

Prépositions et noms de régions anciennes : évolution des emplois et représentations socio-culturelles

1. Introduction

Les subtilités de l'emploi des prépositions ont jusqu'ici, en général, fait l'objet d'hypothèses concernant leur identité en langue, définition qui, en retour, doit rendre compte des emplois observés en discours – dont l'observation permet d'élaborer les hypothèses en question...

« Ainsi, la préposition *avec* peut indiquer l'instrument (*Il mange avec des baguettes*), l'accompagnement (*Il est avec Paul*), la participation (*Il l'a traduit avec son collègue*), la manière (*Il l'a fait avec empressement*), la matière (*Il l'a préparé avec de la farine et du sucre*), différents types de possession (*Le grand blond avec une chaussure noire* (film), *une jeune fille avec une belle dot*), etc. Derrière la diversité des effets de sens, on peut souvent restituer un sens de base invariant (ici : la concomitance, c.-à-d. la coprésence de deux termes reliés par la préposition [...]) » (Riegel *et al.*, 2009⁴ : 643).

Mais cette même grammaire présente ainsi l'emploi des prépositions devant les « noms propres géographiques » (*ibid.* : 644), sans aucunement, donc, en distinguer l'apport sémantique :

« [à + article] devant les noms propres masculins à initiale consonantique et pluriels (*au Brésil, aux Etats-Unis, aux Bermudes*) ;

[*en* sans article] devant les noms féminins singuliers et les masculins à initiale vocalique (*en France, en Afghanistan*), mais avec de notables exceptions (*en Guyane, à la Martinique*). »

De surcroît, si cette présentation a le mérite de la simplicité, elle constitue en même temps un facteur d'erreurs et se heurte d'ailleurs à de nombreux contre-exemples : en ce qui concerne la première « règle », elle laisse entendre que la préposition *à* s'emploie devant les noms masculins pluriels, or *Bermudes* est du féminin et on lit dans *La Nouvelle République* du 8 décembre 2016, p. 10 : « Voici notre petite sélection non exhaustive des spectacles, à l'approche des fêtes en Deux-Sèvres » (et non *aux Deux-Sèvres*). *Limousin* est un nom propre géographique masculin, mais on dit *en Limousin* ou *dans le Limousin* et non *au Limousin* ; en quoi *en Guyane* illustre-t-il « de notables exceptions », puisqu'il s'agit d'un nom féminin singulier ? Et il y a bien longtemps que *en Martinique* l'emporte sur *à la Martinique* ou *en Guadeloupe* sur *à la Guadeloupe* (Reboul, 1994).

L'hypothèse ici soutenue, dans la continuité des travaux de Harris 1954, 1988 (voir l'excellente présentation que fait Cartier 2016 de la sémantique distributionnelle) ou Blumenthal (par exemple Blumenthal, 2008) et de ceux que nous avons conduits au sein

du projet PRESTO¹, est que la fréquence des distributions observée en discours est pertinente pour la compréhension du fonctionnement linguistique, à quoi nous ajoutons qu'elle est en outre significative des représentations sociales à un moment donné (Leeman, 2015, 2016a, 2016b, 2016c ; Leeman & Falaise, 2016 à par.). Ainsi, pour prendre l'exemple des noms propres géographiques, la Gestapo, entité extérieure à la France et dans laquelle plus personne ne se reconnaît « [sévît] *jusque dans le Languedoc* » (*Le Monde*, 1968 – c'est nous qui soulignons), alors que Georges Guingouin, homme du pays et référence valorisée en tant que personnalité de la Résistance, dirige les Francs-tireurs et partisans « *en Limousin* » (*Le Monde*, 1984 – souligné par nous).

2. Les régions anciennes

Un article précédent (Leeman & Falaise, *op. cit.*) est consacré à l'emploi des prépositions *à*, *dans*, et *en* avec les noms propres de départements et de régions administratives² en France ; la présente étude, forte de ces travaux antérieurs étayant l'hypothèse selon laquelle l'évolution des emplois linguistiques est liée aux changements des représentations sociales et culturelles, concerne la combinaison des prépositions *à*, *dans* et *en* avec les noms de régions anciennes.

Le terme « région ancienne » regroupe ici des territoires de tailles diverses (allant des anciennes provinces aux « terroirs » locaux) et aux dénominations bien établies dans la langue de la seconde moitié du XX^e siècle – ne seront donc pas traités les noms d'entités plus récentes que les départements établis à la fin du XVIII^e siècle³ mais en revanche seront pris en compte aussi bien des entités administratives formelles (anciennes provinces, anciens comtés, départements...), que des « terroirs », notion plus subjective qui n'a pas toujours eu de concrétisation sur le plan administratif (par exemple *Médoc*).

1. Le cas des « terroirs »

Le terroir n'est pas une entité administrative, comme la région ou le département, mais une portion du territoire français plus anciennement délimitée de par sa configuration géographique (déterminée naturellement par des montagnes, des rivières ou autres), ses particularités géologiques (dont dépendent la flore, la faune ainsi par conséquent que les activités locales : agriculture, élevage, pêche, bûcheronnage... et donc leurs productions caractéristiques : tabac, vin, foie gras, pipes...), les us et coutumes de ceux qui y vivent (habitat, habillement, cuisine, etc.), l'installation d'une circonscription ecclésiastique (évêché par exemple) ou d'un ordre religieux, etc. Souvent le nom du terroir est formé sur

1

<http://presto.ens-lyon.fr/>

2

Au sens des 22 régions administratives de France métropolitaine de 1970-2015.

3

Nous avons déjà abordé la question des noms de régions administratives, créées dans le dernier tiers du XX^e siècle, dans (Leeman & Falaise, *op. cit.*).

celui de la ville dont il constitue l'environnement – ainsi la région désormais dite « Hauts de France » comporte-t-elle l'Amiénois, le Beauvaisis, le Boulonnais, le Calaisis, l'Odomarois... (mais aussi la Flandre, l'Artois, le Marquenterre...).

