'entendez ? »), attirer l'attention de l'interlocuteur ou à s'assurer qu'elle ne se relâche pas (« Dites, vous m'écoutez ? ») ou, en style shakespearien, « Prêtez-moi l'oreille ! » - et, à l'autre bout du fil, « Hm-hm ! »). Cette accentuation du contact - **la fonction phatique**, dans les termes de Malinowski – peut donner lieu à un échange profus de formules ritualisées, voire à des dialogues entiers dont l'unique objet est de prolonger la conversation. L'effort en vue d'établir et de maintenir la communication est typique du langage des oiseaux parleurs; ainsi la fonction phatique du langage est la seule qu'ils aient en commun avec les êtres humains. C'est aussi la première fonction verbale à être acquise par les enfants; chez ceux-ci, la tendance à communiquer précède la capacité d'émettre ou de recevoir des messages porteurs d'information.

Une distinction a été faite dans la logique moderne entre deux niveaux de langage, le « langage-objet », parlant des objets, et le « métalangage » parlant du langage lui-même. Mais le métalangage n'est pas seulement un outil scientifique nécessaire à l'usage des logiciens et des linguistes; il joue aussi un rôle important dans le langage de tous les jours. **Chaque fois que le destinataire et/ou le destinataire jugent nécessaire de vérifier s'ils utilisent bien le même code, le discours est centré sur le code: il remplit une fonction métalinguistique (ou de glose)** (« Je ne vous suis pas, que voulez-vous dire ? » demande l'auditeur, ou, dans le style relevé: « Qu'est-ce à dire ? »). Et le locuteur, par anticipation, s'enquiert : « Comprenez-vous ce que je veux dire ? »). L'information que fournissent toutes ces phrases équationnelles porte uniquement sur le code lexical du français : leur fonction est strictement métalinguistique.

La visée du message en tant que tel, l'accent mis sur le message pour son propre compte, est ce qui caractérise **la fonction poétique** du langage. La fonction poétique n'est pas la seule fonction de l'art du langage, elle en est seulement la fonction dominante, déterminante, cependant que dans les autres activités verbales elle ne joue qu'un rôle subsidiaire, accessoire. Cette fonction, qui met en évidence le côté palpable des signes, approfondit par là même la dichotomie fondamentale des signes et des objets. Aussi, traitant de la fonction poétique, la linguistique ne peut se limiter au domaine de la poésie.

4.2 La communication non-linguistique

Les échanges non verbaux sont nombreux et si divers qu'on peut les considérer comme un moyen de communication autonome. Ce langage silencieux présente certains traits du langage verbal et notamment celui d'être un code, en partie arbitraire, différent d'une culture à l'autre.

4.2.1 Culture et communication

On considère comme une donnée évidente le fait que la langue est le principal système de communication entre les hommes. En réalité, la langue est toujours accompagnée d'autres systèmes ou moyens de communication. Et toute culture est un réseau complexe de systèmes signifiants permettant divers types de communication dont la langue n'est qu'une composante. En effet, la culture est la somme des répertoires des comportements codés, accomplis et interprétés par les membres de l'organisation sociale dans des situations communicatives, ... bref, culture = communication.

Ces modèles de comportement, très structurés, sont organisés en séquences articulées de façon différente selon les cultures. Prenons l'exemple des gestes, dont l'étude constitue la *kinésique*.

L'étude des gestes français contemporains tels qu'ils apparaissent dans un corpus bien délimité de pièces de théâtre, entre 1900 et 1945, ainsi que de leurs rapports avec la parole, aboutit à la distinction : gestes accompagnateurs, ou complémentaires, ou substitutifs de la parole.

La comparaison des aspects spatio-temporels des gestes des Juifs et des Italiens vivant à New York montre que les Juifs orientaux traditionnels, vivant en guetto, ont une gesticulation particulière, de même que les Italiens du Sud non assimilés. Les Juifs et les Italiens assimilés ont perdu les caractéristiques gestuelles de leurs communautés respectives et se rapprochent des Américains.

Le comportement varie donc avec l'environnement social, indépendamment de toute variation raciale. Un même signifié (« Salut ! ») peut avoir d'innombrables signifiants gestuels. Les mêmes signifiants (/hocher la tête/, /se toucher le front/) peuvent avoir des signifiés différents selon les cultures.

4.2.2 La communication, système de codes interdépendants

La communication est un système de codes interdépendants transmissibles à travers des canaux influençables à base sensorielle. La conversation, par exemple, est le système total de communication employé par les participants dans une action intersubjective. Trois types d'information sont échangés entre les participants :

- (a) Information cognitive : contenu des signes linguistiques échangés ;
- (b) Information indicielle, qui porte sur le locuteur dans le but de définir et de contrôler le rôle qu'il joue dans la conversation ;

(c) Information injonctive échangée par les participants pour faire progresser la conversation, varier les locuteurs, aboutir à un résultat.

Les comportements, les moyens utilisés pour communiquer cette information, peuvent être classés selon deux oppositions : *vocal vs. non vocal* (parole vs. gestes, attitudes) et *verbal vs. non verbal* (mots vs. non mots).

1. Vocal-verbal	Le mot phonétique comme unité linguistique
2. Vocal-non verbal	Intonation, qualité de la voix, emphase
3. Non vocal-verbal	Le mot graphique comme unité linguistique
4. Non vocal-non verbal	Expression du visage, geste, attitude ...

Ces moyens peuvent aussi être classés de la manière suivante :

1. Moyens linguistiques :	Langue doublement articulée et de manifestation vocale
2. Moyens paralinguistiques	Non verbal, vocal (ton de la voix, intonation, ...), non vocal (gestes...); plus ou moins conscients, compris par les membres de même culture.
3. Moyens extralinguistiques :	Échappent au contrôle du locuteur pendant la conversation; vocal (qualité de la voix apportant à l'auditeur des renseignements des caractéristiques biologiques, psychologiques ou sociaux sur le locuteur), non vocal (façon de se vêtir ...).

Aussi, est-il possible de relier le type d'information et le moyen utilisé :

1. Information cognitive	Moyens linguistiques et paralinguistiques (gestes substitutifs de la parole).
2. Information injonctive	Moyens paralinguistiques surtout (gestes, mouvements, intonation pour suggérer à un participant de prendre la parole) et moyens linguistiques parfois (phrases rituelles de bienvenue ...)
3. Information indiciaire :	Transmise par tous les moyens. Dans la parole, les détails biologiques sont fournis par les moyens extralinguistiques; l'information psychologique et sociale est inférée des moyens linguistiques et paralinguistiques. La convergence des trois permet à l'auditeur de cerner les caractéristiques du locuteur

La possibilité permanente de varier les types et les moyens d'information, la nécessité pour les participants de connaître les mêmes codes, expliquent que la communication entre membres de cultures différentes est généralement imparfaite.

Pour rendre compte du phénomène communicatif dans son ensemble, plusieurs aspects doivent être tenus en considération. En effet, à l'intérieur d'une symbolique généralisée s'ordonnent : a) des faits pré-linguistiques (événements physico-biologiques) ; b) la communication qui étudie le langage et les phénomènes qui lui sont strictement associés : phénomènes paralinguistiques (vocalisations et qualité de la voix) et kinésiques (mouvements et gestes). La proxémique, une branche qui étudie la structuration signifiante de l'espace humain, est un exemple de ces recherches. Le comportement humain dans l'espace fournit des informations. Les inconnus maintiennent entre eux une certaine distance ; si les places dans un lieu donné sont en nombre suffisant, les sièges qu'ils occupent sont séparés ou même éloignés. Une famille, au contraire, a tendance à se regrouper, les enfants sur les genoux des adultes.

4.2.3 Les procédés de communication non linguistique

George Mounin distingue 4 types de procédés de communication non linguistique :

1. Les procédés de signalisation **substitutifs** du langage parlé, qui sont sans autonomie réelle à l'égard de la communication linguistique : les écritures, les enseignes-idéogrammes, les alphabets, etc. ;
2. Les procédés de communication **systématiques** : emploi des chiffres, des symboles mathématiques ; la signalisation routière, etc.
3. Les procédés de communication **systématiques se lisant sur la trame de l'espace** : cartes, plans, diagrammes, schémas de toutes sortes, etc. ;
4. Les procédés de communication **asystématiques se lisant sur la trame de l'espace** : illustrations, images, publicité, etc.

4.3 Complexité de la communication humaine et limites du Schéma de communication proposé par Roman Jakobson

Il importe de remarquer que les fonctions jakobsoniennes du langage n'épuisent pas leur objet.

- (1) Le schéma de 6 fonctions qu'il propose est à la fois trop général pour permettre une taxinomie et une syntaxe appropriées, et, en même temps trop particulier du fait qu'il ne porte que sur la seule communication verbale à l'exclusion de tous les autres systèmes de communication.
- (2) Le schéma de communication proposé par Jakobson ne concerne que le faire informatif, articulable, selon le rapport destinateur/destinataire, en faire émissif/faire réceptif. Or, il existe d'autres manières de concevoir la transmission du savoir : tel est le cas du faire persuasif et du faire interprétatif qui relèvent, plus que de la communication, de la manipulation. Si le langage est communication, il est aussi production du sens, de signification. Il ne se réduit pas à la simple transmission d'un savoir sur l'axe « je »/ « tu » ;

complémentaire, il se développe pour lui-même, pour ce qu'il est, avec une organisation interne propre dont la seule théorie de communication, qui prend en quelque sorte le point de vue externe, ne semble pas pouvoir rendre compte.

- (3) Les activités humaines, dans leur ensemble, sont généralement considérées comme se déroulant sur deux axes principaux : (a) l'axe de production qui est celui de l'action sur les choses, par laquelle l'homme transforme la nature et (b) l'axe de la communication qui est celui de l'action sur les autres hommes, créatrice des relations intersubjectives, fondatrices de la société. En effet, de nos jours, l'homme est devenu un être « communicationnel » par excellence, pour lequel communiquer signifie exister. Il ne cesse de tisser autour de lui d'innombrables fils qui le rattachent à l'univers environnant. En général, on ne peut ne pas communiquer, car communiquer c'est vivre, et la situation de non communication est en soi une communication. Toute communication suppose une rétroaction, ce qui entraîne un dialogue. Le concept d'échange qui, dans la tradition anthropologique française recouvre cette seconde sphère d'activités peut donc être interprétée de deux façons différentes : soit comme le transfert d'objets de valeur, soit comme la communication entre sujets.
- (4) Le langage est le moyen par excellence de communication entre les sujets. Il tient sans aucun doute une place importante dans la société, nos institutions, notre vie. Mais le langage verbal coexiste avec d'autres modes de communication : la mimique et les langages gestuels. Aussi l'homme communique à ses semblables par une grande variété de langages. Il en résulte que le modèle de communication humaine est plus complexe. Elle suppose l'émission et la réception de signes (verbaux ou non-verbaux) a fait l'objet d'une convention préalable entre l'émetteur et le récepteur.

ANNEXES

5 DES « MOTS » AUX UNITÉS MORPHOLOGIQUES ET LEXICALES

5.1 Problèmes posés par la notion de mot

Peut-on considérer le mot, notion évidente pour les usagers de la langue, comme l'unité linguistique minimale dans le domaine du lexique ? Notre perception du mot est essentiellement graphique. C'est une unité pratique et de sens commun, issue de la pratique de la lecture et de l'écriture. On peut en donner une définition fondée sur la tradition orthographique : un mot est un segment de discours compris entre deux espaces blancs, éventuellement entre une apostrophe et un espace blanc (Apothéloz 2002 : 7)

Saussure (1972 : 54) voyait dans le mot "malgré la difficulté qu'on a à le définir, [...] une unité qui s'impose à l'esprit, quelque chose de central dans le mécanisme de la langue.

Toutefois, le découpage des mots est moins évident qu'il n'y paraît à première vue.

