

# **Compter les langues pour gouverner : enjeux épistémologiques et conséquences socio-politiques des statistiques linguistiques officielles**

Philippe HUMBERT

Institut de plurilinguisme, Université de Fribourg (Suisse)

De tous les pays non-anglophones, les Pays-Bas seraient les meilleurs en anglais, loin devant la Suisse (14<sup>e</sup>) et la France (29<sup>e</sup>).<sup>1</sup> Avec plus de 12 millions d'habitant·e·s, la ville la plus peuplée de francophones dans le monde ne serait plus Paris mais Kinshasa.<sup>2</sup> Ces chiffres ne circulent pas uniquement au sein d'un cercle fermé de sociolinguistes passionné·e·s de comptabilité linguistique, mais ils sont relayés et réappropriés dans l'espace public. Dans cet article, j'interroge les coulisses de la réalisation de ce genre de statistiques pour chercher à comprendre les raisonnements politiques et scientifiques contribuant à dresser une image quantifiée des langues.

Se popularisant dès le XIX<sup>e</sup> siècle, la statistique est une discipline scientifique dont l'histoire est étroitement liée à des enjeux de gouvernance politique et économique (Desrosières 2010). Les analyses des raisonnements démographiques de Foucault (2004) et Hacking (1982) démontrent comment la statistique s'est transformée en outil de domination politique et de (re)production des inégalités sociales, par exemple en élaborant des statistiques du chômage, de la pauvreté et de la criminalité, utilisées à la fois pour créer des catégories

---

<sup>1</sup> « Les Suisses classés parmi les bons élèves en anglais », *RTS Info* du 22.11.2016, consulté le 03.05.2018 : <https://www.rts.ch/info/suisse/8175791-les-suisses-classes-parmi-les-bons-eleves-en-anglais.html>

<sup>2</sup> « Kinshasa a dépassé Paris comme plus grande ville francophone du monde », par Pierre Magnan, *France Info* du 02.06.2017, consulté le 03.05.2018 : <http://geopolis.francetvinfo.fr/kinshasa-a-depasse-paris-comme-plus-grande-ville-francophone-du-monde-145213>

socio-économiques de citoyen·ne·s et en justifier l'isolement social. Historiquement, les recensements de population ont aussi un rapport étroit avec le colonialisme, puisque les colons avaient besoin de catégoriser les populations qu'ils découvraient pour les répertorier et les cartographier en communautés imaginées homogènes sur les territoires qu'ils administraient (Anderson 2006, 163-85). Depuis les années 1970, la statistique n'est plus uniquement produite par l'Etat et pour l'Etat, mais elle s'inscrit dans des constellations de gouvernance de plus en plus complexes, dans lesquelles ce qui relève de l'Etat ou du privé n'est pas évident à identifier (cf. Merry 2016). Perçues et conçues comme « objectives », l'analyse des ancrages culturels, politiques et économiques de la construction de cette objectivité scientifique permet de saisir les intérêts économiques et politiques qui se trament derrière la réalisation des chiffres (Merry 2016; Porter 1995).

Dans cette optique, cet article aborde la question de la quantification des langues en lien avec des enjeux de gouvernance socio-politiques et économiques. Mon objectif n'est pas de comprendre comment produire de « meilleures » statistiques, mais de saisir les raisonnements privilégiant une approche quantitative des langues plutôt qu'une autre, reflétant la formation d'idéologies langagières, à savoir des conceptions et croyances contribuant à définir ce qui compte comme langue et comme locuteur (cf. Woolard et Schieffelin 1994). Il s'agit d'aborder les processus contribuant à rendre certaines pratiques linguistiques plus visibles ou au contraire, à effacer certaines langues et groupes de locuteurs des résultats statistiques. En m'inscrivant dans une démarche sociolinguistique *critique* (Heller 2002), mes questions de recherche essaient de dépasser la dimension descriptive du sujet : ce n'est pas seulement « comment compter les langues ? » que je questionne, c'est « pourquoi compter les langues de telle manière ? Dans quel contexte, dans quel intérêt et avec quelles conséquences, pour qui ? ».

Pour répondre à ces questions, je me concentre sur trois axes thématiques qui inscrivent la quantification des langues dans des enjeux épistémologiques fondamentaux pour les sciences du langage, puisqu'il ne s'agit pas de questionnements concernant uniquement statisticien·ne·s et démographes, mais d'objets pouvant faire débat en sciences du langage.

1. Les situations linguistiques abordées dans les recensements et enquêtes officielles sur les langues ;
2. La quantification du plurilinguisme ;
3. Les degrés de variation linguistique sondés.

Ces trois axes permettent d'aborder des enjeux systématiquement rencontrés dans la littérature scientifique : pour quelles raisons privilégie-t-on le relevé d'une situation linguistique plutôt qu'une autre ? Recense-t-on une seule ou plusieurs langues par individu ? À partir de quel moment une langue est-elle comptée comme une langue – ou comme un dialecte ou un patois – dans la statistique ?

Mon analyse s'appuie sur un recueil d'études empiriques et théoriques que j'ai réalisé avec Renata Coray et Alexandre Duchêne (voir Humbert et al. 2018).<sup>3</sup> Après avoir consulté plusieurs centaines de travaux de recherche issus de la sociolinguistique, de la sociologie, de la politologie, de la démographie et d'autres disciplines transversales comme la démolinguistique et la géolinguistique, j'ai sélectionné des épisodes que je considère emblématiques des multiples problématiques émergeant de la quantification des langues, afin de problématiser l'ancrage idéologique des enjeux épistémologiques.

