

Compter les langues: histoire, méthodes et politiques des recensements de population

Sprachen zählen: Geschichte, Methoden und Politik der Volkszählungen

Contare le lingue: storia, metodi e politiche dei censimenti della popolazione

Dumbrar linguas: istorgia, metodos e politica da dumbraziuns dal pievel

Counting languages: history, methods and politics of population censuses

Executive Summary

**Philippe Humbert, Renata Coray, Alexandre Duchêne
2018**

Publié par | Herausgeber

Institut de plurilinguisme
www.institut-plurilinguisme.ch

—
Institut für Mehrsprachigkeit
www.institut-mehrsprachigkeit.ch

Auteurs | AutorInnen

Philippe Humbert, Renata Coray, Alexandre Duchêne

Traduction | Übersetzung

Susanne Obermayer, Francesco Screti, Lia Rumantscha, Corinne Feghali

Le projet dont il est question a été réalisé dans le cadre du programme de travail 2016–2020 du Centre scientifique de compétence sur le plurilinguisme. La responsabilité du contenu de la présente publication incombe à ses auteurs.

Das vorliegende Projekt wurde im Rahmen des Arbeitsprogramms 2016–2020 des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit durchgeführt. Für den Inhalt dieser Veröffentlichung sind die AutorInnen verantwortlich.

Fribourg | Freiburg, 2018**Layout**

Billy Ben, Graphic Design Studio

Avec le soutien de | Unterstützt von

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Dipartimento federale dell'interno DFI
Departament federal da l'intern DFI
Bundesamt für Kultur BAK
Office fédéral de la culture OFC
Ufficio federale della cultura UFC
Uffizi federal da cultura UFC

Compter les langues : histoire, méthodes et politiques des recensements de population

Une synthèse de la revue de la littérature

Sprachen zählen: Geschichte, Methoden und Politik der Volkszählungen

Ein kurzer Abriss

Contare le lingue: storia, metodi e politiche dei censimenti della popolazione

Sintesi della rassegna della letteratura

Dumbrar linguas: istorgia, metodos e politica da dumbraziuns dal pievel

Ina sintesa da la revista da litteratura

Counting languages: history, methods and politics of population censuses

Executive Summary of literature review

Philippe Humbert, Renata Coray, Alexandre Duchêne

2018

Rapport du Centre scientifique de compétence sur le plurilinguisme
Bericht des Wissenschaftlichen Kompetenzzentrums für Mehrsprachigkeit
Rapporto del Centro scientifico di competenza per il plurilinguismo
Rapport dal Center scientific da competenza per la plurilinguitad
Report of the Research Centre on Multilingualism

Index

Français	Compter les langues : histoire, méthodes et politiques des recensements de population	5
-----------------	--	---

1	Introduction	6
2	Les recensements des langues à travers l'histoire	8
3	Les manières de recenser les langues	10
4	Réception et appropriation des résultats statistiques sur les langues	12
5	Conclusions	14

Deutsch	Sprachen zählen: Geschichte, Methoden und Politik der Volkszählungen	17
----------------	---	----

1	Einleitung	18
2	Sprachenerhebungen: ein historischer Überblick	20
3	Wie Sprachen erhoben werden	22
4	Sprachenstatistiken: Rezeption und Aneignung	24
5	Fazit	26

Italiano	Contare le lingue: storia, metodi e politiche dei censimenti della popolazione	29
-----------------	---	----

1	Introduzione	30
2	I censimenti delle lingue nella storia	32
3	I modi di censire le lingue	34
4	Ricezione e appropriazione dei risultati dei censimenti linguistici	36
5	Conclusioni	38

Rumantsch	Dumbrar linguas: istorgia, metodas e politica da dumbraziuns dal pievel	41
------------------	--	----

1	Introducziun	42
2	Las relevaziuns da datas davart las linguas en il decurs da l'istorgia	44
3	Las modas da relevar datas davart las linguas	46
4	Recepziun ed appropriaziun dals resultats statistics davart las linguas	48
5	Conclusiuns	50

English	Counting languages: history, methods and politics of population censuses	53
----------------	---	----

1	Introduction	54
2	Language censuses through history	56
3	Methods to collect data on language	58
4	Reaction to and appropriation of statistical results on languages	60
5	Conclusions	62

Compter les langues: histoire, méthodes et politiques des recensements de population

Une synthèse de la revue de la littérature

—
Philippe Humbert, Renata Coray, Alexandre Duchêne

1 Introduction

Dans de nombreuses sociétés plurilingues, les statistiques officielles des langues jouent un rôle déterminant dans l'organisation des espaces langagiers. En effet, dans certains pays, les chiffres sur le nombre de locuteurs monolingues ou bilingues ont des conséquences politico-économiques directes, notamment dans le cadre de l'attribution de soutiens financiers aux minorités linguistiques. Les statistiques linguistiques constituent également un savoir constamment consulté – et parfois remis en question – par de nombreux scientifiques qui s'en servent pour analyser des phénomènes démographiques en lien avec les langues. Étant donné l'omniprésence des statistiques des langues dans le monde scientifique et politique, le Centre scientifique de compétences sur le plurilinguisme s'est donné la tâche d'élaborer un « State of the Art » présentant l'état de la littérature scientifique sur les recensements des langues et leurs interprétations dans l'espace social.

La présente synthèse de cette revue de littérature ne contient aucune information bibliographique ; nous n'avons pas souhaité réduire l'ensemble des références à une part peu représentative de la littérature. La bibliographie complète figure dans la version intégrale de la revue de littérature (en français, accessible en ligne sur <https://goo.gl/iRBxKr>). La revue de littérature, dont nous résumons ici le contenu, est d'utilité pour les chercheurs travaillant avec des données issues

de recensements : elle offre une meilleure compréhension des enjeux sociopolitiques liés à la création et à l'utilisation de statistiques linguistiques. En décrivant l'état de la recherche dans divers contextes, elle fournit un point de comparaison international nécessaire à une meilleure compréhension de la position helvétique en matière de recensement des langues. Il convient de préciser que seules les statistiques des langues élaborées par une autorité officielle – au niveau national ou régional – ont été prises en compte. En général, il s'agit de recensements de la population au sein desquels la langue figure au menu des catégories démographiques comme l'âge, le sexe ou encore la nationalité.

La revue de littérature couvre des travaux de recherche issus de disciplines variées : sociolinguistique, sociologie, politique, démographie et géographie figurent parmi les matières dans lesquelles les recensements et les statistiques des langues ont été le plus fréquemment abordés. Partant d'ouvrages et d'auteurs les plus cités, nous avons systématiquement consulté les bibliographies proposées et avons redoublé d'efforts pour offrir aux lecteurs et aux lectrices un aperçu des plus internationaux. Les auteurs consultés se sont en effet tous penchés sur les données linguistiques des recensements de population réalisés à divers échelons – nationaux ou régionaux – dans diverses parties du monde. À cette première

immersion globale dans le sujet s'ajoute la consultation complète d'une trentaine de revues de sociolinguistique, de politique et de démographie en français, en anglais, en allemand, et sporadiquement en russe, en espagnol, en italien, en romanche et en serbo-croate. À l'aide d'une série de mots-clés plurilingues, nous avons encore approfondi la récolte sur des moteurs de recherche spécialisés. Dans plusieurs cas, la langue n'apparaissait pas comme le sujet d'étude principal, mais comme une composante dite « ethnique », « indigène » voire « raciale ». Par conséquent, même si la langue occupait un rôle central dans l'élaboration de statistiques ethniques et dans la mise en œuvre de projets politiques, les auteurs y voyaient plutôt une caractéristique non linguistique. Ces critères de sélection rendent compte de l'étendue de la recherche que nous avons menée tout autant que de ses limites.

Certains pays et certaines régions apparaissent plus fréquemment que d'autres dans la revue de littérature. Cela n'est pas seulement dû aux langues que nous avons utilisées pour notre recherche, mais aussi à la place qu'occupent les recensements des langues dans la société. En effet, tous les pays n'effectuent pas de recensement des langues avec autant d'assiduité que la Suisse ou le Canada. Dans certains États, pourtant composés de quantité de variétés linguistiques comme la France et la Chine, aucun recensement de population ne contient de questions sur les langues. On peut cependant y trouver des statistiques linguistiques récoltées par le biais d'estimations ou sur la base d'enquêtes par échantillonnage. Comme nous avons pu le constater, l'apparition ou non d'une question sur la ou les langues des

répondants est souvent justifiée par des motivations politiques, car suivant le contexte, la statistique des langues – tout comme son absence – peut devenir source de conflit.

Dans un premier temps, la revue de littérature et la présente synthèse abordent le sujet dans sa dimension historique, afin de discerner les fonctions initiales et l'évolution du recensement des langues en divers contextes. Dans un deuxième temps, les aspects méthodologiques de la production des statistiques linguistiques sont traités pour offrir l'éventail des manières de catégoriser et de circonscrire les variétés linguistiques des locuteurs. Dans un troisième temps, la réception et l'appropriation des résultats statistiques sur les langues figurent au centre des analyses, lesquelles engendrent parfois des débats hautement politisés – comme au Pays basque, en Belgique ou au Canada – et reposent sur des arguments scientifiques et politiques. Pour terminer, nous synthétisons trois aspects des recensements des langues – l'histoire, la méthode et la réception des résultats – afin de dégager les principaux enjeux et les limites des statistiques linguistiques officielles.

2 Les recensements des langues à travers l'histoire

Dès le XIX^e siècle, les statistiques des langues s'inscrivent dans des projets de gouvernance qui allient sciences et politique dans les processus de comptage. Les méthodes de quantification changent en fonction des objectifs des enquêtes et des recensements. Dans tous les cas, les chiffres sur les langues participent de la construction des États-nations, de la définition d'une identité nationale collective et de la distribution des ressources.

Les premiers relevés statistiques exhaustifs sur les langues ont été réalisés au milieu du XIX^e siècle en Europe. Auparavant, la France avait déjà effectué des estimations sur la présence des dialectes et des patois sur son territoire dans le but d'identifier les régions dans lesquelles le français n'était pas encore majoritairement utilisé. L'objectif était à la fois de constituer un savoir encyclopédique sur les pratiques linguistiques en France et de désigner les régions où les dialectes et les patois devaient être éradiqués à l'aide des structures scolaires et administratives.

Durant les sessions du *Congrès international de statistique* (1853-1876), deux méthodes de recensement des langues sont discutées : celle qui cherche à sonder « la langue maternelle » des répondants et celle qui se penche sur « la langue parlée ». Ces deux conceptions sont principalement débattues entre les statisticiens autrichiens et

prussiens. Les Prussiens avancent que le relevé de la « langue maternelle » constitue le moyen le plus objectif de révéler la nationalité des répondants ; les Autrichiens sont d'avis que langue et nationalité ne doivent pas être confondues, puisqu'elles ne coïncident pas nécessairement. Plusieurs problèmes, encore débattus de nos jours, sont discutés durant les sessions du Congrès. Les principales préoccupations concernent les degrés de distinctions linguistiques (langues, dialectes, patois) et la manière de relever les langues des individus bilingues.

Dans les faits, l'Autriche va recenser la « Umgangssprache » (langue parlée communément) et la Prusse la « Muttersprache » (langue maternelle). En relevant la langue parlée communément, le gouvernement autrichien s'attire les foudres de nombreuses minorités slaves qui dénoncent une tentative de les occulter des résultats : cette question contribue à souligner les pratiques de l'allemand – langue majoritairement utilisée dans les rapports formels – et à effacer la présence d'autres langues au sein de l'Empire. Quant à la Prusse, si les statisticiens comptent démontrer l'hégémonie de la langue allemande sur l'ensemble du territoire avec les résultats de la « Muttersprache », certaines minorités slaves émettent des revendications à l'est du pays sur la base des mêmes chiffres. Dans les deux cas, les statisticiens avancent

une vision « objective » du dénombrement des langues ; dans les deux cas, les chiffres deviennent une source de conflits politiques.

Durant le XX^e siècle, de nombreux recensements posent une/des question(s) sur les langues. La langue des individus n'est pas uniquement interprétée comme une donnée linguistique, mais aussi comme un substitut de la nationalité, de l'ethnicité, voire de la race. À ce titre, les langues font l'objet d'une attention particulière dans les recensements coloniaux. En témoignent les pratiques britanniques en Inde, où le relevé des langues constitue une entreprise anthropologique colossale dont les résultats forment une référence encyclopédique autant pour les scientifiques que pour l'administration (post-)coloniale.

L'URSS entame aussi des recherches anthropologiques de grande envergure par le biais des recensements de population. Si les statistiques des langues servent d'abord à clarifier l'appartenance à des groupes ethniques de nombreux individus, elles sont également utilisées pour démontrer la progression du russe en tant que langue de communication inter-ethnique. Après la dissolution de l'URSS, la majorité des recensements post-soviétiques renverse les priorités : l'objectif est généralement de minimiser la présence du russe et de démontrer que la langue autrefois secondaire est devenue la langue dominante dans ces pays en transition. Aux États-Unis, les recensements des langues cherchent aussi à obtenir plus d'informations d'ordre à la fois ethnique, racial et linguistique. Si l'un des objectifs est d'identifier des groupes d'individus non anglophones pour déterminer les zones où il faudrait fournir du matériel de vote dans d'autres langues que l'anglais, les statistiques américaines

font également l'objet de dérives discriminatoires par le biais de la création du concept de « foyers linguistiquement isolés ».