- chez l'enfant, le découpage des mots ne va pas de soi...
- **les formes *marcher, marchons, marchèrent* correspondent-elles à des mots différents ?**
- spontanément, on dira que l'expression *laps de temps* est formée de trois mots... mais est-ce que *laps* est bien un mot ?
- selon la définition donnée ci-dessus, *coffre-fort* est un mot, alors que la forme *château fort* est composée de deux mots. Pourtant on pourrait montrer que ces deux formes fonctionnent de la même manière. Deux formes qui suivent des **principes de fonctionnement identiques ne doivent-elles pas être considérées comme appartenant au même type d'unité ?**
- L'autre mot pour désigner la *patate* est *pomme de terre*. Il apparaît clairement que *patate* et *pomme de terre* sont interchangeable. Doit-on alors considérer que *pomme de terre* ne correspond qu'à un seul mot, même si notre conscience graphique nous invite à identifier trois mots ?

Siouffi & Van Raemdonck (1999) distingue le mot graphique du mot phonétique ; le « mot phonétique » renvoyant à une suite de sons entre deux pauses. Mais cette définition pose également problème, car les mots phonétiques ne correspondent pas nécessairement aux mots graphiques. En outre, les pauses ne correspondent pas toujours aux blancs graphiques. Lorsque des mots commencent par des voyelles, ils peuvent être unis au mot précédent, entre autres, par des phénomènes de liaison (les-enfants) ou d'élision (Eh, l'-ami, t'exagères !).

5.2 Des critères formels pour définir le mot

On tente donc de définir le mot selon des critères dits *distributionnels* : **stabilité interne, mobilité de position, non-séparabilité des constituants**. Soit la phrase « Le chat miaulait tristement ». Comment peut-on découper cette phrase pour identifier les unités qui la constituent ?

- (1) Le-chat-triste-ment-miau-lait
 Triste-ment-le-chat-miau-lait
 *Ment-triste-miau-lait-le-chat

Certains groupes peuvent être décomposés et leur ordre peut être inversé (tristement miaulait / miaulait tristement), alors que d'autres peuvent être décomposés (triste-ment) mais sans que

l'ordre de leurs composants puisse être modifié : ces unités, les mots, se caractérisent par leur **stabilité interne** (non-modifiabilité de l'ordre des composants du mot) et par leur **mobilité de position** (un mot peut être déplacé à l'intérieur de la phrase). Cependant, selon ces deux critères, la suite *le chat* constitue un mot.

(2) Le chat miaulait tristement

Le chat tristement miaulait

Tristement le chat miaulait

Tristement miaulait le chat

*Le tristement chat miaulait

Il faut donc un critère supplémentaire : la **non-séparabilité** des constituants c'est-à-dire un mot est une unité à l'intérieure de laquelle il n'est pas possible d'insérer un autre mot (même sans modifier l'ordre des constituants) : *le chat* → *le petit chat* → *le joli petit chat* ...

Remarque : le mot *le* n'est pas aussi pleinement un mot que par ex. *tristement* (il ne répond qu'à un des trois critères). C'est le cas des morphèmes clitiques que la tradition orthographique sépare comme des mots. Du point de vue graphique, le mot est une unité à géométrie variable. Du point de vue de la cohésion interne, certaines unités graphiques sont perçues comme étant "moins que des mots".

Selon les critères établis ci-dessus, *pomme de terre* est un mot au même titre que *patate*. Pour éviter les ambiguïtés liées à la notion de mot, les linguistes diront que ces deux formes constituent chacune une **lexie**. Les formes *en avoir par-dessus la tête* ou *faire de son nez* sont des lexies complexes. La discipline de la linguistique qui étudie les lexies est la **lexicologie**. Pourtant, on perçoit bien que contrairement à *patate*, la lexie *pomme de terre* peut être décomposée en plusieurs unités. Quelles sont donc les unités à partir desquelles les lexies sont formées ?

5.3 Le morphème

La notion de morphème désigne les unités de première articulation, plus petites unités dotées de sens

5.3.1 Les catégories de morphèmes

Les morphèmes peuvent être de deux types :

- les **morphèmes lexicaux ou lexèmes** constituent des unités lexicales et sémantiques qui possèdent un sens « plein », et ont pour fonction de désigner une réalité du monde (un objet, une action, une idée, etc.). Les lexèmes forment les unités de base du lexique d'une langue et constituent une classe ouverte et en évolution constante ;
- les **morphèmes grammaticaux** constituent des unités qui servent à déterminer, à préciser, le sens d'un morphème lexical ; ils ne sont jamais autonomes (ils fonctionnent toujours en association avec un lexème) et forment, pour chaque langue, une classe fermée et stable.

Au sein de la classe des morphèmes grammaticaux, on distingue notamment les **affixes** (morphèmes liés, agglutinés à des lexèmes) :

- selon leur fonction : les affixes flexionnels ou dérivationnels ; les affixes flexionnels (presque exclusivement des suffixes en français) servent à réaliser la flexion, c'est-à-dire la variation des formes des lexies qui permet de leur conférer des marques (nombre, temps) et qui correspond à, la *déclinaison* pour les noms et adjectifs, et à la *conjugaison*

pour les verbes). Les affixes dérivationnels permettent de réaliser la dérivation (procédé qui consiste à ajouter des morphèmes à une lexie pour en créer de nouvelles ;

- selon la position qu'ils occupent : les préfixes, infixes et suffixes (à gauche, au centre ou à droite du lexème de base).

(3) Préfixe : **in-**connu

(4) Suffixe : maquill-**age** (dérivationnel) ; travail-**ons** (flexionnel)

(5) Infixe : tagalog (Philippines) : *bumbili* 'acheté' = *bili* 'acheter' + *um* inséré après la première consonne de la base (issu de <http://www.sil.org>)

Le **radical** d'une lexie est le morphème lexical qui subsiste lorsqu'on a enlevé tous les affixes dérivationnels et flexionnels de cette lexie :

(6) *-fai-* est le radical de *défaire* ;

(7) *nomade* est le radical de *nomadisme*

Le radical ne se confond pas avec le **lemme**, qui est la forme considérée conventionnellement comme non marquée d'une lexie (celle que l'on trouve comme entrée de la lexie dans les dictionnaires, l'infinitif pour les verbes, le masculin singulier pour les noms et les adjectifs). Le lemme de *défais*, *dé faisons* est *défaire*, le lemme de *nomades* est *nomade*.

5.3.2 Morphème, morphe, allomorphe

Le morphème est, comme toutes les unités de la langue, une unité abstraite, qui n'existe que dans la compétence des locuteurs : dans *nous aimons*, il y a pas moins de quatre marques grammaticales (2^e personne + pluriel + indicatif + présent) qui sont amalgamées dans un morphème *-ons*, c'est-à-dire qui ne sont pas exprimées explicitement et successivement dans la matérialité du langage. Ce morphème est néanmoins *marqué, actualisé, réalisé*, par la forme graphique concrète *-ons* à l'écrit, et par le son [ɔ̃] à l'oral. La réalisation concrète du morphème est le **morphe**. Un même morphème peut avoir différentes réalisations, que l'on appelle des **allomorphes**.

le verbe *aller* (→ *ir-* dans « j'irai » ; → *va-* dans « tu vas » ; → *all-* dans « nous allons »)
 les deux morphes « fleur » et « flor » dans les mots fleuriste / floral
 les deux morphes « labeur » et « labor » dans les mots labeur / laborieux

Un morphe est dit **libre** s'il peut exister de manière autonome (ex. « choix », « voiture »). Un morphe est dit **lié** s'il n'a pas d'existence autonome mais qu'il ne se retrouve que combiné à d'autres morphèmes dans des lexies construites (ex. « ez » dans *aimez* , « douleur- » et « -eux » dans *douloureux* , « in- » et « -dolare » dans *indolore* , etc.).

Les réalisations écrites des morphèmes grammaticaux sont appelées **morphogrammes** (le 's' graphique est le principal morphogramme du pluriel en français, à côté de 'x', par ex. *des* chevaux, *les* enfants, etc.).

Les linguistes distinguent trois types de marquage des morphèmes : le morphème peut être réalisée de façon **discontinue** (présent sur plusieurs éléments, ex. *il* ne *vient* pas, *il* a *mangé*), **amalgamée** (non distinguable ; ex. *au* = « à + le », *du* = « de + le ») et il peut être non réalisé, on parle alors de **morphème (morphe) zéro** (ex. la valeur grammaticale « singulier » dans *enfant* est réalisée par une absence de marque). La discipline qui étudie les morphèmes et les règles de leur combinaison est la **morphologie**.

5.4 La formation des unités lexicales

Il existe deux procédés principaux pour former de nouvelles unités lexicales à partir d'unités existantes, grâce à notre savoir morphologique spontané : la dérivation et la composition. À ces procédés principaux s'ajoutent des procédés secondaires.

5.4.1 Formation par dérivation

Alors que la flexion associe un morphème grammatical (affixe flexionnel) au radical d'une lexie sans créer une lexie nouvelle (*aimer* et *aimons* correspondent à la même lexie), la dérivation permet d'enrichir le lexique de la langue par l'association d'une lexie existante, qu'on appellera alors **base**, avec des affixes dérivationnels. La base est l'élément sur lequel opère un affixe dérivationnel.

Prenons l'exemple du mot *renationalisation*. Il est dérivé de la base *nationalisation*, par l'adjonction du préfixe *re-*, base qui est elle-même dérivée par l'ajout du suffixe *-ation* à la base verbale *nationaliser*, elle-même issue de l'association entre la base *national* et le suffixe *-iser* ; *national* étant lui-même le dérivé de *nation*. *nationalisation* : nationalis-ation ; national-is-er ; nation-al => radical/racine = *nation* ; base = *nationalis-*, *national-*, *nation-*... suivant le dérivé

Cet exemple montre que la récursivité est une propriété des mécanismes de dérivation; on parlera dans un cas comme celui de *renationalisation* de dérivation multiple. Les préfixes et les suffixes modifient le sens de la base à partir de la valeur sémantique (signifié) qui leur est associée.

5.4.1.1 Suffixation

La suffixation consiste à ajouter un suffixe dérivationnel à une base (exemples de suffixes : -age dans *partage*, *chantage*, -able dans *faisable*, -aille dans *trouvaille*, *bleusaille*) (cf. le *Bon usage*, §160 sv.).

(8) *événement* (nom) + *-iel* (suffixe) > *événementiel* (adjectif)

(9) *organis-* (verbe) + *-ation* (suffixe) > *organisation* (nom)

La suffixation entraîne parfois une modification du radical (le lexème se réalise alors sous la forme d'un allomorphe) :

(10) *contrat* (nom) + *-uel* (suffixe) > *contractuel* (adjectif) [*contract-* sur le modèle latin *contractum*]

(11) *air* (nom) + *-ien* (suffixe) > *aérien* (adjectif) [racine latine *aer-*]

Dans le cas des **suffixes**, la nouvelle unité formée par dérivation modifie fréquemment la catégorie de la base. Il s'agit d'une formation **exocentrique**. On parle de formation **endocentrique** quand le dérivé appartient à la même catégorie grammaticale que la base.

(12) *chante-* (verbe) + *-age* > *chantage* (nom)

(13) *charme* (nom) + *-ant* > *charmant* (adjectif)

(14) *sel* (nom) + *-in* > *salin* (adjectif) ; *sel* (nom) + *-age* > *salage* (nom)

Le suffixe qui sera ajouté à une base donnée pour créer un mot dérivé dépend principalement de deux critères : **la catégorie lexicale du dérivé** (pour produire un adjectif, on ne peut utiliser que certains types de suffixes), **la catégorie lexicale de la base** (les suffixes s'ajoutent tantôt à des noms, des verbes, etc.), **la valeur sémantique ajoutée par le suffixe** (voir p. suivante). Un quatrième critère peut cependant intervenir : il arrive en effet qu'un suffixe sélectionne des

unités lexicales dont le sens doit présenter une certaine propriété, ce qui fait que toutes les combinaisons théoriquement envisageables ne sont pas réalisées (par ex. **mange-ation* alors qu'on a *damn-ation*).

