



HAL
open science

Construction et usages de typologies de trajectoires dans une perspective comparative

Zora Mazari, Alexie Robert

► To cite this version:

Zora Mazari, Alexie Robert. Construction et usages de typologies de trajectoires dans une perspective comparative. 39, Centre d'études et de recherches sur les qualifications (Céreq). 2023, 28 p. halshs-03979217

HAL Id: halshs-03979217

<https://shs.hal.science/halshs-03979217v1>

Submitted on 8 Feb 2023

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

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



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

Céreq ÉTUDES

39
2023

Construction et usages
de typologies de
trajectoires dans une
perspective comparative

Le cas des enquêtes Génération
1998, 2004 et 2010

Zora MAZARI, Alexie ROBERT

Synthèse

Les sources de données longitudinales basées sur des calendriers se sont multipliées ces dernières années, et avec elles l'ambition d'appréhender les trajectoires des individus dans leur globalité plutôt que par le biais d'indicateurs se restreignant à une description de moments précis des parcours individuels. Les typologies de trajectoires sont un outil précieux pour traiter les données longitudinales de calendriers complets : elles permettent d'en réduire la complexité en les synthétisant en principaux « parcours-types ». Cette méthode peut s'appliquer à des thèmes de recherches et données variés en sciences sociales et permet la comparaison entre plusieurs populations. Cette étude présente une méthodologie statistique pour élaborer des typologies d'itinéraires individuels retraçant la dynamique du parcours d'une population donnée. Il propose également une méthode permettant de produire une typologie comparable lorsque l'objet de l'étude nécessite de comparer des individus issus d'échantillons indépendants. Ces techniques seront mises en œuvre sur des données de plusieurs cohortes de Générations du Céreq, chacune permettant le suivi mensuel sur les sept premières années de vie active de jeunes sortis du système de formation initiale français, respectivement en 1998, 2004 et 2010. La typologie de trajectoires professionnelles construite¹ met en évidence l'importance de la dimension temporelle dans la construction des parcours d'insertion sur le marché du travail. Elle offre une aide à la lecture de la diversité des cheminements observés, facilitée par le biais de représentations graphiques adaptées.

¹ La mise en œuvre a été réalisée avec le logiciel SAS et le programme utilisé est téléchargeable au lien suivant : https://github.com/AlexieRobert/typologie_comparable

Sommaire

Synthèse	1
Introduction	4
1. Méthodologie de construction d'une typologie de trajectoires dans une enquête longitudinale	5
1.1. Présentation des données de l'enquête 2017 auprès de la Génération 2010	5
1.2. Constitution de la base d'analyse à partir d'un calendrier d'activité mensuel	7
1.3. Analyse des correspondances multiples	9
1.4. Méthode de classification	10
1.5. Mise en application	12
2. Construction d'une typologie de trajectoires comparables entre plusieurs enquêtes Génération	15
2.1. Constitution d'un jeu de données comparable à partir des enquêtes Génération 1998, 2004 et 2010	16
2.2. Une classification pour trois générations.....	18
2.3. Exemples d'analyses des données à partir de la typologie comparable.....	19
Conclusion	24
Bibliographie	25

Introduction

Les méthodes typologiques (ou l'analyse typologique) sont fréquemment utilisées en complément ou en alternative aux modélisations économétriques pour analyser les parcours individuels. Elles permettent de dépasser la simple juxtaposition d'un grand nombre d'indicateurs en constituant des groupes d'individus décrits par un ensemble de caractéristiques. Les groupes (ou types) sont définis de manière à rassembler les individus statistiquement proches, et à classer dans des groupes distincts les individus ayant des caractéristiques éloignées. « *Contrairement à l'analyse biographique, les méthodes adoptant cette approche sont le plus souvent "non paramétriques" : elles ne font pas d'hypothèse stochastique sur la genèse des parcours. Leur objectif consiste principalement à décrire et explorer les parcours, à en identifier les régularités ou les différences* » (Robette, 2014). Les typologies permettent ainsi de synthétiser de nombreuses informations et d'offrir d'autres perspectives de lecture des données.

Cette étude décrit la méthode typologique mobilisée sur les données Génération², un dispositif d'enquêtes qui permet de suivre des cohortes de jeunes sortis de formation initiale une même année, tous niveaux de diplômes et spécialités confondus, pendant leurs sept premières années sur le marché du travail. L'objectif est de résumer une grande variété de trajectoires d'insertion professionnelle en un nombre réduit de profils. L'intérêt de la méthode est de faire émerger des « parcours-types » (ou « trajectoires-types ») rendant compte des plus ou moins grandes difficultés rencontrées par les jeunes dans leur insertion professionnelle et des formes variées que celle-ci peut emprunter. La création d'une telle typologie permet ainsi de décrire et quantifier les principaux parcours d'insertion de nouveaux entrants sur le marché du travail en regroupant au sein d'une même trajectoire les jeunes qui ont connu des parcours proches.

L'objectif de cet article est de mettre en évidence, tant sur le plan théorique et méthodologique que sur le plan illustratif, l'intérêt de la construction d'une typologie pour comparer des parcours d'individus au sein d'une même cohorte, mais également entre des cohortes indépendantes observées à des dates différentes. La première partie détaillera ainsi la mise en œuvre de la construction d'une typologie de trajectoires professionnelles à partir des données d'une cohorte : celle des sortants de formation initiale en 2010 (« Génération 2010 »). La seconde partie explicitera, dans une perspective comparative, les étapes de construction d'une typologie basée sur trois cohortes différentes observées sur leurs sept premières années de vie active et sorties de formation à 6 années d'intervalle (cohortes des Générations 1998, 2004 et 2010). Des exemples d'analyses mobilisant cette typologie comparable illustreront l'intérêt de cette méthode pour étudier l'évolution des parcours d'insertion des jeunes au gré des contextes conjoncturels des différentes époques et sous l'effet des évolutions plus structurelles ayant pu affecter les comportements sur le marché du travail.

² Ces enquêtes produites par le Centre d'études et de recherches sur les qualifications (Céreq) sont représentatives au niveau national de l'ensemble des sortants de formation initiale une année donnée. Des interrogations, à intervalle régulier, limitent la perte d'information et dressent la situation professionnelle des jeunes à différents moments de leur parcours. Pour réaliser trois interrogations espacées dans le temps, à 3, 5 et 7 ans, un échantillon a été construit de manière à rester représentatif de la structure des sortants de formation initiale l'année scolaire étudiée en anticipant la non-réponse et l'attrition. L'objectif étant de garantir la qualité des données requises dans les enquêtes longitudinales.

1. Méthodologie de construction d'une typologie de trajectoires dans une enquête longitudinale

1.1. Présentation des données de l'enquête 2017 auprès de la Génération 2010

La méthodologie de construction d'une typologie de trajectoires est présentée à partir de sa mise en œuvre sur les données de l'enquête Génération 2010, à savoir les jeunes sortis de formation initiale en 2010, tous niveaux de diplômes et spécialités confondus. Cette génération a été interrogée à trois reprises : en 2013, 2015 et 2017. Pour les individus ayant répondu aux trois vagues de l'enquête (8 882 jeunes), nous disposons d'un calendrier d'activité décrivant mois par mois la situation professionnelle principale des individus entre la fin de leur formation initiale et leur interrogation à l'automne 2017, soit 84 mois en moyenne.

Le principe de remplissage du calendrier d'activité est la saisie successive des situations professionnelles (appelées séquences dans la suite de l'article) rencontrées durant la période d'observation. Les cinq types de situation possible sont les suivants : emploi, recherche d'emploi, formation, reprise d'études, autre situation. Pour décrire les situations d'emploi, le calendrier propose une entrée « établissement employeur », afin d'observer les mobilités professionnelles. En conséquence, au sein d'une période d'emploi continu, chaque changement d'employeur donnera lieu à une nouvelle séquence. Chacun des emplois occupés est décrit précisément : contrat de travail, temps de travail, salaire perçu, caractéristiques de l'entreprise (nature, secteur, taille), etc. Certaines de ces informations peuvent changer au cours d'une séquence, ce que saisit l'enquête en collectant les données à l'embauche et à la débauche, ou au moment de l'enquête pour les emplois en cours. Les jeunes sont également interrogés sur leur parcours de formation, leurs caractéristiques sociodémographiques ainsi que sur des thématiques spécifiques qui varient à chaque enquête (exemples : séjours à l'étranger pendant les études, acquisition de diplôme en post-initial, attractivité de la fonction publique, intermédiaires du marché du travail, etc.).

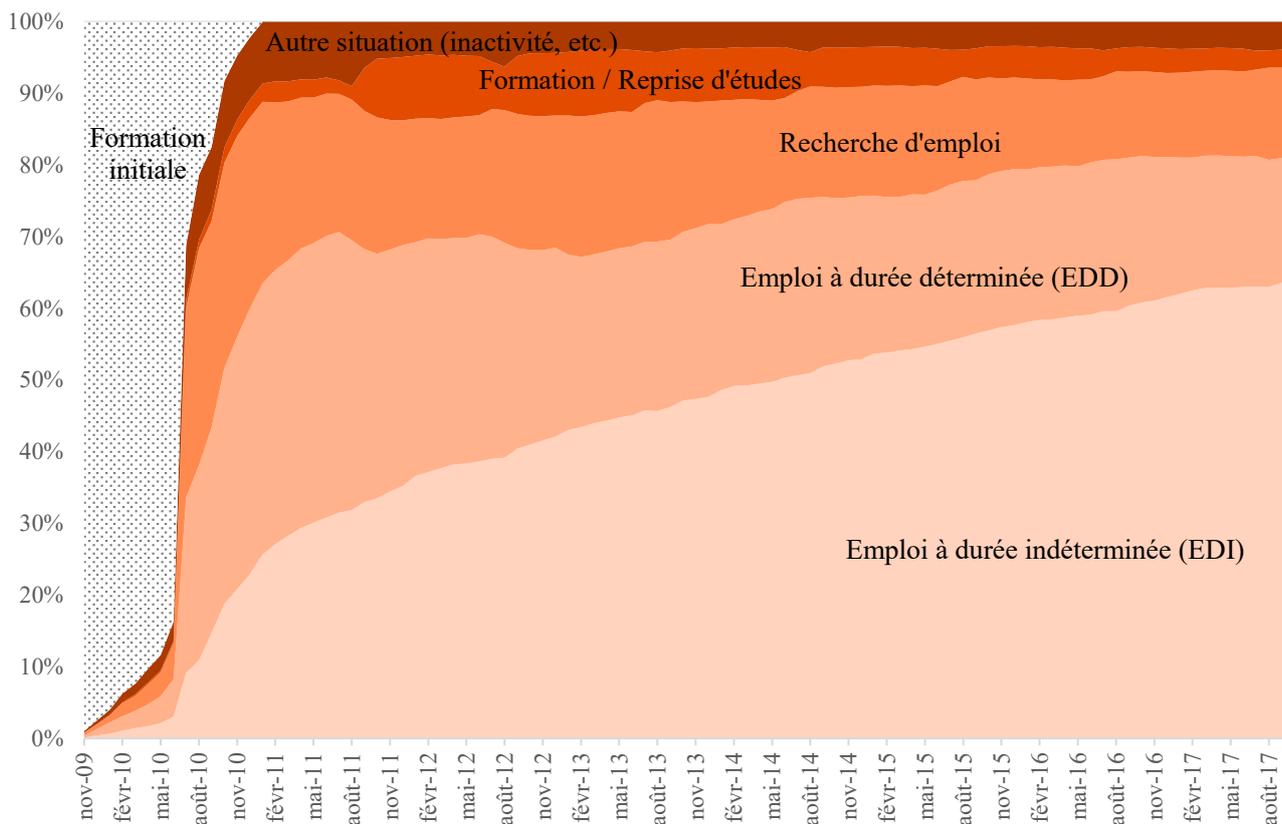
À l'issue des différents traitements post-enquête (Cisse *et al.*, 2019), les fichiers de diffusion de chaque vague d'enquête auprès de la Génération 2010 se répartissent en trois tables :

- Une table « individus » (une ligne par individu), comprenant l'ensemble des données sur le parcours scolaire, les caractéristiques sociodémographiques ainsi que les variables (une par mois) contenant la situation sur le marché du travail du calendrier d'activité mensuel ;
- Une table des séquences d'emploi (une ligne par période continue chez un même employeur), contenant la description des emplois occupés successifs ;
- Une table des séquences dites de « non-emploi » (une ligne par période distincte), représentant les épisodes soit de recherche d'emploi, de formation, reprise d'études ou d'autres situations, avec leur description propre.

