

# Structuration du lexique et principe d'économie : le cas des ethniques

Michel Roché

CLLE-ERSS (Axe DUMAL)  
CNRS et Université de Toulouse-Le Mirail  
[mroche@univ-tlse2.fr](mailto:mroche@univ-tlse2.fr)

## 1 Introduction

Les noms et les adjectifs que nous regrouperons sous l'étiquette commune d'« ethniques » posent à la morphologie constructionnelle des problèmes qui passent généralement inaperçus dans la mesure où ils n'affectent pas la forme des unités concernées. Comment construire l'un par rapport à l'autre *Français* et *français*, *Russe* et *russe* ? Comment relier ces lexèmes à un nom de pays qui est tantôt en amont (*France*) tantôt en aval (*Russie*) ? Comment rendre compte du fait que, quelle que soit sa forme, l'adjectif renvoie aussi bien au pays (*les paysages français*, *la plaine russe*) qu'à ses habitants (*le caractère français*, *l'âme russe*) ? Dans une publication antérieure (Roché, 2005), nous avons ébauché une description d'ensemble du système et évoqué le rôle qu'y joue la conversion. Après avoir rappelé ces données, on se propose de poursuivre ici cette exploration en observant, dans un premier temps, la double « dimension paradigmatique » (van Marle, 1985) dans laquelle s'intègrent ces formations : celle du paradigme lexical que constituent les ethniques dans leur ensemble et celle du paradigme dérivationnel qui regroupe chacun d'entre eux avec le nom de pays, le nom de langue et les adjectifs correspondants. On s'intéressera ensuite à deux séries de dérivés construits sur les ethniques – préfixés en *anti-* et suffixés en *-isme* et en *-iste* – qui héritent de certaines caractéristiques du système et manifestent à leur tour le rôle de la marque morphologique dans la structuration du lexique.

## 2 Le système des ethniques en français

On appellera « ethniques » les noms de peuples (*Russe*, *Tzigane*), les gentilés construits sur un nom de pays ou de région historique ou géographique (*Italien*, *Piémontais*) et les adjectifs correspondants (*russe*, *tzigane*, *italien*, *piémontais*), à l'exclusion des gentilés construits sur un nom de ville ou de circonscription administrative. Le corpus constitué pour ce travail rassemble 1090 dénominations, dont 830 suffixales. Pour simplifier les notations, les noms de personnes et les adjectifs seront donnés uniquement sous leur forme de citation (masculin singulier). Il va de soi que, sauf à être épécènes, ils sont systématiquement variables en genre.

Le système que constituent ces lexèmes peut se présenter morphologiquement, en français contemporain, sous quatre formes :

- (1) *Tzigane* N → *tzigane* Adj 'de Tzigane'
- (2) *Russe* N → *russe* Adj 'de Russe'  
→ *Russie* → *russe* Adj 'de Russie'
- (3) *France* → *français* Adj 'de France'  
→ *Français* N → *français* Adj 'de Français'
- (4) *Hongrois* N → *hongrois* Adj 'de Hongrois'  
↓  
*Hongrie* → *hongrois* Adj 'de Hongrie'

Dans le type (1), il n'y a pas de nom de pays (ou il n'est pas usuel). Dans le type (2), le nom de pays est formé sur le nom de peuple, par suffixation (*Russie*, *Arabie*, *Birmanie*, *Turquie*...) ou par composition (*Angleterre*, *Pays Basque*, *Thaïlande*, *Kurdistan*...). Dans le type (3), le nom des habitants est formé, par

suffixation, sur le nom de pays (*Français, Zaïrois, Tunisien, Andorran, Chypriote, Emirati...*). Dans le type (4), les finales de l'un et de l'autre ne permettent pas d'orienter la relation constructionnelle. Formellement, les deux dénominations peuvent être identiques (*Corse / Corse, Suisse / Suisse...*), prendre les apparences d'une alternance flexionnelle (*Lorrain / Lorraine, Argentin / Argentine...*), faire alterner deux suffixes par troncation réciproque (*Hongrois / Hongrie, Albanais / Albanie...*), opposer une finale quelconque à une finale suffixoïde (*Breton / Bretagne, Catalan / Catalogne...*), etc. Le plus familier est le type (3), parce qu'il est commun aux gentilés formés sur les noms de villes et qu'il correspond à la vision contemporaine du monde, structuré en Etats bien délimités. Mais le type (1) et le type (2) sont plus fondamentaux, et ils ont longtemps été majoritaires. Le type (4) résulte de divers aléas historiques sur lesquels nous reviendrons.

Dans tous les cas, on observe un certain nombre de points communs que nous présenterons schématiquement (pour une argumentation plus détaillée, voir Roché (2005)) :

- Quelle que soit la relation génétique entre le nom de personne et le nom de pays, quand il existe, les deux dénominations sont, d'un point de vue fonctionnel, réciproquement motivées, comme dans le type (4). La perception qu'en ont les locuteurs, reflétée par les dictionnaires, est la même dans tous les cas de figure. *Russe* est défini par le *Grand Robert* comme 'de Russie', malgré une relation constructionnelle inverse.

(5)	<i>Hongrois</i> 'habitant de la Hongrie'	↔	<i>Hongrie</i> 'pays des Hongrois'
	<i>Russe</i> 'habitant de la Russie'	↔	<i>Russie</i> 'pays des Russes'
	<i>Français</i> 'habitant de la France'	↔	<i>France</i> 'pays des Français'

- Les adjectifs sont des adjectifs de relation et, pour cette raison, ne peuvent pas constituer le primitif. C'est le nom de personne qui est premier, non seulement dans les types (1) et (2) où les adjectifs sont formés par conversion mais aussi dans le type (3) où *Français* N n'est pas une nominalisation de *français* Adj mais construit parallèlement sur *France*.
- Dans tous les cas, les adjectifs ont la même forme que le nom de personne : *Français* N / *français* Adj, *Russe* N / *russe* Adj, etc. Les exceptions (*germanique, hispanique, helvétique...*) sont des emprunts au latin utilisés dans des conditions très particulières (*infra*, § 2.3).
- L'adjectif qui renvoie au nom de personne (*français* dans *le caractère français, russe dans l'âme russe*) et celui qui renvoie au nom de pays (*français* dans *les paysages français, russe dans la plaine russe*) ont systématiquement la même forme. Aussi apparaissent-ils, lexicalement, comme un seul « mot ». Mais dans la mesure où il s'agit d'adjectifs de relation, qui n'ont de raison d'être que par rapport au nom, ils constituent bien, d'un point de vue constructionnel, deux unités distinctes. C'est la raison pour laquelle, dans les schémas (2) et (3), on a posé *Russie* → *russe* Adj 'de Russie' et *Français* N → *français* Adj 'de Français' comme des relations constructionnelles.
- Le nom de langue, quand il existe, est presque toujours une nominalisation par conversion de l'un de ces deux adjectifs – *français* Adj 'de Français' → *français* N 'langue des Français' – et l'adjectif de relation qui y renvoie – *français* 'du français' (*le vocalisme français*) – est formé lui aussi par conversion.

Ainsi, en dépit des différences entre les divers types et des multiples variantes de détail, le système contemporain des ethniques et des noms de langues est remarquablement unifié et cohérent. Cette structure est le résultat de la pression lexicale qui s'exerce sur le lexème construit du fait de son appartenance à deux paradigmes :

- un *paradigme dérivationnel* (Bauer, 1997), équivalent synchronique de la « famille de mots », qui regroupe les mots construits sur une même base, directement ou indirectement (*France, Français, français, franciser, francophone...*) ;
- un *paradigme lexical* qui tend à être, plus précisément, un *paradigme lexico-morphologique* quand les lexèmes appartenant à une même classe sémantico-référentielle ont en commun une même marque formelle (*français, irlandais, congolais, javanais...*)<sup>1</sup>.

Nous verrons successivement comment ces deux types de paradigmes se sont constitués pour les ethniques, et comment ils peuvent entrer en conflit.

## 2.1 Des paradigmes dérivationnels économes

Même quand il est complet et qu'il comporte un nom de pays et un nom de langue, le paradigme dérivationnel dans lequel s'inscrit un ethnique est, en français comme dans les autres langues romanes, particulièrement économe. Six lexèmes (nom de peuple ou d'habitants, nom de pays, nom de langue, plus les trois adjectifs de relation correspondants) partagent deux formes seulement, voire une seule (*Corse / corse*). Comment en est-on arrivé là ?

