

# Enregistrer, quantifier et contrôler les migrations dans une ville industrielle du Luxembourg : Dudelange, fin XIX<sup>e</sup>-XX<sup>e</sup> siècles

Alex RELICOVSKI

59

L'immigration dans les villes luxembourgeoises à partir de la fin du XIX<sup>e</sup> siècle a depuis plusieurs années attiré l'attention des chercheurs et les travaux disponibles sont désormais importants. La question centrale a été celle de comprendre la croissance urbaine en lien avec l'arrivée des migrants en ville, mais les études portant sur la mobilité intra-urbaine et l'insertion des nouvelles populations dans le tissu urbain sont-elles peu nombreuses. Les sources mobilisées, issues le plus souvent des recensements de population et des estimations des arrivées et des départs établies par les communes, de même que les sources démographiques qui restent à dépouiller, demeurent éparpillées et lacunaires<sup>1</sup>. En résulte que l'historiographie en question, bien qu'en plein essor, obéit à des logiques nationales, car il est difficile d'établir avec justesse les lieux d'arrivée et de sortie des migrants à partir des recensements, qui ne mentionnent que l'origine nationale des migrants. Pour l'histoire contemporaine, les recensements se font tous les cinq ans après 1875, mais se poursuivent à une cadence moins régulière au cours du XX<sup>e</sup> siècle<sup>2</sup>. Ces derniers indiquent l'origine géographique des individus, notamment à partir du recensement de 1871, le premier à mentionner leur nationalité, mais ne donnent aucune information sur leurs mouvements vers et au-delà de la ville. Pour mesurer cette mobilité, il convient de connaître avec précision les adresses successives des individus, et de croiser les statistiques officielles et leur évolution dans le temps avec des sources nouvelles, émises à d'autres échelles de l'État. Au tournant du siècle, la ville de Dudelange, au sud du pays, a accumulé de nombreux documents relatifs à l'identification de ses résidents. Son développement rapide lié à l'industrie sidérurgique, qui prend son essor dans la région de la Minette dans la seconde moitié

1. Dans l'ouvrage issu de sa thèse consacrée à l'étude des dossiers de naturalisation au Luxembourg, Denis Scuto résume ainsi cette situation : « On ne dispose donc, pour évaluer l'immigration, que d'informations éparpillées, discontinues, issues de sources aussi différentes que les documents communaux, les archives judiciaires, les relevés de la Bienfaisance publique, les témoignages de contemporains... ». Cf. SCUTO Denis, *La nationalité luxembourgeoise (XIX<sup>e</sup>-XX<sup>e</sup> siècles). Histoire d'un alliage européen*, Bruxelles, Éditions de l'Université de Bruxelles, 2012, p. 72.

2. LEHNERS Jean-Paul, « À propos de la démographie historique au Grand-Duché de Luxembourg : sources, problèmes, perspectives, bibliographie », dans Jean-Claude MULLER (éd.), *Annuaire - Jahrbuch* 1993, Luxembourg, Imprimerie Centrale S.A., 1993, p. 259.

du XIX<sup>e</sup> siècle<sup>3</sup>, présente un cas de figure similaire à d'autres « villes-usine<sup>4</sup> » de la Grande Région, caractérisées par leur forte mobilité liée à un contexte de grands flux de main-d'œuvre frontalière. Dans le cas français, les exemples de Longwy et de Jœuf, étudiés respectivement par Gérard Noiriel<sup>5</sup> et Piero Galloro<sup>6</sup>, ont fait école. Au Luxembourg, la recherche sur les migrations s'est structurée plus récemment. Si les premiers travaux remontent aux années 1970-1980 et concernent surtout les mouvements d'émigration, les chantiers de recherche initiés par Denis Scuto et d'autres historiens, à partir des années 1990-2000, ont contribué à l'émergence d'un champ en voie d'institutionnalisation, visant à décrire les migrations dans leur ensemble<sup>7</sup>. Alors que certains travaux insistent à présent sur l'aspect transnational du phénomène migratoire dans un espace avant tout frontalier<sup>8</sup>, les études centrées sur les groupes nationaux, comme les Italiens<sup>9</sup>, ou plus récemment les Russes<sup>10</sup>, demeurent l'axe privilégié par la recherche. En parallèle, des recherches commencent à exploiter des fonds d'archives unifiés, comme les formulaires de déclaration d'arrivée dans les communes du sud du pays, pour reconstituer les trajectoires migrantes de manière transversale, dans une perspective d'utilisation critique des outils numériques<sup>11</sup>. Ces approches mériteraient d'être étendues à d'autres périodes et à un corpus plus large. C'est là l'apport que nos sources espèrent amener, à partir de l'étude des mobilités urbaines, vues sous l'angle administratif à l'échelle locale. Ces dernières constituent un fonds encore inexploré mais dont l'étude s'avère essentielle pour une meilleure compréhension de la mobilité des individus, à Dudelange sur le temps-long : il s'agit des plus de 50 000 cartes d'enregistrement des résidents dans la commune, dont la production va de la fin du XIX<sup>e</sup> siècle aux années 1970, à l'exception de la période d'occupation allemande de 1940-1945. Ces archives urbaines ont bénéficié des efforts de numérisation entrepris ces dernières années par l'université du Luxembourg pour aboutir à un traitement informatique des données sur les migrations dans le pays<sup>12</sup>.

3. LEBOUTTE René, PUISSANT Jean et SCUTO Denis, *Un siècle d'histoire industrielle (1873-1973)*, Paris, Éditions Sedes, 1998.

4 Ce qualificatif est néanmoins à nuancer. Malgré la centralité de l'ARBED pour le développement industriel de Dudelange, la ville regroupe en réalité une pluralité d'usines et de fabriques différentes, mobilisant une main d'œuvre variée. Pour une étude de cas récente, cf. JACQUOT Lionel et al. (dir.), *Bataville (1931-2001). Ville-usine de la chaussure*, Grenoble, Presses Universitaires de Grenoble, 2023.

5 NOIRIEL Gérard, *Immigrés et prolétaires, Longwy, 1880-1980*, Marseille, Éditions Agone, 2019 (1ère éd. 1984).

6 GALLORO Piero D., *La main-d'œuvre des usines sidérurgiques de Lorraine, 1880-1939. Étude des flux. L'exemple des Forges de Jœuf*, thèse d'histoire, dir. Alain Wahl, université de Metz, 2 vol., 1993-1996 ; GALLORO P., *Ouvriers du fer, princes du vent. Histoire des flux de main-d'œuvre dans la sidérurgie lorraine (1880-1939)*, Metz, Éditions Serpenoise, 2001.

7. SCUTO D., « Historiographie de l'immigration au Luxembourg », *Hémécht. Revue d'Histoire luxembourgeoise*, vol. 60, no. 3-4, 2008, p. 391-413.

8 LEINER Stefan, *Migration und Urbanisierung. Binnenwanderungsbewegungen, räumlicher und sozialer Wandel in den Industriestädten des Saar-Lor-Lux-Raumes (1856-1910)*, Saarbrücken, Kommissionsverlag: Saarbrücker Druckerei und Verlag GMBH, 1994 ; SAUER Arnaud, *Entre les lignes. Flux, mobilités et réseaux de la main-d'œuvre étrangère dans le bassin transfrontalier de la Minette durant l'entre-deux-guerres*, thèse d'histoire, dir. Denis Scuto, ENS Paris/université du Luxembourg, 2024 ; KACI Maxime et al., « L'espace frontalier franco-luxembourgeois du Moyen Âge à nos jours : nouveaux éclairages sur les interactions entre pouvoirs et pratiques », *Hémécht*, vol. 72, no. 2, 2025, p. 133-138.

9 GALLO Benito, *Les Italiens au Grand-Duché de Luxembourg. Un siècle d'histoire et de chroniques sur l'immigration. Luxembourg*, Imprimerie Saint-Paul, 1987. Cf. également la thèse de Sophie Schram, la première à aborder l'histoire des migrations féminines dans la région à partir de sources d'histoire orale, mais se limitant aux Italiennes de Dudelange : SCHRAM Sophie, *Weibliche Lebenswelten und Erinnerungskulturen. Das Beispiel « Kleinitalien » – ein Einwandererviertel in Luxemburg*, Opladen – Berlin – Toronto, Budrich UniPress Ltd., 2016.

