

Quelles transitions professionnelles les demandeurs d'emploi effectuent-ils après une formation ?

Kevin M. Frick, Yagan Hazard, Damien Mayaux, Thomas Zuber

1^{er} mars 2024

Résumé

La formation professionnelle contribue-t-elle à la résorption des déséquilibres structurels du marché du travail ? Nous testons cette hypothèse en étudiant les trajectoires professionnelles et le retour à l'emploi faisant suite à une formation, à partir de données administratives françaises. En utilisant une nouvelle mesure d'écart de compétences entre métiers, nous observons que les retours à l'emploi suite à une formation se font dans des métiers plus éloignés en termes de compétences par rapport au dernier poste occupé. D'un point de vue purement réallocationnel, l'effet sur le retour à l'emploi de la formation professionnelle ne semble cependant pas être tiré par des redirections plus nombreuses vers des métiers en forte tension.

Codes JEL : J62 J68 J24

Mots-clés : déséquilibre structurel, formation, compétences

Affiliations :

Frick : Toulouse School of Economics. 1 Esplanade de l'Université, 31000 Toulouse, France. kevin-michael.frick@tse-fr.eu

Hazard : Collegio Carlo Alberto. Piazza Vincenzo Arbarello, 8, 10122 Torino TO, Italie. yagan.hazard@carloalberto.org

Mayaux : Paris School of Economics. 48 Boulevard Jourdan, 75014 Paris, France. mayaux.damien@psemail.eu

Zuber : Banque de France. 31 Rue Croix des Petits Champs, 75001 Paris, France. thomas.zuber@banque-france.fr

Les opinions exprimées dans cette étude ne reflètent ni les positions de la Banque de France ni celles de l'Eurosystème. Ces travaux ont bénéficié d'un accès aux moyens de calcul de l'IDRIS au travers de l'allocation de ressources 2022-AD011013484R1 attribuée par GENCI. Nous tenons à remercier Cyril Verluise pour ses conseils avisés lors de la genèse de ce projet, Pauline Givord et Anne Bucher pour leurs nombreuses remarques et leur soutien dans l'accès aux données, ainsi que l'ensemble des participants aux différentes réunions de co-pilotage de ce projet.

Introduction

Le marché du travail français souffre d'un taux de chômage plus élevé que la moyenne européenne conjugué à d'importantes difficultés de recrutement. En théorie, le marché du travail devrait se rééquilibrer si les personnes en recherche d'emploi se tournaient vers les métiers en tension. En pratique, de nombreux facteurs empêchent ces mobilités professionnelles d'avoir lieu, notamment la spécificité des compétences requises par chaque métier.

Les programmes de formation visent à apporter des éléments de réponse à ce double problème. D'abord en facilitant l'accès à la formation des personnes en recherche d'emploi, puis en orientant autant que possible l'offre de formation existante vers les métiers en tension pour lesquels les difficultés de recrutement sont les plus importantes. La combinaison de ces deux leviers devrait permettre d'exploiter à leur maximum les gains que l'on peut attendre de la formation professionnelle en termes de retour à l'emploi. En France, le Plan d'investissement dans les compétences (PIC) mis en place depuis 2018 s'inscrit dans cette logique.

Les effets directs de la formation sur le retour à l'emploi sont mieux connus que ceux, indirects, sur l'équilibre du marché du travail. De nombreuses études documentent l'effet de la formation sur le retour à l'emploi [Card et al. \[2018\]](#), notamment le cas du PIC en France [Chabaud et al. \[2022\]](#). En revanche, seules quelques rares études considèrent les effets indirects que la formation professionnelle peut avoir sur le marché du travail en contribuant à réduire les déséquilibres structurels d'offre et de demande de travail existant entre différents métiers [Şahin et al. \[2014\]](#), [Barnichon and Figura \[2015\]](#), [Marinescu and Rathelot \[2018\]](#). Une importante difficulté dans cette littérature consiste à quantifier à quel point un facteur donné, comme la mobilité géographique ou les écarts de compétences entre métiers, est responsable des déséquilibres entre métiers. L'enjeu est pourtant de taille : en 2021 en France, ces déséquilibres structurels pourraient être responsables d'environ 15% du niveau de chômage observé [[Fontaine and Rathelot, 2022](#)].

Dans cet article, nous étudions dans quelle mesure la formation professionnelle permet de réduire ces déséquilibres en permettant aux personnes en recherche d'emploi de se reconvertir vers des métiers en tension. En utilisant des données administratives françaises et une nouvelle mesure d'écart de compétences entre métiers, nous comparons les trajectoires professionnelles des personnes en recherche d'emploi selon si elles ont ou non suivi une formation. Nous observons que les retours à l'emploi suite à une formation longue se font dans des métiers plus éloignés par rapport au dernier poste occupé, lesquels sont également plus en tension. Cela suggère que les formations longues contribuent bien à résoudre les déséquilibres structurels du

marché du travail.

Afin d'évaluer l'impact de la formation professionnelle sur les transitions professionnelles des demandeurs d'emploi, nous comparons les trajectoires d'emploi des demandeurs d'emploi ayant recours à la formation professionnelle et des demandeurs d'emploi n'y ayant pas recours entre 2018 et 2020.¹ Cette comparaison est réalisée grâce à une méthode machine learning avec orthogonalisation en deux étapes [Chernozhukov et al., 2018].² Ce travail est rendu possible grâce à l'utilisation de la base ForCE permettant chaque trimestre de rassembler des informations sur le recours à la formation et le retour à l'emploi de l'ensemble des demandeurs d'emploi ayant été inscrits à Pôle emploi depuis moins de 10 ans. À bien des égards, cette étude prolonge donc celle de Chabaud et al. [2022] à laquelle nous renvoyons pour une description plus détaillée de la base ForCE. Contrairement à Chabaud et al. [2022] qui étudient l'effet du recours à la formation sur le retour à l'emploi, cette étude se concentre sur la question plus spécifique de l'impact de la formation sur les transitions professionnelles des demandeurs d'emploi et certaines caractéristiques de l'emploi retrouvé. Alors que l'étude de Chabaud et al. [2022] prend pour population de référence l'ensemble des inscrits à Pôle emploi, notre méthodologie nous impose de nous restreindre à ceux ayant connu un épisode d'emploi stable au cours des 12 derniers mois. Cette sous-population est plus jeune, mieux formée, et plus proche de l'emploi, ce qui limite la portée des résultats obtenus.

Afin de pouvoir mieux décrire la nature des trajectoires professionnelles des demandeurs d'emplois, une importante composante de cette évaluation a consisté à construire une mesure fiable et continue des écarts de compétences entre métiers à partir des données textuelles contenues dans les vacances d'emploi postées à Pôle emploi. Nous entraînons un réseau de neurones sur des tâches prétextes liées aux compétences des différents métiers de la classification du Répertoire Opérationnel des Métiers et des Emplois (ROME V3). Cet entraînement nous permet ensuite de placer chaque métier dans un espace de compétences de dimension réduite et de définir une notion de distance inter-métiers comme la distance angulaire entre deux métiers donnés. Ces compétences qui forment les dimensions de l'espace ne sont pas interprétables individuellement, mais il est pertinent de comparer le vecteur des compétences pour deux métiers donnés. Grâce à cette mesure de distance entre métiers, nous sommes capables de distinguer les transitions professionnelles que l'on pourrait qualifier de locales qui relient entre eux des métiers dont les compétences sont relativement proches, et des transitions professionnelles plus

1. Plus précisément, nous comparons des demandeurs d'emploi ayant recours *pour la première fois* à la formation professionnelle à des demandeurs d'emploi n'y ayant jamais eu recours à un instant t – ce groupe de contrôle contenant donc des individus qui auront recours à la formation professionnelle lors d'une période ultérieure.

2. Nous vérifions que l'utilisation de la méthode des appariements sur score de propension [Rosenbaum and Rubin, 1983] donne des résultats en tout point comparables.

lointaines dirigées vers des métiers dont les compétences s'écartent beaucoup plus nettement des compétences utilisées dans le métier d'origine du demandeur d'emploi. [Dawson et al. \[2021\]](#) utilisent une méthodologie similaire à la nôtre.³

Nous montrons que le recours à la formation professionnelle permet aux demandeurs d'emploi de parcourir une plus grande distance entre les compétences mobilisées dans leur métier de départ et celles requises par leur métier d'arrivée. D'un point de vue purement réallocatif, l'effet sur le retour à l'emploi de la formation professionnelle ne semble cependant pas être tiré par des redirections plus nombreuses vers des métiers en forte tension. Ce résultat laisse présumer qu'un ciblage plus systématique de l'offre de formation professionnelle sur les compétences demandées par les métiers en tension permettrait d'accroître l'effet réallocatif de la formation professionnelle.

Nous contribuons également à une abondante littérature qui montre que les transitions entre métiers sont associées à un fort coût en capital humain spécifique [[Becker, 1964](#)]. Dans leur étude pionnière [Jacobson et al. \[1993\]](#) montraient déjà que les pertes de revenus liées à un épisode de chômage étaient plus importantes pour les personnes décidant de manière plus ou moins contrainte de changer de secteur. [Shaw \[1984\]](#) est le premier à mettre en évidence le rôle spécifique joué par la trajectoire professionnelle accomplie à travers différents métiers plutôt que du secteur d'emploi dans la détermination du revenu d'un travailleur. Dans le contexte du système d'apprentissage allemand, [Eckardt \[2022\]](#) montre que les personnes travaillant dans un domaine pour lequel elles n'ont pas été spécifiquement formées accusent une perte de revenu par rapport à leur collègues bénéficiant d'une formation adéquate. Ces coûts sont positivement corrélés à une mesure de distance en compétences inter-métiers. Les coûts liés aux transitions professionnelles forcées ont par ailleurs été étudiés dans le cadre de l'évaluation des effets de la concurrence internationale sur les marchés du travail locaux. [Traiberman \[2019\]](#) pour le Danemark et [Mestieri et al. \[2020\]](#) pour la France montrent que les changements de métiers liés à une forte concurrence internationale aboutissent à des pertes de revenus. Ces pertes sont d'autant plus importantes que le nouveau métier du travailleur est éloigné de son métier d'origine. Toujours dans le cadre de l'étude des conséquences des chocs de commerce international [Hyman \[2018\]](#) montre que les politiques de formation permettent aux travailleurs touchés de réaliser des transitions professionnelles vers des métiers moins exposés à la concurrence internationale. Enfin, reprenant l'intuition de [Shaw \[1984\]](#), une littérature abondante montre que la prise en compte du caractère multidimensionnel des compétences (par opposition à un simple index linéaire) est cruciale pour la compréhension des transitions sur le marché du travail, du processus

3. Voir également [Bana et al. \[2021\]](#) et [Gentzkow et al. \[2019\]](#).

d'appariement entre travailleurs et entreprises et de la détermination du niveau de rémunération des salariés [Gathmann and Schönberg, 2010, Lindenlaub and Postel-Vinay, 2021, Guvenen et al., 2020, Baley et al., 2020].

Le reste de l'article est structuré de la manière suivante. Dans une première partie nous présentons les principes que nous avons suivis et les résultats obtenus dans la construction d'une mesure de distance "inter-métiers" à partir des données textuelles contenues dans les offres d'emploi postées à Pôle emploi. Dans une deuxième partie nous présentons la base ForCE et les choix méthodologiques que nous avons adoptés dans la construction de notre échantillon d'étude et des différentes variables d'intérêt. Enfin dans une dernière partie nous présentons les résultats issus d'une comparaison des trajectoires professionnelles réalisées par les demandeurs d'emploi ayant recours à la formation et celles des inscrits à Pôle emploi n'y ayant pas eu recours.

1 Une nouvelle mesure de distance inter-métiers

Notre analyse de l'effet de la formation sur le retour à l'emploi et les mobilités professionnelles repose sur une nouvelle mesure de distance inter-métiers, qui traduit les obstacles liés aux différences de compétences entre métiers rencontrés lors d'une reconversion professionnelle.

Il existe déjà de nombreuses sources de données de qualité reliant métiers et compétences professionnelles. Aux États-Unis, le système O*NET fournit un répertoire des compétences détaillé, et pour chaque métier de la nomenclature américaine (la "Standard Occupational Classification"), un indicateur de la fréquence d'utilisation et du niveau de maîtrise attendu dans chacune des compétences. Il contient aussi une table des métiers connexes (la "Related Occupation Matrix") qui donne, pour chaque métier de départ, une liste ordonnée de dix métiers vers lesquels il est le plus facile de se reconvertir eu égard aux compétences attendues. En France, la classification du Répertoire Opérationnel des Métiers et des Emplois (ROME V3) joue un rôle analogue : en plus de la définition de 532 métiers et 14 secteurs d'activité, elle fournit un registre de compétences qu'elle fait correspondre à la nomenclature des métiers et suggère une liste de métiers de reconversion possibles en partant d'un métier donné (rubrique "Mobilités" du ROME). Le ROME permet ainsi d'approcher la notion de distance en compétences inter-métiers soit en étudiant les intersections existantes entre les compétences recensées pour chaque métier, soit en s'appuyant sur l'ensemble des trajectoires professionnelles suggérées par les transitions vers des métiers "proches" recensés dans la rubrique "Mobilités".

Toutefois, l'usage du ROME V3 pour étudier l'effet de la formation professionnelle présente

plusieurs limites. Tout d’abord, la liste des mobilités professionnelles suggérées ne vise pas à l’exhaustivité, ce qui fait qu’elle ne mentionne pas certaines transitions tout à fait possibles et pertinentes et ne dit rien des transitions difficiles à réaliser. De plus, elle classe ces suggestions de mobilité professionnelle en deux catégories seulement, alors qu’on souhaiterait plutôt mesurer de manière continue la difficulté à se reconvertir d’un métier à un autre, en particulier pour des métiers assez “lointains” en termes de compétences. Enfin, les compétences identifiées dans le ROME sont souvent spécifiques à un ou deux métiers seulement, ce qui rend impossible la distinction entre des paires de métiers assez différents et des paires de métiers complètement différents. Ces trois points sont autant d’obstacles à l’étude quantitative de l’effet de la formation sur les mobilités professionnelles, puisqu’on peut supposer que cette dernière permet justement des mobilités plus “lointaines” en termes de compétences.

Dans cette partie, nous expliquons comment nous avons construit une mesure de distance inter-métiers en entraînant un réseau de neurones à partir de textes d’offres d’emploi. Nous verrons que le recours à un large corpus de données textuelles nous a permis de produire une mesure continue et capable de formuler des prédictions pertinentes même pour des paires de métiers plus éloignés. Après avoir détaillé les données utilisées, nous rentrerons dans le détail de la méthodologie de construction de la mesure, puis nous expliquerons comment nous avons validé les résultats obtenus.

1.1 Données

Nous avons utilisé différentes sources de données pour construire puis valider notre distance.

Le principal jeu de données utilisé est le texte de plus de 4 millions d’offres postées sur le site de Pôle Emploi entre décembre 2018 et octobre 2020 ainsi que les codes métier ROME associés⁴. Nous utilisons ces données pour pré-entraîner notre modèle de langage sur des textes d’offres d’emploi, puis pour entraîner notre réseau de neurones à prédire le code métier à partir du texte de l’offre. Nous avons rééquilibré le nombre d’offres de chaque métier par rapport à l’échantillon initial afin de ne pas introduire de biais lors de l’entraînement.

Un de nos objectifs était de proposer une alternative aux métriques basées sur le ROME V3. Nous avons donc utilisé deux champs du ROME afin de produire des mesures de distance inter-métiers alternatives pour tester notre propre mesure :

- Le champ ‘Mobilités’ associe à des couples de métiers la valeur 1 si une mobilité inter-métiers semble possible sans formation et 2 si une mobilité inter-métier semble possible

4. Base [Jocas](#) d’offres d’emplois collectées par la Dares.

avec une formation minimale des compétences sous-jacentes. Nous utilisons notamment ce champ pour produire une distance inter-métiers alternative, appelée "Graph distance", qui est la distance sur le graphe orienté dont les sommets sont les codes ROME, dont la présence d'une arête indique qu'une mobilité est suggérée dans le champ 'Mobilités' et dont la pondération de chaque arête correspond au niveau 1 ou 2 défini précédemment. Nous avons également utilisé le champ Mobilité dans la construction de notre mesure de distance inter-métiers.

- Le champ "Compétences" associe une liste d'items de compétences, générales ou spécifiques, à chacun des 532 métiers de la classification ROME V3. Nous utilisons ce champ pour produire une distance inter-métiers alternative, appelée "Skill distance", qui est le cosinus de l'angle entre les vecteurs représentant les métiers dans l'espace, dont la composante i vaut 1 si la compétence i est associée au métier et 0 sinon. Cela correspond à l'analogie de notre mesure de distance, mais à partir d'une représentation dans un espace de dimension égale au nombre total de compétences listées dans ROME et qui repose sur énormément de "dire d'expert".

En plus de ces champs issus de la classification du ROME nous avons intégré à l'entraînement du réseau de neurones des informations sur le marché du travail français. Pour faire en sorte que notre mesure de distance inter-métiers reflète uniquement des écarts en matière de compétences requises, et non d'autres facteurs affectant les comportements des agents sur le marché du travail comme les stéréotypes de genre associés à des métiers, le prestige social de certaines professions ou encore les différentiels de tension., nous nous sommes limités à des données qui nous semblaient clairement interprétables en termes de compétences professionnelles, en particulier le niveau moyen d'éducation des travailleurs dans une occupation donnée, ainsi qu'une liste de couples de métiers de départ et d'arrivée pour lesquels nous avons observé des transitions fréquentes au sein d'une même entreprise, associées à des hausses de salaire, que nous avons interprété comme des mobilités professionnelles "verticales". Pour cela, nous nous sommes appuyés sur les Déclarations Annuelles de Données Sociales (DADS) sur l'année 2019.