Au départ, il y a le fait que les principaux suffixes spécialisés (*-ais, -ois*) et privilégiés (*-ien, -ain*) sont sous-spécifiés quant à l'opposition Nom / Adjectif. Cela ne va pas de soi *a priori*. L'allemand oppose systématiquement un suffixe *-er* nominal à un suffixe *-isch* adjectival (*Österreicher* N 'Autrichien' / *österreichisch* Adj 'autrichien'). En latin, si les dérivés en *-anus* sont à la fois nom et adjectif, ceux en *-icus* sont seulement adjectifs ; et lorsque les deux sont en concurrence sur une même base, ils ont tendance à se spécialiser (*asianus* comme nom désignant une personne, par exemple, et *asiaticus* comme adjectif). En français donc, formés avec le même suffixe, *Français* N 'habitant de la France' et *français* Adj 'de la France' ont la même forme.

C'est sans doute cette caractéristique qui a entraîné la deuxième : la formation d'un adjectif de relation par conversion quand le primitif est le nom de peuple : *Russe* N → *russe* Adj 'de Russe'. Ceci non plus n'était pas évident. Le recours à la conversion, pour ce type de formations, est généralement soumis à des conditions qui ne sont pas toutes réunies ici<sup>2</sup>. Ce qui nous fait dire que la conversion est en l'occurrence un moyen de sauvegarder le parallélisme avec le type *Français* N / *français* Adj. L'allemand (*Russe* N 'Russe' → *russisch* Adj 'de Russe'), le latin (*Gallus* N 'Gaulois' → *gallicus* Adj 'de Gaulois') recourent systématiquement, dans ce cas, à la suffixation.

Troisième caractéristique : une fois posé ce premier adjectif de relation, il est réinvesti, ou plutôt sa forme est réinvestie pour l'autre adjectif de relation. Dans le type (3), la forme construite – par suffixation – pour *français* 'de France' va être réutilisée pour *français* 'de Français'. Dans le type (2), la forme construite – par conversion – pour *russe* 'de Russe' va être réutilisée pour *russe* 'de Russie'. Ce phénomène, à la différence des précédents, semble assez général. Il y a quelques exemples de suffixation en chaîne quand la base est un nom de ville – it. *Roma* → *Romano* N 'Romain' → *romanesco* Adj 'de Romain' (distinct de *romano* Adj 'de Rome', mais réservé à certains emplois), al. *Berlin* → *Berliner* N 'Berlinois' → *berlinerisch* Adj 'de Berlinois' (distinct de *berlinisch* Adj 'de Berlin') –, mais l'allemand, comme l'italien ou le français, opère généralement le même transfert. Quand le nom de personne est construit (*Österreicher* 'Autrichien'), l'adjectif construit sur le nom de pays (*österreichisch* 'd'Autriche') renvoie aussi au nom de personne (*österreichisch* 'd'Autrichien'). Quand c'est le nom de pays qui est construit (*Russland* 'Russie'), l'adjectif construit sur le nom de personne (*russisch* 'de Russe') renvoie aussi au nom de pays (*russisch* 'de Russie'). *Österreicherisch* est attesté (à propos de la langue autrichienne), *russländisch* également, mais ils ne sont pas usuels. En néerlandais<sup>3</sup>, où le nom d'habitant est également tantôt construit (*Amerikaan* 'Américain', *Israeliet* 'Israélien'), tantôt non construit (*Noor* 'Norvégien', *Belg* 'Belge'), l'adjectif construit sur ce nom (*Amerikaans, Israelietisch, Noors, Belgisch*) est repris systématiquement pour renvoyer au nom de pays (resp. *Amerika, Israel, Noorwegen, België*). En latin enfin, *gallicus* pourrait, formellement, être construit sur *Gallia* aussi bien que sur *Gallus* mais l'antériorité de *Gallus* invite à y voir le même processus : formé d'abord sur *Gallus*, *gallicus* est réinvesti pour renvoyer à *Gallia*. Symétriquement, *asiaticus* construit sur *Asia* renvoie également à *Asianus*.

Dernière étape : le nom de langue et l'adjectif qui y renvoie. L'un et l'autre, on l'a dit, ont la même forme que les autres adjectifs de relation, quelle que soit la façon dont ils sont construits. Les rares exceptions (*hébraïque*) sont des restes du système latin sur lequel nous reviendrons.

Ainsi, à plusieurs niveaux, se manifeste un « principe d'économie » de portée plus générale, qui fait que la langue, plutôt que de construire une forme nouvelle, réutilise une forme déjà existante.

L'adjectif de relation *électoral*, par exemple, formé sur *électeur*, renvoie normalement à ce nom (les *listes électorales* sont des 'listes d'électeurs'), mais il sert également d'adjectif de relation à *élection* (le *droit électoral* traite de tout ce qui concerne les élections) à la place de *°électionnel*, qui devrait logiquement prendre place dans l'abondante série des dérivés en *-el* construits sur un nom en *-ion*.

- (6) *élire* → *électeur* → *électoral* 'qui concerne les électeurs' (*listes électorales*)  
 ↗ " 'qui concerne les élections' (*droit électoral*)  
*élire* → *élection* → *°électionnel*

Résultat : une discordance entre la compositionnalité formelle et la compositionnalité sémantique du mot construit.

Ce décalage peut en entraîner un autre par rapport à l'instruction propre au suffixe. Hors contexte, *débats éducatifs* serait compris comme '(débats) qui éduquent'. Mais dans la phrase « En France, les débats éducatifs sont trop souvent réduits aux débats sur l'école. » (*Le Monde*, 23.03.2007 : 23), il s'agit de débats 'qui concernent l'éducation'.

- (7) *éduquer* → *éducatif* 'qui éduque' (*jeux éducatifs*)  
 ↗ " 'qui concerne l'éducation' (*débats éducatifs*)  
*éduquer* → *éducation* → *°éducationnel*

Le dérivé en *-if* construit sur la base verbale est réutilisé comme adjectif de relation pour renvoyer au dérivé en *-ion* appartenant au même paradigme. Autre exemple : les « problèmes érectiles » évoqués par un sexologue sur France-Inter (06.01.2005) ne sont pas les problèmes susceptibles de surgir mais ceux qui concernent l'érection.

A côté de ces manifestations ponctuelles, les adjectifs en *-iste* donnent un exemple d'application systématique du principe d'économie comparable à celui des ethniques. Leur valeur propre est identique à celle des noms en *-iste* : axiologique (*esclavagiste* 'favorable à l'esclavage') ou agentive (*récidiviste* 'qui récidive'). Mais ils servent également d'adjectif de relation pour renvoyer tantôt à la base (*l'époque tsariste*), tantôt au dérivé en *-isme* (*les théories tsaristes*), tantôt au dérivé nominal en *-iste* (*les milieux tsaristes*). Vérification indirecte : il n'y a pratiquement pas d'adjectifs de relation en *-al* ou en *-ique* construits sur des dérivés en *-isme* ou en *-iste*. Sauf exception, les adjectifs en *-ismal* ou en *-ismique* sont construits sur des bases démotivées (*rhumatismal*, *cataclysmique*...). Quant aux finales en *-istique*, elles ne sont pas le résultat d'une dérivation en chaîne (*linguistique* renvoie à *langue*, pas à *linguiste*), *-istique* est simplement une variante de *-ique*. Mais – autre manifestation du principe d'économie – s'il y a déjà, dans le même paradigme, un adjectif de relation construit sur le primitif, il pourra être réutilisé pour renvoyer au dérivé en *-isme*. C'est ainsi que les adjectifs *alcoolique*, *palustre* ou *gigantesque*, qui, normalement, renvoient respectivement à *alcool*, *marais* et *géant*, peuvent servir d'adjectif de relation à *alcoolisme*, *paludisme* et *gigantisme* :

- « [des] excès alcooliques répétés » (Robida, *in Rob.*)  
 « des accès palustres courts et rapprochés » (Quillet, *in TLF*)  
 « Gigantisme et acromégalie sont en effet une seule et même affection [...] elle donne naissance au type gigantesque lorsqu'elle se développe avant la soudure des épiphyses; elle évolue au contraire vers le type acromégalique quand elle apparaît plus tard. » (*TLF s.v. acromégalie*).