Lexies dérivées	Suffixes	Valeur
1. lav-age, truqu-age, tiss-age... 2. gaspill-age, papot-age, radot-age...	-age	1. Suff. formateur de subst. d'action 2. S'accole à des verbes intrans. plus ou moins péj.
identifi-able, mange-able, mani-able, néglige-able; ris-ible, fermentesc-ible, faill-ible	-able, -ible	Sert à former des adj. à partir de verbes, exprime la possibilité
cri-ailler, rim-ailler (faire de mauvais vers), tir-ailler, train-ailler (var. de train-asser)...	-ailler	Suff. formateur de verbes fréquentatifs et gén. péj., à partir de bases essentiellement verbales (plus rarement subst.).
richard, binoclard, pantouflard...	-ard	Suff. péj. formateur d'adj. ou de subst. qualifiant ou désignant des personnes
dévidoir, miroir, tranchoir, balançoire, abreuvoir, bougeoir	-oir(e)	Servent à constr. des subst. désignant des instruments ou des lieux.
Alpin, andin; Bernardin, ursuline; féminin, masculin; etc.	-in(e)	Suff. formateur d'adj. de relation
antimonite, arsenite, nitrite ... appendicite, bronchite, bibliophilite...	-ite	Suff. formateur de subst. appartenant au vocab. des sc. nat. et de la médecine (plusieurs signifiés : sel d'acide; affection de nature inflammatoire)

5.4.1.2 Préfixation

La préfixation consiste à ajouter un préfixe dérivationnel à une base. Dans le cas de l'ajout de préfixes, la nouvelle unité formée par dérivation prend généralement la même catégorie que la base (formation **endocentrique**), sauf rares exceptions.

- (15) fermer > enfermer
 (16) content > mécontent
 (17) char > antichar (nom → adj)

Tableau (1) Exemples de dérivation préfixale

Exemples	Préfixe	Valeur
cokoter, coexister, concitoyen	co(n)-	simultanéité
décharger, dénatalité, désagréable, désordre	dé-	différence, séparation
inégal, incassable	in-	négation
mécontent, mésalliance, mésestimé	mé(s)-	négation, valeur péjorative
prétraite	pré-	antériorité
rechanter, ressortir	re-	répétition, retour à un état ancien, renforcement

Comme les suffixes, les préfixes ne s'associent pas à n'importe quelle catégorie de mot. Le préfixe *re-* ne s'applique qu'aux verbes, le préfixe *in/im-* ne s'applique qu'aux adjectifs. C'est ce qui permet de déterminer l'ordre de la dérivation dans les cas suivants :

(18) mobile - immobile - immobiliser - immobilisation
*mobile - mobiliser - *immobiliser

(19) évaluer - réévaluer - réévaluation
évaluer - évaluation - *réévaluation

5.4.1.3 Remarques

1) Il est important de ne pas confondre l'opération de dérivation et l'analyse morphologique des dérivés (qui se base sur les relations abstraites entre les bases et leurs dérivés), telles qu'elles peuvent être réalisées en synchronie, et le processus historique de dérivation (la façon dont le dérivé a été effectivement introduit dans la langue). C'est cette capacité d'analyse morphologique, mobilisable face à chaque mot mais non mise en œuvre dans l'usage spontané de la langue, qui explique notre capacité à produire et à comprendre de nouveaux dérivés.

2) La productivité ou la vitalité d'un affixe est sa propriété d'être encore exploité aujourd'hui pour former de nouveaux dérivés. Le plus souvent, cette vitalité dépend de la capacité des locuteurs à associer un signifié aux affixes, ce qui suppose qu'ils perçoivent la motivation des dérivés (qu'ils soient capables de les analyser).

3) La dérivation peut être réalisée sans l'intervention d'un affixe (sans marque de trace morphologique visible). Elle est alors (a) soit purement sémantique (on étend ou on modifie l'acception d'un mot, ex. *canard* ou *café*), soit (b) grammaticale, quand une unité change de catégorie, mais sans affixe (infinitif substantivé, nominalisation de l'adjectif, etc.). On parle alors de **conversion** (anciennement : dérivation impropre).

(20) parler → le parler

(21) rouge → le rouge

(22) bleu → un bleu

(23) contre → le contre

(24) opposer, opposant → un opposant

4) Le mode de représentation le plus fréquent de la base et des affixes consiste à utiliser des parenthèses étiquetées. Les parenthèses progressent de la base vers la forme construite, du noyau vers les ajouts. Elles indiquent l'ordre d'application des processus de formation. Il faut en fermer autant qu'on ouvre et veiller à mentionner la catégorie grammaticale des formes.

(25) inavouable = (avou-)_V + (-able)_{Suf} + (in-)_{Préf}
= ((avou-)_V + (-able)_{Suf})_{Adj} + (in-)_{Préf}
= ((in-)_{Préf} ((avou-)_V + (-able)_{Suf})_{Adj})_{Adj}

5) Il existe des affixes homonymes. Comparez par exemple :

(26) pommier, poirier, amandier VS aventurier, boursier, caissier
ainsi que

(27) lentement, rapidement, solidement VS lancement, fonctionnement, déploiement
Une même forme peut correspondre à plusieurs types de morphème différents :

(28) Il est contre (préposition) cette mesure / Il a subi le contrecoup (préfixe) / Il faut peser le pour et le contre (nom)

(29) Il est mal (adverbe) dans sa peau / Il est malpoli (préfixe)

6) « En français, comme dans la plupart des langues, les affixes dérivationnels opèrent toujours avant les affixes flexionnels. Et, parmi ces derniers, les flexifs de genre opèrent avant ceux de nombre, et ceux de temps avant ceux de personne. » (Apothéloz 2002 : 14).

(30) (((morphème lexical) suffixe dérivationnel) flexif de genre) flexif de nombre →
talentueuses

(31) (((morphème lexical) suffixe dérivationnel) flexif de temps) flexif de personne
→ banaliserons

5.4.2 Formation par composition

La composition est un procédé qui associe plusieurs lexèmes en une nouvelle unité complexe, qui fonctionne comme une unité simple. Celle-ci sera souvent appelée **locution** ou lexie complexe.

5.4.2.1 Les principes de la composition

En principe, la composition associe plusieurs morphèmes autonomes (vs affixes et lexèmes sous la forme d'un morphe lié) existant dans la langue, dont au moins un est une lexie (un lexème autonome). Les morphèmes associés constituent autant de *bases* de la composition, le cas échéant sous une forme fléchie particulière (ex. *ramasse-poussière*, *garde-manger*). Cette association de morphèmes crée une nouvelle lexie complexe, appelée aussi locution (mot composé dans la terminologie courante) : *en* (morphème grammatical libre) + *fait* (lexème libre) → *en effet* ; *monte* (lexème libre) + *charge* (lexème libre) → *monte-charge*, etc. (*entre temps*, *va-t-en-guerre* sont aussi des locutions). On distingue un composé graphiquement disjoint, graphiquement conjoint et graphiquement soudé.

(32) pomme de terre, pause café (ou pause-café)

(33) lave-vaisselle, pot-de-vin

(34) autoroute, contrordre

Il est important de bien distinguer composé et syntagme (suites libre de mots). **Le composé est une association figée, dont les éléments ont perdu leur autonomie** (ils ne sont plus analysables ; ex. *pomme de terre*, *petit-fils*). **Le syntagme est une association temporaire d'unités autonomes, même s'il peut être très fréquent** : ex. *petit garçon*, *développement durable*, qui constitue une collocation (cooccurrence privilégiée, mais non figée, de deux termes).

Remarque. Il y a des degrés de figements, et donc une gradation entre les syntagmes et les composés. En français, la plupart des composés sont des noms. Les bases qui les composent sont de catégories diverses ; et tous les composés ne sont pas des noms :

(35) plage horaire, centre commercial, libre-échange, lave-vaisselle, ceinture de sécurité, je ne sais quoi

(36) jusqu'au bout de (locution prépositionnelle)

(37) d'autant plus que (locution conjonctive)

(38) en effet (locution adverbiale)

Dans certains cas, les composés sont des syntagmes qui se sont figés avec le temps ; on parle alors de **lexicalisation** (terme qui désigne de manière générale le processus par lequel une suite de morphème, un syntagme devient une unité lexicale autonome). Une fois lexicalisée, ces formes peuvent faire l'objet d'une dérivation : *un jusqu'au-boutiste* (voir aussi le cas particulier

de *je-m'en-foutiste*, lexicalisé par la dérivation).

5.4.2.2 La composition savante

Dans certains cas, la composition résulte de la combinaison de bases savantes, c'est-à-dire de lexies empruntées à d'autres langues mais qui n'existent pas de façon autonome dans le système de la langue emprunteuse. On parle alors de composition savante. Dans la plupart des cas, les formations savantes sont étymologiquement homogènes (*calorifuge* est issu de deux racines latines, *thalassothérapie*³⁵ est grec, mais *sociologie* est hybride).

Bases latines :

- (39) -culteur [lat. cultor, cultura] : agriculteur, apiculteur.
- (40) -fuge [lat. fuga, de fugare 'mettre en fuite'] : vermifuge, centrifuge.

Bases grecques :

- (41) bibli- (livre) : bibliophile
- (42) log- (discours) : logopédie
- (43) péd- (enfant) : pédiatrie

Autres langues :

- (44) italien : parasol, paravent.
- (45) anglais : self- (self-service), -man : tennisman, wattman (en Belgique, faux anglicisme).

A première vue, on pourrait penser que certains de ces composés savants fonctionnent comme des dérivés :

- (46) biologie, biocarburant, bio-cosmétique
- (47) écologie, éco-responsable, écotaxe

Dans ces exemples, *eco-* et *bio-* constituent-ils des préfixes ? Même si elles ne constituent pas des lexies autonomes en français, les bases savantes fonctionnent comme des morphèmes lexicaux et non comme des morphèmes grammaticaux, à l'instar des affixes : elles correspondent à des noms (-logie), des verbes (-phage, -phone) et leur signifié renvoie effectivement à un référent particulier, alors que la valeur sémantique des affixes renvoie à des notions atemporelles et universelles.

Il existe des cas qui peuvent créer des confusions : dans *télévision* (étymologiquement « transmission à distance d'images » *télévision* → *télé*) et *automobile* (étymologiquement « qui se meut de soi-même » : *voiture automobile* → *automobile* → *auto*), *télé-* et *auto-* sont à l'origine des préfixes. Aujourd'hui, ils sont devenus des lexies en français et entrent dans la formation de lexies composées. Il y a donc deux morphèmes *télé* et deux morphèmes *auto*, comme l'attestent les exemples suivants :

- (48) téléportation, téléchargement, télésurveillance VS téléachat, télé-jeu, téléreportage
- (49) autographe VS autoroute, autostoppeur

5.4.3 Autres procédés de renouvellement du lexique

5.4.3.1 Mots valises

Le terme « mots-valises » est un calque de l'anglais « portmanteau word ». Il désigne **la réunion,**

³⁵ On remarquera que *thalasso* et *thérapie* sont utilisables comme des morphes libres en français.

à l'oral et à l'écrit, de deux mots, dont on garde le début (pour le premier) et la fin (pour le deuxième) pour créer une lexie combinant le sens de ces deux mots (cf. *Bon usage*, § 177d).

- (50) *français* = français + anglais (à partir duquel on a créé *francolof* 'français + wolof', *camfrançais* 'camerounais + français + anglais)
- (51) *maroxellois* = marocain + bruxellois
- (52) *foultitude* = foule + multitude [apparaît en français en 1848 (Petit Robert)].
- (53) *motel*, emprunt à l'anglais américain, formé à partir de motor[car] 'automobile' + hôtel ; apparaît en français en 1946 (Petit Robert).