## **1. Les types de situations linguistiques recensés**

Les auteurs ayant synthétisé la recherche sur les manières de relever les langues dans les recensements distinguent principalement trois situations linguistiques (Liebersohn 1966, 267; de Vries 1990, 61-63; Arel 2002b, 97) :

---

<sup>3</sup> Cet article est le fruit d'une réflexion personnelle présentée lors du colloque international *L'Image des langues* (Neuchâtel, 10.11.2017). Je profite de ces lignes pour remercier chaleureusement Renata Coray et Alexandre Duchêne de nos échanges stimulants et constructifs, ainsi que les organisateur·trice·s et les participant·e·s au colloque pour leurs commentaires. Merci aux deux réviseur·e·s anonymes pour leurs précieuses suggestions.

Cette recherche a été menée dans le cadre du projet « Compter les langues : histoire, méthodes et politiques des recensements de population » au Centre scientifique de compétence sur le plurilinguisme (Université et Haute école pédagogique de Fribourg, Suisse), dirigé par Alexandre Duchêne et Renata Coray.

La rédaction de cet article a pu être achevée dans le cadre du projet "Gouverner les langues par les nombres : analyse historiographique, ethnographique et interactionnelle de la production de statistiques officielles des langues en Suisse » (n°181377) Doc.Mobility du Fonds national suisse de la recherche scientifique

- a) La première langue apprise durant l'enfance (souvent associée aux origines des individus) ;
- b) La langue d'usage régulier au moment du recensement ;
- c) Le niveau de connaissance ou de compétence dans une langue officielle.

Dans cet article, je démontre que chacune de ces situations linguistiques est justifiée par des choix à la fois scientifiques et politiques, et surtout que chaque situation ne profite pas à tous les groupes de locuteur·trice·s.

### **1.1 Aux origines des trois situations linguistiques : langue, nationalité, race, politique et territoire**

C'est à partir du XIX<sup>e</sup> siècle que les premiers recensements exhaustifs des langues sont réalisés en Europe, à une époque où les Etats-nations tentent d'associer une ou plusieurs langues à leur projet de construction nationale. Selon Labbé (2009), à cette époque, les statisticiens rassemblés durant les sessions du *Congrès international de statistique* (1853-1876) ne s'accordent pas sur la manière de compter les langues. Ces désaccords transparaissent dans les pratiques déjà si variées entre les Etats : certains souhaitent que la langue reflète les origines des répondant·e·s, d'autres veulent souligner les aspects plus fonctionnels du langage en société.

La Prusse défend le relevé de la *Muttersprache* [la langue maternelle<sup>4</sup>], parce que Richard Böckh, le statisticien à l'origine de cette idée, est persuadé que la *Muttersprache* est le moyen le plus objectif de relever la nationalité des individus. Selon lui, on peut mentir sur ses origines ou sa nationalité, mais pas sur sa langue maternelle (Labbé 2003, 48-49). Fortement influencé par les concepts développés par les romantiques du nationalisme allemand, Böckh est persuadé que la langue maternelle des individus contient les caractéristiques culturelles inhérentes des peuples (Leuschner 2004, 393-97). Parmi ses projets, Böckh veut prouver scientifiquement que l'allemand est la vraie langue de cœur des Alsacien·ne·s, dans le but de justifier une annexion de l'Alsace (idem, 407). Mais Böckh n'a pas vu venir les choses à l'Est de la Prusse, où les minorités slaves commencent à revendiquer des portions de territoire sur la base des résultats de cette même *Muttersprache* (idem, 411-13). D'ailleurs, d'autres minorités européennes se mettent à exiger un relevé de la langue maternelle pour gagner en visibilité dans les nombres officiels.

---

<sup>4</sup> Toutes les traductions françaises sont de moi.

Composée de nombreuses minorités slaves, l'Autriche opte dès 1880 pour un recensement de la *Umgangssprache*, c'est-à-dire la langue d'usage courant, celle qu'on utilise au quotidien dans la plupart des échanges.<sup>5</sup> Ces minorités utilisant plutôt l'allemand que le tchèque, le slovaque ou le slovène dans les échanges habituels en dehors du cadre strictement familial, la question suscite un tollé auprès des représentant·e·s politiques de ces minorités qui accusent le gouvernement de chercher à les effacer du paysage politique (Brix 1982). Dès lors, ces minorités exigent un recensement de la langue maternelle dans la veine des relevés prussiens. En effet, elles avancent que la *Umgangssprache* ne fait qu'accroître la proportion de germanophones, puisque l'allemand était la langue la plus utilisée pour communiquer à travers l'Autriche (ibidem).

En somme, autant pour la *Muttersprache* que pour la *Umgangssprache*, le relevé de caractéristiques linguistiques génère des revendications politiques de la part des minorités linguistiques et des gouvernements. Curieusement, les statisticiens autrichiens et prussiens affirment tous deux détenir la manière la plus « objective » de relever des caractéristiques linguistiques et/ou nationales. Pour les Autrichiens, on ne peut pas objectivement associer la langue à la nationalité d'un individu ; pour les Prussiens, la langue maternelle est le reflet le plus objectif de sa nationalité.

Le relevé des connaissances ou des compétences d'une langue officielle traduit aussi des intérêts politiques. Les Britanniques sont les premier·ère·s à s'y être intéressé dans le recensement d'Irlande de 1851. Les administrateur·trice·s anglais·es souhaitent ainsi observer l'évolution du gaélique irlandais pour constater la progression de l'anglais sur le territoire. À cause de la grande famine, l'administration britannique peut effectivement constater une disparition accélérée du gaélique, en lien avec les décès et émigrations massives d'Irlandais·es (Christopher 2011, 539). En dehors du Royaume-Uni, les Britanniques relèvent sporadiquement les caractéristiques linguistiques des régions colonisées. Ce qui leur importe le plus, c'est de savoir s'il existe un réservoir suffisant d'individus capables de communiquer en anglais et/ou dans une autre langue locale (cf. Christopher 2011). En outre, les colons britanniques interprètent la langue comme un indicateur de la race des communautés qu'ils découvrent, mais il semblerait que les catégories langue, religion, race aient été utilisées de façon interchangeable (cf. Manoharan 2001; Christopher 2011).