10 GANSCHOW Inna, *100 Jahre Russen in Luxemburg. Geschichte einer atomisierten Diaspora*, Luxembourg, Fondation Lydie Schmit, 2020.

11 VENKEN Machteld et SAUER Arnaud, « Arrival declaration forms. A new gateway for mapping migration to Luxembourg », *Frontiers. Hum. Dyn.*, vol. 4, 2022, p. 1-19 ; VENKEN Machteld et JACQUET David, « Remigration to Luxembourg. Examining a New Research Question by Means of Digital Hermeneutics », dans Sarah OBERBLICHER, Eva PFANZELTER et Valerio LARCHER (dir.), *Return and Circular Migration in Contemporary European History*, Berlin, De Gruyter, 2022, p. 67-91.

12. Notre thèse bénéficie du soutien de la commune de Dudelange, dans le cadre du projet Local Sources qui vise à la prise en compte et à la valorisation de sources locales pour l'étude de l'histoire contemporaine du Luxembourg : [<https://www.uni.lu/c2dh-en/research-projects/local-sources>].

L'exploitation de ce fonds d'archives soulève plusieurs questions. D'une part, il rend nécessaire l'élaboration d'une chaîne de traitement souple et documentée pour l'extraction et l'analyse des données, nous permettant d'en relever les étapes, qui vont de la transcription des sources à leur mise en base de données. Loin d'être un processus linéaire, cette opération est néanmoins nécessaire, à l'heure où des voix parmi les *digital historians* en appellent à plus de transparence dans le travail quotidien sur les données, qui passe par un retour réflexif nécessaire sur les méthodes utilisées<sup>13</sup>. D'autre part, il révèle l'enjeu de travailler à partir d'un corpus d'archives extensif, difficilement appréhendable par la seule main humaine. Nous rejoignons alors les discussions de Claire Lemerrier et Claire Zalc à propos du « vertige<sup>14</sup> » induit par le travail sur des bases de données étendues. En évoquant le travail pionnier des *Galériens de Louis XIV*, les autrices disent bien : « André Zysberg traite 60 000 fiches de galériens, quand 1 000 auraient suffi à son propos<sup>15</sup> ». Si les méthodes dont nous disposons aujourd'hui semblent à première vue pouvoir nous aider à dépasser un tel écueil par les possibilités de traitement qu'elles offrent, leur usage suppose toujours une médiation, dont il faut interroger les effets sur la construction du corpus et son interprétation. Dans cet article, nous cherchons donc à montrer qu'il est possible d'atteindre un tel objectif, mais au prix de moins de contrôle sur les données, et au risque de reproduire des effets de sources.

## I. Fichiers nominatifs et archives urbaines au service des États à partir de la fin du XIX<sup>e</sup> siècle

L'étude des dispositifs matériels de gestion du savoir dans les sociétés européennes, à la croisée de l'histoire intellectuelle et de l'histoire et de la sociologie des sciences, connaît actuellement des renouvellements. Si elle a longtemps plutôt préoccupé les modernistes<sup>16</sup>, les enjeux qu'elle analyse se sont étendus depuis plusieurs années à l'histoire contemporaine, notamment sous l'impulsion des travaux de Bruno Latour sur l'écrit comme facteur de stabilisation du savoir scientifique et administratif<sup>17</sup>. Le champ des *paper studies*<sup>18</sup>, ainsi que les tournants matériels (*material turn*) et

13 DU MOUZA Cédric, LAMASSÉ Stéphane et RYGIEL Philippe, « De l'histoire numérique à l'histoire données ? », *Les Cahiers de Framespa*, 42, 2023 ; HOESTRA Rik et KOOLEN Marijn, « Data scopes for digital history research », *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 52, n° 2, 2018, p. 1-16 ; REISINGER Gunther, « Digital source criticism: net art as a methodological case study », dans Dieter DANIELS et Gunther REISINGER (dir.), *Netpioneers 1.0: Contextualizing Early Net-Based Art*, Berlin, Sternberg Press, 2010, p. 123-142 ; KARILA-COHEN Karine et al., « Nouvelles cuisines de l'histoire quantitative », *Annales HSS*, vol. 73, n° 4, 2018, p. 771-783.

14 LEMERCIER Claire et ZALC Claire, *Méthodes quantitatives pour l'historien*, Paris, Éditions La Découverte, coll. « Repères », 2008, p. 12.

15 *Id.*  
16 DENIS Vincent et LACOUR Pierre-Yves, « La logistique des savoirs. Surabondance d'informations et technologies de papier au XVIII<sup>e</sup> siècle », *Genèses*, vol. 102, n° 1, 2016, p. 107-122 ; KAFKA Ben, *The Demon of Writing: Powers and Failures of Paperwork*, Princeton, Princeton University Press, 2020 (1<sup>ère</sup> éd. 2002). Voir également la synthèse de WACQUET Françoise, *L'ordre matériel du savoir. Comment les savants travaillent, XVI<sup>e</sup>-XXI<sup>e</sup> siècles*, Paris, CNRS Éditions, 2022, (1<sup>ère</sup> éd. 2015), p. 72-85.

17 Voir à titre d'exemple LATOUR Bruno et WOOLGAR Steve, *La vie de laboratoire. La production des faits scientifiques*, Paris, Éditions La Découverte, 1996 (1<sup>ère</sup> éd. 1979) ; LATOUR Bruno, *La science en action. Introduction à la sociologie des sciences*, Paris, Éditions La Découverte, 1989 (1<sup>ère</sup> éd. 1987) ; *Id.*, *La fabrique du droit. Une ethnographie du conseil d'État*, Paris, Éditions La Découverte, 2004 (1<sup>ère</sup> éd. 2002).

18 Pour une bonne synthèse en allemand, voir HESS Volker et MENDELSON J. Andrew, « Paper Technology und Wissensgeschichte », *NTM: Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin*, vol. 21, 2013, p. 1-10.

informationnels (*information turn*) dans l'histoire des savoirs à l'échelle mondiale<sup>19</sup> nous intéressent ici, puisqu'ils permettent de considérer la matérialité de sources comme un objet d'analyse en soi, en tant qu'infrastructure déterminante dans la constitution des États modernes. Ces derniers traversent une période, au tournant des XIX<sup>e</sup> et XX<sup>e</sup> siècles, d'homogénéisation, à l'échelle européenne, de leurs techniques administratives, liée au développement de la bureaucratie<sup>20</sup>. Un des objets importants de cette « révolution de papier<sup>21</sup> » est le fichier nominatif, dont le support est la fiche<sup>22</sup>. Produite en série, elle assigne à chaque individu une identité déterminée par des classifications forgées par les agents étatiques, grilles de lecture du monde social dont il convient d'interroger l'historicité, la validité et l'évolution dans le temps. Dans l'histoire des migrations, la période est marquée par une nationalisation des pratiques d'identification, liée à l'apogée de l'État-nation territorial<sup>23</sup>. Fondée sur une distinction entre nationaux et étrangers, notamment pour l'accession aux droits sociaux, la construction de l'existence juridique des individus inclus dans cette nouvelle forme d'organisation socio-politique dépend des grilles forgées en amont par les administrations<sup>24</sup>. Dans la pratique, cela se voit au regard des nouveaux types de documents qui émergent aux différents échelons de l'État. Au-delà du passeport et de la carte d'identité, on voit apparaître les premiers registres de police des étrangers ainsi que les documents d'enregistrement qui les complètent. Leurs modalités de classement, d'archivage et de consultation permettent aux administrations d'accéder facilement à l'état civil des individus enregistrés. Implémentées au niveau de la ville, il y a interrelation entre histoire urbaine et histoire de la régulation des migrations, la nation reposant sur les autorités locales pour l'identification des individus<sup>25</sup>. Ainsi, les formes d'application des réglementations sur les étrangers doivent être comprises à la lumière des archives produites par les administrations communales. Le « point de vue communal<sup>26</sup> » est donc à prendre en compte. Pourtant, à l'échelle des administrations locales, peu d'études se sont intéressées aux pratiques concrètes de contrôle des migrations<sup>27</sup>, bien que les circuits administratifs

19 NIJENHUIS Ida *et al.* (dir.), *Information and Power in History. Towards a Global Approach*, Londres, Routledge, 2020 ; CHANDLER Alfred D. Jr. et CORTADA James (dir.), *A Nation transformed by Information. How Information has shaped the United States from Colonial Times to Present*, Oxford – New York, Oxford University Press, 2000.