La notion de « terroir » ainsi conçue peut se rapprocher de celle de « pays » (nous remercions Peter Blumenthal pour cette suggestion), telle qu'elle apparaît dans les expressions *vin de pays, il est connu dans le pays, on dit au pays que...*, etc. Toutefois, le terme *pays* est problématique car polysémique, étant susceptible de désigner des états (France, Allemagne, etc.), mais aussi les « pays » en tant qu'entités administratives créées par la *Loi d'orientation pour l'aménagement et le développement du territoire* de 1995. Le terme *terroir* nous semble moins sujet à ambiguïté.

Peuvent être perçus comme des appellations de terroirs certains noms de départements (comme *Gers*) ou de régions administratives, tel *Limousin*, province historique et culturelle devenue région au cours du processus de régionalisation entre les années cinquante et quatre-vingt (statut administratif qu'elle a perdu avec la réforme de 2016).

2. Prépositions et noms de régions

Selon la *doxa*, l'emploi de la préposition *dans* s'est étendu depuis le XVI^e siècle, en particulier pour indiquer le lieu, au détriment de celui de *en* – beaucoup plus fréquente en ancien français qu'aujourd'hui, selon le *Robert historique* (Rey *et al.* 1992). Mais en réalité c'est souvent l'évolution inverse qui s'observe, lorsque l'on examine les attestations concernant les noms géographiques (Eskénazi, 1987). À titre d'exemple, parmi quelques-unes des appellations citées plus haut, des noms de ville ou de village témoignent de l'emploi ancien de *en* (premier extrait ci-dessous), mais cette préposition se trouve aussi dans des formulations contemporaines informant sur des activités diverses :

« *Cet été aura lieu le premier Festival Equestre de Rue-en-Marquenterre (F.E.R.M.), qui se tiendra les 27/28 et 29 Juillet prochain.* » (site Web *Baie de Somme zéro carbone*, pour désigner la commune de Rue)

« *Bienvenue en Boulonnais où mer et nature se rejoignent pour vous ouvrir des horizons de découverte et de bien-être vastes et authentiques.* » (site Web de la Communauté d'agglomération du Boulonnais)

« *L'établissement qui nous intéresse est situé en Boulonnais dans la commune de Ferques au nord du bourg de Marquise.* » (site Web de l'éditeur *Histopale*)

Conformément à sa définition (Leeman & Falaise, *op. cit.*), la préposition *en* se généralise, y compris pour les noms masculins singuliers, lorsqu'il s'agit de valoriser la région en tant que telle, par le biais des stéréotypes censés la caractériser. Nous montrons en outre que cette évolution linguistique est liée à celle des représentations qu'a la société française de ses territoires, depuis qu'a été enclenchée la grande réforme de décentralisation de l'État français, laquelle a permis une réappropriation, par chacun, de ses « racines ».

3. Constitution d'une liste de noms

Afin de pouvoir étudier en corpus les noms de régions anciennes, il faut au préalable en établir une liste, ce pour quoi a été choisie une méthode automatique, destinée à rassembler rapidement une grande quantité de noms. L'ensemble obtenu agrège quatre listes différentes :

1. Les 34 anciennes provinces françaises. Par exemple *Île-de-France*, *Berry*, *Languedoc*, *Normandie*.
2. Une liste de 711 *régions naturelles de France*, « des territoires d'étendue limitée présentant des caractères physiques et une occupation humaine homogènes », publiée sur Wikipédia⁴, issue d'un travail collectif fondé sur le *Guide des Pays de France* (Zégierman, 1999) et le *Dictionnaire des pays et provinces de France* (Fénié & Fénié, 2000). Par exemple *Albanais*, *Albigeois*, *Alpes mancelles*, *Armagnac*, *Baie de Somme*.
3. Une liste de 335 noms de « régions européennes » issue de DBPedia (Lehmann & al., 2015). Par exemple *Auxerrois*, *Basse-Normandie*, *Bresse*, *Camargue*, *Charolais*.
4. La liste des 1429 noms d'intercommunalités françaises, elle aussi issue de DBPedia. Par exemple : *Communauté de communes de l'Argentonnois*, *Communauté de communes des Terres du Gâtinais*, dont nous extrayons automatiquement le dernier mot, qui correspond souvent à un nom de région ancienne (ainsi, dans l'exemple précédent : *Argentonnois*, *Gâtinais*).

L'agrégat de ces listes, obtenu automatiquement, n'est ni exhaustif, ni exempt d'erreurs, mais la quantité prime ici sur la qualité du fait que peu de ces noms sont présents avec une fréquence statistiquement significative dans notre corpus *Le Monde* (voir plus loin la partie 3.1) ; il est donc plus économique de ne contrôler que ces derniers *a posteriori*.

Ainsi, après contrôle des noms apparaissant fréquemment dans le corpus *Le Monde*, nous avons retiré de la liste ceux qui ne correspondent qu'à des régions administratives modernes (par exemple *Basse-Normandie*), ainsi que ceux qui sont fréquemment précédés des prépositions *à*, *dans* ou *en*, mais presque jamais en tant que noms de régions, comme *Albanais*, *Comté*, *Double*, *La Hague*, etc. En revanche ont été ajoutés quelques noms d'îles françaises : *Guadeloupe*, *Martinique*, *Mayotte* et *Réunion*.

3. Corpus et démarche méthodologique d'analyse

Il nous est apparu nécessaire, pour plus de précaution, de varier les genres de textes lors de la constitution du corpus afin d'éviter le risque de ne décrire qu'un certain type d'usage.