---

<sup>5</sup> Le terme allemand *Umgangssprache* est difficile à traduire. Dans la littérature, les auteur·e·s relèvent souvent la notion d'usage et une référence à sa dimension plutôt orale (cf. Arel 2002b, 116-17; Busch 2016, 4).

## 1.2 Combiner des situations linguistiques pour répondre à des enjeux de gouvernance diversifiés

Il n'est pas exclu de relever plusieurs situations linguistiques pour tenter de répondre à des enjeux de gouvernance différents. Recensant traditionnellement la « *rodnoj jazyk* » [langue natale] dans le but d'affiner les informations sur la nationalité titulaire des habitant·e·s,<sup>6</sup> l'URSS ajoute une question sur la maîtrise d'une deuxième langue pour vérifier la progression du russe en tant que langue de communication sur l'ensemble du territoire soviétique (Silver 1986). Après la dissolution de l'URSS, les rapports de force sont inversés : les minorités constituent généralement la majorité dans leur nouvelle entité territoriale (Pavlenko 2011). Les manières de relever les langues dans ces nouveaux Etats reflètent une réappropriation des enjeux de politiques linguistiques. Désireux de se distancer du passé soviétique, le Kazakhstan se focalise uniquement sur les connaissances en langue kazakh pour occulter le passé russophone (Bhavna 2004). En revanche, la Biélorussie combine deux types de situations linguistiques pour mettre en évidence la coexistence du russe et du biélorusse dans la société, dans le but de maintenir un lien étroit avec leur voisin russe (Arel 2002a, 813-14).

Les Etats-Unis d'Amérique sondent aussi plusieurs situations linguistiques. En fait, les questions changent énormément durant le XXe siècle, reflétant la prévalence d'intérêts politiques variables sur la nécessité de conserver une continuité statistique pour produire des données comparables dans le temps (cf. Leeman 2004). Dès les années 1960, la question linguistique sert à compléter celle de la race pour identifier les « *Latinos-as* » qui ne seraient ni vraiment Blanc·he·s, ni vraiment Noir·e·s (Solórzano et Ahlén 2010, 20-23). Dès 1980, le recensement américain relève la langue à la maison (uniquement si celle-ci diffère de l'anglais). Un individu inscrivant une autre langue que l'anglais à la maison est prié de répondre à une question supplémentaire sur ses compétences en anglais. À partir de cette question, tout foyer où personne de plus de 14 ans ne mentionne savoir *très bien* l'anglais est considéré comme « *linguistically isolated* » [isolé linguistiquement] (Leeman 2004, 528). Bien que ce concept ne soit plus utilisé depuis 2011 (Leeman 2013, 322), l'idée initiale de cette question était de lutter contre les discriminations linguistiques en identifiant les foyers qui auraient besoin d'obtenir du matériel administratif dans d'autres langues. Il s'avère que ces statistiques ont

---

<sup>6</sup> Les passeports soviétiques contenaient une inscription de la nationalité titulaire (tatar, ouzbèk, etc.) afin de distinguer les peuples composant l'URSS (voir Arel 2002a).

plutôt produit l'effet inverse, relayant une image disproportionnée de non anglophones dans la société américaine et associant la diversité linguistique à une forme de menace pour la nation (Leeman 2004; 2013).

### **1.3 La statistique, un outil réapproprié et développé par les minorités**

L'histoire des recensements prussiens et autrichiens montre à quel point la question du comptage est cruciale pour les minorités (voir section 1.1). Aujourd'hui, d'autres minorités ont recours aux statistiques linguistiques pour étayer des arguments politiques. À ce titre, les activistes basques entretiennent un lien étroit avec la rhétorique des nombres, tant et si bien qu'ils-elles produisent leurs propres données pour prouver que leur langue nécessite un soutien politique et économique pour ne pas disparaître (Urla et Burdick 2018). En Espagne, ils-elles ne sont pas la seule minorité à produire des statistiques sur les pratiques et compétences langagières (cf. Moreno-Fernández 2007). Ils-Elles ont toutefois passablement œuvré à développer l'instrument quantitatif, notamment par le biais du *Kale Neurketa* [l'enquête du basque dans la rue], qui consiste à envoyer des observateur·trice·s se promener discrètement dans l'espace public et prendre des notes de la présence du basque tel qu'il est parlé (Altuna et Urla 2013, 216-19). Si les activistes basques s'attellent à produire des données de plus en plus détaillées en combinant plusieurs types d'enquêtes, c'est aussi une manière de s'adapter à une rhétorique néolibérale prônant la réalisation d'audits et d'évaluations quantitatives, dans le but est de prouver scientifiquement que le basque est encore bien vivant dans certaines régions et qu'il constitue un atout économique pour ses locuteur·trice·s (Urla et Burdick 2018).

## **2. La quantification du plurilinguisme individuel**

Le plurilinguisme constitue l'une des raisons fondamentales pour lesquelles les statistiques linguistiques sont produites. Dans les relevés statistiques officiels, les bilingues sont souvent contraint·e·s de sélectionner une seule langue. Il existe une grande variabilité dans les degrés d'ouverture à l'inscription de plusieurs langues, reflétant des enjeux socio-politiques ancrés dans des idéologies langagières en lien avec la construction et l'organisation des territoires linguistiques.

