



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

• VALORISATION DE LA RECHERCHE

DÉCEMBRE 2024
N°10

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

Kevin M. Frick Toulouse School of Economics

Yagan Hazard Collegio Carlo Alberto, Turin

Damien Mayaux Paris School of Economics

Thomas Zuber Banque de France

Les rapports *Valorisation de la recherche* sont des travaux de chercheurs et chercheuses ayant bénéficié d'une contribution financière de la Dares. À ce titre, ils n'engagent que leurs auteurs et autrices, et ne représentent pas la position de la Dares ni celle du ministère du Travail et de l'Emploi.

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

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

3 décembre 2024

Résumé

La formation professionnelle contribue-t-elle à la résorption des déséquilibres structurels du marché du travail ? En utilisant une nouvelle mesure d'écart de compétences entre métiers construite grâce à l'application d'un algorithme de traitement du langage à un échantillon d'offres d'emploi, nous observons que les retours à l'emploi faisant 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 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. Au-delà de la nouvelle mesure de distance inter-métiers que nous proposons, nos résultats empiriques reposent sur une hypothèse d'indépendance conditionnelle du recours à la formation et doivent donc être interprétés avec prudence.

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 ainsi que la participation financière de la Dares dans le cadre de l'appel à projet de recherche "la formation des personnes en recherche d'emploi" (3ème édition), lancé sous l'égide du comité scientifique d'évaluation du PIC. 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.

1 Synthèse

Les déséquilibres structurels d’offre et de demande sur le marché du travail contribuent à augmenter aussi bien le niveau que la persistance du taux chômage. En France, le « mismatch » inter-métiers pourrait ainsi expliquer jusqu’à 15% du niveau de chômage observé. L’un des objectifs du Plan d’investissement dans les compétences (PIC) consiste donc à pallier ce problème de déséquilibres structurels en facilitant la réorientation de la main d’œuvre disponible vers les métiers les plus en tension. En effet, de telles mobilités inter-métiers impliquent souvent l’acquisition de nouvelles compétences, d’où la nécessité de concevoir des parcours de formation professionnelle adaptés aux personnes en recherche d’emploi qui envisagent ces transitions. Cet article cherche à déterminer dans quelle mesure l’effet de la formation professionnelle sur l’emploi passe par des retours à l’emploi (i) associés à une transition professionnelle, (ii) concentrés dans des métiers en plus forte tension que les métiers d’origine des demandeurs d’emploi. Pour ce faire, nous construisons une mesure de distance en compétences inter-métiers en utilisant le contenu textuel des offres d’emploi publiées par Pôle emploi. Nous développons un algorithme de traitement du langage (natural language processing) permettant d’isoler le contenu propre aux compétences dans ce corpus de texte. Ceci nous permet de produire une cartographie des métiers et de leurs compétences dont les prédictions (notamment en matière de transitions professionnelles) sont plus performantes que celles de mesures de distance inter-métiers alternatives. L’appariement de données administratives réalisé par le dispositif Formation, Chômage et Emploi (ForCE) nous permet ensuite de suivre les trajectoires des personnes en recherche d’emploi ayant eu recours à la formation professionnelle entre 2018 et 2020 sur une période de 24 mois. Notre étude se concentre sur un sous-échantillon de demandeurs d’emploi ayant exercé un emploi stable au cours de l’année écoulée pour lesquels la notion de transition professionnelle peut être précisément définie. Nous excluons par ailleurs du champ de notre étude les formations (POEI, POEC, AFPR) qui sont associées à un projet de recrutement préalable et qui concernent souvent un métier en tension. La comparaison des trajectoires professionnelles réalisées par les demandeurs d’emploi formés et non-formés, associée à la présence d’un grand nombre de variables de contrôle présentes dans les données administratives, permettent d’identifier (sous des hypothèses usuelles de sélection sur les observables) l’effet de la formation sur l’emploi, la distance inter-métiers parcourue, et les gains de tension associés aux transitions professionnelles permises par la formation. Ces gains en tension sont mesurés par le différentiel de taux de retour à l’emploi moyen entre le métier d’origine et le métier d’arrivée des demandeurs d’emploi, après correction pour les caractéristiques observables des demandeurs d’emplois présents dans ces métiers. Les

résultats de nos travaux confirment les effets de la formation professionnelle sur le retour à l'emploi des demandeurs d'emploi et montrent, par ailleurs, que ces effets s'expliquent en partie par des transitions professionnelles et des changements en compétences plus importants accomplis par les personnes entreprenant une formation professionnelle. Du point de vue réallocatif l'effet sur le retour à l'emploi de la formation professionnelle ne semble cependant pas être tiré par un retour à l'emploi plus fréquent vers des métiers en plus forte tension que le métier d'origine des demandeurs d'emploi. Ceci laisse penser que, sur l'échantillon de demandeurs d'emploi étudié et pour les formations que nous avons prises en compte, un meilleur ciblage des métiers en tension par l'offre de formation professionnelle à destination des demandeurs d'emploi permettrait d'améliorer les effets réallocatifs de la formation sur le marché du travail.

Introduction

Les déséquilibres structurels d'offre et de demande sur le marché du travail contribuent à augmenter aussi bien le niveau que la persistance du taux chômage. Parmi ces déséquilibres, l'hypothèse du "mismatch", c'est-à-dire de l'inadéquation entre les compétences détenues par les travailleurs et celles demandées par les entreprises, est souvent mise en avant. Cette inadéquation a pour conséquence la coexistence de "poches" résiduelles de chômage et de marchés "tendus" dans lesquels les entreprises cherchant à recruter se heurtent à la rareté de l'offre de travail qualifiée. Si en théorie ces déséquilibres pourraient se résorber grâce à une réorientation de l'offre de travail vers les marchés en tension, en pratique de nombreux facteurs empêchent ces mobilités professionnelles d'avoir lieu.

Les programmes de formation professionnelle ciblant les personnes en recherche d'emploi visent à apporter des éléments de réponse à ce 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, opérant sur l'équilibre du marché du travail via les transitions professionnelles que le recours à la formation peut faciliter. De nombreuses études documentent l'effet de la formation sur le retour à l'emploi (Card et al., 2018), notamment dans le cadre 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).