20 DREYFUS Françoise, *L'Invention de la bureaucratie. Servir l'État en France, en Grande-Bretagne et aux États-Unis (XVIII<sup>e</sup>-XX<sup>e</sup> siècles)*, Paris, Éditions La Découverte, 2000.

21 GARDEY Delphine, *Écrire, calculer, classer. Comment une révolution de papier a transformé les sociétés contemporaines, 1800-1940*, Paris, Éditions La Découverte, 2008.

22 ABOUT Ilsen et DENIS Vincent, *Histoire de l'identification des personnes*, Paris, Éditions La Découverte, coll. « Repères », 2010, p. 60 ; KRAJEWSKI Markus, *Paper Machines. About Cards & Catalogues, 1548-1929*, Cambridge – Londres, The MIT Press, 2011 (1<sup>ère</sup> éd. 2002) ; VAN FAASSEN Marijke et OPREL Marieke, « Paper trails to private lives. The performative power of card indexes through time and space », dans I. Nijenhuis *et al.* (dir.), *op. cit.*, p. 254-273.

23 ABOUT I. et DENIS V., *op. cit.*, p. 56-70.

24 ROSENTAL Paul-André, « Migrations, souveraineté, droits sociaux. Protéger et expulser les étrangers en Europe du XIX<sup>e</sup> siècle à nos jours », *Annales HSS*, vol. 66, no. 2, 2011, p. 335-373.

25 GREEFS Hilde et WINTER Anne (dir.), *Migration Policies and Materialities of Identification in European Cities. Papers and Gates, 1500-1930s*, Abingdon, Routledge, 2019.

26 DE BARROS Françoise, « Les acteurs municipaux et "leurs" étrangers (1919-1984) : gains et contraintes d'un détour communal pour l'analyse d'un travail de catégorisation étatique », *Genèses*, n° 72, vol. 3, 2008, p. 42.

27 À titre d'exemple, cf. VIAL Éric, « Pratiques d'une préfecture : les demandes d'expulsion de ressortissants italiens dans l'Isère de 1934 à la Seconde Guerre mondiale », dans Marie-Claude BLANC-CHALÉARD *et al.* (dir.), *Police et migrants. France 1667-1939*, Rennes, Presses Universitaires de Rennes, 2001, p. 167-180. Cf. également ZALC Claire, *Dénaturalisés. Les retraits de nationalité sous Vichy*, Paris, Éditions du Seuil, 2016.

soient aujourd'hui mieux connus, notamment pour le second XX<sup>e</sup> siècle<sup>28</sup>. Comme le rappelle Delphine Diaz : « Les sources étatiques doivent être prises pour ce qu'elles sont : des archives foisonnantes, produites par et pour le contrôle<sup>29</sup> ». Avec la mise en place de l'État-nation, les communes ont bien perdu de l'autonomie héritée de l'époque moderne, au profit d'une centralisation incarnée par les préfetures, les ministères ou les commissions nationales. Mais dans le cas de la création du migrant comme objet juridique à partir de la fin du XIX<sup>e</sup> siècle, celui-ci est désigné avant tout au niveau local<sup>30</sup>, par des pratiques classificatoires (attribution d'un statut réglementaire, dénombrement), par deux types distincts de tâches : celles qui appliquent aux individus des règles administratives, qu'elles soient juridiques ou techniques, et celles qui contrôlent la bonne application de ces règles. Dans le cas luxembourgeois, le service de la police des étrangers est centralisé au parquet de la cour depuis mai 1884, mais l'application effective des mesures prises par ce dernier restent, au moins en théorie, débattues au niveau communal<sup>31</sup>. C'est notamment le cas à Dudelange.

## II. Dudelange, ville industrielle en pleine expansion au tournant du XX<sup>e</sup> siècle

Dudelange passe de 1600 habitants en 1880 à plus de 10 000 habitant·e·s dans les premières années du XX<sup>e</sup> siècle<sup>32</sup>. La commune est élevée au statut de ville par ordonnance grand-ducale, le 4 août 1907<sup>33</sup>. Avec la fondation de la Société des Hauts Fourneaux et Forges de Dudelange (SHFFD) le 5 juillet 1882, des ouvrières et ouvriers non-qualifié·e·s, souvent issu·e·s du secteur agricole, affluent vers les sites de construction de la future usine, qui ne cessera ensuite de s'agrandir. Tournée dès l'origine vers la production d'acier, dont l'exportation profite de l'aménagement du chemin de fer au Luxembourg à partir de la fin du XIX<sup>e</sup> siècle, l'usine continue de s'agrandir au tournant du siècle, aboutissant à la création des Acières réunies de Burbach-Eich-Dudelange (ARBED) en 1911-1912<sup>34</sup>. Devenu un des premiers exploitants d'acier européen, l'ARBED est dès le départ tourné vers l'Allemagne du Zollverein, union douanière à laquelle le Luxembourg, sous pression, se rallie en 1842, et notamment vers la Sarre, ainsi que vers la Lorraine des Wendel<sup>35</sup>. Les

28 BRUNO Anne-Sophie, RYGIEL Philippe, SPIRE Alexis et ZALC Claire, « Jugés sur pièces. Le traitement des dossiers de séjour et de travail des étrangers en France (1917-1984) », *Population*, vol. 61, n° 5, 2006, p. 737-762.

29 DIAZ Delphine, *En exil. Les réfugiés en Europe, de la fin du XVIII<sup>e</sup> siècle à nos jours*, Paris, Éditions Gallimard, coll. « Folio Histoire », 2021, p. 24.

30 NOIRIEL Gérard, « Surveiller les déplacements ou identifier les personnes ? Contribution à l'histoire du passeport en France de la I<sup>re</sup> à la III<sup>e</sup> République », *Genèses*, vol. 30, 1998, p. 82-84.

31 SCUTO D., *op. cit.*, p. 150. Voir notamment les circulaires du ministère de l'Intérieur relatives aux écritures administratives : Archives nationales de Luxembourg (ANLux), fonds du ministère de l'Intérieur, INT-0012, dossier « Simplification des écritures (1911-1924) ».

32 STATEC, *Statistiques historiques 1839-1989*, Luxembourg, Éditpress, 1990, p. 15.

33 NEUMAN Henri, *Les Communes. Leur administration, leurs finances et leurs domaines (1840-1891)*, Luxembourg, Imprimerie V. Bück, 1894 ; CONRADY Jean-Pierre et KREPS Robert, *Dudelange. Passé et présent d'une ville industrielle*, t. 1, Luxembourg, Éditpress, 1991, p. 114.

34 CHOMÉ Félix, ARBED. *Un demi-siècle d'histoire industrielle, 1911-1964*, Luxembourg, Éditions ARBED, 1972 (1<sup>re</sup> éd. 1964).

35. ROTH François, « Réflexions sur l'industrialisation de la Lorraine et du Luxembourg (1850-1950) », *Hémécht*, vol. 46, n° 3, 1994, p. 638-639.

seules aciéries de l'ARBED mobilisent ainsi plus de 60% des travailleur·euse·s industriel·le·s du pays en 1913, dont une part majoritaire est étrangère<sup>36</sup>. Cette industrialisation rapide entraîne une transformation profonde de la composition démographique de Dudelange<sup>37</sup>. Ainsi, une population mobile, souvent peu qualifiée, vient répondre aux besoins croissants d'une industrie lourde en pleine expansion. Face à cette croissance, l'État luxembourgeois met en place dès 1893 un dispositif de surveillance des étranger·ère·s : leur enregistrement dans la commune au plus tard deux jours après leur arrivée est décidé par la loi du 30 décembre sur la police des étrangers. Reprenant celle de 1880, la nouvelle loi y ajoute la nécessité pour tout individu migrant de remplir une déclaration d'arrivée détaillée (Fig. 1), renseignant l'état civil et les changements de résidence antérieurs et futurs<sup>38</sup>. La déclaration d'arrivée vient ainsi compléter les plus de 600 000 fiches individuelles spécifiquement dédiées aux migrant·e·s pour la période 1880-1940, conservées aux archives du Ministère de la justice, section « Police des étrangers<sup>39</sup> ». Ces dernières sont en usage jusqu'aux années 1970. Ces mesures rejoignent les dispositions analogues prises en

Fig. 1 : Déclaration d'arrivée à Dudelange d'un migrant d'origine italienne (Archives communales de Dudelange, n. p.)