1. Corpus : quatre genres textuels

Quatre corpus ont été construits à partir du Web :

1. Un corpus diachronique issu du journal *Le Monde*, formé à partir des premiers paragraphes de 10 % des articles, diffusés gratuitement par le journal sur le Web,

4

https://fr.wikipedia.org/wiki/Liste_des_régions_naturelles_de_France

- pour la période 1950-2015 (16,5 millions de mots).
2. Un corpus contemporain de commentaires touristiques, extraits du site *TripAdvisor* (7,2 millions de mots). Il s'agit de textes courts (rarement plus de deux phrases), accompagnés des « lieux⁵ d'origine » des auteurs des commentaires⁶, information qui s'avère pertinente pour la caractérisation des prépositions introduisant les noms propres géographiques.
 3. Un corpus contemporain issu de 722 sites Web d'offices de tourisme (3,6 millions de mots).
 4. Un corpus contemporain de guides touristiques, extraits de la base de données du projet *Wikitravel* (5,3 millions de mots).

Les trois premiers corpus ont servi de base à Leeman & Falaise (*op. cit.*), mais le deuxième a été entre-temps enrichi de 3 millions de mots.

Ces pages Web ont été téléchargées, et les textes et métadonnées pertinentes pour notre étude ont été extraits, puis annotés morphosyntaxiquement et lemmatiquement ; les corpus ainsi obtenus ont été intégrés à nos outils d'étude de corpus (ScienQuest : Falaise & al. 2016 ; et TXM⁷). En complément au corpus diachronique *Le Monde* est aussi recouru à *Google n-gram Viewer*⁸, qui effectue des recherches dans la base *Google Books*. Il ne s'agit pas d'une base contrôlée : elle mêle tous les genres textuels sans souci d'équilibre, et ne permet pas une vérification fine des concordances⁹, mais, étant de très grande taille (47 milliards de mots), elle constitue un contrepoint utile aux attestations extraites du journal *Le Monde*, corpus mieux contrôlé mais beaucoup plus petit, et qui s'avère donc parfois insuffisant.

2. Méthodologie

5

Généralement une ville, plus rarement un département (nous considérons alors l'auteur originaire du chef-lieu du département pour nos calculs, par exemple Nîmes pour le Gard), voire dans un cas un pays (en l'occurrence, la Belgique ; nous considérons alors l'auteur originaire de Bruxelles). Dans la suite, nous ne traitons que les lieux situés en Europe Occidentale (deux lieux étaient situés en dehors de cette zone).

6

Ils sont renseignés par les auteurs eux-mêmes, avec toute la subjectivité que cela implique. Nous parlerons de « région d'origine » par simplicité.

7

<http://textometrie.ens-lyon.fr>

8

<https://books.google.com/ngrams>

9

En particulier, il n'est pas possible de retirer les occurrences où le nom de région appartient à un nom composé (par exemple *Languedoc-Roussillon* au lieu de *Languedoc*), comme nous le faisons sur les autres corpus.

L'observation consiste à comparer le nombre d'occurrences des distributions attestées (la fréquence de telle ou telle autre combinaison de type Prép + N) selon les genres de texte et les époques où ces derniers ont été produits ainsi que suivant l'origine géographique du sujet parlant – lorsqu'elle est disponible.

2.1. Variables observées

Afin de déterminer si certaines préposition sont favorisées, et, dans l'affirmative, selon quelle répartition, ont été comparées les fréquences d'emploi de *à*, *en* et *dans* devant les noms de région ancienne (de l'ancienne province au « terroir » local, en passant par les départements anciens), sur trois plans :

- un plan chronologique (années entre 1944 et 2015), mesuré sur le corpus journalistique ;
- un plan générique (quatre genres : articles de journal, commentaires touristiques, descriptifs fournis par les offices de tourisme, par les guides de voyage), mesuré sur les quatre corpus ;
- un plan géographique (concernant la distance entre la région considérée et la région d'origine du scripteur), mesuré sur le corpus de commentaires touristiques, car une relation semble se dessiner entre l'emploi de la préposition et la distance (*dans*) ou la proximité (*en*) du locuteur avec ce dont il parle (ce qui a déjà été observé *supra* à propos de ce qu'assume ou n'assume pas le sujet parlant).

Chaque fois qu'est effectuée une extraction du motif {*à*, *dans*, *en*}+Nom, il est vérifié que chaque concordance correspond à une attestation pertinente pour notre étude : sont alors éliminés en particulier les cas où le nom n'est clairement pas celui d'une région (par exemple quand ce dernier est aussi le nom d'une rivière, comme dans « *noyé dans l'Isère* »), les noms propres composés (*Poitou-Charentes*, *Languedoc-Roussillon*, *Lans-en-Vercors*, *Communauté de communes Chamousset en Lyonnais*), qui émanent d'une décision administrative, et les cas où la préposition n'est pas sélectionnée par le nom de lieu mais liée à la construction (« *dans la Bretagne d'aujourd'hui* », « *sa réputation ne s'arrête pas au Poitou* », « *fut attribué au Dauphiné* »).

2.2. Sélection de noms de régions

Au sein des données ainsi extraites, un grand nombre de noms de régions s'emploient avec une préposition stable (table 1), et ne seront pas retenus ici.

Nom de région	À (à la/l', au, aux)	EN	DANS (dans la/l'/les)
Bretagne			171
Gers			13
Mercantour			8
Réunion		36	
Vercors			35
Vosges			20

Yvelines			31
----------	--	--	----

Table 1. Fréquences absolues de quelques noms de régions, à préposition stable, dans les quatre corpus.

Les noms de régions dont la préposition n'est pas fixée dans l'usage (table 2) sont beaucoup plus prometteurs pour notre étude, dans la perspective de déterminer si l'un des plans définis en 3.2.1 joue un rôle dans le choix de la préposition.

