



UFR HSS
HUMANITÉ
& SCIENCES SOCIALES

UNIVERSITÉ
CAEN
NORMANDIE

TYPOLOGIE DES LANGUES

COURS

RICHARD RENAULT

Maître de conférences



PLAN DU COURS

1. INTRODUCTION.....	3
2. LA CLASSIFICATION DES LANGUES.....	4
2.1. LA CLASSIFICATION GÉOGRAPHIQUE.....	4
2.2. LA CLASSIFICATION GÉNÉTIQUE.....	5
2.3. LA CLASSIFICATION ARÉALE.....	8
2.4. CLASSEMENT DES LANGUES SELON LE SYSTÈME D'ÉCRITURE.....	9
2.5. CORRÉLATIONS ENTRE LES DIFFÉRENTES CLASSIFICATIONS.....	10
3. LES DIFFÉRENTES APPROCHES EN TYPOLOGIE DES LANGUES.....	17
4. LA CLASSIFICATION MONOTHÉTIQUE.....	18
4.1. LES CRITÈRES.....	18
4.1.1. Classification des langues au moyen d'un critère phonétique.....	18
4.1.2. Classification des langues au moyen d'un critère phonologique.....	19
4.1.3. Classification des langues au moyen d'un critère morphologique.....	19
4.1.4. Classification des langues au moyen d'un critère syntaxique.....	20
4.2. LA TYPOLOGIE MORPHOLOGIQUE.....	20
4.2.1. Le degré de synthèse.....	21
4.2.1.1. Type analytique.....	22
4.2.1.2. Type synthétique.....	23
4.2.1.3. Type polysynthétique.....	24
4.2.2. Types morphologiques.....	25
4.2.2.1. Type isolant.....	25
4.2.2.2. Type agglutinant.....	26
4.2.2.3. Type fusionnant.....	26
4.2.2.4. Type incorporant.....	28
4.3. LANGUES NOMINATIVES ET LANGUES ERGATIVES.....	30
4.3.1. L'identification des fonctions.....	30
4.3.1.1. L'identification par le cas morphologique.....	31
4.3.1.2. L'identification par l'accord du verbe.....	32
4.3.1.3. L'identification par la place des arguments.....	34
4.3.2. Prototype nominatif et prototype ergatif.....	34
4.3.3. Statut du sujet des phrases intransitives.....	36
4.3.4. Systèmes hybrides et type binaire.....	40
5. TYPOLOGIE POLYTHÉTIQUE.....	44
5.1. TYPOLOGIE MORPHOLOGIQUE DE SAPIR.....	44

1. INTRODUCTION

La typologie linguistique a pour but la description et la classification des langues. En linguistique, le mot typologie est utilisé avec deux significations : on parle de typologie d'un fait linguistique, lorsqu'on classe un ensemble d'objets au regard d'une propriété différentielle. Chaque classe définit ainsi un type d'objets. Par exemple, en morphologie, on obtient une typologie des mots en distinguant plusieurs classes à partir de la nature des morphèmes impliqués dans le mot. On définit ainsi des mots simples (un seul morphème), des mots composés (plusieurs morphèmes libres) des mots dérivés (un morphème lexical avec un ou plusieurs affixes) etc. Toute typologie repose donc sur une répartition d'objets en fonction d'une propriété partagée, et chaque classe d'objets définit un type. Lorsqu'on parle de typologie des langues, il s'agit cette fois de la classification des langues du monde en fonction d'une propriété structurale partagée par un ensemble de langues.

En s'intéressant aux différences entre les langues, la typologie relève également de la linguistique générale dans la mesure où elle s'appuie sur des propriétés communes aux langues (une propriété n'est pertinente pour la typologie que si elle est partagée par plusieurs langues) et dans la mesure où elle vise également à dégager des invariants dans les langues (une propriété différentielle qui se distribue sur l'ensemble des langues peut être l'expression d'une propriété plus générale présente dans toutes les langues). Un premier exemple simple : les langues se répartissent selon qu'elles ont des prépositions ou des postpositions. En français, nous avons des prépositions (*à la maison*) alors que le japonais a des postpositions (*uti de = à la maison*). Sur le plan général, cette différence d'organisation des constituants syntaxiques, donne lieu à trois observations : 1) toutes les langues ont des adpositions dans leur système des catégories syntaxiques (une adposition désigne une catégorie syntaxique qui recouvre les prépositions et les postpositions). 2) La différence de place respective de l'adposition et de son complément dans une langue donnée est peut être une propriété plus générale concernant la relation de dépendance entre deux constituants ; selon les langues, le terme dépendant se place avant ou après le terme dont il dépend. On s'attend alors à retrouver cette différence de place avec d'autres catégories de la langue qui ont également des propriétés relationnelles. 3) Les deux remarques précédentes montrent que les langues manifestent des propriétés de symétrie ; ce qui dans une langue est placé à gauche d'un constituant, peut être à droite dans une autre langue.

Les propriétés différentielles utilisées pour la classification des langues concernent tous les niveaux de la description linguistique ; elles peuvent être de nature phonétique, phonologique, morphologique, syntaxique ou lexicale. Le plus souvent, les propriétés retenues sont de nature oppositive et binaire (les langues sont alors de deux types ; avoir ou non telle propriété. Quelques exemples simples, les langues peuvent être monosyllabiques ou non, les langues peuvent être tonales ou non, les langues peuvent avoir ou ne pas avoir de genre grammatical, les langues peuvent avoir ou ne pas avoir de système casuel, les langues ont ou n'ont pas d'accord du sujet avec le verbe... Comme on le verra par la suite lorsqu'il s'agira de

typologie morphologique et syntaxique, la classification des langues n'est pas exclusivement de nature binaire ; la classification fondée sur des critères d'ordre morphologique comporte trois ou quatre types selon les descriptions et selon le point de vue adopté. D'une part, on fait la distinction entre des langues analytiques, synthétiques et polysynthétiques et d'autre part on fait la distinction entre des langues isolantes, agglutinantes, fusionnantes ou incorporantes.

2. LA CLASSIFICATION DES LANGUES

La classification des langues à partir d'une propriété structurale n'est pas la seule possible. Les langues du monde se distribuent également en fonction 1) de leur répartition géographique, 2) de leur relation de parenté (classification génétique), 3) de propriétés partagées entre langues géographiquement contiguës (linguistique aréale).

2.1. LA CLASSIFICATION GÉOGRAPHIQUE

Breton [1976] nous invite à bien faire la distinction entre la géographie linguistique et la géographie des langues. Seule la première concerne directement la linguistique car il s'agit de représenter et d'étudier la distribution d'une propriété qui relève de la structure interne des langues. La seconde discipline intéresse également le linguiste mais elle est avant tout le fait du géographe qui étudie pour sa part la distribution d'une propriété de l'humain : parler telle ou telle langue. À ce titre, il prend en compte la dynamique des langues, l'ethnolinguistique, les politiques linguistiques, les flux migratoires et évidemment l'origine des langues qui est du ressort du linguiste (linguistique génétique). La répartition géographique des langues (langues d'Europe, langues d'Afrique, d'Asie, du Caucase...) au delà de la géographie des langues n'apporte que peu d'éléments à l'étude des différences entre langues car un fait de structure dans une langue ne pourra pas s'expliquer par la localisation géographique de ses locuteurs ni par leur identité ethnique. Par ailleurs, les grands ensembles géographiques de langues ne constituent que rarement des groupes de langues homogènes avec des propriétés structurales partagées qui les distinguent les uns des autres. Cela étant, on associe souvent un ensemble géographique de langues à une ou plusieurs propriétés saillantes dans ces langues. Mais dans ce domaine, il convient d'être prudent quant à la portée des observations car les contre-exemples sont systématiques. Quelques exemples ; 1) on associe les langues d'Asie avec les langues à tons (voir plus loin), mais cette caractéristique ne concerne qu'une partie seulement des langues d'Asie (le khmer ou cambodgien n'a pas de tons). De plus, la variation tonale est attestée ailleurs qu'en Asie, en suédois notamment, même s'il s'agit dans cette langue d'un phénomène très limité. 2) On associe les langues d'Afrique avec les langues à registres (voir plus loin) et avec les langues à classes nominales (voir plus loin). Mais ces deux propriétés ne sont pas présentes dans toutes les langues africaines ; le swahili (Kenya, Tanzanie, République du Congo, République démocratique du Congo...) n'a pas de registres et le yoruba

(Nigéria) n'a pas de classes nominales. De plus, les classes nominales des langues africaines ne sont qu'une expression particulière de la catégorie du genre grammatical que l'on trouve aussi dans des langues d'Europe et dans les langues du Caucase. Quant à ces dernières, elles ne forment pas un ensemble homogène sur le plan génétique puisqu'on y trouve des langues kartvéliennes (géorgien, laze...), des langues abkhaz-adyghéennes (abkhaz, kabardien...), des langues nakho-dagestaniennes (avar, tchéchéne...), des langues altaïques (azeri, koumik...) et des langues indo-européennes (arménien, ossète...). Étant donné leur origine différente, les langues du Caucase constituent un bel échantillon de la diversité linguistique sur un territoire relativement réduit. Le cas des langues du Caucase est à mettre en contraste avec celui des langues aborigènes d'Australie qui forment un ensemble très cohérent à la fois sur le plan génétique (toutes les langues aborigènes d'Australie se répartissent en deux familles dont on suppose de surcroît qu'elles ont un passé commun) et sur le plan typologique (beaucoup de similitudes, phonétiques, morphologiques et syntaxiques). Bien entendu, ces deux cas de figure opposés que représentent les langues aborigènes d'Australie et les langues du Caucase sont à mettre en rapport avec les particularités géographiques et humaines correspondantes ; insularité d'un vaste territoire d'un côté et région montagneuse au carrefour de l'Europe et de l'Asie de l'autre côté.

Plus intéressante est l'approche selon laquelle un ensemble géographique de langues présente un faisceau de propriétés partagées avec cependant des écarts importants pour certaines de ces propriétés. Néanmoins ce faisceau de propriétés contribue à une caractérisation globale du groupe de langues dès lors que les écarts sont relativisés par une approche quantitative.

2.2. LA CLASSIFICATION GÉNÉTIQUE

La classification des langues la plus connue et la plus souvent utilisée pour identifier une langue est la classification génétique. Le français est une langue romane et une langue indo-européenne ; l'anglais est une langue germanique et indo-européenne ; le finnois est une langue fennique, finno-ougrienne et ouralienne. Les divers regroupements font référence à un passé commun (langue mère), restitué à partir de régularités phonétiques qui ont fait le succès de la méthode comparative. Plus la période commune restituée est proche et plus les régularités phonétiques sont nombreuses et probantes. Chronologiquement, les divers niveaux de regroupement sont : la branche, le groupe, la classe, et la famille. Le tableau suivant met en parallèle les différents regroupements auxquels appartiennent le français et le finnois. Pour les deux filiations, chaque regroupement est donné avec deux autres langues du même niveau (voir à ce propos l'[arbre des langues indo-européennes](#) (document 1) et l'[arbre des langues ouraliennes](#) (lien 1)) :

FAMILLE	indo-européen + <i>farsi</i> + <i>hindi</i> + ...	ouralien + <i>nganasan</i> + <i>nenets</i> + ...
CLASSE	Langues "centum" + <i>allemand</i> + <i>gallois</i> + ...	finno-ougrien + <i>hongrois</i> + <i>mansi (vogoul)</i> + ...
GROUPE	italique + * <i>ombrien</i> + * <i>osque</i> + ...	finno-permien + <i>oudmour</i> + <i>mordve</i> + ...
BRANCHE	romane + <i>italien</i> + <i>roumain</i> + ...	fennique + <i>estonien</i> + <i>vepse</i> + ...
LANGUE	<i>français</i>	<i>finnois</i>

Le document 2 est une version graphique du tableau ci-dessus.

La méthode comparative repose sur la régularité des correspondances phonétiques dans la comparaison des langues et non sur la simple ressemblance des mots.

LANGUES ROMANES				
français	italien	espagnol	portugais	roumain
<i>lait</i>	<i>latte</i>	<i>leche</i>	<i>leite</i>	<i>lapte</i>
<i>nuit</i>	<i>notte</i>	<i>noche</i>	<i>noite</i>	<i>noapte</i>

Il importe peu que les mots *leche* de l'espagnol, *latte* de l'italien et *lapte* du roumain se ressemblent. Par contre, ce qui est pertinent pour établir une relation de parenté entre ces trois langues, c'est ici la correspondance entre la consonne "ch" de l'espagnol, "tt" de l'italien et "pt" du roumain ; correspondance que l'on retrouve respectivement dans les mots *noche*, *notte* et *noapte (nuit)* et dans *hecho*, *fatto* et *fapt (fait)*.

De même pour le rapprochement entre les mots *kieli* du finnois et *keel* de l'estonien (*langue*).

LANGUES FENNIQUES					
finnois	estonien	carélien	vepse	live	
<i>kieli</i>	<i>keel</i>	<i>kel'</i>	<i>kel'</i>		<i>langue</i>
<i>silmä</i>	<i>silm</i>	<i>silmä</i>	<i>silm</i>	<i>si:lma</i>	<i>œil</i>

Ici, on peut observer deux correspondances qui se retrouvent dans de nombreux mots : 1) à la diphtongue du finnois correspond une voyelle longue en estonien, 2) la voyelle finale du finnois fait l'objet d'une apocope en estonien. Autres exemples illustrant cette régularité phonétique :

	FINNOIS	ESTONIEN	
1) APOCOPE	<i>kaksi</i>	<i>kaks</i>	<i>deux</i>
	<i>laulu</i>	<i>laul</i>	<i>chant</i>
	<i>jumala</i>	<i>jumal</i>	<i>Dieu</i>
	<i>lintu</i>	<i>lind</i>	<i>oiseau</i>
	<i>talvi</i>	<i>talv</i>	<i>hiver</i>
2) DIPHTONGUE VS LONGUE	<i>juo</i>	<i>joo</i>	<i>il/elle boit</i>
	<i>tie</i>	<i>tee</i>	<i>chemin</i>
	<i>syön</i>	<i>söön</i>	<i>je mange</i>
	<i>vieras</i>	<i>võõras</i>	<i>étranger</i>
3) APOCOPE ET DIPHTONGUE VS LONGUE	<i>vielä</i>	<i>veel</i>	<i>encore</i>
	<i>suoni</i>	<i>soon</i>	<i>tendon</i>
	<i>kieli</i>	<i>keel</i>	<i>langue</i>

Exemples montrant cette fois la simple ressemblance entre mots de la même famille :

LANGUES INDO-EUROPÉENNES							
français	anglais	allemand	breton	russe	sanscrit	farsi	hindi
<i>frère</i>	<i>brother</i>	<i>Bruder</i>	<i>breur</i>	<i>brat'</i>	<i>bhratar</i>	<i>baradar</i>	<i>bhrata</i>

LANGUES OURALIENNES						
finnois	estonien	mordve	same	khanti	mansi	hongrois
<i>kala</i> (<i>poisson</i>)	<i>kala</i>	<i>kal</i>	<i>guolle</i>	<i>Xul</i>	<i>Xuül</i>	<i>hal</i>

La parenté des langues indo-européennes est fondée notamment sur la fameuse loi de Grimm qui fait correspondre à une forme restituée de l'indo-européen "*bh", "b" en germanique, "φ" en grec et "F" en latin :

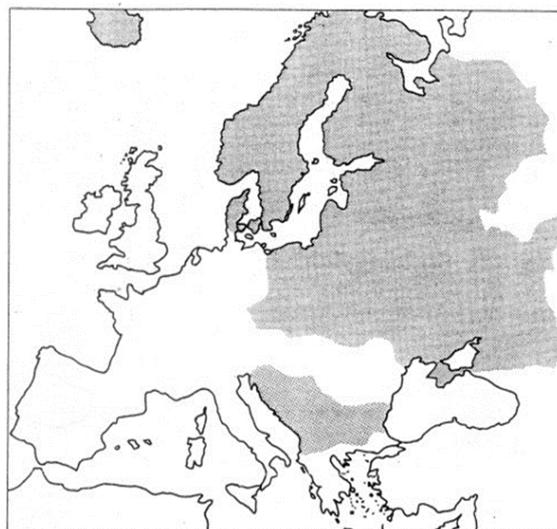
- frère : sanscrit = *bhratar-*
- germanique = *Bruder*
- grec = *φέρω*
- latin = *frater*

(Voir à ce propos le commentaire sur la méthode comparative de J.-C. Milner dans *Introduction à une science du langage* [1989], pp 93-99)

Pour les langues ouraliennes, on notera la correspondance entre la consonne occlusive sourde à l'initiale en finnois, estonien et mordve qui renvoie à la même consonne sonore en same, et à une fricative de même point d'articulation dans les langues ougriennes : hongrois, kanti et mansi.