France la même année, suite à la loi sur la « protection du travail national<sup>40</sup> », témoignant de la méfiance des autorités pour ces populations mobiles. Elles sont modifiées à plusieurs reprises au cours de la période. La Première Guerre mondiale marque un tournant<sup>41</sup>. Comme en France, les sorties de guerre de 1918-1919 renforcent les suspicions envers les populations étrangères<sup>42</sup>. Dans les années 1920, le Luxembourg durcit encore ses politiques migratoires, avec des arrêtés successifs imposant des autorisations de travail, des pièces justificatives, ou encore une surveillance sanitaire. Cette logique de filtrage repose sur une tension entre les besoins économiques du pays, en pleine reconstruction, et une volonté de contrôle : ainsi en 1929, les étranger·ère·s représentent 40% de la main-d'œuvre dans les usines, avant de retomber à 25% en 1933<sup>43</sup>. L'année d'après, la carte d'identité est rendue obligatoire

pour tout individu migrant<sup>44</sup>. Alors que les mouvements de population étaient

36. SCUTO D., « Histoire des immigrations au Luxembourg (XIX<sup>e</sup>-XXI<sup>e</sup> siècles) », dans *25 ans d'action pour l'immigration, 1985-2010*, OGBL, 25<sup>ème</sup> anniversaire du Département des Immigrés, Luxembourg, OGBL, 2010, p. 25.

37. TRINKAUS Fabian, *Arbeitereistenzen und Arbeiterbewegung in den Hüttenstädten Neunkirchen, Saar und Düdelingen, Luxemburg: 1880-1935/40: ein historischer Vergleich. Saarbrücken*, Kommission für Saarländische Landesgeschichte, 2014.

38. Loi du 30 décembre 1893, concernant la police des étrangers, *Journal officiel du Grand-Duché de Luxembourg*, Mémorial A n<sup>o</sup>. 1, 4 janvier 1894, p. 1-5.

39. ANLUX, min. Justice, Police des étrangers, « Fichier alphanumérique ». Cité dans SCUTO D., *La Nationalité luxembourgeoise*, op. cit., p. 148.

40. NOIRIEL Gérard, *La Tyrannie du national. Le droit d'asile en Europe (1793-1993)*, Paris, Calmann-Lévy, 1991, p. 169-176 ; *Id.*, *État, nation et immigration. Vers une histoire du pouvoir*, Paris, Éditions Gallimard, coll. « Folio histoire », 2001, p. 301-302.

41. RYGIEL Philippe, *Le Temps des migrations blanches. Migrer en Occident (1840-1940)*, Montreuil, Aux lieux d'être, 2007, p. 127-128. Pour la France, cf. la synthèse de DORNEL Laurent, *Les Étrangers dans la Grande Guerre*, Paris, Musée de l'Histoire de l'immigration – La Documentation française, 2014. Pour le Luxembourg, cf. SCUTO D., « Identifier, contrôler, réprimer. L'effet de 14-18 sur la circulation des personnes », *Mutations. Mémoires et perspectives du Bassin Minier*, vol. 10, 2018, p. 133-143.

42. Cf. à ce sujet VIDAL-NAQUET Clémentine, « Une si longue sortie de guerre. Première guerre mondiale », dans Guillaume PIKETTY (dir.), *Sortir de la guerre. Des guerres de Religion aux conflits asymétriques*, Paris, Passés composés – Ministère des armées, 2025, p. 257-291.

43. Notamment l'arrêté grand-ducal du 30 novembre 1929. Cf. SCUTO D., *La Nationalité luxembourgeoise*, op. cit., p. 89-93.

44. *Ibid.*, p. 194.

stables dans la première moitié du XX<sup>e</sup> siècle, avec des soldes migratoires compensés, le Luxembourg devient, après la Seconde guerre mondiale, un pays d'immigration durable. Cette situation culmine avec la signature, en 1972, d'accords avec le Portugal et la Yougoslavie pour l'accueil de travailleurs étrangers<sup>45</sup>.

Les cartes de résident·e viennent progressivement compléter les mesures mises en place à la fin du XIX<sup>e</sup> siècle (Fig. 2). Classées alphabétiquement, elles sont à l'origine pensées pour faciliter la manipulation des registres d'état civil, permettant à l'administration communale de garder la trace des mobilités successives des résident·e·s et d'y ajouter les changements d'état civil, comme le mariage. Elles s'affirment pourtant comme spécifiquement utiles à la surveillance des migrant·e·s, permettant *a fortiori* de renvoyer l'agent à la déclaration d'arrivée correspondante. En ce sens, leur généralisation à l'ensemble des communes du Grand-Duché est souvent discutée, et nombre de communes qui leur sont au départ réticentes, militent ensuite pour leur extension<sup>46</sup>.

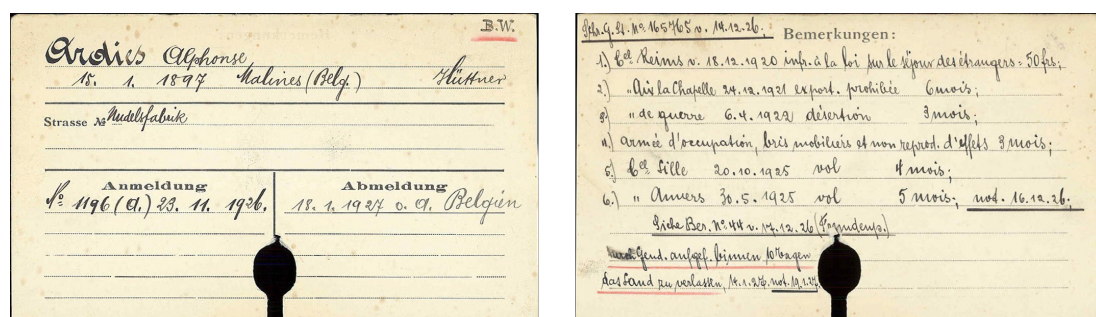


Fig. 2 : Exemple d'une carte d'enregistrement. État civil sur le recto et parcours criminel sur le verso

Pourtant, bien que le régime des cartes ait constitué un outil important pour les administrations du Grand-Duché, on ne peut évaluer avec justesse leur généralisation à l'ensemble des communes. Leur étude peut néanmoins nous aider à mieux comprendre le rapport administratif à l'écrit dans le pays. Pour ce qui est de la professionnalisation des métiers administratifs, notamment à l'échelle locale, celle-ci semble s'être faite sur l'ensemble de la période : les compétences des secrétaires communaux en matière de maniement de l'état-civil sont, par exemple, discutées à plusieurs reprises dans les premières années du XX<sup>e</sup> siècle<sup>47</sup>. Dans le cas du fichier de la Police des étrangers, centralisé dans la capitale, la confection des plus de 140 000 fiches nominatives est dévolue, encore à la même période, à un « soldat volontaire<sup>48</sup> ». Bien que ce ne soit qu'une hypothèse, elle révèle néanmoins les enjeux de travailler avec un corpus de sources large, à une période où la standardisation des techniques de l'écrit en est encore à ses débuts. L'exploration des moyens numériques à notre disposition peut alors nous fournir plusieurs pistes et constitue une étape à part entière de notre raisonnement, nous

45 SCUTO D., « Histoire des immigrations... », *art. cit.*, p. 33.

46 ANLux, min. Intérieur, INT-0012, Circulaire aux administrations communales, 26 septembre 1918, p. 5.

47 ANLux, Intérieur, Caisse de prévoyance des employés communaux, INT-0031-1a.

48 ANLux, Justice, Police des étrangers, J-073-33, Application des « systèmes des fiches » au service de la police des étrangers, lettre du procureur général Victor Thorn au ministre d'État Paul Eyschen, 7 mai 1905, n. p.

permettant en retour de mieux cerner ce que les sources ont à nous dire sur la nature de l'écrit administratif dans le pays. Notre chaîne de traitement s'articule en trois temps, de l'océrisation complète du corpus à son traitement quantitatif, en passant par la phase de mise en données et les questions qu'elle pose aux chercheur·euses·s.