## 2.1 Le bilinguisme à l'épreuve d'un imaginaire homogénéisant

Selon la conception prussienne associant la langue à l'origine des individus, les bilingues sont problématiques. S'il est possible de s'exprimer dans deux langues, il est en revanche inconcevable d'appartenir à deux peuples pour les défenseur·e·s de la *Muttersprache* (Haarmann 1979, 24-25). Les bilingues sont d'abord ignoré·e·s ou assimilé·e·s aux germanophones dans les premiers résultats (Labbé 2003, 54-55). Avec retenue, la statistique prussienne offre la possibilité d'indiquer plusieurs langues par la suite, comme l'indiquent les instructions de 1900 :

« Chaque homme possède une langue maternelle qui est celle dans laquelle il parle couramment et il pense. Seulement quelques personnes possèdent deux langues maternelles, cependant les cas de ce type apparaissent chez les personnes dont les parents ont des langues maternelles différentes. » (Labbé 2003, 55)

Le bilingue est ici une juxtaposition de deux monolinguisms dans une même personne. La référence aux parents est révélatrice d'une transmission linguistique imaginée comme fondamentalement génétique et familiale, écartant d'autres facteurs sociaux qui pourraient influencer sur la connaissance d'une langue (le travail, le commerce, la scolarité, etc.).

Dans des exemples plus récents, quand les recensements relèvent plusieurs situations linguistiques, la première question incite – voire oblige – souvent les individus à n'inscrire qu'une seule langue (cf. Laur 2013 pour le Canada; Duchêne et al. 2018 pour la Suisse; Leeman 2004 pour les Etats-Unis; Arel 2002a pour les recensements post-soviétiques), conférant une dimension politique intrinsèque à la question linguistique. En revanche, dans les questions complémentaires concernant les langues d'usage à la maison ou au travail, par exemple, le bilinguisme ne fait pas nécessairement l'objet de restrictions particulières (cf. Laur 2013 pour le Canada; Duchêne et al. 2018 pour la Suisse).

En somme, quand la question est associée à des enjeux identitaires, le bilinguisme est plutôt restreint. Cela reflète le besoin de quantifier des unités imaginées homogènes dans un souci de clarté politique où la mixité devient une variable statistique plus complexe à interpréter. Ce constat est particulièrement frappant lorsqu'on observe comment les statistiques linguistiques sont transposées en cartes géographiques pour illustrer la répartition des langues dans un pays : les aires linguistiques sont souvent dessinées en fonction du monolinguisme imaginé de chaque individu, occultant leur potentiel



bilingue (cf. Humbert sous presse). Toutefois, la section suivante démontre que le recensement du bilinguisme peut aussi être interprété comme une tentative de s'approprier des locuteur·trice·s, engendrant aussi des conflits politiques identitaires.

## **2.2 Le bilinguisme : menace ou atout ?**

La littérature scientifique mentionne des cas où des individus et institutions reprochent au relevé du plurilinguisme de favoriser la visibilité d'une langue dominante pour gommer la présence de langues minoritaires. En offrant la possibilité d'indiquer plusieurs langues parlées par personne dans les recensements belges entre 1866 et 1947, les autorités ont dû faire face à de vives contestations au sein de la population (Lévy 1964). Pour cause, les résultats avaient une influence directe sur la détermination des régimes linguistiques dans chaque municipalité. La langue française jouissant d'un certain prestige social, les politicien·ne·s flamand·e·s craignent de perdre du terrain, puisqu'ils·elles estiment compter plus de bilingues franco-néerlandais que chez leurs compatriotes francophones wallon·ne·s (Prévost et Beaud 2002, 95-99). La situation devient si tendue que le gouvernement belge décide de supprimer la question linguistique des recensements dès 1960 et de fixer définitivement les frontières linguistiques (idem, 98).

Au Canada, ce sont les activistes québécois·es qui craignent que tous les francophones du pays ne s'assimilent aux anglophones. Dès les années 1960, les scientifiques québécois·es se penchent sur les « transferts linguistiques et le maintien des langues » par le biais des données démographiques à disposition. Concrètement, les « démolinguistes » (les démographes des langues) souhaitent observer le nombre d'individus déclarant utiliser une langue à la maison différente de leur langue maternelle, afin de constater combien de francophones de naissance seraient devenus anglophones (pour en savoir plus sur les calculs des transferts linguistiques, consulter par ex. de Vries 1990). D'un recensement à l'autre, démolinguistes et politicien·ne·s québécois·es exigent que des questions soient ajoutées pour obtenir des informations plus détaillées sur les compétences et pratiques linguistiques de la population à la maison, au travail et auprès des « allophones » (Prévost et Beaud 2002, 99-108). L'enjeu ne devient plus uniquement territorial, mais économique et social, puisque les démolinguistes essaient de calculer les écarts salariaux entre anglophones et francophones, ou encore à vérifier si les allophones sont bien « assimilé·e·s » aux francophones au Québec (ibidem).

Dans ces discours statistiques hautement politisés, la question des bilingues reste relativement en retrait : les analyses statistiques et discussions politiques se concentrent principalement sur la dominance d'une langue dans la société, mais pas tant sur le fait de pratiquer ou posséder plusieurs langues par individu. Certes, que ce soit en Belgique, au Canada ou en Suisse, les chiffres sont/ont été utilisés pour déterminer le régime monolingue, bilingue voire trilingue de certaines régions, mais ce sont généralement des statistiques fondées sur le monolinguisme individuel des habitant·e·s qui prévalent.

