



**HAL**  
open science

# Écrire l'histoire des migrations dans les universités françaises. Les thèses de doctorat soutenues dans les universités françaises depuis 1980

Philippe Rygiel

## ► To cite this version:

Philippe Rygiel. Écrire l'histoire des migrations dans les universités françaises. Les thèses de doctorat soutenues dans les universités françaises depuis 1980. Exilés, réfugiés, étrangers en France 1847-1986, université des pays de l'ADOUR, Feb 2025, Pau, France. ⟨halshs-05629848⟩

**HAL Id: halshs-05629848**

**<https://shs.hal.science/halshs-05629848v1>**

Submitted on 22 May 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Écrire l'histoire des migrations dans les universités françaises.

## Les thèses de doctorat soutenues dans les universités françaises depuis 1980

Exilés, réfugiés, étrangers en France 1847-1986, colloque de Pau, 5-7 février 2025

Philippe Rygiel<sup>1</sup>

« *Once I thought to write the history of the immigrants in America. Then I discovered that immigrants were American history.* »

Oscar Handlin, *The Uprooted*, 1951

Je remercie Laurent Dornel et Delphine Diaz pour leur invitation à ces journées, même si je ne suis pas certain de pouvoir les remercier de m'avoir posé la question à laquelle je dois m'attaquer aujourd'hui puisque j'ai la rude tâche de dresser un panorama de la recherche sur le thème qui nous occupe au cours de ces dernières décennies.

Je me heurte en effet pour aborder ce sujet à quelques difficultés. Au fil des décennies, j'ai produit plusieurs textes évoquant l'histoire et l'historiographie consacrée aux migrations ayant touché la France<sup>2</sup>. Je ne suis pas le seul en ce cas : nous disposons de nombreux bilans successivement établis<sup>3</sup>, dont le dernier ouvre le récent volume publié à l'occasion de la mise

---

<sup>1</sup>École Normale Supérieure de Lyon, LARHRA.

<sup>2</sup>Philippe Rygiel, « L'historiographie des migrations », *Mémoires publiés par la fédération des sociétés historiques et archéologiques de Paris et l'Île-de-France* 61 (2010) : 7–17 ; Philippe Rygiel, « Archives et historiographie de l'immigration », *Migrations* 33 (2009) : 50–59 ; Philippe Rygiel, « Les migrations internationales au prisme d'une historiographie nationale », *Il mestiere di storico*, 2018.

<sup>3</sup>Marie-Claude Blanc-Chaléard, « Des logiques nationales aux logiques ethniques ? », *Le Mouvement Social*, n° 188 (1999) : 3–16 ; Linda Guerry, « Femmes et genre dans l'histoire de l'immigration », *Genre & histoire*, n° 5 (2009) ; Nancy L. Green, « Quatre âges des études migratoires », *Clio. Femmes, genre, histoire*, n° 51 (2020) : 185–206 ; Gérard Noiriel, « Histoire de l'immigration en France. État des lieux, perspectives d'avenir », *Hommes & migrations* 1255, n° 1 (2005) : 38–48.

au programme du concours des ENS<sup>4</sup>. Je ne peux me résoudre à les actualiser simplement, ce serait révéler trop crûment que je n'écris plus tout à fait cette histoire-là depuis au moins une ou deux décennies, et que je n'en suis plus l'actualité d'aussi près qu'autrefois.

La formule « écrire l'histoire des migrations » dissimule en outre quelques questions redoutables. La première concerne l'identité du locuteur absent de cette phrase. Nous vivons dans un monde saturé de discours sur les migrations internationales : politiques, journalistes, experts, représentants de multiples groupes et institutions produisent un océan de textes qui inscrivent presque toujours les migrations les situant, implicitement au moins dans une temporalité voire une historicité. Elle peut servir à paraître savant, à déplorer que les migrants d'aujourd'hui soient moins désirables que ceux d'hier, ou à annoncer qu'ils n'ont jamais été aussi nombreux.

Quelle place occupent, dans cet univers, les textes produits par les historiens et les universitaires ? Quelle est leur portée ? Quels rapports entretiennent-ils avec ces autres discours qui fonctionnent comme un contexte, parfois comme une intertextualité plus ou moins assumée ? Voici les questions qui se posent à nous dans un contexte politique marqué par la progression de mouvements néo-nationalistes qui font de la fermeture des frontières et de la purification des nations une obsession.

La pente naturelle du bilan historiographique est d'autre part normative et implicitement programmatique : allant de grand texte en perspectives novatrices, il dit peu de ce qu'on pourrait appeler la science normale, qui s'écrit dans les universités et les laboratoires sans nécessairement avoir l'ambition d'infléchir la trajectoire d'un champ. Or il est possible, si l'on se pose la question de l'insertion de cette production dans l'espace social, que ce soit précisément cette masse de travaux qui nous fournisse les informations les plus utiles.

J'ai donc choisi une voie différente : tenter de décrire les contours de l'ensemble des textes produits dans les universités françaises traitant de l'histoire des migrations, avant d'énoncer quelques hypothèses sur les contextes dans lesquels ils s'inscrivent. La voie la plus simple et la plus efficace pour s'attaquer à ce type de question est de partir des thèses soutenues dans les universités françaises, qui offrent un corpus assez homogène et facile à constituer par l'intermédiaire du répertoire des thèses, consultable en ligne, theses.fr, qui permet de repérer 275 thèses soutenues dans les universités françaises depuis 1980 ayant pour objet l'histoire des migrations, du moins que la présence d'un mot de la liste retenue<sup>5</sup> dans l'ensemble textuel constitué par le résumé, le titre et les mots-clés a permis de définir comme telle.

Le choix de ce corpus introduit bien sûr de nombreux biais. Le répertoire lui-même n'est pas exempt d'erreurs, voire en fourmille, les thèses sont loin de représenter la totalité des discours savants produits sur ce sujet, la production de savoir par un historien ne s'arrête pas toujours à la soutenance de sa thèse mais peut se poursuivre par le biais d'articles et d'ouvrages. Enfin l'usage même de cette base de données induit des biais. L'inscription d'une référence dans le domaine de l'histoire est parfois incertaine lorsque l'on l'utilise, les rubriques n'étant pas toujours remplies, les dénominations changeantes. De plus nous avons pour constituer le corpus eu recours à une recherche par mots clés explorant les titres et les résumés. Les titres cependant sont parfois peu explicites, les résumés plus que succincts et les mots clés pas toujours bien

---

<sup>4</sup>Delphine Diaz et al., *Exilés, réfugiés, étrangers : France 1848–1986* (Neuilly-sur-Seine : Atlande, 2024).

<sup>5</sup> Il faut pour qu'une thèse entre dans le corpus que les métadonnées présentes conduisent à la ranger dans le domaine historique et que l'ensemble textuel la décrivant comporte le lemme « migr », « refuge » ou « asile ».

choisis et parfois absents. Il est probable qu'une partie de la production, dont le volume et les propriétés sont difficiles à estimer nous échappe. Nous travaillons donc avec un proxy dont la qualité n'est pas certaine, ce qui bien sûr pèse sur la solidité de nos conclusions mais dont le mérite est d'exister et d'autoriser à formuler quelques hypothèses pertinentes.

## I. La production d'une histoire des migrations dans les universités françaises

### Une production en augmentation constante

Le premier constat permis par ce corpus est que le nombre de thèses soutenues augmente légèrement mais de manière significative au fil du temps — à raison de 0,22 thèses supplémentaires par an ( $p$ -valeur :  $1,65 \times 10^{-5}$ ). Cette croissance n'est pas linéaire mais s'opère par paliers, avec une forte augmentation au milieu des années 1990, puis à la fin des années 2000.

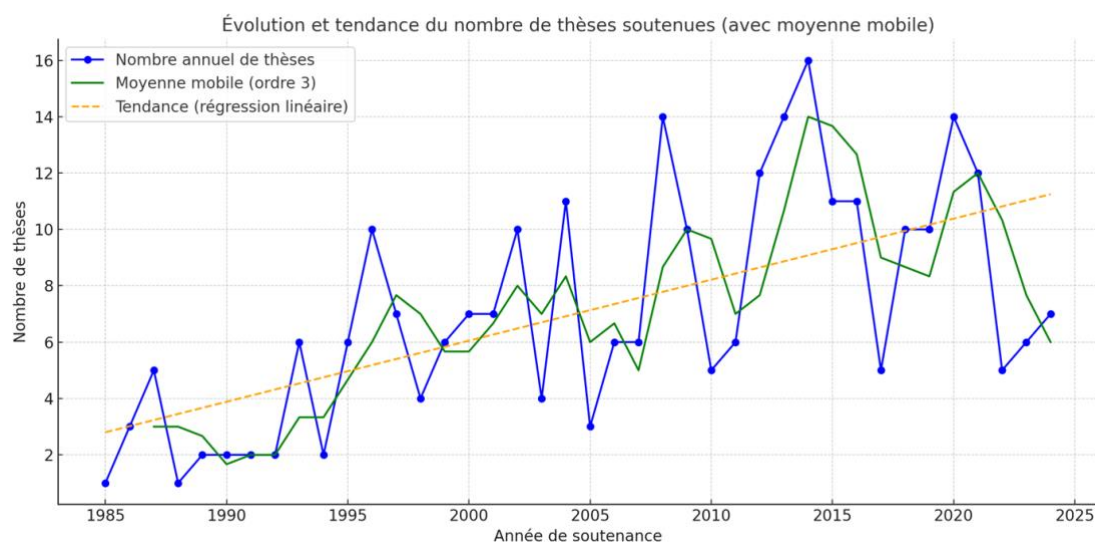


Figure 1. Évolution et tendance du nombre de thèses soutenues (avec moyenne mobile d'ordre 3). Source : theses.fr.

Il est permis de penser qu'il s'agit là de la manifestation locale d'une tendance générale à la multiplication, tous espaces académiques et toutes disciplines confondues, des études s'inscrivant dans le champ des *migration studies*<sup>6</sup>, mais aussi de la croissance en volume de la production académique. La production française d'études d'histoire migratoire s'inscrit donc dans un contexte global, dont elle constitue très petit segment.