Nom de région	À (à la/l', au, aux)	EN	DANS (dans la/l'/les)
Cévennes		14	29
Creuse		16	9
Drôme		11	35
Guadeloupe	11	36	
Isère		19	27
Languedoc		12	6
Limousin		28	15
Martinique	47	421	
Médoc	2	3	
Périgord		40	48
Poitou		8	4

Table 2. Fréquences absolues de quelques noms de régions, à préposition variable, dans les quatre corpus.

4. Résultats

L'observation générale est que la « règle » telle que donnée par Riegel *et al.* (*op. cit.*) ne correspond pas aux énoncés attestés : ainsi qu'en témoigne la table 2, d'une part, la préposition *à* ne se trouve que rarement devant les noms masculins, et ne se rencontre pas devant les noms féminins pluriels ; d'autre part, *en* apparaît devant des noms masculins commençant par une consonne.

1. Noms d'îles féminins

Sur les plans chronologique et générique, les deux noms d'îles concernés par des changements de préposition dans nos corpus, *Guadeloupe* et *Martinique*, ont déjà été traités dans Leeman & Falaise (*op. cit.*) : *Martinique* révèle dans le corpus *Le Monde* un remplacement complet de *à* par *en* au cours des années 1970 et 1980, *Guadeloupe* suivant la même évolution quoique un peu moins nettement ; en particulier, la période 1970-1980 n'y apparaît pas comme charnière, contrairement à ce que montre *Google Books*. Si les prépositions devant ces deux noms d'îles évoluent globalement de la même manière, cette évolution n'est toutefois pas symétrique et ne s'effectue pas nécessairement sur le même rythme, comme en atteste par exemple « *Le forfait d'une semaine [...] en Martinique ou à la Guadeloupe pour 1 099 euros [...] fait partie de l'opération "Prix de crise" lancée par Nouvelles Frontières pour l'hiver* » (*Le Monde*, 2008).

Sur le plan géographique, nos données ne couvrent que la Martinique (notre corpus ne comportant pas de commentaires pour la Guadeloupe) : 35 attestations de *en Martinique* proviennent de Martiniquais (8,3 % du total des 421 attestations avec *en* sur le corpus *Trip Advisor*), ainsi que 6 *dans la Martinique* (17,1 % des 35 attestations avec *dans* au sein du corpus *Trip Advisor*). Les Martiniquais privilégient ainsi la préposition *en*, comme mentionné dans Leeman & Falaise (*op. cit.*), mais c'est aussi le cas de locuteurs originaires d'autres régions. En tout état de cause, *en* tend au fil du temps à remplacer *à*.

2. Noms féminins (autres que ceux des îles)

Parmi les noms féminins, sont distingués d'une part *Cévennes*, au pluriel, et d'autre part *Creuse*, *Drôme* et *Isère*, au singulier et qui correspondent aussi à des noms de département et de rivière¹⁰.

2.1. Cévennes

Le corpus *Le Monde* n'offre que six occurrences de {*dans les*, *en*} *Cévennes*, ce qui est trop maigre pour se risquer à tirer des conclusions. Mais le recours à *Google Books* montre, comme avec *Martinique* et *Guadeloupe*, que la préposition *en* progresse, cette fois aux dépens de *dans* – augmentation qui est aussi particulièrement forte dans les années 1970-1980. Toutefois, on observe ici clairement une courbe classique en diachronie : après une période de forte présence (1970-1995), la nouvelle expression « *en Cévennes* », sans doute parce qu'elle perd peu à peu son caractère innovant et donc saillant, tend à se stabiliser à une fréquence plus faible (années 2000), et coexiste avec l'ancienne expression « *dans les Cévennes* ». L'hypothèse d'interprétation est que la mise en place de la décentralisation a eu pour effet une réappropriation et une valorisation, par les locuteurs, de l'identité régionale – laquelle se marque avec la préposition *en* – puis, la réforme perdant de son actualité, la dénomination s'est peu à peu démunie de son caractère innovant et revendicatif – et donc la préposition *en* tend à être moins utilisée.

Sur le plan générique (tableau 3), on constate que la préposition *en* est plus fréquente dans les textes d'offices de tourisme et les commentaires touristiques, qui mettent en avant le lieu, que dans des textes journalistiques et les guides touristiques, censés être plus neutres – ce qui rejoint l'interprétation précédente.

Genre textuel	en Cévennes	dans les Cévennes
Journal	2	4
Office de tourisme	3	1
Commentaire touristique	17	7

10

Rappelons ici que les résultats sont nettoyés de toutes les occurrences correspondant clairement au nom de la rivière et non au nom de la région, une différence en général facile à établir grâce au contexte.

Guide touristique	0	5
-------------------	---	---

Tableau 3. Fréquences absolues de *en Cévennes* et de *dans les Cévennes* dans les quatre corpus.

Le plan géographique (figure 2) confirme les observations *supra*, faisant apparaître que la préposition *en* est préférentiellement utilisée par les personnes proches des Cévennes (distance médiane 39 km), alors que celles qui sont plus loin (distance médiane 252 km) ont tendance à utiliser la préposition *dans*.

2.2. Creuse

Le corpus *Le Monde* ne présente que 11 attestations de {*en, dans la*} *Creuse*, sans claire progression temporelle, mais les données issues de *Google Books* montrent une progression de *en* à la fin des années 1960, suivie d'une rapide stabilisation qui dure ensuite jusqu'à la fin des années 2000. Si ces changements restent d'une ampleur assez limitée, ils vont dans le même sens que ce qui a été constaté pour Cévennes.

Sur le plan générique, *en* est toujours favorisé dans les commentaires touristiques ; il se confirme donc que le genre textuel influe sur le choix de la préposition.