Néanmoins, le plurilinguisme peut aussi constituer un atout que la statistique va chercher à mettre en valeur. Après des décennies de recensements du monolinguisme, la statistique suisse cherche désormais à fournir des données plus détaillées sur les pratiques plurilingues de la population. Cette récente quantification du plurilinguisme helvétique fait écho à une évolution des paradigmes politiques : il s'agit désormais de prouver que la Suisse dispose d'un réservoir linguistique diversifié et adapté aux défis d'une économie de marché globalisée, que la diversité linguistique ne constitue pas une entrave à la cohésion nationale mais un atout (cf. Duchêne et al. 2018; Humbert sous presse). Les idées sur le plurilinguisme changent par endroit, la statistique s'y adapte.

### **3. Les degrés de variations linguistiques sondés**

La statistique constitue un terrain propice à l'analyse des processus de catégorisations linguistiques. Dans les enquêtes et recensements, l'élaboration d'unités quantifiables identifiées comme des « langues » est ancrée dans des contextes de production empreints d'idéologies langagières. Dès lors, ce qui est communément appelé un dialecte, un patois, un parler ou un créole dans certaines sociétés ne fait pas toujours l'objet de distinction, c'est-à-dire que certaines variétés linguistiques n'apparaissent pas dans les résultats parce qu'elles ne sont pas comptées comme des langues. L'histoire montre que certaines catégories linguistiques font preuve d'une élasticité remarquable, engendrant des conséquences socio-politiques réelles pour leurs usager·ère·s.

#### **3.1 De l'entrave à l'unité nationale à la production d'un savoir encyclopédique universel**

La France est l'un des premiers pays à avoir cherché à quantifier des variétés linguistiques sur son territoire. Qu'il s'agisse de l'enquête de l'abbé Grégoire (1790-1792) ou de celle de Coquebert de Montbret (1806-1812), ces deux études partagent un objectif similaire : produire un savoir encyclopédique quantifié sur la base de questionnaires, afin de localiser les patois, dialectes et langues régionales qui constitueraient une entrave à la logique centralisatrice de l'Empire (Bulot 1989; Filhon 2016). Si les deux enquêtes ne sont pas réalisées avec la même rigueur ni dans les mêmes conditions, elles partent toutes deux du principe que les variétés linguistiques déviant du standard national posent un problème de gouvernance politique.

Aujourd'hui, d'autres pays relèvent des variétés linguistiques et ils cherchent plutôt à valoriser ces différences. L'exercice s'avère particulièrement complexe pour de multiples raisons. Calvet (2007) démontre que les répondant·e·s ne savent pas toujours quel nom donner à leur langue, ou bien qu'ils-elles emploient des glottonymes renvoyant à des caractéristiques non linguistiques mais nationales ou géographiques (par exemple, « africain » ou « congolais »). En outre, certains noms de langue deviennent problématiques suite à des conflits politiques. Si l'utilisation du terme « serbo-croate » fait encore débat aujourd'hui, cela est avant tout dû aux guerres de Yougoslavie qui ont incité les nouveaux Etats à se réapproprier les enjeux de politique linguistique en divisant cette langue en quatre : le bosniaque, le croate, le monténégrin et le serbe (cf. Bieber 2015). En Suisse, les autorités essaient de quantifier les dialectes nationaux (Duchêne et al. 2018) et à l'Ile Maurice, le recensement explicite que le créole compte comme une langue à part entière (Rajah-Carrim 2005). Toutes ces études démontrent que, vu le caractère formel des sondages nationaux et la dépréciation historique de certaines variétés linguistiques, il n'est pas toujours aisé de relever ces données avec précision, même si les autorités souhaitent décomplexer ses locuteur·trice·s en les incitant à nommer leur dialecte ou leur créole.

Pourtant, internet regorge d'informations quantifiées sur des langues perçues comme des variétés dans certaines sociétés, indépendamment de jugements de valeurs influençant leur statut socio-politique. Il suffit par exemple de pianoter « cucapá », « arbëreshë » ou encore « francoprovençal » sur Google ou Wikipédia pour découvrir dans quelles régions du monde ces langues sont parlées et combien de locuteur·trice·s elles comptent. Les internautes ont rarement conscience que ces informations proviennent presque systématiquement de la même source encyclopédique universelle : *Ethnologue*.

### 3.2 Un héritage colonial

À ce jour, *Ethnologue* est l'inventaire le plus complet des langues du monde (Kamusella 2012). À l'origine, les données d'*Ethnologue* sont récoltées par le Summer Institute of Linguistics (SIL) pour permettre à la Wycliffe Bible Translators de planifier son travail de traduction et de diffusion de l'évangile. L'idée de William Cameron Townsend, le fondateur de ces deux organisations, était de créer une complémentarité entre un projet scientifique et une entreprise missionnaire (ibidem). Les travaux du SIL s'inscrivent dans la continuité de ce que les colons avaient entamé, c'est-à-dire une succession de travaux disparates peu coordonnés. En effet, dans chaque colonie, les pratiques scientifiques divergeaient sensiblement (cf. Makoni et Mashiri 2007; Errington 2008).