2.3. LA CLASSIFICATION ARÉALE

Des langues d'origine différente mais géographiquement contiguës peuvent présenter des propriétés partagées du fait du contact des langues (emprunts ou convergences). Ce type de regroupement de langues qui prend en compte à la fois la distribution géographique et un fait de structure qui relève de la typologie appartient à la linguistique des aires. Un exemple bien connu en Europe est la distribution de la palatalisation (appelée aussi mouillure) des consonnes. Les langues slaves se caractérisent sur le plan phonétique par une série de consonnes avec comme articulation secondaire une palatalisation. On oppose ainsi en russe des consonnes "dures", sans palatalisation, et des consonnes "molles", avec palatalisation (exemples : *brat* (*frère*) et *brat'* (*prendre*)). L'opposition de mouillure se retrouve dans d'autres langues qui n'appartiennent pas au domaine slave mais qui sont à la périphérie géographique du russe : le roumain pour les langues romanes, l'estonien pour les langues fenniques, le lituanien pour les langues baltes, et d'autres encore selon R. Jakobson. Autre exemple, plusieurs langues d'Europe sans affinité génétique ont un possessif réfléchi. La carte suivante est empruntée à Bechert, G. Bernini, & C. Buridant (ed.) [1990].



Distribution du possessif réfléchi

Pour comprendre de quoi il s'agit, considérons tout d'abord le cas du français dans la phrase *Marie a vendu sa maison*. Le déterminant possessif *sa* peut renvoyer aussi bien à *Marie*, le sujet de la phrase, qu'à quelqu'un d'autre. C'est là une propriété générale du déterminant possessif de troisième personne. En letton, langue balte, en danois, langue germanique, et en finnois, langue fennique, l'ambiguïté précédente

(que l'on a également en anglais dans *Mary sold her house*) n'existe pas dans ces langues car il y a deux formes du déterminant possessif ; une forme réfléchie (qui renvoie au sujet de la phrase) et une forme non réfléchie qui renvoie à quelqu'un d'autre. Exemples :

LETTON	<i>tēvs redzēja savu māju</i> père a vu sa maison <i>Père a vu sa maison</i>	(forme réfléchie)
	<i>tēvs redzēja viņa māju</i> père a vu sa maison <i>Père a vu sa maison</i>	(forme non réfléchie)
DANOIS	<i>Søren har solgt sit hus</i> Søren a vendu sa maison <i>Søren a vendu sa maison</i>	(forme réfléchie)
	<i>Søren har solgt hans hus</i> Søren a vendu sa maison <i>Søren a vendu sa maison</i>	(forme non réfléchie)
FINNOIS	<i>Jussi myi talonsa</i> Jussi a vendu maison+3SG <i>Jussi a vendu sa maison</i>	(forme réfléchie)
	<i>Jussi myi hänen talonsa</i> Jussi a vendu de lui maison+3SG <i>Jussi a vendu sa maison</i>	(forme non réfléchie)

La comparaison entre le finnois et les deux autres langues montre que la propriété partagée par ce groupe de langues est plus d'ordre fonctionnel que structural ; les trois langues disposent d'un moyen pour distinguer la coréférence de la non coréférence avec le sujet (distinction entre la forme réfléchie et la forme non réfléchie), mais le moyen utilisé n'est pas de même nature. En letton et en danois, il s'agit de deux formes différentes du déterminant possessif, alors qu'en finnois la différence se fait par l'opposition entre présence et absence du pronom génitif.

2.4. CLASSEMENT DES LANGUES SELON LE SYSTÈME D'ÉCRITURE

Pour terminer cette présentation des différentes classifications, je mentionnerai également la classification des langues selon leur système d'écriture, pour les langues qui en possède une, évidemment. Pour l'essentiel, il y a trois systèmes d'écriture qui se subdivisent en deux groupes selon que l'écriture représente les sons (le signifiant) ou ce à quoi ils sont associés dans la langue (le signifié) : 1) les systèmes alphabétiques où les symboles écrits représentent des sons ou des phonèmes (écriture romane pour le français, le polonais, le finnois, le vietnamien..., écriture cyrillique pour le russe le bulgare, l'ouzbek..., écriture arabe pour l'arabe, le farsi ou persan, l'ourdou..., écriture arménienne, géorgienne...), 2) les systèmes syllabiques où les symboles correspondent à des syllabes (hiragana et katakana du japonais, syllabaires

coréen, cherokee, cree...), et 3) les systèmes idéographiques où les symboles représentent le signifié des morphèmes (chinois, kanji du japonais...). Les langues étant avant tout un moyen de communication oral, l'examen des systèmes d'écriture dans le cadre de la linguistique n'est pertinent que s'il existe des corrélations intéressantes entre écriture et structure de la langue. Il est évident que l'écriture idéographique est adaptée au chinois car les unités de signification se présentent sous la forme de mots monosyllabiques sans variation morphologique (voir à ce propos la partie consacrée à la typologie morphologique). Par contre, ce système d'écriture est inadapté pour les langues dans lesquelles les unités lexicales peuvent avoir des formes différentes ou des affixes, ce qui explique pourquoi le japonais – langue morphologiquement très différente du chinois - a un second système d'écriture (syllabique) pour compléter les kanjis d'origine chinoise. Autre exemple, l'écriture arabe ne note pas les voyelles brèves mais seulement les voyelles longues, et de manière diacritique. Ce qui correspond à un fait de structure des langues sémitiques ; les voyelles à valeur distinctive ne sont que trois (/i/, /a/, /u/) et la variation vocalique à l'intérieur des mots sert avant tout à la flexion des mots alors que la structure consonantique est le support de l'identification lexicale des mots. Si l'écriture arabe est ainsi en rapport avec la nature infixionnelle de la langue, elle est inadaptée pour une langue génétiquement et typologiquement différente comme le turc dans laquelle les voyelles sont plus nombreuses et assurent une fonction distinctive sur le plan lexical. C'est pourquoi, le turc et l'ouzbek (langues altaïques) ont abandonné l'écriture arabe au début du XX^e siècle au profit d'un système alphabétique. Du fait qu'il s'agit d'un système d'écriture relativement récent, le turc est ainsi aujourd'hui une des langues dont le système alphabétique est très proche de son système phonétique (une lettre correspond à un son et un son correspond à une lettre). Les autres langues non sémitiques avec écriture arabe tel le persan et l'ourdou (langues indo-européennes) ont dû aménager le système originel en ajoutant des signes. Autre exemple d'adaptation, le vietnamien : cette langue est de type monosyllabique et tonale comme le chinois (voir plus loin pour un développement de cette caractérisation typologique) mais le système alphabétique adopté au cours des XVII^e et XVIII^e siècles a dû être complété par tout un jeu de signes diacritiques suscrits et souscrits pour rendre compte des différences tonales (exemples de voyelles écrites vietnamiennes : a, á, à, ả, ạ ...). Un petit exposé systématique des systèmes d'écriture est proposé dans le *Dictionnaire encyclopédique des sciences du langage* [1972] d'O. Ducrot et T. Todorov.

2.5. CORRÉLATIONS ENTRE LES DIFFÉRENTES CLASSIFICATIONS

Mis à part la classification aréale qui repose sur la distribution géographique d'une propriété typologique due à un emprunt ou à un phénomène de convergence, il n'y a pas de liens nécessaires entre les différentes classifications des langues. Tout au plus, peut-on dégager des tendances ou des singularités. La distribution géographique des langues dans le monde ne coïncide pas avec une répartition équivalente sur le plan génétique et typologique. En Europe, par exemple, on trouve aussi bien des langues indo-européennes – la majorité – (français, anglais, suédois, russe, letton, irlandais...) que des langues finno-ougriennes (finnois,

hongrois, estonien...). De même sur le plan typologique, il y a aussi bien des langues avec genre grammatical (français, italien, allemand, russe...) que des langues sans genre grammatical (anglais, basque, finnois...). On observe cependant des tendances ; les langues d'Europe sont surtout d'origine indo-européenne (le finnois, le hongrois, le basque font notamment exception), et elles sont, à une exception près (le basque), des langues nominatives. Les mêmes observations valent pour les autres grandes répartitions géographiques. L'Australie, avec ses langues aborigènes, comme on l'a vu précédemment, apparaît comme une singularité ; à un territoire donné correspond un ensemble relativement homogène de langues sur le plan génétique et typologique, si bien que le terme même de langues aborigènes d'Australie désigne à la fois une distribution géographique et une famille de langues, voire même un type de langues si l'on prend en compte tout un faisceau de propriétés communes à ces langues. Inversement, on entend parfois parler à tort de langues d'Afrique comme un ensemble qui serait homogène alors qu'il existe une très grande diversité génétique et typologique.

Si on laisse de côté le cas particulier des langues indo-européennes dont la diffusion dans le monde est liée à la colonisation, les familles de langues et les divers regroupements génétiques (cette fois on adopte le point de vue inverse) ont une distribution géographique relativement homogène dont l'étendue territoriale, qu'elle soit continue ou qu'elle résulte d'un essaimage, s'explique par le passé commun des langues de la famille et par les grands mouvements de populations. Voir à ce propos la carte des [grandes familles de langues](#) (document 3) ainsi que la [page personnelle de Merritt Ruhlen](#) (lien 2) qui contient une carte des 12 super-familles proposées par J.H. Greenberg. Une carte telle que celle publiée dans l'*Atlas des langues* de Bernard Comrie, Stephen Matthews et Maria Polinsky (dir.) en 2004, est plutôt une carte de géographie des langues puisqu'elle prend en compte les langues indo-européennes parlées en Amérique (langues ayant un statut officiel dans les divers pays de ce continent). Notons que les langues indo-européennes parlées en Afrique et ayant aussi un statut officiel ne sont pas systématiquement mentionnées sur la carte, sauf l'anglais et l'afrikaans (langue dérivée du néerlandais) de l'Afrique du sud. Certaines familles de langues des deux premières cartes ont une étendue plus vaste que la somme des territoires ayant des locuteurs natifs car la distribution géographique des langues de ces familles se caractérise plus par un essaimage que par une continuité territoriale. C'est particulièrement vrai des langues ouraliennes dont la plupart sont parlées par de petites communautés disséminées sur une zone qui doit être interprétée comme l'extension maximale de la famille, sans compter qu'elles sont parfois éparpillées sur des territoires où dominant une langue indo-européenne. Ce qui est vrai d'une famille, l'est également de langues particulières comme le rom ou romani – langue indo-européenne dérivée du sanskrit et parlée par la communauté tzigane – qui est dispersé aux quatre coins de l'Europe.

Alors que la quasi totalité des familles et super-familles présentent une continuité territoriale, la super-famille déné-caucasienne, proposée par Ruhlen, qui englobe notamment les langues sino-tibétaines d'Asie et les langues na-déné d'Amérique du nord présente un cas intéressant d'essaimage puisqu'une partie des langues se situe en Asie et l'autre en Amérique (on en trouve également en Europe et dans le Caucase). Cet

essaimage correspondrait à une des plus anciennes grandes migrations de populations (la seconde après le premier mouvement migratoire à partir de l'Afrique), d'Asie vers l'Amérique du nord par le détroit de Béring. L'hypothèse récente du regroupement des langues sino-tibétaines et des langues na-déné (initiée par Sapir au début du XX^e siècle) et défendue récemment par Ruhlen [1997] n'est toutefois pas acceptée par tous et fait l'objet de nombreuses controverses.

L'évaluation du rapport entre la classification génétique et la classification typologique doit tenir compte du fait que les langues sont plutôt conservatrices et que les changements structuraux ne s'observent qu'à long terme, contrairement aux variations et modifications lexicales qui se font beaucoup plus facilement. On peut ainsi observer que les groupes de langues, telles les langues romanes, les langues germaniques, les langues celtiques, ou les langues slaves, présentent de nombreuses similitudes dans leur structure interne. Notons cependant, que l'anglais et l'allemand, deux langues germaniques de la même branche (branche occidentale) se distinguent par deux propriétés pertinentes pour la classification typologique : l'allemand a un genre grammatical et un système de cas, alors que l'anglais n'a ni l'un ni l'autre. Pour être plus précis sur point, on nuancera en faisant intervenir un point de vue historique ; l'anglais n'a plus de genre ni de cas. Il en reste seulement des traces dans le système pronominal (*he/she/it* et *him/her/it*). Plus le niveau de regroupement est bas dans la hiérarchie de la classification génétique, et plus le faisceau de propriétés typologiques est conséquent. À l'autre bout de l'échelle – celui de la famille ou de la super-famille – il n'est pas rare de constater une différence significative entre une ou plusieurs langues et toutes les autres langues de la même famille. Hjelmlev [1966] mentionne le cas de l'arménien dont la déclinaison des noms en fonction de la catégorie du nombre et du cas, manifeste une structuration des morphèmes typique des langues agglutinantes alors que les langues indo-européennes auxquelles il appartient sont dans leur ensemble des langues plutôt fusionnantes (ces notions de typologie morphologique seront développées par la suite. On retiendra ici le fait qu'une langue d'une famille donnée peut présenter des éléments de structuration propres à une autre famille de langues.) Ce changement dans la forme des mots en arménien s'explique par la proximité géographique des langues altaïques qui sont typiquement des langues agglutinantes. De plus, la forme du pluriel en arménien (*hayer = hay+er* "les/des arméniens", *hay = "arménien"*) n'est pas sans rappeler celle du turc (*evler = ev+ler* "les/des maisons", *ev = "maison"*). Autre exemple : les langues finno-ougriennes sont connues pour être des langues agglutinantes. On observe cependant en estonien une évolution de la forme fléchie des mots qui tend vers la fusion ; on opposera ainsi l'exemple agglutinant du génitif finnois (*jalan = jala+n* "du pied", *jalka = "pied"*) avec ce même génitif en estonien (*jala* "du pied", *jalg = "pied"*) où cette fois il n'est pas possible de segmenter le mot en deux morphèmes. L'identification du génitif se fait alors à partir de la forme globale du mot. Le morphème lexical et le morphème de génitif fusionnent en une forme de mot qui n'est plus décomposable en segments distincts (la forme estonienne s'explique par la disparition de la consonne finale, marque du génitif dans les langues fenniques).

Compte-tenu de ces remarques, on observera que les regroupements génétiques ne sont pas nécessairement liés à un type donné, mais que des tendances plus ou moins fortes s'observent selon les familles. Les deux cas de figure extrêmes sont les langues indo-européennes et les langues altaïques. Dans les langues indo-européennes on trouve des langues qui tendent vers l'isolation (anglais, suédois...), des langues fusionnantes (français, letton...) ainsi qu'une langue agglutinante (arménien). À l'opposé, les langues altaïques sont toutes du type agglutinant.