Ces fichiers contiennent les données collectées lors des trois interrogations. L'intérêt de disposer de sept années d'observation après la fin de la formation initiale est d'offrir un recul important pour observer les cheminements des jeunes. En effet, ce temps rétrospectif est nécessaire pour analyser l'accès au marché du travail et la stabilisation en emploi de la majeure partie des individus, y compris des individus ayant les parcours les plus difficiles. L'accès à un emploi stable se fait, pour une partie des jeunes, après plusieurs années en emploi à durée déterminée. Une durée longue d'observation permet ainsi de suivre cette dynamique temporelle tout en permettant l'analyse des premières mobilités professionnelles.

Le graphique suivant (figure 1), appelé chronogramme, est construit par juxtaposition des histogrammes cumulés donnant pour chaque mois la part relative de chaque état au sein de la population. Il représente ainsi l'évolution dans le temps (axe des abscisses) de la contribution de chaque état au contingent de la cohorte. Ce type de diagramme donne une bonne idée de l'évolution temporelle d'un ensemble d'individus. Le cheminement de la Génération 2010 dans son ensemble peut ainsi être décrit à l'aide de ce chronogramme global qui retrace, au fil des mois, l'évolution de la part de jeunes pour chaque état donné.

Figure 1 • Évolution de la situation professionnelle des jeunes de la Génération 2010 au cours de leurs sept premières années de vie active



Note de lecture : en janvier 2012, 37 % des jeunes de la Génération 2010 sont en EDI, 32 % en EDD, 17 % en recherche d'emploi, 9 % en formation ou reprise d'études et 5 % dans une autre situation (inactivité, etc.).

Source : Céreq, Génération 2010³, bases comparables Générations 1998, 2004 et 2010, interrogation à 7 ans.

Champ : ensemble des jeunes, France métropolitaine.

On observe que la Génération 2010 est marquée au moment de son entrée dans la vie active par un contexte conjoncturel dégradé (Epiphane *et al.*, 2019) avec des parcours durablement marqués par la crise des dettes souveraines (dès 2012). Après deux années sur le marché du travail, un tiers des jeunes est confronté à des situations de « non-emploi » telles que le chômage ou l'inactivité. En fin d'observation, 13 % d'entre eux sont toujours au chômage et 17 % sont dans un emploi à durée déterminée (EDD). Pour ce qui est de l'accès à l'emploi stable (EDI), on visualise une progression lente et constante pour atteindre 64 % des jeunes en septembre 2017.

³ L'ensemble des traitements commentés dans cet article (y compris dans la partie 1 portant uniquement sur l'enquête Génération 2010) sont effectués à partir du jeu de données comparables entre les Générations 1998, 2004 et 2010 (décrites dans la partie 2.1) car l'intérêt de cette étude se situe dans la perspective de rendre comparable les trajectoires-types de ces trois cohortes (partie 2).

1.2. Constitution de la base d'analyse à partir d'un calendrier d'activité mensuel

Le calendrier d'activité est le cœur du questionnaire de l'enquête Génération. Composé de séquences d'emploi et/ou de non-emploi dans son format initial (en moyenne 3 séquences d'emploi et 2 de non-emploi sur les sept années d'observations), il est converti en une liste de variables indiquant pour chaque mois la situation de la personne et la séquence de son parcours à laquelle elle se rattache. On dispose ainsi d'un calendrier d'activité mensualisé de 98 mois (de novembre 2009 à décembre 2017). Cette fenêtre est la plus large possible, son amplitude dépend de la date de fin des études déclarée, qui oscille entre novembre 2009 et décembre 2010 (champ de l'enquête Génération 2010).

Ces données synthétiques sur le parcours, présentes dans la table « individus » évoquée précédemment, contiennent les informations nécessaires pour la construction de la typologie des trajectoires. Le choix des informations à mobiliser pour construire la typologie peut varier selon les besoins de l'étude. L'enquête Génération dispose d'une information riche et dense en termes de description des emplois. Dans le cadre d'une étude portant sur l'analyse des trois premières années de vie active, l'utilisation des cinq types de situations proposées dans le calendrier d'activité peut s'avérer suffisant pour étudier l'entrée sur le marché du travail. En observant les sept premières années sur le marché du travail, il s'agit cependant de comprendre les mécanismes favorisant l'insertion professionnelle et les logiques de stabilisation durable dans l'emploi. Plusieurs couches d'informations⁴ peuvent être ajoutées pour enrichir l'analyse. Ainsi, il est possible de mobiliser l'information sur le contrat de travail considéré comme un marqueur de la stabilité des positions d'emploi. Cet enrichissement de l'analyse, dans une perspective de comparaisons intergénérationnelles, peut donner des éléments sur les transformations structurelles à l'œuvre sur le marché du travail en observant les effets des évolutions en matière de recrutement des débutants (Milin, 2018) sur leur stabilisation dans l'emploi.

Pour construire cette typologie, chaque situation d'emploi a été caractérisée selon la nature du contrat afin de distinguer les emplois dits « précaires » (toutes les formes de contrat à durée déterminée - EDD), de ceux dits « stables » (CDI, emplois de fonctionnaires et non-salariés). Ce dernier ensemble constitue la catégorie des emplois à durée indéterminée (EDI).

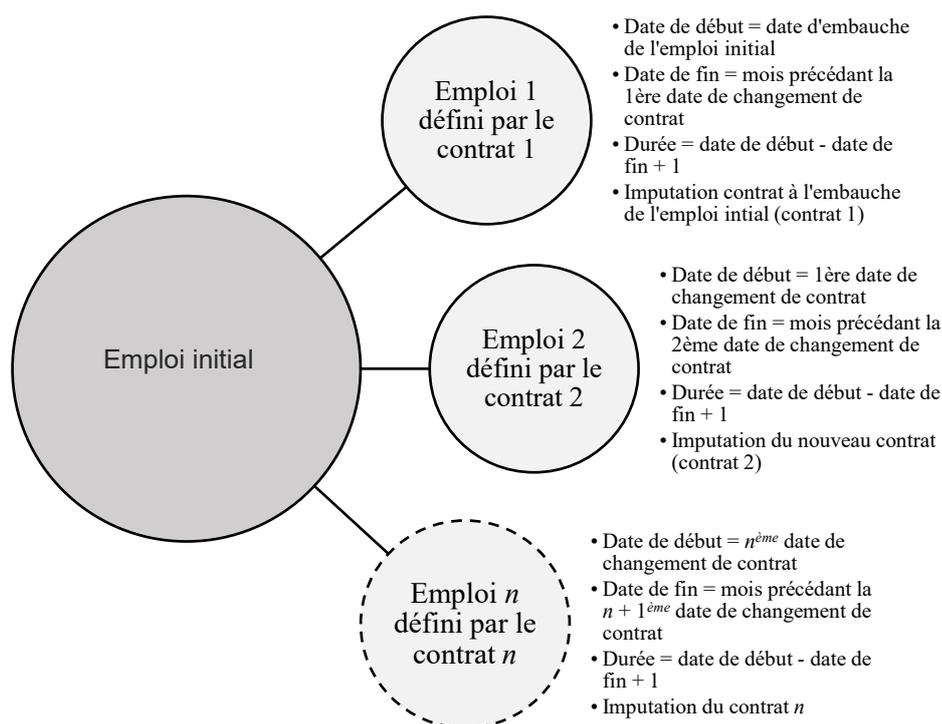
Pour intégrer l'information sur les contrats de travail, nous complétons le calendrier mensualisé avec les données descriptives⁵ de chaque emploi, issues de la base des « séquences d'emplois » (l'une des trois bases sus-évoquées de l'enquête). Dans cette dernière, à partir de ces informations, nous divisons chaque séquence d'emploi en sous-périodes pour distinguer les séquences de contrat de travail. La lecture de cette base est donc modifiée : initialement, le calendrier s'appuie sur une entrée « une ligne dans la base = une séquence continue dans un établissement employeur » ; on retient *in fine* tous les changements d'états avec une entrée « une ligne dans la table = une séquence continue sur un type de contrat de travail dans l'établissement employeur ». Pour exemple, un emploi qui débute en mars 2010 et se termine en juillet 2015 avec un changement de contrat de juin 2015 (une seule séquence au départ) va donner lieu à deux lignes d'emplois successifs (et donc deux séquences au lieu d'une précédemment) : le premier de mars 2010 à mai 2015 avec un contrat de CDD d'une durée de 74 mois et le second, de juin 2015 à juillet 2015 en CDI d'une durée de 2 mois.

Une codification du séquençage du calendrier est opérée afin de valider sa cohérence temporelle (figure 2).

⁴ Par exemple, utiliser les salaires pour approcher les emplois de « qualité » (Mora, 2018).

⁵ Les dates de début et de fin de l'emploi, les dates de changements de contrat (*le cas échéant*, au maximum trois), le type de contrat de travail déclaré au début et à la fin de l'emploi, mais également les contrats liés aux possibles changements survenus dans cet emploi (à la date d'enquête en 2013 et/ou 2015).

Figure 2 • Séquençage du calendrier d'activité mensuel



Après le séquençage, une phase de structuration des données conduit à la création d'un tableau multidimensionnel qui a la forme d'un nouveau calendrier. Cette opération consiste à grouper plusieurs types d'informations dans chaque case avec une double entrée individu et mois d'observation. Pour le parcourir et le remplir, on utilise des boucles pour y imputer du contenu. Pour chaque individu, disposant de 98 cases possibles (une par mois d'observation, de novembre 2009 à décembre 2017), il s'agit, pour chaque mois où l'individu a été en emploi, de spécifier les informations issues du séquençage de manière itérative pour aboutir au calendrier complet. Comme évoqué précédemment, pour chaque mois d'emploi est précisé s'il s'agit d'un emploi à durée déterminée (EDD) ou indéterminée (EDI) pour que la notion de passage d'emploi « précaire » à emploi « stable » puisse *in fine* apparaître dans la typologie que l'on souhaite construire.