5.4.3.2 Réduction (troncation)

La réduction ou la troncation est la suppression, à l'oral et à l'écrit, d'une partie (syllabe) d'un mot ou d'une locution.

- (54) automobile > auto
- (55) métropolitain > métro
- (56) star academy > starac
- (57) un bateau à trois mâts > un trois-mâts
- (58) encaisser des coups > encaisser
- (59) à plus tard > à plus (abrégé en à +)

L'apocope est une réduction à la fin du mot (*ciné* pour *cinéma*, *télé* pour *télévision*), l'aphérèse est une réduction au début du mot (*bus* pour *autobus*, *car* pour *autocar*).

Une question se pose néanmoins : peut-on réellement inclure ces mécanismes de réduction dans les procédés de création lexicale ? Autrement dit, dire *ciné* ou *auto* plutôt que *cinéma* ou *automobile* est-ce vraiment produire une lexie différente, ou s'agit-il simplement d'une variante de prononciation (à l'instar par exemple de la variation de la prononciation de *mettre* [metʁ(ə)] ou [mɛt]) ? La réponse à cette question n'est pas simple : dans certains cas, la troncation a donné naissance à une nouvelle lexie qui n'existe que sous la forme tronquée et qui a un sens spécifique par rapport au mot d'origine : le terme *métro*³⁶, qui désigne le moyen de transport en commun, est bien un mot différent de *métropolitain*. Dans ce cas, la forme tronquée a connu un processus de lexicalisation achevé. Le cas de *ciné* est moins net : d'un côté, les formes tronquées et d'origine peuvent alterner dans le même contexte (ex. *je vais au ciné//cinéma*) – ce qui laisse penser qu'il s'agit simplement de deux variantes du même mot –, mais d'un autre côté, le sens de *ciné* est plus restreint que celui de *cinéma*, il désigne davantage le lieu où le peut voir un film que le cinéma en tant qu'activité artistique (on dira moins souvent *l'industrie du ciné* ou *il travaille dans le ciné* que *l'industrie du cinéma* ou *il travaille dans le cinéma*, comme le montrent par exemple les occurrences de ces syntagmes dans google) – ce qui suggère qu'on a affaire à deux lexies différentes qui ne sont pas de parfaits synonymes.

En outre, il est important de ne pas confondre la troncation avec l'abréviation, procédé exclusivement **graphique** qui consiste à écrire un mot en n'utilisant qu'une partie de ses lettres.

- (60) Monsieur > M.
- (61) numéro > n°
- (62) manuscrit > ms

5.4.3.3 Sigles et acronymes

Les sigles et les acronymes sont des mots formés sur base d'abréviations constituées par des

³⁶ De « chemin de fer métropolitain (= d'une métropole) ».

initiales, mais qui ont été traitées ensuite comme des lexies, de deux manières.

Dans le cas du sigle, on épelle les lettres :

- (63) A.S.B.L. = association sans but lucratif
- (64) P.D.G. = président-directeur-général
- (65) U.S.A. = United States of America

Dans le cas de l'acronyme, on enchaîne les lettres dans la prononciation :

- (66) OTAN= Organisation du Traité de l'Atlantique Nord.
- (67) SIDA = Syndrome d'Immuno-Déficience Acquise

Cette distinction entre acronyme et sigle tend, aujourd'hui, à s'estomper dans les usages graphiques. Les graphies courantes comme *asbl*, *onu*, *sida* ou *sms*, témoignent du fait que ces formes sont perçues comme des lexies à part entière. Dans le même sens, on notera qu'elles peuvent donner lieu à des dérivations : *onusien*, *sidéen*, mais aussi *smicard*, *énarque*, etc.

5.4.3.4 La reduplication

La reduplication est le procédé général qui consiste à répéter une même forme (tiens tiens, mauvais mauvais) ; la reduplication morphologique permet de créer de nouvelles unités lexicales en répétant la première syllabe d'une lexie existante : foufou, bêtebête, nunul, guéguerre, pépère, etc.

5.4.3.5 L'emprunt

On appelle emprunt un élément qu'une langue emprunte à une autre, au cours de son histoire (passée ou récente, même actuelle). Les emprunts à date ancienne ne sont souvent plus perçus comme tels (les locuteurs n'ont plus conscience qu'ils sont originaires d'une autre langue).

- (68) *camarade* < espagnol *camarada* (terme militaire = chambrée), lequel vient de *camara* 'chambre', du latin *camera*.
- (69) *caméra* (créé vers 1900) est un emprunt direct (savant) au latin *camera*. [preuve de l'emprunt : traitement de k + a, qui aurait dû aboutir à la chuintante : cf. *caballu* > cheval).
- (70) *caricature* (< italien *caricatura*)
- (71) *carnaval* (< italien *carnelevale* 'mardi gras' < *carnelevare* : *levare* + *carne* 'lever l'abstinence de viande')
- (72) *cauchemar* (< picard *quauquemaire*, composé de *cauchier* 'fouler, presser' + néerl. *mare* 'fantôme' [cf. *nightmare*], etc.
- (73) *caravansérail* (< persan *karwanserai* 'logement de caravanes')
- (74) *casaque* (< turc *kazak* 'aventurier' (cavalier du bord de la Mer Noire).

Un cas particulier concerne les doublets : alors qu'une unité lexicale a suivi l'évolution phonétique régulière depuis l'étymon (latin, germanique...) et ne constitue donc pas un emprunt, l'autre unité de la paire, dit savante, est empruntée plus tardivement sous une forme proche de l'étymon.

- (75) lat. *aquarium* > *aquarium* / évier
- (76) lat. *captivum* > *captif* / chétif

(77) lat. *potionem* > potion / poison

(78) lat. *pastor / pastorem* > pasteur / pâtre

On distingue plusieurs **types** d'emprunts (typologie appliquée ici aux anglicismes) : l'emprunt lexical ; l'emprunt sémantique ; le calque, correspondant respectivement aux trois exemples suivants :

(79) *happy few*, *hit-parade* [recommandation officielle de 1985 : palmarès], *forwarder*...

(80) domestique, évidence

(81) « to fall in love » en anglais → « tomber en amour » en français québécois.

Les emprunts deviennent parfois productifs et sont à l'origine de nouveaux mots, conformes aux critères morphosyntaxiques de la langue emprunteuse. On parle également dans ce cas de processus de lexicalisation (le mot emprunté devient une lexie à part entière).

(82) Sponsor -> sponsoriser

(83) Stress -> stresser

(84) Bug -> débogage

Un emprunt, une fois bien intégré, peut également élargir son domaine d'emploi. Certains termes passent d'un langage de spécialité au langage courant.

(85) *Black-out* n'est plus seulement réservé au domaine de la guerre

(86) *Buldozer* peut être utilisé dans un sens métaphorique

6 LA SEMANTIQUE LEXICALE

Nous avons à plusieurs reprises évoqué jusqu'ici le sens des morphèmes et des unités du lexique en général. **La sémantique est la discipline de la linguistique qui étudie les signifiés associés aux formes linguistiques et la façon dont ces signifiés se combinent dans des unités minimales et dans des unités plus complexes (la phrase).** La sémantique lexicale décrit l'organisation du sens des mots *dans la langue*.

6.1 Du signifiant au signifié: sémasiologie

On peut caractériser le sens sémantique associé à chaque mot. L'approche sémasiologique décrit le lexique en partant des formes (voir par exemple les entrées du dictionnaire) et en énumérant les différentes significations. En général, les sens propres (premiers, concrets) sont placés avant les sens figurés (dérivés, métaphoriques). C'est dans cette approche que sont construits les dictionnaires de langue: les mots sont classés par ordre alphabétique (de formes) et sont définis.

Le sens associé à une unité lexicale est décomposable dans la mesure où il est constitué à partir du sens de chaque morphème qui compose l'unité en question (**compositionnalité**). Dans le cas de la flexion et de la dérivation, l'ajout d'un morphème grammatical va ajouter un trait sémantique (**sème**) au sens du morphème lexical qui sert de radical/base. Dans le cas de la composition, une des deux bases possède un statut particulier, elle joue le rôle de tête, en ce sens qu'elle relie le mot à une classe sémantique particulière : un *roman-photo* est une sorte de roman, un *feu rouge*, une sorte de feu. En français la tête est en général à gauche, mais elle peut être à droite (ex. *libre-échange*). Dans le cas de la réduplication, la répétition d'une forme tronquée de la base ajoute une valeur sémantique diminutive/atténuative voire dépréciative (un homme *foufou* est un peu excentrique, une *guéguerre* est une petite guerre ou une guerre peu sérieuse, ridicule).

6.1.1 La polysémie

Un même mot présente bien souvent plusieurs sens (par ex. *fruit* "produit comestible sucré" ou "résultat" voire "enfant, progéniture") – en moyenne, chaque mot de la langue dispose de trois significations. C'est ce que l'on appelle « la polysémie d'une unité lexicale » :

- C'est l'analyse de la distribution dans différents contextes sémantiques d'une même unité qui permet d'en connaître les différentes valeurs sémantiques.
- Ces différents sens sont reliés entre eux et sont dérivés les uns les autres en fonction de différents procédés comme la métaphore/métonymie, la généralisation, la spécialisation, etc.

On peut dès lors représenter les différents sens d'un mot par un réseau radial (avec des rayons) – parce que les sens sont reliés entre eux et sont regroupés en fonction des traits de signification qu'ils actualisent.

- (3) **canard** (n.m.)
- a. oiseau aquatique
 - b. viande de cet oiseau
 - c. fausse nouvelle
 - d. journal (contenu ou matérialité)
 - e. fausse note criarde
 - f. morceau de sucre trempé dans le café, l'alcool, etc.
 - g. (loc.) un "vilain petit **canard**"

Ces huit **sens** différents de **canard** peuvent être illustrés par les contextes suivants :

- a. Le seul **canard** que je suis capable de reconnaître est le colvert
- b. Hier j'ai mangé du **canard** à l'orange
- c. Chaque mois cette revue fait courir des canards
- d. J'ai lu cette nouvelle dans différents canards
- e. J'ai fait un couac de **canard** dans le premier morceau que j'ai joué à l'examen
- f. Il ne peut s'empêcher de faire un **canard** en prenant le pousse-café
- g. À l'école, c'est toujours lui qui est le vilain petit **canard**

Figure (1) Extension des sens du mot « canard » (Delbecq 2006)

Quels sont les processus qui relient entre eux ces différents sens à l'intérieur du réseau. **Un réseau sémantique est toujours organisé autour d'un sens central ou prototypique** qui permet d'expliquer les autres sens, dérivés (cf. chapitre 3 de ce cours). Pour un mot donné ("fruit", "oiseau", "meuble", "lien de parenté"...), il y a **trois manières pour déterminer quel sens est le plus central**:

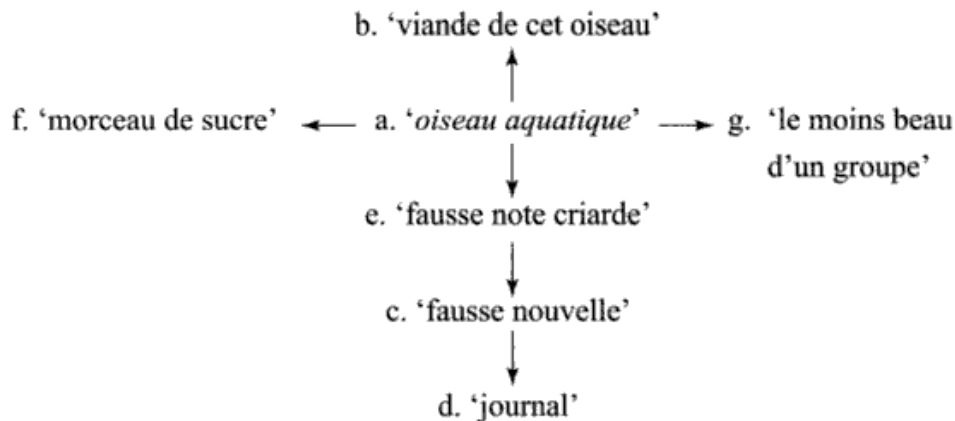
- La première méthode consiste à enquêter pour savoir quel est le sens qui vient le premier à l'esprit (pour *fruit* c'est "comestible sucré" plutôt que "enfant").
- La seconde méthode consiste à compter dans un corpus de texte quel sens du mot est utilisé le plus souvent.
- La troisième méthode est de chercher quel sens est à l'origine des autres, de quel sens premier ou original les autres sens du mot dérivent. Cette troisième méthode repose sur l'évolution historique (diachronique) du sens du mot.