1.2 *Le réseau de neurones*

La construction de la distance inter-métiers se fait en trois temps : nous nous appuyons tout d'abord sur un modèle de langage pour extraire une représentation sémantique riche du contenu des offres d'emploi, puis sur un réseau de neurones pour ne retenir de cette représentation sémantique que le contenu pertinent par rapport aux compétences et positionner les métiers dans

un espace de grande dimension, et enfin nous utilisons la distance angulaire dans cet espace comme mesure de distance inter-métiers.

Notre contribution méthodologique principale est la construction d'une représentation spatiale en grande dimension des compétences mobilisées et exigées par les 532 métiers de la classification ROME V3 qui soit pertinente quantitativement pour analyser les mobilités professionnelles tout en restant interprétable. La mesure de distance inter-métiers utilisée dans la suite de cette note est un sous-produit immédiat de cette construction. Le modèle de langage utilisé pour analyser le texte des offres d'emploi, FlauBERT (Le et al. [2019]), est une référence pour la langue française et nous nous en servons d'une manière tout à fait standard dans la littérature en traitement automatique du langage.

Le choix d'utiliser un réseau de neurones pour obtenir notre représentation nous distingue de la littérature existante sur la multi-dimensionalité des compétences. Celle-ci repose majoritairement sur des méthodes de construction d'index ou de réduction de dimension (analyse en composante principale, analyse de correspondance), qui sont simples à définir mathématiquement mais produisent une représentation des compétences dont la pertinence par rapport à la finalité souhaitée n'a rien d'évident – en particulier pour un usage dans des analyses quantitatives. Au contraire, nous avons fait le choix d'une approche supervisée dans laquelle les tâches prétextes permettant l'entraînement de la représentation sont facilement explicables, en lien étroit avec la finalité de la représentation et imposent une cohérence géométrique au résultat. Ce parti-pris permet une meilleure interprétabilité de la représentation obtenue, au prix d'une complexité accrue dans sa construction.

Le fait que la construction de la représentation spatiale des compétences s'appuie sur des données textuelles fait courir certains risques. Il est possible que les textes des offres d'emplois contiennent des éléments sans lien avec les compétences ayant un faible pouvoir prédictif sur le métier, comme le style de rédaction de l'offre, le nom de l'entreprise ou le lieu d'exercice de l'activité, qui pourraient affecter la représentation. On peut également craindre que, selon les métiers, les offres aient plus ou moins tendance à demander plus de compétences dans l'offre que le métier n'en exige vraiment. La taille du corpus d'offre d'emplois et la faible profondeur du réseau de neurones permettent cependant d'espérer que ce dernier exploitera surtout les éléments du texte ayant un fort pouvoir prédictif sur le métier et parviendra à faire abstraction du bruit dans la description des compétences.

L'utilisation des textes des offres d'emploi présente également deux avantages majeurs sur le long-terme. D'une part, elle montre qu'il est possible de s'affranchir au moins partiellement

du "dire d'expert" pour associer des métiers et des compétences, ce qui pourrait permettre de suivre l'évolution des besoins en compétence des métiers à des échelles de temps beaucoup plus courtes que le rythme actuel de renouvellement des registres de compétences comme celui du ROME. D'autre part, elle permet de rendre la représentation obtenue plus "explicable", dans la mesure où l'on peut déterminer la position dans cet espace de n'importe quelle offre d'emploi. Nous n'avons pas pleinement exploité cette possibilité d'explorer l'espace des compétences en générant des variations dans nos offres d'emplois, mais cela pourrait constituer une méthode de validation supplémentaire si notre représentation spatiale des métiers devait être réutilisée. Enfin, on peut noter qu'un certain nombre d'articles récents utilisent également le traitement automatique du langage sur un large corpus de textes pour produire une représentation spatiale d'un ensemble de produits ou de marques dans un espace de faible dimension (Wang et al. [2021], Ringel [2022]).

1.2.1 Objectifs

Nous utilisons un réseau de neurones afin de placer les 532 métiers de la classification ROME dans un espace de dimension 20 de telle sorte que le positionnement relatif de ces métiers nous renseigne sur les compétences et la possibilité d'effectuer telle ou telle transition professionnelle. Les dimensions de cet espace n'ont pas vocation à être considérées individuellement, comme si elles mesuraient l'importance d'une compétence spécifique (ou même facilement interprétable) pour un métier donné. En revanche, le vecteur formé de ces 20 dimensions contient une représentation synthétique des compétences requises par le métier et peut quant à lui être interprété. Une fois les métiers positionnés dans l'espace, nous utilisons la distance cosinus entre vecteurs comme mesure de la distance inter-métiers en termes de compétences.

Formellement, on appellera représentation R la fonction qui à chaque métier de la classification ROME associe le vecteur correspondant dans cet espace.

$$\begin{aligned}
 R : \{\text{Codes ROME}\} &\longrightarrow \mathbb{R}^{20} \\
 x &\longmapsto R(x)
 \end{aligned}$$

La représentation est construite de sorte que la géométrie dans cet espace ait une interprétation claire en matière de compétences et de mobilité professionnelle.

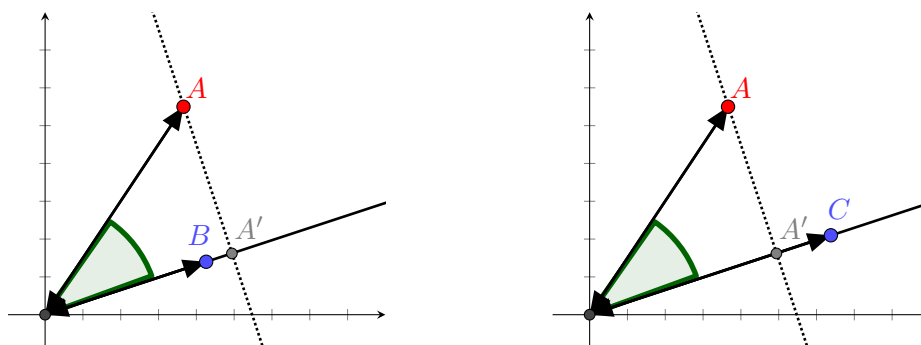
- L'angle entre les vecteurs $R(x)$ et $R(y)$ doit refléter à quel point les compétences mobilisées dans les métiers x et y sont similaires. Le cosinus de cet angle est notre mesure de distance inter-métiers.

- La norme $\|R(x)\|$ doit refléter le niveau de maîtrise attendu dans les compétences mobilisées dans le métier x .
- La projection de u sur la droite engendrée par $R(x)$ doit refléter le niveau de maîtrise dans les compétences mobilisées par le métier x d'une personne dont les compétences détenues sont représentées par le vecteur u .
- Toutes les composantes du vecteur $R(x)$ sont positives et représentent une certaine dimension des compétences mobilisées par le métier x .

La distinction entre "quelles sont les compétences mobilisées" (la direction du vecteur) et "quel est le niveau de maîtrise dans les compétences mobilisées" (la norme du vecteur) s'inspire des questions posées dans plusieurs jeux de données existants reliant les compétences et les métiers, comme ceux de la classification O*NET aux Etats-Unis. En pratique, nous basons notre distance entre métiers sur la direction des vecteurs dans la représentation des métiers, mais leur norme joue également un rôle dans le processus de construction de cette représentation.

Le graphique 1 illustre l'interprétation géométrique de notre représentation. Une conséquence de la troisième propriété, qui porte sur le transfert de compétences lors d'une transition professionnelle, est que la dimension de l'espace majore le cardinal du plus grand ensemble de métiers dont les compétences mobilisées n'ont deux à deux absolument rien en commun. Autrement dit, en passant de l'un de ces métiers à un autre, il faut absolument tout réapprendre de zéro. Nous avons choisi de travailler en dimension 20, ce qui impose à la représentation de capturer des compétences partagées entre les 532 métiers de la classification ROME, tout en laissant plus de flexibilité pour représenter ces compétences que dans un modèle de très faible dimension.

Graphique 1 – Interprétation géométrique de la représentation R des métiers



Notes : Représentation du passage d'un métier de départ A à un métier d'arrivée B dans l'espace des compétences. Notre mesure de distance inter-métiers, qui traduit l'écart entre les compétences mobilisées par les métiers A et B , correspond au cosinus de l'angle vert. Le métier A nécessite un niveau de maîtrise des compétences qu'il mobilise plus élevé que le métier B , d'où une plus grande distance à l'origine. Si un individu dont les compétences sont exactement celles correspondant au métier A venait à se reconverter dans le métier B , son niveau de maîtrise des compétences mobilisées par B serait la distance à l'origine du point A' . Ici, le niveau de maîtrise est suffisant pour envisager une mobilité de A à B . A l'inverse, il n'est pas possible pour ce même individu de se reconverter dans le métier C , qui mobilise les mêmes compétences que le métier B mais nécessite un niveau de maîtrise plus élevé.

Cette traduction géométrique de la faisabilité des transitions professionnelles dans l'espace des compétences n'est pas utilisée directement dans la construction de la mesure. Cependant, elle illustre le mécanisme de la réutilisation partielle des compétences déjà acquises lors d'une mobilité professionnelle. Plus on s'éloigne du métier initial en termes de compétences mobilisées, c'est-à-dire plus l'angle est grand, plus la norme du projeté est petite, c'est-à-dire moins on peut réutiliser ses compétences déjà acquises. Ainsi, pour un niveau de maîtrise des compétences donné dans le métier d'arrivée, il est plus faisable de se reconvertir dans un métier avec lequel on a une faible distance angulaire. Cela explique l'intérêt d'utiliser la distance angulaire comme mesure de la difficulté à se reconvertir.

En plus de cette représentation des métiers, notre réseau de neurones permet également d'associer à tout texte d'une offre d'emploi une position dans ce même espace de dimension 20. On notera F la fonction associée.

$$\begin{array}{rcl}
 F : \{\text{Textes d'offres d'emploi}\} & \longrightarrow & \mathbb{R}^{20} \\
 x & \longmapsto & F(x)
 \end{array}$$

Cette représentation des textes d'offre d'emploi a deux intérêts.

- Elle est utilisée pour apprendre la représentation R des métiers. La principale tâche d'entraînement du réseau de neurones consiste à comparer le texte t d'une offre d'emploi à un métier x , via leurs représentants $F(t)$ et $R(x)$, afin de prédire le code ROME du métier de l'offre d'emploi.
- Une fois l'entraînement terminé, elle peut être utilisée pour analyser qualitativement la représentation obtenue. On peut modifier le texte t d'une offre d'emploi en entrée et observer comment cela affecte son représentant $F(t)$.

1.2.2 L'architecture du réseau de neurones

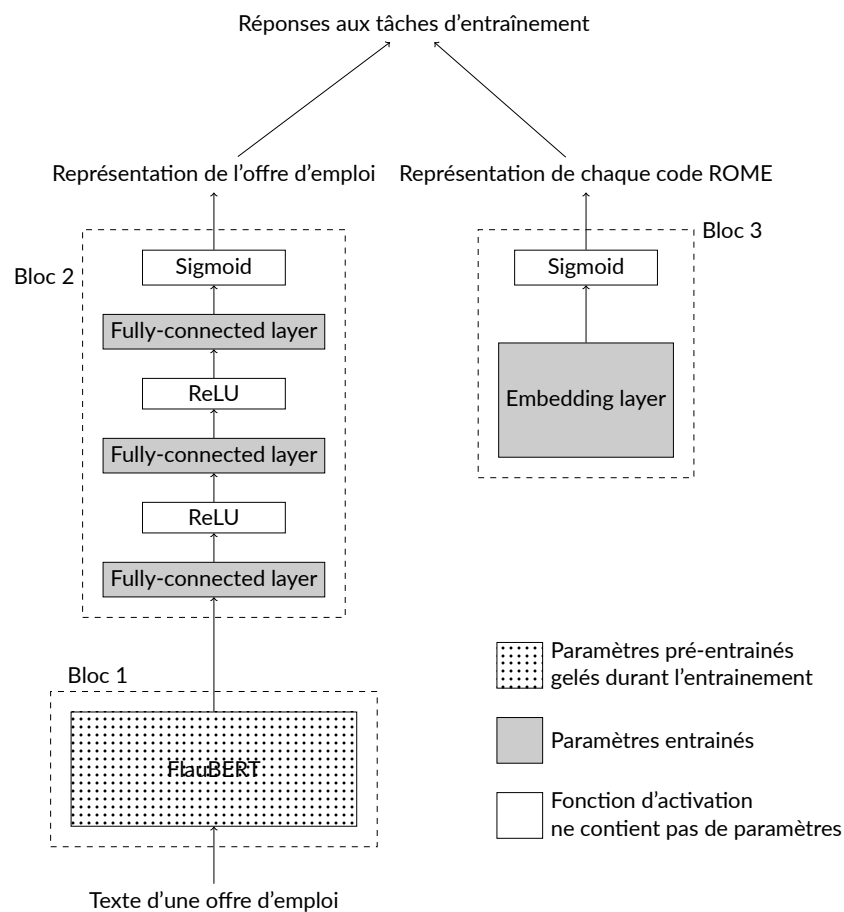
Le réseau de neurones est composé de trois blocs, comme illustré sur le graphique [2](#)

- Le bloc 1 est un modèle de langage. Il prend en entrée un texte et produit une représentation de ce texte en 768 dimensions qui permet de répondre à une large classe de problèmes sémantiques. Nous utilisons une version pré-entraînée de ce modèle et ses paramètres sont figés pendant l'entraînement des autres blocs.
- Le bloc 2 est un réseau de neurones à trois couches. Il prend en entrée la représentation d'une offre d'emploi en sortie du bloc 1 et produit une représentation de cette offre en 20 dimensions. Cette étape permet de ne conserver des 768 dimensions initiales que l'infor-

mation pertinente en terme de compétences et de mobilité professionnelle. Associé avec le bloc 1, il forme la fonction F qui associe au texte d'une offre d'emploi sa représentation en dimension 20.

- Le bloc 3 est une couche supplémentaire qui stocke la représentation R des 532 métiers de la classification ROME en 20 dimensions. La comparaison de la sortie du bloc 2 avec la sortie du bloc 3 permet d'évaluer la performance du réseau aux différentes tâches et d'entraîner ses paramètres en conséquent.

Graphique 2 – Schéma du réseau de neurones utilisé



Les paramètres du bloc 1 (fond en pointillé sur le schéma) sont récupérés sur un modèle déjà entraîné (FlauBERT [Le et al. \[2019\]](#)), puis légèrement ré-entraînés de manière semi-supervisée sur notre corpus d'offres d'emploi séparément du reste du modèle, avant d'être gelés durant l'entraînement des blocs 2 et 3. Les paramètres des blocs 2 et 3 (couches en gris foncé sur le schéma) sont initialisés aléatoirement, puis entraînés grâce au texte des offres d'emplois tel que détaillé dans la sous-section suivante.

1.2.3 Tâches d'entraînement et pénalités associées

La représentation R des métiers fait partie des paramètres du réseau de neurones, qui évoluent à mesure que le réseau de neurones s'entraîne à réaliser certaines tâches. C'est le choix de tâches d'entraînement pertinentes et de pénalités associées qui permet d'imposer à la représentation les propriétés géométriques énoncées précédemment.

Nous utilisons les couples tâche/pénalité suivants pour l'entraînement du réseau de neurones :

– Prédire le code ROME d'une offre d'emploi - WARP loss

Nous avons adapté la pénalité (Weighted Approximate-Rank Pairwise, WARP) de l'article [Weston et al. \[2011\]](#) à notre problème.

Pour chaque texte t d'une offre d'emploi, nous calculons l'angle de son représentant $F(t)$ avec le représentant $R(x)$ de chaque métier x de la classification ROME. On s'attend à ce que l'angle entre $F(t)$ et le représentant $R(x_0)$ du métier correspondant à l'offre soit très faible. On pénalise donc le réseau pour chaque métier x telle que l'angle entre $F(t)$ et $R(x)$ est inférieur à celui entre $F(t)$ et $R(x_0)$.

$$l_{\text{WARP}} = \sum_{\text{ROME codes } x \neq x_0} |F(t) \cdot (R(x) - R(x_0))|_+$$

– Prédire les mobilités suggérées dans la classification ROME - Triplet loss

La triplet loss vient à l'origine de la reconnaissance d'image. Si l'on a trois photos de visage et que les deux premières viennent d'une même personne, alors un algorithme de reconnaissance faciale devrait prédire que la première ressemble plus à la deuxième qu'à la troisième.

De la même manière, nous formons des triplets de métiers en prenant un métier de départ x_0 , un métier x vers lequel une transition depuis le métier de départ est suggérée dans le champ "Mobilités" de la classification ROME, et un autre y vers lequel aucune transition n'est suggérée. Nous prenons ensuite une offre au hasard dans chacun de ces métiers.