Si l'on prend en considération maintenant non pas une propriété unique mais un ensemble de propriétés, alors il est possible d'obtenir une caractérisation plus satisfaisante des familles. Les langues ouraliennes sont des langues agglutinantes, nominatives, dont l'ordre neutre des constituants de la phrase est SOV (sujet-objet-verbe), et ces langues ont des postpositions. Ce type de caractérisation admet cependant des écarts significatifs. Dans l'exemple précédent, le finnois est une langue SVO et non SOV, il n'est pas pour autant un contre-exemple dans la mesure où il vérifie toutes les autres propriétés. Voir à ce propos, la caractérisation que donne Lazard (1990) des langues d'Europe. Il s'agit cette fois de la distribution d'un ensemble de propriétés sur une zone géographique.

Une caractérisation d'une famille de langues à partir d'un ensemble de propriétés structurales peut cependant échouer si ces propriétés sont interprétées comme suffisantes pour identifier une langue quelconque comme appartenant à cette famille.

Le **linguiste N. S. Trubetzkoy** (1939) avait proposé de caractériser les langues indo-européennes par les six propriétés suivantes :

1. "Il n'y a pas d'harmonie vocalique."
2. "Le consonantisme à l'initiale n'est pas plus pauvre que celui de l'intérieur ou de la finale."
3. "Le mot ne doit pas nécessairement commencer par la racine."
4. "Les formes ne sont pas constituées seulement par des affixes, mais aussi par des alternances vocaliques à l'intérieur des morphèmes radicaux."
5. Outre les alternances vocaliques, les alternances consonantiques libres jouent aussi un rôle morphologique."
6. "Le sujet d'un verbe transitif est traité comme le sujet d'un verbe intransitif."

Ces différentes propriétés sont de nature phonologique (1 et 2), morphologique (3, 4 et 5) et syntaxique (6).

Commentaire et illustration à partir des données du français (et de l'anglais) :

Propriété 1

L'harmonie vocalique est bien connue notamment dans langues ouraliennes et altaïques et dans certaines langues d'Afrique comme le yoruba ou l'igbo. Dans ces langues, la distribution des voyelles au sein du mot n'est pas libre ; la présence d'une voyelle donnée conditionne la présence des autres voyelles (un mot turc qui contient la voyelle "a" ne peut pas contenir la voyelle "e"). En outre, la suffixation est soumise à un processus d'ajustement du vocalisme du suffixe (les mots turcs *ev* "maison" et *kar* "neige" ont pour pluriel

respectif *evler* et *karlar*, *-ler/-lar* étant le suffixe de pluriel sous deux formes différentes selon la nature des voyelles du radical). En français, la distribution des voyelles dans le mot ne connaît aucune contrainte et l'affixation (préfixation ou suffixation) est insensible au vocalisme du radical.

Propriété 2

La gamme des consonnes qui peuvent être en début de mot est exactement la même que celle qui se rencontre dans le mot ou en fin de mot. La consonne /ŋ/, par exemple, se trouve en début de mot dans /ʝol/ *gnôle*, en fin de mot dans /viŋ/ *vigne* et en position intervocalique dans /oŋɔ̃/ *oignon*.

Propriété 3

Autrement dit, l'affixation, qu'elle soit de nature syntaxique (flexion) ou lexicale (dérivation) n'est pas uniquement une suffixation. En dérivation française, on a aussi bien des suffixes (*acceptable*, *vendeur*...) que des préfixes (*défaire*, *injuste*...). De même pour les clitiques ; on a aussi bien des enclitiques (*donne-m'en*, *prends-le*...) que des proclitiques (*le dire*, *je le lui donne*...).

Propriété 4

La flexion des mots ne se fait pas uniquement par affixation mais également par la variation vocalique du mot. Sur ce point, les exemples français ne sont pas probants. Ceux de l'anglais sont plus explicites ; le prétérit se forme soit par suffixation du morphème *-ed* (*arrived*, *stopped*...) soit par variation de la voyelle du radical verbal (*sing/sang*, *give/gave*...).

Propriété 5

De nouveau, les exemples anglais seront plus évidents pour illustrer cette propriété. La formation du prétérit peut également se faire par l'alternance de la consonne finale (*send/sent*, *make/made*...). En français, les exemples font défaut ; on a cependant l'opposition entre *savent* et *sachent*.

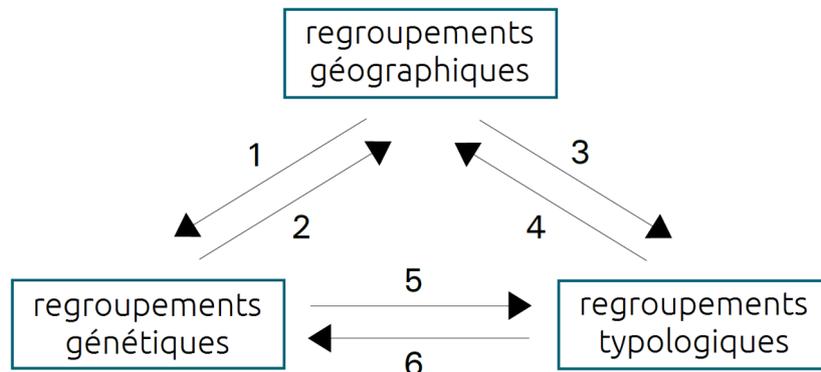
Propriété 6

Cette propriété est l'une des caractéristiques des langues nominatives. Une autre formulation de cette propriété serait par conséquent ; les langues indo-européennes sont des langues nominatives. Telle qu'elle est énoncée par Troubetzkoy, la propriété semble banale. En fait, l'identité de comportement entre le sujet d'une phrase transitive et le sujet d'une phrase intransitive ne va pas de soi. En français, comme dans beaucoup de langues, le sujet des phrases transitives et intransitives s'accorde avec le verbe en nombre et en personne, alors que dans les langues ergatives le sujet d'une phrase intransitive s'accorde avec le verbe dans les mêmes conditions que l'objet de la phrase transitive. Dans ce second type de langues, ce n'est pas le sujet de la phrase transitive qui s'accorde avec le verbe mais l'objet. Le statut des langues nominatives et des langues ergatives sera amplement développé plus loin.

Benveniste a montré que l'ensemble de ces propriétés étaient présentes en takelma, langue de la famille penutia, parlée dans l'Oregon (Voir à ce propos l'article d'E. Benveniste "La classification des langues" publié dans *Problèmes de linguistique générale I et II* (1974), tome I). Cette critique ne remet pas en cause la possibilité d'une caractérisation d'une famille au moyen d'un ensemble de propriétés structurales ; il suffit simplement de trouver le bon ensemble de propriétés. Par contre, il est tout à fait hasardeux de déduire l'appartenance à une famille de langues à partir d'un faisceau de propriétés typologiques. Au mieux, c'est un indice, et l'appartenance à une famille ou à un quelconque regroupement génétique ne peut être établi que par la méthode comparative.

Si l'on prend maintenant le rapport inverse entre la classification génétique et la classification typologique, les choses sont plus nettes ; les différents types de langues se rencontrent dans de nombreuses familles de langues ; on trouve des langues agglutinantes dans les langues ouraliennes (Europe : hongrois...), dans les langues altaïques (Asie : mongol...), dans les langues bantoues (Afrique : swahili...), dans les langues amérindiennes (Amérique du sud : quechua...), dans les langues austronésiennes (Pacifique : indonésien...) et dans des langues génétiquement isolées telles que le basque, le coréen et le japonais. Autre exemple, on trouve le genre grammatical sous des formes diverses (genres, classes, classificateurs) dans des langues indo-européennes (Europe : français, espagnol, gallois...) dans des langues nigéro-congolaises (Afrique : swahili, lingala, zoulou...), dans des langues nakho-dagestaniennes (Caucase : avar, lak, lesghien...), dans des langues sino-tibétaines (Asie : chinois, vietnamien, lao...). Les langues ergatives se rencontrent dans les langues kartvéliennes (Caucase : géorgien...), dans les langues sino-tibétaines (Asie : tibétain...), dans les langues amérindiennes (Amérique du sud : quiché...) et dans les langues aborigènes d'Australie (dyirbal...).

Les relations entre les différentes classifications des langues peuvent se résumer en six propositions. On s'en tiendra aux propositions essentielles, chacune d'elles étant illustrée par un exemple.



1.
 - a. Un regroupement géographique ne correspond pas à un regroupement génétique.
(Les langues d'Europe ne sont pas toutes des langues indo-européennes.)
 - b. Un regroupement géographique peut être génétiquement homogène.
(Les langues d'Australie sont génétiquement homogènes.)
2.
 - a. Un regroupement génétique n'est pas nécessairement lié à une distribution géographique.
(Les langues indo-européennes se distribuent sur plusieurs continents.)
 - b. Les grandes familles de langues ont une distribution géographique homogène.
(Les langues ouraliennes s'étendent d'Asie en Europe.)
3.
 - a. Un regroupement géographique ne correspond pas à un type linguistique.
(Les langues d'Afrique ne sont pas toutes des langues avec classes nominales.)
 - b. Un regroupement géographique peut présenter un faisceau de propriétés typologiques.
(Les langues d'Europe partagent plusieurs propriétés typologiques (G. Lazard, 1990).)
4. Un type de langues n'est pas lié à une distribution géographique homogène.
(Les langues agglutinantes sont présentes un peu partout dans le monde.)
5. Les regroupements génétiques sont relativement homogènes quant à leur type linguistique.
(Les langues ouraliennes sont de type agglutinant et sont du type SOV.)
6. Un type linguistique n'est pas spécifique à un regroupement génétique donné.
(Les langues ergatives se rencontrent dans plusieurs familles de langues.)

3. LES DIFFÉRENTES APPROCHES EN TYPOLOGIE DES LANGUES

Quatre approches caractérisent la typologie linguistique :

- ◆ l'approche monothétique
- ◆ l'approche polythétique
- ◆ l'approche implicationnelle
- ◆ l'approche paramétrique

La typologie monothétique est fondée sur le recours à un seul critère pour la caractérisation d'un groupe de langues définissant un type. Il y a alors autant de types de langues qu'il y a de critères classificatoires.

La typologie polythétique définit un type de langues à partir d'un ensemble de critères classificatoires. Il peut s'agir d'une collection de propriétés sans aucune relation entre elles ou d'un ensemble hiérarchisé de propriétés.

La typologie implicationnelle propose une classification des langues en fonction de propriétés liées par une relation d'implication logique. L'existence d'une propriété observée dans une langue permet de déduire l'existence d'une autre propriété. Un type de langues est alors un ensemble de langues ayant une propriété commune dont on peut inférer d'autres propriétés. Exemple : les langues VSO – langues dont l'organisation linéaire des constituants majeurs de la phrase est Verbe-Sujet-Objet – ont des prépositions (et non des postpositions) et elles placent l'auxiliaire devant le verbe et non après.

La typologie paramétrique propre à la grammaire générative classe les langues en fonction de variables définies par la théorie. La variation prend place au sein d'une propriété générale commune à toutes les langues. Exemple : dans toutes les langues, les relations de dépendance entre les constituants syntaxiques s'expriment en termes de tête et de complément. La variation porte alors sur la place respective de ces deux termes. On a ainsi deux types de langues ; les langues dont la tête syntaxique se place à gauche du complément et celles dans lesquelles la tête se place à droite du complément. Le français est une langue du premier type ; le verbe et l'adposition se placent à gauche de leur complément et les adpositions sont donc des prépositions, tandis que le japonais est une langue du second type ; le verbe et l'adposition se placent à droite de leur complément et les adpositions sont donc des postpositions.

Ces quatre approches correspondent à quatre grandes périodes de l'histoire de la typologie. La typologie monothétique prend son essor aux XVIII^e et XIX^e siècles avec les premières tentatives de classification des langues. Cette période est marquée principalement par les travaux de Wilhelm Von Humboldt (1767-1835) à qui l'on doit la classification des langues, en langues isolantes, agglutinantes et flexionnelles. La typologie polythétique trouve son aboutissement avec les travaux de Sapir (1884-1939). Ce dernier complète la précédente classification d'ordre morphologique en hiérarchisant les critères classificatoires (les types morphologiques seront définis à partir d'un autre critère classificatoire : le degré de

synthèse des mots). La typologie implicationnelle est issue des travaux de Joseph Greenberg (1915-2001) sur les universaux linguistiques fondés notamment sur l'ordre des mots (Greenberg [1963]). Cette approche de la typologie fait l'objet encore de nombreux travaux et le nombre des universaux d'implication (recensé sur le site de l'université de Constance : <http://ling.unikonstanz.de/pages/proj/sprachbau.htm>) approche actuellement les 2000 (dans son travail fondateur Greenberg en avait proposé 45). La typologie paramétrique est apparue dans les années 1980 à la suite des travaux fait en grammaire générative sur la variation linguistique au sein du modèle théorique élaboré par Noam Chomsky (1928-) dans les années 50.

Les textes fondateurs (pour information) :

- ◆ Typologie monothétique : Wilhelm Von Humboldt, 1836, *Über die Verschiedenheit des menschlichen Sprachbaus und seinen Einfluss auf die geistige Entwicklung des Menschengeschlechts* [Introduction à l'œuvre sur le kavi et autres essais, traduit par Pierre Caussat, Paris : Seuil, 1974]
- ◆ Typologie polythétique : Edward Sapir, 1921, *Language : an Introduction to the Study of Speech*, New York [traduction française : *Le langage*, Payot, Paris, 1953].
- ◆ Typologie implicationnelle : Joseph H. Greenberg (ed.), 1966, *Universals of Language*, Second edition, The MIT Press, Cambridge Mas.
- ◆ Typologie paramétrique : Chomsky, N. and H. Lasnik, 1993, *The theory of principles and parameters*, in: J. Jacobs A. von Stechow, W. Sternefeld, and T. Vennemann (eds.) *Syntax: an international handbook of contemporary research*, De Gruyter, Berlin.

4. LA CLASSIFICATION MONOTHÉTIQUE

Dans cette partie je présenterai et illustrerai rapidement quelques classifications possibles à partir d'un seul critère typologique. Par contre, la typologie morphologique, fondée sur le rapport entre les mots et les morphèmes, ainsi que la distinction syntaxique entre langues nominatives et langue ergatives seront amplement développées, car ces deux classifications des langues peuvent être tenues pour représentatives de cette première approche typologique.

4.1. LES CRITÈRES

Dans ce qui suit, je propose quelques exemples de la classification typologique des langues au moyen d'un critère structural. Je limite à trois le nombre de critères pour chacun des domaines de l'analyse linguistique (phonétique, phonologie, morphologie et syntaxe).

4.1.1. CLASSIFICATION DES LANGUES AU MOYEN D'UN CRITÈRE PHONÉTIQUE

Tons et registres : la présence de tons (variation mélodique pendant la réalisation des voyelles) ou de registres (les voyelles sont réalisées avec des hauteurs différentes) permet de répartir les langues du monde

en deux types selon qu'elles contiennent ou pas des tons ou des registres. Les tons caractérisent une bonne partie des langues d'Asie (chinois, vietnamien, thai...) tandis que les registres se rencontrent dans une partie des langues d'Afrique (yoruba, igbo, éwé...).

Diphthongues : langues à diphthongues (anglais, finnois, chinois, khmer...) et langues sans diphthongues (français, turc...).

Clics : les clics sont des bruits de succion comme celui qu'on utilise pour faire avancer les chevaux. Certaines langues – très peu – les utilisent comme consonnes (tout bruit produit par les organes phonatoires est susceptible de fonctionner comme une consonne dans une langue). Les langues à clics sont les langues khoisan, hottentot (nama) et boshiman (on entend ces clics dans le film botswanais de Jamie Uys : *Les dieux sont tombés sur la tête*).

4.1.2. CLASSIFICATION DES LANGUES AU MOYEN D'UN CRITÈRE PHONOLOGIQUE

Cette classification fait référence aux traits phonologiques et aux propriétés prosodiques utilisées à des fins distinctives.

Labialité : l'opposition entre voyelles arrondies et voyelles étirées est pertinente dans certaines langues (français, turc, finnois, suédois, coréen...) mais pas dans d'autres (italien, espagnol, anglais, arabe, hindi...)