Dans le cas de *canard*, on pourrait sans risque d'erreur poser que le sens prototypique est "oiseau aquatique élevé pour sa chair savoureuse – qui plonge – maladroit dans sa démarche terrestre – qui produit un cri malsonnant – comparable à un bavardage? – qui a un bec proéminent"

- un premier sens concret est dérivé par **métonymie** (relation de contiguïté, qui permet de distinguer les parties d'un tout: "la voile", "un verre"...): la viande de canard – la couleur vert canard ;
- un second processus pour dériver un sens, par élargissement, est **la métaphore** (du gr. "transposer"), qui consiste à transposer un élément d'un premier domaine (source) vers un autre domaine (cible) – à cause d'une ressemblance qui est subjectivement établie (par ex. *ped de montagne*) : le sens "cri malsonnant" est

transposé dans le domaine musical pour "fausse note criarde", ... dans le domaine de la rumeur publique avec le sens de "fausse nouvelle", l'aspect "plongeon" ... dans le domaine culinaire donne le sens "morceau de sucre trempé dans le café" ;

- ensuite, un **sens peut être dérivé par spécialisation**: "mauvais journal" est une spécialisation à partir de "fausse nouvelle" (on restreint la signification initiale de sorte qu'elle s'applique à un nombre plus réduit de référents spécifiques) – la presse est un moyen particulier de diffuser des fausses nouvelles (ex. *boisson* "tout liquide qui se boit" => "boisson alcoolisée) ;
- le processus inverse est **la généralisation/extension**: le sens "le moins beau" (*le vilain petit canard*) peut être une généralisation à partir de l'aspect "démarche maladroite" ... (ex. *atterrir* à l'origine "toucher la planète terre" => on peut atterrir sur la lune (ou *alunir*)).



Réseau radial des sens de « canard » (Delbecque 2006)

6.1.2 Polysémie vs homonymie

La polysémie d'une même lexie ne doit pas être confondue avec l'existence de plusieurs unités homonymes ayant par définition des sens différents. On parle d'homonyme lorsque deux mots, d'origine différente, présentent exactement la même prononciation (homophones) ou la même graphie (homographes) :

bar "débit de boissons" (de l'angl. *bar-room* "barre de comptoir") ou *bar* (du gr. *baros* "pesanteur") "unité de mesure de pression des fluides" ou *bar* (du néer. *baer*) "poisson marin (v. loup)" – *charme* "qualité de ce qui attire" ou *charme* "arbre feuillu" – *bise* "vent du nord" ou *bise* "baiser"

Un **paronyme** est un mot dont la prononciation et l'orthographe ressemblent à celles d'un autre mot, mais dont le sens diffère. Cette ressemblance formelle entre deux mots est une source d'erreurs, ou de jeux de mots : *inculper et inculquer* ; *éruption et irruption* ; *collision et collusion* ; *amnistie et armistice* ; *éminent et imminent* ; *prescrire et proscrire* ; *papille et pupille*...

6.2 Des signifiés aux signifiants : onomasiologie

L'approche onomasiologique part d'un concept et s'applique à décrire des mots ou les locutions qui permettent de le désigner. L'exemple typique est le thésaurus (ou l'encyclopédie) qui est un "ouvrage dans lequel les mots sont regroupés en fonction de leur signification, par thèmes".

- (87) Par ex. pour le concept "fruit", on trouve les expressions: 1. fruits rouges, graines, noix, pépin, semence, agrume... ou 2. effet, conséquence, impact... suite, produit, production... rejet, rejeton, descendance.

6.2.1 Synonymie et antonymie

*L'onomasiologie traite donc des relations de synonymie et d'antonymie. La **synonymie** décrit la relation entre deux mots ayant (presque) la même signification (*beau, joli, ravissant, charmant...*). On considère en général qu'il n'existe pas de synonymie "parfaite" au sein d'une même variété de langue.*

- (88) salaire / traitement [fonctionnaire], appointements [employés], honoraires / émoluments [profession libérale], gages [domestique]

Il y a des cas de synonymie partielle.

- (89) magazine / revue [publication périodique], mais pas revue dans le sens d'inventaire (revue de presse) ou de pièce satirique (la revue de fin d'année)

L'**antonymie** décrit la relation entre deux mots de significations "opposées" – on peut parler de manière plus générale d'**oppositions** de sens. On distingue principalement l'antonymie polaire ou binaire (nier l'un est affirmer l'autre (ex. mâle vs femelle; vivant vs mort)) et l'antonymie scalaire ou graduelle (ex. *chaud vs froid ; beaucoup vs peu*). Notez qu'en raison de la polysémie, deux termes peuvent entretenir une opposition binaire pour un de leur sens et une opposition scalaire pour un autre sens : ex. *il est vivant ou il est mort* (opposition binaire) ; *cette soirée est plutôt vivante/morte* (opposition scalaire). À cela s'ajoute le cas des antonymes **récioproques** : ces antonymes n'expriment pas deux propriétés opposées d'un même objet (*grand vs petit*) mais bien une relation de symétrie inversée entre deux entités : *acheter* « s'oppose » à *vendre* ; *père* « s'oppose » à *fil*.

6.2.2 Champs lexicaux et traits sémantiques

Des mots qui appartiennent à la même catégorie syntaxique (nom, adjectif, verbe) font partie, sur le plan syntaxique, d'une même classe paradigmatique (ils sont dès lors substituables, commutables le long de l'axe paradigmatique dans un contexte donné). Si en outre, ils partagent, à l'instar des synonymes et antonymes, un élément de sens commun, ils forment alors sur le plan sémantique un ensemble que l'on appelle champ lexical. A l'intérieur de ce réseau de formes et de significations proches, les unités sont organisées en fonction des relations qu'elles entretiennent (ex. synonymie vs antonymie, hyperonymie vs hyponymie), ainsi qu'en fonction des connotations³⁷ qu'elles véhiculent.

On se souvient que pour Saussure et les structuralistes qui ont suivi, les signes se définissent comme la somme de ce que les autres ne sont pas (une somme de différences, de substance et de contenu). Les signes dans une langue s'organisent en un système et c'est au sein du système que les éléments se définissent les uns par rapport aux autres. En ce sens, les unités qui composent le champ lexical (appelé aussi conceptuel) ont un signifié qui est limité par celui des autres unités du même champ (= définition relationnelle et oppositionnelle) ; un élément du champ se définit par opposition à tout ce qui n'est pas lui.

Dans cette logique relationnelle, chaque unité est néanmoins caractérisée « positivement » par des traits distinctifs qui servent à la distinguer des autres unités et qui font que le sens de chaque

³⁷Il s'agit des significations secondaires qui viennent s'ajouter, en contexte et selon des associations relativement particulières, au sens premier, dit dénotatif, plus stable et partagé par la communauté linguistique (*baraque* a le sens dénotatif de 'maison', pour tout francophone, mais pourra connoter une maison de peu de valeur chez certains, et signaler simplement le recours à un registre familier chez d'autres).

unité lexicale peut se décomposer en unités de sens plus réduites. On peut ainsi représenter ce qui différencie les éléments d'un même champ lexical sous la forme d'une matrice qui présente les « traits de significations » (**sèmes**) et la manière dont ils sont actualisés ou non. Un trait peut prendre une valeur positive (+), négative (-) ou neutre (∅).

	<i>pour s'asseoir</i>	<i>sur pied</i>	<i>pour une personne</i>	<i>avec dossier</i>	<i>avec bras</i>	<i>matériau rigide</i>
siège	+	∅	∅	∅	∅	∅
chaise	+	+	+	+	-	+
fauteuil	+	+	+	+	+	+
tabouret	+	+	+	-	-	+
canapé	+	+	-	+	∅	+
pouf	+	-	+	-	-	-

Représentation selon Pottier (2000).

C'est à nouveau grâce à l'opération de commutation que l'on peut identifier la fonction exercée par différentes unités et les effets de sens qu'elles permettent de produire grâce à leurs traits distinctifs particuliers.

De la même manière, dans une phrase comme « *Belle Marquise, vos beaux yeux me font mourir d'amour* », on peut identifier spontanément plusieurs groupes de mots ou constituants grâce à l'opération de commutation. Tous ces groupes de mots n'ont cependant pas le même statut. En effet, on peut distinguer des groupes de tailles différentes : « distribue le courrier » est un constituant, mais c'est également le cas pour « le courrier » ; « me font mourir d'amour » est un constituant, mais c'est également le cas de « d'amour ». Il s'agit donc d'identifier dans un premier temps les constituants *immédiats* d'une unité, c'est-à-dire les plus grands constituants qui forment l'unité que l'on cherche à décomposer (au départ, la phrase). Chacun de ces constituants peut ensuite être décomposé en constituants plus petits, en répétant l'opération de substitution (commutation).

En résumé, le **constituant** (ou **groupe** ou **syntagme**) est l'unité syntaxique de base: il s'agit d'un groupement provisoire d'éléments dans le cadre d'une phrase, selon des règles particulières. Selon Soutet (1995: 249), un syntagme³⁸ est une suite de mots :

- isolable par le procédé de segmentation/commutation,
- inférieure à la phrase,
- composée d'unités n'ayant pas de solidarité préalable à la solidarité qui les lie, de manière provisoire, dans un cadre discursif particulier.

7.2.2 La structure hiérarchique de la phrase

Comme le montrent les exemples en (2), toute phrase est formée de deux constituants ou **syntagmes de base** : le syntagme (groupe) nominal et le syntagme (groupe) verbal. Ces deux syntagmes de base sont eux-mêmes composés d'autres constituants selon des **relations hiérarchiques**. De ce fait l'ACI fait apparaître la notion de **rang** ou de **niveau d'organisation** : la phrase n'est pas une suite linéairement ordonnée de mots mais un emboîtement d'unités à différents niveaux. Le niveau supérieur est celui de la phrase complète ; vient ensuite le niveau des constituants immédiats de la phrase (le GN et le GV) ; chaque constituant est analysable en constituants immédiats d'un niveau inférieur, etc.

L'ACI se représente dès lors le plus souvent sous la forme de **boîtes**, qui se lisent depuis la phrase complète jusqu'à ses constituants les plus petits (les mots ou les morphèmes, selon la finesse de l'analyse).