On s'attend à ce que la mobilité de x_0 à x soit un meilleur choix que celle de x_0 vers y . En particulier, la transition doit être faisable en terme de compétences sans que le niveau de maîtrise attendu n'ait trop diminué par rapport à l'emploi initial. Cela conduit à utiliser la métrique suivante pour dire à quel point une mobilité est susceptible d'être recommandée.

$$d(x_0, x) = \max \left(1 - \frac{\|R(x)\|}{\|R(x_0)\|}, 1 - \frac{R(x_0) \cdot R(x)}{\|R(x)\|^2} \right)$$

On définit alors la triplet loss comme l'écart, par rapport à cette métrique, entre les couples

(x_0, x) et (x_0, y) .

$$l_{\text{Triplet}} = |d(x_0, x) - d(x_0, y)|_+$$

- **Prédire le niveau de maîtrise attendu dans les compétences mobilisées par le métier x - Norm loss**

Pour chaque métier x , on compare la norme de son représentant $R(x)$ à une valeur e_x (normalisée entre 0 and 1) qui reflète le niveau de qualification moyen de travailleurs occupants ce métier.

$$l_{\text{Norm}} = \sum_{\text{ROME codes } x} (\|R(x)\| - e_x)^2$$

- **Prédire le niveau de maîtrise - Vertical loss**

Dans de nombreux secteurs, les mobilités verticales nécessitent d'avoir acquis les compétences des échelons inférieurs. On s'attend donc à ce que, parmi les couples de métiers de départ et d'arrivée souvent associés dans les données du marché du travail à des mobilités verticales, le premier nécessite moins de compétences que le second, dans toutes les dimensions que recouvre la notion de compétences.

En notant $x \prec y$ le fait que beaucoup de mobilités verticales impliquent de passer du métier x au métier y , on définit la vertical loss par

$$l_{\text{Vertical}} = \sum_{x \prec y} \sum_{i=1}^{20} |R(x)_i - R(y)_i|_+$$

Les différentes pénalités associées à ces tâches sont normalisées, passées au logarithme puis sommées pour former la pénalité totale l , soit :

$$l = \log(1 + l_{\text{WARP}}) + \log(1 + l_{\text{Triplet}}) + \log(1 + l_{\text{Norm}}) + \log(1 + l_{\text{Vertical}})$$

1.3 Résultats

Nous avons validé qualitativement et quantitativement la mesure de distance obtenue à partir de la représentations spatiale des métiers.

1.3.1 Performance sur les tâches prétextes

Durant l'entraînement du réseau de neurones, nous surveillons l'évolution de la fonction de pénalité générale définie précédemment, ainsi que la capacité du réseau de neurones à prédire correctement le code ROME d'une offre d'emploi donné (sa "précision").

Ces deux métriques sont calculées sur l'ensemble d'entraînement et de test, et l'entraînement est arrêté pour la valeur des paramètres qui donne la plus haute précision sur le jeu de données d'entraînement – environ 80% dans notre cas. Nous avons observé que le fait de ré-entraîner le modèle de langage de manière non-supervisée sur notre corpus d'offre avant de réaliser l'entraînement de notre propre réseau de neurones augmente significativement, de l'ordre de 20%, la précision que nous parvenons à atteindre. Cela suggère que la disponibilité de larges corpus d'offre d'emplois comme la base JOCAS peut s'avérer précieuse pour l'entraînement de modèles, même si le code ROME correspondant aux offres d'emplois est inconnu (il est aujourd'hui imputé par des techniques d'apprentissage statistique).

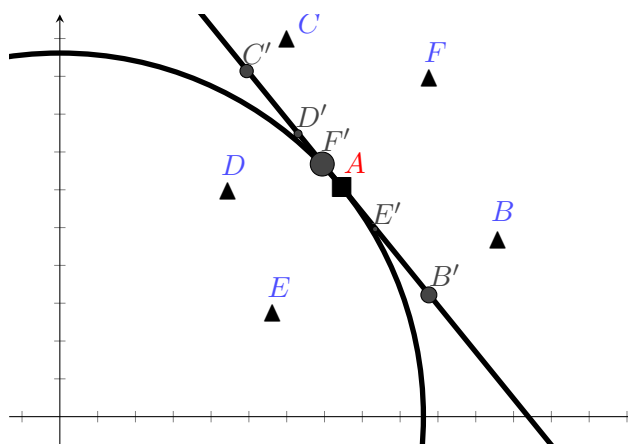
Nous avons observé une baisse de la pénalité générale durant l'entraînement. La décomposition selon chaque pénalité montre que, malgré la normalisation, c'est la tâche de prédiction du code ROME à partir d'une offre d'emploi qui joue le rôle le plus important dans l'entraînement. En effet, les autres pénalités décroissent assez rapidement, sans doute car les contraintes qu'elles imposent sont plus simples à satisfaire dans la géométrie de l'espace.

1.3.2 Validation qualitative de la mesure de distance inter-métiers

Pour l'analyse qualitative, nous nous sommes concentrés sur le métier le plus représenté parmi les demandeurs d'emploi dans chacun des 14 secteurs d'activité définis par la classification ROME. Pour chacun de ces 14 métiers "types", nous affichons en Annexe [A.1](#) les 5 métiers les plus proches selon notre mesure, ainsi qu'en Annexe [A.2](#) ceux parmi eux vers lesquels aucune mobilité n'est suggérée dans la classification ROME. Cette deuxième liste permet d'illustrer la capacité du réseau de neurones à prédire des transitions professionnelles plausibles en dehors de celles qui lui ont été présentées durant l'entraînement. Les résultats suggèrent que, non seulement notre mesure de distance retrouve les mobilités suggérées par la classification ROME V3, mais également qu'elle parvient souvent à faire d'autres suggestions qui semblent qualitativement cohérentes. Cela conforte notre intuition initiale selon laquelle les suggestions du champ "Mobilités" de la classification ROME sont relativement étroites et ne permettent pas d'explorer l'ensemble des mobilités professionnelles pertinentes. On notera toutefois que notre mesure de distance performe moins bien pour les métiers moins représentés dans notre corpus d'offre, comme "Musique et chant" (L1202).

Nous avons également développé une méthodologie de visualisation de la représentation spatiale des métiers qui permet d'illustrer la norme et le positionnement relatif des métiers au voisinage d'un métier de référence. Pour visualiser un espace de dimension 20, on utilise généra-

Graphique 3 – La méthode de visualisation de la représentation utilisée en Annexe A.3



Notes : Pour visualiser dans l'espace des compétences les plus proches voisins B , C , D , E et F d'un métier A donné, pour commencer (1) on les projette sur l'hyperplan tangent à la sphère de centre l'origine et de rayon la norme de A pour former B' , C' , D' , E' et F' , puis (2) on applique une analyse en composante principale pour réduire la dimension, et enfin (3) on grossit le point correspondant à chaque métier en fonction de sa norme initiale.

lement des méthodes de réduction de dimension comme l'analyse en composante principale. Cependant, la géométrie particulière de la représentation contient beaucoup d'information, que nous risquerions de perdre si l'on projetait directement un nuage de points de la dimension 20 à la dimension 2. Pour résoudre ce problème, nous choisissons un métier de référence dont nous souhaitons représenter le voisinage, puis (1) nous projetons les métiers proches sur l'hyperplan tangent à la sphère de centre l'origine passant par le point de référence, suite à quoi (2) nous appliquons une analyse en composante principale sur la projection obtenue. Pour finir, nous réintégrons dans le graphique l'information sur la norme "perdue" lors de la projection sur l'hyperplan tangent en modifiant la taille des points selon la norme du vecteur correspondant.

Le graphique 3 illustre la procédure utilisée pour construire la visualisation. Le carré A en rouge représente le métier de référence et les triangles bleus B , C , D , E et F ses plus proches voisins. Les points bleus sont projetés sur l'hyperplan tangent en A pour former les cercles gris B' , C' , D' , E' et F' (1). On voit que la taille des cercles gris varie selon la norme du vecteur initial. La dimension du dessin ne permet pas de représenter l'effet de l'analyse en composante principale (2), mais on peut retenir qu'elle est appliquée aux points gris et on choisit de la centrer sur le point A .

L'Annexe A.3 illustre par cette méthode le cas d'un des métiers types évoqués précédemment et de ses plus proches voisins. Elle permet d'explorer qualitativement la représentation spatiale, notamment en vérifiant que les normes relatives des métiers semblent cohérentes et que les directions dans lesquels dévient les voisins d'un métier donné peuvent donner lieu à une interprétation. Ce mode de représentation pourrait à l'avenir être exploité pour observer comment

se déplace le représentant $F(t)$ d'une offre d'emploi t quand on change légèrement son texte.

1.3.3 Validation quantitative de la mesure de distance inter-métiers

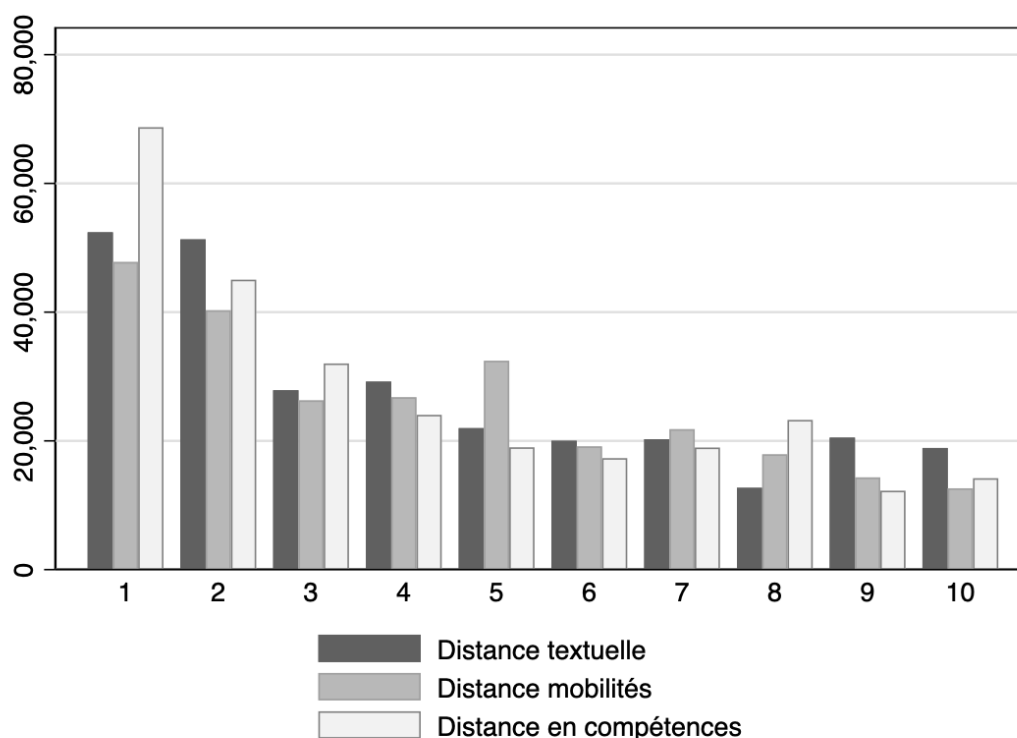
Nous avons cherché à valider quantitativement notre mesure de distance inter-métiers en utilisant d'autres métriques que celles liées aux tâches prétextes de l'entraînement.

Nous avons choisi d'étudier la capacité de notre mesure à prédire les transitions professionnelles effectivement réalisées sur le marché du travail. Prédire parfaitement les transitions observées n'est pas la vocation de la mesure, car ces dernières peuvent dépendre de nombreux facteurs qui n'ont rien à voir avec les compétences professionnelles tels que les aspirations personnelles, les stéréotypes de genre ou encore le contexte local du marché du travail. Cependant, il semble raisonnable de supposer qu'il y ait en général plus de transition professionnelles entre des métiers plus proches en termes de compétences mobilisées.

Nous avons comparé le pouvoir prédictif de notre mesure de distance inter-métiers par rapport à deux alternatives : d'une part, la distance sur le graphe des mobilités professionnelles suggérées dans la classification ROME, d'autre part, la distance cosinus entre les vecteurs de compétences associés au métier de départ et d'arrivée dans la classification ROME.

Le graphique 4 représente, pour chacune des trois mesures de distance inter-métiers et pour chaque rang k de 1 à 10, le nombre de transitions professionnelles observées qui vont d'un métier de départ vers le k -ième métier le plus proche selon la mesure. Cela nous permet d'évaluer la capacité à retrouver les transitions les plus probables et d'observer la qualité de la mesure pour des métiers plus éloignés. On observe que notre mesure de distance (appelée "Distance textuelle" dans le graphique) fait mieux que la distance construite à partir des suggestions de la classification ROME ("Distance mobilités") pour 7 des 10 rangs. D'ailleurs, si l'on cumule les transitions expliquées par les k métiers d'arrivée les plus proches, notre mesure la surpasse pour tout k entre 1 et 10. Concernant la mesure basée sur les compétences définies dans la classification ROME V3 ("Distance en compétences"), on remarque qu'elle a un très bon pouvoir prédictif pour les métiers très proches, mais un faible pouvoir prédictif pour les métiers plus lointains. Cela s'explique simplement par le fait que les notions de compétences utilisées dans la classification ROME V3 sont assez étroites et manquent de pouvoir de généralisation (ce constat a d'ailleurs orienté les travaux sur la refonte de la classification ROME vers sa version 4.0). A l'inverse, notre mesure construite à partir de données textuelles diverses et non structurées semble plus à même d'identifier des notions large de compétence et de généraliser puisque c'est elle qui performe le mieux pour des rangs élevés.

Graphique 4 – Nombre de transitions expliquées par les mesures de distance inter-métiers



Notes : Cet histogramme représente le nombre de transitions réalisées sur le marché du travail en 2019 vers les 10 métiers d'arrivée les plus proches du métier d'origine selon différentes mesures de distance inter-métiers. Nous comparons la distance textuelle que nous construisons aux distances construites grâce aux rubriques "Mobilités" et "Compétences" du ROME V3.

Source : DADS Postes 2019 ; calcul des auteurs.

Le tableau 1 donne la régression, sur l'ensemble des couples de métiers pour lesquels au moins une transition a eu lieu, du logarithme du nombre de transitions réalisées par rapport à différentes mesures de distance inter-métiers normalisées. Cela nous permet de comparer la qualité de la mesure à prédire les transitions "à longue distance", c'est-à-dire même pour des métiers de départ et d'arrivée qui peuvent être assez éloignés en terme de compétences, par opposition à la figure précédente qui se concentrait sur les transitions les plus probables. On remarque que la part de la variance expliquée par notre mesure de distance inter-métiers est plus grande que pour les deux autres alternatives et que le coefficient de corrélation est plus élevé. Il n'est pas surprenant que les mesures de distance inter-métiers centrées sur les compétences n'expliquent qu'une faible part des transitions observées, dans la mesure où elles ne prennent pas en compte les nombreux autres facteurs mentionnés précédemment qui peuvent conduire à une mobilité professionnelle.

Tableau 1 – Pouvoir explicatif des mesures de distance inter-métiers sur les transitions réalisées

	log(transitions)	log(transitions)	log(transitions)	log(transitions)
Distance mobilités	-0.409 (0.0052)			-0.288 (0.0049)
Distance en compétences		-0.4501 (0.0051)		-0.287 (0.0049)
Distance textuelle			-0.609 (0.0052)	-0.495 (0.0052)
<i>N</i>	149,209	149,209	149,209	149,209
<i>R</i> ²	0.0409	0.0506	0.0935	0.1388

Notes : Cette table donne le résultat de quatre régressions dont la variable dépendante est le logarithme du nombre de transitions entre métiers. Les deux premières variables explicatives sont des mesures de distance construites à partir de des rubriques “Mobilités” et “Compétences” du ROME V3, que nous comparons à la distance textuelle construite pour les besoins de cette étude. Chaque mesure de distance inter-métiers a été normalisée de sorte que le coefficient donne directement la corrélation entre variable expliquée et explicative. Les données de transition utilisées sont tirées du fichier DADS Postes 2019.

2 Quel impact de la formation professionnelle sur le retour à l’emploi et les trajectoires professionnelles ?

Cette section étudie l’effet de la formation sur les trajectoires professionnelles des demandeurs d’emploi (retour à l’emploi, changement de métier etc.). Dans un premier temps, nous décrivons les données administratives exploitées et leur traitement préalable. Puis nous expliquons brièvement la stratégie empirique adoptée, avant de présenter les résultats obtenus.

2.1 Données

Nous utilisons les données ForCE arrêtées au T2 2022, qui regroupent⁵

1. le Fichier historique des demandeurs d’emploi (FH) qui rassemble pour chaque demandeur d’emploi des informations sur ses épisodes d’inscription à Pôle emploi sur les 10 années précédents le millésime de ForCE considéré ;
2. la base régionalisée des stagiaires de la formation professionnelle (Brest), qui rassemble pour chaque personne en recherche d’emploi (inscrit ou non à Pôle emploi) ayant été stagiaire de la formation professionnelle les caractéristiques des formations suivies depuis 2017 ;
3. la base Mouvements de main-d’œuvre (MMO, dont les données sont issues de la Décla-

5. Nous renvoyons à la description du dispositif faite par la Dares à la page web suivante pour plus d’informations : <https://dares.travail-emploi.gouv.fr/enquete-source/dispositif-force>

ration sociale nominative), qui rassemble depuis 2017 les informations sur les contrats de travail pour chaque salarié du privé,⁶ pour l'ensemble des demandeurs d'emploi présents dans le FH du millésime de ForCE considéré ;

4. la base I-Milo (issue du système d'information des missions locales), qui rassemble pour chaque jeune suivi en mission locale des informations sur les contacts avec la mission locale et les dispositifs suivis en mission locale, depuis 2016. Cette base n'est pas exploitée dans cette étude.