Genre textuel	en Creuse	dans la Creuse
Journal	5	7
Commentaire touristique	11	2

Table 4. Fréquences absolues de *en Creuse* et *dans la Creuse* dans les deux corpus dans lesquels des occurrences sont attestées.

Le critère géographique apparaît ici nettement plus marqué que le critère chronologique, même si ne se trouvent que deux occurrences de *dans la Creuse* dans le corpus de commentaires touristiques – mais il est remarquable, en relation avec ce qui a été avancé sous (2.1) et ci-dessus pour *Cévennes*, que ces deux attestations proviennent d'auteurs originaires des régions les plus éloignées de la Creuse, alors que tous ceux qui sont issus d'endroits proches utilisent *en*.

2.3. Drôme

Le nom *Drôme* correspond à un nom de rivière et à celui d'un département, mais aussi à celui de « terroirs » : la *Drôme provençale* et la *Drôme des collines* – dénominations qui spécifient une identité (définie par des paysages, une culture, une gastronomie (etc.) typiques). Sur les trente-cinq occurrences de *dans la Drôme*, une seule concerne *Drôme provençale*, et une seule *Drôme des collines*. À l'inverse, toutes les attestations de *en Drôme* comportent en fait *Drôme provençale* et *Drôme des collines*, ce dont on peut inférer que la préposition *en* est principalement dédiée à des noms de parties caractéristiques de la Drôme, alors que le seul *Drôme* (qui peut alors renvoyer au département) apparaît surtout combiné avec la préposition *dans*.

2.4. Isère

Dans le cas de *Isère*, l'évolution des emplois est nette, et comparable à ce qui précède. Ainsi, sur le corpus *Le Monde* (déjà évoqué dans Leeman & Falaise, *op. cit.*) comme dans celui émanant de *Google Books*, la préposition *en* gagne progressivement du terrain sur *dans* depuis les années 1970, et surtout 1990. Le critère chronologique est ici le plus significatif dans le choix de la préposition.

Du point de vue des genres, *en* est partout préféré, sauf dans les textes journalistiques, où toutefois cette préposition tend à devenir majoritaire dans la période récente.

Genre textuel	en Isère	dans l'Isère
Journal	7	24
Office de tourisme	1	0
Commentaire touristique	11	1
Guide touristique	2	0

Table 5. Fréquences absolues de *en Isère* et de *dans l'Isère* dans les quatre corpus.

Il y a peu d'occurrences de *dans l'Isère* au sein des commentaires touristiques, et la distance de la région d'origine de l'auteur ne semble pas un critère très pertinent : les Grenoblois utilisent bien de préférence *en*, mais des auteurs originaires de l'Isère emploient aussi *dans*, tandis que la plupart de ceux qui sont issus de régions plus lointaines utilisent *en*.

Pour *Isère*, le critère le plus significatif dans le choix de la préposition semble ainsi surtout chronologique, comme c'était déjà le cas avec *Guadeloupe* et *Martinique*.

3. Noms masculins singuliers commençant par une consonne

Selon la référence citée (Riegel *et al.*, *op. cit.*), les noms propres géographiques masculins commençant par une consonne seraient (dans le cadre d'un emploi spatial) introduits par *à* (*au Portugal*, *aux États-Unis*) de même que les noms propres géographiques féminins pluriels (*aux Bermudes*) – mais ont déjà été signalées les limites de cette « règle » avec la caution d'attestations comme celles de *en Deux-Sèvres* (nom de département féminin pluriel excluant la préposition *à* : ?? *aller aux Deux-Sèvres*) ou de *en Limousin* (nom propre géographique masculin commençant par une consonne et par là présenté comme devant être introduit par *à*).

3.1. Sur le plan chronologique

Contrairement aux noms que nous avons vus jusqu'ici : *Languedoc*, *Limousin*, *Médoc*, pour lesquels il ne se trouve que trop peu d'attestations dans notre corpus diachronique pour conclure de manière pertinente, *Google Books* montre que *Languedoc* (figure 7), *Limousin* (figure 8) et *Périgord* (figure 10) ont, eux aussi, entretenu des liens grandissants avec la préposition *en* au détriment de *dans* entre le début des années 1960 et la fin de la décennie 1980 ; cette progression de *en* est toutefois moins marquée qu'avec les noms féminins, avec lesquels la préposition *en* était déjà majoritaire auparavant. Dans le cas de *Poitou* (figure 11) la progression de *en* est très faible, et inexistante pour *Médoc* (figure 9).

3.2. Sur le plan générique

L'analyse des prépositions dans les textes d'offices de tourisme montre le grand nombre de *en* (ce qui se conçoit si l'on adopte l'hypothèse guillaumienne concernant le rôle de cette préposition) – en dehors de quoi le partage entre prépositions est globalement assez équilibré, contrastant en cela avec *Google Books*, où *en* est nettement majoritaire. Un élément d'explication de cette différence sera proposé en partie 4.3.4.

Languedoc, Limousin, Médoc, Périgord¹¹, Poitou

	En
Journal	10
Office de tourisme	8
Commentaire touristique	69
Guide touristique	4

Table 6. Fréquences absolues dans les quatre corpus.

3.3. Sur le plan géographique

Comme avec *Cévennes* et *Creuse*, les personnes d'origine proche du Languedoc (figure 12), du Limousin (figure 13) et du Périgord (figure 14), ont tendance à employer *en*. L'exemple de *Périgord*, à travers un grand nombre d'occurrences, montre que, si, au-delà de 100 km de Périgueux, la fréquence d'emploi de *en* est assez proche de celle de *dans*, en deçà de cette limite, c'est *en* qui domine. Émanant de Périgueux même, cinq

11

Pour *Périgord*, on ne comptabilise pas les Périgord « colorés » comme *Périgord vert*.

attestations comportent *en*, pour une seule de *dans*.