### Une production très parisienne

Celui-ci présente quelques caractères marquants, en particulier une très forte centralisation géographique et thématique. Soixante et un pour cent des thèses ont été soutenues dans un établissement parisien, et la production est dominée par quelques grands établissements. Une lecture directe des données pourrait conduire à diagnostiquer une légère diminution de cette

<sup>6</sup>Asya Pisarevskaya et al., « Mapping migration studies : an empirical analysis of the coming of age of a research field », *Migration Studies* 8, n° 3 (2020) : 455–481.

domination au cours du temps, mais les écarts sont statistiquement non significatifs (p-valeur : 0,165) : la centralisation parisienne est persistante et robuste.

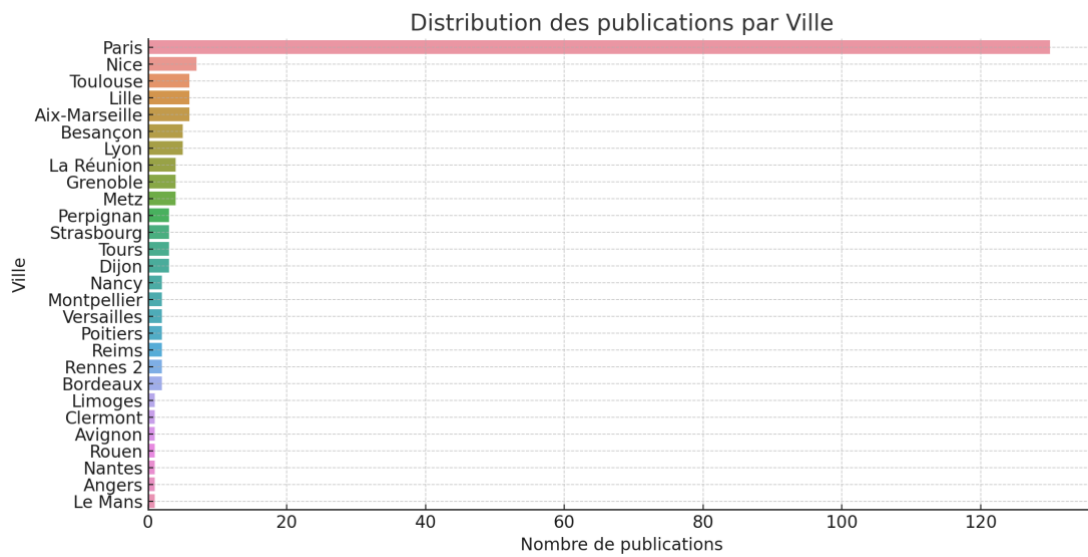


Figure 2. Distribution des thèses par ville de soutenance. Source : theses.fr.

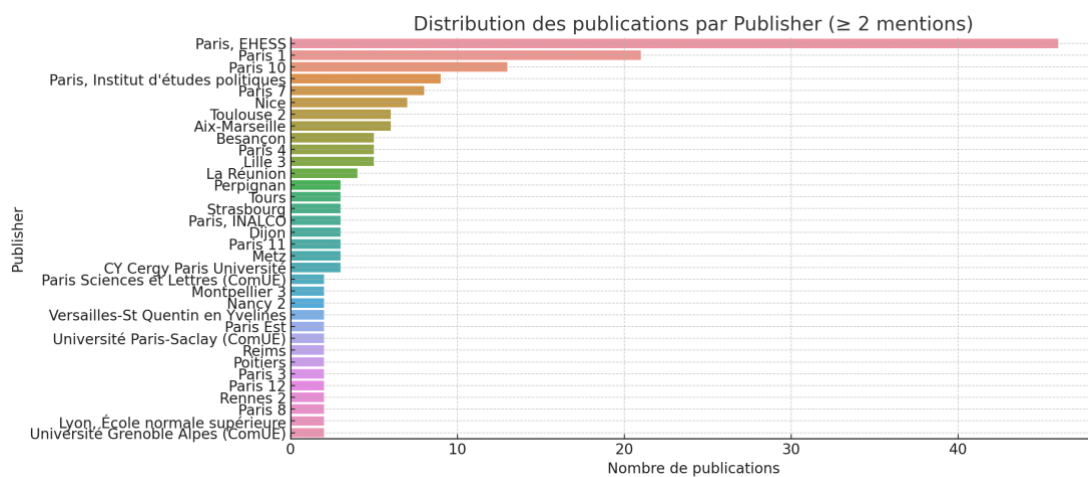


Figure 3. Distribution des thèses par établissement (au moins 2 thèses). Source : theses.fr.

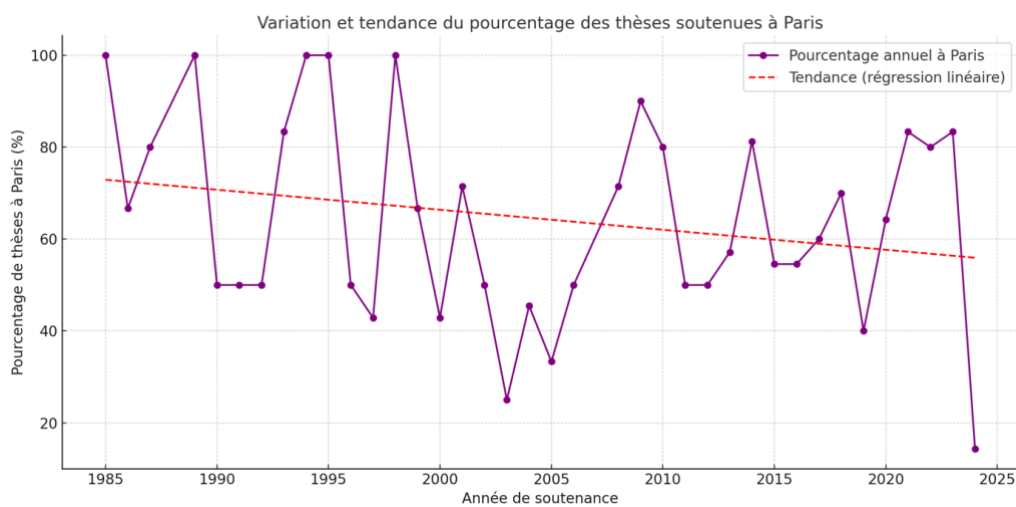


Figure 4. Variation et tendance du pourcentage de thèses soutenues à Paris (*p*-valeur : 0,165 ; tendance non significative). Source : theses.fr.

Cette centralisation parisienne se double d'une concentration remarquable autour d'un petit nombre de directeurs et directrices de thèse. Gérard Noriel et Nancy Green, tous deux ayant passé l'essentiel de leur carrière à l'EHESS ont fait soutenir 30 des 210 thèses du corpus, soit environ 15%.

Rang	Direction	Établissement	Thèses dirigées
1	Noiriel, Gérard	EHESS (Paris)	16
2	Green, Nancy L.	EHESS (Paris)	14
3	Milza, Pierre	Institut d'études politiques (Paris)	6
4	Blanc-Chaléard, Marie-Claude	Paris 10	5
4	Schor, Ralph	Université de Nice	5
6	Rygiel, Philippe	ENS de Lyon	3
6	Weil, Patrick	Paris 1	3
6	Wahl, Alfred	Université de Metz	3
6	Thénault, Sylvie	Paris 1	3
10	Guillon, Jean-Marie	Aix-Marseille	2

Tableau — Les dix directeurs-rices de thèse les plus représenté-e-s dans le corpus (1985–2024). Source : theses.fr.

## Une histoire très française

La plupart des thèses définissent un espace de référence, qui est souvent défini comme le territoire de destination d'un flux migratoire, qui correspond dans la grande majorité des cas à tout ou partie du territoire de la France métropolitaine (trois thèses sur quatre), plus marginalement aux territoires sous autorité française (5 %). Seulement 9 % des thèses évoquent un mouvement migratoire qui n'ait pas pour point de destination ou d'origine un territoire sous autorité française. Quelques-unes, en nombre croissant mais modeste, adoptent une perspective comparative qui leur permet d'évoquer d'autres espaces.

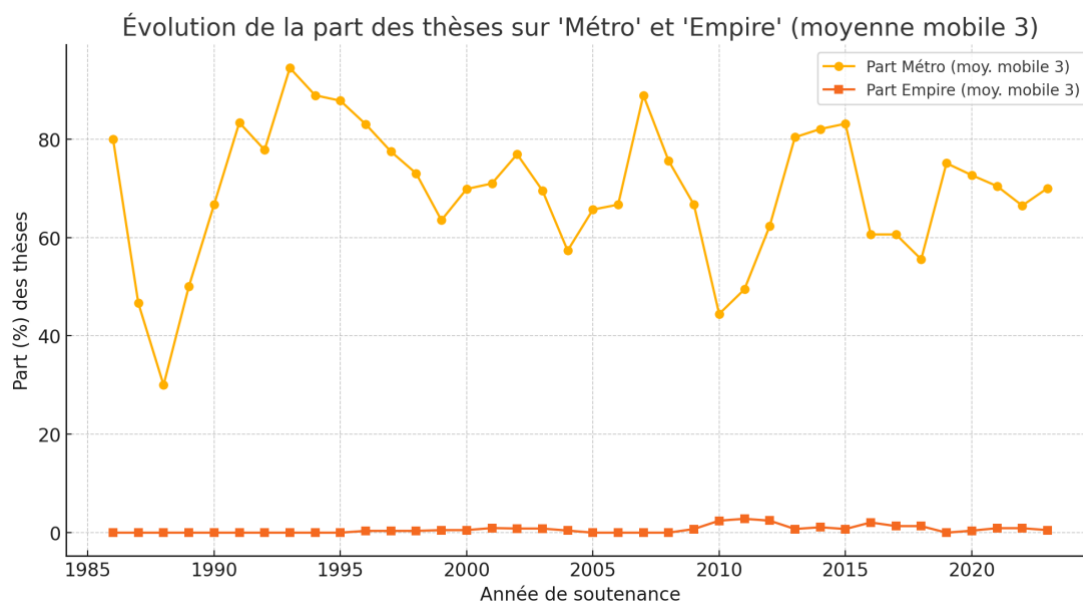


Figure 5. Évolution de la part des thèses consacrées au territoire métropolitain et à l'empire (moyenne mobile d'ordre 3). Source : theses.fr.