Une récupération des informations du calendrier de la table « individus » permet de compléter les cases vides du nouveau calendrier avec les informations des périodes de non-emploi (recherche d'emploi, formation, reprise d'études ou inactivité) pour les mois dans la fenêtre d'observation (premier mois après la fin des études au dernier mois observé). Un regroupement des situations de formation et reprise d'études en une même catégorie est effectué pour évoquer, dans les analyses, les retours en formation dans leur ensemble.

Pour permettre la mise en œuvre de l'analyse factorielle, une dernière opération d'harmonisation de la durée d'observation globale, commune à l'ensemble des répondants de la Génération 2010 est réalisée. En effet, comme la date de sortie de formation initiale diffère pour une partie de la population, les premiers mois d'observations sont vides pour certains individus. De même pour le dernier mois d'observation, les cases du calendrier correspondant à novembre et décembre 2017 sont vides pour certains individus, la collecte ayant été menée durant 3 mois (d'octobre à décembre). Pour résoudre ce problème, on crée un nouveau calendrier identique au précédent mais qui commence pour chaque individu au premier mois d'observation « réel » c'est-à-dire celui qui suit le mois correspondant à la fin de sa formation initiale, peu importe à quel mois calendaire cela correspond.

En définitive, on conserve les 82 premiers mois d'observation qui suivent la sortie (nombre de mois minimum remplis pour l'ensemble des individus) pour chacun des individus. Chacun des mois est caractérisé par l'une des cinq situations professionnelles possibles : emploi à durée indéterminée, emploi à durée déterminée,

recherche d'emploi, formation-reprise d'études ou autre situation. Cette table servira de base d'analyse pour les étapes suivantes.

1.3. Analyse des correspondances multiples

Partant d'une base individuelle composée d'un nombre important de variables qualitatives à analyser, nous avons choisi de réaliser une analyse des correspondances multiples (ACM). Cette méthode a pour but de résumer l'ensemble des combinaisons possibles des cinq situations différentes pour chacun des 82 mois (soit 5 puissance 82 trajectoires différentes possibles) sur un nombre limité de dimensions (*via* les axes factoriels). Cela permet de restreindre l'analyse à un sous-ensemble d'informations pertinentes pour l'élaboration, en aval, d'une classification des trajectoires individuelles. Avant de retenir l'ACM pour réaliser l'analyse des données, deux autres méthodes ont été envisagées.

Dans le cadre du traitement de calendriers (ou séries temporelles) décrivant des changements d'états successifs, Deville et Saporta (1979) proposent une variante de l'analyse des correspondances qui tient compte du temps : l'analyse « harmonique » (AHQ). Elle est adaptée à l'exploration de trajectoires complexes avec l'idée d'introduire la notion de durée dans l'explication des phénomènes sociaux. Toutefois, sa spécificité consiste notamment à lever la contrainte de simultanéité dans le rapprochement des parcours en affectant un poids plus élevé au temps passé dans un état quel qu'il soit. *In fine*, les trajectoires peuvent regrouper des individus dont le cheminement, dans les différents états, diffère dans le temps (Saporta, 1996 ; Robette, 2011). Cette variante n'a pas été retenue pour l'exploitation des données de l'enquête Génération car le principe de cette méthode est de comparer la proportion de temps passé dans chaque état modulo la durée ; ainsi, ce sont les états les plus longs qui ont un poids plus important. De plus, l'intérêt de cette enquête est d'observer les mécanismes d'insertion, on souhaite donc comparer les états mensuels simultanément avec un objectif de classement selon la temporalité et l'ordre dans lequel ils surviennent (exemple : le temps d'accès au premier emploi et l'ordre d'apparition entre un emploi à durée indéterminée et d'autres types d'emploi a son importance dans notre analyse). Cette technique statistique est toutefois utilisée dans d'autres enquêtes du Céreq disposant d'un calendrier professionnel (Stéphanus, 2018).

Une autre approche qui a été envisagée est l'*optimal matching*. Elle consiste, dans une première étape, à chercher à mesurer la dissemblance entre les trajectoires, fondée sur la notion de séquence, avant de réaliser la phase de classification. Cette méthode aurait pu être mobilisée mais est complexe dans sa mise en œuvre par rapport à nos attendus d'exploitation : observer les transitions. Cette méthode est plus souple et plus riche mais requiert de définir en amont les critères d'application et fixer les coûts de rapprochement entre deux trajectoires (paramètres qui vont influencer sur la distance calculée entre ces deux trajectoires). Elle nécessite l'introduction de deux dimensions pour décrire les éléments constitutifs de la trajectoire : le type de séquence et sa durée. Le choix des coûts est complexe, pouvant créer potentiellement une polarisation sur un segment de la trajectoire. Elle peut accorder davantage de poids à certaines séquences selon leur positionnement dans la trajectoire, car le début de trajectoire est en moyenne plus perturbé (nombreux emplois ou alternance emploi-chômage en début de vie active). Par conséquent, par souci de contrôle on choisit une méthode simple pour ne pas favoriser une partie de la trajectoire.

Pour appliquer la technique choisie (ACM), une étape préalable consiste à créer un tableau disjonctif complet à partir des 82 variables correspondant aux mois d'observation, décomposées en autant d'indicateurs qu'il y a de modalités possibles (cinq situations professionnelles : EDI, EDD, recherche d'emploi, formation-reprise d'études ou autre), soit un total de 5x82 indicateurs résumant l'information d'un individu. On effectue ensuite une analyse des correspondances multiples sur la matrice obtenue en utilisant la « proc corresp » du logiciel SAS. Les données sont assorties d'une pondération qui a été intégrée dans l'analyse. Il revient au même de réaliser l'ACM pondérée ou l'ACM usuelle sur le tableau disjonctif complet dont les modalités de la variable sont affectées d'un poids (Cazes, 1980).

Encadré 1 • Utilisation d'une ACM pour mesurer les distances entre les individus

Nous disposons de n individus, de Q variables. Chacune de ces variables nommée q (q variant de 1 à Q) possède J_q modalités. Le total des modalités est noté J où $J = \sum_{q=1}^Q J_q$.

On appelle tableau disjonctif complet Z le tableau à n lignes et J colonnes dont le paramètre générique est défini comme suit :

$$\begin{cases} Z_{ij} = 1 \text{ si l'individu } i \text{ a sélectionné la modalité } j \\ Z_{ij} = 0 \text{ sinon} \end{cases}$$

Distance du X^2 entre deux individus i et k

$$d^2(i, k) = \frac{1}{Q} \sum_{j=1}^J \frac{n}{n_j} (Z_{ij} - Z_{kj})^2$$

Cette distance permet de mesurer les ressemblances entre deux individus : plus ceux-ci auront répondu aux mêmes modalités, plus ils seront proches l'un de l'autre. A partir de la matrice de ces distances, l'ACM consiste à projeter le nuage de points qui se trouve dans un espace de dimension $J-Q$, sur un premier axe qui maximise la part de la distance projetée. Le deuxième axe, orthogonal au premier, constitue avec le premier axe le plan sur lequel l'inertie du nuage de points est maximale. Le troisième axe, orthogonal aux deux premiers, constitue avec les deux autres axes l'espace de dimension 3 dans lequel l'inertie du nuage projeté est maximale. Et ainsi de suite, jusqu'à obtenir $J-Q$ axes (Le Lan, 2005).

Des macros-SAS pour l'analyse des données et la visualisation des résultats sont mises à disposition par l'Insee⁶. Nous avons notamment utilisé la macro `%AideACM` pour visualiser les résultats issus de l'ACM.

Cette étape est destinée à calculer les coordonnées des modalités des variables qualitatives, ainsi que les coordonnées des individus dans un espace de représentation optimal (sur plusieurs dimensions) selon le critère d'inertie. En effet, l'inertie du nuage de points est la somme des distances au carré de tous les points deux à deux. Elle mesure la dispersion du nuage de points et permet de réduire l'hétérogénéité des données tout en conservant une part essentielle de l'information. En outre, pour conserver au mieux la richesse de l'information de départ, on sélectionne un nombre d'axes suffisamment important pour atteindre *a minima* 85 % de l'inertie totale observée. Cette réduction de l'information et le passage de données qualitatives à quantitatives étant l'objectif principal à l'utilisation de cette méthode, il n'est pas apparu utile d'interpréter les résultats obtenus sur les premiers axes factoriels comme il est classique de le faire.

1.4. Méthode de classification

Après avoir réduit l'espace d'analyse en limitant la perte d'information grâce à l'ACM, nous cherchons à regrouper les individus selon des critères de proximité statistique, en suivant une logique d'agrégation pas à pas.

L'ACM, comme prérequis pour la construction de la typologie, a permis d'obtenir une information synthétique au travers des coordonnées des individus sur les axes factoriels : la classification va donc pouvoir se faire à partir de ces coordonnées factorielles quantitatives. Une classification ascendante hiérarchique (CAH) a été réalisée à partir des 32 premiers axes factoriels, axes qui ne représentent que 10 % des dimensions de l'espace factoriel mais qui portent 85 % de l'inertie. La procédure d'agrégation choisie pour la constitution des classes est la méthode de Ward, généralement considérée comme l'une des deux plus performantes dans les méthodes de classification hiérarchiques (Malhotra 2004, p. 548). En effet, cette méthode a été conçue dans un objectif de maximisation de l'inertie interclasses et de minimisation de l'inertie intra-classe (analyse approchée de la variance) : l'idée est donc de constituer les groupes les plus homogènes possible, rassemblant des individus aux caractéristiques proches. Inversement, les individus appartenant à deux groupes distincts doivent être les plus différents possibles (encadré 2).

⁶ <https://www.insee.fr/fr/information/2021906>

L'usage de la pondération lors de cette phase est indispensable : d'une part, pour corriger les biais de représentativité et par conséquent éviter que les résultats ne soient biaisés et, d'autre part, afin d'éviter d'isoler les *outliers* (valeurs extrêmes) dans une classe donnée.

Encadré 2 • Caractériser des trajectoires-types pour appréhender l'insertion professionnelle

Une CAH (classification ascendante hiérarchique) est un algorithme de classification d'individus*. Elle peut être effectuée à partir des coordonnées de ces individus sur les axes d'une ACM. En général, on garde un nombre d'axes limité. En pratique, soit on en garde un nombre communément admis (par exemple un seuil de 10 axes), ou alors le nombre d'axes nécessaires pour satisfaire une condition : soit les axes nécessaires pour conserver une proportion fixée de l'inertie totale, soit les axes qui portent une inertie supérieure à celle d'un axe moyen. Dans tous les cas, les axes non gardés pour classer les individus sont alors considérés comme peu informatifs et portant majoritairement du « bruit ».