Cet article cherche à déterminer si, et dans quelle mesure, les personnes en recherche d'emploi ayant suivi une formation professionnelle ont réalisé des transitions professionnelles (i) associées à un changement de compétences plus important que celui qu'elles auraient connu en l'absence

de formation, et (ii) orientées vers des marchés du travail en plus forte tension que ceux dans lesquels elles auraient exercé en l’absence de formation. Pour ce faire, nous construisons une mesure de distance en compétences inter-métiers en utilisant le contenu textuel des offres d’emploi publiées par Pôle emploi. Nous développons à cette fin un algorithme de traitement du langage (*natural language processing*) permettant d’isoler le contenu propre aux compétences dans ce corpus de texte. 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 et qui relient entre eux des métiers dont les compétences sont relativement proches, et des transitions professionnelles plus 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 nôtre¹. L’appariement de données administratives réalisé par le dispositif Formation, Chômage et Emploi (ForCE) nous permet ensuite de suivre les trajectoires des personnes en recherche d’emploi ayant eu recours à la formation professionnelle entre 2018 et 2020. Notre étude se concentre sur un sous-échantillon de demandeurs d’emploi ayant exercé un emploi stable au cours de l’année écoulée pour lesquels les notions de “métier d’origine” et donc de “transition professionnelle” éventuelle sont bien définies. La comparaison des trajectoires professionnelles réalisées par les demandeurs d’emploi formés et non-formés, associée à la présence d’un grand nombre de variables de contrôle présentes dans les données administratives, permet de d’évaluer l’effet de la formation sur l’emploi, la distance inter-métiers parcourue, et les gains de tension associés aux transitions professionnelles permises par la formation. Cette comparaison est réalisée grâce à une méthode de "machine learning avec orthogonalisation en deux étapes" (Chernozhukov et al., 2018) permettant de corriger les différences observables entre le groupe “traité” et le groupe de “contrôle”². Bien entendu le résultat de cette comparaison n’est interprétable comme un effet causal de la formation professionnelle que dans la mesure où l’hypothèse sous-jacente d’indépendance conditionnelle (CIA) du recours à la formation professionnelle est vérifiée – ce qui a peu

1. Voir également Bana et al. (2021) et Gentzkow et al. (2019).

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.

de chance d'être le cas en pratique. Malgré cet écueil bien connu de la littérature sur l'évaluation des politiques de formation, nos résultats permettent d'apporter un premier éclairage sur la question peu étudiée du lien entre la formation professionnelle et la résorption des déséquilibres structurels du marché du travail.

Par rapport à la littérature existante, nos résultats confirment les effets de la formation professionnelle sur le retour à l'emploi qui ont pu être mis en évidence sous des hypothèses comparables d'indépendance conditionnelle du recours à la formation. À bien des égards, cette étude prolonge donc celle de [Chabaud et al. \(2022\)](#). Toutefois, contrairement à [Chabaud et al. \(2022\)](#) qui étudie l'effet du recours à la formation sur le retour à l'emploi, notre 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. Nous montrons ainsi que les effets de la formation sur le retour à l'emploi passent presque entièrement par un retour à l'emploi vers des métiers éloignés des métiers d'origine des personnes formées. Du point de vue réallocatif ces transitions professionnelles plus nombreuses des personnes formées ne semble cependant pas être systématiquement dirigées vers des métiers en plus forte tension que les métiers d'origine. Ceci laisse penser qu'un meilleur ciblage des métiers en tension par l'offre de formation professionnelle à destination des demandeurs d'emploi permettrait d'améliorer les effets réallocatifs de la formation sur le marché du travail. 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 aux demandeurs d'emploi 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. En plus de la validité incertaine attachée à l'hypothèse d'indépendance conditionnelle, cette restriction nécessaire de notre échantillon circonscrit d'autant plus le champ d'interprétation de nos résultats.

Nous contribuons 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 entre différents métiers, plutôt qu'entre secteurs, 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 dis-

tance 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 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'étude 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 puis 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.

2 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 détaillé des compétences, 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, la liste ordonnée des 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 l'accès 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.

2.1 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 associés.

Nous utilisons le texte de ces offres pour pré-entraîner notre modèle de langage, 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 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 principalement sur du "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 de formation des travailleurs dans un métier donné³, ainsi qu'une liste de couples de

3. Pour ce faire nous utilisons le niveau de formation renseigné dans le Fichier historique de Pôle emploi pour les inscrits en 2018. Cette information est tirée de l'appariement ForCE.

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 interprétées comme des mobilités professionnelles “verticales”⁴.

2.2 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-dimensionnalité 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

4. Ces transitions verticales sont construites sur données administratives de l’année 2019 (DADS Postes).

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 tendance à demander plus de compétences que le métier n’en exige vraiment, ou à demander des compétences différentes selon le cycle du marché du travail comme observé dans [Deming and Kahn \(2018\)](#). 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.

2.2.1 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 deux métiers nous renseigne sur la possibilité d’effectuer une transition professionnelle de l’un vers l’autre. 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

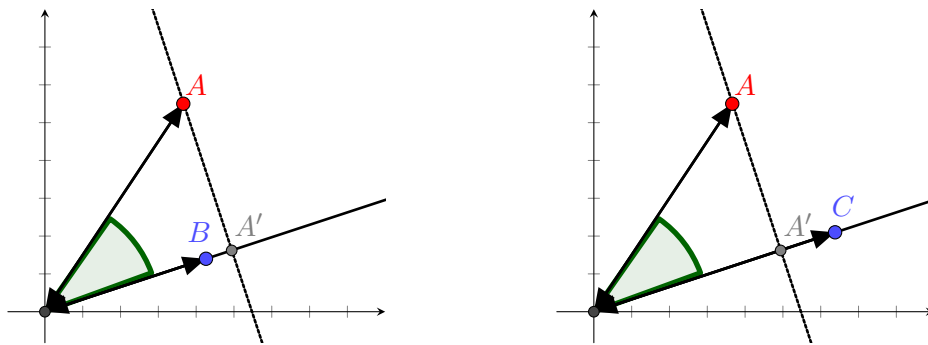
dans les compétences mobilisées par le métier x d'une personne dont les compétences détenues sont représentées par le vecteur u .

- 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.

FIGURE 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 reconvertir 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 reconvertir 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.

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

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 $F(t)$.