Un	très	joli	garçon	siffle	la	Traviata	GN + GV
Un	très	joli	garçon	siffle	la	Traviata	GN + V + GN
Un	très	joli	garçon	siffle	la	Traviata	Dét. Gadj + N + V + Dét + N
Un	très	joli	garçon	siffle	la	Traviata	Dét + Adv + Adj + N + V + Dét + N

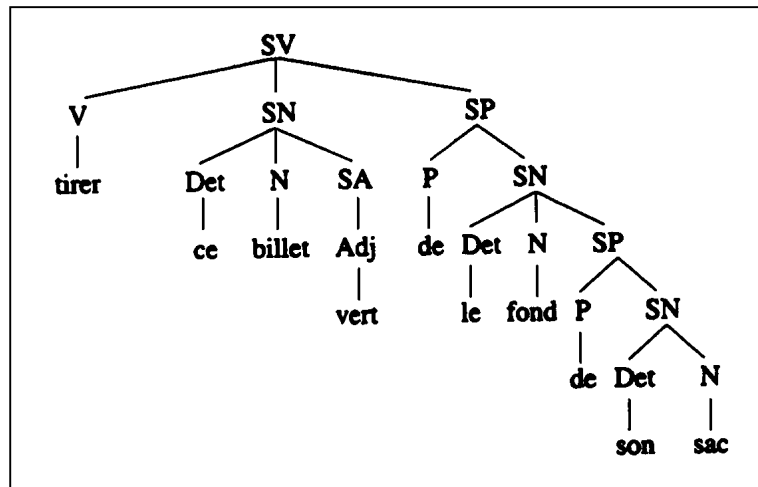
Analyse en constituants immédiats (Béguelin 2000: 106)

La structure hiérarchique des phrases se représente également sous forme de parenthèses étiquetées ou sous forme de schéma arborescent. L'exemple du Groupe Verbal (ou Syntagme Verbal) ci-dessous illustre la représentation par parenthèses (on note sous chaque parenthèse fermante le nom du constituant qu'elle contient) et dans la figure 2 sous la forme d'un arbre syntaxique.

(92) [(tirer)_V (ce billet vert)_{GN} (du fond (de son sac)_{GPrép})_{GPrép}]_{GV}

³⁸ En général, on désigne le syntagme à partir de son unité noyau (syntagme nominal *le petit chat*, syntagme adverbial: *très joliment*, etc.).

Dans la représentation arborescente, la racine de l'arbre – nœud supérieur – représente le groupe complet ; les nœuds intermédiaires représentent les sous-constituants ; les feuilles de l'arbre (éléments terminaux) représentent les catégories grammaticales.



Représentation arborescente d'un Groupe Verbal (extrait de Tellier 1995)

Précisons que la **fonction** des relations hiérarchiques entre les syntagmes est de nature sémantique : la façon dont les constituants de la phrase se combinent entre eux permet de marquer les relations sémantiques qui les lient, chaque phrase assurant par ailleurs l'intégration de ces constituants au sein d'unités de signification délimitées qui permettent la transmission progressive de l'information, tandis que les relations entre les phrases successives permettent de tisser progressivement la trame du discours. Le sens au niveau des phrases est donc **compositionnel** : la signification d'une unité complexe est construite à partir (i) de la signification que ses parties ont indépendamment et (ii) de règles de combinaisons syntaxiques et sémantiques.

7.3 Les classes de mots (analyse distributionnelle)

Les types de mots qui composent un constituant syntaxique ne sont pas sélectionnés et organisés de façon aléatoire. Pour être grammatical (pour respecter les règles de la grammaire d'une langue), un constituant doit être composé de mots de certaines catégories, ordonnés selon des règles syntagmatiques définies. Nous commençons par examiner quelles sont les catégories de mots pertinentes du point de vue de la syntaxe avant d'envisager leurs règles d'agencement (voir §4).

7.3.1 La notion de distribution

L'analyse distributionnelle vise, à partir d'un corpus suffisamment important de phrases grammaticales d'une langue, à segmenter chaque phrase en unités constitutives et à décrire la distribution des unités qu'on a identifiées.

La **distribution** d'un élément se définit comme l'ensemble des environnements possibles (contexte linguistique à gauche et à droite de l'unité étudiée) pour cet élément. Les éléments qui partagent la même distribution forment une **classe (distributionnelle)** ou une **catégorie**. Par exemple, si on part du petit corpus suivant (4), on peut établir une liste d'exemples (5) où l'élément mis en gras se trouve chaque fois dans l'environnement suivant :

___ Nom ou

___ Adjectif Nom (remarque : le signe # désigne une frontière de phrase, de groupe...).

(93) Il était une fois une petite fille qu'on appelait Chaperon Rouge, on lui avait donné ce nom parce sa maman lui avait fait un très joli bonnet rouge. Et, à l'époque, les bonnets s'appelaient des chaperons. Au village, quand on la voyait arriver, on disait: Tiens voilà le Petit Chaperon Rouge. Le petit Chaperon Rouge avait une grand-maman qui vivait seule à l'autre bout de la forêt. La vieille dame très âgée, ne sortait pas beaucoup de sa maison. Un jour sa maman lui dit: Grand-maman est malade, j'ai fait ce matin des galettes que ta grand-maman aime tant, sois gentille, va donc les lui porter avec ce fromage et ce petit pot de confiture. Mais ne quitte pas le sentier, car papa a vu des loups qui rôdaient dans la forêt.

(94) Liste de groupes syntaxiques ayant une structure interne comparable

une fois
une petite fille
sa maman
un très joli bonnet
rouge
l'époque
les bonnets
des chaperons
le Petit Chaperon
Rouge
une grand-maman
l'autre bout
la forêt
la vieille dame
sa maison
un jour
sa maman
ce matin
des galettes
ta grand-maman
ce fromage
ce petit pot
le sentier
des loups
la forêt
au (=à **le**) village

Ces éléments en gras (des **déterminants**, nous le savons par notre culture grammaticale) fonctionnent comme des accompagnateurs obligatoires du nom. Ils entretiennent donc une relation syntagmatique étroite avec le nom, et s'accordent avec celui-ci (voir morphologie flexionnelle). Leur absence entraîne l'agrammaticalité de la construction du nom : **il était fois petite fille*.

7.3.2 Catégorie, sous-catégorie et extension d'une catégorie

Les éléments partageant une même distribution forment une catégorie grammaticale (et forment une classe paradigmatique). En conséquence, ces éléments sont commutables au sein de la chaîne syntagmatique (moyennant le respect des règles d'accord avec le nom) et leur commutation n'entraîne pas d'agrammaticalité. Contrairement à l'approche grammaticale traditionnelle, ce type d'analyse ne se fonde pas sur la fonction ou le sens d'une unité pour en définir la catégorie (voir ci-dessous avec la question des adverbes), mais uniquement sur sa distribution.

- (95) { une , sa , la , cette , ... } petite fille
 { un , mon , le , ce , ... } loup

Il arrive que des éléments aient à la fois des contextes identiques et des contextes différents. L'exemple (7) montre que certains déterminants acceptent de se combiner entre eux (*les deux arbres, les quelques arbres*) tandis que d'autres ne le peuvent jamais (**les ses arbres, *certains deux arbres*). Ces deux combinatoires (ou distributions) distinctes définissent, dans la classe des déterminants, deux sous-classes ou sous-catégories.

- (96) les { deux , quelques } arbres
 les { *ses, *certains } arbres

On distingue ainsi, à côté des déterminants classiques, un prédéterminant (*tout*), des post-déterminants (*deux, quelques...*), etc.

Un autre exemple (8) concerne l'analyse de *très* et *ici*, traditionnellement décrits comme des adverbes. En effet, leur distribution semble en disjonction : *ici* peut servir de complément à un verbe, alors que *très* ne le peut pas; inversement, *très* peut modifier un adjectif, ce qui n'est jamais le cas de *ici*. Cette analyse remet en cause la définition homogène de la catégorie des adverbes.

- (97) Marie travaille **ici**.
 *Marie travaille **très**.
 Gustave est **très** fatigué.
 *Gustave est **ici** fatigué.

Si l'on revient aux déterminants, l'opération de commutation démontre que certaines unités, non identifiées a priori comme des déterminants, apparaissent dans le même environnement (9). Ces éléments doivent également être intégrés dans la catégorie des déterminants, qui s'en trouve étendue.

- (98) { les , une foule de , beaucoup de , quels ... } gens

7.3.3 Autres critères de caractérisation d'une catégorie grammaticale

Si les éléments d'une même catégorie grammaticale se caractérisent essentiellement par leur distribution, ils partagent également certaines propriétés morphologiques : les unités de certaines catégories grammaticales subissent en effet des variations morphologiques particulières, d'autres pas, en fonction des éléments qui les régissent. Le déterminant varie ainsi en genre, en personne et en nombre selon le nom qu'il accompagne. Les éléments d'une même classe partagent également une même fonction (critère fonctionnel ou syntagmatique) : ils se

définissent par le rôle qu'ils jouent dans le syntagme auquel ils appartiennent. Le déterminant fonctionne comme un spécificateur du nom et il est indispensable à la formation du syntagme nominal.

7.3.4 Catégories grammaticales couramment admises

La définition des catégories grammaticales de base peut s'avérer problématique. On reconnaît généralement les catégories suivantes.

(1) Catégories majeures ou lexicales (peuvent former la tête de syntagmes)

- Nom (propres et communs)
- Verbes
- Adjectifs
- Adverbes (avec plusieurs sous-catégories)
- Prépositions

(2) Catégories mineures ou grammaticales (ne se réalisent pas de manière autonomes ; entrent dans la composition des groupes)

- Déterminants
- Marqueurs d'enchaînement ou de coordination (si, parce que... ; mais, et...)
- Auxiliaires et verbes modaux
- Indicateurs de transformation de la phrase (ne... pas, est-ce que...)

7.3.5 Conclusion

L'analyse distributionnelle part du principe que les unités lexicales se définissent au niveau syntaxique par les relations syntagmatiques et paradigmatisées qu'elles entretiennent les unes avec les autres. L'avantage est de poser des bases méthodologiques explicites et précises pour le classement des unités, alors que la classification traditionnelle regroupe parfois, sur une base notionnelle, des éléments qui n'ont pas le même comportement syntaxique. L'analyse distributionnelle a ainsi remis en cause un certain nombre de classifications grammaticales traditionnelles (adverbes, déterminants, mais aussi pronoms, etc.).

7.4 Les règles syntagmatiques

Les unités syntaxiques (les syntagmes) sont constituées selon des règles (syntagmatiques) qui président à la mise en relation des mots sur l'axe syntagmatique et qui dépendent de propriétés catégorielles, c'est-à-dire liées à l'appartenance des mots à une catégorie grammaticale.

7.4.1 La structure interne des groupes syntaxiques

Chaque type d'unité syntaxique présente ainsi une structure interne définie et qui se caractérise par la catégorie de l'élément qui lui sert de **tête** (de noyau). La nomenclature actuelle (Béguelin 2000: 111) identifie cinq types de syntagmes (ou groupes): le syntagme nominal, le syntagme verbal, le syntagme adjectival, le syntagme adverbial et le syntagme prépositionnel. Chacun de ces groupes est construit autour d'une catégorie syntaxique: le nom, le verbe, l'adjectif, l'adverbe et la préposition. La phrase présente également une structure interne définie, mais on la traite à part, étant donné qu'elle constitue le niveau maximal d'intégration syntaxique (voir §5).

Pour chaque groupe, on mentionne ses constituants obligatoires et facultatifs (ces derniers entre parenthèses). L'ordre d'apparition des constituants à l'intérieur de chaque type de syntagme est également défini par des règles propres à chaque langue, même s'ils ne sont pas détaillés ici. Les schémas suivants sont extraits de Béguelin (2000).

7.4.1.1. Le Groupe Nominal (GN)

Le groupe nominal comporte deux constituants de base :

1. le nom (ou un « équivalent du nom ») ;
2. le déterminant.

Il accepte les constituants secondaires suivants :

3. groupe adjectival ;
4. groupe prépositionnel ;
5. phrase subordonnée.

Gr. Nominal → Détermination + Nom + (GAdj⁷)
+ (GPrép) + (Phrase subordonnée)

Lorsqu'un constituant facultatif peut être équivalent à une unité simple ou complexe (dans ce cas, l'adjectif ou le groupe adjectival), la règle ne mentionne que le groupe.

Les pronoms de conjugaison sont considérés comme des équivalents du Nom, dans cette analyse, même si leur position dans la phrase n'est pas toujours identique à celle des Noms pleins.