De cet héritage colonial – combinant science, religion et politique – découle des écarts gigantesques entre les diverses sources encyclopédiques. Selon Baggioni et Vanche-Robby (1997, 37) il existerait entre 3'000 et 20'000 langues dans le monde en fonction des sources consultées. Ces écarts sont principalement dus à la perspective occidentale des langues : les gens qui documentent, classent et comptent des langues qu'ils ne connaissent pas le font sur la base d'une vision essentialiste des langues. Ils·Elles relient la langue à la nationalité, à l'ethnicité, à la race ou encore à des caractéristiques dites « tribales » (Baggioni et Vanche-Roby 1997; Makoni et Mashiri 2007; Miller 2015). L'extrait suivant, issu de *Composition linguistique des nations du monde* (1974), est révélateur de ce genre de raisonnement scientifique :

« Pour la statistique linguistique, les données sur les races sont utiles seulement quand nous savons qu'une nation est plurilingue et que les communautés linguistiques correspondent en gros aux couches raciales. » (Kloss et McConnell 1974, 20)

Cet extrait s'inscrit dans le volet méthodologique initiant une série de six volumes encyclopédiques dans lesquels sont listées toutes les langues du monde. Disposant de sources hétéroclites (recensements officiels, enquêtes indépendantes, correspondances avec le SIL...), les deux linguistes proposent une stratégie inverse aux recensements de la Prusse lorsqu'ils ne trouvent pas de chiffres sur les langues : ce n'est pas la langue qui détermine origines, ce sont les origines qui déterminent la langue. À noter que cette pratique est encore d'usage dans certains travaux statistiques récents, y compris dans le domaine de la sociolinguistique (par ex. Deumert 2010). Ainsi, pour certain·e·s scientifiques, langue, ethnicité, race et tribalité sont des caractéristiques interchangeables selon les contextes.

### 3.3 Un compte à rebours pour qui ?

Malgré la complexité et la diversité des sources sur les langues, plusieurs institutions exploitent les statistiques de langues dites « en danger » à des fins politiques sur la scène internationale. Il s'agit principalement d'organisations comme l'UNESCO, le SIL ou encore Terralingua, qui comparent la disparition des langues à une menace pour l'écosystème planétaire (Duchêne et Heller 2007).

Moore et al. (2010) sont d'avis que l'attention excessive prêtée aux statistiques des langues en danger a tendance à faire de l'ombre à l'étude des mécanismes menant à leur disparition. À cet égard, l'ethnographie de Muehlmann (2012) sur les locuteur·trice·s du cucapá au Mexique illustre en détail ce qu'elle appelle « *the malaise of enumeration* » [« le malaise de l'énumération »] : si ONG et scientifiques manifestent leur fascination pour la langue cucapá en la documentant au plus près des individus qui la parlent, les Cucapás sont en fait préoccupé·e·s par des problèmes plus fondamentaux pour leur survie en tant qu'êtres humains, comme avoir accès à l'eau courante, à la nourriture et à des soins qualité. En outre, de nombreux individus se déclarant cucapá estiment que la langue n'est pas l'élément d'identification au groupe le plus pertinent (ibidem).

Si la statistique est primordiale pour certaines minorités (cf. section 3.1), d'autres ont des problèmes plus urgents que sauver leur langue et elles ne considèrent pas nécessairement la langue comme le marqueur le plus fondamental de leur identité.

## 4. Pour une approche critique des statistiques linguistiques

À travers ces trois axes thématiques, j'ai cherché à démontrer que la statistique officielle des langues constitue un terrain de recherche propice à l'étude d'enjeux fondamentaux en sciences du langage. Dans les trois axes, nous constatons que la combinaison de raisonnements scientifiques, politiques et économiques tend à rendre certaines langues ainsi que leurs locuteur·trice·s plus visibles et à en effacer d'autres avec les nombres. Des relations de pouvoir se dessinent dans le processus de production d'un savoir quantitatif considéré comme objectif. Mon analyse cherche ainsi à faire écho à celles de Tabouret-Keller (1997, 1999) sur les processus de dénomination linguistique,

dans lesquelles elle met en exergue l'émergence de rôles individuels et institutionnels.

Dans chaque chapitre, des enjeux définitoires se manifestent à travers l'interaction d'acteur·trice·s et d'institutions engagé·e·s dans le processus de quantification des langues, chacun·e contribuant à sa manière à alimenter la part idéologique des raisonnements méthodologiques statistiques. Le chapitre 1 (supra) démontre que le choix des situations linguistiques et de leurs éventuelles combinaisons n'est pas anodin : il peut autant servir à asseoir la légitimité et la domination de la/des langue(s) de l'autorité que constituer une preuve soulignant les inégalités dont seraient victimes les minorités. Le chapitre 2 (supra) permet de comprendre que la catégorisation des bi-/plurilingues pose un problème centré sur les locuteur·trice·s et sur leur insertion dans la collectivité. Indépendamment de ses pratiques ou de ses compétences linguistiques, le sujet bilingue est tantôt perçu comme une menace pour la nation, tantôt comme un atout dont le capital linguistique pourrait se transformer en ressource culturelle et/ou économique bénéfique pour la société. Le chapitre 3 (supra) démontre que cette volonté de regrouper ou de séparer des variétés linguistiques en fonction de critères géopolitiques, linguistiques, ethniques, raciaux ou tribaux, est le fruit d'intérêts scientifiques et politiques qui ne sont pas neutres, mais ancrés dans des projets de gouvernance conditionnés par une interdépendance entre ces intérêts multiples.

Ainsi, dans l'acte de quantification des langues, on retrouve les dynamiques interactionnelles entre locuteur·trice·s, linguistes et institutions que Tabouret-Keller (1999, 89-90) décrit. Sans que leurs actes soient nécessairement prédéterminés ou intentionnels, tous ces individus tentent d'agir sur la réalité sociolinguistique qui les entoure en utilisant des nombres et en établissant des conventions sur les façons de la quantifier, tout en évoluant dans un contexte socio-politique et économique conditionnant une partie de leurs actes. Les chiffres, tout comme les raisonnements qui les façonnent, n'apparaissent plus comme une réalité quantitative déconnectée des êtres humains, mais comme l'arrangement d'intérêts scientifiques et politiques dont l'objectif est d'essayer de déterminer le rôle des langues et des locuteur·trice·s dans la société, d'associer la langue des individus à des phénomènes complexes comme « l'intégration » ou la « discrimination » sociale, en floutant parfois la frontière entre langue, ethnicité, nationalité, race et tribu.