Les adjectifs, logiquement, sont plus souvent concernés par le principe d'économie : dénués de capacité référentielle directe, ils ont davantage de plasticité sémantique. Mais on trouve également des exemples parmi les noms, avec les mêmes effets de décalage dans la compositionnalité du mot construit ou par rapport aux caractéristiques (sémantiques ou catégorielles) du suffixe. Le développement récent d'*assistantat* au sens de 'condition d'assisté', 'système d'assistance'<sup>4</sup>, est sans doute une « récupération » de *assistantat* 'fonction d'assistant'. Dans la série des noms de sportifs construits sur le nom de l'arme avec laquelle ils pratiquent l'escrime (*épée*, *fleuret*, *sabre*), à côté de *épéiste* et de *fleurettiste* on trouve *sabreur*, qui reprend une dénomination antérieure construite (avec un autre sens) sur le verbe *sabrer*. Echappant au paradigme des dérivés en *-iste*, elle rejoint celui des noms de sportifs construits avec le

suffixe *-eur* sur un verbe ou un nom d'activité. Application systématique de ce principe dans la série des dérivés en *-isme* et en *-iste* : l'existence, dans le même paradigme dérivationnel, d'un adjectif de relation en *-ien* bloque la formation d'un dérivé en *-iste* (la forme en *-ien* en tient lieu) et réciproquement l'existence d'un dérivés en *-iste* bloque généralement la formation d'un adjectif de relation en *-ien* (la forme en *-iste* en tient lieu), alors que les deux suffixes ne sont pas équivalents (cf. Roché, 2007).

- (8a) *Staline* → *stalinien* Adj 'de Staline'  
      ↗ " N Adj 'partisan de Staline', 'favorable à Staline'  
      *Staline* → °*staliniste*
- (8b) *Lénine* → *léniniste* N Adj 'partisan de Lénine', 'favorable à Lénine'  
      ↗ " Adj 'de Lénine'  
      *Lénine* → °*léninien*

Ici encore, la dimension lexicale – l'influence du paradigme dérivationnel – l'emporte sur la logique constructionnelle.

## 2.2 Un paradigme lexical unifié

Revenons aux ethniques pour observer maintenant le paradigme lexical qu'ils constituent. Lorsqu'ils sont formés sur le nom de pays, ils sont marqués comme tels soit par un suffixe spécifique – *-ais* ou *-ois* le plus souvent (*Congolais*, *Béninois*), *-(i)ote* (*Chypriote*), *-ite* (*Yéménite*), *-i* (*Emirati*), etc. –, soit par un suffixe dont c'est un emploi privilégié – *-ien* et *-ain* principalement (*Canadien*, *Cubain*). Lorsque le nom de peuple est le primitif, en revanche, sa forme est imprévisible (*Russe*, *Turc*, *Arabe*...) et ne permet pas de le caractériser. D'où la tendance à doter le nom de peuple originel d'une finale suffixale qui va l'intégrer dans le paradigme : *Anglais* se substitue à *Angle*, *Danois* à *Dane*, *Finnois* à *Finn*, *Hongrois* à *Hongre*, etc. Il s'agit là du phénomène d'« hypercaractérisation diachronique » depuis longtemps mis en évidence (cf. Malkiel, 1957-1958) : un lexème déjà caractérisé par son sens tend à l'être également par sa forme, si les deux se sont pas conjointement marqués d'emblée. Ou, en d'autres termes, d'une « intégration paradigmatique » (Corbin, 1987). Cette forme très particulière de suffixation – sémantiquement tautologique – est exactement semblable à l'exemple classique des noms d'arbres du type *peuplier* (formé sur l'afr. *peuple*, qui avait déjà le même sens). Dans certains cas, le suffixe sert directement d'habillage à un emprunt (*Malais*, *Iroquois*, *Illinois*), comme pour le non moins classique *palétuvier*. D'autres dénominations (*Gascon*, *Breton*) sélectionnent un ancien cas régime de préférence au cas sujet (*Gasc*, *Bret* sont également attestés) parce qu'il leur donne une apparence suffixale et les agrège à un paradigme en *-on* (*Wallon*, *Frison*, *Saxon*, *Teuton*) qui s'étend lui aussi par intégration paradigmatique (*Lapon*, *Letton*).

Une variante consiste à construire par suffixation un nouvel ethnique sur le nom de pays construit lui-même sur le primitif : *Malais* → *Malaisie* → *Malaisien*. D'où les nombreux doublets : *Finnois* / *Finlandais*, *Somali* / *Somalien*, *Azéris* / *Azerbaïdjanais*, *Thaï* / *Thaïlandais*, etc. Ils peuvent servir à distinguer le groupe ethnique proprement dit, d'un côté, et les citoyens d'un Etat, de l'autre : tous les Azéris ne sont pas Azerbaïdjanais et tous les Malaisiens ne sont pas Malais. Mais dans la pratique ils tendent à être interchangeable, avant que l'un chasse l'autre. On trouve sur la Toile de nombreuses attestations de « gouvernement thaï », « gouvernement malais », « gouvernement azéri » alors qu'on attendrait normalement « thaïlandais », « malaisien », « azerbaïdjanais ». Ainsi se continue sous nos yeux le processus historique qui a fait passer les ethniques du type (1) au type (2) puis au type (3). Un peuple sans territoire attitré (type (1)) s'installe dans un « pays » auquel il faut donner un nom (type (2)) ; et quand ce pays devient un Etat au sens moderne du terme les individus ne sont plus perçus comme les membres d'un peuple mais comme les habitants de cet Etat (type (3)). Si les *Espagnols* sont aujourd'hui les habitants de l'*Espagne*, l'*Espagne* (*Hispania*) a d'abord été le pays des *Hispani*. Et ainsi de l'Italie, de la France, etc.

La place manque pour étudier le rôle des différents suffixes dans ce double mouvement. D'une façon générale, leur distribution dépend en grande partie de critères phonologiques. Les bases en /i/ donnent à 88 % des dérivés en *-ien* (*italien*, *estonien*, *bolivien*, *tanzanien*, *fidgien*...). Que ces dénominations soient, souvent, empruntées à d'autres langues européennes n'y change rien : il est vraisemblable que les

mêmes conditionnements y sont à l'œuvre. Proprement française, la répartition des deux variantes qu'étaient, à l'origine, *-ais* et *-ois* est identique à celle qu'a observée Plénat (à paraître) pour les gentilés construits sur un nom de commune française. Dans notre corpus, sur 152 dénominations en *-ais* ou en *-ois*, un radical dont la dernière voyelle est une voyelle d'avant donne 31% de dérivés en *-ais* et 69% de dérivés en *-ois* tandis qu'un radical dont la dernière voyelle est une voyelle d'arrière donne 76% de dérivés en *-ais* et 24% de dérivés en *-ois*. Ainsi en Afrique, face à *soudanais*, *gabonais*, *congolais*, *burundais*, *sénégalais*, etc. les seuls ethniques en *-ois* sont *béninois* et *zaïrois*. En Asie orientale, *chinois* et *tonkinois* s'opposent de façon prévisible à *japonais* et *taïwanais*. Mais *balinais* a suivi sans doute le modèle de *javanais*, comme *atjehnaï* et *sumatranais* (même consonne épenthétique<sup>5</sup>). Pour l'Allemagne, seul *-ois* est présent, quelle que soit la voyelle qui précède (*badois*, *hessois*, *brandebourgeois*...). Parce que des bases en *-bourg* ne pouvaient donner que des dérivés en *-bourgeois* et que ceux-ci ont entraîné les autres ? Pour des raisons historiques, si tous ces ethniques sont apparus au même moment dans l'évolution phonétique de *-ais* / *-ois* ? Quoi qu'il en soit, ces exemples font apparaître un « effet de série », voire un simple effet de rime, une diffusion de proche en proche où l'histoire se mêle à la géographie. Les dénominations en *-on* énumérées plus haut dessinent un arc de cercle qui va des Pyrénées à la Baltique, dans lequel s'insèrent parfaitement *percheron*, *augeron*, *beauceron* et *brabançon* (et en s'écartant un peu *berrichon* et *bourguignon*). Toutes les bases en *-land(e)* donnent des dérivés en *-ais* (*hollandais*, *irlandais*, *finlandais*, *thaïlandais*, *groenlandais*...), alors que séparément des radicaux en *-an(e)* ou en *-ad(e)* peuvent être suivis d'un autre suffixe (*iranien*, *tchadien*).

### 2.3 Conflits au croisement des paradigmes

Dans certains cas, le mouvement qui s'était amorcé selon l'un des deux schémas précédents n'a pas abouti mais a laissé des traces dans des expressions complexes (*feu grégeois*, *poulet basquaise*...), des nominalisations démotivées (*mongolien* 'trisomique', *turquoise* 'pierre fine', *persiennes* 'contrevents'...), des emplois particuliers (*slavon* comme nom de langue). Pendant un certain temps, une dénomination suffixale, construite sur le nom de peuple ou le nom de pays, a concurrencé la dénomination originelle. Elle avait vocation à la supplanter dans tous ses emplois, y compris nominal. *Mongolien* s'est dit des habitants de la Mongolie et *Russien* de ceux de la Russie<sup>6</sup>. *Basquaise*, d'après les dictionnaires, sert de féminin à *Basque* comme *Suisse* à *Suisse*. Mais le nom de peuple primitif a résisté, la nouvelle dénomination ne s'est pas imposée et ce qu'il en reste est tombé ou tombe en désuétude. Les Russes sont toujours des *Russes* et les Basques (des deux sexes) sont restés (ou redevenus) des *Basques*. Cas particulier : *persien* ne s'est pas imposé mais son concurrent *persan* a eu plus de chance, jusqu'à ce qu'*iranien* le relègue dans un rôle historique ou comme nom de langue tandis que *perse* subsistait pour la Perse antique.