Une fois le nombre d'axes de l'ACM fixé, nous travaillons dans un sous-espace de dimension égale à ce nombre. Cet espace est défini par une métrique euclidienne classique, puisque chaque individu y est représenté par des coordonnées avec des valeurs non plus qualitatives mais quantitatives.

Pour initialiser la CAH, un nombre de classes, supérieur au nombre de classes retenues à l'issue de la CAH, est en général choisi en amont et un premier jeu de centres de classes est tiré aléatoirement pour initier le processus de nuées dynamiques. Chaque individu est alors affecté au centre de classe le plus proche. L'algorithme de la CAH consiste ensuite à recalculer les nouveaux centres de classes et regrouper les individus en fonction de ces nouveaux centres de classes. Une fois la convergence du processus atteint, le jeu du rapprochement des classes, en commençant par les plus proches, peut s'opérer jusqu'à constituer l'arbre de classification dans son intégralité. La distance mobilisée est celle dite de Ward définie comme suit (Le Lan, 2005) :

$$d(C_i, C_j) = I(C_i \cup C_j) - [I(C_i) + I(C_j)]$$

où $I(C_i)$ désigne l'inertie de la classe C_i , avec $I(C_i) = \sum_{k \in C_i} d^2(k, G_i)$

où G_i est le centre de gravité de C_i

Cette distance est également égale à :

$$d(C_i, C_j) = \frac{m_i m_j}{m_i + m_j} d^2(G_i, G_j)$$

où m_i et m_j désignent les masses des classes C_i et C_j .

L'opération d'agrégation est répétée jusqu'à obtenir une seule classe. À partir de cet algorithme sont élaborés l'arbre hiérarchique et l'histogramme des indices de niveaux qui mettent en évidence, lorsque l'on passe d'une configuration de n classes à $n-1$ classes, le gain d'inertie intra-classe (ou perte d'inertie interclasses**) et donc un degré de perte d'homogénéité de la partition. Plus cette perte est élevée, plus on aura tendance à estimer que la configuration à n classes est statistiquement une « bonne » configuration***.

* Il existe également des CAH pour classer des variables, mais dans le cas présent, ce sont les individus que nous souhaitons classer.

** En effet, pour toute partition des individus, l'inertie inter classes plus l'inertie intra classe est égale à l'inertie totale qui est constante. Un gain d'inertie intra classe lors du passage d'une configuration de n classes à $n-1$ classes équivaut donc à une perte égale d'inertie inter classes.

*** Cette méthode a une limite : en général, ces « sauts » d'inertie ont mécaniquement plus de chance d'être élevés pour des configurations avec des plus petits nombres de classes. Aussi lorsque nous observons deux sauts d'inertie équivalents pour deux nombres de classe différents, nous considérons que la « meilleure » des deux configurations est celle ayant le nombre de classes le plus élevé.

Pour représenter les étapes successives de la CAH, il est courant d'utiliser un arbre hiérarchique (à l'aide de la macro-programme %cahnum de l'Insee) regroupant les classes deux à deux à chaque étape, en fonction de la distance entre ces dernières, pour n'en former plus qu'une dans l'ultime étape. Rappelons que pour un nombre de classes donné, la partition idéale est celle qui minimise l'inertie intra-classe, permettant ainsi de former des classes les plus homogènes possibles. Ainsi, le dendrogramme nous indique l'ordre dans lequel les agrégations successives ont été opérées. Il nous indique également la valeur de l'indice d'agrégation à chaque niveau d'agrégation. Il est généralement pertinent d'effectuer la coupure après les agrégations correspondant à des valeurs peu élevées de l'indice et avant les agrégations correspondant à des valeurs élevées. En coupant l'arbre au niveau d'un *saut important* de cet indice, on peut espérer obtenir une partition de bonne qualité car les individus regroupés en-dessous de la coupure étaient proches, et ceux regroupés après la coupure sont éloignés. Dans notre cas, six classes se dégagent. On s'appuie notamment sur un indicateur de mesure de la qualité (critère décisionnel) pour conforter ce choix : le *cubic criterion clustering*⁷ (Sarle, 1983), dont la valeur calculée est supérieure à 2. On peut alors considérer que la classification est bonne.

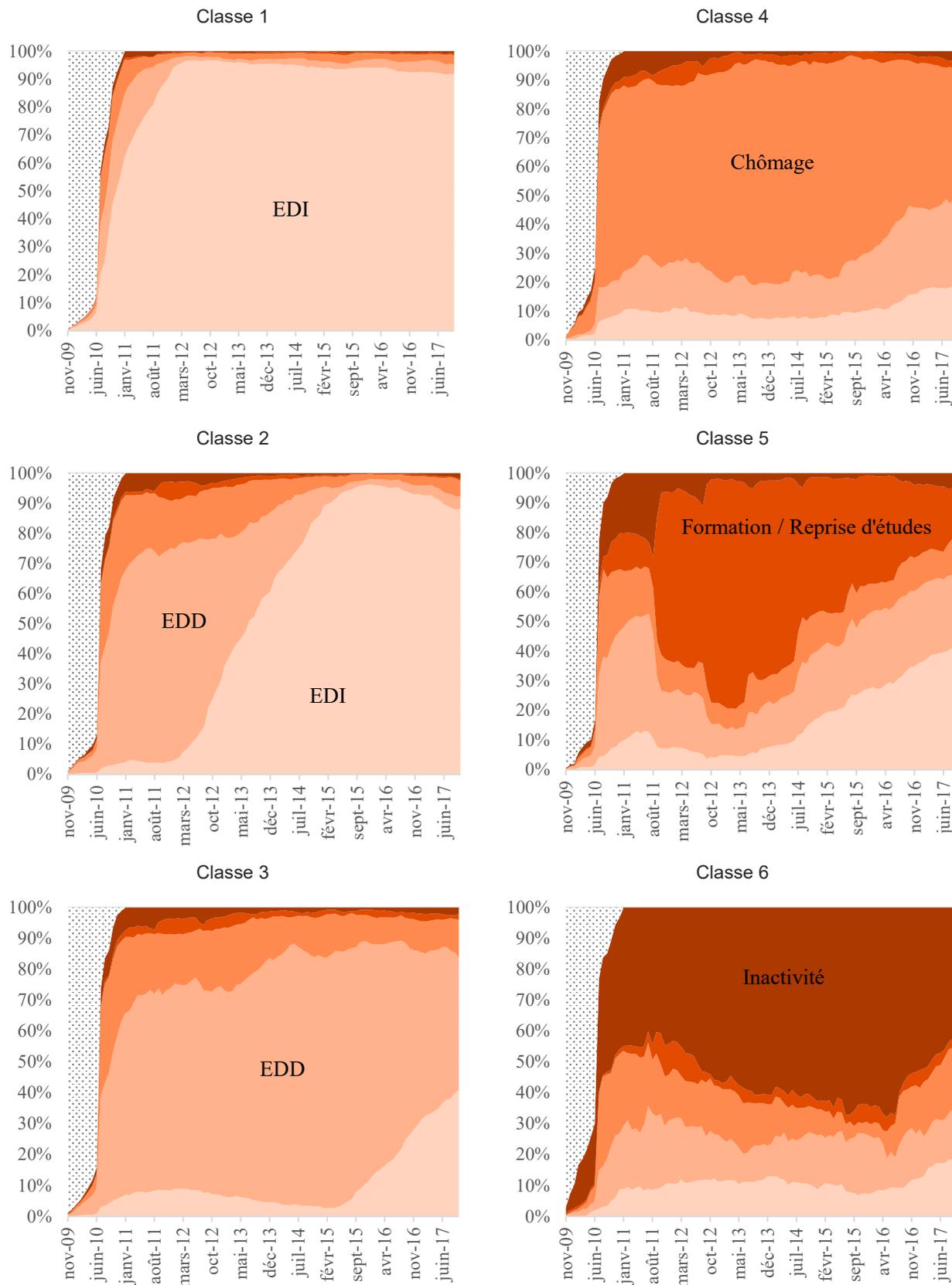
1.5. Mise en application

La mise en œuvre de la méthode est effectuée à partir de la base de données de la Génération 2010, issue du jeu de données comparables des enquêtes Génération 1998, 2004 et 2010 lors desquelles les jeunes sortis de formation initiale ces années-là sont suivis pendant leurs sept premières années de vie active. Six trajectoires-types d'insertion professionnelle ont ainsi été identifiées. Sont regroupés au sein d'une même trajectoire les jeunes qui ont connu des parcours proches. *A contrario*, les parcours seront considérés d'autant plus éloignés qu'ils diffèrent sur un nombre important de mois au cours des trois années qui suivent la formation initiale.

L'interprétation des trajectoires-types se fait grâce à la lecture de représentations graphiques de type chronogrammes (figure 3) ainsi que le calcul d'indicateurs décrivant le parcours des jeunes de chacune des classes (nombre de mois au chômage, nombre d'emplois occupés, conditions d'emploi à la date de l'enquête, etc.).

⁷ Le Cubic Clustering Criterion (CCC) de Sarle indique si la classification est éloignée de celle qui aurait été obtenue par une distribution uniforme du nuage des individus, et si elle peut être jugée bonne (CCC > 2), à vérifier (CCC entre 0 et 2), ou si elle risque d'avoir été affectée par des individus hors normes (CCC < 0).

Figure 3 • Représentation des trajectoires-types



Source : Céreq, bases comparables Génération 1998, 2004 et 2010, interrogations à 7 ans.
 Champ : ensemble des jeunes de la Génération 2010, France métropolitaine.

Ces trajectoires-type peuvent alors être décrites au travers de dénominations résumant le parcours sur 7 années de vie active des jeunes de chaque classe.

- Classe 1. Stabilisation *rapide* en emploi à durée indéterminée (EDI)

Pour un tiers des jeunes débutant sur le marché du travail de la Génération 2010 (34 %), le processus d'insertion est le plus direct, avec un accès rapide à l'emploi à durée indéterminée et un maintien dans ce statut d'emploi par la suite (si l'individu a changé d'employeur sur la période, il est resté dans ce statut). Dès janvier 2012, 94 % des jeunes concernés par cette trajectoire étaient en emploi à durée indéterminée. En moyenne, ils ont accédé à leur premier EDI au bout de 5 mois sur le marché du travail et ils ont enregistré trois mois de chômage seulement au cours de ces sept années. Ces trajectoires ne vont pas toutes de pair avec un employeur unique puisque la moitié de ces jeunes ont eu au moins deux emplois différents sur la période.