Hauteur : la hauteur est distinctive dans les langues à tons et dans les langues à registres.

Intensité : l'intensité est pertinente dans les langues qui font un usage distinctif de l'accent de mot. La place de l'accent étant variable, deux mots peuvent être en opposition uniquement par la place de l'accent (en Europe : anglais, italien, russe...). Par contre, l'accent d'intensité ne joue aucun rôle distinctif en français, allemand, tchèque, swahili... Dans ces langues, l'accent d'intensité est invariable (sur la première syllabe en tchèque, sur la dernière en français, sur l'avant-dernière en swahili...) et sa fonction est alors démarcative (il sert à l'identification des frontières de mots).

4.1.3. CLASSIFICATION DES LANGUES AU MOYEN D'UN CRITÈRE MORPHOLOGIQUE

Genre grammatical : ce critère concerne également par extension tous les systèmes qui comportent une classification des noms (classes nominales, classificateurs). On distingue ainsi les langues avec genre grammatical (français, italien, danois, gallois, pashto, hindi, avar, swahili, lingala, chinois, vietnamien...) et les langues sans genre (finnois, turc, mongol, éwé, yoruba, indonésien, quechua...)

Cas grammaticaux : certaines langues identifient les fonctions syntaxiques de manière morphologique (désinence, clitique, particule, adposition). On distingue ainsi les langues avec cas (latin, allemand, polonais, lituanien, finnois, turc, hindi, géorgien, arabe, hébreu, japonais, coréen...) et des langues sans cas (français, anglais, suédois, gallois, irlandais, guarani, chinois, vietnamien, éwé, wolof...)

Nature des adpositions : les langues se répartissent en deux ensembles ; celles qui ont des prépositions (français, irlandais, russe, lituanien, arabe...) et celles qui ont des postpositions (finnois, turc, hongrois, hindi, japonais...). Ce critère sera repris et illustré plus loin, dans la présentation de la typologie implicationnelle.

4.1.4. CLASSIFICATION DES LANGUES AU MOYEN D'UN CRITÈRE SYNTAXIQUE

Ordre des constituants majeurs de la phrase : dans cette classification non binaire, les langues sont classées en fonction de l'agencement linéaire du sujet, du verbe et de l'objet. Les trois principaux types sont l'ordre SOV (turc, japonais, hindi...), SVO (français, anglais, finnois...) et VSO (arabe, gallois, maori...). Ce critère sera développé dans la partie consacrée à la typologie implicationnelle.

Nominativité/ergativité : ce critère qui fait référence à la forme des arguments du verbe dans la phrase transitive et intransitive est développé plus loin.

Nature du passif : la construction passive peut se présenter sous deux formes ; une forme analytique ou périphrastique et une forme synthétique. Dans les langues où le passif est de type analytique, la morphologie passive est répartie entre un auxiliaire et une forme nominale du verbe (participe). En français, la forme verbale passive est caractérisée par la combinaison *être*+participe passé. Les langues ayant ce type de passif sont notamment : le français, l'anglais, le gallois, l'hindi... Dans les langues ayant une construction passive de type synthétique, la morphologie passive est un morphème verbal qui se combine avec les autres morphèmes de la flexion du verbe. Le swahili est une de ces langues. Finnois, estonien, turc, japonais sont également des langues avec passif synthétique.

SWAHILI	<i>mwana amekuliwa</i>	<i>na simba</i>
	enfant ...+manger+ PASSIF +...	par lion
	<i>L'enfant a été mangé par le lion</i>	

Cette phrase passive est à comparer avec la phrase active correspondante :

<i>simba amekula</i>	<i>mwana</i>
lion ...+manger+...	enfant
<i>Le lion mange l'enfant</i>	

Dans cet exemple, les morphèmes de flexion (temps et accord en personne) ne sont pas glosés. La voyelle "i" devant le morphème de passif -w est une voyelle épenthétique (voyelle de transition entre deux consonnes).

4.2. LA TYPOLOGIE MORPHOLOGIQUE

Les deux dimensions classificatoires que je présenterai dans cette partie sont 1) le degré de synthèse et 2) le type morphologique. L'une et l'autre classification reposent sur la nature de la relation entre les mots et les morphèmes. Le degré de synthèse s'applique au mot et prend en compte la quantité d'information qu'il contient en termes de morphèmes, tandis que le type morphologique procède de manière inverse en prenant appui sur le morphème et en classant les langues en fonction du mode de réalisation des morphèmes dans le mot. Ces deux approches complémentaires peuvent être schématisées ainsi :

degré de synthèse : mot → morphème

type morphologique : mot ← morphème

Les deux notions utiles à cette typologie sont le mot et le morphème. Quelques remarques d'ordre général à ce propos. Le mot existe dans toutes les langues, mais ce à quoi il correspond varie beaucoup d'une langue à l'autre. Une définition générale est donc difficile à mettre en œuvre. La meilleure définition du mot est celle de Mel'čuk [1993]. On en fera l'économie ici car elle repose sur plusieurs notions qui demanderaient un développement qui dépasse le cadre de ce cours. On admettra donc ici la réalité du mot dans les langues. Le morphème, quant à lui, procède de l'analyse – ce n'est pas une donnée immédiate des langues – et fait l'objet d'une définition universelle (le morphème est la plus petite unité de signification). Étant donné la variabilité du mot dans les langues, on comprend par conséquent la pertinence d'une typologie. Toutes les langues n'articulent pas de la même manière les mots et les morphèmes.

4.2.1. LE DEGRÉ DE SYNTHÈSE

Une comparaison entre deux énoncés équivalents en finnois et en français permettra de comprendre ce qui est en jeu avec le degré de synthèse. Soit les deux énoncés suivants :

FINNOIS	<i>taskuissani</i>
FRANÇAIS	<i>dans mes poches</i>

Ces deux énoncés sont rigoureusement équivalents ; ils contiennent la même information, à la fois sur le plan quantitatif et sur le plan qualitatif. Et pourtant ils n'ont pas la même forme ; un mot en finnois et trois mots en français. Procédons maintenant à l'analyse morphologique de l'exemple finnois :

FINNOIS	<i>taskuissani</i> = <i>tasku</i> + <i>i</i> + <i>ssa</i> + <i>ni</i> = "poche" + PL + INESSIF + 1SG PL = pluriel INESSIF = "dans" SG = première personne du singulier
----------------	---

Autrement dit, les "ingrédients" de cet énoncé sont 1) la signification du lexème *poche*, 2) une signification de pluralité exprimée par le nombre, 3) une signification correspondant à une localisation et 4) une indication de personne. Il n'y a pas lieu de recourir à une quelconque signification correspondant à la possession, car cette notion n'est rien d'autre que l'interprétation résultant de la mise en relation d'un objet (*poche*) et d'une personne (1SG). On dira par conséquent que le mot finnois contient quatre segments et chaque segment représente un morphème.

Les quatre informations contenues dans l'énoncé finnois se retrouvent évidemment en français puisque les deux énoncés ont la même signification globale. Il y a donc les mêmes morphèmes, mais répartis autrement. Le morphème lexical forme un mot avec le morphème de nombre (*poches*), le morphème de

personne est un mot indépendant (*mes*) et le morphème correspondant à l'expression d'une localisation est représenté en français par une préposition (*dans*).

Remarquons, que le nombre en français se manifeste deux fois dans l'énoncé. Du point de vue du contenu (ce que l'on cherche à exprimer) il n'est pertinent qu'une seule fois, mais sa présence dans le déterminant possessif est le produit d'une relation de dépendance syntaxique entre le nom et le déterminant. De même, on ne compte pas ici le genre dans les morphèmes exprimant un contenu, car le genre grammatical n'est pas une catégorie qui fait l'objet d'un choix de la part du locuteur. Le genre est simplement une manifestation d'une relation de dépendance entre constituants syntaxiques ; les locuteurs ne cherchent pas à exprimer une quelconque notion (liée notamment à la distinction de sexe) en procédant à un accord en genre.

Résumons : à un même contenu, exprimé en termes de morphèmes, correspond plusieurs réalisations dans les langues. Avec le degré de synthèse, les langues sont classées selon le contenu morphologique des mots. Il s'agit à la fois d'une évaluation quantitative (un ou plusieurs morphèmes dans le mot) et qualitative (nombre de lexèmes dans le mot).

Le degré de synthèse comporte trois types :

- ◆ le type analytique
- ◆ le type synthétique
- ◆ le type polysynthétique

4.2.1.1. Type analytique

Dans les langues analytiques, les mots ne contiennent qu'un seul morphème et sont par conséquent invariables. Les mots complexes ne peuvent être formés que par composition (mise en relation de deux mots). Nombreuses langues d'Asie sont de ce type. Exemples vietnamiens :

VIETNAMIEN	<i>Tôi cho anh moi donner toi Je te donne</i>
	<i>Anh cho tôi toi donner moi Tu me donnes</i>
	<i>Họ cho tôi sách il donner moi livre Il me donne un livre</i>
	<i>làng tôi village moi mon village</i>

Comme le montrent ces quatre exemples la forme du pronom personnel de première personne est invariable ; sa forme ne dépend pas de sa fonction dans la phrase. La place des constituants est le seul

procédé d'identification des fonctions syntaxiques. Le verbe est également invariable ; l'indication de temps du procès décrit par le verbe est donné éventuellement par des morphèmes indépendants comme *demain hier maintenant...*

Puisque les mots ne contiennent qu'un seul morphème, il ne peut y avoir ni flexion (variation en temps, personne, nombre, genre...) ni dérivation (formation de mots au moyen d'affixes). Les mots complexes sont donc formés exclusivement par composition. Quelques exemples :

VIETNAMIEN	<i>thợ cày = laboureur (thợ = homme, cày = labourer)</i> <i>khai trường = rentrée scolaire (khai = ouvrir, trường = école)</i> <i>làm việc = travailler (làm = faire, việc = travail)</i>
-------------------	---

4.2.1.2. Type synthétique

Dans les langues synthétiques, les mots contiennent plusieurs morphèmes mais un seul lexème. Outre le lexème, les autres morphèmes sont des affixes ou des traits morphologiques. Dans ce dernier cas, le morphème – unité minimale de signification n'est pas associé à un segment. L'exemple finnois précédent (*taskuissani = dans mes poches*) est un exemple typique de mot synthétique. Autres exemples :

JAPONAIS	<i>urite = vendeur</i> ur+(i)te vendre+suffixe agentif
	<hr/> <i>tabesaserareru = a été fait mangé</i> tabe+sase+rare+ru manger+CAUSATIF+PASSIF+PRÉSENT
TURC	<i>gölgeli = ombragé</i> gölge+li ombre+suffixe adjectival
	<hr/> <i>evlerimde = dans mes maisons</i> ev+ler+(i)m+de maison+PLURIEL+1SG+LOCATIF
QUECHUA	<i>wañuna = mort</i> wañu+na mourir+suffixe nominal
	<hr/> <i>yanapawashkanki = (tu) m'as aidé</i> yanapa+wa+shka+nki aider+1SG+PASSÉ+2SG

Dans tous ces exemples, les morphèmes sont réalisés sous la forme de segments et les affixes sont soit des préfixes soit des suffixes (selon la place qu'ils occupent par rapport au radical (lexème)).

Un mot synthétique peut aussi contenir des morphèmes non segmentaux. Les morphèmes (ou traits morphologiques) sont révélés alors par l'analyse.

Le mot *ma* du français contient trois morphèmes : la personne (1SG), le nombre (SG), et le genre (féminin). Chacun de ces morphèmes est mis en évidence par les oppositions avec : *ta* (2SG), *mes* (PL), et *mon* (masculin). Les amalgames du français sont aussi des formes synthétiques puisqu'ils contiennent en un mot exactement les mêmes morphèmes que leur équivalent non amalgamé : *du* (= *de le*) vs *de la*, *au* (= *à le*) vs *à la*...

La prosodie peut également être utilisée dans une forme synthétique :

ÉWÉ	<i>vǎ = viens !</i> vǎ + venir+IMPÉRATIF
------------	--

En ajoutant une modulation montante au ton haut de la voyelle du radical verbal on obtient la forme impérative.

4.2.1.3. Type polysynthétique

Dans les langues polysynthétiques, un mot peut contenir plusieurs morphèmes et aussi plusieurs lexèmes, si bien que le mot peut parfois être assimilé à une phrase complète comme l'exemple suivant emprunté à Hjemlev [1966] :

GROENLANDAIS	<i>kavfiliorniarumagaluarpunga = je voudrais faire du café</i>
---------------------	--

Malheureusement, l'auteur ne donne pas l'analyse morphologique de cet exemple. On peut cependant supposer la présence de trois lexèmes *vouloir*, *faire* et *café*.

Autre exemple en yupik, autre langue de la famille esquimau-aléoute :

YUPIK	<i>tuntussuqatarniksaitengqiggtuq</i> = <i>Il n'a pas encore dit de nouveau qu'il allait chasser le renne</i> <i>tuntu+ssur+qatar+ni+ksaite+ngqiggte+uq</i> renne+chasser+FUTUR+dire+NEG+de nouveau+3SG/IND (NEGation, INDicatif)
--------------	---

Exemple plus simple en nahuatl, langue amérindienne :

NAHUATL	<i>ooninakakwah = j'ai mangé de la viande</i> <i>oo+ni+naka+kwa+h</i> PASSÉ+1SG+viande+manger+ACCOMPLI
----------------	--

Ce mot contient deux lexèmes (*viande* et *manger*) et trois morphèmes grammaticaux (personne, temps et aspect).

Quel est le statut du français au regard de cette typologie ? Sans aucun doute, le français est une langue synthétique comme le montrent les deux exemples suivants représentatifs d'une partie de la flexion des verbes et de la dérivation nominale :

FRANÇAIS	<i>chantions</i> chant+i+ons chanter+IMPARFAIT+1PL
	<i>dénationalisation</i> dé+nation+al+is+ation préfixe+nation+suffixe+adjectival+suffixe verbal+suffixe nominal

Ces deux mots contiennent un lexème (respectivement *chant-* et *nation*) et plusieurs morphèmes flexionnels ou dérivationnels.

Notons que la distinction entre type analytique et type synthétique donne lieu parfois à une opposition en français. On oppose ainsi un futur synthétique (un seul mot pour *chantera*) et un futur analytique (deux mots pour *va chanter*). Par ailleurs, on oppose également la forme synthétique – quelque peu obsolète – *doctoresse* et la forme équivalente analytique *femme médecin*. De même en anglais, on a une forme synthétique du comparatif (*bigger*) et une forme analytique (*more expansive*). De même, dans plusieurs langues, on distingue une conjugaison synthétique où les morphèmes de temps, et de personne, forment un seul mot avec le radical verbal, et une conjugaison analytique où ces mêmes morphèmes fonctionnels sont répartis entre un auxiliaire et le radical verbal (cette distinction est présente notamment en basque et en gallois).