Nous construisons notre base d'analyse en suivant les mêmes principes que ceux adoptés par [Chabaud et al. \[2022\]](#) dans leur étude sur l'impact de la formation professionnelle sur le retour à l'emploi des demandeurs d'emploi – nous renvoyons à cette étude pour plus de détails quant à la construction de l'échantillon d'intérêt et la construction des variables. Ainsi, pour chaque mois entre janvier 2018 et décembre 2020, nous créons une base contenant l'ensemble des demandeurs d'emploi inscrits durant ce mois m (hors catégorie E / catégorie administrative 5).⁷, après avoir préalablement abouté les épisodes administratifs de demande d'emploi séparés de moins de 30 jours. Les demandeurs d'emploi dont nous pouvons inférer (via les informations présentes dans le FH) qu'ils ont suivi une formation avant 2017 sont exclus de l'analyse, qui se focalise sur les primo-formés. Cet échantillon est apparié à Brest afin de déterminer l'ensemble des demandeurs d'emploi entrant en formation dans le mois m , ce qui constitue le groupe de traitement pour ce mois-ci – tout en excluant du groupe de contrôle l'ensemble des individus ayant suivi une formation avant le mois m . Enfin, l'appariement avec la base MMO permet de récupérer des informations sur les contrats détenus par les demandeurs d'emploi (i) préalablement au mois m , et (ii) pour chacun des 24 mois suivant le mois m . Dans la mesure où cette étude se concentre sur les trajectoires professionnelles accomplies par les demandeurs d'emploi, notre échantillon principal restreint la population des demandeurs d'emploi aux personnes ayant détenu un emploi stable au cours des 12 mois précédant le mois m . Nous identifions ainsi pour chaque demandeur d'emploi un métier de référence qui, lorsque nous l'associons aux caractéristiques de l'emploi retrouvé après le mois m , nous permet de construire les principales variables d'intérêt de cette étude (distance occupationnelle parcourue, différentiel de tension, nature du contrat retrouvé...).

Nous conservons un grand nombre d'informations sur les individus en provenance du FH, telles que leur âge, sexe, situation familiale, nombre d'enfants, niveau de formation, qualification, situa-

6. Depuis 2022, les informations sur les contrats en cours d'une grande partie des salariés de la fonction publique sont également collectées.

7. La catégorie E / catégorie administrative 5 contient l'ensemble des demandeurs d'emploi inscrits et ayant un emploi à temps plein, n'ayant de ce fait aucune obligation de recherche d'emploi.

tion de handicap, habitation dans une zone rural ou urbaine, volonté de mobilité géographique, catégorie d'inscription, motif d'inscription, nationalité, métier recherché, salaire de réserve, ancienneté au chômage au mois m , temps passé au chômage avant l'épisode en question, nombres d'épisodes de chômage avec l'épisode en question.

2.2 Stratégie empirique

Notre stratégie d'identification et d'estimation tout au long de cette note repose en premier lieu sur une hypothèse d'indépendance conditionnelle. Formellement, définissons une variable $D_i \in \{0, 1\}$ indiquant si l'individu i est entré en formation durant le mois m ($D_i = 1$) ou non ($D_i = 0$). Pour une variable d'intérêt Y_i (par exemple, le retour à l'emploi dans les 24 mois suivants le mois m d'entrée potentielle en formation), on peut définir deux valeurs potentielles (selon le langage usuel de l'inférence causale) désignant la valeur prise par Y_i pour l'individu i dans le cas où i n'est pas entré en formation durant le mois m ($Y(0)$) ou dans le cas où l'entrée en formation a eu lieu ($Y(1)$). Intuitivement, l'hypothèse d'indépendance conditionnelle suppose qu'un individu traité ($D_i = 1$) aurait eu, en l'absence d'entrée en formation, un devenir similaire à un individu i' contrôle ($D_{i'} = 0$) ayant des caractéristiques observables X similaires à i . On note

$$Y_i(0) \perp D_i \mid X$$

Cette hypothèse permet d'identifier l'*average treatment effect on the treated* (ATT), c'est-à-dire l'effet moyen de l'entrée en formation au mois m parmi les individus étant effectivement entrés. Il est important de garder à l'esprit, dans l'interprétation de nos résultats, que le traitement étudié ici est bien l'entrée en (première) formation *durant le mois m* , par opposition à tout autre scénario (incluant l'absence totale d'entrée en formation, ou une entrée postérieure au mois m).

L'ATT peut être formellement défini de la manière suivante

$$\theta_{ATT} \equiv \mathbb{E}[Y(1) - Y(0) \mid D = 1]$$

et la preuve d'identification de ce paramètre est suffisamment simple pour être reportée ci-

dessous :

$$\begin{aligned}
 \theta_{ATT} &= E[Y(1) | D = 1] - E[E[Y(0)|D = 1, X]|D = 1] \quad (\text{Loi des espérances itérées}) \\
 &= E[Y | D = 1] - E[E[Y(0)|D = 0, X]|D = 1] \quad (\text{Indépendance conditionnelle}) \\
 &= \underbrace{E[Y | D = 1]}_{\text{Observé}} - \underbrace{E[E[Y|D = 0, X] | D = 1]}_{\text{Observé}}
 \end{aligned}$$

L'estimation associée à cette stratégie d'identification peut cependant se révéler plus complexe, dans la mesure où la dimensionalité de X (grand nombre de variables, variables continues) peut empêcher de construire de simples analogues empiriques aux quantités affichées ci-dessus. Ce problème peut-être dépassé de plusieurs manières différentes.

La première consiste à avoir recours à du *matching* sur score de propension. Ces méthodes sont bien connues comme solutions au problème de dimensionalité de X depuis [Rosenbaum and Rubin \[1983\]](#), et nous renvoyons à [Imbens and Rubin \[2015\]](#) (partie III) pour un traitement exhaustif de ces méthodes (ainsi qu'un traitement complet de la stratégie d'identification d'effet causaux sous hypothèse d'indépendance conditionnelle). Il s'agit de l'une des méthodes d'estimation que nous avons explorées, en suivant là encore les choix d'exécution faits par [Chabaud et al. \[2022\]](#) sur des données similaires. Nous renvoyons le lecteur à ce document pour plus de détails quant aux choix de variables de contrôles, méthode de *matching* choisie (mixte entre *matching* exacte sur le sexe, la catégorie d'âge et l'ancienneté au chômage), tests de robustesse, etc.⁸

La seconde méthode d'identification (et d'estimation) repose sur la méthode dite de *double/debiased machine learning*. Formellement, on note

$$\begin{aligned}
 E[Y | X, D] &= g(D, X), \\
 E[D | X] &= m(X).
 \end{aligned}$$

Sous cette hypothèse et ces notations, l'ATT peut être défini comme

$$\theta_{ATT} = E[g(1, X) - g(0, X) | D = 1]$$

et l'ATE, l'effet moyen dans l'ensemble de la population, peut être défini comme

$$\theta_{ATE} = E[g(1, X) - g(0, X)]$$

Les possibles facteurs confondants X affectent la variable de traitement D via le score de pro-

8. Concernant l'exécution concrète de cette méthode, nous avons eu recours au package R `MatchIt`, développé par [Ho et al. \[2011\]](#).

propension $m(X)$ et la variable d'intérêt Y via la fonction $g(D, X)$. Ces deux fonctions sont inconnues et potentiellement compliquées, et peuvent être estimées par *machine learning*. On utilise ensuite ces fonctions pour construire des moments identifiants les paramètres d'intérêts. Dans le cas de l'ATE, on a par exemple :

$$\theta_{ATE} = E \left[(g(1, X) - g(0, X)) + \frac{D(Y - g(1, X))}{m(X)} - \frac{(1 - D)(Y - g(0, X))}{1 - m(X)} \right]$$

puis l'on prend un analogue empirique à ce moment – en remplaçant les fonctions $g(D, X)$ et $m(X)$ par leurs estimateurs par *machine learning*. Les conditions nécessaires sur les méthodes d'estimation par *machine learning* pour que l'estimateur résultant de cette procédure soit asymptotiquement normal sont relativement faibles – nous renvoyons à Chernozhukov et al. [2018] pour un traitement complet de cette méthode.⁹ Par ailleurs, l'intuition derrière la propriété de *double robustness* de tels estimateurs est la suivante. Dans le cas où les modèles de l'espérance de $Y(1)$ et $Y(0)$ conditionnellement à X , $g(1, X)$ et $g(0, X)$, sont bien spécifiés, le terme $E[g(1, X) - g(0, X)]$ suffit à identifier l'ATE tandis que le terme restant, $E\left[\frac{D(Y - g(1, X))}{m(X)} - \frac{(1 - D)(Y - g(0, X))}{1 - m(X)}\right]$, est de moyenne nulle – quand bien même le modèle pour le score de propension $m(X)$ se révélait inadéquat. Au contraire, dans la situation opposée où le modèle du score de propension est bien spécifié – tandis que $g(1, X)$ et $g(0, X)$ seraient arbitrairement éloignés de l'espérance conditionnelle de $Y(1)$ et $Y(0)$ – le second terme $E\left[\frac{D(Y - g(1, X))}{m(X)} - \frac{(1 - D)(Y - g(0, X))}{1 - m(X)}\right]$ suffit à lui seul à identifier l'ATE par re-pondération et annuler le biais causé par le premier terme $E[g(1, X) - g(0, X)]$. Ainsi, la validité de l'un des deux modèles – espérance(s) conditionnelle(s) ou score de propension – est suffisante pour l'identification, d'où le nom de *double robustness*.

2.3 Description de l'échantillon étudié

Cette sous-partie décrit l'échantillon de demandeurs d'emplois étudié lors de nos analyses. Comme nous l'avons déjà souligné, notre population d'intérêt se restreint à des personnes pour lesquelles nous sommes en mesure de définir un métier de référence et qui sont, par conséquent, relativement proches de l'emploi.

Le tableau 2 compare les demandeurs d'emploi pour lesquels un contrat de travail stable (CDI ou CDD de plus de 6 mois) terminé il y a moins de 12 mois est observé (notre population d'intérêt), à ceux pour lesquels ce n'est pas le cas. La première sous-population est la seule pour laquelle nous pouvons étudier l'effet de la formation sur le changement de compétences et le différentiel

9. Les estimateurs par *machine learning* des fonctions $g(X, D)$ et $m(X)$ doivent "seulement" être $n^{1/4}$ -convergentes. Concernant l'exécution concrète de cette méthode, des packages sont disponibles sous Python et R (langage utilisé pour cette étude) à l'adresse suivante : <https://docs.doubleml.org/stable/index.html>.

de tension entre métier d'origine et métier d'arrivée. Cette sous-population apparaît plus jeune, avec une proportion plus faible de femmes, davantage de diplômés du supérieur, est inscrite depuis moins longtemps au chômage et se forme davantage.

Tableau 2 – Statistiques descriptives selon l'observation (ou non) d'un contrat stable dans les 12 mois précédents le mois d'étude

	Sans emploi stable dans l'année écoulée	Avec un emploi stable dans l'année écoulée
Femme (%)	53,7	50,2
Age (%)		
< 25	10,8	12
[25, 50]	61,5	66,7
> 50	23,1	16,4
Études (%)		
< Bac	50,2	42,3
Bac	23	24
> Bac	26,8	33,7
Durée d'inscription (mois)	31,1	12,8
Formation (%)	1,3	2
Formation certifiante (%)	0,6	1
Formation > 420h (%)	0,7	1,2
Observations (milliers)	4614	1007

Source : Données ForCE. Ensemble des demandeurs d'emploi inscrits au cours du mois de septembre 2018.

Le tableau 3 présente les caractéristiques des demandeurs d'emploi composant notre population d'intérêt selon le type de formation suivie. Nous distinguons les demandeurs d'emploi n'entreprenant aucune formation des demandeurs d'emploi entreprenant une formation de plus de 30 heures selon le type (certifiante ou non) et la durée de celle-ci (supérieure ou inférieure à 420 heures, soit 3 mois à temps complet).¹⁰

10. Il existe une saisonnalité importante dans le nombre d'entrées en formation parmi les demandeurs d'emploi. Les mois de septembre correspondent à un pic du nombre d'entrées. Nous choisissons de présenter les statistiques pour le mois de septembre 2018 parce qu'il contient un grand nombre de formés, et donne ainsi une image informative de cette sous-population.

Tableau 3 – Statistiques descriptives selon le statut de formation pour la population des demandeurs d'emploi avec un emploi stable au cours de l'année écoulée

	Sans formation	Formation	Formation certifiante	Formation > 420h
Femme (%)	49,3	59,8	63,8	66,2
Age (%)				
< 25	11,3	18,3	20,7	22,8
[25, 50]	66,5	67,9	68,1	67,1
> 50	17,3	8,6	5,9	4,7
Études (%)				
< Bac	43,4	31,8	25,9	24,5
Bac	23,2	32,2	35,7	36,6
> Bac	33,4	36,0	38,3	38,9
Durée d'inscription (mois)	13,2	8,1	7,9	7,8
Changement de métier désiré (%)	73,9	74,8	73,3	73,6
Distance si changement	0,51	0,53	0,52	0,53
Observation (milliers)	986,8	20,4	10,5	12,3

Source : Données ForCE. Ensemble des demandeurs d'emploi inscrits au cours du mois de septembre 2018 pour lesquels un contrat de travail (CDI ou CDD de plus de 6 mois) terminé il y a moins de 12 mois est observé.

Comparée aux demandeurs d'emploi n'entretenant pas une formation de plus de 30 heures, la sous-population des demandeurs d'emploi débutant une première formation de plus de 30h est en moyenne plus jeune, au chômage depuis moins longtemps, et contient une proportion plus importante de femmes et de diplômés du supérieur. Ces observations sont toutes encore plus prononcées dans les sous-populations des individus débutant une formation certifiante ou une formation longue de plus de 420 heures. Par ailleurs, un fait probablement surprenant est l'absence de différence notable de l'aspiration à changer de métier entre ces différentes populations. Nous mesurons cette aspiration comme la part des demandeurs d'emploi déclarant rechercher un métier différent du dernier métier exercé de manière stable.¹¹ On observe que trois quarts des individus indiquent chercher un métier différent de celui précédemment exercé – cette proportion n'évoluant pas ou très peu quel que soit le statut de traitement. De plus, l'ampleur des changements en compétences envisagés – tels que mesurés par la distance en compétence entre le métier recherché et le métier précédemment exercé – est comparable entre individus débutant ou non une formation, quel que soit le type de celle-ci.

2.4 Résultats

Nous présentons ici les résultats de nos analyses concernant l'effet de la formation sur le retour à l'emploi et la réallocation de main d'œuvre entre marchés du travail. Tous les résultats présentés dans cette partie sont obtenus grâce la technique du "Double Debaised Machine Learning"

11. Nous retenons dans cet exercice un métier exprimé dans la version la plus détaillée de la classification FAP, soit 225 postes. Un passage à un niveau moins granulaire, par exemple la FAP 87, ne change pas le résultat. Le passage à la classification FAP permet de faire concorder le ROME utilisé par le FH et la PCS utilisée par les MMO.

(voir supra) et sont largement comparables à ceux obtenus en appliquant une méthode plus “classique” d'appariement sur score de propension.

Le tableau 4 représente l'effet de la formation professionnelle sur le retour à l'emploi à différents horizons temporels (6, 12 et 24 mois après l'entrée en formation), pour l'ensemble des demandeurs d'emploi primo-formés entre janvier 2018 et juin 2020 pour lesquels nous observons un emploi stable au cours de l'année écoulée. Comme largement documenté dans la littérature [Card et al., 2018], nous retrouvons un effet dit de *lock-in*, c'est-à-dire un effet négatif de l'entrée en formation sur le retour à l'emploi à court terme lié au temps consacré à la formation. L'effet de la formation sur le retour à l'emploi est positif 12 mois après l'entrée en formation, et continue à augmenter jusqu'à 24 mois après l'entrée en formation. Nos résultats, qui portent sur une sous-population relativement proche de l'emploi diffèrent tant qualitativement que quantitativement de ceux rapportés par Chabaud et al. [2022]. D'un point de vue qualitatif, l'effet de *lock-in* apparaît plus prononcé et plus durable sur notre population d'intérêt qu'il ne l'est pour l'ensemble des inscrits à Pôle emploi. D'un point de vue quantitatif, l'effet de la formation professionnelle à 24 mois pour notre population d'intérêt est environ 25% plus faible que le résultat que nous obtenons sur l'ensemble de la population des inscrits à Pôle emploi. Ces différences peuvent s'expliquer par le fait que les individus de notre échantillon principal sont par construction plus proche de l'emploi. Ainsi, une entrée en formation est davantage susceptible de réduire significativement les opportunités de retour à l'emploi pour cette sous-population, pour laquelle de telles opportunités sont nombreuses même en l'absence de formation.

Le tableau 4 montre que l'effet à 24 mois de la formation sur le retour à l'emploi est bien positif lorsque l'on considère l'ensemble des formations, cet effet est légèrement inférieur pour les formations certifiantes. La deuxième moitié du tableau montre que cette logique est respectée si l'on restreint la variable d'intérêt au retour à l'emploi dans un contrat stable (CDI ou CDD de plus de 6 mois). Cette différence de taux de retour à l'emploi entre l'ensemble des formations et les formations certifiantes doit être interprété avec prudence. D'une part on constate des effets de “lock in” nettement plus marqués pour les formations certifiantes que pour l'ensemble des formations en général. D'autre part beaucoup de formations courtes sont associées à un projet de recrutement sur des contrats stables (CDI ou CDD supérieurs à 12 mois pour les POEC et les POEI). Il n'est donc pas surprenant que ce genre de formations courtes dont la durée n'excède pas 400h aboutisse à un fort taux de retour à l'emploi, y compris dans des emplois stables.