### III. Retour d'expérience sur la Reconnaissance automatique de texte (ATR) avec Transkribus

L'idée d'extraire du contenu textuel à partir d'images est une préoccupation scientifique ancienne, remontant au début du XX<sup>e</sup> siècle<sup>49</sup>, mais c'est à partir des années 1990 que la reconnaissance automatique de caractères (ATR) a fait un bond en avant, bénéficiant des avancées du *machine learning* et des réseaux de neurones artificiels profonds<sup>50</sup>. Ces avancées ont permis de passer d'une reconnaissance par règles à un apprentissage automatique ou semi-automatique, par la machine assistée ou non par l'humain, du langage visuel et contextuel. L'importance de l'ATR pour l'étude de documents anciens en fait une étape de nos jours souvent nécessaire pour la recherche historique, qu'il s'agisse de l'OCR (*Optical Text Recognition*), ou océrisation, pour les textes imprimés, ou de l'HTR (*Handwritten Text Recognition*) pour l'écriture manuscrite.

Des développements récents ont permis de réduire les besoins en données d'entraînement tout en améliorant considérablement la précision, atteignant des taux d'erreur de caractères (CER) inférieurs à 2% pour certains manuscrits (on considère qu'un taux d'erreur de maximum 5% est acceptable). Des plateformes comme eScriptorium<sup>51</sup> ou Transkribus<sup>52</sup>, développées à partir du milieu des années 2010, ont démocratisé l'usage de ces modèles grâce à des interfaces conviviales et à des infrastructures collaboratives, auxquelles prennent part des chercheurs à la fois en histoire et en sciences de l'information ou encore en informatique. Aujourd'hui, l'enjeu ne réside plus uniquement dans la reconnaissance elle-même, mais dans la capacité à constituer de grands corpus cohérents, à harmoniser les traitements semi-automatisés, et à exploiter ces données à des fins de recherche. Nous avons choisi Transkribus, logiciel développé à partir de 2013 par des chercheur·euse·s issus de plusieurs universités européennes réuni·e·s au sein de l'associa-

49. Ce développement reprend en partie PINCHE Ariane et SROKES Peter A., « Historical Documents and Automatic Text Recognition: Introduction », *Journal of Data Mining and Digital Humanities*, 19 mars 2024, p. 2-3.

50. RICE Stephen, KANAI Junichi et NARTKER Thomas, « An evaluation of OCR accuracy », *Information Science Research Institute*, vol. 20, n° 9, 1993, p. 1-25.

51. Sur eScriptorium, cf. STOKES Peter A. et al., « The eScriptorium VRE for Manuscript Cultures », *Classics@Journal*, vol. 18, 2021.

52. NOCKELS Jo et al., « Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research », *Archival Science*, vol. 22, no. 3, 2022, p. 367-392.

53. [<https://readcoop.org>].

tion READ-COOP, à l'origine un projet de l'université d'Innsbruck en Autriche<sup>53</sup>. La force de cet outil réside en particulier dans les nombreuses techniques qu'il emploie à divers moments du travail de reconnaissance de caractères : segmentation et prétraitement des lignes de texte, combinaison de modèles différents selon les étapes, possibilité d'extraction larges et grande souplesse dans le traitement des erreurs, enfin des modèles de reconnaissance de caractères mis en ligne par READ-COOP et accessibles publiquement, permettant des bonnes possibilités de transcription de sources historiques gratuites. C'est également un outil utilisé dans l'étude des migrations, à partir de *card index* similaires<sup>54</sup>.

Le modèle utilisé a été d'emblée pensé comme suffisamment versatile et s'adaptant bien à une approche mixte, ce qui est utile pour des documents qui allient à la fois écriture manuscrite et imprimée, comme les archives des administrations. Basé sur un modèle qui a été entraîné sur des centaines de documents transcrits à la main, la marge d'erreur liée à l'HTR a pu être réduite à quelque 4%, ce qui veut dire que dans 96% des cas, la machine reconnaît les régions texte et transcrit à peu près justement l'écriture manuscrite. Le fichier TIFF (*Tagged Image File Format*), plus lourd que le JPG, permet une plus grande précision, notamment pour la prise en compte de détails typographiques comme les signes de ponctuation ou les accents. Mais bien que la marge d'erreur ait été réduite au maximum, des doutes demeurent sur la capacité du modèle à fonctionner avec un tel nombre de sources.

Nous constatons vite que Transkribus offre des opportunités d'extraction de l'information qui varient selon les cas. À mesure que notre modèle est entraîné, ce dernier est de plus en plus efficace avec les chiffres, comme c'est le cas pour les dates, et les abréviations (dans la plupart des cas les pays d'origine et les rues de résidence). Ces typographies sont probablement plus facilement reconnaissables, ce qui n'est pas le cas de portions de texte plus longues et non-standardisées, rendant les résultats de la reconnaissance de caractères difficilement opérants pour une analyse quantitative. Ce sont là des écueils bien connus et liés à l'HTR en tant que tel, puisque la haute variabilité de l'écriture manuscrite entre les individus induit des difficultés pour le modèle qui a été entraîné, ce qui n'est pas le cas pour l'OCR<sup>55</sup>.

Mais les résultats variables donnés par notre modèle s'expliquent également par la difficulté de traitement des cartes, qui présentent tellement de cas particuliers qu'il est difficile, avec les moyens à notre disposition, de les traiter uniformément à base d'un modèle, aussi entraîné qu'il ait pu être en amont. Quelques exemples seulement. Le

54. PFANZELTER Eva, « Digital Historical Remigration: The Example of the South Tyrolian Return Option », dans S. OBERBLICHER et al., *Return and Circular Migration*, op. cit., p. 51.

55. Cf. à titre d'exemple RETSINAS George et al., « Best Practices for a Handwritten Text Recognition System », dans Seiichi UCHIDA, Elisa BARNEY et Véronique EGLIN (dir.), *Document Analysis Systems: 15th IAPR International Workshop, DAS 2022, La Rochelle, France, May 22-25, 2022, Proceedings*, Cham, Springer, 2022, p. 247-259. À titre de comparaison, cf. ABADIE Nathalie et al., « A Benchmark of Named Entity Recognition Approaches in Historical Documents. Application to 19th s Century French Directories », dans *Id.*, *Document Analysis Systems*, op. cit., p. 445-460.

bilinguisme inhérent à l'administration luxembourgeoise, qui fait usage à la fois de l'allemand et du français, a constitué une première difficulté, nécessitant la correction manuelle des cartes dans les deux langues. Surtout, le travail dans les deux langues semble induire des rapports à l'écrit différenciés. L'allemand est ainsi utilisé jusqu'à la Seconde Guerre mondiale, et alors que les germanophones écrivent au stylo à plume, qui conserve une bonne lisibilité, les francophones écrivent au stylo à bille à partir de l'après-guerre, compliquant l'HTR. Des difficultés supplémentaires s'ajoutent à cela, dont le remplissage inégal des zones de texte est le plus problématique (blancs, ratures nombreuses, statut matrimonial inscrit dans le champ profession, abréviations de lieux difficilement trouvables ensuite).