En ce qui concerne *Médoc* et *Poitou* (non représentés), les occurrences manquent dans le corpus de commentaires, mais, pour *Médoc*, il est tout de même remarquable que, sur les sept occurrences du corpus¹² avec *en* ou *dans*, la seule employant *en* provienne d'un auteur originaire de Saint-Yzans-de-Médoc, et celles avec *dans* de Vensac (également dans le Médoc), Eysines (en banlieue bordelaise), Limoges, Nîmes et Lille. Pour *Poitou*, nous n'avons qu'une attestation de *dans* (émanant de Poitiers) et cinq avec *en* (provenant de Poitiers, de Saint-Benoit dans la Vienne, et trois d'Île-de-France) – cela dit, nous ne sommes pas en mesure de connaître les conditions de toutes les productions linguistiques : par exemple, est-ce que le natif du Médoc qui écrit *dans le Médoc* n'a pas des comptes à régler avec son milieu d'origine, une distance à prendre avec ce dernier, qui l'empêcheraient (inconsciemment bien sûr) de dire *en Médoc* (puisque cette préposition véhicule l'adhésion à un milieu non seulement géographique mais aussi historique, culturel, affectif) ?

3.4. Sur le plan sémantique

Si des tendances chronologiques et spatiales se dégagent des attestations dans le cas de *Languedoc*, *Limousin*, *Médoc*, *Périgord*, et *Poitou*, autorisant une évaluation critique des ouvrages de référence, elles ne suffisent pas, à elles seules, à expliquer le choix de la préposition (sauf à procéder par hypothèse interprétative).

En revanche, la comparaison des emplois de *en* avec ceux de *dans* au sein du corpus *Le Monde*, relativement équilibré en termes de prépositions, et présentant l'intérêt d'offrir un contexte discursif plus étendu que les commentaires touristiques, montre que *en* est plutôt utilisé pour marquer un lien, une certaine intimité avec la région en question. Ainsi, sept des dix occurrences avec *en* apparaissent dans des passages de longueur substantielle, dédiés à des personnalités ([1], [4], [8], [9], et [10]), une rivière ([2]), et une institution ([6]), présentées comme ancrées localement – ce qui ne se produit pas avec *dans*, lequel, à l'opposé, apparaît dans un contexte plus neutre, consistant souvent en énumérations ou en comparaisons entre lieux ([12], [15], [16], [20] et [23]), à titre d'illustration ([13]) ou dans le cadre d'une interaction avec des éléments appartenant à un autre ensemble, plus vaste (les loups de France [11], les scouts de France [14], les rois carolingiens [18], les opposants à une candidate à une élection nationale [19]), voire extérieurs (la Gestapo [21], la garance, originaire d'Asie [22]).

Ainsi est-elle validée l'hypothèse que *en* met en avant un lien subjectif, intime, avec le lieu (en tant qu'il se définit par des personnalités, des rivières ou des montagnes, une histoire et des institutions locales, etc.), alors que *dans* témoigne plutôt d'un rapport

12

Dont une provient d'un auteur qui n'a pas renseigné de lieu d'origine.

purement objectif (par exemple lors d'énumérations) et d'une non-appartenance (assumée par le locuteur) à la région (mais à un ensemble plus vaste, voire à une entité extérieure).

[1] M. André Chandernagor, député de la Creuse [...] écrit dans la Lettre du Limousin: [...] "Nous avons trop souvent dit en Limousin, que l'élaboration du Plan aurait dû être une occasion privilégiée."

[2] La Vienne partage en deux [...] le département. Elle est, là encore, "haute", avant de reprendre son nom unique en Poitou.

[3] M. Vidal invoque, à l'appui de son projet, le succès rencontré par la ligne touristique Coutras-Saint-Mariens, en Gironde. A noter, toutefois, qu'un autre projet de train touristique en Périgord, sur une ligne désaffectée entre Thiviers et Saint-Pardoux, a échoué l'an dernier, malgré l'appui d'un jeune industriel parisien.

[4] "Vous savez, tous les Corrèziens étaient chiraquiens, même les communistes." Georges Pérol a notamment été conseiller régional en Limousin sur une liste RPR-UDF.

[5] De magnifiques exemples de ces fers forgés à palmettes se trouvaient à l'église de Saint-Léonard, en Limousin. Entre 1913 et 1921 elles en ont disparu et se sont depuis transportées au musée dit des Cloîtres, à New-York, où elles sont mises en honneur.

[6] La mission a principalement visité la Fondation de l'eau, qui, en Limousin, assure une formation internationale de techniciens spécialisés ; la seconde intéresse l'industrie porcelainière.

[7] En chemin, [Cees Nooteboom] ne manque pas de faire étape en Poitou, chez son grand ami l'écrivain argentin Alberto Manguel.

[8] La reconnaissance est aussi affective. C'est en Limousin que ce jeune réfugié juif [lituanien], est devenu véritablement photographe. Caché à Ambazac, à quelques kilomètres au nord de Limoges, il a rejoint la Résistance et participé à la libération de la ville. Il y a ensuite photographié [...] ses camarades de maquis.

[9] Le lieutenant-colonel Georges Guingouin, compagnon de la Libération, ancien chef des FTP en Limousin.

[10] C'est que l'homme – Lorrain exilé en Languedoc – est présent, avec ce don de raconter qui était, chez lui, une façon essentielle de communiquer.

[11] Les derniers loups de France furent tués, dit-on, dans le Limousin. Sans doute, d'autres régions, plus hautes en solitude, plus riches en taillis profonds, pourraient-elles se vanter, elles aussi, d'un tel "privilège".

[12] S'il est une région dont l'unité géographique, géologique et économique ne peut être discutée, c'est bien le Massif Central. [...] pas d'autoroutes, ni en Auvergne ni dans le Limousin.