Tableau 4 – Effets de la formation professionnelle à différents horizons

	(1)	(2)	(3)
	6 mois	12 mois	24 mois
Tout type d'emploi			
Toute formation	-0.007 (0.0018)	0.064 (0.0019)	0.094 (0.0016)
Formation certifiante	-0.096 (0.0026)	0.034 (0.003)	0.093 (0.0025)
Emploi stable			
Toute formation	0.011 (0.0016)	0.044 (0.0018)	0.065 (0.0019)
Formation certifiante	-0.073 (0.0028)	-0.0059 (0.0037)	0.043 (0.0038)

Notes : Ce tableau présente les résultats de régressions séparées de 3 variables dépendantes différentes (en colonne, correspondant au retour à l'emploi pour différents horizons temporels) sur 2 variables explicatives (en ligne, correspondant aux différents types de formation considérés). La partie haute du tableau considère le retour à tout type d'emploi à différents horizons temporels, tandis que la partie basse répète les analyses en ne considérant comme variable dépendante que le retour à l'emploi stable (CDI, CDD de plus de 6 mois). Les erreurs standard clusterisées au niveau métier/zone d'emploi sont présentées entre parenthèses.

Le tableau 5 résume nos principaux résultats sur l'effet de la formation après 24 mois sur les trajectoires professionnelles des demandeurs d'emploi pour différents types de formation (toutes formations confondues et formations certifiantes). Comme précédemment nous présentons l'ensemble de ces résultats sur l'échantillon des demandeurs d'emploi pour lesquels nous observons un contrat stable au cours de l'année écoulée. Dans ce tableau la variable dépendante de retour à l'emploi est décomposée en fonction de la distance entre l'emploi retrouvé et l'emploi d'origine. On distingue ainsi un retour à l'emploi dans le métier d'origine ($d = 1$), dans un métier très proche du métier d'origine ($d \in [2, 5]$), dans un métier proche du métier d'origine ($d \in [6, 20]$), et finalement dans un métier dont les compétences sont éloignées de celles du métier d'origine ($d > 20$). En procédant ainsi on constate que la formation professionnelle réduit la probabilité de retour à l'emploi dans le métier d'origine des demandeurs d'emploi, est presque neutre pour les métiers proches, et augmente de manière substantielle la probabilité de retour à l'emploi dans

les métiers dont les compétences sont éloignées de celles du métier d'origine. Ces conclusions sont encore plus prononcées lorsque l'on restreint les formations considérées aux seules formations certifiantes. Ce résultat est frappant dans la mesure où il implique que la totalité de l'effet sur le retour à l'emploi de la formation professionnelle passe par des transitions professionnelles dirigées vers des métiers éloignés du métier d'origine des demandeurs d'emplois.

Tableau 5 – Effets de la formation professionnelle sur le retour à l'emploi (à 24 mois) dans des métiers plus ou moins proches du métier d'origine

	(1) Métier d'origine ($d = 1$)	(2) Métier très proche ($d \in [2, 5]$)	(3) Métier proche ($d \in [6, 20]$)	(4) Métier éloigné ($d > 20$)
Toute formation	-0.034 (0.0012)	0.0077 (0.001)	0.0097 (0.0011)	0.087 (0.0018)
Formation certifiante	-0.043 (0.0018)	0.012 (0.0016)	0.006 (0.0017)	0.097 (0.0037)

Notes : Ce tableau présente les résultats de régressions séparées de 4 variables dépendantes différentes (en colonne) sur 2 variables explicatives (en ligne, correspondant aux différents types de formation considérés). A titre d'exemple, les coefficients de la colonne (2), "Métier très proche ($d \in [2, 5]$)", correspondent à l'effet de l'entrée en formation (certifiante ou non) sur le retour à l'emploi (à 24 mois) dans l'un des 4 métiers les plus proches du métier précédemment exercé, selon notre mesure de distance inter-métiers. Les erreurs standard clusterisées au niveau métier/zone d'emploi sont présentées entre parenthèses.

Les redirections professionnelles permises par le recours à la formation permettent-elles aux demandeurs d'emploi de réorienter leur effort de recherche vers des marchés en forte tension ? Les tableaux 6 et 7 prolongent l'analyse de l'impact de l'entrée en formation sur les trajectoires professionnelles en distinguant le retour à l'emploi dans des marchés en plus forte ou en plus faible tension que celle du métier précédemment exercé par le demandeur d'emploi. Dans l'ensemble, l'entrée en formation (certifiante ou non) semble augmenter de manière comparable la probabilité d'une transition vers un métier en plus forte ou en plus faible tension. Dans la perspective d'un usage du système de formation visant à réduire les déséquilibres d'offre et de demande de travail entre différents marchés du travail, ce résultat peut apparaître comme relativement décevant. Il suggère en effet qu'une plus grande attention portée au ciblage de l'offre de formation permettrait de rediriger plus encore la main d'œuvre vers les marchés les plus en tension, quand à ce jour les transitions générées par le passage en formation apparaissent comme relativement neutres de ce point de vue.

Tableau 6 – Effets de la formation professionnelle sur le retour à l'emploi (à 6, 12, 24 mois) dans des marchés en faible et forte tension

	(1) 6 mois	(2) 12 mois	(3) 24 mois
Emploi retrouvé dans un marché...			
... en plus faible tension			
Toute formation	0.013 (0.0011)	0.043 (0.0015)	0.057 (0.0017)
Formation certifiante	-0.016 (0.0016)	0.032 (0.0023)	0.053 (0.0027)
... en plus forte tension			
Toute formation	0.011 (0.0013)	0.038 (0.0016)	0.047 (0.0018)
Formation certifiante	-0.013 (0.0016)	0.037 (0.0023)	0.059 (0.0026)

Notes : Ce tableau reprend les analyses du tableau 4 en distinguant, dans la construction des variables dépendantes, le retour à l'emploi dans un marché en plus faible (ou plus forte) tension que le métier précédemment exercé par le demandeur d'emploi. Les erreurs standard clusterisées au niveau métier × zone d'emploi sont présentées entre parenthèses.

Ce constat agrégé cache cependant une hétérogénéité importante, révélée dans le tableau 7 où nous distinguons les demandeurs d'emplois selon que leur marché d'origine est en faible ou en forte tension. Le panel A montre que la formation a un effet plus important sur les transitions vers des métiers en plus forte tension que le métier d'origine lorsque ce dernier est un métier en faible tension. Cela suggère que pour ce public initialement situé dans des marchés où la demande de travail est faible, le système de formation professionnelle génère très clairement des transitions allant dans le sens d'un rééquilibrage entre offre et demande de travail, un constat encourageant. A l'inverse le panel B du tableau 7, montre que le résultat agrégé relativement décevant du tableau 6 s'explique par une situation tout à fait opposée pour les publics formés depuis des marchés du travail relativement tendus (tableau 7, panel B). D'un point de vue purement réallocatif, l'effet sur le retour à l'emploi de la formation professionnelle ne semble donc pas être tiré par des redirections plus nombreuses vers des métiers en forte tension.

Tableau 7 – Effets de la formation professionnelle sur le retour à l’emploi (à 6, 12, 24 mois) dans des marchés en faible et forte tension, selon la tension du marché d’origine

	(1) 6 mois	(2) 12 mois	(3) 24 mois
Emploi retrouvé dans un marché...			
<i>A - Marché d’origine en faible tension</i>			
	... en plus faible tension		
Toute formation	0.0025 (0.0012)	0.014 (0.0015)	0.020 (0.0018)
Formation certifiante	-0.011 (0.0015)	0.009 (0.0024)	0.018 (0.0028)
	... en plus forte tension		
Toute formation	0.02 (0.0022)	0.062 (0.0027)	0.089 (0.0030)
Formation certifiante	-0.017 (0.0031)	0.059 (0.0042)	0.099 (0.0047)
<i>B - Marché d’origine en forte tension</i>			
	... en plus faible tension		
Toute formation	0.018 (0.0018)	0.057 (0.0022)	0.072 (0.0024)
Formation certifiante	-0.021 (0.0025)	0.041 (0.0033)	0.071 (0.0037)
	... en plus forte tension		
Toute formation	0.007 (0.0013)	0.024 (0.0016)	0.031 (0.0019)
Formation certifiante	-0.012 (0.0018)	0.022 (0.0026)	0.034 (0.003)

Notes : Ce tableau reprend les analyses du tableau 6 en distinguant deux sous-échantillons. Le panel A (respectivement B), “Marché d’origine en faible (resp. forte) tension”, se restreint aux demandeurs d’emploi dont le précédent métier exercé était en faible (resp. forte) tension.

Conclusion

La formation professionnelle permet-elle de réduire les déséquilibres structurels du marché du travail? Cette étude vise à apporter des éléments de réponse à cette question en comparant les trajectoires professionnelles réalisées par les demandeurs d'emploi ayant ou non recours à la formation professionnelle. Sur un échantillon de demandeurs d'emploi relativement proches de l'emploi, nos résultats montrent que, par rapport aux demandeurs d'emploi n'ayant pas recours à la formation professionnelle, les demandeurs d'emploi formés accomplissent des transitions professionnelles plus lointaines dans l'espace de compétences. D'un point de vue purement ré-allocatif, l'effet sur le retour à l'emploi de la formation professionnelle ne semble cependant pas être tiré par des redirections plus nombreuses vers des métiers en forte tension. Ce résultat laisse présumer qu'un ciblage plus systématique de l'offre de formation professionnelle sur les compétences demandées par les métiers en tension permettrait d'accroître l'effet réallocatif de la formation professionnelle.

Références

- Isaac Baley, Ana Figueiredo, and Robert Ulbricht. Mismatch cycles. 2020.
- Sarah H. Bana, Erik Brynjolfsson, Daniel Rock, and Sebastian Steffen. job2vec : Learning a representation of jobs. 2021.
- Regis Barnichon and Andrew Figura. Labor market heterogeneity and the aggregate matching function. *American Economic Journal : Macroeconomics*, 7(4) :222–49, 2015.
- Gary S. Becker. *Human Capital : A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. Number beck-5 in NBER Books. National Bureau of Economic Research, Inc, June 1964. URL <https://ideas.repec.org/b/nbr/nberbk/beck-5.html>.
- David Card, Jochen Kluge, and Andrea Weber. What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3) :894–931, 2018.
- Mathieu Chabaud, Anne Bucher, Pauline Givord, and Alexandra Louvet. Quelles sont les chances de retour à l'emploi après une formation? *Document d'études*, 261, 2022.
- Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 2018.
- Nikolas Dawson, Mary-Anne Williams, and Marian-Andrei Rizoiu. Skill-driven recommendations for job transition pathways. *PLOS One*, 2021.
- Dita Eckardt. Are chemists good bankers? returns to the match between training and occupation. Technical report, Working Paper, 2022.
- François Fontaine and Roland Rathelot. Le marché du travail français à l'épreuve de la crise sanitaire. *Notes du conseil danalyse economique*, 71(2) :1–12, 2022.
- Christina Gathmann and Uta Schönberg. How general is human capital? a task-based approach. *Journal of Labor Economics*, 28(1) :1–49, 2010.
- Matthew Gentzkow, Bryan Kelly, and Matt Taddy. Text as Data. *Journal of Economic Literature*, 57(3) :535–574, 2019.
- Fatih Guvenen, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. Multidimensional skill mismatch. *American Economic Journal : Macroeconomics*, 12(1) :210–44, 2020.

- Daniel Ho, Kosuke Imai, Gary King, and Elizabeth A. Stuart. Matchit : Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8) :1-28, 2011. doi : 10.18637/jss.v042.i08. URL <https://www.jstatsoft.org/index.php/jss/article/view/v042i08>.
- Benjamin G Hyman. Can displaced labor be retrained? evidence from quasi-random assignment to trade adjustment assistance. In *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, volume 111, pages 1-70. JSTOR, 2018.
- Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences : An Introduction*. Cambridge University Press, 2015. doi : 10.1017/CBO9781139025751.
- Louis S Jacobson, Robert J LaLonde, and Daniel G Sullivan. Earnings losses of displaced workers. *The American economic review*, pages 685-709, 1993.
- Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoît Crabbé, Laurent Besacier, and Didier Schwab. Flaubert : Unsupervised language model pre-training for french. *arXiv preprint arXiv :1912.05372*, 2019.
- Ilse Lindenlaub and Fabien Postel-Vinay. The worker-job surplus. Technical report, National Bureau of Economic Research, 2021.
- Ioana Marinescu and Roland Rathelot. Mismatch unemployment and the geography of job search. *American Economic Journal : Macroeconomics*, 2018.
- Martí Mestieri, Sergi Basco, Gabriel Smagghue, and Maxime Liegey. The heterogeneous effects of trade across occupations : A test of the stolper-samuelson theorem. 2020.
- Daniel M. Ringel. Multimarket Membership Mapping. *Journal of Marketing Research*, 2022.
- Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1) :41-55, 1983.
- Ayşegül Şahin, Joseph Song, Giorgio Topa, and Giovanni L Violante. Mismatch unemployment. *American Economic Review*, 104(11) :3529-64, 2014.
- Kathryn L Shaw. A formulation of the earnings function using the concept of occupational investment. *Journal of Human Resources*, pages 319-340, 1984.
- Sharon Traiberman. Occupations and import competition : Evidence from denmark. *American Economic Review*, 109(12) :4260-4301, 2019.

Xin (Shane) Wang, Jiaxiu He, David J. Curry, and Jun Hyun (Joseph) Ryoo. Attribute Embedding : Learning Hierarchical Representations of Product Attributes from Consumer Reviews. *Journal of Marketing*, 2021.

Jason Weston, Samy Bengio, and Nicolas Usunier. Wsabie : Scaling Up To Large Vocabulary Image Annotation. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*, 2011.

A Caractéristiques de la mesure de distance en compétences

A.1 Métiers les plus proches selon la mesure de distance pour quelques métiers types

Voisins de Entretien des espaces verts, code A1203

- Entretien et surveillance du tracé routier, $d = 0.440$
- Horticulture et maraîchage, $d = 0.440$
- Maintenance des bâtiments et des locaux, $d = 0.453$
- Création en arts plastiques, $d = 0.461$
- Sylviculture, $d = 0.467$

Voisins de Réalisation de vêtements sur mesure ou en petite série, code B1803

- Tapisserie - décoration en ameublement, $d = 0.352$
- Retouches en habillement, $d = 0.448$
- Coupe cuir, textile et matériaux souples, $d = 0.473$
- Assemblage - montage d'articles en cuirs, peaux, $d = 0.478$
- Préparation de fils, montage de métiers textiles, $d = 0.486$

Voisins de Transaction immobilière, code C1504

- Téléconseil et télévente, $d = 0.472$
- Relation commerciale auprès de particuliers, $d = 0.494$
- Assistanat commercial, $d = 0.504$
- Assistance et support technique client, $d = 0.511$
- Gérance immobilière, $d = 0.514$

Voisins de Vente en habillement et accessoires de la personne, code D1214

- Vente en gros de matériel et équipement, $d = 0.264$
- Achat vente d'objets d'art, anciens ou d'occasion, $d = 0.280$
- Coiffure, $d = 0.446$
- Vente en articles de sport et loisirs, $d = 0.464$
- Vente en décoration et équipement du foyer, $d = 0.465$

Voisins de Communication, code E1103

- Conception de contenus multimédias, $d = 0.395$
- Animation de site multimédia, $d = 0.419$
- Organisation d'évènementiel, $d = 0.439$
- Journalisme et information média, $d = 0.465$
- Développement local, $d = 0.492$

Voisins de Peinture en bâtiment, code F1606

- Montage de réseaux électriques et télécoms, $d = 0.267$
- Installation d'équipements sanitaires et thermiques, $d = 0.281$
- Réalisation et restauration de façades, $d = 0.284$
- Pose de revêtements rigides, $d = 0.286$
- Pose de revêtements souples, $d = 0.288$

Voisins de Personnel de cuisine, code G1602

- Management du personnel de cuisine, $d = 0.229$
- Fabrication de crêpes ou pizzas, $d = 0.362$
- Personnel polyvalent en restauration, $d = 0.420$
- Management d'établissement de restauration collective, $d = 0.467$
- Service en restauration, $d = 0.475$

Voisins de Opérations manuelles d'assemblage, tri ou emballage, code H3302

- Montage-assemblage mécanique, $d = 0.335$
- Conduite d'équipement de conditionnement, $d = 0.335$
- Conduite d'équipement de production alimentaire, $d = 0.353$
- Préparation de matières et produits industriels (broyage, mélange, ...), $d = 0.393$
- Conduite d'installation de production des métaux, $d = 0.433$

Voisins de Maintenance de bâtiments et des locaux, code I1203

- Entretien des espaces naturels, $d = 0.414$
- Aménagement et entretien des espaces verts, $d = 0.453$
- Nettoyage des espaces urbains, $d = 0.486$
- Gardiennage de locaux, $d = 0.486$
- Nettoyage de locaux, $d = 0.491$

Voisins de Personnel polyvalent des services hospitaliers, code J1301

- Assistance médico-technique, $d = 0.343$
- Préparation en pharmacie, $d = 0.350$
- Imagerie médicale, $d = 0.394$
- Analyses médicales, $d = 0.413$
- Pharmacie, $d = 0.428$

Voisins de Assistance auprès d'enfants, code K1303

- Accompagnement médicosocial, $d = 0.409$
- Services domestiques, $d = 0.410$
- Éducation de jeunes enfants, $d = 0.416$
- Assistance auprès d'adultes, $d = 0.472$
- Personnel polyvalent des services hospitaliers, $d = 0.476$

Voisins de Musique et chant, code L1202

- Biologie médicale, $d = 0.512$
- Kinésithérapie, $d = 0.519$
- Gestion des opérations de circulation internationale des marchandises, $d = 0.523$
- Danse, $d = 0.534$
- Navigation commerciale aérienne, $d = 0.556$