D'autres dénominations (*gomme arabe*, *Golfe Persique*, *acide prussique*, *italiques*...) représentent, avec les doublets néo-classiques déjà mentionnés (*germanique*, *hispanique*, *helvétique*...), un cas de figure nettement différent. Le suffixe *-ique* n'est pas, à la différence des précédents, un suffixe spécialisé, il forme des adjectifs de relation en général. Conséquence essentielle : ces dénominations sont restées, sauf exception<sup>7</sup>, cantonnées à un usage adjectival. La plupart sont des formations grecques ou latines utilisées pour des référents anciens ou antiquisants (*guerres médiévales*, *limes germanique*, *colonnes doriques*), comme formes supplétives (*Confédération Helvétique*, *relations germano-polonaises*) ou pour des emplois particuliers (dans *pays germaniques*, *études hispaniques*, par exemple, l'adjectif a une portée plus large que celle d'*allemand* ou *espagnol* et sert en quelque sorte de collectif). Les dénominations figées ou démotivées citées plus haut témoignent, jusqu'en français classique, d'un usage plus large de certains de ces adjectifs hérités (*arabique*, *indique*, *persique*...). *Italique*, par exemple, est dans (*lettres*) *italiques* un équivalent d'*italien*. D'autres sont des formations modernes, via le latin scientifique en général, mais ne se sont implantées plus ou moins durablement que dans des expressions complexes du vocabulaire savant : *acide prussique* en chimie, *selle turcique* en anatomie, *mer Baltique*... Seul, semble-t-il, *balkanique* reste d'un emploi libre mais il constitue un cas particulier dans la mesure où il réfère non à un Etat ou à un peuple mais à une région géographique<sup>8</sup>.

A côté du suffixe « savant » *-ique*, le suffixe « populaire » *-esque* a joué un temps le même rôle. *Arabesque, barbaresque, grecquesque, mauresque, sarrasinesque, tartaresque, tudesque, turquesque* sont attestés et s'étaient un temps plus ou moins diffusés<sup>9</sup>. Il n'en reste que des traces infimes, mais ils sont intéressants à la fois par leur existence et par leur échec. Malgré quelques emplois nominaux (*les Barbaresques, Mauresque* comme féminin de *Maure*), ils sont comme les dérivés en *-ique* essentiellement adjectivaux. Ils témoignent donc que la tendance à construire des ethniques spécifiquement adjectivaux n'est pas seulement une subsistance néo-classique, une continuation du système latin dans le registre savant. Elle répond, sous une autre forme, au même besoin de caractérisation, d'« intégration paradigmatique » que la formation de dérivés en *-ais, -ois, -on* à partir d'un nom de peuple (*supra*, § 2.2). L'adjectif est « marqué » comme tel par son appartenance aux vastes paradigmes des adjectifs de relation en *-ique* ou en *-esque*.

Mais cette logique d'intégration de l'adjectif dans un paradigme lexical typiquement adjectival entre en conflit avec celle du paradigme dérivationnel qui tend à imposer dans chaque famille une forme identique pour l'adjectif et pour le nom de personne (*supra*, § 2.1). *Arabique* et *arabesque* ont vocation à remplacer *arabe* comme adjectif, mais ils ne peuvent pas s'imposer si *Arabe* subsiste comme nom, parce que leur coexistence donnerait un paradigme dérivationnel hétérogène. Pour la même raison, *grégeois* et *turquois*, qui avaient commencé à se diffuser comme adjectifs, n'ont pas pu se maintenir parce que *Grec* et à *Turc* avaient résisté comme noms. Il n'y a pas de place pour *arabesque* et *grégeois* à côté d'*arabe* Adj et de *grec* Adj, et ils ne peuvent pas les éliminer si *Arabe* N et *Grec* N subsistent.

### 3 Les dérivés en *anti-* construits sur un ethnique

Les dérivés en *anti-* qui expriment une idée d'opposition (*médicament antigrippe / antigrippal* 'contre la grippe')<sup>10</sup>, essentiellement adjectivaux mais souvent nominalisés, sont caractérisés, comme les dérivés en *pré-* ou en *post-*, par un « principe de copie » (Dell, 1970 ; Corbin, 1987 ; Levkovych, 2004) qu'on peut schématiser ainsi : s'il existe déjà un adjectif formé sur la base, le dérivé lui emprunte son suffixe (*antiparlementaire* 'contre le parlement', *antigouvernemental* 'contre le gouvernement', etc.) ; s'il n'y en a pas, le préfixe est la seule marque (*antihalo* 'contre l'effet de halo', *antipuces* 'contre les puces'). Les nombreux doublets (*antigrippe / antigrippal*) attestent que le suffixe est (généralement) facultatif et que la base de *antiparlementaire*, par exemple, est bien le nom *parlement* et pas l'adjectif *parlementaire*.

Le principe de copie est une manifestation exemplaire de la double dimension paradigmatique de la dérivation et de la tendance à l'hypercaractérisation. Tous les dérivés en *anti-* sont marqués sémantiquement par le préfixe. Sur le plan catégoriel, la marque n'est pas indispensable : (*brigade*) *antigang*, (*centre*) *antipoison(s)* sont clairement des adjectifs. Mais quand il est présent, le suffixe traduit la composante catégorielle de la dérivation et intègre le résultat dans un paradigme lexical homogène. Que cette marque catégorielle soit obligatoirement empruntée à un autre membre du paradigme dérivationnel illustre d'une autre façon le principe d'économie énoncé plus haut. Y compris au détriment de l'instruction sémantique du suffixe mobilisé. Il s'agit, le plus souvent, d'un suffixe formateur d'adjectifs de relation, donc sémantiquement neutre, mais *antidépresseur* 'contre la dépression', par exemple, reprend le suffixe instrumental de *dépresseur* 'qui abaisse', *anti-fermentescible* 'contre la fermentation' le suffixe modalisateur de *fermentescible* 'susceptible de fermenter', etc.

La dérivation en *anti-* est une des plus productives et les dictionnaires ne recueillent qu'une petite partie des dérivés construits à partir d'un ethnique ou d'un nom de pays qu'on peut relever sur la Toile. Ils partagent les caractéristiques qu'on vient de résumer. Formellement, ils sont généralement construits sur l'adjectif. *Anti-allemand, anti-russe, anti-américain* sont beaucoup plus fréquents que *anti-Allemagne, anti-Russie, anti-Etats-Unis* ou *anti-Allemands, anti-Russes, anti-Américains*. Mais comme l'adjectif qui renvoie au nom de pays est formellement identique à celui qui renvoie au nom de personne, *anti-américain*, par exemple, signifie à la fois 'contre les Etats-Unis' et 'contre les Américains'. Il y a des contextes où la distinction ne serait pas pertinente, mais ailleurs l'ambiguïté de ces dérivés oblige à expliquer, par exemple, si l'on est taxé d'*anti-américanisme*, qu'on peut être contre les Etats-Unis, en tant qu'Etat représenté par son gouvernement, sans être contre les Américains.

## 4 Les dérivés en *-isme* et en *-iste* construits sur un ethnique

Comme la préfixation en *anti-*, la suffixation en *-isme* ou en *-iste* à partir d'un ethnique, d'un nom de pays ou d'un nom de langue (*italianisme, japonisme, angliciste...*) est particulièrement féconde. Et tout autant la combinaison des deux (*anti-américanisme*). Dans ces formations, les confusions formelles et sémantiques propres au système des ethniques sont encore accrues du fait de la polyvalence de ces deux suffixes, auxquels plusieurs modèles constructionnels sont attachés (cf. Roché, en préparation). C'est cette imbrication que nous voudrions maintenant essayer de démêler.