Période	Total thèses	Thèses comparatives	Part (%)
1980–1999	65	1	1,5
2000–2009	75	5	6,7
2010–2024	135	9	6,7

Tableau 1. Évolution de la part des thèses comparatives dans le corpus (source : theses.fr).

## Une histoire très contemporaine

Les thèses sont dans leur grande majorité consacrées à l'histoire migratoire du XX<sup>e</sup> siècle, plus particulièrement de sa partie centrale. La quasi-absence de thèses d'histoire consacrée à la période récente est remarquable. A peine cinq thèses (2 %) étudient une période débutant après 1980, toutes soutenues depuis 2014. L'histoire des migrations très contemporaines, fort agitée pourtant, est presque inexistante, ce qui surprend malgré les contraintes d'accès aux archives.

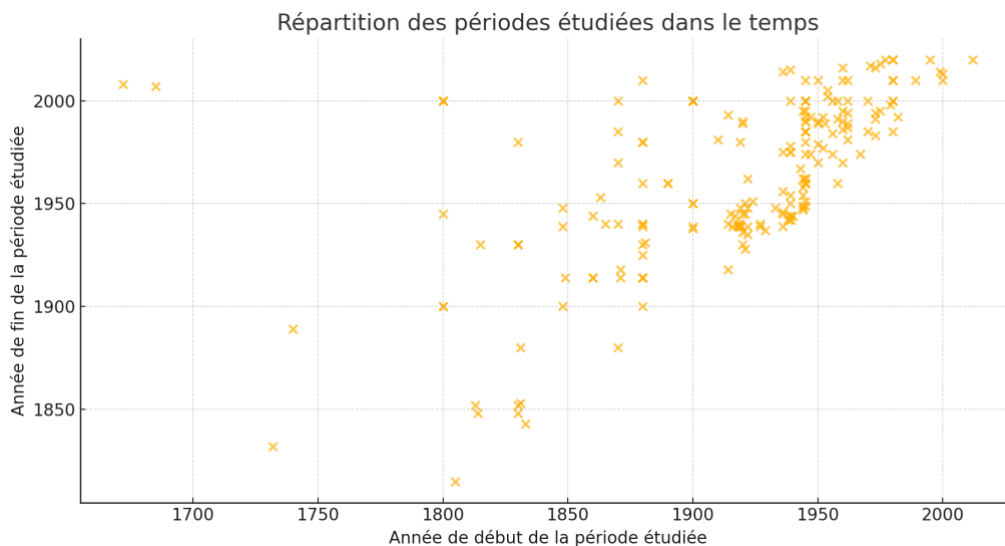


Figure 6. Répartition des périodes étudiées dans le temps. Source : theses.fr.

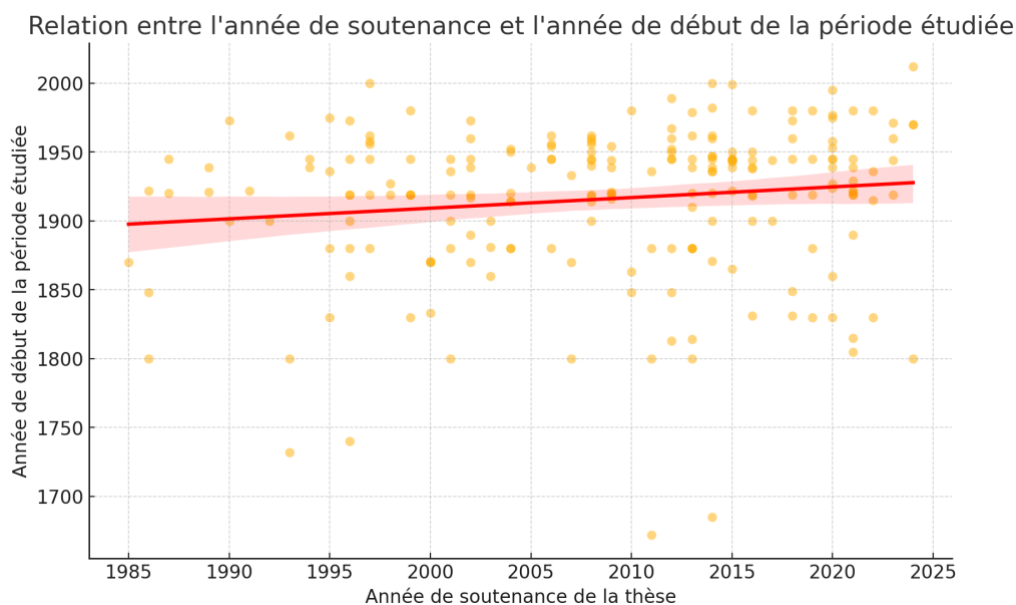


Figure 7. Relation entre l'année de soutenance et l'année de début de la période étudiée. Source : theses.fr.

## II. Un lexique très stable et très politique

Une étude lexicométrique rapide des titres et des résumés déposés sur theses.fr permet de disposer d'aperçus sur le vocabulaire de ces travaux. Précisons d'emblée les limites de l'exercice : nous ne travaillons pas sur le texte même des thèses, mais sur des éléments déposés dans un répertoire institutionnel, qui ne sont pas tous de même longueur ni rédigés avec le même soin.

Terme	Occurrences	Fréquence relative
France	1 058	0,232
immigration	355	0,078

Terme	Occurrences	Fréquence relative
politique	237	0,052
guerre	233	0,051
histoire	214	0,047
réfugiés	198	0,043
émigration	183	0,040
politiques	163	0,036
travail	159	0,035
français	146	0,032
étrangers	129	0,028
intégration	127	0,028
paris	126	0,028
immigrés	109	0,024
exil	99	0,022
italiens	91	0,020
algérie	86	0,019
espagnols	86	0,019
juifs	81	0,018
accueil	79	0,017

Tableau 2. Termes signifiants les plus fréquents dans les titres et résumés du corpus (source : theses.fr).

Si nous employons le vocabulaire du moment nous sommes tentés d'écrire que les termes les plus fréquemment rencontrés renvoient à un vocabulaire stato-centré. Deux absences sont particulièrement frappantes. La première est celle de la dimension économique dans les titres : ni les acteurs économiques collectifs ni les mécanismes de marché n'y apparaissent. Cela ne signifie pas que la dimension économique des migrations est absente du corpus. Une analyse des résumés révèle que le travail, l'emploi et la main-d'œuvre sont mentionnés dans 62 à 73 % des thèses selon les périodes<sup>7</sup>. Par contre l'économique y est traité comme contexte, plus que comme objet. La deuxième absence est celle du vocabulaire de la race, bien moins présent qu'il

---

<sup>7</sup>La dimension économique est présente dans 62 à 73 % des résumés selon les périodes (travail, emploi, main-d'œuvre, marché du travail), alors qu'elle est quasi absente des titres. Elle est traitée comme contexte, pas comme objet : l'immigration est située dans une économie, elle n'est pas étudiée comme phénomène économique.

ne le serait dans le monde anglophone, il n'apparaît que dans 9 % des résumés, proportion stable sur l'ensemble de la période.<sup>8</sup>

Aucun des termes du tableau ne voit sa fréquence varier significativement au cours de la période. Le lexique avec lequel s'écrit l'histoire des migrations dans les universités françaises est stable dans le temps

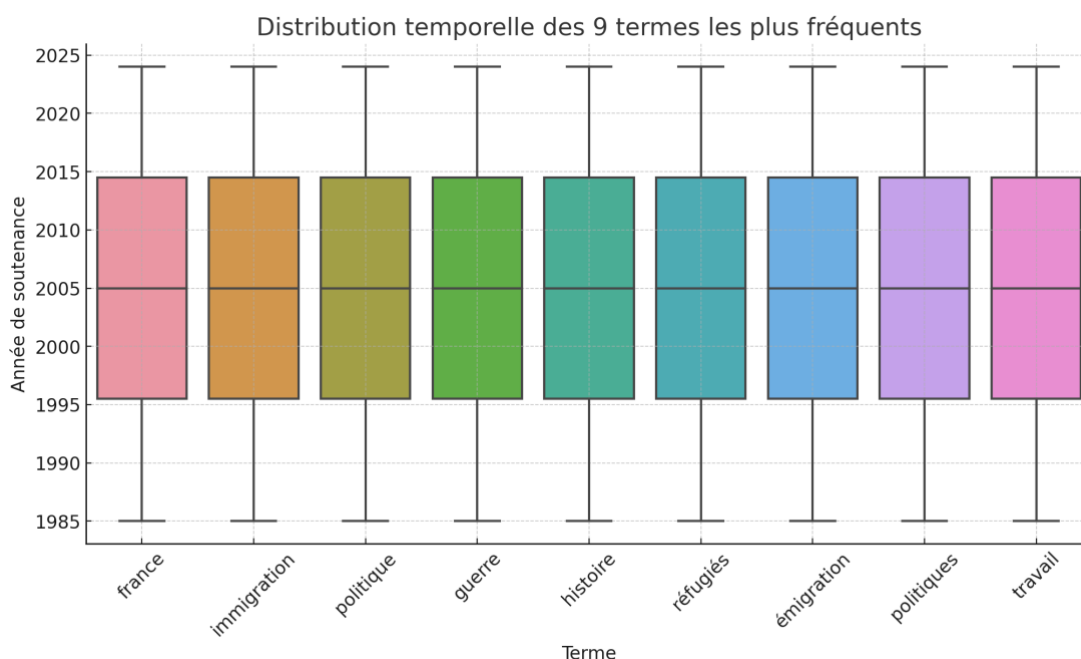


Figure 8. Distribution temporelle des 9 termes les plus fréquents dans le corpus. Source : theses.fr.