Voisins de Secrétariat, code M1607

- Assistanat de direction, $d = 0.238$
- Assistanat technique et administratif, $d = 0.270$
- Accueil et renseignements, $d = 0.278$
- Secrétariat comptable, $d = 0.352$
- Secrétariat et assistanat médical ou médico-social, $d = 0.394$

Voisins de Magasinage et préparation de commandes, code N1103

- Conduite d'engins de déplacement des charges, $d = 0.361$
- Manutention manuelle de charges, $d = 0.378$
- Intervention technique d'exploitation des transports routiers de marchandises, $d = 0.493$
- Vente en gros de matériel et équipement, $d = 0.494$
- Opérations manuelles d'assemblage, tri ou emballage, $d = 0.501$

A.2 *Métiers les plus proches sans mobilité suggérée dans ROME V3 pour quelques métiers types*

Voisins de Entretien des espaces verts, code A1203

- Entretien et surveillance du tracé routier, $d = 0.440$
- Horticulture et maraîchage, $d = 0.440$
- Maintenance des bâtiments et des locaux, $d = 0.453$
- Création en arts plastiques, $d = 0.461$

Voisins de Réalisation de vêtements sur mesure ou en petite série, code B1803

- Tapisserie - décoration en ameublement, $d = 0.352$
- Retouches en habillement, $d = 0.448$

Voisins de Transaction immobilière, code C1504

- Téléconseil et télévente, $d = 0.472$
- Relation commerciale auprès de particuliers, $d = 0.494$
- Assistanat commercial, $d = 0.504$
- Assistance et support technique client, $d = 0.511$

Voisins de Vente en habillement et accessoires de la personne, code D1214

- Vente en gros de matériel et équipement, $d = 0.264$
- Achat vente d'objets d'art, anciens ou d'occasion, $d = 0.280$
- Coiffure, $d = 0.446$
- Vente en articles de sport et loisirs, $d = 0.464$

Voisins de Communication, code E1103

- Animation de site multimédia, $d = 0.419$
- Organisation d'évènementiel, $d = 0.439$
- Journalisme et information média, $d = 0.465$

Voisins de Peinture en bâtiment, code F1606

- Montage de réseaux électriques et télécoms, $d = 0.267$
- Installation d'équipements sanitaires et thermiques, $d = 0.281$
- Réalisation et restauration de façades, $d = 0.284$
- Pose de revêtements rigides, $d = 0.286$

Voisins de Personnel de cuisine, code G1602

- Management du personnel de cuisine, $d = 0.229$
- Fabrication de crêpes ou pizzas, $d = 0.362$
- Personnel polyvalent en restauration, $d = 0.420$
- Management d'établissement de restauration collective, $d = 0.467$

Voisins de Opérations manuelles d'assemblage, tri ou emballage, code H3302

- Montage-assemblage mécanique, $d = 0.335$
- Conduite d'équipement de conditionnement, $d = 0.335$
- Conduite d'équipement de production alimentaire, $d = 0.353$
- Préparation de matières et produits industriels (broyage, mélange, ...), $d = 0.393$

Voisins de Maintenance de bâtiments et des locaux, code I1203

- Entretien des espaces naturels, $d = 0.414$
- Aménagement et entretien des espaces verts, $d = 0.453$
- Nettoyage des espaces urbains, $d = 0.486$
- Gardiennage de locaux, $d = 0.486$

Voisins de Personnel polyvalent des services hospitaliers, code J1301

- Assistance médico-technique, $d = 0.343$
- Préparation en pharmacie, $d = 0.350$
- Imagerie médicale, $d = 0.394$
- Analyses médicales, $d = 0.413$

Voisins de Assistance auprès d'enfants, code K1303

- Accompagnement médicosocial, $d = 0.409$
- Services domestiques, $d = 0.410$
- Éducation de jeunes enfants, $d = 0.416$
- Assistance auprès d'adultes, $d = 0.472$

Voisins de Musique et chant, code L1202

- Biologie médicale, $d = 0.512$
- Kinésithérapie, $d = 0.519$
- Danse, $d = 0.534$

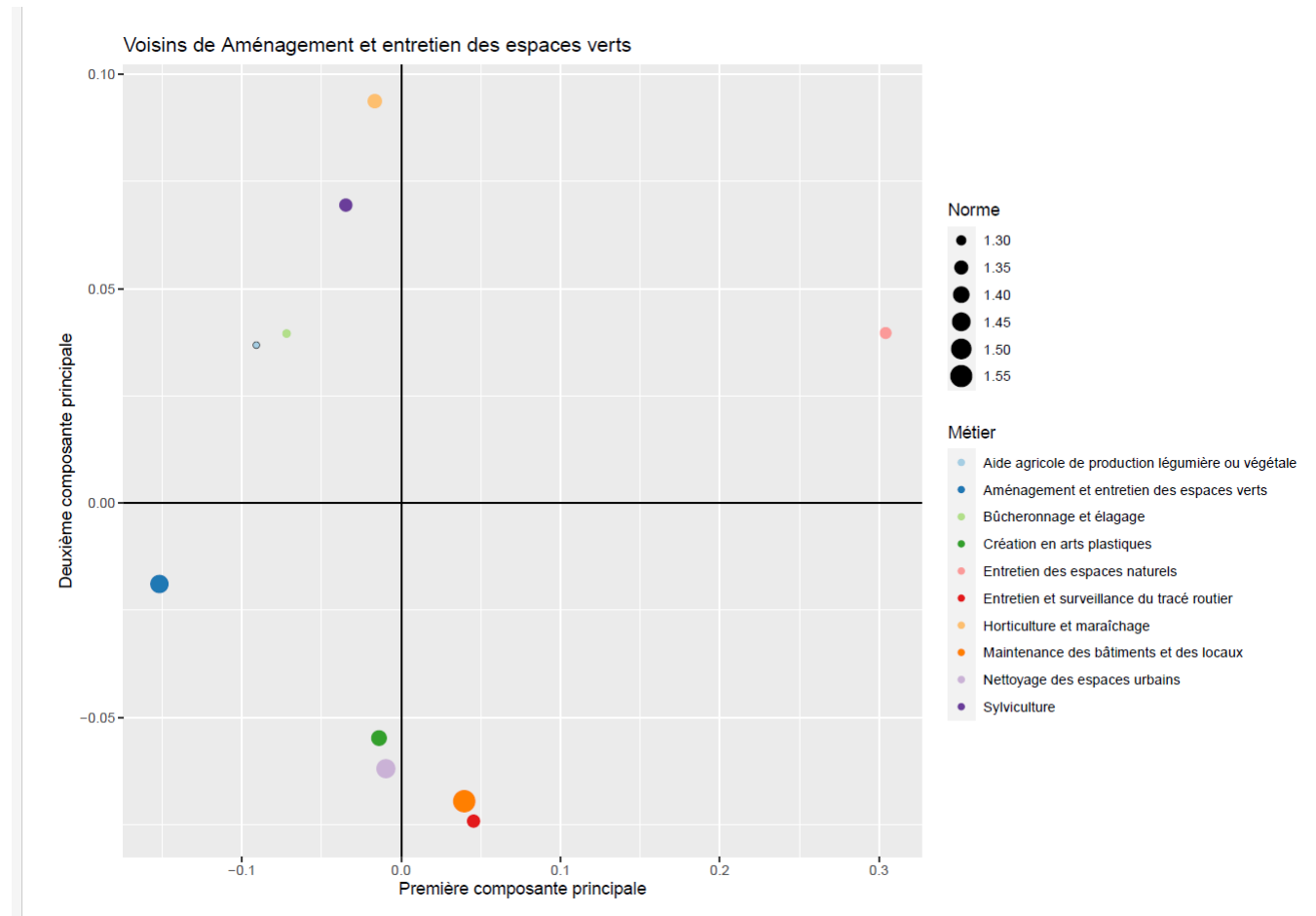
Voisins de Secrétariat, code M1607

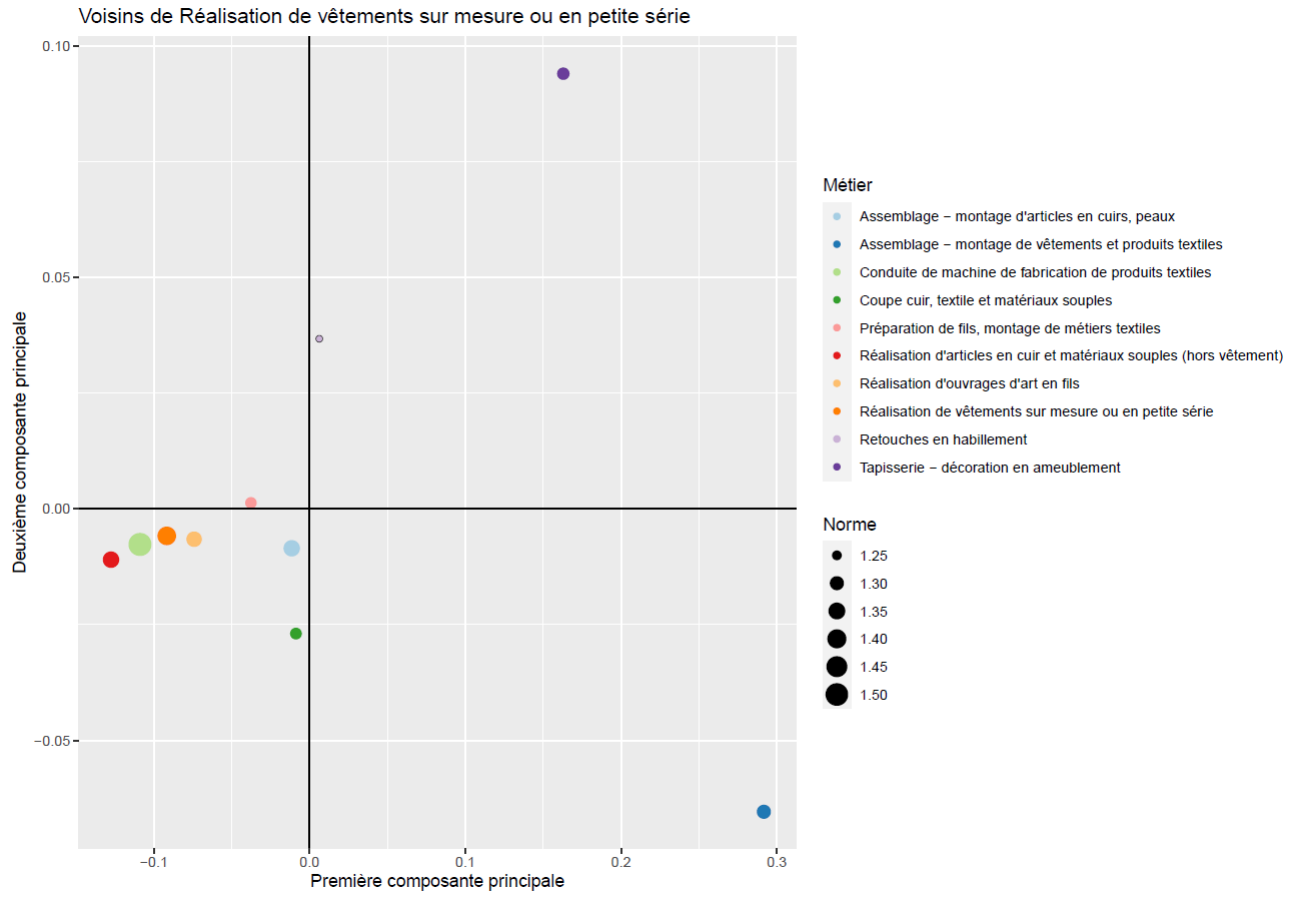
- Assistanat de direction, $d = 0.238$
- Assistanat technique et administratif, $d = 0.270$
- Accueil et renseignements, $d = 0.278$
- Secrétariat comptable, $d = 0.352$

Voisins de Magasinage et préparation de commandes, code N1103

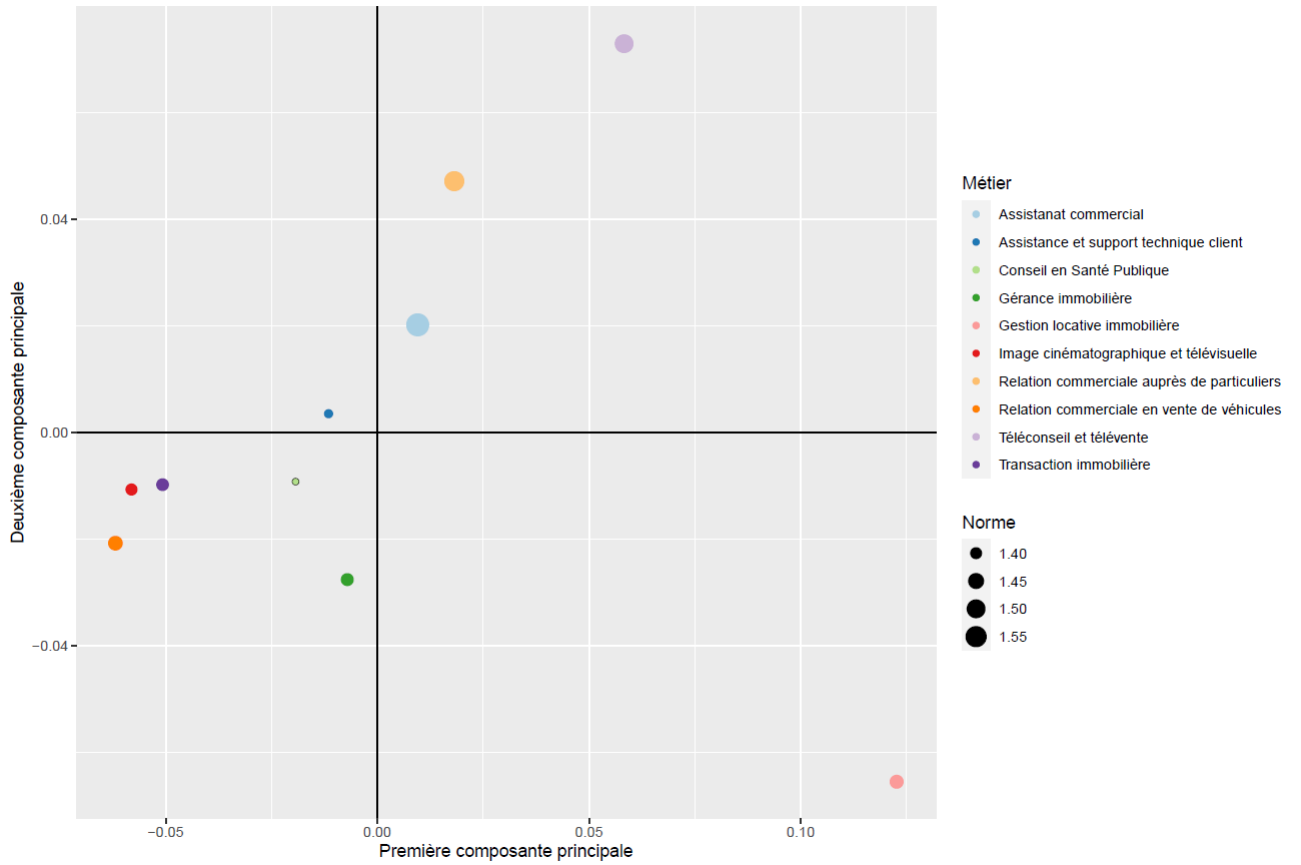
- Conduite d'engins de déplacement des charges, $d = 0.361$
- Manutention manuelle de charges, $d = 0.378$
- Intervention technique d'exploitation des transports routiers de marchandises, $d = 0.493$
- Vente en gros de matériel et équipement, $d = 0.494$

A.3 Visualisation du positionnement relatif des métiers au voisinage de quels métiers types