(99) *Vient **son amie** demain ?

Vient-**elle** demain ?

On a vu précédemment que certaines sous-catégories de noms (comme les noms propres) se construisent sans détermination (voir différence entre « le frère de Jean » et « frère Jean »).

7.4.1.1 7.4.1.2. Le Groupe Verbal (GV)

On distingue deux sous-catégories de verbes qui ne se construisent pas de la même manière

- les verbes de types *être* (sembler, paraître...) ou verbes copules (Vcop)
- les autres verbes, à complément Vcpl

Le groupe verbal comporte un constituant de base :

1. le verbe.

Il accepte les constituants secondaires suivants :

2. groupe nominal, ou un *équivalent du nom* ;
3. groupe adverbial (Vcpl) ;
4. groupe adjectival (Vcop) ;
5. groupe prépositionnel ;
6. phrase subordonnée.

Gr. Verbal (cpl) → Vcpl + (GN) + (GPrép) + (GAdv) + (Phrase subordonnée)
Gr. Verbal (cop) → Vcop + GAdj ou GN ou GPrép ou (Phrase subordonnée)

Le verbe *être* est généralement un Vcop qui demande un attribut ; ce dernier se pronominalise par *le* ou *la* (11). Mais le verbe *être* peut aussi être un verbe à complément, qui se pronominalise alors par *y* (12). Dans ce cas, le verbe *être* peut être paraphrasé par *se trouver*.

(100) Jean est intelligent.

Jean l'est.

(101) Jean est dans le verger.

Jean y est.

Les verbes peuvent régulièrement prendre une forme complexe, à cause de leur flexion (c'est le domaine de la conjugaison) ou de l'adjonction de verbes modaux ou de particules (domaine de la modalisation) :

(102) il *a été parti* pendant longtemps
il *semble avoir pu convaincre* son frère
il *peut se réjouir* de son succès
il *ne croit pas* en ce projet

7.4.1.2 Le Groupe Adjectival (GAdj)

Le groupe adjectival comporte un constituant de base :

1. l'adjectif.

Il accepte les constituants secondaires suivants :

2. groupe prépositionnel ;
3. groupe adverbial ;
4. phrase subordonnée

Gr. Adjectival → (GAdv) + Adjectif + (GPrép) + (Phrase subordonnée)

Le groupe adjectival est fréquemment composé d'un adjectif seul, ou d'une suite d'adjectifs, qui sont le plus souvent coordonnés mais pas systématiquement.

(103) Nous avons acheté des gants **blancs**.

Les sommets dominaient des espaces **enneigés et sauvages**.

Une **belle** maison **traditionnelle rénovée** avec goût.

7.4.1.3 Le Groupe Adverbial (GAdv)

Le groupe adverbial comporte un constituant de base :

1. l'adverbe ;

Il accepte les constituants secondaires suivants :

2. groupe adverbial ;
3. groupe prépositionnel ;
4. phrase subordonnée.

Gr. Adverbial → (GAdv) + Adverbe + (GPrép) + (Phrase subordonnée)

Comme pour l'adjectif, le Groupe Adverbial est souvent composé d'un adverbe seul. Contrairement aux autres classes analysées jusqu'à présent, l'adverbe est invariable.

(104) Il parle **méchamment**.

Manuel est **très récemment** parti en voyage.

Contrairement à toi, je m'abstiendrai.

7.4.1.4 Le Groupe Prépositionnel (GPrép)

La structure du groupe prépositionnel se distingue des autres par la présence obligatoire d'au moins un autre constituant qui accompagne la préposition. En effet, un élément comme *contre* dans (109) s'analyse syntaxiquement comme un adverbe et non comme une préposition.

(105) J'ai voté **contre**.

Le groupe prépositionnel comporte un constituant de base :

1. la préposition.

Celle-ci est nécessairement accompagnée d'un des constituants suivants :

2. groupe nominal (ou *équivalent du nom*) ;
3. groupe adjectival ;
4. groupe adverbial ;
5. phrase subordonnée

Gr. Prép. → Prép + GN ou GAdj ou GAdv ou phrase subordonnée

La préposition a obligatoirement un élément qui la suit.

(106) Le chapeau **de mon oncle** est magnifique. (2)

Nous avons mangé quelque chose **de bon**. (3)

Paul a acheté le journal **d'hier**. (4)

Avant que l'hiver s'installe, il fallait nettoyer le jardin. (5)

7.4.2 L'ambiguïté, un problème classique en syntaxe

Une phrase est ambiguë si elle est passible de plusieurs interprétations.

(107) *They can fish*

→ ils peuvent / savent pêcher (*to fish*)

→ ils mettent du poisson en conserve (*to can*)

(108) *La belle porte le voile.*

→ (La belle porte)_{GN} (le voile)_{GV}.

→ (La belle)_{GN} (porte le voile)_{GV}.

Dans le même genre : *Le cuisinier sale la note. La petite brise la glace.*

La belle ferme le voile...

(109) *Cette doublure est inefficace.*

Doublure : (1) étoffe dont une autre est doublée ; (2) personne qui remplace un comédien...

Si l'ambiguïté naît de différents réseaux de relations déployés entre les constituants phrastiques, elle est de nature syntaxique et dite structurale. Si elle résulte d'unités polysémiques ou homonymiques, elle est lexicale (Garric 2001 : 162).

8.1 Les phonèmes

Nous avons vu plus tôt dans le chapitre 5 que les unités maximales du lexique, les lexies, étaient composées de morphèmes, unités minimales de la langue possédant deux faces : une face représentante – signifiant – et une face représentée – le signifié. Dans le présent chapitre, nous nous posons la question de savoir si ces morphèmes peuvent eux-mêmes être décomposés et selon quels principes sont formés.

Prenons la série de mots suivants : *balle*, *Bill*, *bulle*, *boule*. Dans tous ces mots, seule la voyelle diffère et cela suffit pour distinguer quatre lexies différentes. Chaque paire de cette série de mots constitue ce que l'on appelle une **paire minimale**. La commutation nous permet de voir que cette voyelle constitue une unité indépendante du reste du mot, qui permet de le caractériser en le distinguant des autres mots de cette série. Cela prouve le caractère décomposable des morphèmes et l'existence d'unités linguistiques en deçà du niveau morphologique.

Cette unité, que l'on appelle **phonème** n'a pas de signifié ; c'est une unité de la deuxième articulation du langage³⁹. Le phonème n'a pas de sens⁴⁰, sa *fonction* est de permettre de produire et d'identifier le signifiant des différents morphèmes qui composent la langue : le phonème /a/ permet ainsi de composer le signifiant (le morphe) du lexème *balle*, mais aussi de désigner à lui seul le morphe de la forme fléchie du verbe *avoir* (*a*) ou de la préposition *à*. Deux questions se posent dès lors que l'on a identifié l'existence des phonèmes :

- comment faisons-nous pour produire ou reconnaître un morphème à partir des phonèmes qui le composent ?
- quelles sont les règles d'agencement des phonèmes qui organisent la composition des morphèmes dans une langue donnée ?

Examinons les trois paires minimales suivantes :

(110) espagnol : *pero* 'mais' – *perro* 'chien'

(111) anglais : *sin* 'péché' – *thin* 'fin'

(112) français : *banc* – *bon*

On constate que dans chaque cas, une différence de *son* permet de créer une différence de sens dans une de ces trois langues, alors que la même différence de *son* n'aurait aucun effet dans les deux autres langues. En français, par exemple, la différence de son entre *banc* et *bon* permet d'analyser ces deux sons – que l'on notera respectivement [bã] et [bõ] – comme la **réalisation phonétique** de deux morphèmes qui peuvent être analysés par tout francophone comme étant composés de deux unités, deux phonèmes : /b/ + /ã/ dans le premier cas, /b/ + /õ/ dans le second ;

³⁹ Comme nous l'avons déjà mentionné dans les chapitres précédents, on appelle « double articulation », dans l'hypothèse fonctionnaliste de A. Martinet, l'organisation spécifique du langage humain selon laquelle tout énoncé s'articule sur deux plans. Au premier plan (ou première articulation), tout énoncé s'articule linéairement en unités douées de sens (unités significatives : phrases, syntagmes, mots, etc.) dont les plus petites unités sont appelées monèmes (**ou morphèmes**). Au deuxième plan d'articulation, chaque monème s'articule à son tour dans son signifiant en unités dépourvues de sens (unités distinctives) dont les plus petites sont les **phonèmes**, en nombre limité dans chaque langue.

⁴⁰ Si elle en avait un (si par exemple le phonème /a/ avait un sens défini), ce sens serait nécessairement impliqué dans tout mot comprenant ce phonème, ce qui limiterait grandement la possibilité de créer des unités lexicales de sens variés comprenant ce phonème.

la différence au niveau de l'unité vocalique suffisant pour permettre de distinguer (en production comme en réception) les deux morphèmes. Cela signifie que les locuteurs francophones associent les sons physiques, les réalités acoustiques [ã] et [õ] avec deux unités particulières du système de la langue (ce que les hispanophones et les anglophones ne font pas). En revanche, un francophone qui ne parle pas anglais ne fera pas de différence entre *sin* [sin] et *thin* [θin], parce qu'il n'a pas appris à associer la différence acoustique entre les sons [s] et [θ] à deux unités phonologiques distinctes (le son [θ] étant absent en français).

Prenons en outre l'exemple du mot *coude* prononcé avec le son [t] en final. Le son produit ([kut]) sera sans difficulté associé par un francophone au signifiant du morphème *coude*, de sorte qu'il sera capable d'analyser le son physique [t] comme une réalisation possible de l'unité de la langue (du phonème) /d/.

Cela montre par conséquent :

- que dans chaque langue (chaque système linguistique), certains sons parmi ceux que les humains peuvent produire sont associés à des unités qui n'ont de valeur – « l'information » qui leur permet d'avoir une fonction dans la langue – que les unes par opposition aux autres (un son ne correspond à un phonème ayant une valeur que si ce son renvoie à une unité qui s'oppose à une autre unité dans le système) ;
- que la valeur d'un phonème lui permet d'avoir une fonction *distinctive* en permettant d'identifier les unités morphologiques ;
- que les locuteurs sont capables, dans certaines circonstances, de rapporter plusieurs substances sonores concrètes à une même forme ou unité abstraite (un même phonème) et dès lors de percevoir différentes prononciations d'un même mot comme renvoyant au même signifiant (tout comme ils sont capables, grâce au contexte, de rapporter ce signifiant à différents morphèmes homonymes).

Nous retrouvons donc, au niveau des phonèmes, deux caractéristiques des unités des autres niveaux de la langue :

- ces unités s'organisent en système d'opposition,
- ces unités sont des réalités abstraites qui peuvent être matérialisées par différentes réalisations concrètes.

La **phonologie** est l'étude de l'organisation des phonèmes, de leurs caractéristiques et des conditions qui déterminent leur réalisation phonétique concrète. Le **système phonologique** d'une langue définit l'ensemble des sons qui permettent de distinguer les morphèmes de la langue en question.

8.2 La phonétique

Avant de pouvoir décrire le système phonologique de différentes langues, il nous faut pouvoir identifier, désigner et caractériser les différents sons auxquels les locuteurs d'une langue donnée peuvent attribuer une valeur phonologique. Cette tâche est celle de la **phonétique**. La phonétique est la science qui étudie les sons du langage humain, en tant que réalité physique. Elle mobilise pour cela les connaissances en biologie (anatomie), en physique (acoustique), en neurologie/neuropsychologie (capacités perceptives et cognitives du cerveau humain). La phonétique se subdivise en différentes branches selon qu'elle s'intéresse :

- à la production des sons (phonétique articulatoire),
- à la transmission des sons et à leurs propriétés physiques (phonétique acoustique),

- à la réception et à la perception des sons (phonétique réceptive, auditive ou perceptive).