Si ces raisonnements contribuent à dresser une certaine image des langues, leur approche critique permet d'en dévoiler une autre facette.

Certes, les statistiques linguistiques sont pratiques pour obtenir un aperçu d'une question sociolinguistique. Néanmoins, ignorer l'histoire de ces nombres, c'est risquer faire l'impasse sur la (re)production d'inégalités sociales fondées sur l'instrumentalisation de la langue en société.

## Bibliographie

- ALTUNA Olatz & URLA Jacqueline (2013), The Basque Street Survey: Two Decades of Assessing Language Use in Public Spaces. *International Journal of the Sociology of Language*, 224, 209-227
- ANDERSON Benedict (2006 [1983]), *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Edition révisée. Londres: Verso.
- AREL Dominique (2002a), Démographie et politique dans les premiers recensements post-soviétiques : méfiance envers l'État, identités en question, *Population*, 57 (6), 791-820.
- (2002b), Language categories in censuses: backward- or forward-looking? In David I. Kertzer et Dominique Arel, *Census and identity: the politics of race, ethnicity, and language in national census*, Cambridge, UK ; New York: Cambridge University Press, 92-120.
- BAGGIONI Daniel, et VANCHE-ROBY Marie (1997), Nommer, classer et dénombrer les langues : de l'encyclopédie aux thésaurus documentaires d'aujourd'hui. In Andrée Tabouret-Keller, *Les enjeux de la nomination des langues*. Le nom des langues 1. Louvain-la-Neuve: Peeters, 35-60.
- BHAVNA Dave (2004), Entitlement through numbers: nationality and language categories in the first post-Soviet census of Kazakhstan. *Nations and Nationalism*, 10 (4), 439-59.
- BIEBER Florian (2015), The Construction of National Identity and Its Challenges in Post-Yugoslav Censuses. *Social Science Quarterly*, 96 (3), 873-903.
- BRIX Emil (1982), *Die Umgangssprachen in Altösterreich zwischen Agitation und Assimilation: die Sprachenstatistik in den zisleithanischen Volkszählungen 1880 bis 1910*. Wien: Böhlau.
- BULOT Thierry (1989), L'enquête de Coquebert de Montbret et la glottopolitique de l'Empire français. *Romanischen Philologie*, 287-92.
- BUSCH Brigitta (2016), Categorizing languages and speakers: Why linguists should mistrust census data and statistics. *Working Papers in Urban Language & Literacies*, 1-18.
- CALVET Louis-Jean (2007), La (socio)linguistique au filtre de l'inventaire des langues du monde: Et quelques considérations sur ses rapports avec la sociologie. *Langue et société*, 3 (121-122), 259-73.
- CHRISTOPHER Anthony John (2011), Questions of Language in the Commonwealth Censuses. *Population, Space and Place*, 534-49.
- DESROSIÈRES Alain (2010), *La politique des grands nombres histoire de la raison statistique*. Paris: La Découverte.
- DEUMERT Ana (2010), Tracking the Demographics of (Urban) Language Shift – an Analysis of South African Census Data. *Journal of Multilingual and Multicultural Development*, 31 (1), 13-35.

- DUCHÊNE Alexandre & HELLER Monica (2007), *Discourses of Endangerment: Ideology and Interest in the Defence of Languages*. Londres: Continuum.
- DUCHÊNE Alexandre, HUMBERT Philippe N. & CORAY Renata (2018), How to ask questions on language? Ideological struggles in the making of a state survey. In Alexandre Duchêne et Philippe N. Humbert, *Survey-ing Speakers and the Politics of Census, International Journal of the Sociology of Language*, n° 252, 45-72.
- ERRINGTON James Joseph (2008), *Linguistics in a Colonial World: A Story of Language, Meaning, and Power*. Malden, Massachusetts: Blackwell Publ.
- FILHON Alexandra (2016), Les langues dans les enquêtes de la statistique publique : bilan et perspectives. *Langage et société* 155 (1), 15-38.
- FOUCAULT Michel (2004), *Sécurité, territoire, population: cours au Collège de France, 1977-1978*. Édité par Michel Senellart, François Ewald, et Alessandro Fontana. Hautes études. Paris: Seuil: Gallimard.
- HAARMANN Harald (1979), *Sprachenstatistik in Geschichte und Gegenwart*. Hamburg: Buske.
- HACKING Ian (1982), Biopower and the avalanche of printed numbers. *Humanities in society*, 5, 279-95.
- HELLER Monica (2002), *Éléments d'une sociolinguistique critique*. Paris: Didier.
- HUMBERT Philippe N. (sous presse), Un siècle de cartographie statistique des langues en Suisse: (dé)faire les frontières avec des chiffres et des idées. In Liliane Meyer Pitton et Larissa Semiramis Schedel, *Frontières linguistiques (en Suisse): nouvelles approches, perspectives critiques. Bulletin suisse de linguistique appliquée*, n° 108.
- HUMBERT Philippe N., Renata CORAY & Alexandre DUCHÊNE (2018), *Compter les langues: histoire, méthodes et politiques des recensements de population*. Rapport du Centre scientifique de compétence sur le plurilinguisme. Fribourg: Institut de plurilinguisme. <http://www.institut-mehrsprachigkeit.ch/fr/content/revue-litterature-langues-relevés-statistiques-et-politiques-linguistiques> (consulté le 31.05.2018).
- KAMUSELLA Tomasz (2012), The global regime of language recognition. *International Journal of the Sociology of Language*, 2012 (218), 59-86.
- KLOSS Heinz & MCCONNELL Grant D., (1974), *Linguistic Composition of the Nations of the World. 1: Central and Western South Asia = Composition Linguistique Des Nations Du Monde. 1: L'Asie Du Sud: Secteurs Central et Occidental*. Vol. 1. Québec: Pr. de l'univ. Laval.
- LABBÉ Morgane (2003), Dénombrer les nationalités en Prusse au XIXe siècle: entre pratique d'administration locale et connaissance de la population. *Annales de démographie historique*, 105 (1), 39-61.
- (2009), Internationalisme statistique et recensement de la nationalité au XIXe siècle. *Courrier des statistiques*, n° 127, 39-45.
- LAUR Elke (2013), Minorités majoritaires et majorités minoritaires: des oxymorons?: L'exemple des minorités de langue officielle au Canada. *Minorités linguistiques et société*, n° 3, 19-40.
- LEEMAN Jennifer (2004), Racializing language: A history of linguistic ideologies in the US Census. *Journal of Language and Politics* 3 (3), 507-34.
- (2013), Categorizing Latinos in the history of the US Census: The official racialization of Spanish. In Jose Del Valle, *A Political History of Spanish: The Making of a Language*. Cambridge: Cambridge University Press, 305-23.