L'histoire que racontent ces mots est celle d'une France animée de flux la reliant à son empire ou à son ex-empire, aux périphéries proches de l'Europe méditerranéenne. Histoire très politique, scandée par les guerres, elle a souvent pour héros des exilés distribués entre plusieurs statuts, dont l'accueil et l'intégration constituent des enjeux.

### III. Variations et inflexions thématiques

Si le lexique est stable, les centres d'intérêt se déplacent au sein d'un questionnement dont les structures demeurent assez constantes. Pour le dire de manière plus imagée, le fait de parler la même langue n'empêche pas que les sujets de conversation changent. Cette section retrace ces inflexions en mobilisant, outre l'examen des populations étudiées par les auteurs des thèses, trois approches complémentaires, décrites dans la note méthodologique jointe<sup>9</sup> : une analyse

<sup>8</sup>La quasi-absence du vocabulaire de la race dans les résumés (19 thèses sur 210, soit 9 %, proportion stable sur l'ensemble de la période) contraste avec ce qu'on observerait dans la production anglophone et confirme que ce n'est pas un artefact des conventions de titrage mais une caractéristique structurelle de l'historiographie française.

<sup>9</sup>Les méthodes d'analyse thématique utilisées dans cette section sont décrites en détail dans la note méthodologique jointe : dictionnaires, seuils d'activation, corpus textuel analysé (titre + résumé + mots-clés), distinctions entre méthode par termes-témoins (expertise),

par termes-témoins issus de la connaissance du champ (méthode 1), une analyse par dictionnaires alignés sur la littérature internationale (méthode 2), et une classification automatique (méthode 3).

## Des populations inégalement étudiées

Beaucoup de thèses se définissent par l'étude d'une population définie par son origine. Les populations les plus anciennement étudiées sont les « vieilles migrations » européennes, cependant que les populations en provenance du Sud global suscitent un intérêt plus tardif.

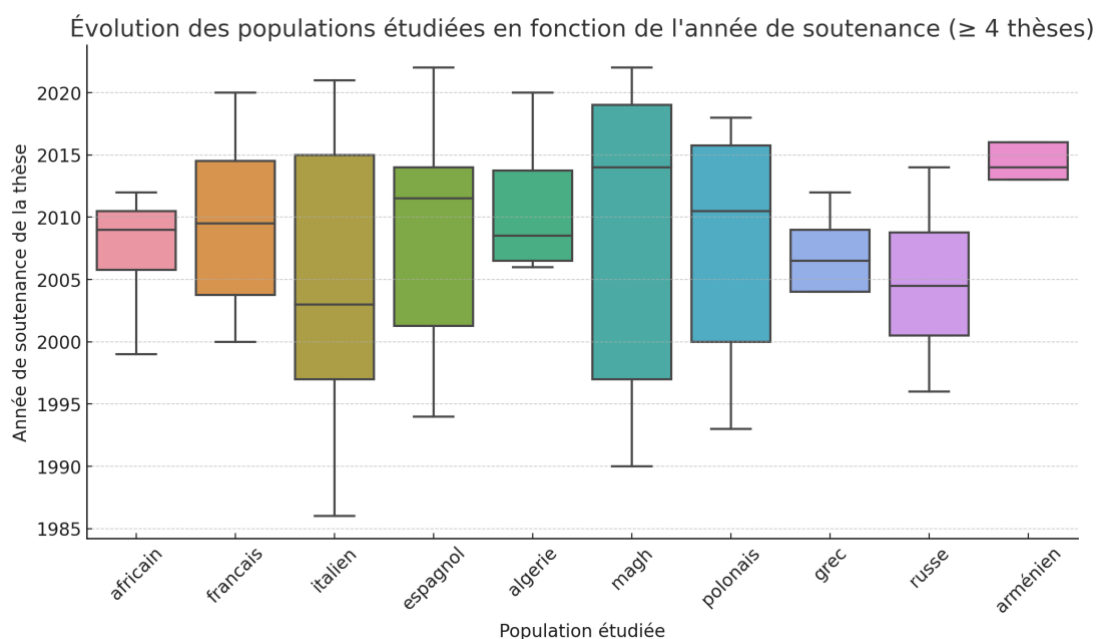


Figure 9. Évolution des populations étudiées en fonction de l'année de soutenance (populations représentées par au moins 4 thèses). Source : theses.fr.

La mise en regard des populations et des périodes historiques étudiées révèle un phénomène que les bilans historiographiques existants avaient peu mentionné : chaque population tend à être associée à une période de prédilection quasi-exclusive, qui constitue une convention implicite du champ plutôt qu'une contrainte archivistique ou une propriété du phénomène étudié. Étudier les Arméniens et les Russes, c'est étudier l'entre-deux-guerres (80 % et 75 % des thèses respectivement) ; étudier les Maghrébins ou les Algériens, c'est étudier le post-1945 (73 % et 70 %) ; étudier les Italiens, c'est avant tout étudier l'entre-deux-guerres (48 %). Les Espagnols constituent la seule exception notable, distribués sur l'ensemble de la chronologie, reflet peut être d'une émigration qui s'est étalée du XIX<sup>e</sup> siècle aux années 1970, mais c'est le cas aussi de l'immigration italienne.

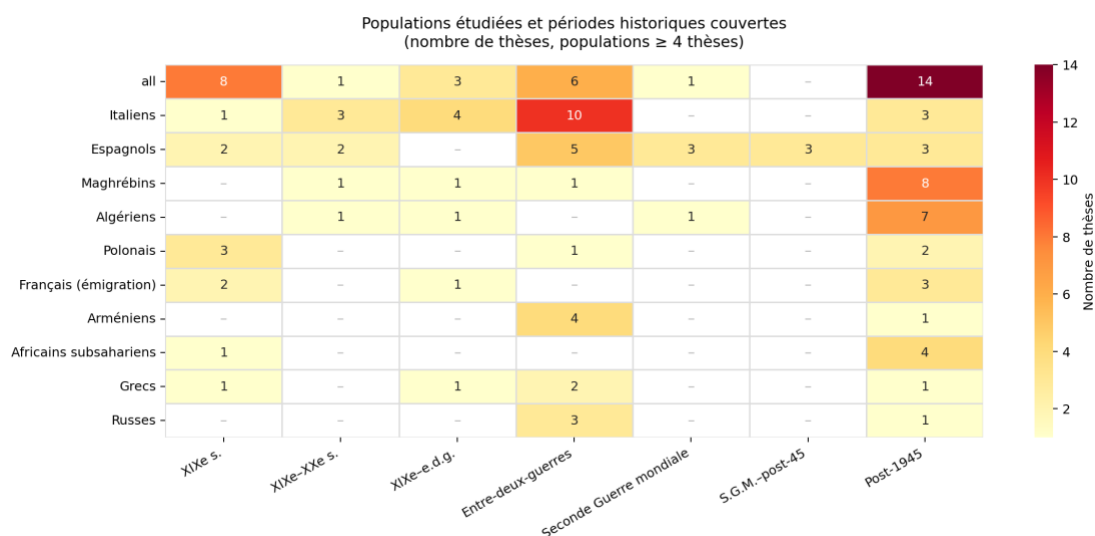


Figure 10. Populations étudiées et périodes historiques couvertes (nombre de thèses ; populations représentées par au moins 4 thèses). Source : theses.fr.

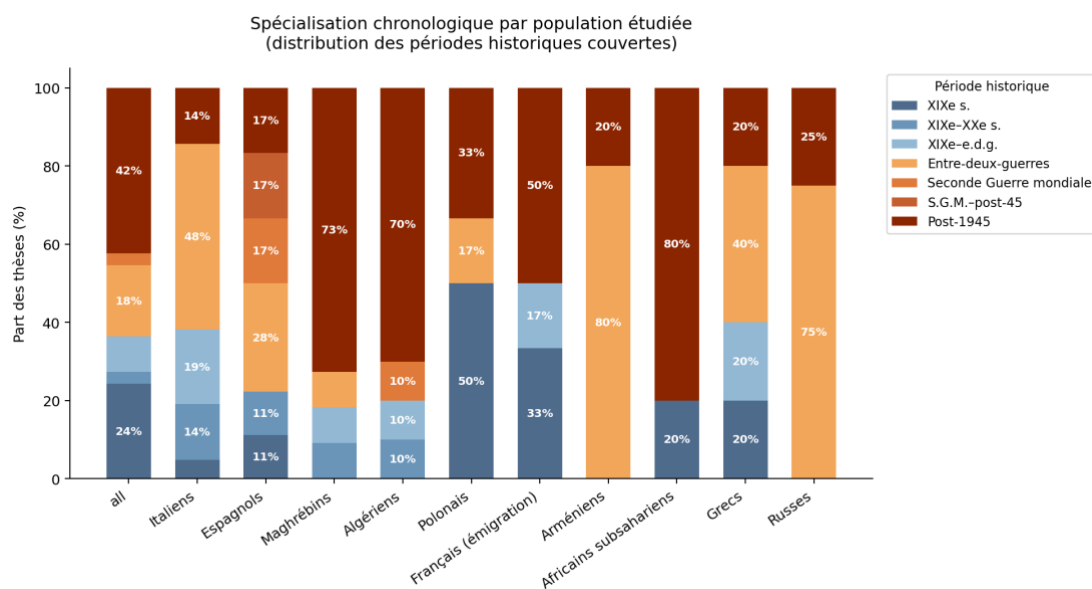


Figure 11. Spécialisation chronologique par population étudiée (distribution des périodes historiques couvertes, en %). Source : theses.fr.

Ces associations population × période témoignent de traditions de recherche qui ont fixé une lecture canonique de chaque groupe, inscrivant chaque population dans une époque, correspondant à une phase d'arrivées massives, plutôt que dans un processus de longue durée. Cette historiographie pense les migrations par population-événement davantage que par trajectoire étudiée sur le moyen ou le long terme.