4.2.2. TYPES MORPHOLOGIQUES

Les langues se répartissent en quatre types :

- ◆ type isolant
- ◆ type agglutinant
- ◆ type fusionnant
- ◆ type incorporant

4.2.2.1. Type isolant

Dans les langues isolantes qui sont également des langues analytiques, le morphème se réalise sous la forme d'un mot invariable. À chaque morphème correspond un mot. Dans les exemples chinois et vietnamiens suivants, tous les mots sont invariables (pas d'accord en nombre) et le nombre est indiqué par un mot indépendant (numéral ou indicateur de pluriel) :

CHINOIS	<i>zhei san ben shu = ces trois livres</i>
	DEM trois CLASS livre

VIETNAMIEN	<i>những chiếc bàn ấy</i> = ces tables-là PL CLASS table DEM (DEMonstratif, CLASSificateur, Pluriel)
-------------------	--

4.2.2.2. Type agglutinant

Dans les langues agglutinantes, les morphèmes sont réalisés sous la forme de segments. Les morphèmes flexionnels et dérivationnels sont attachés à un radical lexical par préfixation ou suffixation. Les langues agglutinantes sont donc nécessairement de type synthétique. Les morphèmes sont facilement identifiables dans le mot en raison du procédé d'agglutination qui consiste à ajouter les segments correspondant aux morphèmes les uns à côté des autres. Mis à part les cas d'allomorphie, dus le plus souvent à des contraintes phonologiques propre à chaque langue, les segments sont invariables. Le type agglutinant se caractérise donc par le procédé de concaténation des morphèmes dans le mot (A + B + C + D + ...), les altérations de la forme des morphèmes intervenant à la jonction des segments étant de nature secondaire. Turc et finnois sont deux langues représentatives de ce type :

TURC	<i>evlerimde</i> = dans mes maisons ev+ler+(i)m+de maison+PLURIEL+1SG+LOCATIF
FINNOIS	<i>taloissammekin</i> = dans nos maisons aussi talo+i+ssa+mm+kin maison+PL+INESSIF+1PL+aussi
	<i>näkymättömydessänsä</i> = dans son invisibilité näky+mä+ttöm+yyde+ssä+nsä être visible+suffixe nominal+suffixe adjectival+suffixe verbal+INESSIF+3SG/PL 3SG/PL troisième personne du singulier ou du pluriel

Dans l'exemple turc, le suffixe de personne *-(i)m* admet deux formes (*-m* et *-im*) selon qu'il est attaché à un mot se terminant par une consonne (*-im*) ou à un mot se terminant par une voyelle (*-m*). Les suffixes de pluriel et de locatif ont deux formes (respectivement *-ler/-lar*, *-de/-da*). Le choix entre les deux variantes est gouverné par un processus d'harmonie vocalique qui aligne la voyelle suffixale sur les voyelles du radical en reprenant une des propriétés phonologiques partagées par les voyelles du radical. Il s'agit là rien de moins qu'un processus d'accord interne au mot.

Dans l'exemple finnois, le suffixe d'inessif (= *dans*) a deux formes (*-ssa* et *-ssä*) dont le choix est également gouverné par un processus d'harmonie vocalique comparable à celui du turc.

4.2.2.3. Type fusionnant

Les langues fusionnantes sont appelées aussi parfois langues flexionnelles. On préférera ici le terme fusionnant en raison de sa similitude avec les autres mots qui désignent le type morphologique. Dans les

langues fusionnantes, qui sont également nécessairement des langues synthétiques, les différents morphèmes du mot peuvent être réalisés sous la forme d'un seul segment. Il est donc impossible dans ce cas de segmenter le mot en parties plus petites, de telle sorte que chaque morphème soit associé à un segment comme dans les langues agglutinantes. Lorsque la fusion est complète, les morphèmes sont identifiables seulement par l'analyse à partir du paradigme des mots de la même catégorie. Dans l'exemple des articles définis de l'espagnol {*el, la, los, las*}, chaque mot représente un seul segment correspondant à chaque fois à une collection de trois morphèmes :

ESPAGNOL	<i>el</i> = [défini, masculin, singulier] <i>la</i> = [défini, féminin, singulier] <i>los</i> = [défini, masculin, pluriel] <i>las</i> = [défini, féminin, pluriel]
-----------------	--

La même chose pour les articles définis du français {*le, la, les*}, à la différence près que le paradigme à trois termes renvoie à une neutralisation (non distinction) de l'opposition de genre au pluriel. On analysera de la même manière, tout un ensemble de mots en français : les pronoms personnels, les articles indéfinis, les déterminants possessifs, le système des démonstratifs...

La fusion concerne les mots grammaticaux mais aussi les radicaux augmentés d'un ou plusieurs morphèmes flexionnels. En anglais la forme de mot *went* doit être analysée comme la fusion du radical verbal de *go* et du morphème de prétérit (*went* = $\sqrt{\text{go}}$ + prétérit). Cet exemple est à comparer avec la forme régulière qui s'analyse comme la concaténation d'un radical et d'un suffixe (*arrived* = *arrive*+*ed* = *arrive*+prétérit). *Went* relève donc de la fusion des morphèmes alors que *arrived* relève de l'agglutination des morphèmes.

Nombreuses sont les langues où la fusion des morphèmes ne concerne que la terminaison du mot, le radical (le lexème) étant alors un segment plus ou moins autonome (la segmentation est alors plus ou moins arbitraire). Quatre exemples typiques :

LATIN	<i>dominus</i> = <i>maître</i> domin+us (-us = nominatif, masculin, singulier)
POLONAIS	<i>studenta</i> student+a (-a = nominatif, animé, masculin, singulier)
ALBANAIS	<i>shëpia</i> = <i>la maison</i> shëpi+a (-a=défini, nominatif, masculin, singulier)
FRANÇAIS	<i>vendeuse</i> vend+euse (-euse = agent, féminin)

On désigne parfois ce type de langues, langues à flexion externe (par opposition aux langues à flexion interne dans lesquelles le radical fusionne avec ses morphèmes fonctionnels).

4.2.2.4. Type incorporant

Les langues incorporantes sont des langues pour lesquelles on peut dire de manière approximative qu'il est possible de rencontrer des "mots à l'intérieur de mots". Ces langues, nécessairement de type polysynthétique, ont la particularité de pouvoir intégrer des lexèmes dans le mot. Il s'agit donc bien de lexèmes dans un mot et non pas d'un mot dans un autre mot. Ce point sera plus clair lorsque nous aurons examiné un exemple d'une langue incorporante. Mais avant cela, un détour permettra de comprendre ce qu'est une incorporation à partir des quelques rares exemples du français qui se situent exclusivement sur le plan de la formation historique des mots. Les mots suivants ont été formés par incorporation : *saupoudrer*, *colporter*, *lieutenant*, incorporation de l'objet (ou d'un complément) au verbe :

saupoudrer ← *poudrer de sel*

colporter ← *porter au cou*

lieutenant ← *qui tient lieu de*

Il s'agit d'une formation historique ; la signification de *saupoudrer* n'a plus rien à voir avec le sel. Seul ici *lieutenant* garde une certaine transparence. Le français n'est pas une langue incorporante, car l'incorporation n'est pas dans cette langue un des procédés productifs pour former des mots complexes ; on utilise plutôt la dérivation (*vendeur...*), la composition (*homme-sandwich...*) et la création de mots valise (*informatique = information automatique*). Dans les langues incorporantes, l'incorporation est un procédé banal utilisé aussi bien pour la formation de mots dans le lexique que pour la construction d'unités syntaxiques. L'exemple le plus connu de construction syntaxique par incorporation est l'incorporation de l'objet indéfini ou générique au verbe. Lorsqu'un objet est défini ou spécifique, il est réalisé sous la forme d'un constituant indépendant ayant la fonction d'objet. Lorsqu'il est indéfini ou générique, il est incorporé au verbe. Ce dernier appartient alors à une construction intransitive. Lorsqu'il y a incorporation, le lexème incorporé perd toutes les caractéristiques nominales qui en font un mot indépendant (pas de nombre, pas d'article, pas de cas). Il s'agit donc bien d'un lexème nu et non pas d'un mot ; c'est pourquoi l'incorporation ne peut pas être définie simplement comme la présence de mots à l'intérieur d'autres mots. Exemples onondaga, langue iroquoise du Canada :

ONONDAGA

Pet waʔhahtuʔtaʔ *neʔ ohwista*
Pat waʔ+ha+htu+ʔt+aʔ *neʔ o+hwist+a*
 Pat PASSÉ+3M/S+3N+perdre+CAUSATIF+ASPECT ART PRÉ+argent+SUF
Pat a perdu l'argent

Pet waʔhahwistahtuʔtaʔ
Pat waʔ+ha+hwist+ahtu+ʔt+aʔ
 Pat PASSÉ+3M/S+argent+perdre+CAUSATIF+ASPECT
Pat a perdu de l'argent

3M/S = troisième personne du masculin/Sujet
 3N = troisième personne du neutre
 ARTicle, PREFixe, SUFFixe

Dans le premier exemple, l'objet est défini et se présente comme un constituant phrastique indépendant ; il est précédé d'un article et contient un préfixe et un suffixe nominal. Dans le second exemple, le lexème *argent* se présente nu, et est incorporé au verbe pour former un verbe intransitif ayant pour signification *perdre de l'argent*. On remarquera par ailleurs que le verbe du premier exemple s'accorde en genre avec le sujet (masculin) et avec l'objet (neutre).

Autre exemple, déjà vu à propos du type polysynthétique. Ici, on observe l'incorporation de l'objet partitif (une certaine quantité de...)

NAHUATL *ooninakakwah = j'ai mangé de la viande*
 oo+ni+naka+kwa+h
 PASSÉ+1SG+viande+manger+ACCOMPLI

Une hirondelle ne fait pas le printemps et un exemple ne suffit pas à se prononcer sur l'appartenance d'une langue à un type, d'autant plus qu'elles sont loin d'être homogènes. Certaines langues peuvent être cependant considérées comme des prototypes dans la mesure où tous les domaines de la langue sont marqués par un des procédés de formation des mots qui caractérisent les types morphologiques. Ces domaines sont la flexion des noms (déclinaison) et des verbes (conjugaison) et la dérivation lexicale. C'est donc à la fois en syntaxe et dans le lexique que se manifestent les propriétés qui touchent à la structuration du mot. Comme cela a été dit précédemment, des langues comme le chinois et le vietnamien peuvent être tenues pour le prototype des langues isolantes. De même que le turc et le finnois représentent des prototypes de langues agglutinantes. D'autres langues ont un statut moins net car on trouve des mots qui relèvent aussi bien d'un type que d'un autre. On définira donc plutôt des tendances pour ces langues. L'anglais est exemplaire sur ce point, car sa dérivation lexicale est clairement de nature agglutinante (*childhood* enfance = *child+hood*, *childless sans enfants* = *child+less*, *baker boulanger* = *bake+er*) alors que sa syntaxe tend très nettement vers le type isolant. On considère d'ailleurs parfois l'anglais comme relativement proche du chinois sur ce point (la même chose vaut pour les langues scandinaves). Cette impression vient du fait que l'anglais a une flexion des mots très pauvre. Le nombre des variations est très réduit ; il y a en tout et pour tout huit morphèmes flexionnels dans cette langue ;

ANGLAIS -s = pluriel
 -s = génitif
 -s = 3SG (présent)
 -ed = prétérit
 -ing = participe présent
 -en = participe passé
 -er = comparatif
 -est = superlatif

Qu'en est-il du français ? On observe aussi bien de l'agglutination que de la fusion, et ce, aussi bien en dérivation (lexique) qu'en flexion (syntaxe) :

	AGGLUTINATION	FUSION
LEXIQUE	<i>invisibilité</i> ...	<i>vend<u>euse</u></i> ...
SYNTAXE	<i>chantions</i> <i>maisons</i> ...	<i>chant<u>âmes</u></i> <i>sont</i> <i>chevaux</i> <i>beau</i> ...

Dans ce tableau, les segments qui relèvent de la fusion de morphèmes sont soulignés. On remarquera que le mot *cheval/chevaux* s'analyse comme un tout ; il serait absurde de segmenter en **chev+al* et **chev+aux* comme on peut le faire pour les mots dérivés du type *normal/normaux* (*norm+al* et *norm+aux*).

La présence de la fusion dans certaines formes nominales, adjectivales et verbales suffit cependant à caractériser le français comme langue fusionnante (ou comme une langue plutôt fusionnante si l'on veut relativiser) car par opposition, les langues agglutinantes ne présentent pas de cas de fusion. Il suffit donc qu'il y ait fusion pour parler de langue fusionnante.

4.3. LANGUES NOMINATIVES ET LANGUES ERGATIVES

Bien que reposant sur un seul critère, la distinction entre langues nominatives et langues ergatives met en jeu deux propriétés : 1) l'identification morphosyntaxique des fonctions sujet et objet dans les phrases transitives, 2) le statut morphosyntaxique du sujet des phrases intransitives. Cette dernière propriété est liée à la précédente car le statut du sujet des phrases intransitives est défini par rapprochement avec les propriétés du sujet et/ou de l'objet des phrases transitives.

4.3.1. L'IDENTIFICATION DES FONCTIONS

Considérons tout d'abord le cas du français dans l'exemple suivant :

Le mécanicien répare les voitures

Cette phrase comporte un verbe et deux arguments. Les arguments sont les expressions référentielles qui renvoient à des individus ou à des objets. L'un des arguments du verbe est celui qui fait quelque chose (l'agent) et le second, la chose ou la personne, affectée ou concernée, par l'activité du premier (le patient). C'est une propriété lexicale du verbe transitif *réparer* que de se construire avec deux arguments, interprétés respectivement comme agent et comme patient. Dans la phrase précédente, l'agent est "*le mécanicien*" et le patient "*les voitures*". En français, l'agent du verbe est représenté en syntaxe par le sujet, tandis que le

patient est exprimé par l'objet. Le rapport entre les arguments du verbe et les fonctions syntaxiques est différent dans la construction passive ; le patient devient le sujet, alors que l'agent occupe une position périphérique facultative, introduite par une préposition :

Les voitures sont réparées (par le mécanicien)

Les deux fonctions syntaxiques de la phrase transitive sont identifiées en français par deux propriétés : le sujet se place devant le verbe et il s'accorde en nombre et en personne avec le verbe. L'objet se place après le verbe et ne s'accorde pas avec le verbe. Rien dans la forme des syntagmes nominaux *le mécanicien* et *les voitures* ne permet d'identifier leur fonction respective. Ces mêmes syntagmes nominaux gardent la même forme dans les phrases suivantes où ils n'ont pas la même fonction :

Le client attend le mécanicien

Les voitures sont accidentées

Il n'en va pas de même si les syntagmes nominaux de la première phrase sont remplacés par des pronoms :

Il les répare

Cette fois la forme du pronom participe à l'identification de la fonction puisqu'en français, les pronoms varient selon la fonction (il/elle, le/la, lui...). Cette variation de la forme des pronoms selon leur fonction est à mettre en rapport avec l'invariabilité des pronoms dans une langue comme le vietnamien ([voir les exemples précédents en 4. 2. 1. 1.](#)).

En résumé, trois propriétés concourent à l'identification des fonctions syntaxiques nécessaires à l'interprétation des arguments du verbe :

1. la place par rapport au verbe
2. l'accord avec le verbe
3. la forme des syntagmes nominaux

Ces trois moyens d'identification des fonctions sont employés diversement selon les langues, et même au sein des langues, comme le montre le cas du français avec le traitement différent de l'objet pronominal.

La distinction entre langues nominatives et langues ergatives sera traitée dans ce qui suit à partir de ces trois propriétés et de manière inégale : 1) identification des fonctions par le marquage casuel, 2) identification des fonctions par l'accord du verbe, et 3) dans une moindre mesure, identification des fonctions par la place des arguments du verbe.

4.3.1.1. L'identification par le cas morphologique

L'identification morphologique des arguments du verbe est notée dans les structures suivantes par un indice qui représente la présence d'un segment (suffixe, clitique, particule, adposition). Les différentes possibilités sont les suivantes :

1. $S_{\alpha} V O_{\beta}$ Les deux arguments sont marqués.

- 2. S_a V O Le sujet seul est marqué.
- 3. S V O_a L'objet seul est marqué.
- 4. S V O Aucun des deux arguments n'est marqué.