- LEUSCHNER Torsten (2004), Richard Böckh (1824–1907): Sprachenstatistik Zwischen Nationalitätsprinzip Und Nationalstaat. *Historiographia Linguistica*, 31 (2), 389–421.
- LÉVY Paul (1964), Quelques problèmes de statistique linguistique à la lumière de l'expérience belge. *Revue de l'Institut de Sociologie* 37, 251–73.
- LIEBERSON Stanley (1966), Language Questions in Censuses. *Sociological Inquiry*, 36 (2), 262–79.
- MAKONI Sinfrey & MASHIRI Pedzisai (2007), Critical Historiography: Does Language Planning in Africa Need a Construct of Language as Part of Its Theoretical Apparatus? In Alastair Pennycook, *Disinventing and Reconstituting Languages*. Clevedon: Multilingual Matters, 62–89.
- MANOHARAN S. (2001), Indian languages in censuses - A study. *Indian Linguistics*, 62 (1-4), 117–40.
- MERRY Sally Engle (2016), *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago, IL London: The University of Chicago Press.
- MILLER Catherine (2015), Retour sur les catégorisations ethno-linguistiques au Soudan: entre construction allogène, appropriation autochtone et perpétuel réajustement. *Canadian Journal of African Studies / Revue canadienne des études africaines*, 49 (1), 127–46.
- MOORE Robert, PIETIKÄINEN Sari & BLOMMAERT Jan (2010), Counting the losses: numbers as the language of endangerment. *Sociolinguistic Studies*, 4(1), 1–26.
- MORENO-FERNÁNDEZ Francisco (2007), Speaker statistics in Spain. *International Journal of the Sociology of Language*, 2007 (184), 139–54.
- MUEHLMANN Shaylih (2012), Rhizomes and Other Uncountables: The Malaise of Enumeration in Mexico's Colorado River Delta: The Countdown at the End of the Colorado River. *American Ethnologist*, 39 (2), 339–53.
- PAVLENKO Aneta (2011), Language Rights versus Speakers' Rights: On the Applicability of Western Language Rights Approaches in Eastern European Contexts. *Language Policy*, 10 (1), 37–58.
- PORTER Theodore M. (1995), *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. 2<sup>e</sup> édition. History and Philosophy of Science. Princeton, N.J: Princeton Univ.Press.
- PRÉVOST Jean-Guy & BEAUD Jean-Pierre (2002), Statistical inquiry and the management of linguistic plurality in Canada, Belgium and Switzerland. *Journal of Canadian Studies*, 36 (4), 88–117.
- RAJAH-CARRIM Aaliya (2005). Language Use and Attitudes in Mauritius on the Basis of the 2000 Population Census. *Journal of Multilingual and Multicultural Development*, 26 (4), 317–32.
- SILVER Brian (1986), The Ethnic and Language Dimensions in Russian and Soviet Censuses. In Ralph S. Clem, *Research Guide to the Russian and Soviet Censuses*. *Studies in Soviet History and Society*. Ithaca, NY: Cornell Univ. Press, 70–97.
- SOLÓRZANO Ramón Jr. & AHLÉN Sonda (2010), Latino questions on race, ethnicity, and language at the advent of the 2010 census. *Harvard Journal of Hispanic Policy*, 22, 17–45.
- TABOURET-KELLER Andrée (1997), *Les enjeux de la nomination des langues*. Le nom des langues 1. Louvain-la-Neuve: Peeters.
- (1999), Le nom des langues : un ambassadeur aveugle, ignorant de ses missions. *Revue des Sciences Sociales de la France de l'Est*, n° 26, 88–93.
- URLA Jacqueline & BURDICK Christa (2018), Counting matters: Quantifying the vitality and value of Basque. In Alexandre Duchêne et

- Philippe N. Humbert, *Survey-ing Speakers and the Politics of Census*, *International Journal of the Sociology of Language*, n° 252, 73-96.
- de VRIES John (1990), On coming to our census: a layman's guide to demolinguistics. *Journal of Multilingual and Multicultural Development*, 11 (1 & 2), 57-76.
- WOOLARD Kathryn A. & SCHIEFFELIN Bambi B. (1994) Language Ideology. *Annual Review of Anthropology*, 23 (1), 55-82.

Preprint