### 4.1 Les données : paramètres sémantiques et catégoriels

Un premier modèle de dérivation en *-isme* instaure une relation axiologique par rapport à la base et nominalise cette relation (l'*esclavagisme* est 'le fait d'être favorable à l'esclavage'), tandis que la dérivation en *-iste* correspondante construit le nom de la personne qui y est impliquée (l'*esclavagiste* est 'celui qui est favorable à l'esclavage'). Conformément à ce modèle, le dérivé en *-isme* peut désigner un mouvement politique, culturel ou d'opinion en relation avec le pays ou ses habitants, et le dérivé en *-iste* ses partisans (ou, en tant qu'adjectif, qualifier ce qui le concerne)<sup>11</sup> :

- (9) *catalanisme* 'autonomisme catalan' / *catalaniste* 'partisan du catalanisme'  
*scandinavisme* 'système politique qui s'inspire de la communauté ethnique et linguistique des pays scandinaves'  
*italianisme* 'goût pour ce qui vient d'Italie'

Un deuxième modèle sert à construire des noms d'action ou d'activité en *-isme* (*exorcisme, alpinisme*) et, parallèlement ou isolément, des noms d'agent en *-iste* (*exorciste, alpiniste, pianiste*). Dans le domaine qui nous intéresse, le dérivé en *-isme* désignera l'étude de la langue et de la civilisation concernées et le dérivé en *-iste* ceux qui s'y adonnent :

- (10) *indianisme* 'étude des langues et des civilisations de l'Inde'  
*océaniste* 'spécialiste des langues et civilisations océaniques'  
*germaniste* 'linguiste spécialisé dans l'étude des langues germaniques'

Le troisième modèle construit, à partir d'un adjectif ou d'un nom de personne, un nom de qualité en *-isme* (*parallélisme, dilettantisme*) sans correspondant en *-iste*. A partir d'un ethnique, le dérivé, comme ceux en *-ité* (*ivoirité*) ou en *-itude* (*belgitude*), désignera le caractère de ce qui est propre au pays et à ses habitants :

- (11) *germanisme* 'caractère germanique et plus particulièrement allemand de quelque chose ou de quelqu'un'  
*italianisme* 'caractère semblable à celui qui est propre à l'Italie et aux Italiens'  
*celtisme* 'ensemble des caractères particuliers aux Celtes'

Ou bien, à partir de l'adjectif correspondant au nom de langue, il désignera un idiotisme ou un emprunt :

- (12) *américanisme* 'idiotisme américain (par rapport à l'anglais)'  
*italianisme* 'expression italienne empruntée par une autre langue'  
*provençalisme* 'particularité propre au provençal'.

La base – le lexème qui sert de point de départ à la dérivation – est logiquement, suivant le cas, tantôt le nom de pays ou celui des habitants, tantôt le nom de langue, tantôt l'un ou l'autre des adjectifs correspondants. Compte tenu des nombreuses homonymies qui caractérisent le système et de divers accidents morphophonologiques, le résultat formel, cependant, est souvent le même. Un *latiniste*, par exemple, étudie le *latin* (N) tandis qu'un *latinisme* est une tournure typiquement *latine* (Adj). Mais le radical – la forme à laquelle est concaténé le suffixe – est identique dans les deux cas. A cause de l'haplogogie, *scandinavisme* peut être construit sur *Scandinavie* aussi bien que sur *scandinave*. Il n'en reste pas moins que, le plus souvent, le nom de pays est différent du nom de personne, du nom de langue et des adjectifs. Des dérivations sur *Italie* et *italien*, sur *Japon* et *japonais* devraient donner



respectivement *italisme* et *italianisme*, *japonisme* et *°japonaisisme*. Or deux faits massifs caractérisent les dérivés construits sur les uns et les autres :

– Quelle que soit la nature de la base et le sens du dérivé, et bien que les différentes dérivations énumérées ci-dessus soient nettement distinctes quant à leur contenu sémantique et catégoriel, la forme du dérivé est toujours la même. *Italisme* est attesté, mais il n'est pas usuel et c'est *italianisme* qu'on retrouve dans tous les emplois répertoriés ci-dessus. De sorte que l'utilisateur de la langue les perçoit comme des acceptions d'un seul et même « mot » alors que, constructionnellement, il s'agit de lexèmes distincts.

– La forme du dérivé est, ou semble être, tantôt construite sur celle de l'adjectif, du nom de personne ou du nom de langue (*italianisme*, *américanisme*, *provençalisme*, *indianiste*...), tantôt sur celle du nom de pays (*japonisme*, *belgicisme*, *sénégalisme*, *océaniste*...). Or ces différences de traitement ne dépendent pas du contenu sémantique et catégoriel des dérivations puisque, dans chaque famille, les diverses acceptions sont regroupées sur une même forme. Toutes choses égales par ailleurs, on devrait avoir logiquement, par exemple, soit *indianiste* et *°océaniste*, construits sur l'adjectif, si l'on comprend ces dénominations comme désignant respectivement un 'spécialiste des langues et civilisations indiennes' et un 'spécialiste des langues et civilisations océaniques' ; soit *°indiste* et *océaniste*, construits sur le nom, si on les comprend comme 'spécialiste des langues et civilisations de l'Inde / de l'Océanie'. Pourquoi ces différences de traitement ? Est-ce *océanien* qui a été tronqué pour former *océaniste*, ou *indien* qui s'est substitué à *Inde* pour servir de base à *indianiste* ?

## 4.2 Les données : paramètres morphophonologiques

Si l'on y regarde de plus près, on constate que la forme du dérivé dépend principalement de celle de l'adjectif (qui est aussi, rappelons-le, celles du nom de personne et du nom de langue). Trois ensembles se dessinent (on laissera de côté les finales rarement représentées et les exceptions, qui ne remettent pas en cause le classement proposé) :

– Si l'adjectif a une finale en *-al* (*provençal*), *-an* (*andorran*), *-ain* (*américain*), *-in* (*latin*), *-on* (*gascon*) ou une finale non suffixoïde (*arabe*, *berbère*, *corse*...), la concaténation se fait, sauf exception, sur cette forme : *provençalisme*, *andorranisme*, *américanisme*, *latinisme*, *gasconnisme*, *arabisme*, *berbérisme*, *corsisme*...

– Si l'adjectif a une finale suffixale ou suffixoïde en *-ien* (*italien*, *autrichien*, *palestinien*), *-éen* (*européen*), *-ique* (*attique*, *gaélique*), *-and* (*allemand*, *romand*), *-ard* (*picard*), *-ol* (*espagnol*), la concaténation se fait tantôt sur cette forme, moyennant le cas échéant une allomorphie (*italianisme*, *européanisme*, *atticisme*, *romandisme*, *picardisme*, *espagnolisme*...), tantôt sur une forme amputée de sa dernière rime (*palestinisme*, *européisme*, *gaélisme*...), tantôt sur une base supplétive (éventuellement tronquée) (*austriacisme*, *germanisme*, *hispanisme*...). Les doublets sont relativement nombreux, plusieurs solutions pouvant être exploitées pour une même base.

– Si l'adjectif a une finale suffixale ou suffixoïde en *-ais* (*anglais*, *japonais*, *libanais*, *portugais*...) ou en *-ois* (*hongrois*, *chinois*, *québécois*...), la concaténation est exceptionnelle. Les dérivations se font sur une base tronquée ou sur le nom de pays (*japonisme*, *libanisme*, *québécoisme*...), ou bien sur une base supplétive (*anglicisme*, *lusitanisme*, *magyarisme*...). Les hésitations entre plusieurs formes, ici encore, sont nombreuses. Pour *Chine* / *chinois*, on en trouve quatre – (*anti*-)*chinoisisme*, (*anti*-)*chinisme*, *sinisme*, *sinicisme* – dont aucune n'est réellement usuelle, ce qui traduit sans doute une situation de blocage.

D'un point de vue morphologique, l'élimination des segments *-ais-* et *-ois-* est logique : ces suffixes « populaires » sont incompatibles avec la suffixation « savante » que sont les dérivations en *-isme* et en *-iste*. Le phénomène est le même devant *-ité* et *-itude* (Dal & Namer, 2005). Devant *-isme* et *-iste*, il se double d'une difficulté phonologique : une finale en *-aisisme* ferait se succéder deux sifflantes trop rapprochées. Pour les mêmes raisons, le traitement de ces bases est identique devant *-iser* : *japonais* donne *japoniser* et non *°japonaisiser*. La troncation de *-ien* – s'il s'agit bien d'une troncation – est plus paradoxale. Son yod est certes indésirable à peu de distance du /i/ du suffixe, mais ailleurs, au contraire, c'est une base en *-ien* ~ *-ian* qui se substitue à la base attendue (dans le type *Hegel* → *hégélianisme* vs

*Kant* → *kantisme*, cf. Roché, 2007). Le traitement de *-ique* est encore plus déroutant. D'un point de vue phonologique, sa troncation est logique et attendue : favorisée par l'identité de la voyelle (pour ce qui la concerne, on pourrait parler d'haplogogie), elle évite la rencontre des deux sifflantes due à la spirantisation. En dehors de la présente série, le type *romantique* → *romantisme* est beaucoup plus fréquent que le type *catholique* → *catholicisme*. Mais parallèlement le segment *-ic-* est maintenu là où il pourrait être facilement évité (*celticisme*, doublet de *celtisme*) ou introduit via une base supplétive latine ou latinisante (*gallicisme*, *anglicisme*, *belgicisme*, *flandricisme*, *sinicisme*, *basquicisme*...), comme le segment *-(t)(a)c-* dans la série *rhotacisme*, *lambdacisme*, *iotacisme* (*ibid.*).