Ces problèmes semblent alors moins liés aux performances de l'outil choisi qu'aux évolutions que connaît la culture de l'écrit administratif et sa diffusion inégale au Luxembourg. Les patronymes varient ainsi fortement en fonction de l'orthographe ou de la prononciation que les déclarants leur donnent lors de leur présentation devant les agents de la commune. À titre d'exemple, la mention du nom et du lieu de naissance des migrants d'Europe centrale et orientale à l'oral impose une orthographe, donc une catégorisation, incertaines, surtout dans le cas de cartes classées alphabétiquement. Celles-ci connaissent pourtant une certaine formalisation au début de la période. Dans le cas français, Alphonse Bertillon substitue, à partir de 1896, au classement alphabétique des fiches, un classement fondé sur l'orthographe phonétique des noms qui y sont déclarés, à l'aide d'un calcul des combinaisons orthographiques des patronymes, ce qui permet d'éliminer les incertitudes dues aux homonymes, à l'illettrisme ou à l'origine étrangère des personnes<sup>56</sup>. Dans le cas de Dudelage, le classement alphabétique semble demeurer la norme jusqu'à nos jours, ce qui peut induire des formes nombreuses d'un même nom. La phase de « nettoyage » des données ne peut donc, dans ce cas être sous-estimée<sup>57</sup>. Celle-ci demande de grands efforts en temps liés au passage d'une catégorie, celle issue de l'export HTR, à une autre, celle qui va créer notre base de données. C'est cette étape que nous allons voir à présent.

## IV. Les problèmes de l'exportation et de la mise en données

La phase d'HTR ou de transcription à la main ne produit pas immédiatement une base de données exploitable, mais un ensemble de blocs de texte plus ou moins bien identifiés. Un des avantages de Transkribus est d'offrir une large palette de moyens d'exportation possibles, et nous avons privilégié dans un premier temps le schéma XML (*Extended Markup Language*). La structure arborescente du format XML rend

56. BERTILLON Alphonse, *Du classement phonétique*, Paris, Société d'Éditions scientifiques, 1896. Cette méthode est encore utilisée de nos jours. Cf. DAVID Martin, *Les fichiers de police*, Paris, Presses universitaires de France, coll. « Que sais-je ? », 1999. Cité dans ABOUT I., « Les fondations d'un système national d'identification policière en France (1893-1914). Anthropométrie, signalements et fichiers », *Genèses*, vol. 54, no. 1, 2004, p. 28-52 [ici p. 34].

57. Cf. la critique pertinente de ce terme, dans BELIGNÉ Max, LOUDCHER Sabine et LEFORT Isabelle, « Du nettoyage des données à la critique des sources », dans Léo DUMONT, Octave JULIEN et Stéphane LAMASSÉ (dir.), *Histoires de mots. Saisir le passé grâce aux données textuelles*, Paris, Éditions de la Sorbonne, 2023, p. 219-232.

facilement possible l'extraction des données, qui obéissent à la structuration par blocs de texte offerte par l'HTR<sup>58</sup>. Le XML-PAGE, standard privilégié par READ-COOP, conserve à la fois une structure logique, par ordre de lecture (qui peut être corrigée à la main avant et pendant l'entraînement du modèle) et une structure en termes de zones de texte (« *Text-Regions* »). Ce faisant, il sépare efficacement ce qui relève de la ligne et de la région de texte, permettant une lisibilité facilitée dans le XML. Chaque zone de texte acquiert un identifiant unique (« *Id* »), rangé par sa position dans le texte (« *reading Order* »). Un score de confiance complète cela, ainsi que des coordonnées (Fig. 3).

```

18     </ReadingOrder>
19     <TextRegion id="tr_1728385488" custom="readingOrder {index:0;} structure {score:0.97; type:CR_name_first-name;}>
20       <Coords points="50,20 1435,20 1435,100 50,100"/>
21       <Textline id="tr_1728385488_t1_2" custom="readingOrder {index:0;}>
22         <Coords points="96,35 1361,37 1361,87 96,85"/>
23         <Baseline points="96,80 126,82 156,82 186,82 216,82 246,84 276,82 306,84 336,82 366,82 396,82 426,82 456,84"/>
24         <TextEquiv>
25           <Unicode>Arensdorff Nikolaus ~ Karl Katharina</Unicode>
26         </TextEquiv>
27       </Textline>
28     <TextEquiv>
29       <Unicode></Unicode>
30     </TextEquiv>
31   </TextRegion>
32   <TextRegion id="tr_1728385489" custom="readingOrder {index:1;} structure {score:0.95; type:CR_Date_place_birth;}>
33     <Coords points="156,95 951,95 951,170 156,170"/>
34     <Textline id="tr_1728385489_t1_1" custom="readingOrder {index:0;}>
35       <Coords points="150,115 929,115 929,165 150,165"/>
36       <Baseline points="150,160 180,160 210,158 240,156 270,158 300,156 330,156 360,158 390,156 420,156 450,154"/>
37       <TextEquiv>
38         <Unicode>27.2.1915 Pfaffenthal</Unicode>
39       </TextEquiv>
40     </Textline>
41   <TextEquiv>
42     <Unicode></Unicode>

```

Fig. 3 : Exemple d'une sortie XML issue d'un traitement par HTR

L'avantage de la sortie XML est que Transkribus structure déjà l'information en amont, ce qui nous permet ensuite d'affiner la catégorisation nous-même selon les cas rencontrés. Ainsi, une zone de texte (« *Text\_Region* ») peut être facilement corrigée, ensuite, directement grâce à un script Python<sup>59</sup>. Mais l'inconvénient est que ce type de structure impose un ordre interprétatif, et ce qui lui résiste dans la structuration est la plupart du temps perdu. Les conditions d'élaboration du code et les choix faits en amont deviennent alors des aspects centraux du travail sur nos données. Pour Camille Paloque-Bergès, le code peut être défini comme « la série d'instructions textuelles qui prévaut au traitement des données dans les programmes et les logiciels, de l'exécution d'une action sur les données (par des algorithmes) à l'affichage d'un contenu (par des

58. Sur le schéma XML, cf. BURNARD Lou, *Qu'est-ce que la Text Encoding Initiative ?*, Marseille, OpenEdition Press, 2015 ; FLANDERS Julia et JANNIDIS Fotis, « Data Modeling », dans Susan SCHREIBMAN, Ray SIEMENS et John UNSWORTH (dir.), *A New Companion to Digital Humanities*, Oxford, John Wiley & Sons, 2016, p. 232.

59. Pour une introduction au langage Python appliqué aux humanités numériques, cf. KARSDORP Folger, KESTEMONT Mike et RIDDELL Allen, *Humanities Data Analysis. Case Studies with Python*, Princeton, Princeton University Press, 2021.

instructions appelées « balisage ») sur les médias informatisés<sup>60</sup> ». Or les instructions textuelles à définir s'avèrent très vite à la fois trop difficiles et imprécises pour rendre compte de la diversité de nos données.

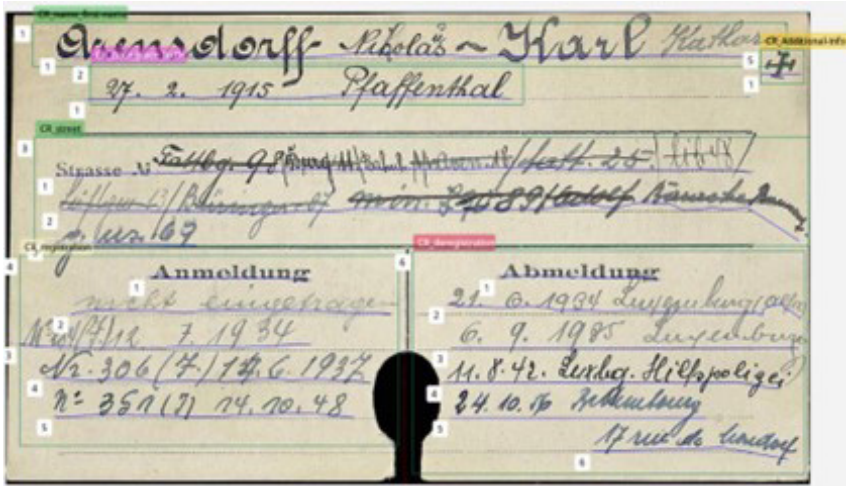


Fig. 4 : La carte correspondant à la sortie XML précédente

Nous avons donc choisi de créer les colonnes de notre base de données sur la base des zones de texte. Pour intégrer ces informations dans une base de données, il faut transformer le fichier XML en structure tabulaire, où chaque colonne ou ligne de notre base correspond à une ligne identifiée sur la carte. Ce passage impose de multiplier les lignes de code pouvant inclure au mieux les zones de texte non-claires dans notre base de données, en effectuant des choix de structure. Dans nos premières versions du code de parsing par XML, ce dernier ne prend qu'une ligne par région texte. L'exemple le plus commun est l'omission des lieux successifs de départ, extrêmement importants puisqu'ils nous permettront, à terme, de calculer les durées de séjour en plus de cartographier les lieux concernés. Si une personne a plusieurs lieux de départ déclarés dans le temps, le risque est de n'en garder qu'un et donc d'en perdre, ou bien de prévoir une structure plus complexe, par exemple à base d'identifiants croisés. Enfin, si une zone de texte est identifiée avant celle qui lui succède sur la carte, cela oblige à vérifier les informations au cas par cas, ce qui est impossible à partir d'un corpus aussi large. C'est le rôle des étiquettes chiffrées identifiées par Transkribus (Fig. 4).