Si les gouvernements peuvent voir dans les statistiques linguistiques un moyen d'imposer des politiques linguistiques, les minorités linguistiques y voient un moyen de défendre leurs intérêts. Devenue indépendante du Royaume-Uni, l'Irlande se réapproprie la question des connaissances de la langue irlandaise pour élaborer des statistiques qui servent à déterminer quelles régions devraient être soutenues par le gouvernement pour revitaliser l'irlandais, alors que les Britanniques voyaient en cette même question un moyen de mesurer la progression de l'anglais sur les territoires qu'ils administraient. L'Amérique latine fournit d'autres exemples de ce type, notamment au Mexique, où les statistiques des langues sont d'abord utilisées pour asseoir l'hégémonie de la langue espagnole, puis pour définir les minorités dites « indigènes » qui nécessitent le soutien du gouvernement.

À partir des années 1960, le dénombrement des minorités linguistiques connaît un engouement international tant sur la scène scientifique que politique. Des disciplines scientifiques spécifiques, utilisant des statistiques linguistiques, se font ainsi une place dans les milieux académiques. Il s'agit principalement de la géolinguistique et de la démolinguistique (ou géographie et démographie des langues). La contextualisation historique de l'émergence de ces matières, plus particulièrement de la démolinguistique au Canada, permet de comprendre l'idéologie politique à la base du développement de ces disciplines scientifiques spécifiques.

3

Les manières de recenser les langues

Les statistiques linguistiques officielles dépendent de nombreux critères méthodologiques. La comparaison des chiffres mentionnés dans plusieurs sources encyclopédiques montre que le nombre de langues et de locuteurs indiqués peut varier énormément. Ces différences sont en grande partie dues aux manières dont les langues ont été documentées et quantifiées. En effet, selon les sources, certaines variétés linguistiques sont regroupées sous une seule catégorie alors que d'autres peuvent apparaître sous plusieurs noms ou encore ne pas figurer du tout. Ces divergences proviennent en partie de la nature des processus de documentation linguistique, qui inscrivent le volet scientifique dans des projets de gouvernance politique. Qu'une langue soit considérée comme telle ou non dans un recensement est également conditionné par un processus de catégorisation des langues et de circonscription des locuteurs, précédant ou accompagnant l'exercice statistique. En témoigne l'ampleur des travaux accomplis par le Summer Institute of Linguistics (SIL) qui cherche à documenter toutes les langues du monde et dont l'objectif initial est de traduire la bible dans un maximum de langues à des fins d'évangélisation.

Au niveau des recensements et des enquêtes statistiques officielles, d'autres aspects méthodologiques diffèrent dans la collecte de données sur les langues. Ils doivent être pris en considération pour com-

prendre les limites de ce qui peut être analysé à l'aide des chiffres. Il s'agit entre autres de la taille et des caractéristiques des échantillons étudiés. En effet, une enquête par échantillonnage ne permet pas de dire les mêmes choses qu'un recensement exhaustif de la population, d'autant plus si l'échantillon écarte certaines franges de la population. De plus, tous les relevés statistiques ne sont pas réalisés à la même fréquence (par exemple, certains recensements de population ont lieu tous les cinq ans, d'autres tous les dix ans). Le fait de poser des questions par écrit ou par oral induit également des différences, tout comme les types de questions (ouvertes ou fermées). L'âge à partir duquel un individu peut répondre à ces questions et la/les langue(s) utilisée(s) pour collecter ces données font partie de la multitude de critères qui ont un impact sur la portée interprétative des résultats. Dans la littérature scientifique, il apparaît que le plurilinguisme individuel n'est pas systématiquement pris en compte. Lorsque les réponses permettent d'indiquer plus d'une langue, le plurilinguisme n'est pas quantifié de la même manière partout non plus.

La formulation des questions sur les langues détermine en grande partie quelle(s) langue(s) du répertoire des individus va/vont figurer dans les résultats. La littérature scientifique distingue les questions qui sont orientées vers : les origines du répondant ;

les pratiques linguistiques du répondant au moment du relevé statistique (en distinguant parfois des domaines d'usage spécifiques) ; les connaissances ou les compétences linguistiques du répondant ; les attitudes et opinions du répondant par rapport à certaines variétés linguistiques. Certains relevés statistiques comptent une seule question sur les langues, d'autres en posent plusieurs. À noter que certaines statistiques linguistiques officielles sont élaborées sur la base de données non linguistiques, telles que la nationalité, l'ethnicité ou la race, ou bien il peut s'agir d'estimations calculées selon le nombre d'habitants domiciliés par région. La terminologie utilisée pour se référer aux questions sur les langues – « langue maternelle », « langue natale » ou encore « langue parlée à la maison » – suit des définitions fluctuantes. Par exemple, le relevé de la « langue maternelle » peut renvoyer à : la première langue apprise durant l'enfance et toujours comprise par le répondant au moment du recensement (Canada) ; la langue parlée par la mère au répondant (Inde, République tchèque et Slovaquie) ; la langue utilisée à la maison (États-Unis) ; la langue dans laquelle le répondant pense et qu'il dit maîtriser le mieux (Suisse).

4 Réception et appropriation des résultats statistiques sur les langues

Les statistiques linguistiques officielles sont utilisées de manière récurrente dans les milieux scientifiques et politiques, sans qu'une frontière nette ne sépare ces deux domaines. Au contraire, dans de nombreux cas, un dialogue s'établit entre ces deux milieux. La littérature scientifique regorge de situations dans lesquelles les statistiques linguistiques constituent un enjeu de reconnaissance politique au sein de la société, notamment par le biais de débats sur la territorialité linguistique et/ou sur les droits linguistiques (accès aux services officiels, traductions, soutien de langues minoritaires en tant que langues de scolarisation, etc.). Les débats politiques renvoient fréquemment à des discussions sur la fixation de seuils ou de quotas pour déterminer à partir de quels pourcentages ou nombres absolus certaines minorités linguistiques se voient octroyer des droits ou subventions spécifiques. La fixation de ces seuils ou quotas soulève généralement des débats hautement politisés. Par ailleurs, il n'existe pas de convention dans l'élaboration de ces seuils : les décisions sont prises différemment au sein de chaque État ou région. Les statistiques des langues sont également exploitées pour identifier des populations. *A contrario*, leur absence peut révéler la volonté d'un gouvernement de ne pas donner de visibilité à certaines mino-

rités, comme l'illustrent les statistiques françaises et turques, qui ne relèvent aucune donnée sur les langues pour éviter de donner une voix à leurs minorités et renforcer le sentiment d'unité nationale derrière une seule langue. Dans les cas américain et australien, l'objectif initial est de repérer la part des non anglophones pour savoir où il conviendrait d'apporter un soutien linguistique dans les relations officielles. Dans ces deux cas, le processus d'interprétation des statistiques linguistiques a connu des dérives discriminatoires.

Dans les milieux scientifiques, les statistiques officielles des langues sont accueillies très différemment selon les chercheurs qui les consultent, voire les exploitent. La plupart des statistiques linguistiques sont critiquées pour le manque d'informations relevées sur le bilinguisme ou le plurilinguisme individuel, car la plupart des relevés statistiques offrent la possibilité de n'indiquer qu'une seule langue. Les lacunes des recensements ont incité de nombreux chercheurs à pallier ces manquements et à vérifier la validité et la pertinence des résultats statistiques en les comparant à d'autres données statistiques, telles que les registres des écoles.

Les chercheurs mentionnent encore d'autres limites selon les statistiques qu'ils consultent. Par exemple, certains auteurs

soulignent que le relevé des langues parlées à la maison n'est pas représentatif de toutes les pratiques linguistiques d'un individu, puisqu'il est possible de s'exprimer dans d'autres langues en dehors du foyer, surtout pour les populations issues de la migration. Les statistiques sur les connaissances ou les compétences linguistiques des répondants nécessitent aussi quelques précautions, car sans descripteurs précis de ce qui est entendu par connaissance ou compétence dans les langues sondées, les déclarations des répondants sont susceptibles de manquer de consistance et de fiabilité.

Les statistiques linguistiques officielles sont utilisées dans de très nombreuses études sur les transferts et le maintien des langues, à savoir sur le fait de continuer à utiliser la même langue toute sa vie ou, au contraire, de passer à une autre langue ou à en acquérir d'autres. Ceci dans le but d'observer l'évolution de groupes de locuteurs au niveau macro-sociolinguistique. Ces analyses distinguent parfois les domaines d'usage et sont croisées à d'autres variables comme le taux de fécondité ou de mortalité. Elles peuvent avoir des répercussions politiques, comme au Canada où certains chercheurs transforment leur argument scientifique en revendication politique.

Les résultats ont aussi été interprétés comme révélateurs d'attitudes par rapport aux langues et aux locuteurs. Dans le contexte des recensements post-yougoslaves par exemple, le fait de répondre « serbo-croate » ou « serbe » ou « croate » aux questions portant sur la langue maternelle est perçu comme une évolution de l'appropriation des dénominations linguistiques en relation avec des situations socio-politiques spécifiques. Au

niveau international, la question du comptage des langues en danger, c'est-à-dire de langues qui seraient de moins en moins utilisées, divise les chercheurs. Certains comparent la disparition de ces langues à une menace sur la biodiversité humaine, d'autres estiment que ces décomptes passent à côté de préoccupations plus importantes pour les populations concernées.

5 Conclusions

À travers tous les recensements et enquêtes sur les langues que nous avons abordés dans notre revue de littérature, l'argument statistique combine des enjeux à la fois scientifiques et politiques. Certains États ont fortement développé l'outil statistique, quitte à le décliner en plusieurs types d'enquêtes. À cet égard, l'historique des statistiques linguistiques telles qu'élaborées aux Canada démontre à quel point les méthodologies se sont affinées sous l'influence de revendications combinant des arguments scientifiques et politiques : suite à de constantes critiques concernant la finesse des données sur les langues dans les recensements canadiens, des questions sont progressivement ajoutées pour répondre à des exigences reposant sur un argumentaire scientifique. À plus petite échelle, les statistiques du basque trouvent leur essor à travers des enquêtes sociolinguistiques officielles, dont les méthodes s'éloignent progressivement des enquêtes traditionnelles pour essayer de saisir quantitativement les usages de la langue basque dans les rues des territoires basquophones.

Malgré une diversité évidente dans les manières de relever des caractéristiques plus ou moins linguistiques, la plupart des recensements et enquêtes poursuivent un objectif relativement similaire : la quantification des langues comme l'expression d'une caractéristique socialement située, laquelle n'est pas sans conséquences dans l'orga-

nisation politique et dans les revendications qui ne se limitent pas à des enjeux linguistiques. Les chiffres peuvent devenir le cheval de bataille de minorités souhaitant obtenir une forme de reconnaissance sur un territoire, avoir accès à certaines ressources et à des droits, comme la possibilité de voir ses enfants scolarisés dans une langue plutôt qu'une autre ou encore de définir la/les langue(s) officielle(s) d'administration à un niveau régional et/ou national. Selon les régimes linguistiques, les statistiques des langues peuvent avoir un impact plus ou moins direct sur l'allocation de ressources – pas uniquement d'ordre linguistique – dans des contextes plurilingues, notamment par le biais de la définition de seuils minimaux à partir desquels une minorité peut obtenir gain de cause.

Selon l'impact des statistiques sur l'organisation de la vie politique, la façon même de récolter les données sur les langues et de fixer des seuils minimaux peut faire l'objet de débats houleux. Par exemple, en Belgique, les résultats des recensements avaient un impact direct sur la fixation des langues officielles des communes. La publication des statistiques linguistiques a tant suscité de tensions entre Wallons et Flamands que les autorités ont décidé de supprimer la question linguistique des formulaires et de fixer les frontières linguistiques une fois pour toute. Dans d'autres pays comme la Suisse, c'est le principe de territorialité

linguistique qui prime, signifiant que les statistiques officielles importent peu : les habitants d'un territoire sont supposés s'adapter à la/aux langue(s) officielle(s) de leur commune ou canton de domicile. Seul le canton des Grisons utilise les données statistiques pour la mise en œuvre des politiques linguistiques élaborées selon des seuils, optant ainsi pour un principe de territorialité dynamique qui continue de nourrir les débats politiques aujourd'hui.

Plusieurs scientifiques ont cherché des moyens de vérifier la validité des données statistiques sur les langues ou ont développé d'autres types d'enquêtes approfondies pour nuancer les chiffres des recensements. De nombreuses applications ont aussi été élaborées. La plus évidente est celle des études des transferts linguistiques et du maintien des langues à échelle démographique, lesquelles peuvent se retrouver sous d'autres intitulés, tels que les tentatives de mesures de la vitalité des langues. Dans les comparaisons internationales, les études sur la vitalité ou les transferts des langues suscitent de nombreux débats épistémologiques qui renvoient à des désaccords d'ordre méthodologique et idéologique.