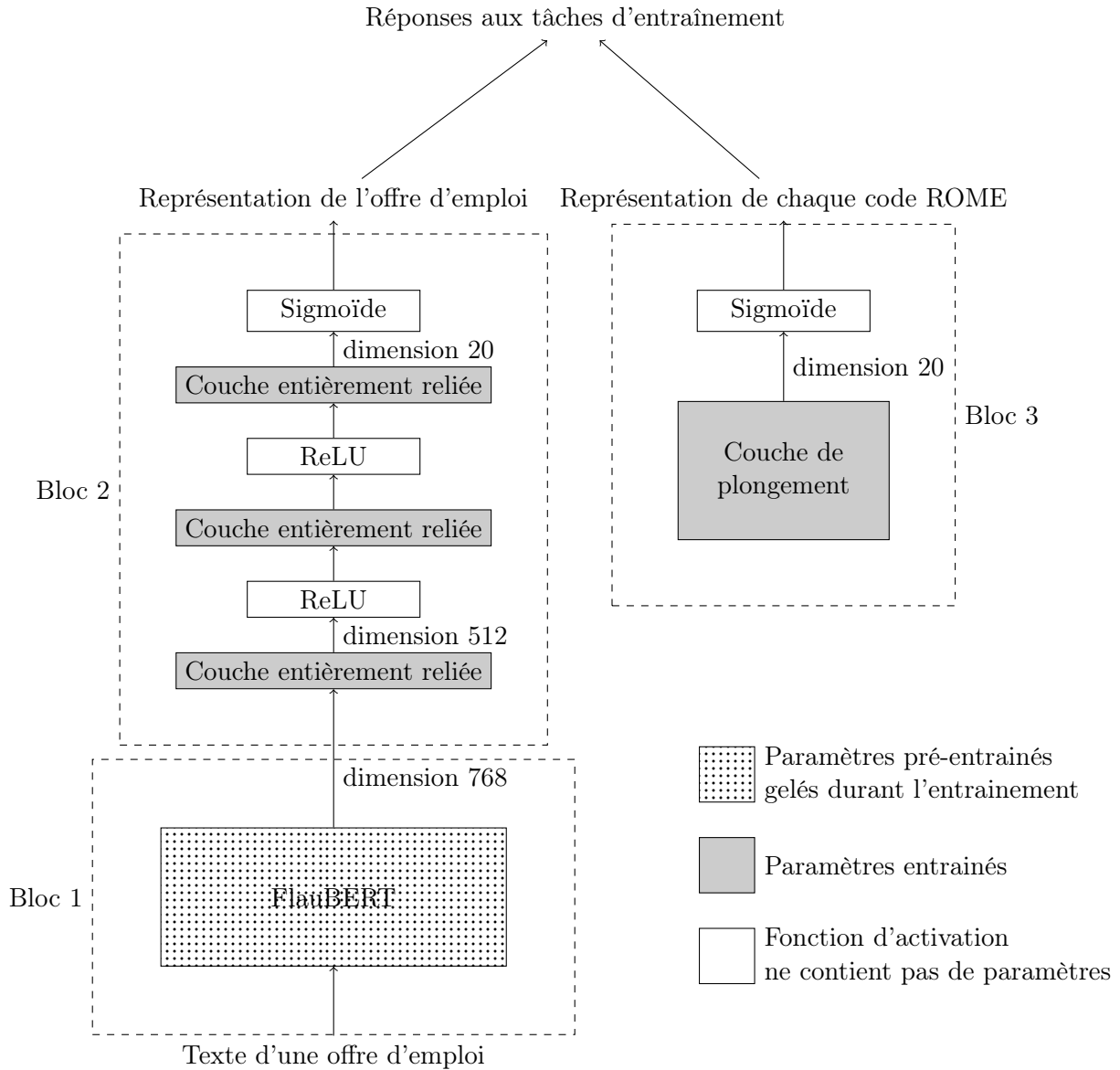
2.2.2 L'architecture du réseau de neurones

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'information pertinente en termes 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.

FIGURE 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.

2.2.3 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 termes 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 :

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

2.3 Résultats

Nous avons validé qualitativement et quantitativement la mesure de distance obtenue à partir de la représentations spatiale des métiers. Nous détaillons dans l'annexe [A.1](#) et dans l'annexe [A.2](#) des premiers éléments sur la performance de l'algorithme sur les tâches prétextes ainsi qu'un exercice de validation qualitative.

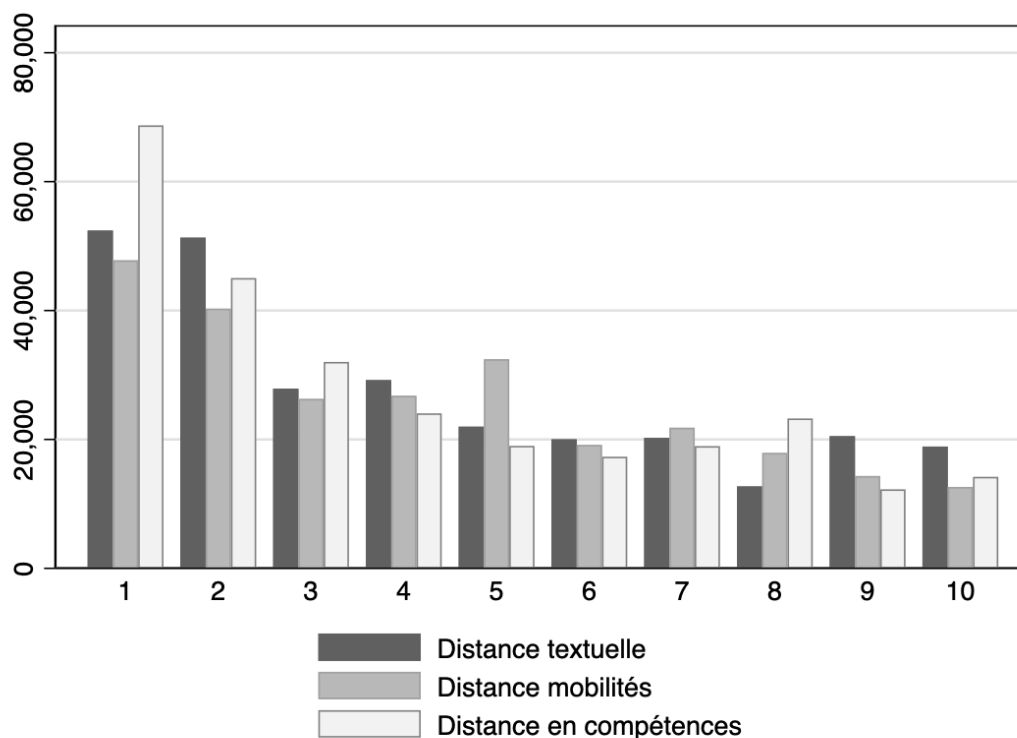
Notre principale démarche de validation quantitative de la mesure obtenue a consisté à évaluer sa capacité à prédire les transitions professionnelles effectivement réalisées sur le marché du travail.