Sur le plan prosodique, certaines différences de traitement semblent logiques. *Méditerranéen* (6 syllabes) a plus de chances d'être tronqué que *coréen* (3 syllabes). D'où *méditerranéisme* face à *coréaniste*, tandis qu'*européen*, de longueur intermédiaire, donne aussi bien *européisme* qu'*européanisme*. Parmi les bases supplétives en *-ique* ~ *-ic-*, les trisyllabes *germanic-*, *hispanic-*, *helvetic-* sont tronquées (d'où *germanisme*, *hispanisme*, *helvétisme*) tandis que les dissyllabes *anglic-*, *gallic-*, *nordic-* ne le sont pas (d'où *anglicisme*, *gallicisme*, *nordicisme*). Tous aboutissent ainsi au même calibre, qui est aussi celui de *latinisme* ou d'*hellénisme* et qui est obtenu dans *flandricisme* et *basquicisme* par l'introduction d'un segment parasite, dans *belgicisme* par le choix de la base *Belgique* plutôt que *belge*. A partir d'une base en *-ien*, en revanche, si l'on s'attend à trouver *lacédémonisme* plutôt que *°lacédémonianisme*, pourquoi *dorisme*, par exemple, plutôt que *°dorianisme*, alors qu'*italianisme*, *algérianisme*, *canadianisme* ont des bases plus longues ? D'une façon générale, un examen systématique des bases en *-ien* montre que, pour elles, le facteur prosodique n'est pas pertinent. S'il intervient, c'est en renfort d'autres éléments.

Il faut tenir compte, en effet, en amont de la finale adjectivale, de la forme qui subsiste lorsque l'adjectif est tronqué ou qu'il est remplacé par le nom de pays. Et revenir, par exemple, sur le couple *Inde* / *indien* → *indianiste* vs *Océanie* / *océanien* → *océaniste*. *Indianiste* et *océaniste* sont plus proches du gabarit optimal, pour un dérivé, que ne le seraient *°indiste* et *°océanianiste* : base dissyllabique, dérivé trisyllabique (cf. Plénat & Roché, 2003). Mais aussi, et surtout, *océaniste* a deux avantages déterminants sur son compétiteur *°océanianiste* :

– Il évite la consécution des segments /anj/ et /ani/, qui se heurte à une contrainte dissimilative (cf. Plénat, à paraître). Devant *-ité* et *-itude*, cette même contrainte fait que *-ien* est tronqué s'il suit /n/ (*estonitude* et non *°estonianitude*) et qu'il ne l'est pas s'il suit une autre consonne (*italianité* et non *°italité*) (Dal & Namer, 2005).

– Il permet d'accrocher le suffixe au segment *-an-*, dont plusieurs indices montrent qu'il constitue une finale privilégiée lorsqu'il suit une consonne. On a vu que, dans ce cas, une finale en *-an-* n'est jamais tronquée (types *andorranisme*, *américanisme*). Comme dans *océaniste*, la troncation d'autres finales le met souvent au contact du suffixe (*germanisme*, *hispanisme*, *lusitanisme*..., *iranisme*, *libanisme*, *pakistanisme*...). Dans *antillanisme*, il s'insère comme un interfixe pour éviter *°antillaisisme* et *°antillisme*, indésirables l'un et l'autre. Dans *maghrébanisme*, qui concurrence *maghrébinisme*, il se substitue à *-in-*, qui est pourtant lui-même un bon accroche. Comme si, pour renforcer la cohérence du paradigme lexical, la rime riche *-anisme* était préférable à la rime « suffisante » que constitue le suffixe seul.

Les facteurs phonologiques, par conséquent, ne sont pas seuls en cause, puisque sur ce plan on relève un certain nombre d'incohérences, mais ils sont souvent déterminants. Et ils le sont d'autant plus s'ils tendent à renforcer certaines séries lexicales.

### 4.3 Interprétation : les « effets de série »

Cette observation invite à se tourner vers les ensembles qu'on voit se dessiner en parcourant le corpus, dans lesquels une identité formelle correspond à une parenté géographique ou culturelle : *dorisme*, *ionisme*, *éolisme*, sur des bases tronquées en *-ien* (ou en *-ique*) désignant (ou qualifiant) des dialectes grecs ; *hébraïsme*, *aramaïsme*, *chaldaïsme* pour le Proche Orient ancien ; *gallicisme*, *belgicisme*,

*flandricisme*... Mais on butte immédiatement sur *atticisme*, par exemple, qui appartient à la troisième série pour la forme, à la première pour le sens. A y regarder de plus près, cependant, le paradoxe n'est qu'apparent. Hérité du grec, *atticisme* a été introduit très tôt en français (16<sup>e</sup> s.) et a servi de modèle immédiat à *gallicisme*, au moment où sont forgés également *francisme*, *latinisme* et *italianisme*. *Anglicisme* (17<sup>e</sup> s.), *hispanisme* et *germanisme* (18<sup>e</sup> s.) ont suivi. Tous ces dérivés utilisent les bases latines ou latinisantes disponibles, comme il est logique pour ce type de formation<sup>12</sup>, avec pour les bases en *-ic-* une différence de traitement (troncation ou spirantisation) conditionnée par le facteur prosodique mentionné plus haut. Les séries de dérivés en *-icisme*, *-ianisme*, *-anisme* sont ainsi en germe. Les érudits qui forgent *flandricisme* et *belgicisme*, à la fin du 18<sup>e</sup> siècle et au début du 19<sup>e</sup>, ne font que transposer un des modèles. Sur la lancée, *wallonicisme* et *bruxellicisme* viendront même concurrencer *wallonisme* et *bruxellisme*. Parallèlement, les hellénistes, qui ont trouvé en grec δωρισμός 'langage dorien, prononciation dorienne', l'adaptent directement en *dorisme* et sur ce modèle forgent *ionisme* et *éolisme*. Quant à *aramaïsme* et *chaldaïsme*, ils sélectionnent comme base *aramaïque* et *chadaïque* plutôt qu'*araméen* et *chaldéen* à cause d'*hébraïsme* (16<sup>e</sup> s.), qui suit lui-même le modèle de *judaïsme*, emprunt médiéval, via le latin, au grec ιουδαϊσμός. Ainsi une même cause produit des effets différents. Trois dérivés – αττικισμός, δωρισμός, ιουδαϊσμός – construits régulièrement en grec et empruntés à des époques et dans des conditions différentes servent de *leader words* pour amorcer des séries lexicales dans lesquelles une parenté formelle renforce la cohérence sémantique ou culturelle. Ce rôle de *leader word* n'est pas réservé, d'ailleurs, au prototype historique de la série : si *atticisme* est à l'origine de celle des idiotismes en *-icisme*, le relais a été pris par *gallicisme* et *anglicisme*, dont la diffusion contribue largement à l'étendre.

A cet effet de série sont liées les quelques différenciations sémantiques qu'on peut observer entre formes concurrentes, dans les doublets ou les triplets mentionnés plus haut. Les dérivés en *-icisme* désignent plutôt des idiotismes, tandis que les noms de mouvements d'opinion privilégient une forme directement associée à l'adjectif ethnique ou au nom de pays. Mais la répartition est loin d'être systématique. *Francisme*, qui a désigné un mouvement politique fascisant dans les années 1930, a été employé dès l'origine pour 'gallicisme' et est encore abondamment attesté dans ce sens. Inversement, *belgicisme*, par exemple, pourra être employé pour un nom de qualité renvoyant à *belge* (« Qu'ils [les écrivains belges] écrivent la langue de leur éducation, mais qu'ils ne rougissent pas si elle est parfois teintée de belgicisme, trempée dans les originalités locales. » (Gourmont, in *TLF s.v.*)) ou pour un courant d'opinion (« La défense des francophones et de la langue française, en dehors du belgicisme unilingue, commence dès la fin de ce cadre. Divers organismes le représentent, comme la Ligue contre la flamandisation de Bruxelles dans les années 30. » (fr.wikipedia.org/wiki/Mouvement\_wallon)). L'effet de série sélectionne une forme liée à une acception (*-icisme* pour un idiotisme, par exemple), mais cette forme est ensuite disponible pour les autres acceptions.