Les débats entraînés par les statistiques sur les langues ne se limitent pas au monde scientifique ou à de hautes sphères politiques. Ces statistiques ont en effet des conséquences réelles sur la distribution des ressources, là où le plurilinguisme est un sujet politique. La création de fonds financiers pour la mise en œuvre de traductions et de services administratifs qui prennent en considération les demandes de minorités linguistiques ne va pas toujours de soi : dans certains cas, les statis-

tiques linguistiques officielles servent à trancher lorsque ces sujets posent problème. Pourtant, alors même que les statistiques sont présentées comme une solution objective à laquelle recourir pour résoudre des tensions politiques, la revue de littérature démontre que la science, lorsqu'elle est utilisée à ces fins, reste fondamentalement influencée par des enjeux de gouvernance.

Sprachen zählen: Geschichte, Methoden und Politik der Volkszählungen

Ein kurzer Abriss

—
Philippe Humbert, Renata Coray, Alexandre Duchêne

1 Einleitung

In zahlreichen mehrsprachigen Gesellschaften spielen amtliche Statistiken eine massgebliche Rolle bei der Organisation von Sprachräumen. In einigen Ländern haben Angaben zur Anzahl von ein- oder zweisprachigen Sprecherinnen und Sprechern sogar direkte politische oder wirtschaftliche Konsequenzen, namentlich auf die finanzielle Unterstützung von Sprachminderheiten. Sprachenstatistiken stellen zudem ein von zahlreichen Wissenschaftlerinnen und Wissenschaftlern laufend konsultiertes – und manches Mal infrage gestelltes – Wissensreservoir für die Analyse von demografischen Phänomenen im Zusammenhang mit Sprachen dar. Angesichts der Allgegenwart von Sprachenstatistiken in Wissenschaft und Politik legt das Wissenschaftliche Kompetenzzentrum für Mehrsprachigkeit nun eine „State-of-the-Art“-Literaturübersicht zu Sprachenerhebungen und ihren gesellschaftlichen Interpretationen vor.

Die vorliegende Zusammenfassung unserer Literaturübersicht verzichtet auf bibliografische Angaben, um die Gesamtheit der Referenzen nicht auf einen wenig repräsentativen Teil der Literatur zu reduzieren. Die vollständige Bibliografie ist in der ausführlichen Fassung der Literaturübersicht zu finden (französischsprachige Version, online zugänglich unter <https://goo.gl/Ga3AJd>).

Unsere Literaturübersicht, die hier kurz umrissen wird, richtet sich an Forschende, die sich in ihrer Arbeit auf statistische

Daten stützen, und soll zu einem besseren Verständnis soziopolitischer Fragen rund um die Erstellung und Verwendung von Sprachenstatistiken beitragen. Ausgehend von der Darstellung des Forschungsstands in unterschiedlichen Bereichen bietet sie einen internationalen Vergleich zum besseren Verständnis des schweizerischen Ansatzes bei Sprachenerhebungen. Wir weisen darauf hin, dass wir lediglich von amtlichen Stellen auf nationaler oder regionaler Ebene erstellte Sprachenstatistiken berücksichtigt haben. In der Regel handelt es sich dabei um Volkszählungen, bei denen Sprache als demografische Kategorie neben Alter, Geschlecht oder Nationalität erhoben wird.

Die Literaturübersicht berücksichtigt Forschungsarbeiten unterschiedlicher Disziplinen; am häufigsten werden Sprachenerhebungen und -statistiken in der Soziolinguistik, Soziologie, Politologie, Demografie und Geografie verwendet. Ausgehend von den meist zitierten Werken und Autoren haben wir vertiefte systematische Literaturrecherchen vorgenommen, um unseren Leserinnen und Lesern eine breite internationale Darstellung vorzulegen. Die berücksichtigten Werke befassen sich alle mit Sprachendaten aus Volkszählungen, die in unterschiedlichen Teilen der Welt auf nationaler oder regionaler Ebene durchgeführt wurden. Ergänzend zu dieser ersten globalen Übersicht haben wir etwa dreissig Zeitschriften aus den Bereichen Soziolinguistik, Politologie

und Demografie in Französisch, Englisch, Deutsch und punktuell Russisch, Spanisch, Italienisch, Rätoromanisch und Serbokroatisch sowie spezialisierte Suchmaschinen anhand mehrsprachiger Schlagworte umfassend konsultiert. In mehreren Fällen war Sprache nicht das Hauptthema von Studien, sondern Hinweis auf die sog. „ethnische“, „indigene“ oder „Rassen“- Zugehörigkeit. Folglich spielte Sprache bei der Erstellung von ethnischen Statistiken und der Umsetzung von politischen Projekten zwar eine zentrale Rolle, wurde von den Autoren jedoch eher als nicht-sprachliches Merkmal angesehen. Unsere Auswahlkriterien verdeutlichen somit die Breite unserer Suche, gleichzeitig aber auch ihre Grenzen.

Einige Länder und Regionen erscheinen häufiger als andere in unserer Literaturübersicht. Dies liegt freilich nicht nur an den für unsere Untersuchung verwendeten Sprachen, sondern auch am gesellschaftlichen Stellenwert von Sprachenerhebungen, die nicht alle Länder mit derselben Akribie wie die Schweiz oder Kanada erstellen. In Staaten wie Frankreich oder China gibt es zwar unzählige sprachliche Varietäten, in Volkszählungen werden jedoch keine Fragen zur Sprache gestellt. Dennoch ist es möglich, in solchen Fällen Sprachenstatistiken zu finden, die mittels Schätzungen oder Stichproben erstellt wurden. Dies verdeutlicht, dass die Tatsache, ob eine Frage zu der oder den Sprachen der Befragten gestellt wird, meist politisch motiviert ist. Je nach Kontext kann das Vorliegen – oder das Fehlen – einer Sprachenstatistik nämlich zu Konflikten führen.

Im ersten Teil unserer Literaturübersicht sowie der vorliegenden Zusammenfassung gehen wir auf historische Aspekte ein, um die

ursprünglichen Funktionen von Sprachenerhebungen sowie ihre weiteren Entwicklungen in unterschiedlichen Zusammenhängen zu umreissen. Der zweite Teil widmet sich methodologischen Aspekten der Erstellung von Sprachenstatistiken, um die vielen möglichen Ansätze zur Kategorisierung und Abgrenzung von Sprachvarietäten der Sprecherinnen und Sprecher aufzuzeigen. Drittens analysieren wir Rezeption und Aneignung sprachentastischer Ergebnisse, die mitunter – etwa im Baskenland, in Belgien oder in Kanada – hochpolitische Debatten auslösen und sich sowohl auf wissenschaftliche als auch politische Argumente stützen. Schliesslich führen wir die drei Aspekte – Geschichte, Methoden und Rezeption von Resultaten – zusammen, um Herausforderungen und Grenzen amtlicher Sprachenstatistiken herauszuarbeiten.

2 Sprachenerhebungen: ein historischer Überblick

Seit dem 19. Jahrhundert sind Sprachenstatistiken Teil von politischen Projekten, die für ihre Zählungen Ansätze aus Wissenschaft und Politik verbinden. Das Vorgehen zur Quantifizierung wird dabei je nach Zielen der Umfragen und Erhebungen angepasst. Ausnahmslos dienen Zahlen zu Sprachen dabei der Konstruktion von Nationalstaaten, der Definition einer nationalen kollektiven Identität und der Verteilung von Ressourcen.

Erste erschöpfende statistische Erhebungen zu Sprachen wurden Mitte des 19. Jahrhunderts in Europa realisiert. Zuvor hatte Frankreich bereits Schätzungen zur Präsenz von Dialekten und Patois auf seinem Territorium vorgenommen, um zu ermitteln, in welchen Regionen das Französische noch nicht Hauptverkehrssprache war. Ziel war es einerseits, umfassendes Wissen zu den Sprachpraktiken in Frankreich aufzubauen und andererseits diejenigen Regionen zu bestimmen, in denen Dialekte und Patois mithilfe der Schule oder von Verwaltungsstrukturen ausgemerzt werden sollten.

Am *Internationalen statistischen Kongress* (1853–1876) debattierten insbesondere die österreichischen und die preussischen Statistiker über zwei Methoden zur Sprachenerhebung, nämlich diejenige zur Ermittlung der „Muttersprache“ vs. die zur „gesprochenen Sprache“. Die Vertreter Preussens vertraten die Ansicht, dass die

„Muttersprache“ am objektivsten Auskunft über die Nationalität der Befragten gebe; die Österreicher hielten dagegen, dass Sprache und Nationalität nicht verwechselt werden dürften, da sie nicht zwingend zusammenfielen. Einige Probleme, die uns heute noch beschäftigen, wurden bereits während der Sitzungsperioden des Kongresses diskutiert, insbesondere die Abgrenzung zwischen Sprachen, Dialekten und Patois sowie das Vorgehen zur Erhebung der Sprachen von zweisprachigen Personen.

Schlussendlich wird Österreich die „Umgangssprache“ erheben und Preussen die „Muttersprache“. Die österreichische Regierung zieht dadurch den Zorn zahlreicher slawischer Minderheiten auf sich, die anprangern, man versuche sie aus den Ergebnissen auszublenden: Der Gebrauch des Deutschen als hauptsächlich verwendeter Sprache auf formeller Ebene werde so hervorgehoben und die Präsenz anderer im Kaiserreich präsenster Sprachen negiert. In Preussen beabsichtigen die Statistiker zwar, anhand der Resultate zur „Muttersprache“ die Vormachtstellung der deutschen Sprache auf dem gesamten Staatsgebiet zu belegen, einige slawische Minderheiten im Osten des Landes erheben allerdings Forderungen auf Basis derselben Zahlen. In beiden Fällen postulieren die Statistiker die Vision einer „objektiven“

Zählung der Sprachen, die in beiden Fällen zum politischen Zankapfel wird.

Im 20. Jahrhundert enthalten zahlreiche Erhebungen eine oder mehrere Fragen zu Sprachen. Die Sprache von Individuen wird nicht nur als sprachliche Grösse interpretiert, sondern auch stellvertretend für Nationalität, Ethnie oder sogar Rasse. In dieser Hinsicht wird den Sprachen in Erhebungen in den Kolonien besondere Beachtung geschenkt, wie etwa das britische Vorgehen in Indien zeigt, wo die Sprachenerhebung eine anthropologische Unternehmung riesigen Ausmasses ist und ihre Resultate ein universales Referenzwerk sowohl für Wissenschaftler als auch die (post-)koloniale Verwaltung darstellen.

In der Sowjetunion werden mittels Volkszählungen ebenfalls anthropologische Untersuchungen in grossem Stil unternommen. Auch wenn Sprachenstatistiken in erster Linie dazu dienen, die ethnische Zugehörigkeit zu klären, werden sie gleichzeitig herangezogen, um die Zunahme des Russischen als inter-ethnische Kommunikationssprache zu belegen. Nach Auflösung der UdSSR kehren die meisten post-sowjetischen Erhebungen diese Logik um und wollen in der Regel die Präsenz des Russischen herunterspielen und aufzeigen, dass die einstmals zweitrangige nun zur dominierenden Sprache in diesen sich im Übergang befindenden Ländern geworden ist. In den USA sollen Sprachenerhebungen ebenfalls mehr Informationen zu Ethnie, Rasse und Sprache liefern. Ziel ist zwar die Identifizierung von nicht englischsprachigen Gruppen von Individuen, um Gebiete zu bestimmen, in denen Wahlunterlagen in anderen Sprachen als Englisch zur Verfügung gestellt werden

müssen. Die amerikanischen Statistiken leisten durch das Konzept der „sprachlich isolierten Haushalte“ aber auch Diskriminierungen Vorschub.

Während Regierungen Sprachenstatistiken also als Mittel zur Durchsetzung von Sprachpolitiken betrachten können, sehen Sprachminderheiten sie als Mittel zur Verteidigung ihrer Interessen. Nach der Unabhängigkeit vom Vereinigten Königreich, für das die Frage nach Kenntnis der irischen Sprache dazu diente, den Fortschritt des Englischen in den von ihm verwalteten Gebieten zu messen, ging es für Irland darum festzustellen, welche Regionen Unterstützung bei der Revitalisierung des Irischen bedurften. Ähnlich in Lateinamerika, etwa in Mexiko, wo es ebenfalls zuallererst darum ging, anhand der Sprachenstatistik die Vormachtstellung des Spanischen zu bekräftigen, und erst in zweiter Linie darum festzulegen, welche der sogenannten „indigenen“ Minderheiten die Unterstützung der Regierung benötigten.