Toute tentative d'harmonisation à partir d'un code informatique implique donc une normalisation, et cela s'applique tant au format XML qu'à son analyse syntaxique dans notre script. En réduisant la diversité du réel à une structure lisible, ce travail permet du même coup de révéler certaines des lacunes des sources. Le cas de la visibilité des femmes dans les sources est un exemple particulièrement pertinent. Quand elles ne sont pas enregistrées avec leur conjoint ou un membre de leur famille, leur profession

60. PALOQUE-BERGÈS Camille, « Les sources nativement numériques pour les sciences humaines et sociales », *Histoire@Politique*, vol. 30, n° 3, 2016, p. 224.

n'est simplement pas marquée, ce qui suppose qu'elles étaient femmes au foyer, et leur contribution à l'économie industrielle, de mieux en mieux connue, ne ressort pas dans nos résultats<sup>61</sup>. Notre base de données enregistre ces absences comme un champ vide. Dans nos premiers calculs statistiques, elles représentent ainsi moins du tiers des migrants dans la commune pendant toute la période, mais sont près de 60% parmi les personnes sans profession (Fig. 5). Leur inclusion dans notre base de données dépend d'un traitement quantitatif, soit d'un comptage au cas par cas pour les identifier, par exemple en les comptant dans les personnes classées sans profession. Sans des sources supplémentaires, il est alors impossible d'attester avec justesse leur profession.

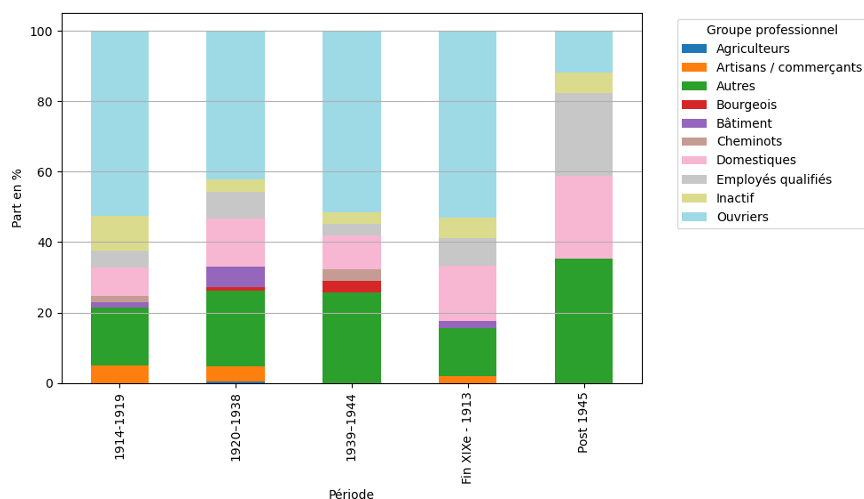


Fig. 5 : Répartition en % des groupes professionnels par période (premier échantillon de 500 cas)

Nous pouvons également mentionner certains cas particuliers, comme le nombre important de personnes déclarées parties vers l'hôpital de Dudelange durant les deux conflits mondiaux, que ce soient les infirmières, sur-représentées, ou les blessés de guerre, ce qui nécessite de créer une catégorie à part dans la base de données. Il s'agit donc d'une interprétation en lecture rapprochée de résultats extraits automatiquement. De tels exemples sont nombreux, et montrent la difficulté de traiter un corpus de données aussi large. Le cas des *Ostarbeiter* mérite en ce sens notre attention. En reprenant fidèlement celle écrite sur la carte, le XML leur assigne une nationalité et un lieu de sortie, ce dernier étant simplement désigné sous le terme de « *Sport* ». Cela correspond au *Sportlager*, camp de travail géré par l'ARBED qui, à partir de 1942 pouvait accueillir jusqu'à plus de 2000 travailleurs forcés, prisonniers de guerre de l'Est<sup>62</sup>. Cela pose l'historien face à un paradoxe, qui est à la fois l'euphémisation du statut de ces individus dans les sources puisqu'ils sont désignés de manière laconique par l'administration communale, et leur visibilité rendue grâce à la chaîne de traitement utilisée.

61. Cf. l'étude classique de DOWNS Laura Lee, *L'inégalité à la chaîne. La division sexuée du travail dans l'industrie métallurgique en France et en Angleterre (1914-1939)*, Paris, Albin Michel, 2002.

62. BAPTISTA BARRA Fernando Ricardo, « Les Ostarbeiter dans la sidérurgie luxembourgeoise (1942-1945) », dans Charles BARTHEL et Joséé KIRPS (dir.), *Terres rouges. Histoire de la sidérurgie luxembourgeoise*, vol. 1, Luxembourg, Centre d'études et de recherches européennes Robert Schuman – Archives nationales de Luxembourg, 2009, p. 16-70 ; GANSCHOW I., 'Keiner weinte. Es gab keine Tränen mehr'. *Ukrainische, russische und belarussische Zwangsarbeiterinnen in Luxemburg im Zweiten Weltkrieg aus transnationaler Sicht*, Luxembourg, CapybaraBooks – C<sup>2</sup>DH, 2025.

Il convient alors de retravailler les catégories, comme les colonnes incluses dans notre base de données.

Le contrôle sur les données devient alors la question centrale. Comment s'assurer que notre base de données finale comportera l'ensemble des données contenues sur les cartes ? Frédéric Clavert définit ce type de pratique comme une « double lecture<sup>63</sup> » des sources à l'ère numérique, entre la lecture distante proposée par Moretti en 2000<sup>64</sup>, qui est une médiation des sources par le quantitatif à l'œuvre devant les larges corpus de données, et la lecture proche à laquelle sont d'avantage habitués les historien·ne·s. En proposant une extraction rigide, demandant un code particulièrement performant, l'usage de Transkribus nous a aidé à envisager des perspectives autres afin d'approcher d'une transcription qui soit la plus fidèle possible. Il s'agit d'une lecture des sources médiée par le code informatique, qui peut être ensuite interprétée et corrigée par la main humaine.



Le tournant digital et la numérisation des sources qui lui est corrélée induisent donc une chaîne de traitement complexe pour transformer la source numérisée en jeu de données. Le travail de l'historien·ne à l'heure du digital se fait alors « entre la poussière de l'archive et la lumière bleue de l'écran<sup>65</sup> ». L'expérience quotidienne de l'utilisation d'outils digitaux fait que les étapes se confondent, entre les difficultés à entraîner un modèle, l'accumulation des données, leur transcription et leur intégration dans une base de données, autant de ce que Caroline Muller et Frédéric Clavert nomment « des pratiques implicites et discrètes<sup>66</sup> ». Ces dernières rapprochent l'historien·ne du digital de l'archiviste, puisqu'à partir du traitement automatisé ressortent des étapes de manipulation des fichiers, de schémas et de structures d'indexation, des concepts propres à une histoire plus large des systèmes d'information à laquelle les fichiers des individus sont eux-mêmes reliés<sup>67</sup>. Avec la mise en données, ce que le système rend visible n'est pas toujours ce qui comptait pour les agents communaux ou les migrant·e·s eux-mêmes.

Une critique rigoureuse des sources est alors essentielle afin de donner à voir ce que les cartes d'enregistrement ne nous disent pas. En retraçant leurs conditions de production et d'émission, nous pourrions à terme interroger les rapports des différentes échelles de l'État luxembourgeois à ses migrations, que ce soit au niveau

63 CLAVERT Frédéric, « Vers de nouveaux modes de lecture des sources », dans Olivier LE DEUFF (dir.), *Le temps des humanités digitales*, Limoges, FYP éditions, 2014.