Les exemples correspondants sont les suivants (les marques casuelles sont en gras) :

1. JAPONAIS	<p><i>Hanako ga hon o katta</i> Hanako NOM livre ACC achète Hanako achète un livre</p>
2. BASQUE	<p><i>Martineek gizona ikusi du</i> Martin+ERG homme a vu Martin a vu l'homme</p>
3. HONGROIS	<p><i>János könyvet olvas</i> János livre+ACC lit János lit le livre</p>
4. CHINOIS	<p><i>Wo chi rou</i> je manger viande Je mange de la viande</p>

NOMinatif, ACCusatif, ERGatif

Commentaire :

Dans l'exemple japonais, les deux arguments du verbe sont identifiés respectivement par les particules casuelles *ga* (sujet) et *o* (l'objet). Dans l'exemple basque, c'est le sujet seul qui est marqué (suffixe de cas ergatif : *-ek*) alors qu'en hongrois, c'est l'inverse : l'objet est marqué (accusatif *-t*) mais pas le sujet. En chinois, aucun des arguments n'est marqué morphologiquement pour l'identification du sujet et de l'objet. Il est évident par conséquent que c'est la place respective des arguments qui identifie les fonctions.

Si l'on s'en tient uniquement au marquage casuel, on dira que le hongrois est une langue nominative ; absence de marquage casuel sur le sujet, et présence d'une marque casuelle sur l'objet (accusatif), alors que le basque est une langue ergative du fait de la présence d'une marque casuelle sur le sujet (ergatif) et de l'absence de marquage de l'objet. On ne peut rien dire pour le moment du japonais et du chinois du fait que ces langues n'utilisent pas de marquage distinctif tel qu'un seul des deux arguments soit marqué.

4.3.1.2. L'identification par l'accord du verbe

Une autre façon de distinguer les arguments du verbe transitif est l'accord du verbe. De nouveau, nous devons envisager toutes les possibilités puisqu'elles sont effectivement représentées dans les langues. Dans les structures suivantes, les indices S et O sur le verbe signale respectivement l'accord avec le sujet et l'accord avec l'objet. La nature de l'accord importe peu ici (accord, en nombre, en personne, en genre...) car c'est la présence de la relation qui est pertinente et non pas son contenu.

- 1. S V_{so} O Les deux arguments s'accordent avec le verbe.

- 2. S V_s O Le sujet seul s'accorde avec le verbe
- 3. S V_o O L'objet seul s'accorde avec le verbe
- 4. S V O Aucun des deux arguments ne s'accorde avec le verbe.

Les exemples correspondants sont les suivants (les marques d'accord sont en gras) :

1. QUICHÉ	<p><i>Koxacukuk</i> ASP+1PL/O+2SG/S+chercher <i>Tu nous cherches</i></p> <p><i>Katkacukuk</i> ASP+2SG/O+1PL/S+chercher <i>Nous te cherchons</i></p>
2. RUSSE	<p><i>On pročital etu knigu</i> il lire+3SG/M le livre+ACC <i>Il lit le livre</i></p> <p><i>Ona pročitala etu knigu</i> elle lire+3SG/F le livre+ACC <i>Elle lit le livre</i></p>
3. AVAR	<p><i>Vasas jas jec:ula</i> garçon+ERG fille 3SG/F+complimenter <i>Le garçon complimente la fille</i></p> <p><i>Jasas vas vec:ula</i> fille+ERG garçon 3SG/M+complimenter <i>La fille complimente le garçon</i></p>
4. CHINOIS	<p><i>Wo chi rou</i> je manger viande <i>Je mange de la viande</i></p> <p><i>wo men chi rou</i> je PL manger viande <i>Nous mangeons de la viande</i></p>

1SG/S = première personne du singulier (sujet)
 1SG/O = première personne du singulier (objet)
 3SG/M = troisième personne du masculin singulier
 3SG/F = troisième personne du féminin singulier
 PLuriel

Commentaire :

En quiché (langue maya du Guatemala), le verbe présente deux jeux de marques d'accord ; une série pour le sujet et une seconde pour l'objet. Dans l'exemple précédent, nous avons deux préfixes dans chacun des deux paradigmes :

sujet {*a-* = 1PL , *ka-* = 2SG}

objet {*at-* = 2SG, *ox-* = 1PL}

La personne du sujet et de l'objet de ces phrases n'est pas représentée par un pronom indépendant car l'accord suffit à l'identification des personnes.

En russe, le sujet seul s'accorde avec le verbe (accord en personne, nombre et genre) : suffixe -Ø pour la troisième personne du masculin singulier et -a pour la troisième personne du féminin singulier.

En avar (langue nakho-daghestanienne du Caucase), c'est l'objet qui s'accorde en personne, nombre et genre : préfixe *v-* pour la troisième personne du masculin singulier et préfixe *j-* pour la troisième personne du féminin singulier.

De nouveau ici l'exemple chinois pour illustrer l'absence d'accord (le chinois étant une langue isolante, noms et verbes sont strictement invariables).

4.3.1.3. L'identification par la place des arguments

Considérons maintenant la place relative des arguments par rapport au verbe pour l'identification des fonctions. Comme précédemment, quatre possibilités selon que les places sont fixes ou non (les arguments ayant une place fixe sont soulignés). L'ordre des constituants de la phrase n'étant pas pertinent ici, on s'en tiendra à une séquence du type SVO. Entre parenthèses, les autres séquences possibles dès lors qu'un argument n'est pas fixe :

- | | |
|---|---|
| 1. <u>S</u> <u>V</u> <u>O</u> | Les deux arguments occupent une place fixe. |
| 2. <u>S</u> V <u>O</u> (<u>S</u> OV) | Le sujet seul occupe une place fixe. |
| 3. S <u>V</u> <u>O</u> , (V <u>O</u> S) | L'objet seul occupe une place fixe. |
| 4. SVO, (SOV, OVS, OSV, VSO, VOS) | Aucun des deux arguments n'a de place fixe. |

Très peu de langues n'utilisent que la place des arguments pour l'identification des fonctions syntaxiques. Seuls le premier et le dernier cas de figure sont sans ambiguïté puisqu'il concerne d'une part les langues qui n'ont pas d'autres moyens d'identification (c'est évidemment le cas du chinois et du vietnamien), et d'autre part, les langues qui disposent d'un moyen morphosyntaxique (cas et/ou accord) pour l'identification des arguments. La place des arguments n'est jamais pertinente pour ces langues telles le finnois, le latin... qui sont connues pour être des langues à ordre des mots libre (pour être plus précis, ce sont les constituants de la phrase qui sont libres et non les mots, car la place relative du déterminant et du nom au sein du syntagme nominal, par exemple, est fixe). Reste les deux cas intermédiaires pour lesquels il est difficile de trouver des exemples probants. On laissera donc de côté ce point.

4.3.2. PROTOTYPE NOMINATIF ET PROTOTYPE ERGATIF

Les trois propriétés précédentes participent à l'identification des arguments des verbes transitifs et concourent par conséquent à la caractérisation des fonctions syntaxiques sujet et objet. Considérons à nouveau le cas du français dans l'exemple très simple :

il le regarde

Les arguments du verbe sont ici des pronoms. Le sujet se caractérise par les trois propriétés : 1) il se place devant le verbe, 2) il s'accorde avec le verbe et 3) sa forme est différente de celle de l'objet (bien que cela ne soit pas très courant dans les descriptions du français, il est tout à fait légitime de parler de cas nominatif pour cette forme du pronom). Le pronom objet : 1) est placé devant le verbe (mais après le sujet), 2) il ne s'accorde pas avec le verbe, et 3) sa forme permet de le distinguer du sujet (*le* = forme accusative du pronom de troisième personne du masculin singulier).

Deux des trois propriétés (place et accord) suffisent à la caractérisation du sujet et de l'objet nominal (pas de cas morphologiques en français pour les noms) :

Le voyageur regarde le passant

Le sujet s'accorde avec le verbe et se place devant. L'objet ne s'accorde pas et se place après le verbe. Comme le montrent les deux exemples suivants, un changement de place implique un changement de fonction et un traitement différent vis à vis de l'accord.

Le passant regarde le voyageur

Les passants regardent le voyageur

Cas et accord étant très souvent complémentaires dans l'identification des fonctions syntaxiques, la prise en compte de ces deux propriétés morphosyntaxiques permet de dégager les prototypes suivants :

Langue nominative : $S V_s O_\beta$

Langue ergative : $S_a V_o O$

Exemples :

LANGUE NOMINATIVE, TURC *ben* *evi* *gördüm*
 je+NOM maison+ACC voir+PAS+1SG
J'ai vu la maison

LANGUE ERGATIVE, AVAR *Vasas* *jas* *jec:ula*
 garçon+ERG fille 3SG/F+complimenter
Le garçon complimente la fille

Dans les langues du premier type, le sujet est non marqué (cas nominatif) et s'accorde avec le verbe. L'objet ne s'accorde pas avec le verbe et il est marqué par le cas accusatif. Dans l'exemple turc, l'accusatif est représenté par le suffixe *-i* et l'accord (personne et nombre) est représenté par le suffixe *-m* (première personne du singulier). On désigne les langues de ce type par le nom des deux cas morphologiques du sujet et de l'objet. On parle alors de langues du type nominatif/accusatif ou plus simplement, de langues nominatives.

Dans les langues du second type, le sujet est marqué (cas ergatif) et ne s'accorde pas avec le verbe. L'objet s'accorde avec le verbe et il est non marqué. Dans l'exemple avar, l'ergatif est représenté par le suffixe *-as* et l'accord (personne, nombre et genre) est représenté par le préfixe *j-*. Dans ces langues, le cas

non marqué est appelé cas absolutif. Ces langues correspondent au type ergatif/absolutif, et on dira plus simplement qu'il s'agit de langues ergatives.

Bien entendu, le cas et l'accord ne se comportent pas toujours exactement comme dans les prototypes précédents. Conformément aux différents cas de figure, on peut trouver quasiment toutes les combinaisons possibles, aussi bien dans la comparaison des langues qu'au sein des langues elles mêmes, car ces dernières ne sont pas toujours homogènes vis à vis de la distinction entre les deux types. Des langues ont, par exemple, une structure nominative pour les arguments nominaux et une structure ergative pour les arguments pronominaux. Autre exemple : des langues ont une structure nominative lorsque le verbe est à un temps donné, mais une structure ergative pour un autre temps...

Des langues nominatives ou ergatives peuvent avoir un double marquage casuel et/ou un accord avec les deux arguments, et à l'opposé, la distinction entre langues nominatives et langues ergatives peut se réduire parfois à une seule des deux propriétés.

4.3.3. STATUT DU SUJET DES PHRASES INTRANSITIVES

Les trois propriétés précédentes (cas, accord et place) peuvent ne pas suffire à la caractérisation des langues au regard de la distinction entre langues nominatives et langues ergatives. Considérons les trois exemples suivants empruntés respectivement au japonais, au swahili et au norvégien :

JAPONAIS	<i>Hanako ga hon o katta</i> Hanako NOM livre ACC achète+PAS <i>Hanako a acheté un livre</i>
SWAHILI	<i>Hamisi alimwona Juma</i> Hamisi 3SG/S +PAS+ 3SG/O +voir Juma <i>Hamisi a vu Juma</i>
NORVÉGIEN	<i>De leser historie</i> ils étudier+PRES histoire <i>Ils étudient l'histoire</i>

3SG/S = troisième personne du singulier (sujet)
3SG/O = troisième personne du singulier (objet)
NOMinatif, ACCusatif, PASsé, PREsent

Ces exemples représentent trois cas extrêmes. En japonais, les arguments sont tous les deux marqués par une particule postposée équivalant à un suffixe de cas morphologique, mais pas d'accord du verbe. En swahili, le verbe s'accorde avec le sujet et l'objet mais ces derniers n'ont aucun marquage distinctif. En norvégien, il n'y a ni accord ni cas. Dans les trois exemples, la place des arguments ne suffit pas à identifier les fonctions. Quel serait alors le critère pertinent, sachant que les langues varient également en fonction de l'organisation linéaire des constituants ? (on abordera ce point en typologie implicationnelle). La place apparaît comme un critère secondaire, jamais comme un critère suffisant.

En outre, le double marquage casuel, ainsi que l'accord du verbe à la fois avec le sujet et l'objet, suffit à identifier les fonctions syntaxiques. L'ordre des constituants dans ces langues admet par conséquent des variations comme le montrent les deux exemples suivants du japonais :

JAPONAIS

*Hanako **ga** susi **o** tabeta*
 Hanako **NOM** sushi **ACC** a mangé
Hanako a mangé des sushis

*susi **o** Hanako **ga** tabeta*
 sushi **ACC** Hanako **NOM** a mangé
Hanako a mangé des sushis

Notons ici que le nom des particules (nominatif et accusatif) n'est évidemment pas un critère pour savoir si le japonais est une langue nominative ou ergative (en dehors de tout autre critère, on pourrait tout aussi bien utiliser les termes ergatif et absolutif).

Puisque la place n'est jamais un critère suffisant, une autre propriété s'avère nécessaire pour la distinction entre le type nominatif et le type ergatif. Cette propriété s'appuie sur les deux précédentes et prend en compte cette fois le sujet des phrases intransitives. La question est alors la suivante : le sujet des phrases intransitives se comporte-t-il comme le sujet ou comme l'objet des phrases transitives ? Étant donné que la référence aux fonctions syntaxiques peut être quelque peu gênante (puisqu'on est amené à dire qu'un sujet peut être comme un objet), on préférera la formulation suivante : l'unique argument des phrases intransitives se comporte-t-il comme le sujet ou comme l'objet des phrases transitives ?

À nouveau deux prototypes :

Langue nominative : SVO et "S"V

Langue ergative : SVO et "O"V

Rappel : l'ordre des constituants n'est pas pertinent dans ces structures.

Les deux caractérisations sont les suivantes :

Langue nominative

Une langue nominative est une langue dans laquelle l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme le sujet d'une phrase transitive.

Langue ergative

Une langue ergative est une langue dans laquelle l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme l'objet d'une phrase transitive.

Dans les structures précédentes, j'ai utilisé les guillemets pour l'unique argument des phrases intransitive : cette notation doit se lire "se comporte comme". Je reviendrai sur ce point plus loin.

Exemples illustrant les deux prototypes :

LANGUE NOMINATIVE, TURC	<i>ben evi gördüm</i> je+NOM maison+ACC voir+PAS+1SG <i>J'ai vu la maison</i>
	<i>ben geldim</i> je+NOM venir+PAS+1SG <i>Je suis venu</i>
LANGUE ERGATIVE, AVAR	<i>Jasas vas vec:ula</i> fille+ERG fille 3SG/F+complimenter <i>La fille complimente le garçon</i>
	<i>Vas vekerula</i> garçon+ABS 3SG/M+courir <i>Le garçon court</i>

En turc, le sujet de la phrase transitive s'accorde avec le verbe, il est non marqué (nominatif) et l'objet est marqué par l'accusatif (-i). L'unique argument de la phrase intransitive se comporte comme le sujet de la phrase transitive : accord du verbe au moyen des mêmes désinences (ici -m pour la première personne) et absence de suffixe casuel (= nominatif).

En avar, l'objet de la phrase transitive s'accorde avec le verbe, il est non marqué (= absolutif) et le sujet est marqué par l'ergatif (-as). L'unique argument de la phrase intransitive se comporte comme l'objet de la phrase transitive : accord du verbe au moyen des mêmes désinences (ici v- pour la troisième personne du masculin singulier) et absence de suffixe casuel (= absolutif).