Seit den 1960er Jahren erfreut sich die zahlenmässige Erfassung von Sprachminderheiten in Wissenschaft und Politik ausserordentlicher Popularität. Wissenschaftliche Disziplinen wie die Geolinguistik und die Demolinguistik (resp. Sprachgeografie und -demografie), die Sprachenstatistiken verwenden, bilden sich im akademischen Bereich heraus. Die historische Kontextualisierung ihrer Entstehung, vor allem der Demolinguistik in Kanada, trägt zum Verständnis der politischen Ideologie bei, die die Herausbildung dieser spezialisierten wissenschaftlichen Disziplinen begünstigte.

3

Wie Sprachen erhoben werden

Amtliche Sprachenstatistiken hängen von zahlreichen methodischen Kriterien ab. Ein Vergleich der in mehreren grossen Nachschlagewerken genannten Zahlen macht deutlich, dass die Anzahl der genannten Sprachen und ihrer Sprecherinnen und Sprecher enorm variieren kann. Diese Unterschiede ergeben sich vor allem aus der Art und Weise, wie Sprachen dokumentiert und quantifiziert werden. Je nachdem werden bestimmte Sprachvarietäten in einer Kategorie zusammengefasst, während sie anderswo unter mehreren Bezeichnungen resp. gar nicht genannt werden. Solche Abweichungen erklären sich aus den Prozessen zur Dokumentation von Sprachen, wobei wissenschaftliche Aspekte und Anliegen der Verwaltung ineinandergreifen. D.h., ob eine Sprache in einer Erhebung als solche angesehen wird oder nicht, wird durch Kategorisierungsprozesse der Sprachen und ihrer Sprecherinnen und Sprecher bestimmt, welche dem statistischen Vorhaben vor- resp. nachgelagert sind. Siehe bspw. die umfangreichen Arbeiten des Summer Institute of Linguistics (SIL), das alle existierenden Sprachen dokumentieren möchte und dessen ursprüngliches Ziel es war, zwecks Missionierung die Bibel in möglichst viele Sprachen zu übersetzen.

Mit Blick auf amtliche statistische Erhebungen und Umfragen gilt es, weitere methodologische Unterschiede bei der Sammlung von Sprachendaten zu berücksichtigen,

um die Grenzen dessen, was mithilfe von Zahlenmaterial analysiert werden kann, zu verstehen. Dies betrifft etwa Umfang und Merkmale von untersuchten Stichproben, denn eine Stichprobenerhebung, insbesondere, wenn sie bestimmte Randgruppen ausschliesst, erlaubt nicht dieselben Aussagen wie eine Vollerhebung der Bevölkerung. Zudem werden nicht alle statistischen Erhebungen mit der gleichen Regelmässigkeit durchgeführt, da bspw. manche Volkszählungen alle fünf Jahre, andere hingegen nur alle zehn Jahre erfolgen. Ob die Fragen schriftlich oder mündlich gestellt werden, führt ebenso wie die Fragetypen (offen vs. geschlossen) zu Unterschieden. Das Alter, ab dem eine Person teilnehmen darf und die Sprache(n), die im Rahmen der Datensammlung verwendet wird/werden, sind Teil der zahlreichen Kriterien, die sich auf die Bedeutung der Resultate auswirken. In der wissenschaftlichen Literatur wird darauf hingewiesen, dass die individuelle Mehrsprachigkeit nicht systematisch erfasst wird. Selbst wenn es die Antworten erlauben, mehr als eine Sprache anzugeben, wird die Mehrsprachigkeit nicht überall gleich quantifiziert.

Welche Sprache(n) des individuellen Repertoires in den Resultaten erscheint/erscheinen, wird grösstenteils dadurch bestimmt, wie die Fragen zu den Sprachen formuliert sind. In der wissenschaftlichen

Literatur wird zwischen Fragen unterschieden, die auf folgende Aspekte abzielen: die Herkunft der Befragten, die Sprachpraktiken zum Zeitpunkt der Erhebung (manchmal aufgeschlüsselt nach Verwendungsbereichen), sprachliche Kenntnisse oder Kompetenzen der Befragten oder Einstellungen und Meinungen zu bestimmten sprachlichen Varietäten. Gewisse statistische Erhebungen stellen eine einzige Frage zur Sprache, andere hingegen mehrere. Es gilt auch zu beachten, dass manch amtliche Sprachenstatistik auf Grundlage von nicht-sprachlichen Daten wie Nationalität, Ethnie oder Rasse erstellt wird oder es sich sogar um Schätzungen auf Grundlage der Anzahl gemeldeter Einwohner einer Region handeln kann. Die in den Fragen zur Sprache verwendete Begrifflichkeit – „Muttersprache“, „Erstsprache“ oder „zu Hause gesprochene Sprache“ – folgt wechselnden Definitionen. So kann die Erhebung der „Muttersprache“ verweisen auf: die in der Kindheit als erste und zum Zeitpunkt der Erhebung immer noch verstandene Sprache der/des Befragten (Kanada), die von der Mutter mit der/dem Befragten gesprochene Sprache (Indien, Tschechische Republik, Slowakei), die zu Hause verwendete Sprache (USA) oder die Sprache, in der die/der Befragte denkt und die er als am besten beherrschte angibt (Schweiz).

4 Sprachenstatistiken: Rezeption und Aneignung

Amtliche Sprachenstatistiken werden regelmässig sowohl in der Wissenschaft als auch in der Politik verwendet, wobei keine klare Abgrenzung zwischen diesen beiden Bereichen möglich ist. Ganz im Gegenteil findet in vielen Fällen ein Dialog zwischen beiden statt. Die wissenschaftliche Literatur ist voll von Fällen, in denen Sprachenstatistiken für die politische Anerkennung innerhalb eines Gemeinwesens eine Rolle spielen, namentlich in Debatten zur sprachlichen Territorialität und/oder zu sprachlichen Rechten (Zugang zu öffentlichen Dienstleistungen, Übersetzungen, Unterstützung von Minderheitensprachen als Schulsprachen usw.). Politische Auseinandersetzungen betreffen häufig Diskussionen zu Schwellenwerten oder Quoten, um festzulegen, ab welchem Prozentsatz oder ab welchen absoluten Zahlen Sprachminderheiten Rechte oder bestimmte Subventionen erhalten. Die Festlegung von Schwellenwerten oder Quoten führt in der Regel zu hoch politisierten Diskussionen. Im Übrigen existieren keine allgemein anerkannten Vorgaben zur Bestimmungen von Schwellenwerten; jeder Staat und jede Region geht anders vor. Sprachenstatistiken werden ebenfalls herangezogen, um Bevölkerungsgruppen zu identifizieren. Freilich kann ihr Fehlen auch die Absicht einer Regierung verraten, bestimmten Minderheiten keine

Sichtbarkeit zu verschaffen. Dies ist in Frankreich und der Türkei der Fall, deren Statistiken keine einzige Angabe zu Sprachen enthalten, damit ihre Minderheiten keine Stimme erhalten und um über die gemeinsame Sprache das nationale Zusammengehörigkeitsgefühl zu stärken. Im Fall der USA und Australiens sollte ursprünglich der Anteil der Anglofonen ermittelt werden, um festzustellen, wo mit Blick auf amtliche Kontakte sprachliche Förderung nötig wäre. Allerdings führte dies bei der Interpretation der Sprachenstatistiken in beiden Fällen auch zu Diskriminierungen.

In der Wissenschaft werden amtliche Sprachenstatistiken je nach Forschenden sehr unterschiedlich aufgenommen resp. verwendet. Da die Mehrzahl der statistischen Erhebungen nur die Angabe einer einzigen Sprache zulässt, wird meist der Mangel an Informationen zur individuellen Zwei- oder Mehrsprachigkeit moniert. Solche Lücken haben dazu geführt, dass viele Forschende diese zu beheben und die Validität und Genauigkeit der statistischen Ergebnisse zu verifizieren suchen, indem sie sie mit anderen statistischen Daten wie etwa Schulregistern abgleichen.

Die Forschenden weisen auf weitere Unzulänglichkeiten der von ihnen konsultierten Statistiken hin, etwa, dass die Erhebung der zu Hause gesprochenen Sprachen

nicht repräsentativ für alle Sprachpraktiken einer Person ist. Vor allem Personen mit Migrationshintergrund können ausser Haus durchaus weitere Sprachen verwenden. Statistiken zu Sprachkenntnissen und -kompetenzen der Befragten sind ebenfalls mit Vorsicht zu behandeln, da es den Aussagen der Befragten ohne präzise Deskriptoren, was mit Kenntnis oder Kompetenz in den erhobenen Sprachen genau gemeint ist, wohl an Konsistenz und Zuverlässigkeit mangelt.

Amtliche Sprachenstatistiken werden vielfach für Studien zu Sprachwechsel und Spracherhalt herangezogen, d.h. zu Untersuchungen dazu, ob Individuen dieselbe Sprache ihr Leben lang verwenden oder zu einer anderen Sprache wechseln resp. weitere erwerben. Derartige Analysen, die die makro-soziolinguistische Entwicklung von Sprechergruppen beobachten, unterscheiden mitunter zwischen Verwendungsbereichen und setzen die Daten zu anderen Variablen wie Geburts- oder Sterberaten in Beziehung. Sie können politische Auswirkungen haben, wie etwa in Kanada, wo einige Forschende wissenschaftliche Argumente in politische Forderungen umwandeln.

Resultate wurden auch als Indizien für Haltungen zu Sprachen und ihren Sprechern interpretiert. In post-jugoslawischen Erhebungen wurde etwa die Antwort „serbokroatisch“, „serbisch“ oder „kroatisch“ auf die Frage nach der Muttersprache als Entwicklung hin zur Aneignung von sprachlichen Bezeichnungen in einer bestimmten soziopolitischen Situation wahrgenommen. Auf internationaler Ebene sind sich die Forschenden hinsichtlich der Zählung von gefährdeten Sprachen, d.h. immer weniger verwendeten Sprachen, uneinig. Während

die einen das Verschwinden von Sprachen mit einer Bedrohung der menschlichen Biodiversität vergleichen, befinden die anderen, dass solche Berechnungen an für die betroffenen Bevölkerungsgruppen wichtigeren Anliegen vorbeizielten.

5 Fazit

Quer durch alle Sprachenerhebungen und -umfragen, mit denen wir uns in unserer Literaturübersicht beschäftigt haben, zieht sich als roter Faden die Vermischung von wissenschaftlichen und politischen Fragen. Einige Staaten haben ihr statistisches Instrumentarium stark ausgebaut und stützen sich sogar auf mehrere Umfragetypen. Die Geschichte der Sprachenstatistiken in Kanada zeigt in dieser Hinsicht auf, in welchem Ausmass die Methodologien unter dem Einfluss von wissenschaftlichen und politischen Forderungen stetig verfeinert wurden: Im Zuge dauernder Kritik an der Genauigkeit von sprachbezogenen Daten in den kanadischen Erhebungen wurden nach und nach Fragen hinzugefügt, um wissenschaftlich gestützte Anforderungen zu erfüllen. In kleinerem Umfang erlebten die baskischen Statistiken einen Aufschwung mittels amtlicher soziolinguistischer Erhebungen, deren Methoden sich sukzessive von denjenigen traditioneller Umfragen entfernen, um den Gebrauch des Baskischen im öffentlichen Raum des Baskenlandes quantitativ zu erfassen.

Trotz offensichtlicher Unterschiede, wie (mehr oder weniger) sprachliche Merkmale erfasst werden, verfolgen die meisten Erhebungen und Umfragen ein recht ähnliches Ziel, nämlich die Quantifizierung von Sprachen als Ausdruck gesellschaftlicher Umstände, wobei diese Quantifizierung Auswirkungen auf die politische Organisation und

auf über rein sprachliche Aspekte hinausgehende Forderungen hat. Zahlen können zum bevorzugten Thema von Minderheiten werden, die nach Anerkennung in einer Region streben, nach Zugang zu bestimmten Ressourcen und Rechten wie die Möglichkeit der Schulbildung in ihrer Sprache, oder die Amtssprache(n) auf regionaler und/oder nationaler Ebene zu bestimmen. Je nach geltender Sprachenordnung können Sprachenstatistiken in mehrsprachigen Kontexten mehr oder weniger direkten Einfluss auf Mittelzuteilungen haben, die nicht nur sprachliche Aspekte betreffen, etwa wenn Minderheiten ab einem Schwellenwert bestimmte Rechte zugestanden werden.

Je nach Auswirkung der Statistiken auf politische Entscheide kann selbst die Form der Sprachdatenerhebung und der Festlegung von Schwellenwerten zu heftigen Auseinandersetzungen führen. In Belgien wurden zum Beispiel die Amtssprachen der Gemeinden jeweils aufgrund der Resultate von Erhebungen festgelegt. Die Veröffentlichung der Sprachenstatistiken führte zu derartigen Spannungen zwischen Wallonen und Flamen, dass die Behörden beschlossen, die Frage zur Sprache ganz aus den Formularen zu streichen und die Sprachgrenzen ein für alle Mal festzulegen. In Ländern wie der Schweiz hingegen spielen die amtlichen Statistiken keine grössere Rolle; es gilt das sprachliche Territorialitätsprinzip und die Einwohner eines Gebiets sind gehalten,

sich an die Amtssprache(n) ihrer Wohngemeinde oder ihres Kantons anzupassen. Einzig der Kanton Graubünden verwendet statistische Angaben bei der Umsetzung einer Sprachenpolitik, die sich auf Schwellenwerte stützt. Dieses dynamisch interpretierte Territorialitätsprinzip führt dort auch heute noch immer wieder zu politischen Diskussionen.