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 transitions 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 3 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 larges de compétence et de généraliser puisque c'est elle qui performe le mieux pour des rangs élevés.

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 termes 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

FIGURE 3 – NOMBRE DE TRANSITIONS EXPLIQUÉES PAR LES MESURES DE DISTANCE INTER-MÉTIER



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. Les transitions utilisées dans ce graphique ne distinguent pas les transitions réalisées avec ou sans formation préalable.

Source : DADS Postes 2019 ; calcul des auteurs.

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.

TABLE 1 – POUVOIR EXPLICATIF DES MESURES DE DISTANCE INTER-MÉTIERES 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.

3 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.

3.1 Données et construction de l’échantillon

Nous utilisons les données du dispositif “Formation, Chômage, Emploi”(FORCE)⁵ qui regroupe notamment :

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 (inscrite 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 Déclaration sociale nominative), qui rassemble depuis 2017 les informations sur les contrats de

5. Millésime 2023T2.

travail pour chaque salarié du privé,⁶ pour l'ensemble des demandeurs d'emploi présents dans le FH du millésime de FORCE considéré ;

Nous construisons notre base d'analyse en suivant les mêmes principes généraux 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. Dans un premier temps, nous créons 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. Afin de concentrer l'analyse sur les primo-formés nous utilisons la table P2 du Fichier Historique pour exclure les demandeurs d'emploi qui ont suivi une formation avant 2017. Dans un second temps cet échantillon est apparié à la base Brest afin de déterminer l'ensemble des demandeurs d'emploi entrant en formation pour la première fois au cours du 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 . Nous excluons du champ des formations considérées les formations qui sont directement associées à un projet de recrutement (POEI, POEC, AFPR)⁸. Nous excluons ainsi du champ de l'étude l'ensemble des demandeurs d'emploi ayant eu recours à l'un de ces trois dispositifs, quelle que soit l'année de ce recours. 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 Fichier historique, telles que leur âge, sexe, situation familiale, nombre d'enfants, niveau de formation, qualification, situation de handicap, habitation dans une zone rural ou urbaine, volonté de mobilité géographique, catégorie d'inscription, motif d'inscription, nationalité, métier recherché,

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.

8. La base BREST permet de distinguer les Préparations Opérationnelles à l'Emploi Individuelles (POEI), les Préparations Opérationnelles à l'Emploi Collectives (POEC) et les Actions de Formation Préalables au Recrutement (AFPR) des Aides Individuelles de Formation (AIF) sur lesquelles nous nous concentrons.

salaire de réserve, ancienneté au chômage au mois m , temps passé au chômage avant l'épisode en question, nombres d'épisodes de chômage avant l'épisode en question. Ces variables de contrôle nous permettent de redresser les différences observables existant entre les demandeurs d'emploi ayant recours à la formation professionnelle et ceux n'y ayant pas recours.

3.2 Stratégie empirique

Notre stratégie d'identification tout au long de cette article 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'estimation associée à cette stratégie d'identification peut cependant se révéler complexe, dans la mesure où la dimensionnalité 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 dimensionnalité de X depuis Rosenbaum and Rubin (1983). La méthode que nous privilégions, dite de "Double Debiased Machine Learning" (DML) (Chernozhukov et al., 2018), repose sur l'utilisation d'estimateurs non paramétriques des outcomes conditionnels et du score de propension qui sont ensuite combinés pour former un estimateur robuste à une mauvaise

spécification de l'un des deux termes⁹. Bien que présentant de meilleures propriétés statistiques que l'estimateur classique d'appariement sur score de propension, le caractère biaisé ou non de l'estimateur DML est tout autant dépendant de la validité de l'hypothèse d'indépendance conditionnelle que nous formulons. Celle-ci étant par nature fragile dans le contexte du recours à la formation professionnelle, nos résultats ne se prêtent pas directement à une interprétation causale.

3.3 Caractéristiques 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. Comme indiqué ci-dessus, nous excluons par ailleurs du champ de l'étude les demandeurs d'emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

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 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.

9. 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>.

TABLE 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 (%) | 55,3 | 53,8 |
| Age (%) | | |
| < 25 | 17,7 | 13,8 |
| [25, 50] | 66,5 | 70,2 |
| > 50 | 10,9 | 11 |
| Études (%) | | |
| < Bac | 45,3 | 38,9 |
| Bac | 27,9 | 27,7 |
| > Bac | 26,8 | 33,5 |
| Durée d’inscription (mois) | 22,6 | 9,3 |
| Formation (%) | 5,9 | 7,8 |
| Formation certifiante (%) | 2,9 | 4,3 |
| Formation > 420h (%) | 3,9 | 5,2 |
| Observations (milliers) | 810 | 222 |

Source : Données FORCE. Ensemble des demandeurs d’emploi inscrits au cours du mois de septembre 2018 à l’exclusion des demandeurs d’emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

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’entretenant 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.

TABLE 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 (%) | 50,5 | 61,5 | 64,5 | 66,3 |
| Age (%) | | | | |
| < 25 | 11,8 | 18,6 | 21,3 | 23,6 |
| [25, 50] | 67,5 | 67,7 | 67,7 | 66,6 |
| > 50 | 11,9 | 8,9 | 5,9 | 4,6 |
| Études (%) | | | | |
| < Bac | 42,1 | 31,2 | 25,3 | 24,3 |
| Bac | 23,2 | 32,2 | 35,7 | 36,6 |
| > Bac | 25,8 | 32,1 | 35,7 | 36,4 |
| Durée d’inscription (mois) | 11,7 | 8,3 | 8,3 | 8,2 |
| Changement de métier désiré (%) | 69,7 | 69,9 | 69,1 | 69,8 |
| Distance si changement | 0,52 | 0,53 | 0,52 | 0,53 |
| Observations (milliers) | 204,9 | 17,3 | 9,6 | 11,5 |

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, à l’exclusion des demandeurs d’emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

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 stable¹¹. 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.