[13] Tous recherchent un meilleur cadre de vie et veulent profiter des bienfaits de la campagne, où [...] la vie sociale et culturelle peut être très riche. "Je n'ai jamais été si souvent au cinéma et au spectacle que depuis que j'habite la Creuse", raconte Véronique Lazerat, banlieusarde installée depuis 2005¹³ dans le Limousin. Selon un sondage, réalisé [...] auprès de 457 citoyens, [...] 39 % des personnes interrogées souhaiteraient s'installer en zone rurale.

[14] Dans le cadre d'un rassemblement des Scouts et Guides de France organisé à Cussac-Fort-Médoc [...], 8 500 scouts [...] s'étaient retrouvés à Bordeaux pour une journée "festive et pédagogique" [...]. Le soir venu, ils sont repartis dans le Médoc. Ce gigantesque camp écocitoyen doit s'achever vendredi 30 juillet après deux jours de découverte des écosystèmes locaux.

[15] Le comité directeur attend de ses fédérations départementales concernées [...] notamment en Aquitaine, dans le Centre, en Ile-de-France, dans le Limousin et en Provence-Alpes-Côte d'Azur.

[16] Les prix peuvent être 2,5 fois plus élevés en Ile-de-France que dans le Limousin, région la moins chère [...] et aller du simple au double entre l'Ile-de-France et l'Auvergne ou la Bourgogne, selon les chiffres de la Fnaim.

[17] Un site naturel abritant de nombreux squelettes d'animaux préhistoriques de diverses espèces, vieux de 30 000 ans à 50 000 ans a été découvert récemment par hasard dans le Périgord.

[18] Retracer l'histoire du franc, c'est un peu retracer l'histoire de France. Les rois carolingiens, disposant d'abondantes mines d'argent à Melle, dans le Poitou, frappèrent un nombre considérable de "deniers" d'argent.

[19] Que ceux qui doutent de la capacité de Ségolène à rassembler viennent dans le Poitou.

[20] De la commanderie des Templiers à Coulommiers (Seine-et-Marne), à la tour médiévale de Lastours dans le Limousin, compter, selon les séjours, de 130 à 600 euros.

[21] Mon père, réfugié d'Alsace, est mort à Toulouse en 1943 du fait de la

Gestapo qui sévissait jusque dans le Languedoc.

[22] En 1698, un marchand de Nîmes, Martin, avait obtenu un privilège royal pour en introduire la culture dans le Languedoc, mais ses tentatives, qui ne durèrent pas plus de deux ou trois ans, restèrent vaines.

[23] Il n' est pas un village qui ne possède pas un trésor roman, en parfait état de conservation : Castelveil, Haux et Blasimon dans l'Entre-Deux-Mers, Moulis dans le Médoc, Saint-Pierre-de-Mons en Sud Gironde, Saint-Georges-de-Montagne dans le Libournais, etc.

Cette différence sémantique peut d'ailleurs apporter un élément d'explication à l'opposition observée entre nos corpus, où, sauf dans le corpus d'offices de tourisme, *en* est à peu près aussi fréquent que *dans* devant les noms masculins, tandis que les attestations fournies par *Google Books* révèlent que *en* est nettement plus fréquent dans ce contexte (voir 4.3.2). *Google Books* contient en effet, contrairement à nos corpus, une grande quantité de littérature où, probablement, le lien avec la région est mis en avant plus souvent que dans nos propres données, lesquelles comportent davantage d'énumérations de lieux (corpus journalistique) ou de commentaires de touristes s'adressant à des interlocuteurs extérieurs au lieu (« *Une visite incontournable si vous êtes de passage dans le Périgord* »).

4. Synthèse

La table 8 synthétise quels critères se sont avérés significatifs pour chacun des noms de régions anciennes présentés ici.

	Critère	
	Chronologique	Géographique
Martinique, Guadeloupe, Isère	+	
?	-	
Cévennes	+	+
Creuse	-	+
Drôme	Drôme vs Drôme provençale / des collines.	
Languedoc, Limousin, Périgord	+	+
Médoc, Poitou	-	-

Table 8. Synthèse des critères significatifs dans le choix de la préposition.

5. Discussion

Comme souvent lorsque le test d'une hypothèse s'opère sur un corpus attesté, nous souhaiterions disposer de davantage de données. Toutefois, en l'état actuel de la

documentation rassemblée, il est possible d'esquisser une organisation de la distribution des prépositions devant les appellations de régions anciennes. Certes la situation est complexe, et, contrairement à ce que laissent croire certaines descriptions (Riegel *et al.*, *op. cit.*), il ne semble pas réalisable actuellement de réduire à une règle simple l'emploi des prépositions devant les noms propres géographiques – mais, en revanche, apparaissent se vérifier les hypothèses que nous avons avancées sur l'identité des prépositions dites « locatives » (Leeman & Falaise, *op. cit.*).

En effet, il apparaît dans certains cas (tel *Isère*) que le facteur chronologique est prépondérant et qu'une préposition inexistante dans ce paradigme il y a quelques décennies y occupe depuis peu une place significative, sinon prépondérante (*Cévennes*), ce que l'on peut relier, d'une part, aux réformes politiques concernant la décentralisation, lesquelles, d'autre part, ont un impact sur les représentations que se fait, à un moment donné, la société sur les territoires concernés.

Pour d'autres noms, l'évolution chronologique vient juste renforcer *en*, déjà majoritaire (*Languedoc*, *Limousin*, *Périgord*). Dans les deux cas, le point de départ de ce mouvement coïncide avec les politiques de décentralisation initiées dans les années 1970. En relation avec l'hypothèse de l'identité affectée aux prépositions *à*, *dans*, *en* (Leeman & Falaise, *op. cit.*), cette progression peut s'interpréter comme une conséquence, sur les esprits, de la politique de décentralisation en ceci que la reconnaissance politique des provinces, régions et terroirs traditionnels autorise celle, subjective, de l'appartenance à un « pays » (au sens restreint du terme) et par là de l'exploitation d'une autonomie laissée libre par un État jusqu'ici centralisateur et souverain (la réforme politique libère les esprits de l'obligation de se reconnaître seulement dans l'appartenance à une unique entité).