Etliche Wissenschaftler haben sich der Überprüfung der Validität von statistischen Sprachdaten gewidmet oder zusätzliche, vertiefende Umfragetypen entwickelt, um die Zahlenangaben von Erhebungen zu differenzieren. Daneben gibt es weitere Anwendungsbereiche: Man denke etwa an demografische Studien zu Sprachwechsel und Spracherhalt, die auch unter anderen Vorzeichen existieren, etwa als Versuche zur Messung der Vitalität von Sprachen. In internationalen Vergleichen führen Studien zur Vitalität oder zum Sprachwechsel regelmässig zu epistemologischen Auseinandersetzungen, welche sich aus methodologischen und ideologischen Meinungsverschiedenheiten ergeben.

Durch Sprachenstatistiken ausgelöste Auseinandersetzungen beschränken sich nicht auf die Wissenschaft oder die politische Elite. Diese Statistiken haben, sobald die Mehrsprachigkeit politisch ein Thema ist, reale Auswirkungen auf die Verteilung von Ressourcen. Die Zurverfügungstellung von Mitteln für Übersetzungen und amtliche Dienstleistungen, welche Ansprüche von sprachlichen Minderheiten berücksichtigen, ist keine Selbstverständlichkeit: In gewissen Fällen dienen die amtlichen Sprachenstatistiken dazu, in Streitfällen zu entscheiden. Auch wenn Statistiken als ob-

jektive Lösung bei politischen Spannungen angepriesen werden, verdeutlicht unsere Literaturübersicht, dass die Wissenschaft in diesen Fällen zutiefst durch politische Fragestellungen beeinflusst ist.

Contare le lingue: storia, metodi e politiche dei censimenti della popolazione

Sintesi della rassegna della letteratura

—
Philippe Humbert, Renata Coray, Alexandre Duchêne

1

Introduzione

In molte società plurilingui le statistiche ufficiali sulle lingue hanno un ruolo determinante nell'organizzazione degli spazi linguistici. Infatti, in alcuni paesi, le cifre sul numero di monolingui o bilingui hanno delle conseguenze politico-economiche dirette, per esempio nell'attribuzione di sovvenzioni alle minoranze linguistiche. Le statistiche linguistiche sono spesso consultate – e a volte contestate – da numerosi studiosi che se ne servono per analizzare dei fenomeni demografici collegati alle lingue. Data l'onnipresenza delle statistiche sulle lingue nel mondo scientifico e politico, il Centro scientifico di competenza per il plurilinguismo si è proposto di elaborare una rassegna della letteratura scientifica sui censimenti linguistici e la loro interpretazione nello spazio sociale.

La presente sintesi della rassegna della letteratura non contiene alcuna informazione bibliografica, perché non abbiamo voluto ridurre l'estesa letteratura ad una porzione che sarebbe stata poco rappresentativa. La bibliografia completa figura nella versione integrale (in francese, disponibile in linea sul sito: <https://goo.gl/jeNZWo>). La rassegna della letteratura, della quale riassumiamo qui il contenuto, è di utilità per i ricercatori che lavorano con dati estratti dai censimenti: offre una migliore comprensione delle implicazioni socio-politiche legate alla creazione ed all'uso delle statistiche linguistiche. Descrivendo lo stato attuale della ricerca in diversi am-

biti, fornisce un punto di paragone internazionale necessario per comprendere meglio la posizione elvetica in materia di censimenti linguistici. Conviene precisare che sono state prese in considerazione solo le statistiche elaborate da autorità ufficiali regionali o nazionali. In generale, si tratta di censimenti della popolazione nei quali la lingua figura tra altre categorie demografiche come l'età, il sesso o la nazionalità.

La rassegna comprende ricerche svolte in diverse discipline: sociolinguistica, sociologia, politologia, demografia e geografia sono le materie tra le quali i censimenti e le statistiche linguistiche sono state trattate con maggiore frequenza. Partendo dalle opere e dagli autori più citati, abbiamo consultato sistematicamente le bibliografie proposte e ci siamo sforzati di offrire ai lettori ed alle lettrici una panoramica più ampia ed internazionale possibile. Infatti, tutti gli autori consultati hanno studiato i dati linguistici dei censimenti della popolazione svolti a diversi livelli – nazionali o regionali – in diverse parti del mondo. A questa prima immersione globale in materia si aggiunge la consultazione completa di una trentina di riviste di sociolinguistica, politologia e demografia, in francese, inglese, tedesco, e sporadicamente in russo, spagnolo, italiano, romancio e serbo-croato. Sfruttando una serie di parole-chiave in diverse lingue, abbiamo ulteriormente approfondito la raccolta su dei motori di ricerca

specializzati. In molti casi, la lingua non compariva come l'argomento principale della ricerca, ma come una componente definita "etnica", "indigena" o "razziale". Di conseguenza, anche se la lingua occupava uno spazio centrale nell'elaborazione di statistiche etniche e nell'implementazione di progetti politici, gli autori la consideravano piuttosto una caratteristica non linguistica. Questi criteri di selezione danno un'idea dell'estensione della ricerca che abbiamo svolto come pure dei suoi limiti.

Alcuni paesi e regioni appaiono più spesso di altre nella rassegna. Ciò è dovuto non solo alle lingue che abbiamo usato nella nostra ricerca, ma anche al ruolo dai censimenti linguistici nella società. Infatti non tutti i paesi conducono censimenti linguistici con la stessa assiduità della Svizzera o del Canada. In alcuni stati, come la Francia o la Cina, pur caratterizzati dalla presenza di numerose varietà linguistiche, i censimenti della popolazione non contengono domande sulle lingue. Tuttavia vi si possono trovare dei dati statistici sulle lingue raccolti sulla base di stime approssimative o di inchieste a campione. Come abbiamo potuto constatare, la presenza o assenza di una domanda sulla lingua o sulle lingue degli intervistati è spesso giustificata da motivi politici, e questo a seconda della situazione, le statistiche linguistiche o la loro assenza possono divenire fonte di conflitto.

In primo luogo, la rassegna della letteratura e la presente sintesi trattano la questione nella sua dimensione storica, per evidenziare le funzioni iniziali e l'evoluzione del censimento delle lingue nei diversi contesti. In secondo luogo, vengono trattati gli aspetti metodologici della produzione delle

statistiche linguistiche, per mostrare la gamma di modi in cui sono categorizzate e circoscritte le varietà linguistiche dei parlanti. In terzo luogo, vengono analizzate la ricezione e l'appropriazione dei risultati statistici sulle lingue, i quali generano dei dibattiti altamente politicizzati, come nei Paesi Baschi, in Belgio o in Canada, e riposano su argomenti scientifici e politici. Infine, vengono sintetizzati i tre aspetti dei censimenti delle lingue, cioè la storia, il metodo, e la ricezione dei risultati, per mostrarne le principali implicazioni ed i limiti delle statistiche linguistiche ufficiali.

2 I censimenti delle lingue nella storia

Sin dal XIX secolo le statistiche linguistiche s'iscrivono in progetti di governance che uniscono scienza e politica nei processi di inventario. I metodi di quantificazione cambiano in funzione degli obiettivi delle inchieste e dei censimenti. In ogni caso, le cifre sulle lingue contribuiscono alla costruzione degli stati-nazione, alla definizione dell'identità nazionale collettiva ed alla distribuzione delle risorse.

I primi rilevamenti statistici esaustivi sulle lingue sono stati svolti intorno alla metà del XIX secolo in Europa. In precedenza, la Francia aveva già effettuato delle stime sulla presenza di dialetti e patois sul suo territorio, al fine di identificare le regioni nelle quali il francese non era ancora usato in maggioranza. L'obiettivo era da una parte di costituire un sapere enciclopedico sulle pratiche linguistiche in Francia e dall'altro di designare le regioni dove i dialetti o i patois dovevano essere sradicati mediante le strutture scolastiche ed amministrative.

Durante le sedute del *Congresso internazionale di statistica* (1853-1876), sono stati discussi due metodi di censimento delle lingue: quello orientato a sondare "la lingua materna" degli intervistati e quello interessato alla "lingua parlata". Queste due concezioni sono dibattute principalmente tra gli statistici austriaci e prussiani. I prussiani sostengono che la "lingua materna" costituisce il mezzo più obiettivo di rilevare la nazionalità degli intervistati, mentre gli austriaci sono dell'avviso che

lingua e nazionalità non devono essere confuse, poiché non sempre coincidono. Durante le sedute del Congresso vengono discussi diversi problemi tuttora dibattuti. Le principali preoccupazioni concernono il grado di distinzione linguistica (lingue, dialetti, patois) ed il modo di rilevare le lingue degli individui bilingui.

Concretamente, l'Austria censirà la "Umgangssprache" (lingua parlata comunemente) e la Prussia la "Muttersprache" (lingua materna). Rilevando la lingua parlata comunemente, il governo austriaco si aggiudica la collera di numerose minoranze slave che denunciano un tentativo di occultarle nei risultati: questa domanda contribuisce ad enfatizzare l'uso del tedesco –lingua usata maggiormente nelle interazioni formali– e a cancellare la presenza di altre lingue parlate nell'Impero. Per quanto riguarda la Prussia, invece, se gli statistici si prefiggono di dimostrare l'egemonia della lingua tedesca sul territorio con i risultati del rilevamento della "Muttersprache", alcune minoranze slave localizzate all'est del paese, esprimono delle rivendicazioni sulla base delle stesse cifre. In entrambi i casi gli statistici avanzano una visione "obiettiva" del conteggio delle lingue; in entrambi i casi le cifre diventano una fonte di conflitto politico.

Durante il XX secolo in numerosi censimenti vengono poste delle domande sulle lingue. La lingua degli individui non è interpretata solo come un dato linguistico, ma come un sostituto della nazionalità,

dell'etnia, o della razza. In questo senso, le lingue sono oggetto di attenzione particolare nei censimenti coloniali. Lo testimonia il caso dell'India, dove i rilevamenti linguistici svolti dai britannici costituiscono un'impresa antropologica colossale, i cui risultati costituiscono un riferimento enciclopedico sia per gli studiosi che per l'amministrazione (post-) coloniale.

Anche l'URSS svolge delle ricerche antropologiche di ampia portata mediante censimenti della popolazione. Le statistiche linguistiche servono principalmente a chiarire l'appartenenza ai gruppi etnici di numerosi individui, ma sono anche usate per dimostrare la progressione del russo come lingua di comunicazione interetnica. Dopo la dissoluzione dell'URSS, la maggior parte dei censimenti post-sovietici inverte l'ordine delle priorità, con l'obiettivo generale di minimizzare la presenza del russo e dimostrare che la lingua una volta secondaria è ora divenuta la lingua dominante nel paese in fase di transizione. Anche negli Stati Uniti i censimenti linguistici cercano di ottenere maggiori informazioni di tipo etnico, razziale e linguistico. Se uno degli obiettivi è identificare gruppi non anglofoni per determinare le zone dove fornire materiale elettorale in lingue diverse dall'inglese, le statistiche statunitensi mostrano anche derive discriminatorie mediante la creazione del concetto di "nucleo familiare isolato linguisticamente" ("linguistically isolated households").

Laddove i governi possono vedere le statistiche linguistiche come un mezzo per imporre delle politiche linguistiche, le minoranze linguistiche vi vedono un mezzo per difendere i loro interessi. Raggiunta

l'indipendenza dal Regno Unito, l'Irlanda si appropria della domanda sulla conoscenza dell'irlandese, per elaborare delle statistiche volte a determinare le regioni che il governo deve sostenere per rivitalizzare l'irlandese, mentre i britannici usavano la stessa domanda per misurare la progressione dell'inglese sui territori da loro amministrati. L'America latina fornisce altri esempi di questo tipo, come il Messico, dove le statistiche linguistiche sono prima usate per imporre l'egemonia della lingua spagnola e poi per definire le minoranze dette "indigene" che hanno bisogno del sostegno del governo.

A partire dagli anni 1960 il conteggio delle minoranze linguistiche conosce un impegno internazionale notevole sia nel campo linguistico che politico. Discipline scientifiche specifiche, che usano le statistiche linguistiche, si fanno spazio nell'ambito accademico. Si tratta principalmente della geolinguistica e della demolinguistica (o della geografia e della demografia delle lingue). La contestualizzazione storica della loro emergenza, più particolarmente della demolinguistica in Canada, permette di comprendere l'ideologia politica alla base dello sviluppo di tali discipline scientifiche specifiche.