- Classe 2. Stabilisation *différée* en emploi à durée indéterminée (EDI)

21 % des jeunes de la Génération 2010 ont accédé à l'emploi à durée indéterminée, mais plus tardivement que dans la première trajectoire. Ils ont conservé ce statut d'emploi dans la suite de leurs parcours. L'accès à un emploi à durée indéterminée est donc progressif pour les jeunes ayant connu ce type de parcours, avec dans la majorité des cas un passage préalable par un ou plusieurs emplois à durée déterminée avant d'avoir un emploi stable. Si en juin 2012, soit 2 ans environ après la fin de leurs études, 64 % avaient un emploi à durée déterminée, seuls 17 % avaient un emploi à durée indéterminée. Cette part augmente progressivement, de sorte que dès mi-2015, leur taux d'emploi à durée indéterminée atteint 95 %. Ainsi, les jeunes de cette trajectoire ont passé 83 % de la période en emploi, ont accédé à leur premier emploi au bout de 5 mois en moyenne sur le marché du travail et ont obtenu leur premier emploi à durée indéterminée au bout de 33 mois en moyenne.

- Classe 3. Emploi à durée déterminée (EDD) dominant, durable ou récurrent

Pour 19 % des jeunes de la Génération 2010, l'accès à l'emploi s'est fait rapidement mais ils restent tout au long de leurs sept premières années de vie active dans des emplois à durée déterminée, soit successifs soit entrecoupés de courtes périodes de recherche d'emploi. La moitié de ces jeunes a eu au moins quatre emplois différents sur la période. Ainsi ces jeunes ont passé 78 % de leurs sept premières années de vie active en emploi et ont en moyenne accédé à leur premier emploi au bout de 5 mois comme les jeunes de la classe précédente. En revanche, seuls 41 % sont en emploi à durée indéterminée au bout de leurs sept premières années de vie active et parmi ceux qui ont eu un EDI dans la période, le temps d'accès moyen était de 50 mois (un peu plus de 4 ans).

- Classe 4. Chômage persistant ou récurrent

Pour 13 % des jeunes de la Génération 2010, les années qui suivent la formation initiale sont dominées par le chômage, qu'il s'agisse d'épisodes récurrents entrecoupés de courts emplois ou de formations brèves, ou d'une recherche d'emploi n'ayant quasiment jamais abouti. En moyenne, ces jeunes ont passé presque deux tiers de leurs sept premières années de vie active en recherche d'emploi (64 %), pour 28 % de temps en emploi seulement. Dans ce groupe, 22 % n'ont même jamais occupé d'emploi pendant les cinq premières années de vie active. La présence d'emploi dans ces parcours devient plus importante en fin de période même si, sept ans après la fin de leurs études, seule la moitié des jeunes sont en emploi et à peine 40 % de ces emplois sont à durée indéterminée.

- Classe 5. Longue(s) période(s) en formation ou reprise d'études

Parmi les jeunes qui avaient mis fin à leurs études en 2010, 9 % se singularisent par une reprise d'études post-initiale plutôt longue pendant les sept années qui suivent leur sortie de formation initiale. Ces jeunes ont passé en moyenne trois de ces sept années en reprises d'études. Leurs parcours d'insertion analysés sur ses autres composantes apparaissent moins défavorables que ceux enregistrés par les jeunes du groupe précédent, restés durablement aux marges du marché du travail, puisqu'ils ont passé 39 % de leur parcours en emploi. Au terme de leurs sept premières années sur le marché du travail, 43 % des jeunes de ce groupe occupent un emploi à durée indéterminée et 22 % un emploi à durée déterminée.

- Classe 6. Inactivité durable

4 % des jeunes de la Génération 2010 ont un parcours durablement marqué par des situations en marge du marché du travail (hors emploi, recherche d'emploi, formation, reprise d'études), majoritairement de l'inactivité,

qui représentent en moyenne 53 % de leurs sept premières années de vie active. Leur temps d'accès au premier emploi est en moyenne de presque deux ans (22 mois). Au moment de l'interrogation, 38 % sont encore en inactivité, seuls 39 % occupent un emploi (dont un peu plus de la moitié à durée indéterminée) et 18 % en recherchent un.

Ainsi, les profils-types de la Génération 2010 témoignent de la grande hétérogénéité des parcours des jeunes au moment de s'insérer sur le marché du travail. Pour la majorité des jeunes de la Génération 2010, l'accès et le maintien dans l'emploi s'est effectué sans difficultés notables et ils ont peu ou n'ont pas été confrontés au chômage (trois premières classes). À l'opposé, une partie des jeunes reste longtemps aux marges du marché du travail, ne parvenant qu'épisodiquement à avoir un emploi. Entre les deux, des expériences variées de transition existent comme les retours en formation en début de vie active.

Ces classes décrites et illustrées ci-dessus pour l'enquête la plus récente, celle de la Génération 2010, serviront de référence pour la réalisation d'une typologie comparable puisque les jeunes des deux précédentes enquêtes Génération (1998 et 2004) seront ajoutés à la typologie comme individus supplémentaires (projection). Le détail de la méthode est abordé dans la partie suivante.

2. Construction d'une typologie de trajectoires comparables entre plusieurs enquêtes Génération

Un intérêt de la typologie présentée précédemment est de pouvoir mettre en évidence l'hétérogénéité des parcours d'insertion et leurs formes-type au sein d'une même cohorte (la Génération 2010), pour comprendre quels profils d'individus s'inscrivent dans tel ou tel parcours-type. Une autre utilisation des typologies de trajectoires cible plutôt une comparaison entre différentes cohortes. L'un des atouts du dispositif des enquêtes Génération est en effet de pouvoir comparer les cohortes successives, qui se sont insérées sur le marché du travail dans des contextes différents. Il s'agit dans notre cas de comparer les sept premières années de vie active des jeunes sortis de formation initiale en 2010 (Génération 2010) avec deux précédentes cohortes sorties en 1998 et en 2004 (« Génération 1998 » et « Génération 2004 »). Cela nécessite alors de créer une typologie commune à ces trois enquêtes. En effet, il n'est pas possible de comparer simplement trois typologies faites indépendamment sur des Générations différentes, même si l'on applique la même méthode à chaque fois (mêmes variables, même nombre d'axes retenus dans la classification, etc.).

La première raison est que les classes obtenues pourraient être intrinsèquement différentes d'une Génération à l'autre, à cause par exemple d'un contexte économique spécifique pour l'une des enquêtes (contexte de crise économique de 2008 pour la Génération 2004 et crise des dettes souveraines de 2012 pour la Génération 2010) ou encore la présence du service national uniquement pour la Génération 1998.

La deuxième raison est qu'une typologie n'est valable que pour les données sur lesquelles elle a été construite. En d'autres termes, ce n'est pas parce que deux classes de deux typologies différentes ont des caractéristiques et des noms proches qu'elles recouvrent exactement la même réalité. Cela peut être illustré en prenant l'exemple des deux typologies de parcours professionnels créées de manière indépendante sur les données des sept premières années de vie active des Générations 2004 et 2010⁸. Dans les deux cas, trois parcours-types ont reçu les mêmes appellations « Accès rapide à l'emploi à durée indéterminée (EDI) », « Emploi à durée déterminée (EDD) durable ou récurrent » et « Chômage durable ou récurrent » mais selon la Génération, les parcours qu'ils recouvrent ne sont pas exactement les mêmes. Ainsi, dans le premier parcours-type, si l'accès au premier EDI se fait en 4 mois en moyenne dans la Génération 2004, il est de près de 11 mois pour la Génération 2010. De la même manière, pour la Génération 2004, parmi les individus dans le parcours « EDD durable ou récurrent », 85 % accèdent à un EDI et ce en 37 mois en moyenne alors que leurs homologues de la Génération 2010 sont seulement 79 % à être concernés et leur temps d'accès moyen à leur premier EDI est de 48 mois. Enfin, pour le parcours-type « Chômage durable ou récurrent », la part de

⁸ Les calendriers d'activité des deux enquêtes qui ont servi de base aux deux typologies ont les mêmes situations (emploi, chômage, formation, reprises d'études, autre situation) mais les bases de données utilisées n'ont pas eu de traitement pour les rendre comparables, comme expliciter dans la partie 2.1.

temps médiane passée au chômage sur les 7 ans par les individus de cette classe est de 70 % pour la Génération 2004 et de 62 % pour la Génération 2010. Ainsi, un individu de la Génération 2004 classé dans une certaine catégorie de trajectoire n'aurait pas forcément été classé dans cette classe dans la typologie de la Génération 2010 si on l'avait ajouté comme individu dans cette enquête.

Pour obtenir une typologie commune et comparable entre différentes cohortes, il faut donc une méthodologie spécifique. La partie qui suit s'attache à décrire la méthodologie employée sur les données des enquêtes Génération 1998, 2004 et 2010.

2.1. Constitution d'un jeu de données comparable à partir des enquêtes Génération 1998, 2004 et 2010

Les enquêtes longitudinales Génération 1998, 2004 et 2010 ont été mobilisées pour construire une typologie de trajectoires comparable. Il s'agit de trois cohortes de jeunes sortis du système éducatif chacune de ces trois années et interrogés sur leurs parcours professionnels mois par mois durant leurs sept premières années de vie active. Les effectifs de répondants à l'ensemble des vagues d'enquêtes⁹ sont de 15 714 jeunes pour la Génération 1998, 11 426 pour la Génération 2004 et 8 461 pour la Génération 2010. La construction d'une typologie de trajectoires commune à ces trois Générations permet d'appréhender le processus d'insertion professionnelle aux prismes de la conjoncture et des évolutions du marché du travail.

Étant donné les évolutions du champ (nouvelles formations, etc.) et parfois du contenu des enquêtes d'une Génération à l'autre, des jeux de données dites « comparables » ont été constitués par le Céreq. Le travail d'harmonisation a consisté à :

- Délimiter un périmètre de champ commun aux trois Générations¹⁰ ;
- Rapprocher leurs méthodes de pondérations (et de calage) par le biais d'un redressement par grands niveaux de formation afin de lisser l'effectif total sur celui des enquêtes Emploi rétropolées ;
- Uniformiser l'information collectée pour tenir compte des évolutions de concepts (exemple : passage des francs en euros), de nomenclatures (diversité des diplômes par exemple) et de questionnement.

Pour ce qui est de l'élément central du questionnement, à savoir le calendrier d'activité, il n'a que très peu évolué. Toutefois, les jeunes enquêtés de la Génération 1998 ont fait parfois partie, selon leur âge, des dernières classes appelées au service militaire. Cette situation professionnelle notée « service national » a donc été proposée lors de l'interrogation de cette Génération, mais pas dans les enquêtes suivantes. Or cette situation est singulière dans la transition formation-emploi puisqu'elle ne constitue ni de l'emploi, ni de l'inactivité, ni de la formation, ni de la recherche d'emploi, c'est-à-dire aucune des quatre situations qui constituent le calendrier d'activité des deux autres Générations. Cette situation est une parenthèse imposée, uniquement pour une partie des jeunes hommes, et qui intervient très vite dans leur parcours d'insertion professionnelle.

Pour que la typologie puisse être comparable entre les trois Générations, il faut que l'on puisse recréer pour les Générations 1998 et 2004 un calendrier d'activité de même nature que celui présenté précédemment pour la Génération 2010, c'est-à-dire avec une longueur de 82 mois et cinq situations possibles : deux correspondant à de l'emploi (EDD/EDI) et trois de non-emploi (formation/reprises d'études, recherche d'emploi, autre situation). Il n'est donc pas possible de laisser la situation du service militaire comme une catégorie à part dans le calendrier d'activité de la Génération 1998. Un traitement spécifique a donc été nécessaire sur les données de cette enquête (cf. encadré 3).