Variables de contrôle utilisées Les stratégies d’identification et d’estimation associées reposent sur la correction des différences observables entre demandeurs d’emploi contrôles et trai-

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.

tés grâce à la méthode « DML »(voir supra). En pratique nous incluons parmi les variables de contrôle utilisées par l’algorithme le niveau de qualification, l’âge, le sexe, le type de contrat recherché, le niveau de formation, l’expérience sur le marché du travail, la résidence dans un quartier prioritaire de la ville, la situation matrimoniale, la catégorie administrative d’enregistrement à Pôle emploi, la nationalité, la résidence dans une zone urbaine sensible, le dernier motif d’inscription à Pôle emploi, le caractère obligatoire ou non des démarches de recherche d’emploi, le niveau du salaire de réserve indiqué, les préférences de mobilité, le métier recherché, la zone d’emploi de résidence, le nombre d’épisodes de chômage distincts ainsi que le temps total d’inscription à Pôle emploi au cours des 10 dernières années.

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

Le tableau 4 synthétise nos résultats sur l’effet de la formation professionnelle sur le retour à l’emploi à différents horizons temporels (3, 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. On distingue ici l’effet de l’ensemble des formations de l’effet des seules formations certifiantes, pour le retour à l’emploi en général puis, dans un second temps, pour le retour à l’emploi dans un emploi stable (CDI ou CDD de plus de 6 mois). 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 dès 6 mois après l’entrée en formation. Cet effet augmente ensuite de manière importante apparaît maximal 12 mois après l’entrée en formation et décroît légèrement à 24 mois tout en restant largement positif. 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 Chabaud et al. (2022) obtiennent 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 proches 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 par ailleurs que si l'effet initial de "lock in" est plus prononcé pour les formations certifiantes (qui sont très souvent des formations longues), l'effet à 12 puis à 24 mois de ce type de formations est plus élevé que l'effet mesuré pour les formations prises dans leur ensemble. Cette conclusion demeure inchangée lorsque l'on s'intéresse au retour à l'emploi dans un emploi stable (CDI ou CDD de plus de 6 mois), auquel cas la montée en charge de l'effet apparaît plus graduelle (effet légèrement inférieur à 12 mois que pour le retour à l'emploi en général) mais plus importante à long terme (effet plus important de l'ensemble des formations et des formations certifiantes sur le retour en emploi stable à 24 mois).

TABLE 4 – EFFETS DE LA FORMATION PROFESSIONNELLE À DIFFÉRENTS HORIZONS

| | (1) | (2) | (3) | (4) |
|-----------------------|--------------------|-------------------|------------------|------------------|
| | 3 mois | 6 mois | 12 mois | 24 mois |
| | Tout type d'emploi | | | |
| Toute formation | -0,026 (0,002) | 0.008 (0,002) | 0.089 (0,003) | 0.062 (0,002) |
| Formation certifiante | -0,042 (0,002) | -0.017 (0,003) | 0.102 (0,004) | 0.082 (0,003) |
| | Emploi stable | | | |
| Toute formation | -0,005 (0,001) | 0,022 (0,002) | 0,082 (0,002) | 0,080 (0,003) |
| Formation certifiante | -0,015 (0,002) | 0,008 (0,002) | 0,086 (0,003) | 0,096 (0,004) |

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. Ainsi le recours à la formation en général et le recours à des formations certifiantes en particulier ont bien un effet réallocatif sur le retour à l'emploi en éloignant les demandeurs d'emploi de leur métier d'origine et les re-dirigeant vers des métiers que l'on peut considérer comme « éloignés » en termes de compétences. 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.

TABLE 5 – EFFETS DE LA FORMATION PROFESSIONNELLE SUR LE RETOUR À L'EMPLOI (À 24 MOIS) DANS DES MÉTIERS PLUS OU MOINS PROCHE 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 | -0,009 (0,002) | 0,007 (0,002) | 0,01 (0,001) | 0,048 (0,001) |
| Formation certifiante | -0,015 (0,002) | 0,017 (0,002) | 0,009 (0,002) | 0,063 (0,003) |

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.

L'effet de la formation professionnelle sur le retour à l'emploi passe-t-il par une redirection accentuée vers les marchés en forte tension ? Pour répondre à cette question le tableau 6 décompose l'effet de la formation selon que l'emploi retrouvé se situe dans un marché en plus faible ou en plus forte tension que le marché de l'emploi d'origine. Les données de tension utilisées pour

réaliser cette décomposition sont obtenues en régressant le retour à l'emploi des demandeurs d'emploi n'ayant pas recours à la formation sur des effets fixes marché (ZE*FAP) en contrôlant pour les caractéristiques individuelles des demandeurs d'emploi présents sur chaque marché. Cette manière de procéder nous permet de nous appuyer sur un indicateur de tension (taux de retour à l'emploi par marché) qui évite les problèmes de mesure habituels liés au caractère non-observable de l'effort de recherche des demandeurs d'emploi et des entreprises¹². Le tableau 6 montre qu'aussi bien dans le cas des formations certifiantes que pour l'ensemble des formations, l'effet à 24 mois sur le retour à l'emploi ne semble pas être porté par des redirections plus nombreuses vers des marchés de l'emploi en plus forte tension que le marché d'origine de chaque 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 que, sur l'échantillon de demandeurs d'emploi étudié et pour les formations que nous avons prises en compte, 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.

TABLE 6 – EFFETS DE LA FORMATION PROFESSIONNELLE SUR LE RETOUR À L'EMPLOI À 24 MOIS DANS DES MARCHÉS EN FAIBLE ET FORTE TENSION

| | (1) plus faible tension | (2) plus forte tension |
|-----------------------|----------------------------|---------------------------|
| Toute formation | 0,034 (0,002) | 0,032 (0,002) |
| Formation certifiante | 0,041 (0,003) | 0,045 (0,003) |

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 \times zone d'emploi sont présentées entre parenthèses.

12. La Dares utilise une mesure comparable (taux d'écoulement de la main d'œuvre par marché) lorsqu'elle construit un indicateur "synthétique" de tension sur le marché du travail.