3

I modi di censire le lingue

Le statistiche linguistiche ufficiali sono svolte secondo diversi criteri metodologici. Il confronto delle cifre menzionate in diverse fonti enciclopediche mostra che il numero di lingue e di parlanti indicati può variare in maniera eccessiva. Queste differenze sono in gran parte dovute al modo in cui le lingue sono documentate e conteggiate. Infatti, a seconda delle fonti, alcune varietà linguistiche sono raggruppate sotto una sola categoria, mentre altre possono comparire sotto diversi nomi oppure non comparire affatto. Tali divergenze provengono in parte dalla natura dei processi di documentazione linguistica, che iscrivono l'opera scientifica in progetti di governance politica. Considerare una lingua come tale in un censimento dipende anche da un processo di categorizzazione delle lingue e dei parlanti, precedente o contemporaneo all'esercizio statistico. Lo testimonia l'ampiezza dei lavori svolti dal Summer Institute of Linguistics (SIL) che prova a documentare tutte le lingue del mondo. Il suo obiettivo iniziale è di tradurre la bibbia nel massimo numero di lingue possibile con fini evangelici.

Per quanto riguarda i censimenti e le statistiche ufficiali, vi sono anche altri aspetti metodologici, che differiscono nella raccolta dei dati sulle lingue, i quali devono essere considerati, per comprendere i limiti di ciò che può essere analizzato mediante le cifre. Si tratta tra l'altro della dimensione e delle caratteristiche del campione stu-

diato. Infatti, un'inchiesta a campione non permette di trarre le stesse conclusioni che un censimento esaustivo della popolazione, a maggior ragione se il campione non comprende alcune fasce della popolazione. Inoltre, le inchieste statistiche non sono tutte effettuate con la stessa frequenza (per esempio, alcuni censimenti sono svolti ogni cinque anni, mentre altri, ogni dieci anni). Il fatto di porre le domande in forma scritta o orale, come pure il tipo di domande (risposta aperta o chiusa) inducono ugualmente delle differenze. L'età a partire dalla quale un individuo può rispondere a queste domande e la lingua o le lingue usate per raccogliere i dati fanno parte della miriade di criteri che hanno un impatto sulla portata interpretativa dei risultati. Nella letteratura scientifica, emerge che il plurilinguismo individuale non è preso in considerazione sistematicamente. Anche quando la domanda permette di rispondere indicando più lingue, il plurilinguismo non è quantificato sempre allo stesso modo.

La formulazione delle domande sulle lingue determina in gran parte quale lingua o quali lingue del repertorio individuale apparirà o appariranno nei risultati. La letteratura scientifica distingue le domande che sono orientate a: le origini dell'intervistato; le pratiche linguistiche dell'intervistato al momento del rilevamento statistico (distinguendo a volte i domini d'uso specifici); le conoscenze o competenze linguistiche dell'intervistato;

gli atteggiamenti e opinioni dell'intervistato rispetto ad alcune varietà linguistiche. Alcune rilevazioni statistiche contano una sola domanda sulle lingue, mentre altre ne contano diverse. Da notare che alcune statistiche linguistiche ufficiali sono elaborate sulla base di dati non linguistici come la nazionalità, l'etnia, o la razza, o si può trattare di stime calcolate sul numero di abitanti residenti per ogni regione. La terminologia usata per riferirsi alle domande sulle lingue –“lingua materna”, “lingua nativa”, o “lingua parlata a casa”– è fluttuante. Per esempio, il rilevamento della “lingua materna” può riferirsi a: la prima lingua appresa durante l'infanzia e sempre compresa dall'intervistato al momento del censimento (Canada); la lingua parlata dalla madre dell'intervistato (India, Repubblica Ceca, e Slovacchia); la lingua usata a casa (USA); la lingua nella quale l'intervistato pensa e che considera di conoscere meglio (Svizzera).

4

Ricezione e appropriazione dei risultati dei censimenti linguistici

Le statistiche linguistiche ufficiali sono spesso usate negli ambienti scientifici e politici, benché non esista una frontiera netta che li separi. Al contrario, in molti casi, esiste un dialogo tra questi due ambiti. La letteratura scientifica abbonda di situazioni in cui le statistiche linguistiche implicano un riconoscimento politico in seno alla società, in particolare in dibattiti sulla territorialità linguistica e/o i diritti linguistici (accesso ai servizi ufficiali, traduzioni, sostegno delle lingue minoritarie come lingue di scolarizzazione, etc.). I dibattiti politici rimandano spesso a delle discussioni sulla costituzione di soglie o di quote per determinare la percentuale o il numero di parlanti a partire dal quale certe minoranze ottengono dei diritti o sovvenzioni specifiche. La costituzione di tali soglie o quote solleva generalmente dei dibattiti fortemente politicizzati. D'altronde non esistono delle convenzioni nell'elaborazione di tali soglie: ogni stato o regione prende le decisioni in modo diverso. Le statistiche sulle lingue sono sfruttate anche per identificare delle popolazioni. Al contrario, la loro assenza può rivelare la volontà di un governo di non dare visibilità ad alcune minoranze, come dimostrano le statistiche francesi o turche, che non rilevano alcun dato sulle lingue, per evitare di dare voce alle minoranze e rinfor-

zare così un sentimento di unità nazionale dietro una sola lingua. Nel caso degli Stati Uniti e dell'Australia, l'obiettivo iniziale era di conteggiare la parte non anglofona, per sapere dove fornire un sostegno linguistico nelle relazioni ufficiali. In questi due casi il processo interpretativo delle statistiche linguistiche è andato incontro a delle derive discriminatorie.

Nell'ambito scientifico, le statistiche ufficiali sulle lingue vengono accolte diversamente a seconda dei ricercatori che le consultano o sfruttano. La maggior parte delle statistiche linguistiche è criticata per la mancanza d'informazioni sul bilinguismo o il plurilinguismo individuale, poiché la maggior parte dei rilevamenti statistici permette di indicare solo una lingua. Le lacune dei censimenti hanno spinto numerosi ricercatori a supplire a tali mancanze ed a verificare la validità e la pertinenza dei risultati statistici, comparandoli con altri risultati statistici, come quelli dei registri delle scuole.

I ricercatori menzionano ancora altri limiti a seconda delle statistiche consultate. Per esempio, alcuni autori sottolineano che il rilevamento delle lingue parlate a casa non è rappresentativo di tutte le pratiche linguistiche di un individuo; è infatti possibile esprimersi in altre lingue al di fuori di casa, soprattutto per le persone con un passato

migratorio. Le statistiche sulle conoscenze o competenze linguistiche degli intervistati necessitano anch'esse di qualche precauzione, perché senza delle definizioni chiare di cosa s'intende per conoscenza o competenza nelle lingue sondate, le dichiarazioni degli intervistati rischiano di essere incoerenti ed inaffidabili.

Le statistiche linguistiche ufficiali sono usate in numerosi studi sulla sostituzione e la conservazione delle lingue, cioè sul fatto di continuare ad usare la stessa lingua tutta la vita o, al contrario, di passare ad un'altra lingua o ad acquisirne altre. Ciò al fine di osservare l'evoluzione di gruppi di parlanti a livello macro-sociolinguistico. Queste analisi distinguono spesso tra i domini d'uso e sono incrociate ad altre variabili come il tasso di fecondità o di mortalità. Esse possono avere ripercussioni politiche come in Canada, dove alcuni ricercatori trasformano i loro argomenti scientifici in rivendicazioni politiche.

I risultati sono anche stati interpretati come rivelatori di atteggiamenti verso le lingue ed i parlanti. Nel caso dei censimenti in ambito post-iugoslavo, per esempio, il fatto di rispondere "serbo-croato" o "serbo" o "croato" alle domande sulla lingua materna è percepito come un'evoluzione dell'appropriazione delle denominazioni linguistiche in relazione a delle situazioni socio-politiche specifiche. A livello internazionale la questione del conteggio delle lingue in pericolo, cioè delle lingue sempre meno usate, divide i ricercatori. Alcuni comparano la scomparsa delle lingue ad una minaccia alla biodiversità umana, altri considerano che tali conteggi trascurano delle preoccupazioni ben più importanti per le popolazioni coinvolte.

5 Conclusioni

Attraverso tutti i censimenti ed indagini sulle lingue che abbiamo trattato nella nostra rassegna della letteratura, la questione statistica combina delle implicazioni scientifiche e politiche. Alcuni stati hanno sviluppato ampiamente lo strumento statistico per declinarlo in diversi tipi d'inchiesta. In questo senso, la storia delle statistiche linguistiche come quelle elaborate in Canada dimostra come le metodologie si sono evolute sotto l'influsso di rivendicazioni che combinavano argomenti scientifici e politici: in seguito a delle costanti critiche sulla precisione dei dati sulle lingue nei censimenti canadesi, si sono aggiunte progressivamente delle domande per rispondere a delle richieste che riposavano su delle prove scientifiche. Su una scala minore, le statistiche del basco prendono slancio dalle indagini sociolinguistiche ufficiali, i cui metodi si allontanano progressivamente dalle inchieste tradizionali, per tentare di cogliere quantitativamente gli usi della lingua basca nelle strade dei territori bascofoni.

Nonostante un'evidente diversità nei modi di rilevare delle caratteristiche più o meno linguistiche, la maggior parte dei censimenti e delle indagini persegue un obiettivo simile: la quantificazione delle lingue come espressione di una caratteristica socialmente situata, che non è priva di conseguenze nell'organizzazione politica e nelle rivendicazioni che non si limitano a delle

implicazioni linguistiche. Le cifre possono divenire il cavallo di battaglia di minoranze che desiderano ottenere un qualche riconoscimento su un territorio, avere accesso a delle risorse e a dei diritti, come la possibilità di scolarizzare i propri figli in una lingua invece che in un'altra o di definire la lingua ufficiale o le lingue ufficiali dell'amministrazione a livello regionale o nazionale. A seconda del regime linguistico, le statistiche sulle lingue possono avere un impatto più o meno diretto sull'allocatione delle risorse –non solo di ordine linguistico– in dei contesti plurilingui, in particolare mediante la definizione di soglie minime a partire dalle quali vengono riconosciuti determinati diritti a una minoranza.

A seconda dell'impatto delle statistiche sull'organizzazione della vita politica, il modo stesso di raccogliere dei dati sulle lingue e di stabilire delle soglie minime può essere oggetto di aspri dibattiti. Per esempio in Belgio, i risultati dei censimenti avevano un impatto diretto sullo stabilimento delle lingue ufficiali dei comuni. La pubblicazione delle statistiche linguistiche ha suscitato tante tensioni tra valloni e fiamminghi, che le autorità hanno deciso di eliminare le domande sulla lingua dai formulari e di fissare le frontiere linguistiche una volta per tutte. In altri paesi, come la Svizzera, è il principio di territorialità linguistica che prevale, per cui le statistiche ufficiali non contano molto: gli abitanti di

un territorio devono adattarsi alla lingua ufficiale o alle lingue ufficiali del loro comune o cantone di residenza. Solo il cantone dei Grigioni usa i dati statistici per l'implementazione di politiche linguistiche elaborate su delle soglie, optando così per un principio di territorialità dinamico, che continua ancora oggi a stimolare i dibattiti politici.

Diversi studiosi hanno cercato dei mezzi per verificare la validità dei dati statistici sulle lingue o hanno sviluppato altri tipi d'inchieste approfondite, per contestualizzare le cifre dei censimenti. Sono state elaborate anche diverse applicazioni. La più evidente è quella degli studi della sostituzione linguistica e della conservazione delle lingue a scala demografica, che si possono trovare anche sotto altri nomi, come i tentativi di misurare la vitalità delle lingue. Nelle comparazioni internazionali, gli studi sulla vitalità o sulla sostituzione delle lingue suscitano numerosi dibattiti epistemologici che rimandano a disaccordi di ordine metodologico e ideologico.

I dibattiti scatenati dalle statistiche linguistiche trascendono i limiti dell'ambito scientifico o politico. Queste statistiche infatti hanno delle conseguenze reali sulla distribuzione delle risorse, laddove il plurilinguismo è una questione politica. Lo stanziamento di fondi per le traduzioni e per servizi amministrativi che prendano in considerazione le richieste delle minoranze non è sempre automatico e pacifico: in alcuni casi le statistiche linguistiche ufficiali servono a prendere delle decisioni quando tali questioni pongono dei problemi. Tuttavia, benché le statistiche vengano presentate come una soluzione obiettiva alle tensioni

politiche, la presente rassegna della letteratura dimostra che la scienza, quando utilizzata a questi fini, resta profondamente influenzata da questioni di governance.

Dumbrar linguas: istorgia, metodas e politica da dumbraziuns dal pievel

Ina sintesa da la revista da litteratura

—
Philippe Humbert, Renata Coray, Alexandre Duchêne

1

Introducziun

En bleras societads plurilinguas giogan las statisticas uffizialas davart las linguas ina rolla essenziala per organisar ils spazis linguistics. En tscherts pajais han las cifras davart il dumber da pledaders monolings u bilings schizunt consequenzas politic-economicas directas, surtut cura ch'i va per l'attribuziun da sustegns finanziels a las minoritads linguistics. Las statisticas linguistics èn er ina funtauna da savida consultada cuntinuadamain – e mintgatant messa en dumonda – da numerus scienziads che sa servan da quella per analizar fenomens demografics en connex cun las linguas. En vista a l'omnipreschientscha da las statisticas linguistics en il mund scientific e politic ha il Center scientific da cumpetenza per la plurilinguitad elavurà in „State of the Art“ che preschenta il stan da la litteratura scientifica davart la relevaziun da datas davart las linguas e lur interpretaziuns en la societad.