⁹ Chacune de ces trois Générations a été interrogée à 3, 5 et 7 ans après la fin des études. Les effectifs cités ici concernent les bases comparables, c'est-à-dire après reconstitution d'un champ commun aux trois Générations.

¹⁰ Suppression des étudiants étrangers, des sortants des Départements d'outre-mer (DOM) et des sortants de 6^{ème} et 5^{ème} et des instituts universitaires de formation des maîtres (IUFM) non présents dans la base de sondage pour toutes les enquêtes.

Encadré 3 • Dilatation du calendrier d'activité de la Génération 1998

Parmi les répondants aux trois vagues de l'enquête Génération 1998, la déclaration d'un service national a concerné 2 127 jeunes hommes, âgés de 20 ans ou plus en 1998 (13,5 % des enquêtés) et qui avaient presque tous atteint un niveau bac et au-delà. Pour l'essentiel, ce service national débute entre 7 mois et 16 mois après la sortie des études avec une durée médiane de 10 mois, débordant donc largement sur la deuxième année de vie active.

Une première solution pour pouvoir avoir un calendrier d'activité comparable entre les trois Générations aurait été d'écarter ces individus concernés par le service national de l'analyse ; mais les effectifs étaient trop importants et cette suppression aurait entraîné une déformation trop grande de l'échantillon, porteuse de biais potentiels dans les analyses subséquentes, en supprimant uniquement des hommes de la Génération 1998.

Une autre approche aurait été de saisir le service national non comme une situation à part entière mais de redéfinir cette séquence pour l'intégrer à l'un des types de situations retenues ; en l'occurrence, il aurait pu être rapproché d'une situation d'emploi à durée déterminée. Cela pose comme nouveau problème un allongement artificiel, éventuellement conséquent, de la seule durée d'emploi dans le parcours et entraîne une déformation des parcours orientée vers un type de situation. Or, le service national n'a pas les attributs de l'emploi sur le marché du travail et ne peut compter comme expérience de travail exactement dans les mêmes termes que les séquences effectives d'emploi. De plus, cette modification n'interviendrait que pour un profil spécifique d'individus entraînant une distorsion en termes de genre et d'âge (et donc aussi de niveau de diplôme atteint en sortie de formation initiale).

Pour obtenir un calendrier de même longueur que ceux des Générations 2004 et 2010, la méthode finalement choisie consiste à supprimer ces déclarations de service national et procéder à une « dilatation » proportionnelle de l'ensemble des périodes déclarées par les individus concernés. Ceci est fait pour ventiler le nombre de mois dont l'information a été supprimée sur l'ensemble du calendrier plutôt que sur un seul état. L'incidence de cette méthode sur les parcours est donc faible puisque l'ensemble des périodes du calendrier sont allongées et le nombre de mois ajoutés est limité, le service national n'affectant que 12 mois maximum sur les 82 mois au total du calendrier (soit 15 % maximum de la durée des parcours).

Ce travail consiste en un recalcul par imputation de la durée des événements du calendrier. Les étapes ont été les suivantes :

- Suppression des situations de type « service national » ;
- Renumérotation des situations dans l'ordre chronologique des situations restantes ;
- Recalcul de la durée de chaque situation : elle est égale à la durée initiale multipliée par la durée totale du calendrier d'activité de l'individu (y compris le service national), divisé par la durée totale hors durée du service national ;
- Recalcul des dates de début et de fin de chaque situation en cohérence avec la nouvelle durée, afin d'obtenir un calendrier complet dont le premier et le dernier mois d'observation concordent avec ceux du calendrier initial.
- Recalcul des dates de changements de contrats (*le cas échéant*) au sein des situations d'emploi, en cohérence avec les nouvelles dates de début et de fin de ces situations et en tenant compte de la temporalité à laquelle intervenait le changement de contrat initialement (exemple : si un tiers de la durée d'emploi s'est écoulé jusqu'au changement, cette proportion sera conservée).

2.2. Une classification pour trois générations

Une fois obtenu ce jeu de données comparables entre les trois Générations (même champ et calendrier de même longueur avec les mêmes situations), le processus de construction des « nouveaux » calendriers d'activité tel que décrit précédemment (partie 1.2) a été appliqué.

À partir de là, deux options sont envisageables pour construire la typologie de parcours. La première consiste à utiliser de la même manière l'ensemble des individus en compilant les données des trois Générations. Dans la deuxième, une seule des enquêtes est utilisée pour faire émerger les types de trajectoires et les individus des deux autres Générations sont quant à eux « projetés » sur ce classement (ils sont traités en « observations supplémentaires »).

Dans le premier cas (données compilées), on aurait pu observer quelle Génération est plus concernée par un type de parcours qu'un autre mais avec le risque de voir des configurations de trajectoires gommées dans chacune des Générations. De plus, le résultat serait plus « abstrait » puisque la typologie ne reflète plus une époque précise mais un mélange de plusieurs époques. On retient donc la deuxième méthode, nous permettant de regarder ce qui se passe sur les Générations antérieures au prisme des constats faits sur la Génération 2010, sur laquelle on choisit de créer la typologie.

À partir de cette nouvelle base de données, nous utilisons la même méthodologie que celle décrite dans la partie 1 (analyse des correspondances multiples suivie d'une classification ascendante hiérarchique). Afin d'obtenir la même typologie de trajectoires sur chacune des trois enquêtes, celle-ci est construite uniquement sur les données de la Génération 2010. Nous obtenons ainsi 6 classes de parcours. Les individus des enquêtes Génération 1998 et 2004 sont introduits en individus supplémentaires (en passant leur pondération en négatif, cf. encadré 4). L'objectif est de rattacher ces individus à l'une des six trajectoires-types déjà construites qui servent de référence et ainsi de pouvoir réaliser des comparaisons intergénérationnelles. Il est alors possible d'étudier l'évolution des proportions de jeunes concernés par chacune des 6 trajectoires-types entre chaque Génération.

Encadré 4 • Projection des individus des Générations 1998 et 2004

Pour représenter les individus des Générations 1998 et 2004 dans le nuage de points formé par les individus de la Génération 2010 (suite à l'analyse des correspondances multiples), on peut projeter* leurs coordonnées sur le plan des axes principaux du nuage. Par défaut, tous les individus ont le même poids. Un individu prenant une valeur négative, nulle ou manquante sur cette variable de poids est traité comme individu supplémentaire. Dans notre cas, le poids des individus des Générations 1998 et 2004 ont été passés en négatif dans l'analyse des correspondances multiples et la classification. Ces individus supplémentaires n'interviennent donc pas dans les calculs de détermination des axes factoriels. Pour les positionner, la combinaison linéaire des variables actives définies par chaque axe est utilisée.

Concernant la qualité de la représentation, lorsque des points de projections des individus sont éloignés sur un axe (ou sur un plan), on peut assurer que les points représentant ces individus sont éloignés dans l'espace. En revanche, deux individus dont les projections sont proches sur un axe (ou sur un plan) peuvent ne pas être proches dans l'espace.

* « Plusieurs auteurs ont déjà souligné que le terme supplémentaire s'applique souvent de manière abusive à tout ce qui ne fait pas partie du tableau des données alors qu'on devrait bien réserver le terme projection en individus supplémentaires à une opération géométrique précise » (Chessel, Lobry, 2017).

2.3. Exemples d'analyses des données à partir de la typologie comparable

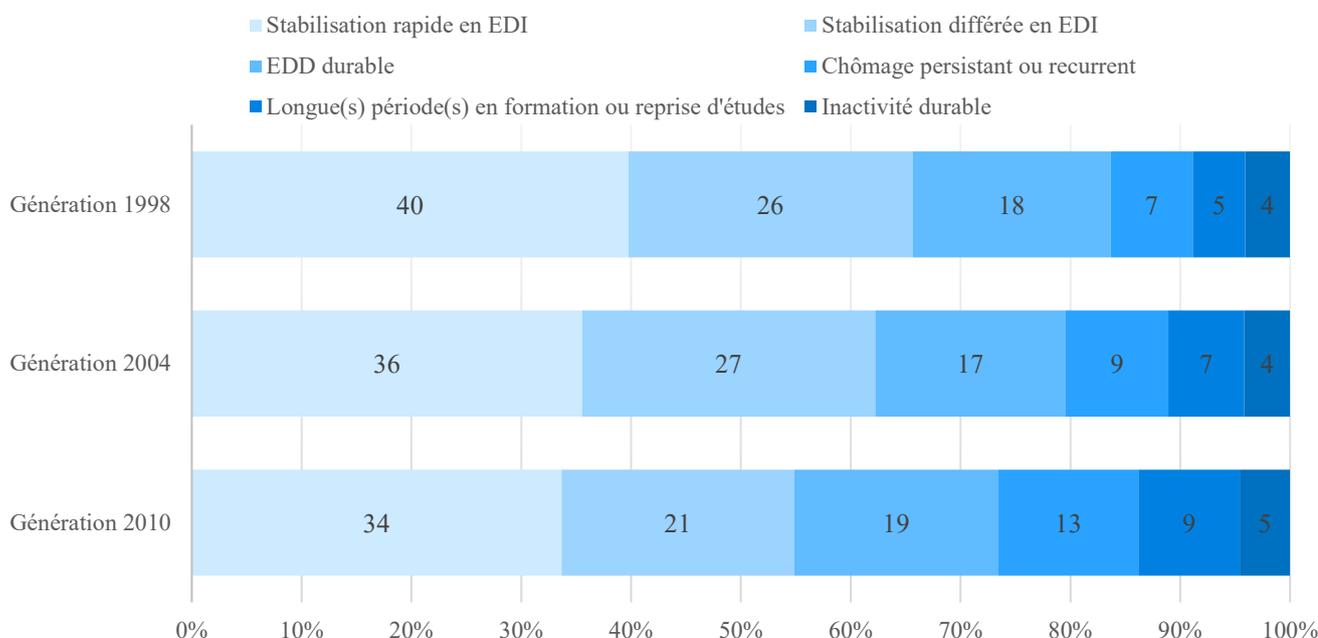
Comme l'indique Yvette Grelet (2002) : « La classification n'apporte aucun élément sur les déterminants des parcours, ce n'est pas son objet, mais le croisement des classes et des caractéristiques individuelles suggère des associations et des hypothèses sur ces déterminants qui peuvent être confirmées, par exemple, au moyen de régressions logistiques (où l'on modélise la probabilité d'appartenance à un type) ». Cette auteure illustre bien que l'intérêt de la construction d'une typologie réside à la fois dans sa mobilisation seule pour comparer les trajectoires de différentes sous-populations mais également en confrontant cette variable synthétique à d'autres variables pour connaître les déterminants de ces différents types de parcours.

La typologie précédemment construite sera utilisée ci-dessous pour illustrer l'évolution de l'insertion professionnelle des jeunes au fil des Générations. Epiphane *et al.* (2019) mobilisent d'ailleurs cette typologie avec cet objectif dans le cadre de leur étude. Ces trajectoires seront également analysées à la lumière du niveau de diplôme, du genre ou de la dimension territoriale.