En fait, le choix des formes, quand plusieurs sont possibles, ou leur distribution, dans les doublets, dépendent surtout du niveau de langue. Les bases supplétives ou allomorphisées sont privilégiées dans le discours « savant » et le niveau soutenu tandis que des locuteurs plus frustes ou d'autres contextes utiliseront des formes plus proches de la base – *québécoïsme* et *turquisme* plutôt que *québécoïsme* et *turcisme*, *autrichianisme* plutôt que *austriacisme*, *allemandisme* et *francisme* (voire *françaisisme*) à la place de *germanisme* et *gallicisme*. Si les formes en *-ianiste* sont fréquentes parmi les noms de spécialistes, y compris sur des bases longues (*indonésianiste*, *calédonianiste*, *polynésianiste*...) ce n'est pas que ces dérivés sélectionneraient spécialement la base adjectivale alors que les autres préféreraient la base nominale, mais simplement parce que ces lexèmes apparaissent dans des contextes « savants » tandis que les formations en *anti-*, par exemple, sont fréquentes dans des contextes journalistiques ou des forums de discussion.

L'effet de série peut aussi s'exercer de l'extérieur, à partir d'un autre paradigme spécialisé de dérivés en *-isme*. A côté de *macédonisme* 'mouvement politique', qui adopte la forme attendue compte tenu de la longueur de la base et du segment qui précède le suffixe, *macédonianisme*, qui désigne spécifiquement un courant religieux, viole les contraintes d'euphonie pour rejoindre la série des dérivés

en *-ianisme* construits sur un anthroponyme et appartenant au même domaine (*arianisme, nestorianisme, pélagianisme...*).

D'une façon générale, dans les deux ensembles, très vastes et diversifiés, que constituent les dérivés en *-isme* et en *-iste*, la dimension paradigmatique de la dérivation se manifeste par une tendance à élargir la marque formelle en amont du suffixe pour constituer des séries spécialisées. Plus la rime est riche, mieux le dérivé est perçu comme appartenant à un paradigme donné. De la même façon qu'une finale en *-ianisme* est privilégiée pour une doctrine philosophique ou religieuse, une finale en *-anisme* (après consonne) ou en *-icisme* le sera pour un dérivé construit sur un ethnique ou sur un nom de langue. Reste à déterminer par quelle manipulation morphophonologique ce résultat est obtenu.

#### 4.4 Interprétation : troncation ou substitution paradigmatique ?

L'intégration du dérivé dans une des séries que nous venons d'évoquer, à l'intérieur du paradigme lexical, peut être réalisée de deux manières : en jouant sur le choix de la base (nom de pays plutôt qu'adjectif ou l'inverse) ou grâce à une manipulation morphophonologique (ajout d'un segment parasite ou, beaucoup plus souvent, troncation). Nous avons jusqu'à maintenant évoqué cette question avec prudence à cause de l'ambiguïté du type *japonisme, palestinisme*, etc. : une dérivation sur le nom de pays (*Japon, Palestine*) donne le même résultat qu'une dérivation sur l'adjectif (*japonais, palestinien*) amputé de sa dernière rime, comme dans la suffixation substitutive (les types *valise* → *valoche, bâtonnier* → *bâtonnat, (Saint) Thomas (d'Aquin)* → *thomisme*, etc.).

Quelques cas particuliers, cependant, permettent de distinguer les deux solutions. *Européisme* et *asiatisme* ne peuvent pas être construits formellement sur *Europe* et *Asie* : une dérivation directe aurait donné *°europisme* et *°asisme*. Ils ne peuvent être construits que sur *européen* et *asiatique*, moyennant une troncation de la dernière rime. *Britannisme* et *gaélisme, hébraïsme* et *finnisme*, peuvent difficilement être construits sur d'autres bases que *britannique* et *gaélique, hébraïque* et *finnois*. La troncation, là aussi, est évidente. Mais elle ne suffit pas à expliquer *hébraïste*, par exemple. Si un *hébraïsme* est une tournure typiquement *hébraïque*, un *hébraïste* est quelqu'un qui étudie l'*hébreu*. Sémantiquement, lexicalement, la base est ici le nom. Il y a donc eu à la fois, dans ce cas, substitution de *hébraïque* à *hébreu* puis troncation pour former le radical de *hébraïste*. Le processus est le même dans un dérivé comme *européiste* lorsqu'il désigne – dans la bouche de leurs adversaires – les partisans de l'Europe intégrée. La base lexicale est le nom *Europe*, auquel l'adjectif *européen* – tronqué – a été substitué.

Inversement, *baviérisme, champagnisme, flandrisme, guatémalisme, monacoïsme, périgordisme* ne peuvent être construits que sur *Bavière, Champagne, Flandres, Guatemala, Monaco, Périgord* et pas sur une troncation de *bavarois, champenois, flamand, guatémaltèque, monégasque, périgourdin*. Mais en passant d'une province à l'autre, c'est *bourguignonisme* qu'on trouvera et pas *°bourgognisme, catalanisme* et pas *°catalognisme, provençalisme* et pas *°provincisme*, même quand il s'agit d'un mouvement politique lié à la région et non d'un idiotisme lié à la langue. Toutes choses égales par ailleurs, le dérivé est construit tantôt sur la base nominale, tantôt sur la base adjectivale sans qu'on puisse rendre compte de ce chassé-croisé par l'effet d'une troncation.

D'une façon générale, les phénomènes de troncation sont incontestablement attestés dans notre corpus mais ils ne suffisent pas à expliquer les faits observés. Il faut revenir à l'autre type de paradigmes, les paradigmes dérivationnels, et admettre, suivant les cas, tantôt que la base nominale (nom de pays ou nom de langue) s'est substituée à la base adjectivale, tantôt que la base adjectivale s'est substituée à la base nominale. Des exemples comparables peuvent être observés dans d'autres secteurs de la dérivation en *-isme* et en *-iste*. Dans le type *hégélianisme* déjà mentionné, le dérivé est construit lexicalement sur les noms de personne *Hegel* et non sur l'adjectif *hégélien*, même nominalisé (l'*hégélianisme* est la philosophie de Hegel, comme le *kantisme* est celle de Kant, et pas celle des hégéliens). De nombreux dérivés en *-alisme* et en *-arisme* ne peuvent pas être construits, sémantiquement, sur l'adjectif en *-al* ou en *-aire* qui semble être leur base : le *personnalisme* est une philosophie fondée sur la *personne* et non sur ce qui est *personnel* (cf. Roché, en préparation). Dans tous ces exemples, comme dans l'exemple néerlandais analysé par Booij (1997 : 50), où l'adjectif de relation *Chomskyaans* 'de Chomsky' est construit non sur

ce nom mais sur *Chomskyaan* ‘partisan de Chomsky’, c’est à une « substitution paradigmatique » qu’on assiste : un membre du paradigme dérivationnel s’est substitué à un autre comme base du dérivé à former.

Substitution paradigmatique et troncation aboutissent au même résultat dans les cas où un mot comme *japonisme* ne peut être construit que sur *japonais* (s’il désigne une particularité de la langue, par exemple). Mais les deux phénomènes ne se situent pas sur le même plan. La troncation est d’ordre morphophonologique et obéit à des conditionnements principalement phonologiques (prosodiques ou segmentaux). La substitution paradigmatique est d’ordre lexical et rejoint le principe d’économie mis en évidence plus haut : là où l’on attendrait deux dérivés, construits l’un sur le nom de pays et l’autre sur l’adjectif ou le nom de langue, une même forme sert pour les deux. Les notions de troncation et de substitution paradigmatique sont donc séparément nécessaires puisque, on l’a vu, certains cas de figure ne relèvent que de l’une ou de l’autre. Et celle de substitution paradigmatique est plus « généralisante » puisqu’elle se traduit par diverses manifestations formelles (dont un allongement de la base dans le type *hégélianisme*, par exemple) et pas seulement par un raccourcissement.