## La sortie des usines

Si les premières thèses soutenues s'inscrivent souvent dans le sillage d'une histoire ouvrière et se caractérisent par l'emploi fréquent du vocabulaire industriel<sup>10</sup>, celui-ci voit son usage

<sup>10</sup>Termes retenus : usine, usines, ouvrier, ouvriers, industrie, industriel, industrielle, fabrication, atelier, production, ouvrières.

décroître à partir des années 1990 (Spearman sur tranches de 5 ans :  $p = 0,015$ ), avant de retrouver une certaine actualité au cours des dernières décennies. Ce reflux traduit un déplacement de l'histoire des migrations du côté des *cultural studies* et des approches par l'identité plutôt que par l'inscription sociale des migrants.

Période	Fréquence cumulée (vocabulaire ouvrier)
1980–1984	0,00226
1985–1989	0,00299
1990–1994	0,00211
1995–1999	0,00173
2000–2004	0,00054
2005–2009	0,00120
2010–2014	0,00076
2015–2024	0,00095

Tableau 3. Évolution de la fréquence du vocabulaire ouvrier dans le corpus ( $p$ -value Spearman : 0,015).

### L'horizon perdu de l'intégration ?

Si la grande question, du milieu des années 1990 et jusqu'au milieu des années 2000, est l'intégration ou l'assimilation des populations d'origine étrangère, l'usage du terme et de ceux qui lui sont associés décroît à partir de 2005, sans jamais disparaître.

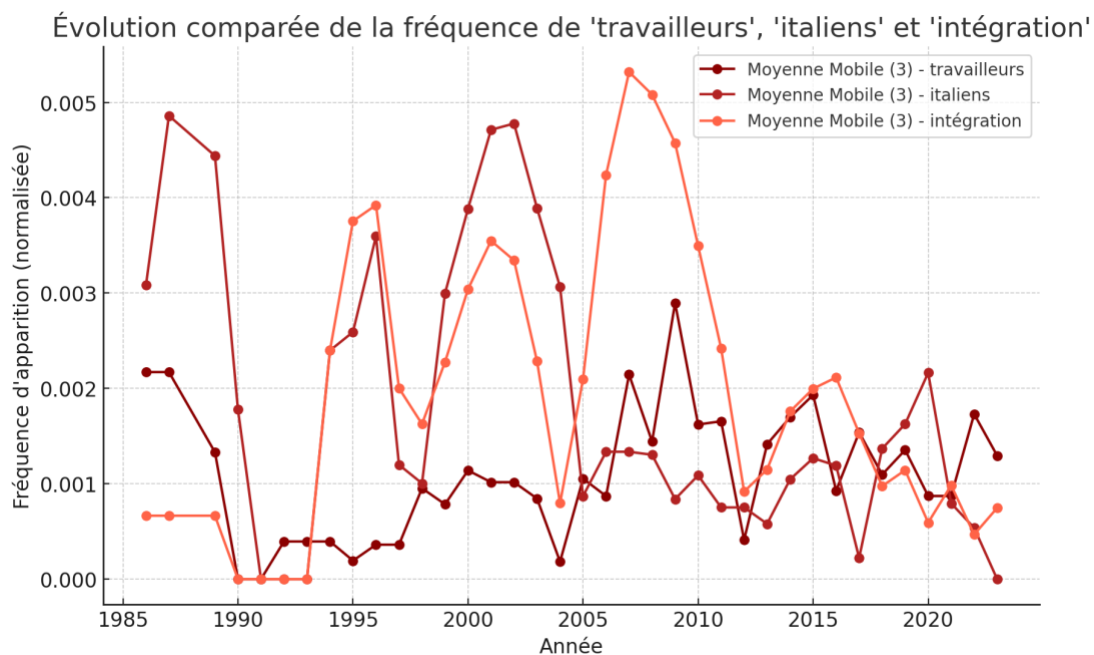


Figure 12. Évolution comparée de la fréquence des termes « travailleurs », « italiens » et « intégration » (moyenne mobile d'ordre 3). Source : theses.fr.

Fréquence cumulée du cluster Intégration et Assimilation par périodes

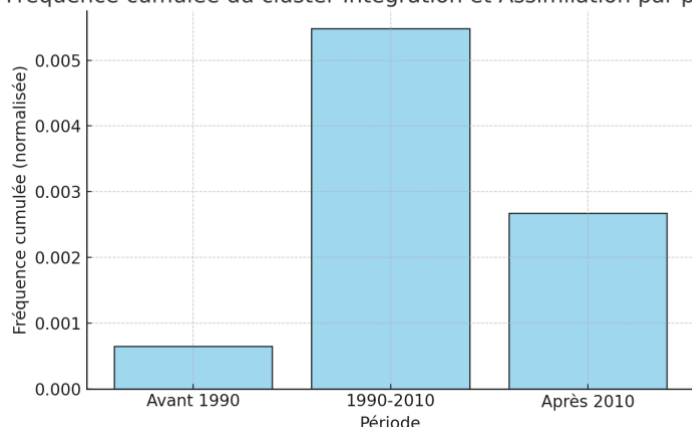


Figure 13. Fréquence cumulée du cluster « Intégration et assimilation » par grandes périodes. Source : theses.fr.

## Une histoire qui se genre timidement

Le vocabulaire lié au genre et à l'identité sexuelle émerge assez tôt dans le corpus, mais avec des fréquences très modestes. Il connaît une forte croissance à partir des années 2000, puis une stabilisation. Les corrélations sont faiblement significatives (Spearman : 0,67,  $p = 0,071$ ). Il est tentant de renvoyer cette évolution aux mutations affectant l'ensemble des productions des sciences sociales au cours de la période. Nous ne disposons pas cependant de données qui nous permettraient de nous prononcer sur l'éventuelle originalité des évolutions affectant ce sous-champ.

Décennie	Fréquence normalisée
1980	0,014
1990	0,145
2000	0,101
2010	0,391
2020	0,348

Tableau 4. Évolution de la fréquence du vocabulaire de genre dans le corpus.

## Un tournant colonial assez précoce

Le vocabulaire colonial<sup>11</sup> enregistre une progression plus nette (Spearman : 0,72,  $p = 0,051$ ). Le premier colloque de la Cité nationale de l'histoire de l'immigration, en 2006, présentait les interactions entre études coloniales et migratoires comme un front pionnier appelé à se développer. Si tel fut effectivement le cas, notre corpus conduit à nuancer ce constat : le colloque inaugural de la CNHI, en ce domaine, consacrait plus qu'il n'initiait. Nos données conduisent en effet à conclure à une timide émergence dès la première moitié des années 1990, puis un usage croissant jusqu'à un plateau vers 2010.

<sup>11</sup>Termes retenus : colonie, colonial, colonisation, empire, métropole, décolonisation, impérial, indigène, indigènes, postcolonial, coloniale.

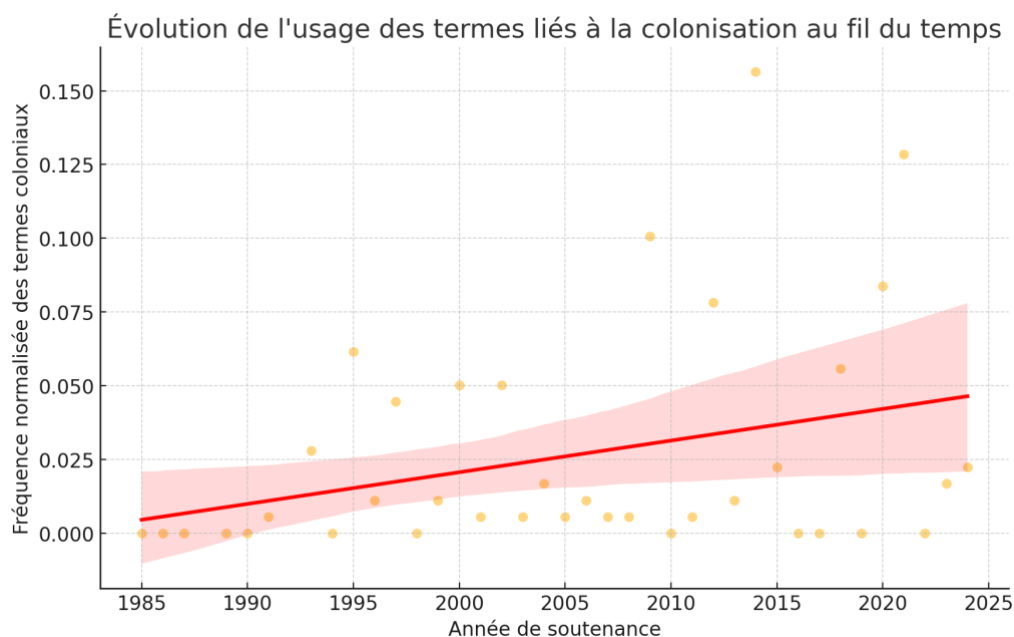


Figure 14. Évolution de l'usage des termes liés à la colonisation au fil du temps. Source : theses.fr.

## Un tournant spatial discret

Il est plus difficile, armé des seuls résumés, de distinguer des évolutions dans les méthodologies ou les techniques de recherche mobilisées. Un cependant fait exception, l'intérêt croissant pour l'inscription spatiale des phénomènes migratoires<sup>12</sup>. La corrélation n'est significative que sur des tranches de dix ans (Spearman : 0,90,  $p = 0,037$ ), ce qui signale une tendance longue plutôt qu'une rupture datée.<sup>13</sup>

## Frontières et contrôles

La prise en compte des résumés permet l'émergence d'un élément que les analyses lexicales sur les seuls titres ne permettaient pas de percevoir. La progression en fréquence du vocabulaire lié à l'administration et au contrôle des migrations constitue l'inflexion thématique la plus marquée des dernières années.<sup>14</sup> La part des thèses activant ce cluster passe de 0 % dans les années 1990 à 10 % dans les années 2000, 16 % dans les années 2010 et 24 % dans les années 2020. C'est le signal le plus net de changement thématique récent, plus que le tournant colonial.

<sup>12</sup>Termes retenus : territoire, région, quartier, espace, zone, agglomération, province, département, commune, banlieue, secteur, cartographie, urbanisation, paysage, habitat, frontière, implantation, répartition.