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. À cette fin nous construisons une mesure originale de distance en compétences entre métiers en nous appuyant sur les textes des offres d'emploi postées à Pôle emploi. Nous utilisons cette mesure de distance pour étudier les transitions professionnelles des demandeurs d'emploi ayant ou non recours à la formation. 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. Nos résultats reposent sur une hypothèse – forte dans ce contexte – d'indépendance conditionnelle du recours à la formation et doivent donc être interprétés avec prudence. D'un point de vue purement réallocatif, l'effet sur le retour à l'emploi de la formation professionnelle ne semble pas être tiré par des redirections plus nombreuses vers des métiers en forte tension. Ce résultat laisse présumer que, sur l'échantillon de demandeurs d'emploi étudié et pour les formations que nous avons prises en compte, 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

- Baley, I., A. Figueiredo, and R. Ulbricht (2020). Mismatch cycles.
- Bana, S. H., E. Brynjolfsson, D. Rock, and S. Steffen (2021). job2vec : Learning a representation of jobs.
- Barnichon, R. and A. Figura (2015). Labor market heterogeneity and the aggregate matching function. *American Economic Journal : Macroeconomics* 7(4), 222–49.
- Becker, G. S. (1964, June). *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.
- Card, D., J. Kluve, and A. Weber (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association* 16(3), 894–931.
- Chabaud, M., A. Bucher, P. Givord, and A. Louvet (2022). Quelles sont les chances de retour à l’emploi après une formation? *Document d’études* 261.
- Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*.
- Dawson, N., M.-A. Williams, and M.-A. Rizoiu (2021). Skill-driven recommendations for job transition pathways. *PLOS One*.
- Deming, D. and L. B. Kahn (2018, January). Skill Requirements across Firms and Labor Markets : Evidence from Job Postings for Professionals. *Journal of Labor Economics* 36(S1), S337–S369. Publisher : The University of Chicago Press.
- Eckardt, D. (2022). Are chemists good bankers? returns to the match between training and occupation. Technical report, Working Paper.
- Fontaine, F. and R. Rathelot (2022). Le marché du travail français à l’épreuve de la crise sanitaire. *Notes du conseil danalyse économique* 71(2), 1–12.
- Gathmann, C. and U. Schönberg (2010). How general is human capital? a task-based approach. *Journal of Labor Economics* 28(1), 1–49.
- Gentzkow, M., B. Kelly, and M. Taddy (2019). Text as Data. *Journal of Economic Literature* 57(3), 535–574.

- Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020). Multidimensional skill mismatch. *American Economic Journal : Macroeconomics* 12(1), 210–44.
- Hyman, B. G. (2018). 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, pp. 1–70. JSTOR.
- Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American economic review*, 685–709.
- Le, H., L. Vial, J. Frej, V. Segonne, M. Coavoux, B. Lecouteux, A. Allauzen, B. Crabbé, L. Besacier, and D. Schwab (2019). Flaubert : Unsupervised language model pre-training for french. *arXiv preprint arXiv :1912.05372*.
- Lindenlaub, I. and F. Postel-Vinay (2021). The worker-job surplus. Technical report, National Bureau of Economic Research.
- Marinescu, I. and R. Rathelot (2018). Mismatch unemployment and the geography of job search. *American Economic Journal : Macroeconomics*.
- Mestieri, M., S. Basco, G. Smagghue, and M. Liegey (2020). The heterogeneous effects of trade across occupations : A test of the stolper-samuelson theorem.
- Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41–55.
- Şahin, A., J. Song, G. Topa, and G. L. Violante (2014). Mismatch unemployment. *American Economic Review* 104(11), 3529–64.
- Shaw, K. L. (1984). A formulation of the earnings function using the concept of occupational investment. *Journal of Human Resources*, 319–340.
- Traiberman, S. (2019). Occupations and import competition : Evidence from denmark. *American Economic Review* 109(12), 4260–4301.
- Weston, J., S. Bengio, and N. Usunier (2011). Wsabie : Scaling Up To Large Vocabulary Image Annotation. In *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI*.

A Autres approches pour l'évaluation du réseau de neurones

A.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 donnée (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.

A.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 avons déterminés les 5 métiers les plus proches selon notre mesure, ainsi que ceux parmi eux vers lesquels aucune mobilité n'est suggérée dans la classification ROME (voir [B](#) et [C](#)). 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).

B Notre distance appliquée aux mobilités présentes dans ROME

V3

B.1 Depuis Aménagement et entretien des espaces verts, code A1203

- Entretien des espaces naturels, $d = 0.248$, rang 2
- Horticulture et maraîchage, $d = 0.440$, rang 4
- Sylviculture, $d = 0.467$, rang 7
- Vente de végétaux, $d = 0.656$, rang 80
- Arboriculture et viticulture, $d = 0.683$, rang 119
- Formation professionnelle, $d = 0.697$, rang 143
- Contrôle et diagnostic technique en agriculture, $d = 0.715$, rang 189

B.2 Depuis Réalisation de vêtements sur mesure ou en petite série, code B1803

- Retouches en habillement, $d = 0.448$, rang 4
- Stylisme, $d = 0.571$, rang 16
- Patronnage - gradation, $d = 0.574$, rang 18
- Études - modèles en industrie des matériaux souples, $d = 0.575$, rang 19
- Montage de prototype cuir et matériaux souples, $d = 0.636$, rang 41
- Costume et habillage spectacle, $d = 0.666$, rang 79
- Vente en décoration et équipement du foyer, $d = 0.692$, rang 120
- Vente en habillement et accessoires de la personne, $d = 0.698$, rang 129
- Achat vente d'objets d'art, anciens ou d'occasion, $d = 0.725$, rang 191
- Relation commerciale grands comptes et entreprises, $d = 0.782$, rang 309
- Formation professionnelle, $d = 0.795$, rang 344
- Management de magasin de détail, $d = 0.815$, rang 382
- Animation d'activités culturelles ou ludiques, $d = 0.826$, rang 408
- Encadrement technique en insertion professionnelle, $d = 0.872$, rang 478
- Enseignement technique et professionnel, $d = 0.889$, rang 495

B.3 Depuis Transaction immobilière, code C1504

- Gestion locative immobilière, $d = 0.445$, rang 2
- Relation commerciale auprès de particuliers, $d = 0.494$, rang 4
- Gérance immobilière, $d = 0.514$, rang 7
- Management de projet immobilier, $d = 0.561$, rang 15
- Relation commerciale grands comptes et entreprises, $d = 0.588$, rang 25
- Gestion de clientèle bancaire, $d = 0.598$, rang 30