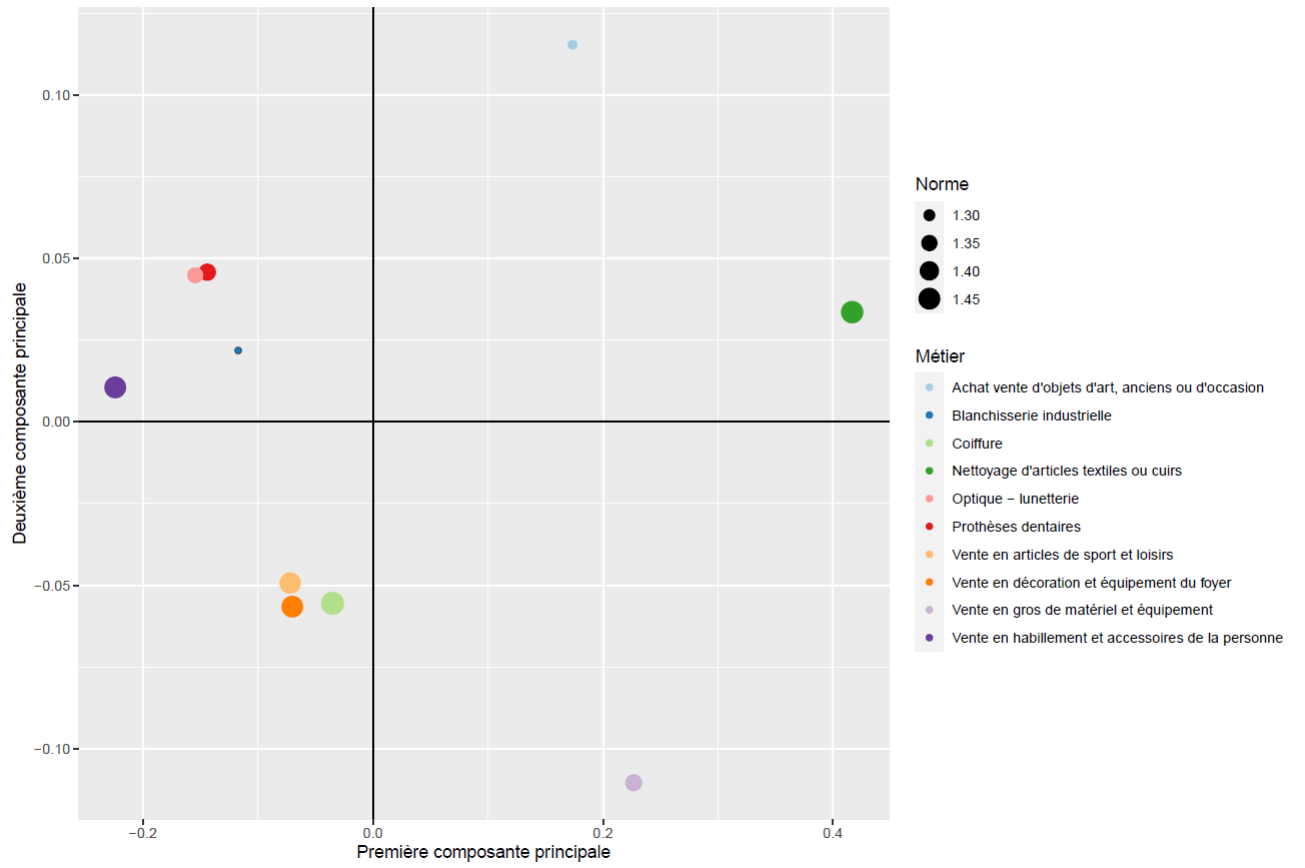


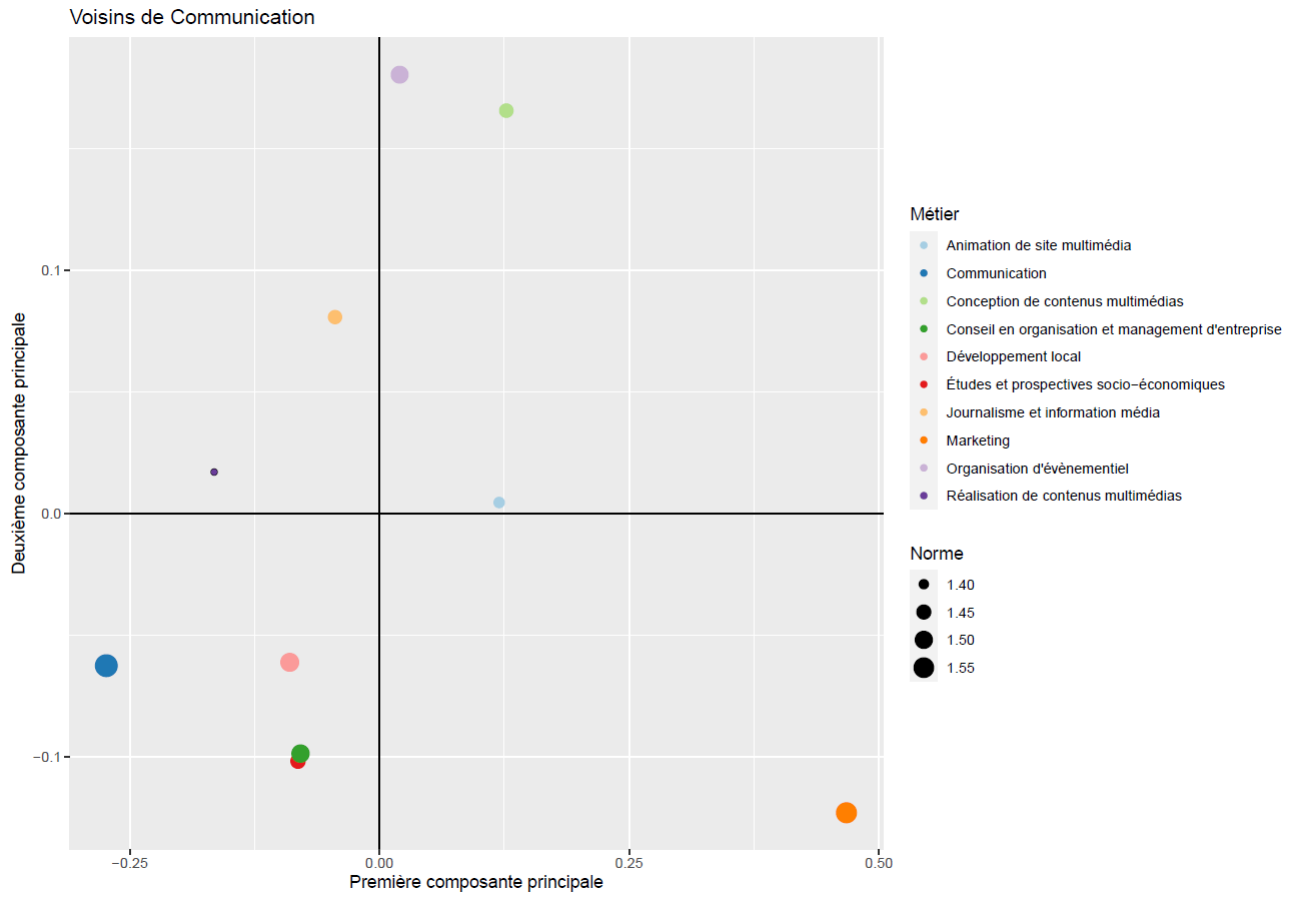


Voisins de Transaction immobilière

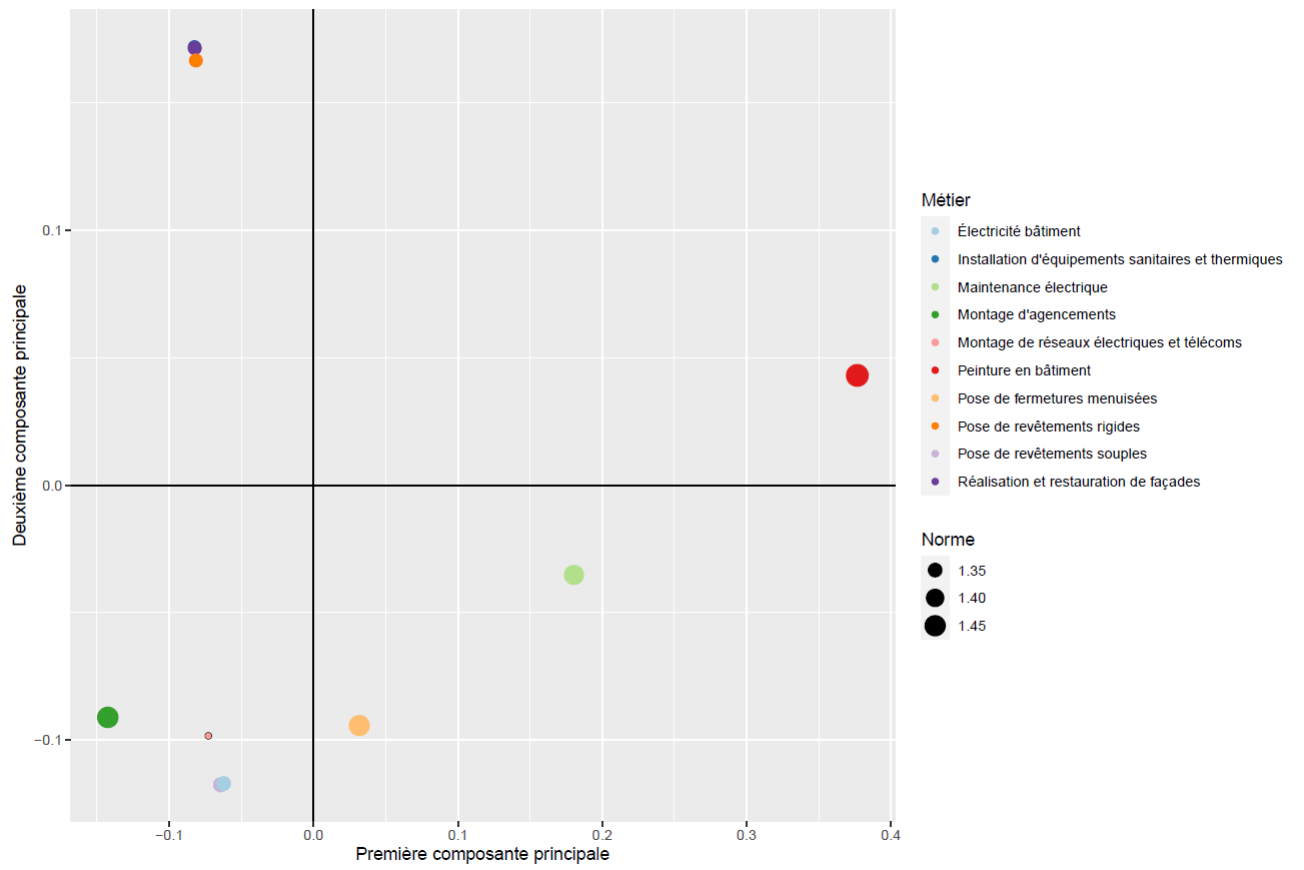


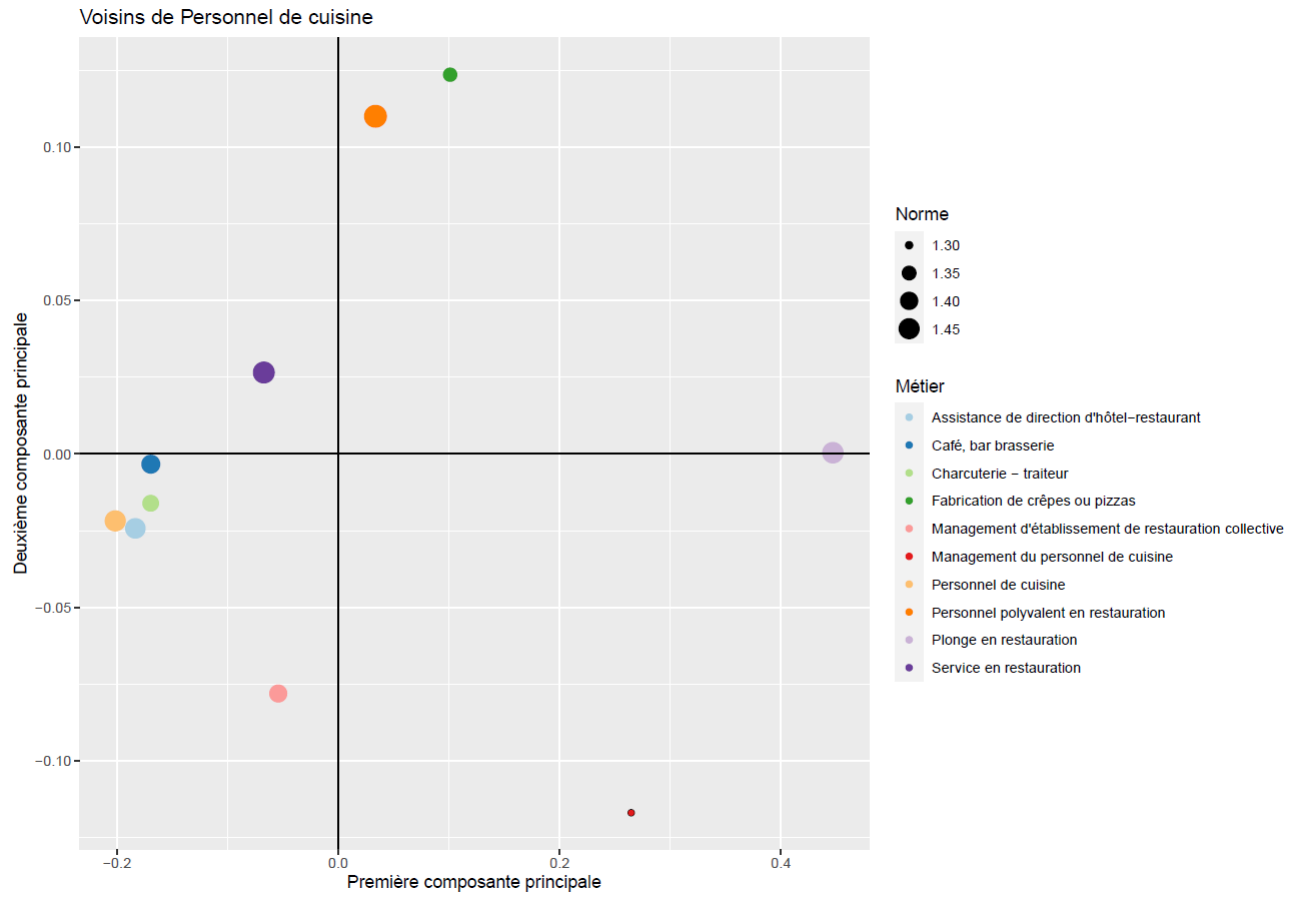
Voisins de Vente en habillement et accessoires de la personne