<sup>13</sup>Dirk Hoerder, « The traffic of emigration via Bremen/Bremerhaven : Merchants' interests, protective legislation, and migrants' experiences », *Journal of American Ethnic History* 13, n° 1 (1993) : 68–101.

<sup>14</sup>Termes retenus pour ce cluster : police, contrôle, expulsion, frontière, titre de séjour, régularisation, sans-papier, rétention, catégorisation. La part des thèses activant ce cluster passe de 0 % dans les années 1990 à 10 % dans les années 2000, 16 % dans les années 2010 et 24 % dans les années 2020.

Au regard de l'importance accordée par les travaux fondateurs à ces dimensions, le résultat surprend même si l'on est intuitivement amené à le rapporter à la prééminence de la question du contrôle et des frontières dans les débats politiques récents

### **Regarder autrement, classifications automatiques**

La classification des thèses soutenues sur la base de dictionnaires préétablis à des limites, l'une, évidente, étant qu'elle ne permet pas de faire apparaître des résultats inattendus parce ce qu'un thème était absent du schéma de connaissance du champ produit a priori par un expert. Nous avons voulu donc compléter cet aperçu en produisant une classification automatique des thèses du corpus utilisant Word2Vec<sup>15</sup>. Appliquée en incluant la variable de période historique étudiée, elle produit quatre clusters bien séparés : migrations coloniales et longue durée ancienne, migrations européennes et entre-deux-guerres, migrations post-coloniales et débats politiques contemporains, études locales et microhistoire. Ces résultats sont cohérents avec ce que l'analyse lexicale révèle par ailleurs.

Mais la confrontation de deux classifications, que l'on confronte l'une à l'autre, la première prenant en compte la variable décrivant la période étudiée, l'autre non, est en elle-même instructive.<sup>16</sup> Lorsqu'on retire la variable temporelle pour ne conserver que le contenu textuel des résumés, les clusters s'effondrent en un nuage dense sans séparation thématique claire : les mêmes mots (guerre, politique, histoire, intégration, émigration) dominent tous les groupes. Le principe organisateur de notre corpus est la période étudiée, plus que les objets ou les thématiques retenues. Nous pouvons être tentés d'en conclure que l'histoire des migrations produite dans les universités françaises pense ses objets d'abord par époque, ce qui rejoint l'observation sur les conventions chronologiques implicites associées à chaque population.

---

<sup>15</sup>Tomas Mikolov et al., « Efficient estimation of word representations in vector space », *arXiv* 1301.3781 (2013).

<sup>16</sup>La classification automatique (Word2Vec + K-Means, k = 4) produit des clusters qui se séparent nettement mieux lorsque la variable de début de période est incluse (score de silhouette 0,41) que lorsqu'on travaille sur le contenu textuel seul (score de silhouette 0,49 mais clusters sans correspondance thématique claire). Les deux solutions sont présentées dans la figure comparative jointe.

### Classification automatique des thèses : effet de la variable chronologique

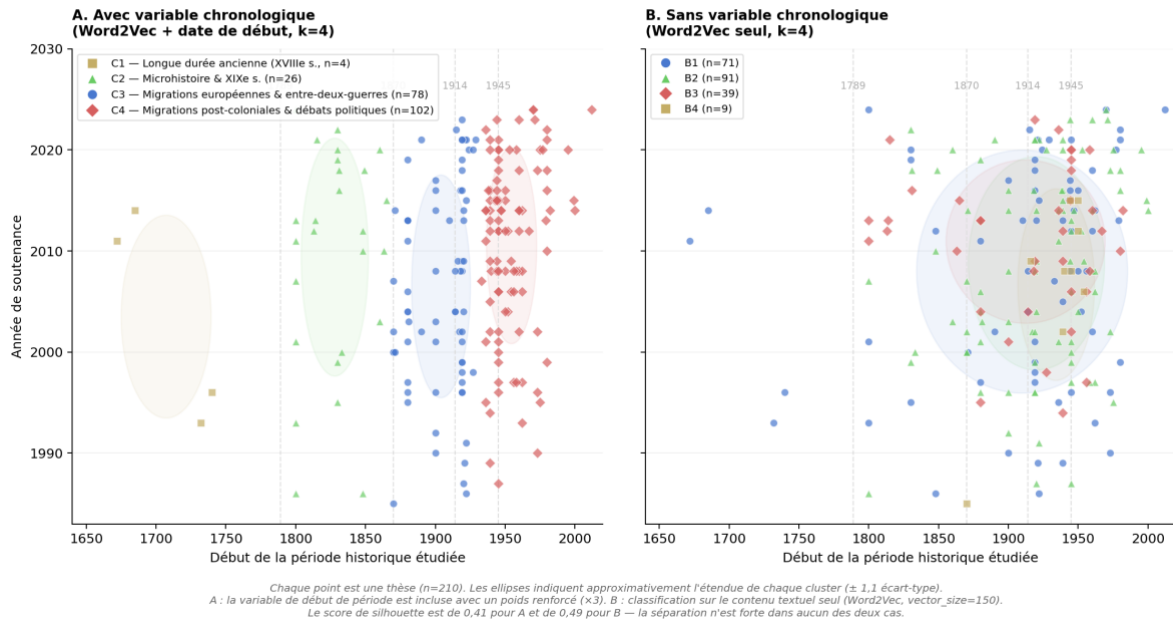


Figure 15. Classification automatique des thèses : effet de la variable chronologique. A : avec variable de période historique. B : Word2Vec seul, sans variable chronologique. La séparation des clusters disparaît quasi entièrement lorsque la chronologie est retirée. Source : theses.fr.

## IV. Une production insérée dans des dynamiques globales

### Les thèmes des migrations studies

Les contextes dans lesquels peuvent être insérés ces productions sont bien sûr multiple. Je mettrai ici l'accent sur ceux dont nous sommes peut-être les moins conscients : l'insertion de nos textes dans des conversations globalisées qui ont les migrations pour objet, au sein d'un univers savant lui aussi globalisé et largement transdisciplinaire.

Pisarevskaya et ses collaborateurs, qui ont systématiquement étudié la littérature savante traitant des migrations internationales publiée entre 1986 et 2017<sup>17</sup>, ont établi une cartographie de grands clusters thématiques permettant de décrire toutes disciplines confondues, cet ensemble, très largement dominé par les productions anglophones et d'autres sciences sociales que l'histoire. La figure ci-dessous présente à gauche les grandes sections thématiques structurant la littérature internationale ; à droite la distribution temporelle de clusters analogues dans notre corpus français. L'analogie est loin d'être parfaite produite par la définition de dictionnaires les plus proches possibles de ceux décrivant les catégories de la littérature internationale.

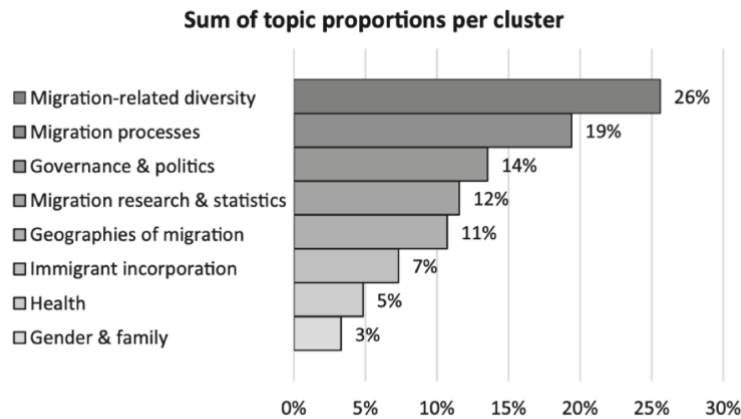


Figure 17. Proportion relative des grands clusters thématiques dans la littérature internationale sur les migrations (1986–2017). Source : Pisarevskaya et al. (2020).

466 • A. PISAREVSKAYA ET AL.



Figure 18. Âge moyen et importance relative des clusters dans la littérature internationale (la taille des bulles correspond à l'importance du cluster ; la position sur l'axe correspond à l'âge moyen des publications). Source : Pisarevskaya et al. (2020).

La comparaison appelle plusieurs observations. La première concerne les **convergences attendues** : la prédominance de « Migration-related diversity » (intégration, assimilation, identité) dans les deux corpus, la progression (assez tardive) de « Gender & family », la montée constante de « Governance & politics ». Ces convergences entre corpus qui ne se recoupent pas, Pisarevskaya et al. ne prennent pas en compte la littérature francophone et l'histoire y pèse très peu de choses dans leur corpus, suggèrent une généralisation des problématiques à l'échelle occidentale et des sciences sociales toutes entières, que nous pouvons supposer liée à la convergence des contextes politiques.

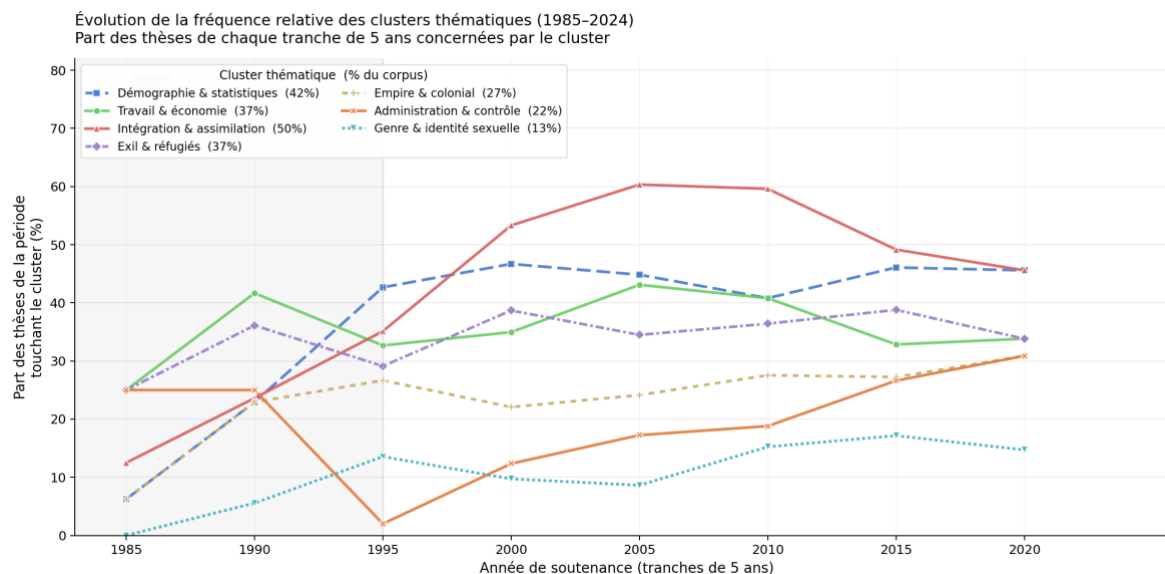


Figure 16. Évolution de la fréquence relative des clusters thématiques (méthode par dictionnaires alignés sur Pisarevskaya et al., 1985–2024). La taille de la légende indique la part du cluster dans le corpus total. Source : theses.fr.