Questa sintesa da la litteratura na cuntegna nagina infurmaziun bibliografica; nus n'avain betg vuli reducir la totalitad dals referiments ad ina part pauc represchentativa da la litteratura. La bibliografia cumpletta figurescha en la versiun integrala da la revista da litteratura (en franzos, accessibla en l'internet sut <https://goo.gl/eUERLc>). La revista da litteratura, che nus preschentan qua en furma resumada, serva als perscrutaders che lavuran cun datas da relevaziuns: ella gida a chapir meglier las dumondas

sociopoliticas colliadas cun la creaziun e l'utilisaziun da statisticas linguistics. Cun descriver il stan da la perscrutaziun en divers contexts furnescha ella ina basa da cumparegliaziun internaziunala ch'è necessaria per chapir meglier la posiziun svizra en connex cun la relevaziun da datas davart las linguas. Igl è impurtant da precisar che nus avain resguardà mo las statisticas linguistics elavuradas d'ina autoritad uffiziala – sin nivel naziunal u regiunal. En general sa tracta da dumbraziuns dal pievel che resguardan la lingua en il rom da categorias demograficas sco la vegliadetgna, la schlattaina u la naziunalitad.

La revista da litteratura cumpiglia lavurs da perscrutaziun da differentas disciplinas. La sociolinguistica, la sociologia, la politologia, la demografia e la geografia èn ils roms che s'occupan il pli savens da relevaziuns da datas davart las linguas e las statisticas correspondentas. A basa da las ovras e dals auturs citads il pli savens avain nus consultà sistematicamain las bibliografias avant maun ed ans essan stentads d'offrir a las lecturas ed als lecturs ina survista la pli internaziunala pussaivel. Ils auturs consultads èn sa basads sin las datas linguistics da las dumbraziuns dal pievel realisadas sin differentes plauns – naziunals u regiunals – en diversas parts dal mund. Sper questa emprima immersiun globala en la materia avain nus consultà ina trentina da survistas da la sociolinguistica,

politologia e demografia en franzos, anglais, tudestg e sporadicamain en russ, spagnol, talian, rumantsch e serbocroat. Plinavant avain nus pavlà maschinas da tschertgar spezializadas cun ina seria da chavazzins-clav per approfondar la retschertga e generar dapli resultats. En plirs cas n'è la lingua betg stada il tema da studi principal, mabain cumpara sco ina cumponenta „etnica“, „indigena“ u schizunt „razziala“. Era sche la lingua ha giugà ina rolla centrala per elavurar statisticas etnicas e per realisar projects politics, han ils auturs damai considerà ella plitost sco ina caratteristica betg linguistica. Quests criteris da seleziun rendan quint da la vastadad da la retschertga che nus avain fatg, mussan però era ses cunfins.

Tscherts pajais e tschertas regiuns cumparan pli savens che auters en la revista da litteratura. Quai n'ha betg da far unicamain cun las linguas che nus avain duvrà per nossa retschertga, mabain era cun l'impurtanza che las relevaziuns da datas davart las linguas han en la societad. Tut ils pajais na rimnan betg datas davart lur linguas cun tanta conscienziusadad sco la Svizra u il Canada. En tscherts stadis cun numerusas varietads linguistics sco la Frantscha u la China na datti schizunt nagina dumonda davart las linguas en la dumbraziun dal pievel. Ins chatta là dentant statisticas linguistics fatgas a basa da stimaziuns u d'enquistas en furma da provas statisticas. Sco quai che nus avain pudì constatar vegn l'existenza u la mancanza d'ina dumonda davart la lingua u las linguas da las persunas interrogadas savens giustifitgada cun arguments politics. Pertge che la statistica linguistica – sco era sia mancanza – po daventar ina funtauna da conflicts.

En in emprim pass preschentan la revista da litteratura e questa sintesa il tema ord perspectiva istorica per scuvrir las funcziuns inizialas ed il svilup da la relevaziun da datas davart las linguas en divers contexts. En in segund pass vegnan dilucidads ils aspects metodologics da l'elavuraziun da las statisticas linguistics per preschentar l'entira paletta da las metodos da categorisar e da circumscriber las varietads linguistics dals pledaders. En in terz pass stattan la recepciun e l'utilisaziun dals resultats statistics davart las linguas en il center da las analisas che sveglian mintgant debattas fermamain politics – sco en il Pajais Basc, en Belgia u en il Canada – e che sa basan sin arguments scientifics e politics. Per finir resumain nus trais aspects da la relevaziun da datas davart las linguas – l'istorgia, la metoda e la recepciun dals resultats – per pudair eruir las sfidas principalas ed ils cunfins da las statisticas linguistics uffizialas.

2 Las relevaziuns da datas davart las linguas en il decurs da l'istorgia

Dapi il 19avel tschientaner cumparan las statisticas davart las linguas en projects guvernamentals che collian la scienza e la politica en ils process da dumbraziun. Las metodos da quantifitgar sa midan tut tenor las finamiras da las retschertgas e dumbraziuns. Las cifras davart linguas participeschan a la construcziun dals stadis naziunals, a la definiziun d'ina identitad naziunala collectiva ed a la distribuziun da las resursas.

Las emprimas relevaziuns cumplexivas davart las linguas èn vegnidas realisadas en la mesadad dal 19avel tschientaner en l'Europa. Avant aveva la Frantscha gia realisà stimaziuns davart la preschientscha da dialects e patuas sin ses territori cun la finamira d'eruir las regiuns, en las qualas il franzos na vegniva anc betg utilisà d'ina maiortad. La finamira era d'acquistar ina savida enciclopedica davart las praticas linguisticas en Frantscha e d'eruir las regiuns, nua ch'ils dialects ed ils patuas stuevan vegnir eliminads cun l'agid da estructuras scolasticas ed administrativas.

Durant las sessiuns dal *Congress internaziunal da la statistica* (1853-1876) han ins discutà duas metodos da relevar datas davart las linguas: quella che emprova da sondar „la lingua materna“ da las personas interrogadas e quella che sa fatschenta cun „la lingua discurredda“. Questas duas

concepziuns èn surtut contestadas tranter ils statistichers austriacs e prussians. Ils Prussians pretendan che la dumonda da la „lingua materna“ saja la moda la pli objectiva d'eruir la naziunalitad da las personas interrogadas; ils Austriacs èn da l'avis che la lingua e la naziunalitad na possian betg vegnir confundidas, perquai ch'ellas na sajan betg necessariamain identicas. Plirs problems, che vegnan discussads anc oz, èn vegnids debattads durant las sessiuns dal Congress. Ils quitads principals concernan ils stgalims da distincziuns linguisticas (linguas, dialects, patuas) e la moda d'eruir las linguas discurreddas da las personas bilinguas.

En l'Austria han las personas interrogadas la finala da responder davart la „Umgangssprache“ (lingua da mintgadi) ed en la Prussia davart la „Muttersprache“ (lingua materna). Cun far la dumonda davart la lingua discurredda en il mintgadi attira la regenza austriaca la critica da numerosas minoritads slavicas che considereschon quai sco emprova da las excluder dals resultats: questa dumonda suttastritga las praticas dal tudestg – lingua utilisada maioritarmain en ils rapports formals – e gida a zuppar la preschientscha dad autras linguas entaifer l'imperi. Ed en la Prussia: entant ch'ils statistichers vulan cumprovar l'egemonia da la lingua tudestga cun ils resultats da la

„Muttersprache“, fan tschertas minoritads slavicas a l'ost dal pajais pretensiuns a basa da las medemas cifras. En omadus cas fan ils statistichers valair ina visiun „objectiva“ da la dumbraziun da las linguas; en tuts dus cas daventan las cifras ina funtauna da conflicts politics.

Durant il 20avel tschientaner cumpara(n) en numerosas dumbraziuns ina u pliras dumonda(s) davart las linguas. La lingua na vegn betg unicamain interpretada sco ina data linguistica, mabain era sco in substitut da la naziunalitad, da l'appartegnientscha etnica, gea schizunt da la razza. En quest connex daventan las linguas spezialmain impurtantas en las dumbraziuns colonialas. Quai demussan las praticas britannicas en l'India, nua che la relevaziun da datas davart las linguas è in'interpresa antropologica immensa ed ils resultats èn ina referenza enciclopedica tant per ils scenziads sco era per l'administraziun (post-)coloniala.

L'URSS fa medemamain retschertgas antropologicas da gronda purtada sin la basa da las dumbraziuns dal pievel. Las statisticas linguisticas servan en emprima lingia a sclerir l'appartegnientscha da personas a gruppas etnicas, ma ellas vegnan era duvradas per cumprovar l'avanzament dal russ sco lingua da comunicaziun inter-etnica. Suenten la dissoluziun da l'URSS sa midan las prioritads en las dumbraziuns postsovieticas: la finamira generala è da minimisar la preschientscha dal russ e da cumprovar che la lingua ina giada secundaria è daventada la lingua dominante en quest pajais en transiziun. Er en ils Stadis Unids emprovan las relevaziuns da datas davart las linguas d'obtegnair dapli infurmaziuns davart l'etnia, la razza e la lingua.

Ina da las finamiras è d'eruir las personas betg anglofonas per determinar las regiuns, nua ch'ins sto furnir material da votaziun en autras linguas che l'anglais. Las statisticas americanas palaisan medemamain tendenzas discriminantas tras la creaziun dal concept da las „chasadas linguisticamain isoladas“.

Las statisticas linguisticas pon esser per las regenzas in med per imponer la politica linguistica, per las minoritads linguisticas però in med per defender lur interess. Cura che l'Irlanda è daventada independenta dal Reginavel Unì, ha ella puspè mess en discussiun la dumonda da las enconuschientschas da la lingua irlandaisa. La finamira era da determinar tge regiuns che duevan vegnir sustegnidas da la regenza per revitalisar l'irlandais; ils Brits percuenter avevan vis en questa medema dumonda in med per mesurar l'avanzament da l'anglais en ils territoris ch'els administravan. L'America latina ha furnì auters exempels, surtut en il Mexico, nua che las statisticas linguisticas vegnivan l'emprim utilisadas per rinforzar l'egemonia dal spagnol, suenter per definir las minoritads numnadas „indigenas“ che basegnan il sustegn da la regenza.

A partir dals onns 1960 survegn la dumbraziun da las minoritads linguisticas ina pli gronda atenziun internaziunala tant sin il champ scientific sco era politic. Disciplinas specificas che utiliseschan statisticas linguisticas s'estableschan en ils circuls academics. I sa tracta surtut da la geolinguistica e da la demolinguistica (geografia e demografia da linguas). La contextualisaziun istorica da la cumparsa da quests roms, surtut da la demolinguistica en il Canada, permetta da chapir l'ideologia politica che sa chatta a la basa dal svilup da questas disciplinas.

3

Las modas da relevar datas davart las linguas

Las statisticas linguisticas uffizialas dependan da numerus criteris metodologics. Cumparegliond las cifras menziunadas en pliras funtaunas enciclopedicas èsi sa mussà ch'il dumber da linguas e da pledaders inditgads po variar fermamain. Questas differenzas èn d'attribuir per gronda part a las modas, tenor las qualas las linguas èn vegnidas documentadas e quantifitgadas. Tut tenor las funtaunas vegnan tschertas varietads linguisticas reunidas en ina suletta categoria, entant che autras cumparan sut plirs nums u na figureschan schizunt insumma betg. Questas divergenzas derivan per part dal caracter dals process da documentaziun linguistica nua che ina dimensiun scientifica daventa part da projects da la politica guvernamentala. Sch'ina lingua vegn considerata u betg en ina dumbraziun, dependa medemamain d'in process da categorisaziun da las linguas e da circumscripziun dals pledaders avant u durant la retschertga statistica. Da quai dat perditga la vastadad da las lavurs realisadas dal Summer Institute of Linguistics (SIL): quel sa stenta da documentar tut las linguas dal mund cun la finamira da translatar la Bibla en tantas linguas sco pussaivel per intents missiunars.

Tar las dumbraziuns e las enquistas statisticas uffizialas sa differenzieschan anc auters aspects metodologics en connex cun la rimnada da datas davart las linguas.