B.4 Depuis Vente en habillement et accessoires de la personne, code D1214

- Vente en gros de matériel et équipement, $d = 0.264$, rang 3
- Achat vente d'objets d'art, anciens ou d'occasion, $d = 0.280$, rang 4
- Vente en articles de sport et loisirs, $d = 0.464$, rang 6
- Vente en décoration et équipement du foyer, $d = 0.465$, rang 7
- Optique - lunetterie, $d = 0.507$, rang 10
- Personnel de caisse, $d = 0.508$, rang 11
- Management de magasin de détail, $d = 0.519$, rang 13
- Animation de vente, $d = 0.547$, rang 21
- Management/gestion de rayon produits non alimentaires, $d = 0.585$, rang 27
- Relation commerciale auprès de particuliers, $d = 0.610$, rang 32
- Relation commerciale grands comptes et entreprises, $d = 0.679$, rang 109
- Décoration d'espaces de vente et d'exposition, $d = 0.722$, rang 200

B.5 Depuis Communication, code E1103

- Journalisme et information média, $d = 0.465$, rang 6
- Développement et promotion publicitaire, $d = 0.613$, rang 22
- Élaboration de plan média, $d = 0.638$, rang 43
- Promotion d'artistes et de spectacles, $d = 0.743$, rang 181

B.6 Depuis Peinture en bâtiment, code F1606

- Montage d'agencements, $d = 0.263$, rang 2
- Réalisation et restauration de façades, $d = 0.284$, rang 5
- Pose de revêtements rigides, $d = 0.286$, rang 6

- Pose de revêtements souples, $d = 0.288$, rang 7
- Maintenance des bâtiments et des locaux, $d = 0.497$, rang 13
- Application et décoration en plâtre, stuc et staff, $d = 0.522$, rang 17
- Entretien d'affichage et mobilier urbain, $d = 0.538$, rang 19
- Direction de chantier du BTP, $d = 0.569$, rang 28
- Vente en décoration et équipement du foyer, $d = 0.661$, rang 118
- Formation professionnelle, $d = 0.701$, rang 195
- Tapisserie - décoration en ameublement, $d = 0.825$, rang 480

B.7 Depuis Personnel de cuisine, code G1602

- Management du personnel de cuisine, $d = 0.229$, rang 3
- Management d'établissement de restauration collective, $d = 0.467$, rang 6
- Charcuterie - traiteur, $d = 0.535$, rang 8
- Café, bar brasserie, $d = 0.536$, rang 9
- Assistance de direction d'hôtel-restaurant, $d = 0.545$, rang 10
- Management d'hôtel-restaurant, $d = 0.567$, rang 15
- Pâtisserie, confiserie, chocolaterie et glacerie, $d = 0.575$, rang 17
- Formation professionnelle, $d = 0.701$, rang 157
- Gestion de structure de loisirs ou d'hébergement touristique, $d = 0.717$, rang 190
- Encadrement technique en insertion professionnelle, $d = 0.733$, rang 228
- Organisation d'évènementiel, $d = 0.831$, rang 447

B.8 Depuis Opérations manuelles d'assemblage, tri ou emballage, code H3302

- Manutention manuelle de charges, $d = 0.319$, rang 2
- Montage-assemblage mécanique, $d = 0.335$, rang 3
- Conduite d'équipement de conditionnement, $d = 0.335$, rang 4
- Préparation de matières et produits industriels (broyage, mélange, ...), $d = 0.393$, rang 6
- Magasinage et préparation de commandes, $d = 0.501$, rang 13
- Assemblage d'ouvrages en bois, $d = 0.504$, rang 15
- Montage de produits électriques et électroniques, $d = 0.539$, rang 24
- Préfabrication en béton industriel, $d = 0.545$, rang 25
- Revalorisation de produits industriels, $d = 0.546$, rang 26

- Aide agricole de production légumière ou végétale, $d = 0.586$, rang 40
- Aide agricole de production fruitière ou viticole, $d = 0.621$, rang 52
- Conduite de machines d'impression, $d = 0.625$, rang 57
- Façonnage et routage, $d = 0.629$, rang 59
- Maintenance mécanique industrielle, $d = 0.633$, rang 63
- Blanchisserie industrielle, $d = 0.648$, rang 77
- Contrôle en industrie du cuir et du textile, $d = 0.652$, rang 81
- Encadrement d'équipe en industrie de transformation, $d = 0.654$, rang 84
- Décoration d'objets d'art et artisanaux, $d = 0.660$, rang 91
- Mise en forme, repassage et finitions en industrie textile, $d = 0.674$, rang 122
- Aide d'élevage agricole et aquacole, $d = 0.677$, rang 129
- Préparation de fils, montage de métiers textiles, $d = 0.682$, rang 136
- Préparation et finition d'articles en cuir et matériaux souples, $d = 0.731$, rang 242

B.9 Depuis Maintenance des bâtiments et des locaux, code I1203

- Aménagement et entretien des espaces verts, $d = 0.453$, rang 4
- Gardiennage de locaux, $d = 0.486$, rang 6
- Nettoyage de locaux, $d = 0.491$, rang 7
- Peinture en bâtiment, $d = 0.497$, rang 8
- Maintenance électrique, $d = 0.537$, rang 14
- Installation d'équipements sanitaires et thermiques, $d = 0.566$, rang 24
- Entretien d'affichage et mobilier urbain, $d = 0.592$, rang 37
- Électricité bâtiment, $d = 0.612$, rang 53

B.10 Depuis Personnel polyvalent des services hospitaliers, code J1301

- Personnel polyvalent d'hôtellerie, $d = 0.436$, rang 8
- Assistance auprès d'enfants, $d = 0.476$, rang 11
- Soins d'hygiène, de confort du patient, $d = 0.500$, rang 14
- Services domestiques, $d = 0.525$, rang 16
- Nettoyage de locaux, $d = 0.533$, rang 18
- Conduite de véhicules sanitaires, $d = 0.645$, rang 99
- Conduite d'opérations funéraires, $d = 0.723$, rang 258
- Hydrothérapie, $d = 0.753$, rang 354