2.3.1. Indicateur synthétique longitudinal d'analyse de l'insertion professionnelle

En préambule, pour mieux appréhender la comparaison de la typologie selon les trois Générations étudiées, il convient de resituer leurs contextes respectifs. D'une part, le niveau de diplôme détenu en fin de formation initiale a augmenté au fil des cohortes. Environ 10 % des jeunes des Générations 1998 et 2004 sortent avec un diplôme de niveau bac+5 et plus, contre 16 % de la Génération 2010. D'autre part, ces Générations ont fait face à un contexte économique de plus en plus dégradé au moment de leur entrée sur le marché du travail. Ce contexte pèse plus particulièrement sur l'accès à l'emploi des jeunes de la Génération 2010, confrontés dès 2012 à un rebond de la crise financière de 2008, ralentissant leur insertion professionnelle et les exposant davantage au chômage en début de vie active (Epiphane *et al.*, 2019). Ainsi, 18 mois après la fin des études, la Génération 1998 comptait 12 % de chômeurs, contre 15 % pour la Génération 2004 et 20 % pour la Génération 2010. De plus, la Génération 2010 est plus concernée que les autres par les embauches sous forme de contrats à durée limitée (Milin, 2018).

Figure 4 • Répartition des 6 types de trajectoires des jeunes pour les trois enquêtes Générations



Source : Céreq, bases comparables Génération 1998, 2004 et 2010, interrogations à 7 ans.
Champ : ensemble des jeunes, France métropolitaine.

Les résultats de la typologie de trajectoires appliquée aux trois Générations (figure 4) mettent bien en évidence les conséquences de ce contexte sur les parcours d'insertion professionnelle des jeunes débutants sur le marché du travail aux trois époques. On retrouve une diminution des trajectoires dominées par l'emploi entre les Générations 1998 et 2010 (de 84 % à 74 %). Les trajectoires de stabilisation en emploi à durée indéterminée sont moins nombreuses (de 66 % à 55 %), et ne sont pas compensées par les parcours enchaînant les emplois à durée déterminée, dont la part reste stable (18 % contre 19 %).

2.3.2. Évolution des parcours d'insertion selon le niveau de diplôme

La mobilisation de la typologie de trajectoires permet de montrer que la distribution des types de parcours professionnels évolue différemment d'une Génération à l'autre selon le niveau de diplôme détenu à la fin des études.

Tableau 1 • Trajectoires-types comparables par niveau de diplôme

	Non diplômés		Diplômés du secondaire		Diplômés bac+2 à +4		Diplômés bac+5 et plus	
	G1998	G2010	G1998	G2010	G1998	G2010	G1998	G2010
Stabilisation en EDI	41	21	61	49	79	74	88	79
<i>Stabilisation rapide en EDI</i>	20	8	33	27	51	47	70	60
<i>Stabilisation différée en EDI</i>	21	13	28	22	28	27	18	19
EDD durable	25	24	21	21	14	14	7	13
Chômage persistant ou récurrent	20	34	7	13	3	4	2	3
Longue(s) période(s) en formation ou reprise d'études	6	10	7	13	2	6	1	3
Inactivité durable	8	11	4	4	2	2	2	2
Ensemble	100	100	100	100	100	100	100	100

Source : Céreq, bases comparables Génération 1998, 2004 et 2010, interrogations à 7 ans.

Champ : ensemble des jeunes des Générations 1998 et 2010, France métropolitaine.

Parmi les sortants de formation initiale sans aucun diplôme, on observe une dégradation plus forte des parcours dans la Génération la plus récente. Les trajectoires marquées par un éloignement de l'emploi – chômage récurrent ou persistant (34 %) et inactivité durable (11 %) – sont aussi fréquentes dans la Génération 2010 que celles dominées par l'emploi, alors qu'elles ne concernaient que 28 % des jeunes sans diplôme de la Génération 1998, arrivés sur le marché du travail douze ans plus tôt. De plus, parmi les parcours dominés par l'emploi, on constate une nette diminution des trajectoires de stabilisation en EDI (21 % contre 41 % pour la Génération 1998), désormais moins nombreuses que les parcours caractérisés par un enchaînement durable d'EDD (24 %).

L'accès à l'emploi, et en particulier à l'EDI, plus difficile pour la Génération 2010 touche aussi les jeunes diplômés de l'enseignement secondaire. Une majorité d'entre eux continuent néanmoins de connaître les trajectoires dominées par l'emploi, notamment la stabilisation en EDI (49 %, contre 61 % pour la Génération 1998).

Pour les sortants de l'enseignement supérieur, l'effet du contexte économique apparaît de plus faible ampleur avec une stabilisation en EDI qui demeure la norme et concerne les trois quarts d'entre eux. Celle-ci n'en est pas moins plus tardive pour les jeunes de la Génération 2010 que pour leurs homologues de la Génération 1998, en particulier pour les sortants de niveaux bac+5 et plus. Les plus faibles évolutions observées pour les jeunes de niveau bac+2 à +4 en comparaison des autres niveaux peuvent s'expliquer par la création des licences professionnelles en 1999, ce qui a modifié la composition de cette catégorie de diplômés entre les Générations 1998 et 2010. En effet, les sortants diplômés de licences professionnelles sont très souvent issus de DUT ou de BTS. La licence professionnelle est donc souvent une année d'étude supplémentaire pour des jeunes qui, jusque-là (et donc dans la Génération 1998), sortaient du système éducatif au niveau bac+2.

L'analyse de l'évolution des trajectoires-types révèle également, à tous les niveaux de diplômes, une augmentation des parcours dominés par la reprise d'études au fil des Générations. Les non-diplômés et davantage encore les diplômés du secondaire sont toujours les plus concernés ; et le poids de cette trajectoire passe, pour ces derniers, de 7 % à 13 % entre les deux Générations (pour les sortants de niveau baccalauréat, ce taux passe même de 9 % à 17 %).

2.3.3. Évolution des parcours d'insertion selon le genre

Une analyse de la typologie comparable des trajectoires professionnelles selon le genre (tableau 2) fait apparaître que l'évolution des parcours-types se distingue selon le genre. Alors que les femmes connaissent moins souvent que les hommes un parcours de « stabilisation rapide en emploi à durée indéterminée (EDI) » dans chacune des Générations, la diminution de la part d'individus concernés au fil des enquêtes est plus forte chez les hommes que chez les femmes. L'écart selon le genre sur ce type de parcours se réduit ainsi au fil des cohortes. De plus, si les femmes et les hommes étaient presque aussi nombreux pour les Générations 1998 et 2004 à avoir un parcours dominé par un chômage persistant ou récurrent, ces derniers sont plus souvent concernés pour la Génération 2010. Enfin, la proportion de jeunes dans des parcours marqués par de la formation ou de la reprise d'études a doublé chez les hommes et presque autant pour les femmes (1,7 fois plus).

Tableau 2 • Trajectoires-types comparables selon le genre

	Hommes			Femmes		
	G1998	G2004	G2010	G1998	G2004	G2010
Stabilisation en EDI	71	65	56	60	59	53
<i>Stabilisation rapide en EDI</i>	45	38	36	34	33	32
<i>Stabilisation différée en EDI</i>	26	27	21	26	26	22
EDD durable	17	18	18	19	17	19
Chômage persistant ou récurrent	7	10	15	8	9	11
Longue(s) période(s) en formation ou reprise d'études	4	5	8	6	9	10
Inactivité durable	1	2	3	7	6	6
Ensemble	100	100	100	100	100	100

Source : Céreq, bases comparables Génération 1998, 2004 et 2010, interrogations à 7 ans.

Champ : ensemble des jeunes, France métropolitaine.

2.3.4. Évolution des parcours d'insertion selon la région de formation

Pour conclure cette partie dédiée aux exemples de mobilisation de la typologie comparable de parcours d'insertion professionnelle, nous présentons ci-dessous l'évolution des parcours en fonction de la région de fin d'études des individus. La représentation cartographique (figure 5) permet de mettre en relief les disparités dans les parcours d'insertion selon la région formatrice¹¹. En se basant sur la Génération 2010, quelques résultats des disparités régionales peuvent être pointés :

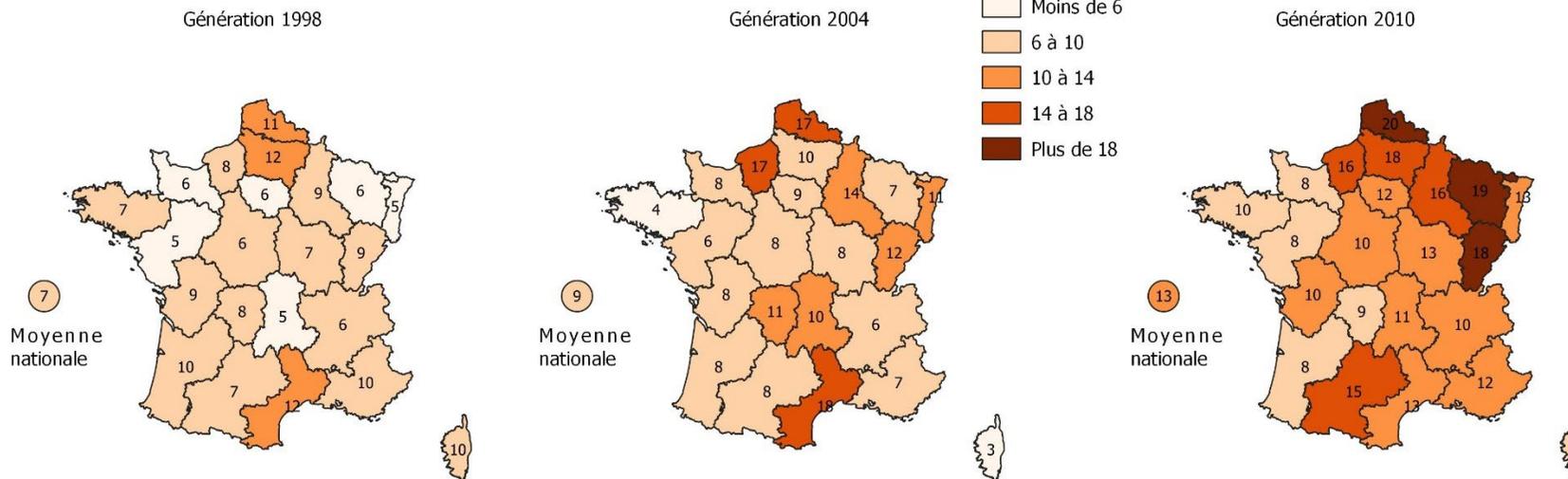
- On remarque que les régions Franche-Comté, Lorraine, Nord-Pas-de-Calais et Picardie se caractérisent par une part importante de jeunes dont le parcours sur les sept premières années de vie active est marqué par du chômage durable ou récurrent (autour de 19 % de jeunes concernés) alors que ce taux est le plus bas pour les régions Aquitaine, Basse-Normandie et Pays-de-la-Loire (8 %).
- Si l'on se focalise sur les trajectoires dominées par une stabilisation en emploi à durée indéterminée, la part est la plus faible dans les régions Haute-Normandie, Nord-Pas-de-Calais, Provence-Alpes-Côte-d'Azur et Picardie (autour de 47 % de jeunes concernés) alors que ce taux est le plus élevé pour les régions Basse-Normandie (64 %) et Midi-Pyrénées (62 %).