Afin de bien comprendre les caractéristiques des sons du langage, il est fondamental de comprendre leurs conditions de production, c'est pourquoi nous intéresserons essentiellement à décrire quelques principes de **phonétique articulatoire**.

La production des sons, étudiée par la phonétique articulatoire, comprend 3 étapes : la respiration (souffle) permet de produire un courant d'air qui va passer par la glotte (cordes vocales), la bouche et / ou le nez ; la phonation, ou production du son, résulte de la mise en vibration des cordes vocales ou du bruit de l'air à travers les résonateurs ; l'articulation transforme la voix ou le bruit informe en sons distincts (consonnes et voyelles) par la mobilisation des articulateurs (langue, lèvres, ...).

En d'autres mots, une source sonore est produite (par la vibration des cordes vocales ou par le passage de l'air dans le conduit bucco-nasal) et ce son est modelé par l'articulation (qui amplifie certaines fréquences et en atténue d'autres).

Du point de vue **acoustique**, le son est produit par une mise en vibration de l'air. Le son laryngé, produit lors de la réalisation des voyelles ou des consonnes voisées, est un son « non pur ». Il est en réalité constitué de plusieurs ondes vibratoires dont chacune possède une fréquence propre, l'onde la plus grave représentant la hauteur ou fréquence fondamentale (notée f0) du son. La forme des cordes vocales donne déjà un *timbre**particulier au son.

Notons par ailleurs que le langage est aussi basé sur des sons qui ne sont pas produits avec le concours des cordes vocales. Ces sons non voisés (consonnes non voisées ou sourdes) se caractérisent également par certaines fréquences particulières. Toutefois, il ne s'agit pas de sons à proprement parler (qui se caractérisent par une onde sonore périodique), mais bien de bruits (cf. Le cours de phonétique, phonologie et morphophonologie).

L'articulation est la transformation en sons distincts – dit *articulés* – de la source sonore produite par les cordes vocales ou par le passage de l'air. Cette transformation du son se fait par le passage dans les *résonateurs* (le pharynx, la cavité buccale et les cavités nasales) dont la forme résulte de la position prise par les *articulateurs* (*les lèvres, les dents, la langue et le voile du palais – qui ouvre ou ferme le passage de l'air vers les fosses nasales*) [toutes ces notions et leur fonctionnement seront exposés en détails dans le cours de Phonétique, phonologie et morphophonologie].

8.3 Des sons aux systèmes phonologiques

8.3.1 L'organisation des systèmes phonologiques

Nous venons de décrire les sons, ou phones, qui pouvaient être produits par l'appareil phonatoire humain. Les locuteurs des langues du monde n'utilisent pas tous la totalité de ces sons, mais seulement un sous-ensemble des sons possibles. En effet, les communautés linguistiques peuvent naviguer entre deux « stratégies » :

- celle qui consiste à n'exploiter que peu de contrastes phonétiques et à n'utiliser que peu de phonèmes, les plus différenciés les uns des autres, ce qui conduit nécessairement à utiliser plus souvent les mêmes phonèmes, mais demande moins de précision dans l'articulation ;
- celle qui consiste à enrichir le système phonologique en utilisant davantage de phonèmes, ce qui implique une précision articulatoire accrue, mais permet de disposer d'un inventaire phonologique plus large pour créer les morphèmes.

Les locuteurs de la plupart des langues utilisent, au minimum, des ensembles de phones qui connaissent des différences importantes ([a, i, u] pour les voyelles ; [k, p, m] pour les consonnes), et enrichissent leur système en ajoutant quelques phonèmes dont les contrastes sont

plus fins (cf. chapitre 1 de ce cours).

La possibilité de marquer des unités distinctives à l'aide de sons relativement proches (ex. [s] – [ʃ]) suppose que les locuteurs veillent à produire l'effort articulatoire nécessaire pour maintenir ces contrastes. Même si on ne prononce jamais un mot exactement deux fois de la même manière, le maintien de la valeur distinctive de deux phonèmes suppose que les locuteurs les prononcent suffisamment distinctement pour qu'ils soient identifiables et renvoient bel et bien à des morphèmes différents, sans quoi l'opposition phonologique devient non productive et progressivement caduque : c'est à cette condition que les mots *seau* vs *chaud*, *sauce* vs *chose*, *bus* vs *buche*, etc. peuvent être distingués.

Pour qu'elles soient suffisamment et systématiquement différentes et qu'elles puissent s'opposer les unes aux autres, les unités du système phonologique doivent présenter au minimum un contraste notable. Comme pour les unités de signification, nous pouvons différencier les phonèmes à l'intérieur du système par une série de **traits distinctifs**, qui correspondent dans ce cas à leurs caractéristiques articulatoires, et qui permettent de les situer les uns par rapport aux autres. Dans un système avec peu de phonèmes, ceux-ci se différencieront pour plusieurs traits distinctifs :

- /a/ [+ouvert, +antérieur, -arrondi] où /a/ et /i/ ne se distinguent que par un trait, mais présentent néanmoins un contraste de trois degrés d'aperture
- /u/ [- ouvert, - antérieur, + arrondi]
- /i/ [-ouvert, + antérieur, - arrondi]

Dans un système avec de nombreux phonèmes, on pourra trouver des oppositions très contrastées, mais aussi des séries de phonèmes qui présentent les mêmes traits distinctifs et ne se distinguent que par un trait particulier (voir les deux séries de voyelles moyennes en français).

Lorsque dans une langue donnée, une caractéristique articulatoire (ex. le voisement des consonnes, ou le caractère nasal des voyelles) permet d'opposer des phonèmes, on dira que ce trait articulatoire est distinctif dans la langue donnée (en français par exemple, le trait de nasalité et le trait de voisement sont distinctifs).

8.3.2 Les règles d'agencement des phonèmes : la syllabe

À partir de l'inventaire phonologique d'une langue donnée, toutes les combinaisons sont en principe possibles pour créer des unités de signification. Or, une analyse de la distribution des différents phonèmes d'une langue à l'intérieur des morphèmes montre que les combinaisons respectent en réalité des principes tantôt universels, tantôt propres à chaque langue. L'analyse phonologique de différentes langues permet de montrer que :

- les phonèmes s'organisent en syllabes,
- ces syllabes ont une structure de base abstraite comprenant trois places (attaque, noyau, coda),
- seule la place du noyau doit (en principe) nécessairement être instanciée dans toute syllabe,
- le noyau doit être nécessairement occupé par un phonème vocalique, l'attaque et la coda par une ou plusieurs consonnes
- certaines structures syllabiques concrètes sont privilégiées (CV),
- que les attaques de syllabes peuvent être complexes, mais qu'elles tendent à respecter certains principes (échelle de sonorité)

Ces principes valent pour toutes les langues. En revanche, il existe des règles propres à chaque langue : par exemple, certaines suites consonantiques ne peuvent se trouver en attaque de

syllabe en français (ex. [mp]), une syllabe dont le noyau est une voyelle nasale n'est jamais fermée par une coda nasale en français standard.

9 REFERENCES BIBLIOGRAPHIQUES

- Austin, J.-L. (1970). *Quand dire, c'est faire*, Paris, seuil.
- Baratin, M., Desbordes F. (1981). *L'analyse linguistique dans l'Antiquité classique*, Paris, Klincksieck.
- Baylon, C. & Fabre, P. (1975). *Initiation à la linguistique*, Editions Fernand Nathan.
- Béguelin, M. -J. 2000. *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. Bruxelles : Duculot-De Boeck.
- Benveniste, E. 1966. *Communication animale et langage humain*, In *Problèmes de linguistique générale* 1, 56-74.
- Bloomfield, L. (1933). *Language*,
- Chomsky, N. (1969). *Structures syntaxiques*, trad.de l'américain, Paris, Le seuil.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*, MIT Press.
- Delbecq, Nicole. 2006. *Linguistique cognitive. Comprendre comment fonctionne le langage*. Bruxelles : De Boeck-Duculot (2e édition augmentée).
- Dik, S. (1989). *The theory of functional grammar. Part 1 : The structure of the clause*. Dordrecht : Foris
- Durand, J. 2009. *L'alphabet phonétique international*, in C. Herren schmidt, R. Mugnaioni, M.-J. Savelli, C. Touratier (éds.) *Le monde des écritures*. Paris : Gallimard.
- Fages, J.-B. (1968). *Comprendre le structuralisme*, Toulouse, Edouard Privat.
- Fauconnier, G. (1984). *Espaces mentaux*, Paris, Editions Minuit.
- François J. & Cornish F. (1995). *Le modèle néerlandais de grammaire fonctionnelle : Esquisse d'un mode d'emploi*. In: *L'Information Grammaticale*, N. 67. pp. 12-20
- Garric, N. 2001. *Introduction à la linguistique*. Paris : Hachette.
- Klinkenberg, Jean-Marie. 1996/2006. *Précis de sémiotique générale*. Paris : Le Seuil
- Labov, W. (1976). *Sociolinguistique*, Paris, Minuit.
- Lakoff, G. (1987). *Women, fire, and other dangerous things*, University of Chicago Press.
- Langacker, R. (1987). *Foundations of the cognitive grammar*, Stanford University Press.
- Mangum, A.W. 2010. *Animal communication. The 'language' of honey bees*. In S. Behrens & J. Parker (eds). *Language in the real world. An introduction to linguistics*. Oxon: Routledge, 255-273.
- Martinet, André. 1970. *La double articulation*. In *Eléments de linguistique générale*. Paris : Armand Colin, 13-15.
- Moeschler, Jacques & Antoine Auchlin. 1997/2000. *Introduction à la linguistique contemporaine*. Paris : Armand Colin.
- Mounin, G. (sous dir.) (1993). *Dictionnaire de linguistique*, Paris, PUF.
- Paveau, M.-A. & Sarfati, G.-E. (2003). *Les grandes théories de la linguistique. De la grammaire comparée à la pragmatique*, Paris, Armand Colin.
- Pottier, Bernard. 2000. *Représentations mentales et catégorisations linguistiques*. Bruxelles: De Boeck-Duculot
- Robins, R.-H. (1976). *Brève histoire de la linguistique. De Platon à Chomsky (1967)*, trad. Fr. de M. Borel, Paris, Seuil.
- Ricken, U. (1978). *Grammaire et philosophie au siècle des Lumières*, Lille, Presses universitaires de Lille.
- Rosier I. (1983). *La grammaire spéculative des modistes*, Lille, Presses universitaires de Lille.

- Sapir, E. (1939). *Nootka Texts: tales and Ethnological Narratives with Grammatical Notes and Lexical Materials*: William Dwight Whitney Linguistics Series. Philadelphia : Linguistic Society of America.
- Schaeffer, J.-M.. 1995. *Nouveau dictionnaire encyclopédique des sciences du langage* ; pp. 213-223)
- Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Publié par Charles Bally et Albert Schehaye. Editions critique préparée par Tullio di Mauro. Paris : Payot.
- Siouffi, G. & Van Raemdonck D. (1999), *100 fiches pour comprendre la linguistique*. Rosny: Bréal.
- Simon, A.C. & Bachy, S. 2009. *Parole & Langue. Parcours et exercices de linguistique*. DVD-ROM, Louvain-la-Neuve, Presses universitaires de Louvain.
- Soutet, O (1996). *Linguistique*. Paris : Presses universitaires de France.
- Soutet, O (1989). *La syntaxe du français*, Paris, PUF, Que-sais-je ?
- Talmy, L. (1988). *The relation of grammar to cognition*. In Brigid a Rudzka-Ostyn (ed), *Topics in Cognitive linguistics*, 165-205. Amsterdam: John Benjamins.
- Tellier, C. 2003. *Éléments de syntaxe du français*. Québec : Gaëtan Morin.
- Troubezko, N.S. (1939). *Grundzüge der Phonologie*, Vol. Traduction française (J. Cantineau): *Principe de Phonologie*, Paris, Klincksieck, 1947.