## 5 Conclusion

Si le conditionnement de la construction des lexèmes était uniquement affaire de règles morphologiques, et commandé par les seules caractéristiques catégorielles et sémantiques de chaque dérivation, il serait identique pour tous les dérivés. On constate au contraire, dès qu’on observe un corpus un peu étendu, que le choix du procédé, celui de l’affixe, le traitement morphophonologique de sa concaténation à la base dépendent aussi de deux séries de facteurs : les contraintes phonologiques de « bonne formation », que l’on n’a fait qu’entrevoir ici sans approfondir leurs effets faute de place, et les contraintes lexicales qui tendent à renforcer l’insertion du mot construit dans les deux paradigmes auxquels il est destiné à appartenir. Une finale en *-icisme*, par exemple, ne devrait résulter que d’une base en /ik/ spirantisée. On a vu cependant qu’elle pouvait être à la fois attractive, pour des raisons lexicales, au point de se trouver dans des dérivés dont la base ne comportait pas elle-même une finale en *-ique* (le type *flandricisme*) ; et répulsive, pour des raisons phonologiques, si bien que la finale de la majorité des bases en *-ique* est tronquée (le type *germanisme*). Dans la formation des ethniques et des dérivés auxquels ils servent de base, la dimension lexicale se traduit par une même tendance à la structuration dans chacun des paradigmes concernés, mais selon des modalités différentes. Les paradigmes dérivationnels tendent à adopter la même structure, dans laquelle le principe d’économie joue un rôle déterminant : même forme, dans les ethniques, pour le nom de pays ; même forme pour les dérivés en *anti-* et les dérivés en *-isme*, qu’ils soient construits sur le nom de pays ou sur l’ethnique. Dans les paradigmes lexicaux, la marque morphologique renforce la cohérence sémantique et un même effet de série, à partir de *leader words*, se traduit par un effet de rime qui contribue tantôt au choix de l’affixe tantôt à son renforcement par une homophonie du segment qui précède. Mettre en évidence le rôle des *leader words* revient à souligner le poids de l’histoire, mais la vision diachronique n’est pas la seule à devoir en tenir compte. La trace de l’histoire est présente dans le lexique existant à un moment donné, donc dans la synchronie actuelle, qui façonne la créativité lexicale et ses modalités morphophonologiques.

## Références bibliographiques

- Bauer, L. (1997). “Derivational paradigms”, in G. Booij & J. van Marle (eds.), *Yearbook of Morphology 1996*. Dordrecht: Kluwer, 243-256.
- Booij, G. (1997). “Autonomous morphology and paradigmatic relations”, in G. Booij & J. van Marle (eds.), *Yearbook of Morphology 1996*. Dordrecht: Kluwer, 35-54.
- Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Niemeyer.
- Crocco Galèas, G. (1991). *Gli etnici italiani : Studio di morfologia naturale*. Padova: Unipress.

- Dal, G. & Namer, F. (2005). « L'exception infirme-t-elle la notion de règle ? Ou le lexique construit et la théorie de l'optimalité ». *Faits de langues* 25, 123-130.
- Dell, F. (1970). *Les règles phonologiques tardives et la morphologie dérivationnelle du français*, Doctoral Diss. Cambridge (Mass.) : MIT, miméo.
- Levkovych, O. (2004). *Les mots français en anti-*. Mémoire de DEA. Toulouse : Université de Toulouse-Le Mirail.
- Malkiel, Y. (1957-1958). "Diachronic Hypercharacterization in Romance". *Archivum Linguisticum* 9, 79-113; 10, 1-36.
- van Marle, J. (1985). *On the Paradigmatic Dimension of Morphological Creativity*. Dordrecht: Foris.
- Plénat, M. (à paraître). « Quelques considérations sur la formation des gentils », in Fradin, B. (éd.), *La raison morphologique. Hommage à la mémoire de Danièle Corbin*. Amsterdam : John Benjamins.
- Plénat, M. & Roché, M. (2003). "Prosodic constraints on suffixation in French", in G. Booij, J. DeCesaris, A. Ralli, S. Scalise (eds.), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting, (Barcelona, September 20-22, 2001)*. Barcelona: IULA-Universitat Pompeu Fabra, 285-299.
- Rob. = P. Robert. *Dictionnaire alphabétique et analogique de la langue française*, 2<sup>e</sup> édition revue par A. Rey. Paris, 1987.
- Roché, M. (2005). « Sur une classe d'adjectifs par conversion », in I. Choi-Jonin, M. Bras, A. Dagnac et M. Rouquier (éd.), *Questions de classification en linguistique : méthodes et descriptions. Mélanges offerts au Professeur Christian Molinier*. Berne : Peter Lang, 321-347.
- Roché, M. (2007). « Logique lexicale et morphologie : la dérivation en *-isme* », in G. Boyé, N. Hathout, F. Montermini (eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*. Somerville : Cascadia Press, 45-58.
- Roché, M. (à paraître). « Pour une morphologie *lexicale* », *Mémoires de la Société de Linguistique de Paris*, Nouvelle série, n° 17.
- Roché, M. (en préparation). « Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? », in M. Roché, G. Boyé, H. Giraud, N. Hathout, S. Lignon, F. Montermini, M. Plénat, *Des unités morphologiques au lexique*.
- TLF = *Trésor de la langue française, Dictionnaire de la langue du XIX<sup>e</sup> et du XX<sup>e</sup> siècles*. Sous la dir. de P. Imbs et de B. Quemada. Paris, Editions du CNRS, 1971-1994.

<sup>1</sup> Sur la notion de paradigme en morphologie lexicale, voir Roché (à paraître).

<sup>2</sup> En général, les adjectifs de relation ne sont formés par conversion – *piéton* → (*rue*) *piétonne*, *paysan* → (*traditions*) *paysannes*, *ouvrier* → (*mouvement*) *ouvrier*... – que si leur base est dotée d'une finale suffixale ou suffixoïde (cf. Roché, 2005), ce qui n'est pas forcément le cas pour les ethniques.

<sup>3</sup> Exemples empruntés à Booij (1997 : 45-46) qui, à partir d'une autre analyse et sans tenir compte de la polyvalence sémantique de l'adjectif, insiste également sur la « dimension paradigmatique » de ces dérivations.

<sup>4</sup> « Ils ont grandi dans une situation d'assistanat [...] » (*Le Monde*, 04.12.2005 : 15), « On a fait du RMI un assistanat. » ([www.liberation.fr/actualite/evenement/evenement2/211366.FR.php](http://www.liberation.fr/actualite/evenement/evenement2/211366.FR.php)).

<sup>5</sup> On peut noter à ce sujet que l'épenthèse, courante quand une base à finale vocalique est dissyllabique (*Congo* → *congolais*, *Togo* → *togolais*, *Java* → *javanais*) est normalement remplacée par une troncation si la base est trisyllabique (*Angola* → *angolais*, *Ouganda* → *ougandais*, *Sri Lanka* → *sri-lankais*). L'épenthèse de *sumatranais* (au lieu de °*sumatrais*) est due vraisemblablement au voisinage de *javanais*.

<sup>6</sup> Surtout dans les composés *Blanc-Russien*, *Petit-Russien*, *Grand-Russien*, qui ont subsisté plus longtemps et désignaient respectivement les habitants de la *Russie Blanche* (Biélorussie), de la *Petite Russie* (Ukraine) et de la *Grande Russie* (Russie proprement dite).

<sup>7</sup> L'emploi nominal de *Britannique*, par exemple, qui sert à subsumer *Anglais*, *Écossais* et *Gallois*. Pour *Asie*, une répartition s'était ébauchée entre *asiatique* N et *asiatique* Adj mais le premier a été éliminé au profit du second.

<sup>8</sup> Autre exception, *voltaïque* 'de la Haute-Volta [aujourd'hui Burkina-Fasso]', seule formation en *-ique* relativement récente sur un nom de pays, doit son existence à l'homonymie de la base avec le nom du physicien *Volta* sur lequel

---

avait été formé l'adjectif *voltaïque*, avec évidemment un tout autre sens. Comme quoi le principe d'économie, au prix d'un calembour, peut aller chercher une forme en dehors même du paradigme dérivationnel.

<sup>9</sup> Logiquement, puisqu'il s'agit d'un suffixe d'origine italienne, la diffusion, dans cet emploi, de l'italien *-esco* lui-même a été un peu plus large. Outre *romanesco* déjà cité, Crocco Galèas (1991) donne quelques exemples de gentilés construits sur un nom de localité. *Tedesco* s'est imposé pour 'allemand' et *arabesco* a gardé un emploi adjectival libre. Mais globalement l'évolution générale a été la même et la plupart des ethniques en *-esco* sont tombés en désuétude (merci à Fabio Montermini qui me les a communiqués).

<sup>10</sup> Ne sont pas concernés ici les dérivés nominaux dans lesquels le préfixe inverse les propriétés de la base (*antihéros*, *antimatière*) ou exprime une relation spatiale (*antichambre*).

<sup>11</sup> Les gloses des exemples qui suivent sont empruntées au *Grand Robert*.

<sup>12</sup> *Francisme*, attesté dès le 16<sup>e</sup> siècle comme synonyme de *gallicisme*, ne fait pas exception : dans la titulature néolatine, les rois de France sont *Rex Francorum*.