¹¹ Le point géographique de référence est la région du dernier établissement de formation fréquenté. Une partie des individus a changé de région au cours de leurs sept premières années de vie active.

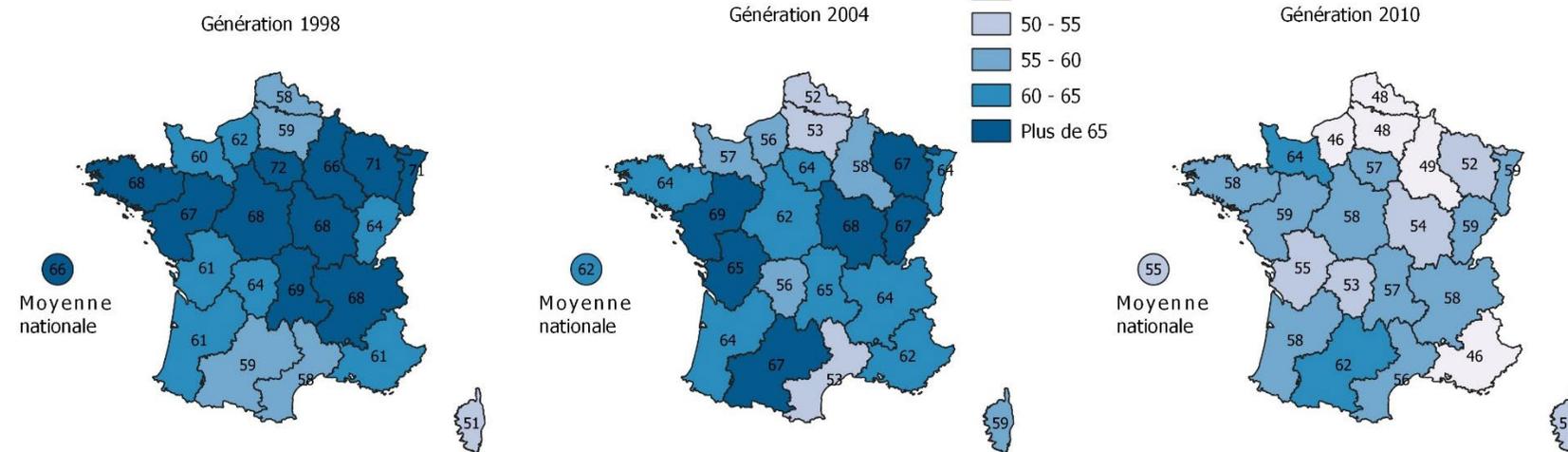
Toutes les régions n'ont pas connu les mêmes évolutions entre les trois enquêtes Génération étudiées. Le cas de la région Lorraine se singularise car elle présente l'évolution de la plus forte ampleur sur les deux indicateurs évoqués entre les Génération 1998 et 2010 : -19 points sur la part des parcours de stabilisation en EDI et +13 points sur la part des parcours dominés par du chômage durable ou récurrent. Cependant, d'autres régions sont marquées par des évolutions en sens inverse ; la région Basse-Normandie connaît une augmentation de la part de stabilisation en EDI (+4 points, la plus forte de toutes les régions) quand la région Aquitaine connaît une baisse de la part de parcours marqués par du chômage durable ou récurrent (-2 pts, baisse la plus forte alors même que la tendance nationale est à l'augmentation du fait de la dégradation de la conjoncture économique au fil des enquêtes).

Figure 5 • Typologie de trajectoire selon la Génération, par région de formation

Taux de chômage durable ou récurrent (en %)



Taux de stabilisation en EDI (en %)



Source : Céreq, bases comparables Génération 1998, 2004 et 2010, interrogations à 7 ans.
Champ : ensemble des jeunes, France métropolitaine.

Conclusion

Les typologies de trajectoires constituent un outil très utile pour toute recherche s'intéressant à des données longitudinales complexes telles que les parcours professionnels. Elles permettent en effet de proposer une description synthétique de ces données et surtout, en prenant en compte différentes situations et leurs enchaînements dans le temps, de repérer les traits communs entre les parcours des individus et d'identifier les principales trajectoires empruntées. Cette méthode peut s'appliquer à des thèmes de recherche et données variés en sciences sociales et permet la comparaison entre individus au sein d'une même cohorte. Elle permet aussi de réaliser des comparaisons entre individus de cohortes différentes. Dans notre cas, la typologie des parcours comparables construite sur les données des Générations 1998, 2004 et 2010 donne à voir l'évolution des parcours d'insertion professionnelle dans le temps entre des cohortes de jeunes arrivés sur le marché du travail à plusieurs années d'intervalles.

La production d'une telle typologie contribue ainsi à une meilleure connaissance des cheminements des individus (dans notre cas, du processus d'insertion professionnelle), difficilement analysable en étudiant les différents évènements du parcours de manière indépendante.

La possibilité de croiser cet indicateur synthétique avec d'autres variables (genre, niveau de diplôme, indicateurs territoriaux, etc.) élargit les perspectives d'analyse. Dans cette étude, nous avons choisi d'illustrer par quelques statistiques descriptives les liens entre les types de parcours et leurs caractéristiques individuelles. Mais on peut imaginer bien d'autres applications possibles comme la mise en œuvre de modèles statistiques pour analyser les déterminants des divers parcours mis en évidence par la typologie, ou dans lesquels les types obtenus sont mobilisés comme variables explicatives. De plus, cette méthode peut être vue comme un outil pouvant aider à cibler des publics vulnérables, notamment pour l'élaboration de politiques publiques.

La réalisation d'une typologie nécessite d'être adaptée en fonction de l'objet de l'analyse, en gardant à l'esprit que certains choix méthodologiques effectués ne sont pas anodins et pourraient influencer le résultat (Giret *et al.*, 2011). Chaque méthode a des paramètres propres qui sont déduits d'hypothèses dont on ne peut garantir qu'ils soient optimaux mais nécessitent d'être justifiés.

Bibliographie

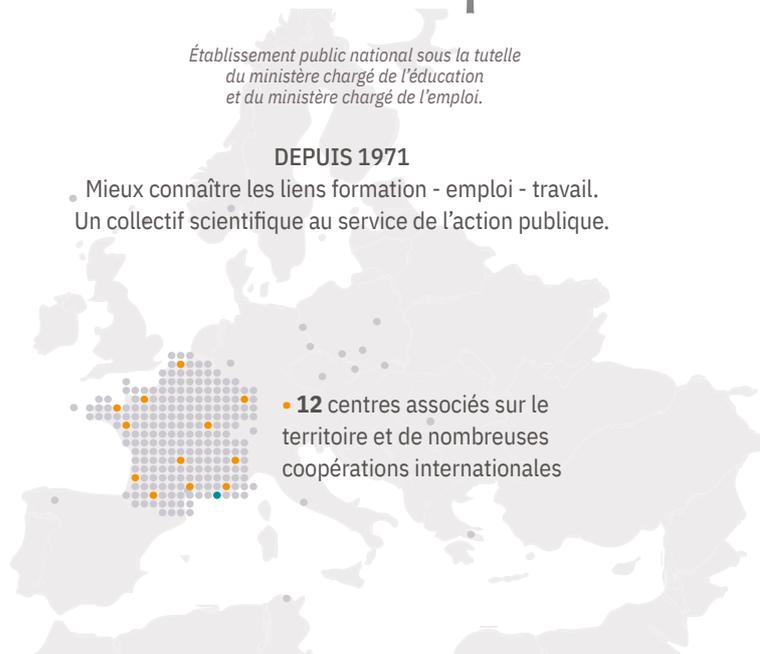
- Cazes, P. (1980). L'analyse de certains tableaux rectangulaires décomposés en blocs. *Les cahiers de l'analyse des données*, 5 (4), 387-403.
- Chessel, D. & Lobry, J. R. (2017). Comment projeter des individus ou des variables supplémentaires ? Consultation statistique avec le logiciel R.
- Cisse, M., Gaubert, E., Mazari, Z. & Olaria, M. (2019). *Enquête Génération 2010 Interrogation 2017. Bilan méthodologique*. Céreq, collection « Études » (n° 27).
- Deville, J.-C. & Saporta, G. (1979). « Analyse harmonique qualitative ». In E. Diday (ed.), *Data Analysis and Informatics* (p. 375-389). North-Holland.
- Epiphane, D., Mazari, Z., Olaria, M., & Sulzer, E. (2019). Des débuts de carrière plus chaotiques pour une génération plus diplômée - Génération 2010. *Céreq Bref*, 382.
- Grelet, Y. (2002). *Des typologies de parcours. Méthodes et usages*. Céreq, coll. « note de travail Génération 92 » (n° 20).
- Le Lan, R. (2005). « Analyse de données et classification sur données d'enquête. Choix sur les variables, le nombre de classes et le nombre d'axes ». *Actes des journées de méthodologie statistiques*.
- Malhotra, N. (2004). *Études marketing avec SPSS*. Paris : Pearson éducation France.
- Milin, K. (2018). CDD, CDI : comment évoluent les embauches et les ruptures depuis 25 ans ? *DARES analyses*, n°026.
- Mora, V. (2018). « Comment les conditions d'insertion des jeunes se sont-elles transformées en 20 ans ? » In T. Couppié, A. Dupray, D. Epiphane et V. Mora (eds), *20 ans d'insertion professionnelle des jeunes : entre permanences et évolution* (p. 51-59). Céreq, coll. « Essentiels » (n° 1).
- Robette, N. (2011). Explorer et décrire les parcours de vie. Les typologies de trajectoires. CEPED.
- Robette, N. (2014). L'étude des parcours de vie, démarche descriptive ou causale ? *Idées économiques et sociales*, 177 (3), 8-15.
- Saporta, G. (1996). « L'analyse harmonique qualitative, une synthèse de la théorie ». Communication au colloque Relección y análisis de datos longitudinales, Universidad Nacional de Colombia & ORSTOM, Bogota, Colombie. (p. 111-120).
- Stéphanus C. (2018). « Construction d'une typologie de trajectoires professionnelles à partir de données de l'enquête Défis ». Communication aux 13es Journées de méthodologie statistique de l'Insee (JMS), 12-14 juin 2018, Paris.

Céreq

*Établissement public national sous la tutelle
du ministère chargé de l'éducation
et du ministère chargé de l'emploi.*

DEPUIS 1971

Mieux connaître les liens formation - emploi - travail.
Un collectif scientifique au service de l'action publique.



 **+ d'infos**
et tous les travaux

À explorer
www.cereq.fr



 **+ de 600 publications**
Accessibles librement