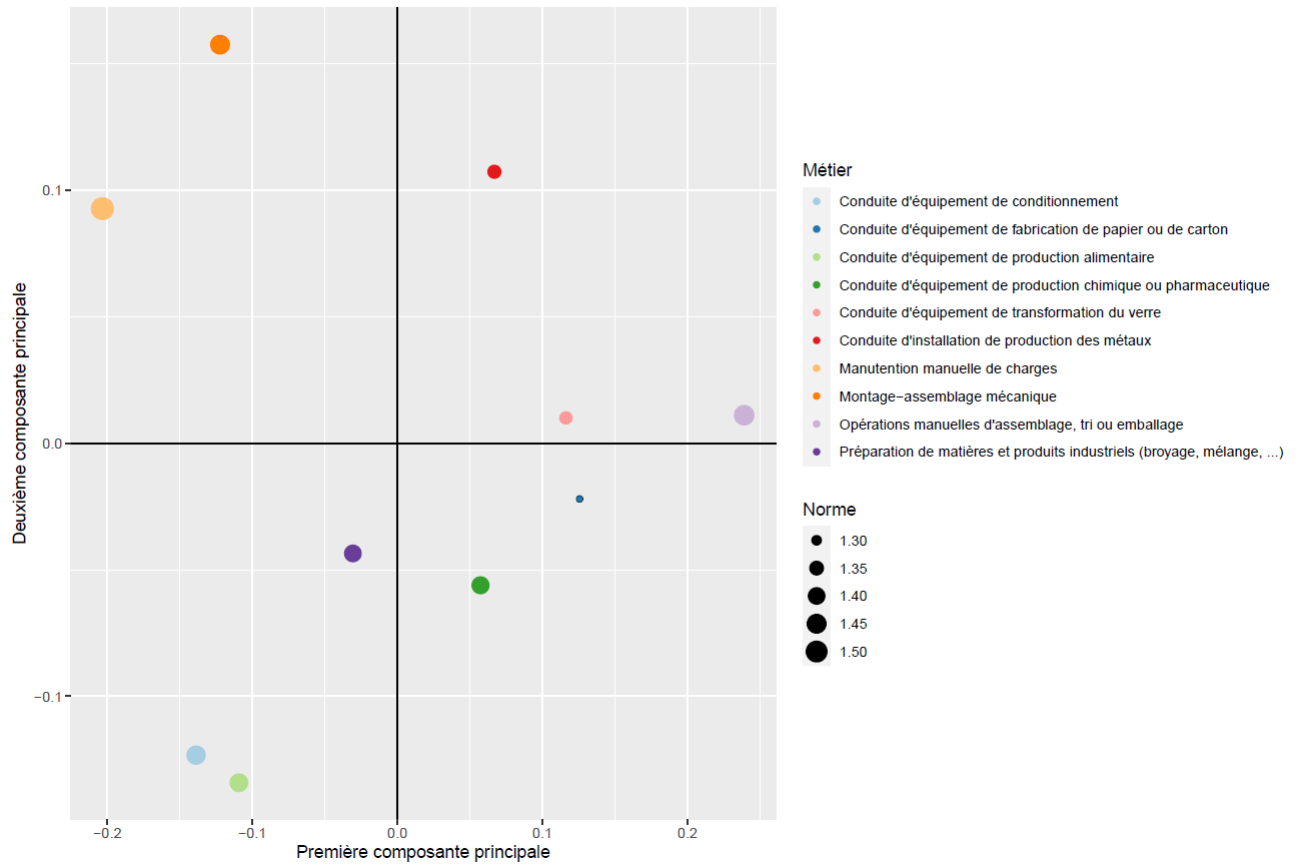


Voisins de Peinture en bâtiment

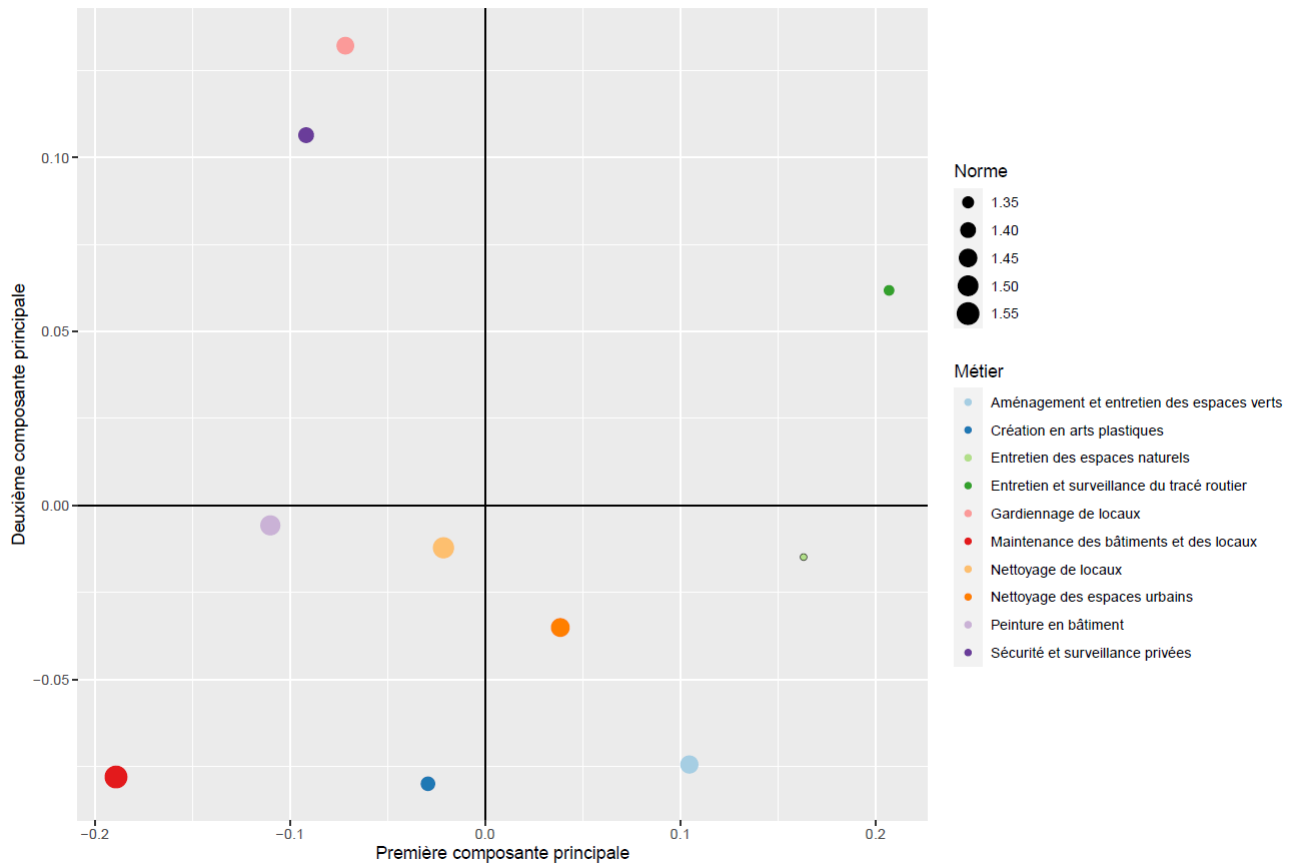




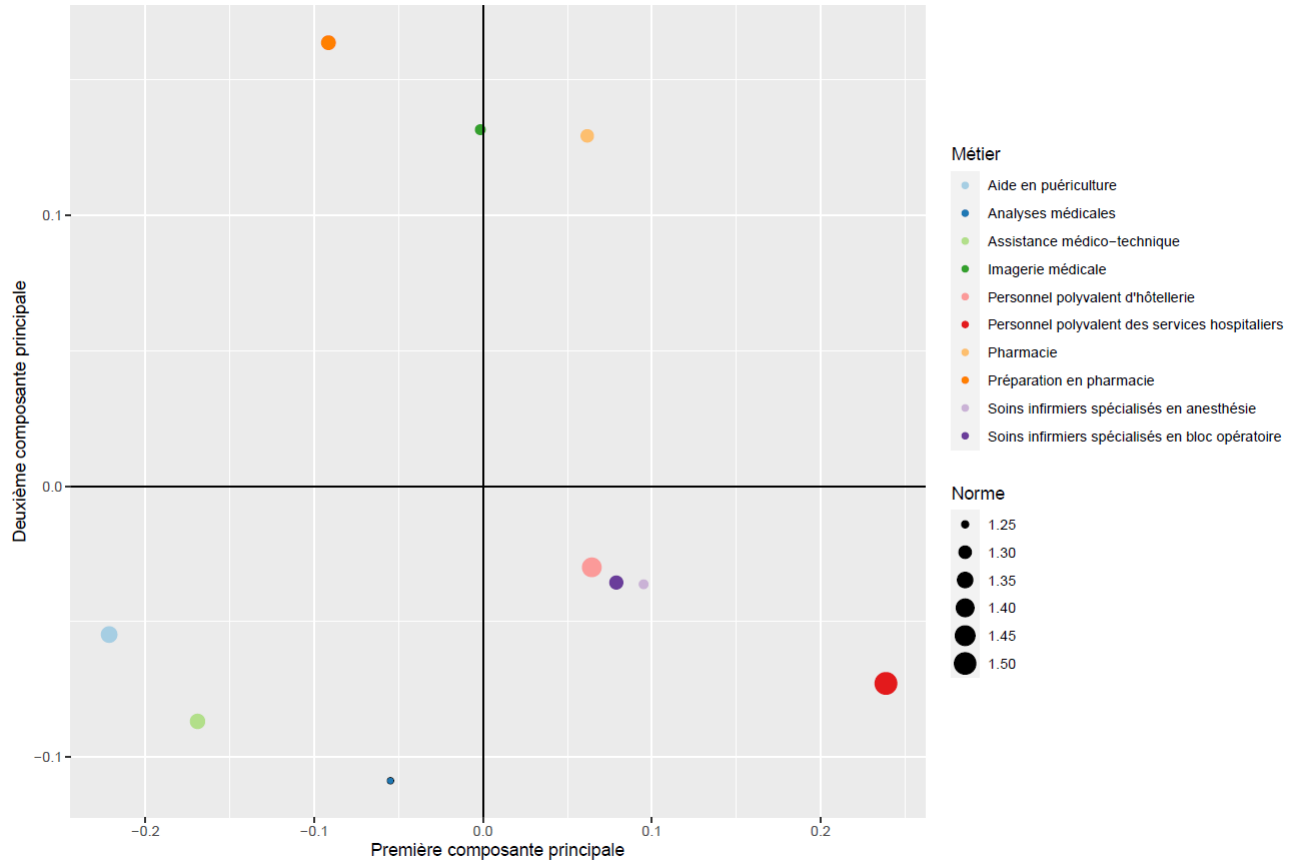
Voisins de Opérations manuelles d'assemblage, tri ou emballage

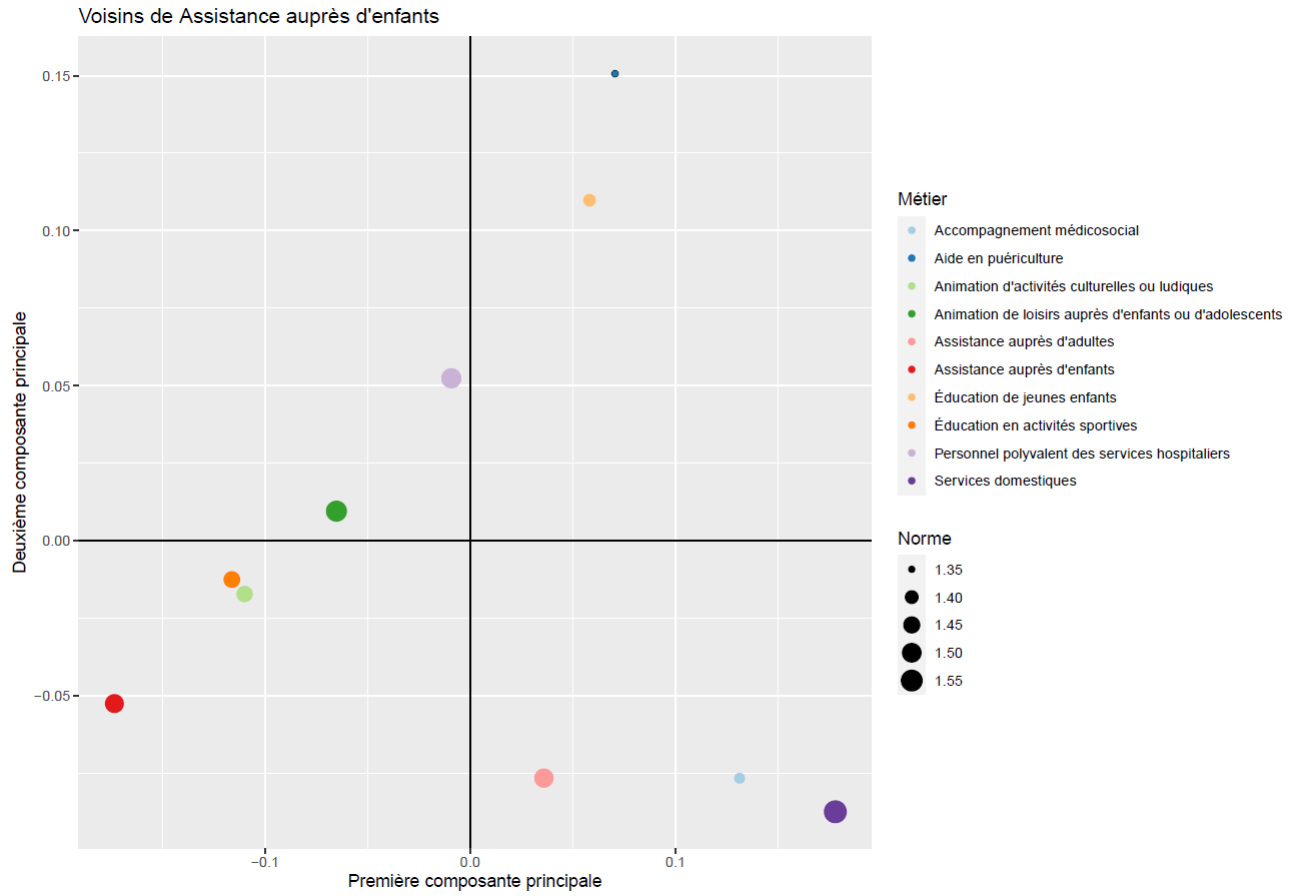


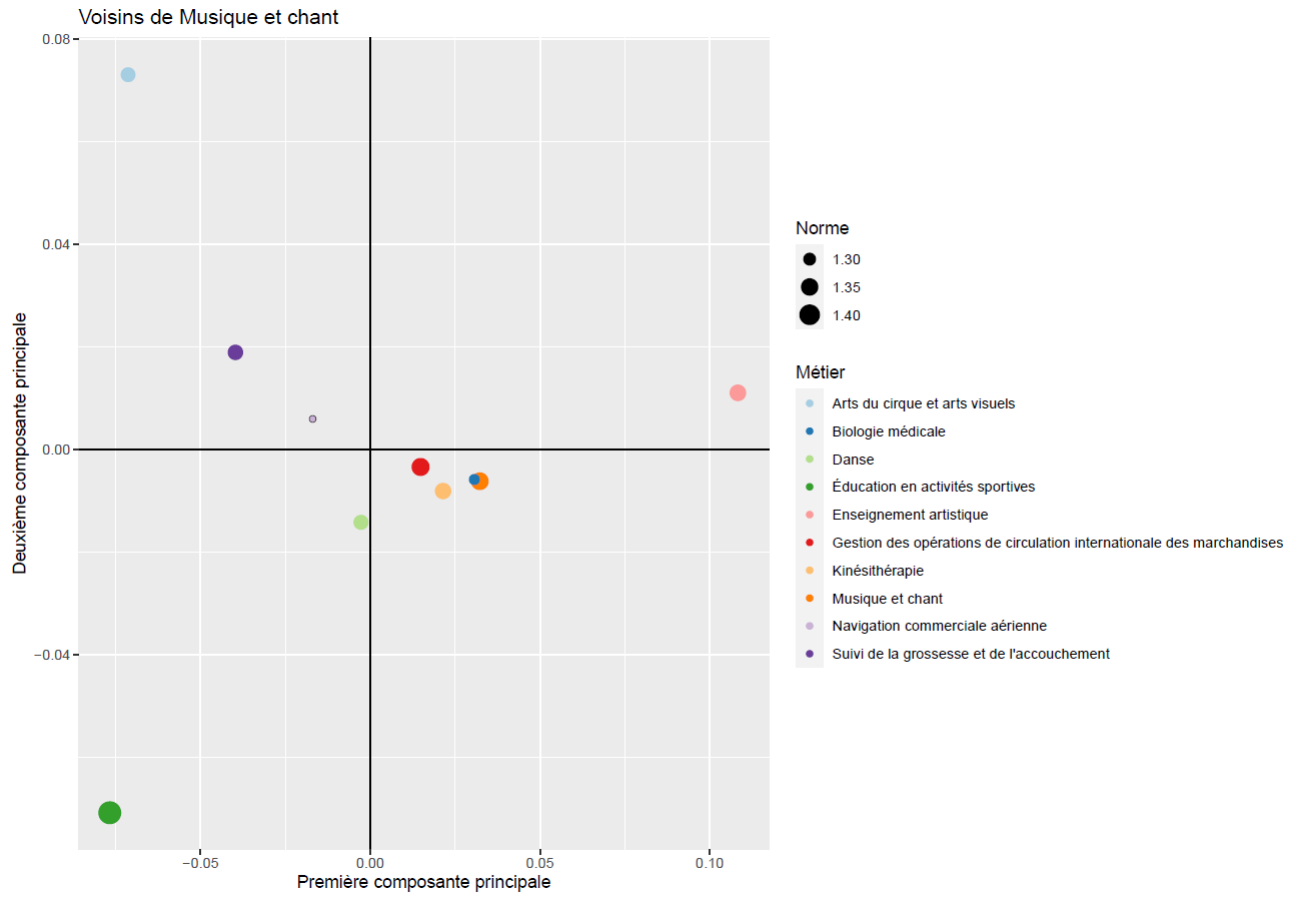
Voisins de Maintenance des bâtiments et des locaux

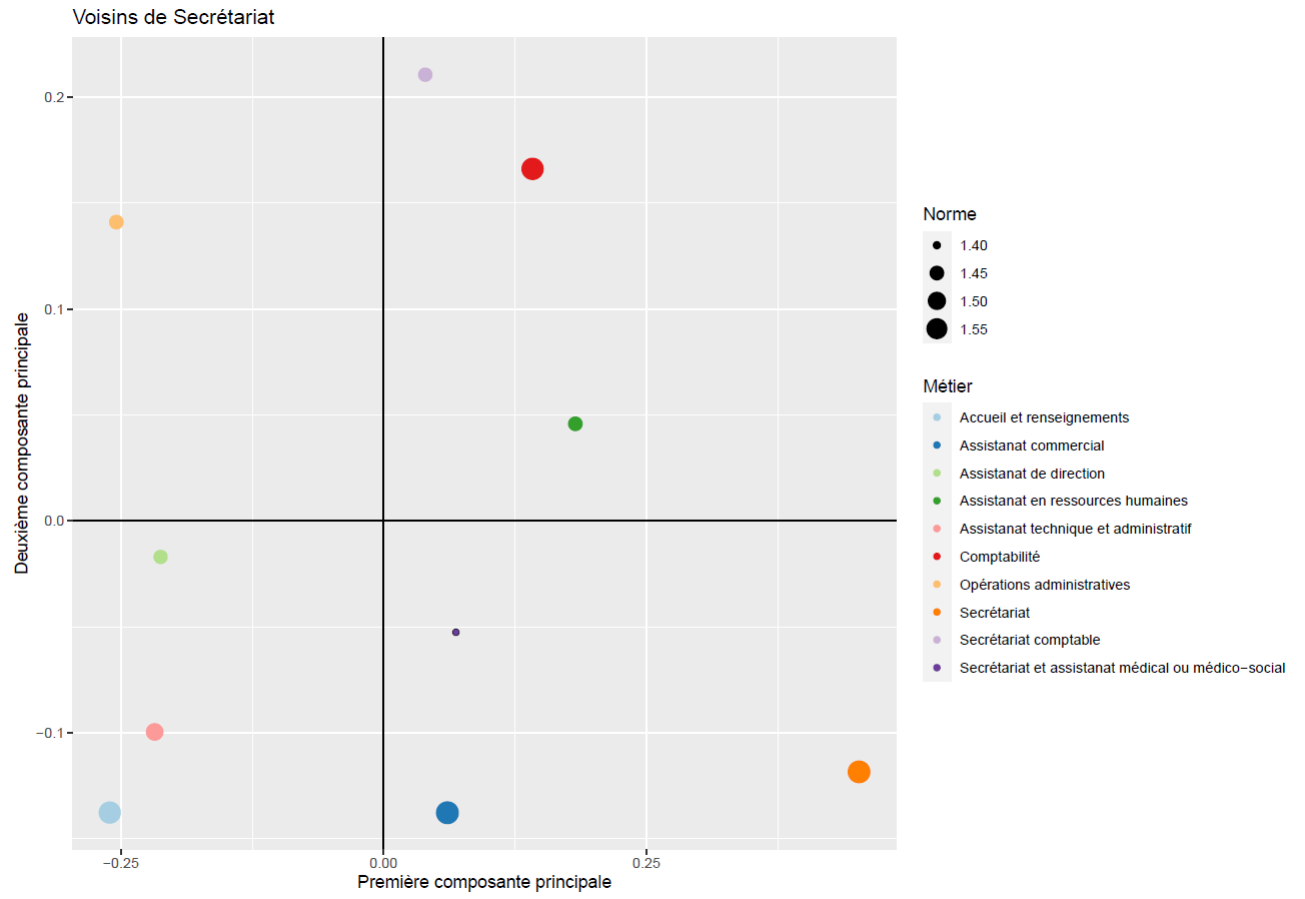


Voisins de Personnel polyvalent des services hospitaliers









Voisins de Magasinage et préparation de commandes