B.11 Depuis Assistance auprès d'enfants, code K1303

- Aide en puériculture, $d = 0.403$, rang 2
- Services domestiques, $d = 0.410$, rang 4
- Éducation de jeunes enfants, $d = 0.416$, rang 5
- Assistance auprès d'adultes, $d = 0.472$, rang 6
- Personnel polyvalent des services hospitaliers, $d = 0.476$, rang 7
- Animation de loisirs auprès d'enfants ou d'adolescents, $d = 0.484$, rang 8
- Soins d'hygiène, de confort du patient, $d = 0.577$, rang 18
- Assistance médico-technique, $d = 0.607$, rang 32
- Personnel polyvalent d'hôtellerie, $d = 0.623$, rang 40

B.12 Depuis Musique et chant, code L1202

- Enseignement artistique, $d = 0.462$, rang 2
- Animation d'activités culturelles ou ludiques, $d = 0.596$, rang 14
- Art dramatique, $d = 0.616$, rang 19
- Promotion d'artistes et de spectacles, $d = 0.641$, rang 32
- Enseignement des écoles, $d = 0.647$, rang 37
- Direction d'établissement et d'enseignement, $d = 0.656$, rang 45
- Enseignement général du second degré, $d = 0.657$, rang 46
- Animation musicale et scénique, $d = 0.733$, rang 122

B.13 Depuis Secrétariat, code M1607

- Assistanat de direction, $d = 0.238$, rang 3
- Assistanat technique et administratif, $d = 0.270$, rang 4
- Secrétariat comptable, $d = 0.352$, rang 6
- Secrétariat et assistanat médical ou médico-social, $d = 0.394$, rang 7
- Assistanat commercial, $d = 0.412$, rang 8
- Collaboration juridique, $d = 0.508$, rang 12
- Gestion locative immobilière, $d = 0.557$, rang 14
- Formation professionnelle, $d = 0.675$, rang 113

B.14 Depuis Magasinage et préparation de commandes, code N1103

- Intervention technique d'exploitation logistique, $d = 0.356$, rang 2
- Conduite d'engins de déplacement des charges, $d = 0.361$, rang 3
- Manutention manuelle de charges, $d = 0.378$, rang 4
- Vente en gros de matériel et équipement, $d = 0.494$, rang 6
- Opérations manuelles d'assemblage, tri ou emballage, $d = 0.501$, rang 7
- Préparation de matières et produits industriels (broyage, mélange, ...), $d = 0.534$, rang 11
- Conduite et livraison par tournées sur courte distance, $d = 0.540$, rang 14
- Blanchisserie industrielle, $d = 0.555$, rang 16
- Conduite d'équipement de conditionnement, $d = 0.601$, rang 28
- Mise en rayon libre-service, $d = 0.601$, rang 29
- Façonnage et routage, $d = 0.683$, rang 111
- Exploitation des pistes aéroportuaires, $d = 0.887$, rang 496

C Mobilités proches selon notre distance mais absentes de ROME

V3

C.1 Depuis Aménagement et entretien des espaces verts, code A1203

- Entretien et surveillance du tracé routier, $d = 0.440$, rang 2
- Maintenance des bâtiments et des locaux, $d = 0.453$, rang 4
- Création en arts plastiques, $d = 0.461$, rang 5

C.2 Depuis Réalisation de vêtements sur mesure ou en petite série, code B1803

- Assemblage - montage de vêtements et produits textiles, $d = 0.251$, rang 1
- Tapisserie - décoration en ameublement, $d = 0.352$, rang 2
- Coupe cuir, textile et matériaux souples, $d = 0.473$, rang 4
- Assemblage - montage d'articles en cuirs, peaux, $d = 0.478$, rang 5

C.3 Depuis Transaction immobilière, code C1504

- Téléconseil et télévente, $d = 0.472$, rang 2

- Assistanat commercial, $d = 0.504$, rang 4
- Assistance et support technique client, $d = 0.511$, rang 5

C.4 Depuis Vente en habillement et accessoires de la personne, code D1214

- Nettoyage d'articles textiles ou cuirs, $d = 0.056$, rang 1
- Coiffure, $d = 0.446$, rang 4

C.5 Depuis Communication, code E1103

- Marketing, $d = 0.270$, rang 1
- Conception de contenus multimédias, $d = 0.395$, rang 2
- Animation de site multimédia, $d = 0.419$, rang 3
- Organisation d'évènementiel, $d = 0.439$, rang 4

C.6 Depuis Peinture en bâtiment, code F1606

- Montage de réseaux électriques et télécoms, $d = 0.267$, rang 2
- Installation d'équipements sanitaires et thermiques, $d = 0.281$, rang 3

C.7 Depuis Personnel de cuisine, code G1602

- Plonge en restauration, $d = 0.012$, rang 1
- Fabrication de crêpes ou pizzas, $d = 0.362$, rang 3
- Personnel polyvalent en restauration, $d = 0.420$, rang 4

C.8 Depuis Opérations manuelles d'assemblage, tri ou emballage, code H3302

- Conduite d'équipement de production alimentaire, $d = 0.353$, rang 4

C.9 Depuis Maintenance des bâtiments et des locaux, code I1203

- Entretien et surveillance du tracé routier, $d = 0.382$, rang 1
- Entretien des espaces naturels, $d = 0.414$, rang 2
- Nettoyage des espaces urbains, $d = 0.486$, rang 4

C.10 Depuis Personnel polyvalent des services hospitaliers, code J1301

- Aide en puériculture, $d = 0.319$, rang 1
- Assistance médico-technique, $d = 0.343$, rang 2
- Préparation en pharmacie, $d = 0.350$, rang 3
- Imagerie médicale, $d = 0.394$, rang 4
- Analyses médicales, $d = 0.413$, rang 5

C.11 Depuis Assistance auprès d'enfants, code K1303

- Accompagnement médicosocial, $d = 0.409$, rang 2

C.12 Depuis Musique et chant, code L1202

- Biologie médicale, $d = 0.512$, rang 2
- Kinésithérapie, $d = 0.519$, rang 3
- Gestion des opérations de circulation internationale des marchandises, $d = 0.523$, rang 4
- Danse, $d = 0.534$, rang 5

C.13 Depuis Secrétariat, code M1607

- Opérations administratives, $d = 0.184$, rang 1
- Accueil et renseignements, $d = 0.278$, rang 4

C.14 Depuis Magasinage et préparation de commandes, code N1103

- Intervention technique d'exploitation des transports routiers de marchandises, $d = 0.493$, rang 4