A l'inverse, le cluster « Histoire coloniale et postcoloniale », absent de la nomenclature de Pisarevskaya et al., est l'un des plus actifs dans le corpus français. Nous pouvons considérer que c'est un écart attendu, lié aux spécificités impériales de l'histoire française et à l'importance des flux migratoires post-coloniaux dans l'histoire française. En sens inverse, le cluster « Health » (5 % de la littérature internationale) est complètement absent du corpus français. Les dimensions sanitaires de la migration, mais aussi l'importance des considérations sanitaires dans la genèse des mécanismes de contrôle actives dans la recherche internationale, et ayant suscité depuis longtemps déjà l'intérêt d'historiens, n'ont pas pénétré l'historiographie nationale. De même, le cluster « Histoire ouvrière histoire du travail », semble spécifique au contexte français. Sous bénéfice d'inventaire, il peut renvoyer au poids de l'histoire sociale d'inspiration marxiste dans la formation du champ en France comme au rôle clé joué par les processus migratoires dans la constitution de la classe ouvrière française et ce très tôt dans son histoire, à la différence de ce que l'on peut observer dans le cas d'un certain nombre d'états européens de taille similaire.

Les propriétés des deux corpus jouent également un rôle lorsqu'il s'agit d'interpréter cet écart. Le corpus de Pisarevskaya et al. est transdisciplinaire et largement dominé par les sciences sociales. Le corpus français est par définition un corpus d'histoire : cela explique également l'importance moindre du cluster « Migration processes » dans notre corpus, plus lié aux approches sociologiques et économiques.

## Des contextes politiques convergents

Pour comprendre ces convergences et ces écarts, il faut garder en mémoire le contexte de production de ces textes : un Occident dont les différents segments connaissent au cours de la période des évolutions politiques et idéologiques très similaires. Les années 1970 voient la politisation de la question migratoires ; le début des années 1980 la réémergence des extrêmes droites ; 1990 est en Europe marqué par le tournant Schengen dont l'un des effets est de redéfinir le migrant qui tendanciellement devient celui qui n'est pas né en Europe. Vers 2015

éclate ce que l'on a nommé la « crise » de l'accueil, qui manifeste aussi, ou rend visible, une militarisation des frontières bien entamée. Partout ensuite progresse ce que Hollifield appelle la posture nativiste<sup>18</sup> dont les tenants circulent, échantent, tendant à former une sorte d'internationale de l'ethno nationalisme.

Il n'y a pas de simultanéité parfaite entre ces inflexions politiques et inflexions académiques, mais il est possible de les relier en utilisant la grille d'analyse proposée par Reinecke<sup>19</sup>. Celui-ci distingue trois usages de l'histoire dans les *migration studies* : le mode « histoire profonde » (régularités sur longue période), le mode « généalogique » (trajectoire des catégories et des politiques), et le mode « contrapuntique » (mise en lumière des histoires négligées). Si la première période de notre corpus est marquée par une comparaison implicite entre le *then* et le *now* avec comme enjeu l'intégration, on observe ensuite un déplacement vers une perspective davantage généalogique. L'étude des mécanismes du contrôle migratoire et de la construction des catégories administratives prend une tournure généalogique. Le pic récent du cluster « Governance & politics » dans les deux corpus en est la manifestation la plus visible.

## **Conclusion : à quoi sert cette histoire ?**

Un constat s'impose. Ni l'accumulation de connaissances ni les stratégies discursives des historiens n'ont beaucoup pesé sur le débat public et un constat similaire pourrait sans doute être posé pour l'ensemble des travaux de sciences sociales. La production historique apparaît plus comme un produit de son contexte que comme une instance permettant de l'infléchir ou de le renommer.

Faut-il pour autant désespérer de cette histoire ? Je ne le crois pas, à condition de préciser à qui elle peut servir quelles sont ces fonctions. La première de celle-ci est une fonction de compréhension, débouchant sur le possible exercice d'une expertise, avec toutes les ambiguïtés associées à ce terme et les illusions parfois sur la puissance d'une intervention experte adossée à une connaissance savante. Cette histoire rend intelligibles des phénomènes que les cadres nationaux tendent à réifier et à naturaliser les enfermant dans les catégories de l'État-nation, toujours pensé de l'intérieur comme une singularité. L'histoire des migrations a aussi une fonction mémorielle et identitaire : écrire l'histoire des migrations, c'est aussi écrire l'histoire de France, c'est marquer surtout pour les groupes et les individus leur appartenance à l'histoire de France. Il est plusieurs façons en effet de comprendre le « leur histoire est notre histoire » de Gérard Noiriel, devenu mot d'ordre pour le Musée national de l'histoire de l'immigration.

Enfin, pour l'avoir, avec d'autre expérimenté, je tiens que cette histoire peut nourrir, localement souvent, un certain nombre d'acteurs de la société civile. Syndicats, associations, collectifs, enseignants également trouvent dans ce corpus des ressources argumentatives, des précédents, des contre-récits, des ressources mobilisables. Encore faut-il que les questions posées résonnent avec le monde auquel ils sont confrontés et que des médiations qui je crois passent plus par la

---

<sup>18</sup>James F. Hollifield, *Immigrants, Markets and States* (Cambridge MA : Harvard University Press, 1992) ; James F. Hollifield, « The emerging migration state », *International Migration Review* 38, n° 3 (2004) : 885–912.

<sup>19</sup>Christoph Reinecke, « Three different uses of history in reflexive migration studies », *Migration Studies*, 2023.

rencontre et l'échange que par la production de contenus permettent à l'histoire de jouer son rôle de dénaturisatrice des évidences.

La production universitaire est inégalement ajustée à ces fonctions, peu comparative, peu attentive aux dimensions économiques comme objet, structurée d'abord par des conventions chronologiques qui fixent chaque population dans son époque, elle reflète les structures de l'université française autant que les réalités des migrations et des débats qu'elles suscitent. La prise de conscience de ces limites est peut-être utile à une réflexion sur sa fonction civique.

# Note méthodologique : analyse thématique du corpus

## 1. Trois méthodes complémentaires

L'analyse thématique du corpus repose sur trois méthodes distinctes qui ne répondent pas aux mêmes questions et ne doivent pas être confondues.

**Méthode 1 — Termes-témoins issus de la connaissance du champ (présente section : § 2).** Des termes ou groupes de termes signalant des inflexions historiographiques identifiées par l'expertise de l'auteur sont sélectionnés et leur évolution temporelle est retracée dans le corpus. La démarche est *confirmatoire* : il s'agit de vérifier quantitativement et de dater des perceptions expertes formées *a priori*.

**Méthode 2 — Présence de thèmes par dictionnaires alignés sur Pisarevskaya et al. (§ 3).** Des dictionnaires sont construits *a posteriori* en s'alignant sur les catégories de la littérature internationale<sup>20</sup>, avec les adaptations nécessaires au vocabulaire français. La démarche permet une comparaison structurelle entre le corpus français et la production internationale.

**Méthode 3 — Classification automatique (Word2Vec + K-Means).** Chaque thèse est représentée par un vecteur appris sur le contenu des résumés, puis assignée à un cluster par l'algorithme K-Means. La démarche est purement *inductive* : les clusters émergent des données sans intervention thématique préalable.

La confrontation des résultats des trois méthodes est elle-même un résultat : là où elles convergent, la conclusion est robuste ; là où elles divergent, la divergence est informative sur la nature de ce qu'elles capturent respectivement.<sup>21</sup>

## 2. Méthode 1 — Termes-témoins et perceptions expertes

### 2.1 Principe

Cette méthode part de la connaissance que l'auteur a du champ historiographique pour identifier des **termes-témoins** : des mots ou groupes de mots dont la présence ou la progression dans le corpus signale une orientation thématique ou une inflexion dont il postule l'existence.

Chaque thème est associé à une **hypothèse historiographique explicite**, formulée avant l'analyse. C'est l'explicitation de cette hypothèse qui distingue la méthode 1 de la méthode 2 : dans la première, le chercheur dit *pourquoi* il cherche ce terme avant de le chercher. Les résultats peuvent confirmer, infirmer ou nuancer l'hypothèse.

---

<sup>20</sup>Asya Pisarevskaya et al., « Mapping migration studies : an empirical analysis of the coming of age of a research field », *Migration Studies* 8, n° 3 (2020) : 455–481.

<sup>21</sup>La méthode est décrite dans : P. Rygiel, *Écrire l'histoire des migrations. Une vue de France*, étude de cas sur la comparaison des résultats selon la méthode utilisée.

Le corpus textuel analysé est le même que pour les autres méthodes : le texte agrégé de chaque thèse, constitué par la concaténation en minuscules du titre, du résumé corrigé et des mots-clés automatiques.<sup>22</sup>

La recherche est menée sur les **occurrences brutes** (à la chaîne, insensible à la casse), normalisées par le nombre total de mots du corpus pour chaque tranche temporelle, afin de produire des **fréquences relatives**. Les corrélations temporelles sont testées par le coefficient de Spearman (non paramétrique), sur des tranches de cinq ans et de dix ans. Les seuils de significativité et les intervalles de confiance sont indiqués dans le texte pour chaque thème.