Revenons maintenant sur le cas des trois langues précédentes, japonais, swahili et norvégien. Les trois phrases transitives précédentes sont mises en rapport maintenant avec des phrases intransitives :

JAPONAIS	<i>Hanako ga hon o katta</i> Hanako NOM livre ACC achète <i>Hanako achète un livre</i>
	<i>Hanako ga kooen made aruita</i> Hanako NOM parc dans promener+PAS <i>Hanako s'est promené dans le parc</i>
SWAHILI	<i>Hamisi alimwona Juma</i> Hamisi 3SG/S+PAS+3SG/O+voir Juma <i>Hamisi a vu Juma</i>
	<i>Hamisi alifika</i> Hamisi 3SG+PAS+arriver <i>Hamisi est arrivé</i>
NORVÉGIEN	<i>Jeg leser historie</i> je étudier+PRES histoire <i>J'étudie l'histoire</i>

Jeg sover
 Je dormir+PRES
Je dors

Ces trois langues sont des langues nominatives puisqu'à chaque fois l'unique argument de la phrase intransitive a le même cas et/ou les mêmes marques d'accord que le sujet de la phrase transitive. Quant au norvégien, on peut à priori s'en tenir uniquement à la place des constituants syntaxiques ; l'unique argument de la phrase intransitive est placé devant le verbe comme le sujet de la phrase transitive. On ajoutera ici un autre argument en prenant en compte la forme des pronoms. L'unique argument de la phrase intransitive a la même forme pronominale que le sujet de la phrase transitive. L'objet pronominal a une forme différente comme le montre l'exemple suivant :

NORVÉGIEN *Hun ser meg*
 elle voir+PRES moi
Elle me voit

L'argument de la forme pronominale (qui vaut également pour le français) est du même ordre que le cas morphologique puisqu'il repose sur la morphologie des constituants. Dans le cas du norvégien on doit donc prendre en compte une autre structure – la structure pronominale – pour statuer sur son type. Quant au chinois, à défaut d'autres arguments, on s'en tiendra à celui de la place des constituants syntaxiques :

CHINOIS *wo chi rou*
 je manger viande
Je mange de la viande

wo lai
 je venir
Je viens

L'unique argument de la phrase intransitive occupe la même place par rapport au verbe que le sujet de la phrase transitive.

On remarquera, comme cela a déjà été dit à propos du degré de synthèse et du type morphologique, que le norvégien est très proche du chinois également en ce qui concerne l'ordre des constituants et l'identification des fonctions. La différence porte finalement sur le comportement des pronoms (variable selon la fonction en norvégien, et invariable en chinois) et sur la présence d'une marque de présent dans le verbe norvégien, mais pas d'accord du verbe dans les deux langues.

Revenons sur le statut des fonctions syntaxiques. Si l'unique argument d'une phrase intransitive "se comporte comme" le sujet ou l'objet, cela ne signifie pas précisément qu'il est un sujet (dans les langues nominatives) ou qu'il est un objet (dans les langues ergatives). Une étude plus avancée de ce problème permettrait de préciser que le sujet des phrases intransitives dans les langues nominatives partage un ensemble de propriétés avec le sujet des phrases transitives, alors que le sujet des phrases intransitives des

langues ergatives partagent un ensemble de propriétés avec l'objet des phrases transitives. Si cette fois je parle de sujet pour les phrases intransitives, c'est peut-être qu'il est possible de définir les fonctions indépendamment des propriétés morphosyntaxiques qui les caractérisent. On peut remarquer que, dans tous les exemples précédents, la propriété de la place est partagée aussi bien par les langues nominatives que par les langues ergatives ; l'unique argument des phrases intransitives est placé devant le verbe. En grammaire générative, la position sujet se définit sur le plan général par une relation structurale hiérarchisée qui ne tient compte ni de l'accord ni du cas. Les trois propriétés – cas, accord et place – représentent des critères utiles à l'identification des fonctions mais pas à leur définition sur le plan général, sans référence à une structure de langue particulière. Autre chose, on ne peut pas assimiler simplement l'unique argument des phrases intransitives des langues ergatives à un objet, car il lui manque des propriétés sémantiques. Ce qui correspond à un objet syntaxique, se définit sur le plan lexical notamment par la propriété d'être concerné ou d'être affecté par le procès décrit par le verbe, ce qui présuppose donc un agent ou une cause efficiente.

Pour ne pas nous engager sur la question du statut théorique des fonctions syntaxiques, des propriétés lexicales des verbes et des propriétés interprétatives des arguments, nous nous en tiendrons à la formulation : l'unique argument des verbes intransitifs "se comporte comme".

4.3.4. SYSTÈMES HYBRIDES ET TYPE BINAIRE

La typologie fondée sur les propriétés morphosyntaxiques du sujet et de l'objet dans les phrases transitives et intransitives est-elle binaire ? Autrement dit, toutes les langues du monde se répartissent-elles obligatoirement en deux types complémentaires ? En fait, les prototypes proposés précédemment admettent des écarts, et il est donc envisageable de prévoir une fois de plus plusieurs cas de figure. Pour l'essentiel, je reprendrai ici la typologie de Lazard [1994] en l'adaptant à la démarche suivie ici pour la notation des structures et en la discutant. Cette typologie comportent cinq types ; deux prototypes et trois formes hybrides. Précisons d'emblée que Lazard ne parle pas de cinq types de langues mais de cinq types de structures. Autrement dit, la typologie qu'il propose ne classe pas les langues mais classe les structures que l'on peut observer dans les langues. Les commentaires que je propose montreront que les cas hybrides n'excluent pas la possibilité d'une partition des langues en deux types. Rappelons tout d'abord les deux prototypes :

type nominatif : SVO et "S"V

type ergatif : SVO et "O"V

Type nominatif

Une langue nominative est une langue dans laquelle l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme le sujet d'une phrase transitive.

Type ergatif

Une langue ergative est une langue dans laquelle l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme l'objet d'une phrase transitive.

Les trois types hybrides répondent aux structures suivantes :

type neutre :	SVO et "SO"V
type mixte :	SVO et "S"V/"O"V
type disjoint :	SVO et XV

Type neutre

Une langue présente un type neutre si l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme le sujet et l'objet d'une phrase transitive.

Type mixte

Une langue présente un type mixte si l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) comme le sujet ou comme l'objet d'une phrase transitive.

Type disjoint

Une langue présente un type disjoint si l'unique argument des phrases intransitives se comporte (vis à vis du cas morphologique et/ou de l'accord) ni comme le sujet ni comme l'objet d'une phrase transitive.

Pour le type neutre, tout repose sur l'importance que l'on accorde à la place des constituants dans les langues dépourvues de procédés morphosyntaxiques, ou pour lesquelles nous ne disposons pas de descriptions suffisantes. Si l'on ne tient pas compte de la place du sujet et de l'objet dans les phrases transitives, et de la place de l'unique argument des phrases intransitives, alors les exemples norvégiens et chinois relèvent du type neutre dès lors que l'unique argument ne présente aucune propriété différentielle par rapport aux arguments de la phrase transitive. Notons cependant que le norvégien présente dans son système pronominal une preuve suffisante pour y voir le comportement d'une langue nominative. Quant au chinois, il est impossible, comme on l'a vu précédemment, de faire appel à la forme des pronoms dans les deux phrases étant donné que ces derniers sont invariables du fait du caractère isolant de cette langue. De ce point de vue, l'exemple chinois est une illustration du type neutre (Lazard donne un exemple birman qu'il est inutile de reprendre ici). Si l'on tient compte maintenant du critère de la place des constituants (à défaut d'autres arguments) alors le norvégien et le chinois sont l'un et l'autre du type nominatif, car l'unique argument de la phrase intransitive se place devant le verbe comme le sujet de la phrase transitive. Du coup, le type neutre n'existerait pas dans les langues du type SVO (et OVS) – où chacun des arguments du verbe transitif est placé de part et d'autre du verbe parce que l'unique argument des verbes intransitifs se place forcément à gauche ou à droite du verbe. Le type neutre vaudrait donc surtout pour les langues de type SOV (et VSO) dans lesquelles sujet et objet occupent la même place relativement au verbe. Si dans une phrase intransitive, l'unique argument précède le verbe, on ne peut pas savoir s'il doit cette propriété au sujet ou à l'objet de la phrase transitive car tous les deux sont placés devant le verbe.

Le type mixte est présent en géorgien et dans les langues australiennes. En géorgien, on a le type nominatif au présent et le type ergatif à l'aoriste. Dans les langues australiennes, comme le dyirbal, la structure ergative concerne les phrases avec syntagmes nominaux tandis que la structure nominative est systématique dès lors qu'il s'agit de pronoms. Les exemples suivants sont empruntés à Lazard :

DYIRBAL	<i>balan d'ugumbil bangul yaṛaṅgu balgan</i> ART/ABS femme/ABS ART/ERG homme/ ERG frapper <i>L'homme frappe la femme</i>
	<i>bayi yaṛa baniṅu</i> ART/ABS homme/ABS venir <i>L'homme vient</i>
ARTicle, ABSolutif, ERGatif	

Dans la phrase transitive avec arguments nominaux, le sujet est à l'ergatif et l'objet est non marqué (absolutif). Dans la phrase intransitive, l'unique argument est à l'absolutif comme l'objet de la phrase transitive. La structure est du type ergatif.

DYIRBAL	<i>ṛad'a ṛinuna balgan</i> moi/NOM toi/ACC frapper <i>Je te frappe</i>
	<i>ṛinda baniṅu</i> toi/NOM venir <i>Tu viens</i>
NOMinatif, ACCusatif	

Dans la phrase transitive avec pronoms, le sujet est au nominatif tandis que l'objet est à l'accusatif. Dans la phrase intransitive, l'unique argument du verbe se comporte comme le sujet de la phrase transitive. La structure est donc du type nominatif.

Le type disjoint est représenté par l'hindi (exemples empruntés à Lazard) :

HINDI	<i>billī-ne ghore-ko dekh-ā hai</i> chat- ERG cheval- ACC voir/PPRES- F AUX/3SG <i>Le chat voit le cheval</i>
	<i>billī gir-ī hai</i> chat tomber/PPAS- F AUX/ 3SG <i>Le chat est tombé</i>
ERGatif, ACCusatif, PPRES = participe présent, AUXiliaire, Masculin, Féminin	

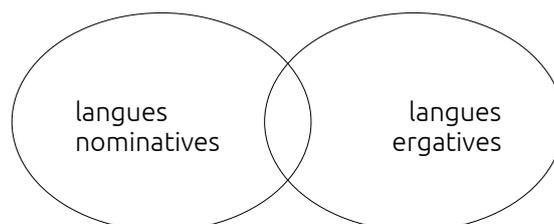
Dans la phrase transitive le sujet est signalé par le suffixe ergatif *-ne* et l'objet est également marqué par un suffixe (accusatif = *-ko*). Le verbe s'accorde en personne, nombre et genre avec le sujet (*chat* est un mot féminin en hindi). Le fait que l'accord se répartisse entre l'auxiliaire et le participe passé importe peu ici. Dans la phrase intransitive, l'unique argument partage avec le sujet de la phrase transitive la propriété de l'accord. Par contre, il n'a ni le cas du sujet ni le cas de l'objet.

Comme le signale Lazard la structure nominative existe également en hindi. Il suffit pour cela de changer le temps du verbe :

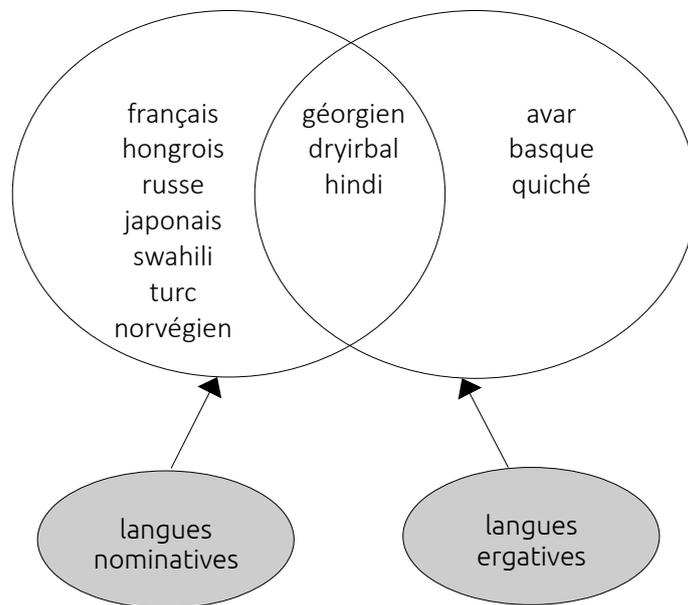
HINDI	<i>billī ghore-ko dekht-ī hai</i> chat cheval- ACC voir/PPAS- F AUX/ 3SG <i>Le chat a vu le cheval</i>
	<i>billī gir-ī hai</i> chat tomber/PPAS- F AUX/ 3SG <i>Le chat est tombé</i>
PPAS=participe passé	

Dans la phrase transitive, le sujet est non marqué, s'accorde avec le verbe en personne, nombre et genre et l'objet est marqué par le cas accusatif. Dans la phrase intransitive, l'unique argument est non marqué et il s'accorde avec le verbe dans les mêmes conditions que le sujet de la phrase transitive. Avec cet exemple, l'hindi répond précisément au prototype des langues nominatives. Les premiers exemples constituent un écart par rapport à ce prototype, à la fois par le comportement du sujet dans la phrase intransitive (la structure est du type disjoint) et par le double marquage des arguments du verbe transitif. L'hindi est donc plutôt une langue qui s'approche du prototype des langues nominatives.

Le type neutre peut être écarté si l'on fait l'hypothèse (raisonnable) qu'il existe toujours un argument possible dans la syntaxe des langues pour savoir si une langue relève du type nominatif ou ergatif (voir le cas du norvégien). Le type disjoint, représenté ici par l'hindi, peut être également écarté puisqu'il existe des structures dans cette langue qui permettent de déterminer si la langue est plutôt de type nominatif ou plutôt de type ergatif. Donc, si l'on retient des cas hybrides uniquement les structures mixtes (une langue peut avoir des structures nominatives et des structures ergatives) alors nous pouvons concevoir la répartition entre langues nominatives et langues ergatives comme deux ensembles non disjoints dont l'intersection correspond aux langues qui ont les deux structurations :



Les langues citées dans cette parties se répartissent ainsi (le nombre d'exemples n'a évidemment aucune valeur indicative) :

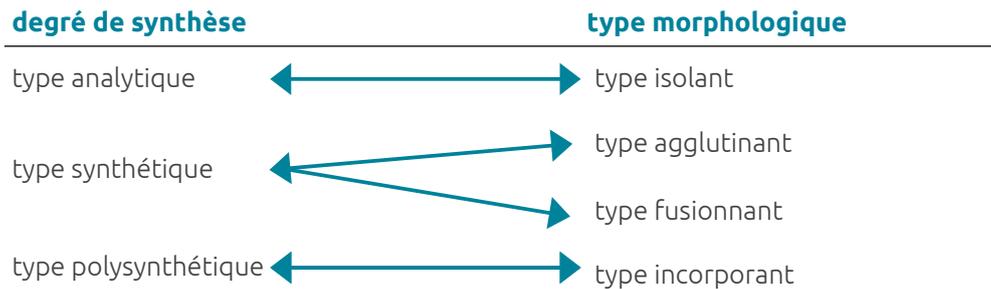


5. TYPOLOGIE POLYTHÉTIQUE

Dans cette partie, nous ne verrons qu'un seul exemple de typologie polythétique. Il s'agit de la typologie de Sapir. La classification des langues qu'il propose constitue en quelque sorte la version la plus aboutie de la typologie à base morphologique.

5.1. TYPOLOGIE MORPHOLOGIQUE DE SAPIR

Nous avons vu précédemment que les langues se répartissent en fonction de leur degré de synthèse et en fonction de leur type morphologique. Ces deux classifications ne sont pas indépendantes car elles reposent l'une et l'autre sur la relation entre morphèmes et mots. Les langues analytiques sont nécessairement des langues isolantes et inversement, les langues isolantes sont nécessairement des langues analytiques, car si dans une langue, un mot ne contient qu'un seul morphème, les morphèmes, en retour, renvoient à des mots simples. De même pour la relation entre le type polysynthétique et le type morphologique incorporant ; si un mot peut contenir plusieurs lexèmes sans qu'il soit question de composition, il y a alors nécessairement un processus d'incorporation. Reste le type synthétique qui correspond à deux modes de réalisation des morphèmes dans le mot : l'agglutination et la fusion.