64 MORETTI Franco, *Graphs, Maps, Trees: Abstract Models for Literary History*, Londres – New York, Verso, 2007 ; *Id.*, *Distant Reading*, Londres – New York, Verso, 2013. Cf. à propos des débats engagés par la lecture distante : CLAVERT Frédéric et MULLER Caroline, « L'histoire au temps des algorithmes. Une réflexion prospective sur l'introduction de l'intelligence artificielle en histoire au 21<sup>e</sup> siècle », 20 & 21. *Revue d'histoire*, vol. 162, no. 2, 2024, p. 13-26.

65 MULLER C. et CLAVERT F., « De la poussière à la lumière bleue. Émotions, récits, gestes de l'archive à l'ère numérique », *Signata*, vol. 12, 2021.

66 *Ibid.*

67. VAN FASSEN M. et OPREL M., *op. cit*

du ministère, du Parquet ou de la commune. Bien que le traitement de ce fonds documentaire extensif soit possible, offrant de nombreuses possibilités d'analyses quantitatives, les données extraites restent, comme nous l'avons vu, imprécises et sujet à interprétation. Cela montre, néanmoins, la nécessité de s'intéresser aux sources locales, qui souvent échappent aux entreprises d'archivage nationales. La masse des sources, fruit d'un effort de numérisation coûteux, induit aussi des biais et des risques dans le traitement comme dans l'écriture du code. C'est pourquoi, du scan à l'analyse historique, il est indispensable de documenter les choix effectués, les étapes précises du code informatique ainsi que les exclusions.

Enfin, c'est par la transparence au niveau des méthodes digitales utilisées, ainsi que par la coopération interdisciplinaire que ces biais peuvent être dépassés. Pour ce qui est de l'effet de sources, ou la tendance à entrevoir le monde social par le seul prisme des catégories utilisées dans les archives, nous en revenons à un des écueils classiques liés au traitement de séries documentaires larges. À la manière de l'histoire sérielle des années 1960/70 et de ses renouvellements dans les décennies suivantes<sup>68</sup>, l'histoire numérique engage l'apport de l'historien·ne pour penser une discipline dont l'usage et les méthodes, souvent considérées comme allant de soi, sont à la base d'un tournant proprement réflexif et herméneutique<sup>69</sup>. Le risque est en effet l'objectivation, et de ne pas donner leur voix aux acteur·rice·s eux-mêmes et elles-mêmes, qui seraient cantonné·e·s au regard que l'État porte sur eux, ou à des séries statistiques. Autant de « schémas cognitifs<sup>70</sup> » qui rapprochent les préoccupations liées aux sources de l'histoire des migrations, de leur écriture en tant que données, plus d'un siècle plus tard.

68. FURET François, « Histoire quantitative et construction du fait historique », *Annales. Économies, Sociétés, Civilisations*, vol. 26, no. 1, 1971, p. 63-75 ; LEMERCIER C. et ZALC C., *op. cit.*, p. 8-18.

69 FICKERS Andreas, TATARINOV Juliane et VAN DER HEIJDEN Tim, « Digital History and Hermeneutics: Between Theory and Practice: An Introduction », dans A. FICKERS et J. TATARINOV (dir.), *Digital History and Hermeneutics. Between Theory and Practice*, Berlin, De Gruyter, 2022, p. 1-19.

70 RYGIEL Ph. (dir.), *Le Bon grain et l'ivraie. La sélection des migrants en Occident, 1880-1939*, Montreuil, Aux lieux d'être, 2008, p. 10.

## Bibliographie

ABOUT Ilsen et DENIS Vincent, *Histoire de l'identification des personnes*, Paris, Éditions La Découverte, coll. « Repères », 2010.

BELIGNÉ Max, LOUDCHER Sabine et LEFORT Isabelle, « Du nettoyage des données à la critique des sources », dans Léo DUMONT, Octave JULIEN et Stéphane LAMASSÉ (dir.), *Histoires de mots. Saisir le passé grâce aux données textuelles*, Paris, Éditions de la Sorbonne, 2023, p. 219-232.

74

CLAVERT Frédéric et MULLER Caroline, « L'histoire au temps des algorithmes. Une réflexion prospective sur l'introduction de l'intelligence artificielle en histoire au 21<sup>e</sup> siècle », *20 & 21. Revue d'histoire*, vol. 162, n<sup>o</sup>. 2, 2024, p. 13-26.

DE BARROS Françoise, « Les acteurs municipaux et “leurs” étrangers (1919-1984) : gains et contraintes d'un détour communal pour l'analyse d'un travail de catégorisation étatique », *Genèses*, n<sup>o</sup>. 72, vol. 3, 2008, p. 42-62.

DU MOUZA Cédric, LAMASSÉ Stéphane et RYGIEL Philippe, « De l'histoire numérique à l'histoire données ? », *Les Cahiers de Framespa*, 42, 2023.

GALLORO Piero D., *Ouvriers du fer, princes du vent. Histoire des flux de main-d'œuvre dans la sidérurgie lorraine (1880-1939)*, Metz, Éditions Serpenoise, 2001.

GARDEY Delphine, *Écrire, calculer, classer. Comment une révolution de papier a transformé les sociétés contemporaines, 1800-1940*, Paris, Éditions La Découverte, 2008.

GREEFS Hilde et WINTER Anne (dir.), *Migration Policies and Materialities of Identification in European Cities. Papers and Gates, 1500-1930s*, Abingdon, Routledge, 2019.

HOESTRA Rik et KOOLEN Marijn, « Data scopes for digital history research », *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 52, n<sup>o</sup>. 2, 2018, p. 1-16.

KARSDORP Folgert, KESTEMONT Mike et RIDDELL Allen, *Humanities Data Analysis. Case Studies with Python*, Princeton, Princeton University Press, 2021.

LEBOUTTE René, PUISSANT Jean et SCUTO Denis, *Un siècle d'histoire industrielle (1873-1973)*, Paris, Éditions Sedes, 1998.

MULLER Caroline et CLAVERT Frédéric, « De la poussière à la lumière bleue. Émotions, récits, gestes de l'archive à l'ère numérique », *Signata*, vol. 12, 2021.

NIJENHUIS Ida *et al.* (dir.), *Information and Power in History. Towards a Global Approach*, Londres, Routledge, 2020.

NOCKELS Jo *et al.*, « Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research », *Archival Science*, vol. 22, no. 3, 2022, p. 367-392.

NOIRIEL Gérard, *Immigrés et prolétaires, Longwy, 1880-1980*, Marseille, Éditions Agone, 2019 (1<sup>ère</sup> éd. 1984).

75

NOIRIEL Gérard, *La Tyrannie du national. Le droit d'asile en Europe (1793-1993)*, Paris, Calmann-Lévy, 1991.

OBERBLICHER, Sarah, PFANZELTER, Éva et LARCHER, Valerio (dir.), *Return and Circular Migration in Contemporary European History*, Berlin, De Gruyter, 2022.

PINCHE Ariane et STOKES Peter A., « Historical Documents and Automatic Text Recognition: Introduction », *Journal of Data Mining and Digital Humanities*, 19 mars 2024, p. 2-3.

ROSENTAL Paul-André, « Migrations, souveraineté, droits sociaux. Protéger et expulser les étrangers en Europe du XIX<sup>e</sup> siècle à nos jours », *Annales HSS*, vol. 66, no. 2, 2011, p. 335-373.

RYGIEL P., *Le Temps des migrations blanches. Migrer en Occident (1840-1940)*, Montreuil, Aux lieux d'être, 2007.

SCUTO Denis, *La nationalité luxembourgeoise (XIX<sup>e</sup>-XXI<sup>e</sup> siècles). Histoire d'un alliage européen*, Bruxelles, Éditions de l'Université de Bruxelles, 2012.

SCHREIBMAN Susan, SIEMENS Ray et UNSWORTH John (dir.), *A New Companion to Digital Humanities*, Oxford, John Wiley & Sons, 2016.