## 2.2 Les cinq thèmes retenus et leurs dictionnaires

Cinq inflexions historiographiques ont été soumises à ce test. Les listes de termes sont reproduites telles qu'elles figurent dans le code d'analyse (voir dépôt GitHub), avec les éventuelles coquilles et répétitions du code original conservées pour assurer la reproductibilité.

### T1. Histoire ouvrière — histoire du travail

**Hypothèse** : L'hypothèse est celle d'un reflux de l'histoire ouvrière, au profit d'approches par la culture ou l'identité.

**Termes retenus** : « usine », « usines », « ouvrier », « ouvriers », « industrie », « industriel », « industrielle », « fabrication », « atelier », « production », « ouvrières »

**Source dans le code** : *termes\_travail\_industriel* dans le code original.

### T2. Intégration et assimilation

**Hypothèse** : L'hypothèse est d'un pic de centralité de cette problématique dans les années 1995–2005, suivi d'un reflux sans disparition, correspondant à la montée puis à la remise en question du paradigme intégrationniste.

**Termes retenus** : « identifiées », « intégrations », « identités », « intégration », « intégrée », « identité », « intégré », « intègrent », « assimilation », « identitaire », « intégrés », « identity », « réintégration », « identitaires », « intégrer »

**Source dans le code** : *termes\_cluster* dans le code original.

### T3. Genre et identité sexuelle

**Hypothèse** : L'hypothèse est d'une émergence tardive et encore incomplète de cette dimension dans l'historiographie des migrations.

**Termes retenus** : « femme », « femmes », « genre », « sexu\* », « masculin\* », « féminin\* », « féminisme », « domestique », « maternité », « prostitu\* »

**Source dans le code** : *Fourni par l'auteur ; liste reconstituée à partir des analyses menées sur le corpus.*

---

<sup>22</sup>Le texte agrégé est la concaténation en minuscules du titre, du résumé corrigé et des mots-clés automatiques (*Automatic Tags*) tels que déposés sur theses.fr.

#### T4. Histoire coloniale et postcoloniale

**Hypothèse :** L'hypothèse est celle d'une progression forte de cette thématique.

**Termes retenus :** « colonies », « colonie », « colonial », « colonisation », « colonisés », « empire », « métropole », « décolonisation », « impérial », « indigène », « indigènes », « postcolonial », « coloniale »

**Source dans le code :** *Fourni par l'auteur.*

#### T5. Tournant spatial

**Hypothèse :** L'hypothèse est d'un intérêt croissant pour l'inscription spatiale des phénomènes migratoires, lié à la généralisation des SIG et à l'influence du *spatial turn* dans les sciences sociales.

**Termes retenus :** « territoire », « région », « quartier », « espace », « zone », « agglomération », « province », « département », « commune », « banlieue », « secteur », « métropole », « cartographie », « urbanisation », « paysage », « habitat », « frontière », « centre », « pôle », « implantation », « répartition »

**Source dans le code :** *termes\_spatiaux\_trouves dans le code original.*

#### 2.3 Précaution : non-exclusivité

Les thèmes de la méthode 1 ne sont pas exclusifs : une même thèse peut contribuer à plusieurs d'entre eux simultanément.<sup>23</sup> La fréquence cumulée d'un cluster est une mesure de sa **présence relative dans le champ**, pas de sa proportion dans une partition du corpus.

### 3. Méthode 2 — Dictionnaires alignés sur Pisarevskaya et al.

#### 3.1 Principe

La méthode 2 vise à rendre le corpus français comparable à la littérature internationale analysée par Pisarevskaya et al.<sup>24</sup>, en utilisant les mêmes grandes catégories thématiques, traduites et adaptées au vocabulaire français. Cette comparaison est **structurelle, non quantitative** : les proportions absolues ne sont pas directement comparables (méthodes différentes, corpus de taille très différente), mais la présence ou l'absence de clusters et leurs trajectoires temporelles respectives font sens.

Deux clusters n'ont pas d'équivalent chez Pisarevskaya et al. et ont été ajoutés : « Histoire coloniale & postcoloniale » (cluster 8) et « Histoire ouvrière — histoire du travail » (cluster 9). Leur présence dans le corpus français sans équivalent international constitue un écart

---

<sup>23</sup>Les clusters n'étant pas exclusifs, la somme des pourcentages dépasse 100 %. Une thèse d'histoire ouvrière traitant de la condition féminine activera simultanément les thèmes T1 et T3. La figure présente des fréquences relatives indépendantes, non une partition.

significatif. Inversement, le cluster « Health » de Pisarevskaya et al. est absent du corpus français.<sup>25</sup>

### 3.2 Corpus analysé et seuils

Le corpus textuel est identique à celui de la méthode 1 (texte agrégé titre + résumé + mots-clés).<sup>26</sup> La recherche de termes utilise la même procédure (correspondance insensible à la casse, normalisation par le nombre total de mots).<sup>27</sup>

Le **seuil d'activation** est fixé à **2 termes distincts** du dictionnaire pour la quasi-totalité des clusters, et à **1** pour le cluster « Genre, famille & sexualité » (cluster 7), dont la sous-représentation lexicale structurelle justifie un seuil plus permissif. Un seuil de 1 est trop permissif en général : avec des dictionnaires de 8 à 12 termes et des résumés de plusieurs centaines de mots, activer un seul terme peut résulter d'une occurrence contextuelle plutôt que d'une orientation thématique réelle.

### 3.3 Dictionnaires retenus

#### 1. Migration research & statistics

**Correspondance Pisarevskaya : *Migration research & statistics* — Seuil : 2**

« démographi\* », « statistiqu\* », « recensement », « flux migratoire », « effectif », « quantitatif », « chiffre », « mesure », « taux », « solde migratoire »

#### 2. Migration-related diversity

**Correspondance Pisarevskaya : *Migration-related diversity* — Seuil : 2**

« intégration », « assimilation », « insertion », « identité », « acculturation », « naturalisation », « diversité », « multicultural\* »

#### 3. Migration processes

**Correspondance Pisarevskaya : *Migration processes* — Seuil : 2**

« trajectoire », « mobilité », « réseau », « passeur », « agent », « modalité », « itinéraire », « circulation », « transnational\* », « chaîne migratoire », « transit »

#### 4. Governance & politics

**Correspondance Pisarevskaya : *Governance & politics* — Seuil : 2**

---

<sup>25</sup>Le cluster « Health » de Pisarevskaya et al. (5 % de la littérature internationale) est absent du corpus français. Cet écart est en lui-même un résultat significatif.

<sup>27</sup>L'astérisque (\*) signale une recherche tronquée couvrant les formes fléchies et dérivées. Les termes sans astérisque sont recherchés en correspondance exacte sur le mot entier (*word boundary*).

« politique migratoire », « contrôle », « frontière », « expulsion », « administration », « titre de séjour », « régularisation », « sans-papier\* », « rétention », « catégorisation », « police des étrangers », « législation »

## 5. Geographies of migration

**Correspondance Pisarevskaya :** *Geographies of migration* — Seuil : 2

« territoire », « quartier », « espace », « région », « banlieue », « ville », « urbanisation », « cartographie », « implantation », « habitat », « département », « commune »

## 6. Immigrant incorporation

**Correspondance Pisarevskaya :** *Immigrant incorporation* — Seuil : 2

« accueil », « parcours », « ascension sociale », « mobilité sociale », « communauté », « association », « milieu »

## 7. Genre, famille & sexualité

**Correspondance Pisarevskaya :** *Gender & family* — Seuil : 1

« femme », « genre », « féminin\* », « masculin\* », « sexualité », « famille », « maternité », « regroupement familial », « enfant », « descendant »

## 8. Histoire coloniale & postcoloniale

**Correspondance Pisarevskaya :** — (*spécifique au corpus*) — Seuil : 2

« colonial\* », « colon\* », « empire », « décolonisation », « indigène\* », « postcolonial\* », « métropole », « travailleur colonial », « engagé »

## 9. Histoire ouvrière — histoire du travail

**Correspondance Pisarevskaya :** — (*spécifique au corpus*) — Seuil : 2

« travailleur\* », « ouvrier\* », « usine\* », « industriel\* », « atelier », « grève », « syndicat », « condition\* de travail », « main-d'œuvre », « recrutement », « salaire »

## 10. Exil & réfugiés

**Correspondance Pisarevskaya :** *Partiellement : Forced migration & asylum* — Seuil : 2

« réfugié\* », « exil\* », « asile », « déplacé\* », « apatride », « persécution », « humanitaire », « demandeur d'asile », « protection »

## 5. Limites communes et précautions d'interprétation

**Le texte analysé est un proxy, pas le texte de la thèse.** Titres, résumés et mots-clés déposés sur theses.fr sont de longueur et de qualité rédactionnelle variables. Une thèse au résumé pauvre peut ne pas activer de cluster, non parce qu'elle est sans thème, mais parce que son vocabulaire déposé est insuffisant. Les analyses menées sur les seuls titres aboutissent à des distributions différentes, parce que le vocabulaire des résumés est nettement plus riche.

**Les dictionnaires reflètent les catégories du chercheur (méthode 1) ou une nomenclature externe (méthode 2).** Dans les deux cas, des thèmes réels du corpus peuvent être absents si leur vocabulaire n'a pas été anticipé. Cette limite est inhérente à toute méthode par dictionnaire et justifie de la présenter en complément, non en substitut, de la classification automatique (méthode 3).

**Les résultats des trois méthodes ne sont pas toujours comparables directement.** La méthode 1 produit des courbes de fréquence sur des termes-témoins spécifiques ; la méthode 2 produit des taux de présence de clusters ; la méthode 3 produit une partition des thèses. Comparer les trois sur un même thème (par exemple le tournant colonial) est informatif précisément parce que les méthodes ne mesurent pas la même chose.