Témoigne de cette (ré)appropriation, par chacun, de son origine le fait que le facteur géographique de la distance ou de l'appartenance apparaît à l'œuvre dans de nombreux énoncés, la préposition *en* étant alors privilégiée par les locuteurs qui, d'une manière ou d'une autre, se réclament de l'identité spécifique du terroir, par opposition à l'emploi de la préposition *dans*, laquelle est plutôt le fait d'une appréciation « extérieure » : c'est ainsi que, en deçà d'une certaine distance, *en* devient fortement majoritaire pour *Cévennes*, *Creuse*, *Languedoc*, *Limousin*, et *Périgord*.

De fait, ces critères de sélection brossent finalement des tableaux assez similaires du paradigme *en* vs *dans*. Sur le plan sémantique, *en* marque la proximité (géographique autant qu'affective) et l'attachement au lieu, qui s'appuie sur le constat que, sur le plan géographique, cette préposition apparaît plus fréquemment chez les scripteurs liés à ce lieu et croît, sur le plan chronologique, avec la montée en puissance de la décentralisation et d'un certain esprit d'ancrage local.

6. Conclusion

Lorsque les noms de régions anciennes sont attestés avec plusieurs prépositions, pour une expression traditionnellement considérée comme d'ordre spatial, chaque distribution relève en réalité d'un certain point de vue. Comme vu entre autres à propos des noms *Languedoc*, *Limousin*, *Médoc*, *Périgord* et *Poitou*, qui peuvent s'associer aussi bien à *dans* qu'à *en*, ces deux prépositions n'ont pas, dans ce cadre paradigmatique, la même valeur sémantique : *en* met en avant un lien subjectif et intime, pour ainsi dire identitaire, avec la région, ou valorise un personnage dans lequel le locuteur se reconnaît ou un

événement qu'il assume, tandis que *dans* est de l'ordre de l'objectif, de la distance et de l'extériorité. Les évolutions observées sur les plans chronologique et géographique peuvent ainsi s'interpréter, en regard de ce contraste sémantique, sur le plan sociétal : les sujets parlants ont tendance à utiliser *en* pour désigner leur région d'origine ou celle à laquelle ils s'identifient, et le lien au « local » est de plus en plus mis en avant dans les textes (ainsi que le montre l'évolution chronologique au sein du corpus *Le Monde*). L'ensemble de ces attestations contribue à valider l'identité des prépositions concernées, telle que l'a esquissée Guillaume dès le début du XX^e siècle.

Références

Blumenthal, P., 2008, « Combinatoire des prépositions : approche quantitative », *Langue française* 157 : 37-51.

Cartier, E., 2016, « Distributionnalisme et sémantique : état des lieux en traitement automatique des langues », in Cl. Martinot *et al.* (eds) *Perspectives harrisiennes*, Paris, CRL : 288-344.

Eskénazi, A., 1987, « Député DE Saône-et-Loire – Préfet DU Rhône – EN Vendée », *LINX* 16 : 28-69.

Falaise, A., Tutin, A., Kraif, O. (2011), « Définition et conception d'une interface pour l'exploitation de corpus arborés pour non-informaticiens : la plateforme ScienQuest du projet Scientext », *Traitement Automatique des Langues*, 52(3) : 103-128.

Fénié, B. & et Jean-Jacques Fénié, J.-J., 2000, *Dictionnaire des pays et provinces de France*, Éditions Sud-Ouest.

Guillaume, G., 1919, *Le problème de l'article et sa solution dans la langue française*, Paris, Nizet.

Harris, Z. S., 1954, « Distributional Structure », *Word* 10 (23) : 146-162.

Harris, Z. S., 1988, *Language and Information*, New York, Columbia University Press.

Leeman, D., 2015, « La préposition *en* et les noms de pays », in *Phraséologie et profils combinatoires : lexicque, syntaxe et sémantique. Hommages offerts à Peter Blumenthal*, Paris, Champion : 189-200.

Leeman, D., 2016a, « Pour une extension du postulat « *la métalangue est dans la langue* » : l'explicitation de la situation comme partie prenante de la distribution », in Cl. Martinot *et al.* (dir.) *Perspectives harrisiennes*, Paris, CRL : 203-210.

Leeman, D., 2016b, « Hypothèse de résolution du problème posé par l'emploi des prépositions devant les noms de pays », in C. Marque-Pucheu *et al.* (dir.) *Linguisticae Investigaciones Supplementa* 32, Amsterdam, Benjamins : 107-124.

Leeman, D., 2016c, à par., "Les paradigmes comme échos / véhicules des représentations sociales : l'exemple de structures de type *Prép + N*". *Signata*. <http://www.danielle-leeman.com>, onglet « recherche »

Leeman & Falaise, 2016, à par. « Les prépositions devant les noms de région et de département français », *Langages*.

Lehmann, J. *et al.*, 2015, « DBpedia – A Large-scale, Multilingual Knowledge Base

Extracted from Wikipedia », *Semantic Web Journal*, 6-2, 167-195.

Reboul, S., 1994, « *A la Guadeloupe / en Guadeloupe : vers une interprétation cognitive ?* », *Langue française* 103 : 68-79.

Rey, A. *et al.*, 1998, *Robert historique de la langue française*, Paris.

Riegel, M. *et al.*, 2009, *Grammaire méthodique du français*, Paris, PUF.

Zégierman, F., 1999, *Le guide des Pays de France*, Paris, Fayard.