La relation biunivoque entre le type analytique et le type isolant, ainsi qu'entre le type polysynthétique et le type incorporant a pour effet que les deux termes sont souvent considérés comme synonymes. On peut trouver notamment des descriptions typologiques fondées sur les types isolant, agglutinant, fusionnant et polysynthétique. On se gardera bien de confondre les deux typologies car l'approche n'est pas la même.

Sapir [1921, 1970 pour la traduction française] propose une typologie à base morphologique fondée sur trois critères :

- ◆ le degré de synthèse
- ◆ le type morphologique
- ◆ la nature des concepts (ou le type conceptuel)

Les deux premiers critères ont été développés dans le chapitre précédent. Sapir distingue en fait cinq types morphologiques et non quatre. Il ajoute aux quatre types courants, le type symbolique à propos duquel il précise cependant qu'il est une variété du type flexionnel (type fusionnant). Il correspond en fait au type fusionnant à flexion interne ([cf. § 4 et 5 de la sous-partie 4. 2. 2. 3 Type fusionnant](#)). Cette distinction ne sera pas prise en compte ici. Reste à voir la nature des concepts.

Les trois critères sont liés. Avec les deux premiers critères il s'agit d'évaluer la nature formelle du rapport entre les mots et les morphèmes. Le troisième critère fait intervenir ce que les morphèmes et les mots véhiculent comme information. Il s'agit donc cette fois d'une typologie fondée sur la signification des morphèmes.

Sapir répartit les concepts associés aux morphèmes en quatre groupes selon leur mode de formation et selon leur fonction :

- ◆ les concepts de base (groupe I)
- ◆ les concepts dérivationnels (groupe II)
- ◆ les concepts relationnels concrets (groupe III)
- ◆ les concepts relationnels abstraits (groupe IV)

Les concepts de base et les concepts dérivationnels s'appliquent aux mots alors que les concepts relationnels s'appliquent aux relations entre constituants syntaxiques.

Les concepts de base forment des unités indépendantes dans les langues. Ces concepts correspondent à des entités, des procès, des propriétés, des actions, des états... Les mots *maison*, *regarder*, *jeune*, *lire*, *joie*... sont des mots indépendants qui expriment des concepts de base.

Les concepts dérivationnels sont représentés dans les langues par des morphèmes ou des traits morphologiques associés à un lexème. La signification globale du mot est alors le produit d'une signification quelconque associée au lexème de base et d'une signification incomplète associée au morphème dérivationnel. Quelques exemples de morphèmes dérivationnels en français dans leur version incomplète (la part correspondant au lexème manquant est représentée par les points de suspension) :

-*eur/-euse* : celui/celle qui...

-*able/-ible/-uble* : qui peut être...

-*dé-/dés-* : action contraire de...

Ces bouts de signification constituent des concepts dérivationnels dès lors que leur emploi implique une construction avec un lexème de base :

construction : [*chant*]+[*eur/euse*] = *chanteur/chanteuse*
signification : [celui/celle qui] + [chante]

construction : [*lav*]+[*-able/-ible/-uble*] = *lavable*
signification : [ce qui peut être] + [laver]

construction : [*dé-/dés-*]+[*monter*] = *démonter*
signification : [action contraire de] + [monter]

Les concepts relationnels concrets ont une double fonction dans la langue ; ils véhiculent un contenu sémantique (ils sont concrets) et ils indiquent une relation de dépendance entre des mots d'un énoncé. Les prépositions spatiales (*sur, devant, contre, après, sous...*) indiquent un certain type de relation spatiale – elles ont donc une signification concrète – et elles sont de nature relationnelle dans la mesure où elles mettent en relation deux entités. Les marques de personne sur le verbe sont également de ce type car elles reprennent l'information concernant les arguments du verbe (personne) et elles indiquent les fonctions de ces arguments (sujet et/ou objet). Le morphème de temps des verbes est également un concept relationnel concret, car il donne une information temporelle sur le procès décrit par le verbe (le morphème de temps a donc un contenu sémantique) et il structure l'énoncé en définissant une structure de phrase (en français, toute phrase implique un verbe conjugué, et tout verbe conjugué implique une structure de phrase).

Les concepts relationnels abstraits sont comparables aux précédents mais sans contenu sémantique. La préposition française "de" n'a pas de signification ; elle établit simplement une relation de complémentation entre deux constituants. La même chose pour la conjonction de subordination "que" ; sa fonction est de relier deux phrases dont l'une est enchâssée dans l'autre. Les suffixes de cas accusatif et de cas génitif des langues à cas morphologiques sont également de nature purement relationnel. Le premier introduit une relation de complémentation entre un verbe et un constituant nominal et le second, une relation de complémentation entre deux noms. Dans les exemples finnois suivants le suffixe de génitif (*-n*) n'est associé à aucune signification particulière :

FINNOIS

presidentin auto = la voiture du président

presidentin vaimo = la femme du président

presidentin puhe = le discours du président

Il est évident que toutes les langues ont des concepts de base et des concepts relationnels abstraits (groupes I et IV) ; les premiers sont représentés par des mots simples (toutes les langues ont des mots simples) et les seconds, par des mots ou des morphèmes grammaticaux (toutes les langues ont des mots-outils sans signification ; adpositions, cas, conjonctions, particules...). Les deux autres types de concepts peuvent être présents ou absents, ou bien encore, seulement un des deux. Les différentes combinaisons définissent quatre types de langues :

- ◆ type A : langues ayant uniquement des concepts des groupes I et IV
- ◆ type B : langues ayant des concepts des groupes I, IV et II
- ◆ type C : langues ayant des concepts des groupes I, IV et III
- ◆ type D : langues ayant des concepts des groupes I, IV, II et III

Les types de langues ainsi définis sont désignés respectivement :

- ◆ type A : langues simples et pures
- ◆ type B : langues complexes et pures
- ◆ type C : langues simples et mixtes
- ◆ type D : langues complexes et mixtes

Les termes *simple* et *complexe* font référence à la distinction entre concepts de base et concepts dérivationnels, tandis que les termes *pure* et *mixte* font référence à la distinction entre concepts relationnels abstraits et concepts relationnels concrets.

Combiné avec le degré de synthèse et le type morphologique, le critère de la nature des concepts est illustré par les prototypes suivants :

- chinois = type {analytique, isolant et A}
- turc = type {synthétique, agglutinant et B}
- swahili = type {synthétique, agglutinant et C}
- latin = type {synthétique, fusionnant et D}

Le chinois est du type A car cette langue n'utilise pas la dérivation pour former des mots complexes (sinon elle serait synthétique) mais uniquement la composition. Quelques exemples chinois de mots composés (transcription pinyin). Voir également les [exemples comparables du vietnamien en 4. 2. 1. 1.](#)

CHINOIS

hǎi yuán : mer + personne = marin

háng tiān yuán : navigation + ciel + personne = astronaute/taïkonaute

diàn nǎo : électricité + cerveau = *ordinateur*

Par ailleurs, un même mot en chinois peut correspondre à un concept de base ou à un concept purement relationnel. Selon son emploi, le mot *gěi* est soit un verbe signifiant "donner" soit une préposition introduisant un bénéficiaire (= "à") :

CHINOIS	<i>nǐ gěi tā shū</i> tu donner lui livre <i>tu lui donnes un livre</i>
	<i>nǐ jì yí fēng xìn gěi tā</i> tu envoyer une lettre à lui <i>tu lui envoies une lettre</i>

Le turc est du type B car cette langue abonde en dérivation. Quelques exemples :

TURC	<i>killi</i> = <i>kil+li</i> : argile+suffixe adjectival = <i>argileux</i>
	<i>ağrısız</i> = <i>ağrı+sız</i> : douleur+suffixe privatif = <i>indolore</i>
	<i>süpürücü</i> = <i>süpür+ücü</i> : balayer+suffixe agentif = <i>balayeur</i>
	<i>hastalanmak</i> = <i>hasta + lan + mak</i> : malade+suffixe verbal+infinitif = <i>devenir malade</i>

Par ailleurs, les relations syntaxiques sont plutôt de nature abstraite car elles ne font pas intervenir de morphèmes avec contenu sémantique. La dépendance de l'adjectif par rapport au nom ne fait pas intervenir d'accord en nombre ; l'adjectif est invariable et seule sa place – devant le nom – permet de l'identifier comme épithète :

TURC	<i>güzel kız</i> joli fille <i>une jolie fille</i>
	<i>güzel kızlar</i> joli fille+PL <i>de jolies filles</i>

Autre exemple d'emploi minimaliste des morphèmes grammaticaux en syntaxe turque : dans une succession de verbes se rapportant à un même sujet, seul le dernier verbe est conjugué. Dans l'exemple qui suit, seul le dernier verbe contient les suffixes de temps (imparfait) et de personne (2PL) :

TURC	<i>erken kalkar, çabuk giyinir, işinize giderdiniz</i> tôt se lever vite s'habiller votre travail aller+IMP+2PL <i>Vous vous levez tôt, vous vous habilliez vite (et) vous alliez à votre travail</i>
-------------	---

Le swahili est une langue du type C : langue simple et mixte. La dérivation lexicale est peu productive dans cette langue. La formation des noms d'agent ou d'appartenance à un groupe ne font pas appel à un affixe dérivationnel comme dans beaucoup de langues (*chanteur, islandais...*) mais à la composition (mise en relation de deux mots indépendants) :

SWAHILI	<i>mwana-shule</i> : enfant+école = <i>écolier</i>
	<i>mwana-sharia</i> : enfant+loi = <i>homme de loi</i>
	<i>mwana-nchi</i> : enfant+pays = <i>citoyen</i>

Les noms d'appartenance à un groupe sont systématiquement formés au moyen de mot *mwana* "enfant" pour indiquer la relation avec une personne.

Exemple de nom d'agent formé à partir d'un verbe :

mpiga-picha : frappe-image = *photographe*

Le swahili est une langue avec des concepts relationnels concrets ; ce point est illustré par l'omniprésence des phénomènes d'accord (en nombre et en classe nominale). Dans l'exemple suivant, l'information sur le nombre du sujet est reprise dans chacun des constituants syntaxiques. Les marques d'accord sont soulignées. La glose contient l'information relative au nombre (PL) et à la classe nominale (les classes nominales des langues bantou sont identifiées par un chiffre) :

SWAHILI	<u>wana</u>	<u>wangu</u>	<u>wale</u>	<u>wamekufa</u>
	PL ² +enfant	PL ² +mien	PL ² +ce	PL ² +PARFAIT+être_mort
	"ces miens enfants sont morts"			

Le latin est une langue de type D, car cette langue est à la fois riche en dérivation et sa syntaxe comporte également de nombreuses relations d'accord qui reprennent notamment le nombre. Nombre, genre grammatical et cas morphologique sont fusionnés dans la désinence des noms et des adjectifs (langue fusionnante à flexion externe). Exemples de dérivation :

LATIN	<i>cantor</i> : chanter + suffixe agentif = <i>chanteur</i>
	<i>petitor</i> : chercher + suffixe agentif = <i>candidat</i>
	<i>quercetum</i> : chêne + suffixe collectif = <i>chênaie</i>
	<i>amicus</i> : aimer + suffixe adjectival = <i>ami</i>
	<i>mistesco</i> : doux + suffixe verbal = <i>s'adoucir</i>

L'accord en nombre, genre et cas est particulièrement explicite dans l'exemple classique suivant (syntagme nominal dont la tête est un nom masculin au nominatif pluriel), où chaque mot – comme en swahili – comporte le même affixe :

LATIN	<i>illi</i> ce+[PL...]	<i>boni</i> bon+[PL...]	<i>domini</i> maître+[PL...]
	<i>ces bons maîtres</i>		

Les exemples proposés par Sapir constituent des prototypes. Les langues retenues représentent des exemples francs du type illustré. Mais comme le précise Sapir, la frontière entre les types A et B d'une part et les types C et D d'autre part – opposition entre le type pur et le type mixte – est ténue dans la mesure où la différence entre ces deux types repose sur l'évaluation du contenu sémantique des morphèmes grammaticaux impliqués dans les relations syntaxiques. Nous n'avons retenu ici que le nombre, le temps et la personne car ces catégories ont clairement un contenu sémantique. En revanche, c'est plus délicat lorsqu'il s'agit d'évaluer le contenu sémantique de la catégorie du cas parce que cette catégorie contient aussi bien des cas sans contenu sémantique (nominatif, accusatif...) que des cas avec contenu sémantique (il s'agit alors de cas morphologiques équivalents à des prépositions (locatif, instrumental...), sans compter les cas dont le contenu sémantique est variable selon les langues ; les cas ergatif et génitif notamment peuvent être purement abstraits (relation entre sujet et un verbe pour l'ergatif, et relation entre deux noms pour le génitif) ou être un tant soit peu concrets lorsque leur emploi est limité à l'agent pour l'ergatif et à la possession pour le génitif.

Par ailleurs, la distinction entre les types A et C d'une part et les types B et D d'autre part – opposition entre le type simple et le type complexe est également une affaire de gradation car l'évaluation de la dérivation lexicale est liée au rendement des affixes dérivationnels. Une langue ayant de nombreux dérivés mais dont le rendement est faible n'est pas considérée comme une langue de type complexe. La difficulté est évidemment accrue lorsque la dérivation n'est pas homogène ; présence à la fois de dérivés productifs et de dérivés à faible rendement. En dérivation française, par exemple, les suffixes dérivationnels productifs tels que *-eur/euse* (*chanteur, blogueur...*) et *-iser/ifier* (*fragiliser, sénégaler...*) côtoient des suffixes dérivationnels obsolètes tels que *-aille* (*trouvaille, fiançailles*) ou *-aison* (*salaison, crevaision...*).

Pour comprendre la notion de productivité d'un suffixe dérivationnel, il suffit de comparer le cas du suffixe agentif français *-eur/-euse* avec son équivalent en finnois, langue dans laquelle la nature complexe du type conceptuel ne fait aucun doute. Le suffixe d'agent est *-ja/-jä* et il peut apparaître quasiment avec tous les verbes, même si la notion d'agentivité est faible :

FINNOIS	<i>ajaa</i> conduire, <i>ajaaja</i> conducteur/conductrice
	<i>myy</i> vendre, <i>myyjä</i> vendeur/vendeuse
	<i>rakastaa</i> aimer, <i>rakastaja</i> amant(e)
	<i>odottaa</i> attendre, <i>odottaja</i> celle qui attend (un enfant)

Comme le montre la traduction, il n'est pas possible d'ajouter systématiquement le suffixe *-eur/-euse* pour obtenir un nom d'agent en français (celui qui apprend le français doit apprendre si le mot existe ou non). Donc, en français, on ne peut pas dire que tous les verbes acceptent le suffixe d'agent *-eur/-euse*, et d'autre part, le suffixe *-eur/-euse* ne s'applique pas uniquement à des agents (humains, par définition) mais également à des instruments (*téléviseur, photocopieur, photocopieuse, agrafeuse...*).

Le tableau des types de structure linguistique d' E. Sapir (1921), [pp. 150-151](#) (document 4) reprend tous les cas de figure de sa typologie polythétique. Dans ce document, le terme *mêlé* remplace *mixte*, la *technique* renvoie au *type morphologique* et le *caractère* renvoie au *degré de synthèse*. Dans les colonnes des groupes, les lettres minuscules signalent le type morphologique utilisé pour l'expression du concept correspondant :

- a = isolation
- b = agglutination
- c = fusion externe
- d = fusion interne

Ainsi pour le turc, les concepts du type II et IV (langue complexe et pure) sont représentés dans la langue par des mots qui se présentent sous la forme d'une agglutination de morphèmes dérivationnels ou flexionnels ajoutés à une base lexicale (b). Lorsqu'il y a plusieurs lettres, l'ordre signale une hiérarchisation des techniques. Le latin, par exemple, est marqué "c, d" pour les groupes II et III ; ce qui signifie qu'il utilise plutôt la fusion externe que la fusion interne pour les mots dérivés et pour les mots fléchis.