Els ston vegnir resguardads per chapir ils cunfins da quai che po vegnir analisà cun l'agid da cifras. I sa tracta tranter auter da l'amplezza e da las caracteristicas da las provas statisticas. Ina retschertga che sa basa sin provas na permetta betg d'exprimer il medem sco ina vasta dumbraziun dal pievel, surtut sche las provas negligeschan tschertas gruppas da la populaziun. Ultra da quai na vegnan betg tut las retschertgas statisticas realisadas tenor la medema frequenza (tschertas dumbraziuns dal pievel han lieu per exempel mintga tschintg onns, autras mintga diesch onns). I dat differenzas, tut tenor, sche las dumondas vegnan fatgas a bucca u en scrit, ma era tut tenor il gener da las dumondas (avertas u serradas). La vegliadetgna minimala per pudair responder a las dumondas e la(s) lingua(s) utilisada(s) per rimnar questas datas fan era part dals numerus criteris che han in'influenza sin la muntada dals resultats. En la litteratura scientifica na vegn la plurilinguitad individuala betg resguardada sistematicamain. Era sche las respostas permettan d'inditgar pli ch'ina lingua, na vegn la plurilinguitad betg quantifitgada dapertut da medema maniera.

La formulaziun da las dumondas davart las linguas determinescha per gronda part tge lingua(s) dal repertori da las persunas interrogadas che cumpara(n) en ils resul-

tats. La litteratura scientifica distingua las dumondas che s'orienteschan: als origins da la persuna interrogada; a las praticas linguisticas da la persuna interrogada il mument da la relevaziun da las datas (cun distinguer mintgatant champs da diever specifics); a las enconuschientschas u a las cumpetenzas linguisticas da la persuna interrogada; a las tenutas ed a las opiniuns da la persuna interrogada areguard tschertas varietads linguisticas. Tschertas retschertgas fan ina suletta dumonda davart las linguas, autras fan pliras. Igl è da resguardar che tschertas statisticas linguisticas uffizialas vegnan elavuradas a basa da datas betg linguisticas, sco la naziunalitad, l'appartegnientscha etnica u la razza; i po dentant era sa tractar da stimaziuns calculadas tenor il dumber d'abitants domiciliads en ina regiun. La terminologia utilisada per sa referir a las dumondas davart las linguas – „lingua materna“, „lingua natala“ u „lingua discurredda a chasa“ – suonda definiziuns fluctuantas. L'enquista areguard la „lingua materna“ po sa referir: a l'emprima lingua emprendida durant l'uffanza ed anc adina chapida da la persuna interrogada il mument da la dumbraziun (Canada); a la lingua discurredda da la mamma da la persuna interrogada (India, Republica Tschecha e Slovachia); a la lingua discurredda a chasa (Stadis Unids); a la lingua, en la quala la persuna interrogada pensa e ch'ella dumogna il meglier (Svizra).

4 Recepciun ed appropriaziun dals resultats statistics davart las linguas

Las statisticas linguisticas uffizialas vegnan utilizadas savens en il champ scientific e politic senza ch'in cunfin cler separia questas duas domenas. Il cuntrari: en la gronda part dals cas exista in dialog tranter quests dus secturs. La litteratura scientifica mussa en abundanza che las statisticas linguisticas èn in element decisiv per la reconuschientscha politica entaifer la societad, surtut en connex cun las debattas davart la territorialitad linguistica e/u davart ils dretgs linguistic (access als servetschs uffizials, a translaziuns, a sustegn da linguas minoritaras sco era linguas da scola, etc.). Las debattas politicas reflecteschan savens las discussiuns davart la fixaziun da limits u da quotas per determinar a basa da quels ils pertschients u dumber absoluts che giustifitgeschan la concessiun da dretgs specifics u da subvenziuns a tschertas minoritads linguisticas. La fixaziun da quests limits u da questas quotas sveglia en general debattas fitg politisadas. Ultra da quai n'exista nagina convenziun per l'elavuraziun da quests cunfins: las decisiuns vegnan prendidas differentamain entaifer mintga stadi u regiun. Las statisticas davart las linguas vegnan era duvradas per identifitgar populaziuns. En cas cuntrari po lur assenza revelar la voluntad d'ina regenza da betg render visiblas tschertas minoritads, sco quai che mussan las statisticas franzosas

e tircas: lezzas na releveschan naginas datas davart las linguas per evitar da dar ina vusch a las minoritads respectivas e per rinforzar il sentiment d'unitad naziunala cun ina suletta lingua. La finamira primara en l'America e l'Australia è d'eruir la part da las personas betg anglofonas per savair, nua ch'i cunvegness da porscher in sustegn linguistic en il contact cun las autoritads. En il cas da quests dus pajais ha il process d'interpretaziun da las statisticas linguisticas gî tratgs discriminants.

En ils circuls scientifics vegnan las statisticas uffizialas davart las linguas recepidas fitg differentamain tut tenor ils perscrutaders che las consulteschan u las utiliseschan. La gronda part da las statisticas linguisticas vegn critigada pervia da la mancanza d'infurmaziuns davart la bilinguitad u la plurilinguitad individuala, perquai che la plipart da las relevaziuns permettan d'inditgar mo ina suletta lingua. Las largias da las dumbraziuns han intimà numerus perscrutaders da cumpensar questas mancanzas e da verifitgar la valaivladad e la precisiun dals resultats statistics cumparegliond els cun autras datas statisticas sco per exempel quellas dals registers da scola.

Ils perscrutaders menziuneschan anc auters limits tut tenor las statisticas ch'els consulteschan. Tscherts auturs per exempel suttastritgan che las datas statisticas

davart las linguas discurrendas a chasa n'èn betg represchentativas per tut las praticas linguisticas d'ina persuna, cunquai ch'igl è pussaivel da s'exprimer en autras linguas ordaifer la chasada, quai vala surtut per personas cun biografias da migraziun. Las statisticas davart las enconuschientschas u las cumpetenzas linguisticas da las personas interrogadas ston era vegnir consultadas cun precauziun: senza descripciun precisa da quai ch'è da chapir sut „enconuschientscha“ u „cumpetenzas“ en las linguas sondadas, èn las decleraziuns da las personas interrogadas betg consistentas e fidaivlas avunda.

Las statisticas linguisticas uffizialas vegnan utilizadas en numerus studis davart ils transfers ed il mantegniment da las linguas per eruir, sch'in pledader dovra la medema lingua tut sia vita u sch'el cumenza a far diever d'ina outra lingua ubain acquista novas linguas. La finamira è d'observer il svilup da gruppas da pledaders sin nivel macro-sociolinguistic. Questas analisis distinguan mintgatant ils champs da diever ed èn cruschadas cun autras variablas sco la quota da naschientschas u da mortalitad. Ellas pon avoir consequenzas politicas sco en il Canada, nua che tscherts perscrutaders transfurman lur argument scientific en ina pretensiun politica.

Ils resultats èn era vegnids interpretads sco indicaturs da tenutas en connex cun las linguas e cun ils pledaders. En il context da las dumbraziuns postjugoslavas per exempel è il fatg da responder „serbocroat“ u „serb“ u „croat“ a las dumondas davart la lingua materna vegnì percepì sco in svilup da l'appropriaziun da las denominaziuns linguisticas en relaziun cun situaziuns

sociopoliticas specificas. Sin nivel internaziunal divide la dumonda davart la dumbraziun da las linguas smanatschadas, vul dir las linguas che vegnan duvradas adina pli pauc, ils perscrutaders. Tschertins cumparegian la dispariziun da questas linguas cun ina smanatscha cunter la biodiversitad umana, auters èn da l'avis ch'ils resultats da talas relevaziuns vegnian survitaltads en vista a quitads pli impurtants da las populaziuns pertutgadas.

5 Conclusiuns

En tut las relevaziuns ed enquistas davart las linguas che nus avain resguardà en nossa revista da litteratura unescha l'argument statistic ton problematicas scientificas sco era politicas. Tscherts stadis han sviluppà fermamain l'instrument da la statistica, era sche quai signifitga da realisar plirs genres d'enquistas. En quest regard mussa l'istorgia da las statisticas linguisticas sco quellas elavuradas en il Canada co che las metodologias èn vegnidas raffinadas sut l'influenza da las pretensiuns scientificas e politicas: sin fundament da la critica constanta concernent la precisiun da las datas davart las linguas en las relevaziuns canadaisas èn vegnidas agiuntadas continuadamain dumondas per responder a pretensiuns che sa basan sin in argumentari scientific. En ina dimensiun pli pitschna chattan las statisticas dal basc gronda atenziun tras las enquistas sociolinguisticas uffizialas. Lur metodas s'allontaneschan successivamain da las retschertgas tradiziunalas cun la finamira da registrar quantitativamain ils dievers da la lingua basca sin las vias dals territoris bascofons.

Malgrà che las modas da relevar caracteristicas pli u main linguisticas sa differenzieschan evidentamain, ha la gronda part da las dumbraziuns e retschertgas ina finamira relativamain sumeglianta: la quantificaziun da las linguas sco expressiun d'ina caratteristica situada en il context social, la quala n'è betg senza consequenzas per l'organi-

saziun politica e per pretensiuns che na sa limiteschan betg a dumondas linguisticas. Las cifras pon daventar in tema central da las minoritads che vulan ina furma da reno-schientscha en in territori definì, avair access a tschertas resursas ed a dretgs, per exempel la pussaivladad da laschar scolarisar lur uffants plitost en ina lingua che en in'atra u era per definir la(s) lingua(s) uffiziala(s) d'administraziun sin nivel regional e/u naziunal. Tut tenor la legislaziun linguistica pon las statisticas davart las linguas avair in'influenza pli u main directa sin l'assegnaziun da resursas – betg mo da gener linguistic – en contexts plurilinguis, surtut cun fixar limits minimalis a partir dals quals ina minoritad po sa far valair.

Tut tenor l'influenza da las statisticas sin l'organisaziun da la vita politica po era gia la moda e maniera da rimnar las datas davart las linguas e da fixar limits minimalis esser l'object da vivas debattas. En Belgia per exempel han ils resultats da las relevaziuns già in'influenza directa sin la determinaziun da las linguas uffizialas da las vischnancas. La publicaziun da las statisticas linguisticas ha sveglià tantas tensiuns tranter ils Vallons ed ils Flams che las autoritads han decidì da stritgar la dumonda linguistica en ils formulars e da fixar ils cunfins linguistic inna giada per adina. En pajais sco la Svizra vala il princip da la territorialitad linguistica, da maniera che las statisticas uffizialas n'han betg gronda impurtanza: i vegn pre-

mess che las abitantas ed ils abitants d'in territori s'adatteschan a la(s) lingua(s) uffiziala(s) da lur vischnanca u da lur chantun da domicil. Sulettamain il chantun Grischun dovra las datas statisticas per realisar ina politica linguistica elavurada tenor dumbers da pledaders e sa decida uschia per in princip territorial dinamic che cuntinuescha a nutrir las debattas politicas odiernas.

Plirs scenziads han tschertgà meds per verifitgar la valaivladad da las datas statisticas davart las linguas u han sviluppà auters genres d'enquistas approfondadas per nianzar las cifras da las relevaziuns. Era numerusas applicaziuns èn vegnidas elavuradas. La pli evidentata è quella dals studis davart ils transfers linguistic ed il mantegniment da las linguas sin stgalim demografic ch'ins po chattar era en lavurs che s'occupan da las emprovas da mesurar la vitalitad da las linguas. En cumparegliaziuns sin nivel internaziunal chaschunan ils studis davart la vitalitad u ils transfers linguistic numerusas debattas epistemologicas che reveleschan discrepanzas metodologicas ed ideologicas.

Las debattas en connex cun statisticas davart las linguas n'èn betg restrenschidas sin il mund scientific u ad auts stgalims politics. Questas statisticas han consequenzas realas per la distribuziun da las resursas, là nua che la plurilinguitad è in tema politic. Meds finanziaus per trans-laziuns e servetschs administrativs che resguardan dumondas da las minoritads linguisticas na sa chapeschan betg adina da sez: en tscherts cas pon las statisticas linguisticas uffizialas sclerir la situaziun, sche i dat problems en connex cun talas dumondas. Era sche las statisticas vegnan

preschentadas sco ina soluziun objectiva, a la quala ins po sa referir per schliar tensiuns politicas, mussa la revista da litteratura che la scienza, sch'ella vegn utilisada per quests intents, resta influenzada fundamentalmain da dumondas politicas.

Counting languages: history, methods and politics of population censuses

Executive Summary of literature review

—
Philippe Humbert, Renata Coray, Alexandre Duchêne

1

Introduction

In many multilingual societies, official statistics on language play a major role in organising linguistic space. Indeed, data on the number of monolingual or bilingual speakers can have direct political and economic consequences in some countries, particularly regarding the allocation of financial support to linguistic minorities. Language statistics also represent a body of knowledge which is consistently referred to – and sometimes called into question – by numerous scholars, who use them to analyse demographic phenomena related to languages. Given the omnipresence of language statistics in the scientific and political spheres, the Research Centre on Multilingualism has taken on the mission of developing a state-of-the-art overview of scholarly literature on language statistics and their interpretation in social spaces.

This literature review summary at hand contains no bibliographic information, as we did not wish to narrow down references so as to no longer be representative of the overall body of literature. A complete bibliography is included in the full version of the literature review (available online in French at <https://goo.gl/FfpFNv>). The full version, which is summarised here, is of interest to researchers working with data collected in censuses as it sheds light on sociopolitical issues surrounding the creation and use of language statistics. By describing the state

of research in various contexts, the literature review provides an international framework necessary for a better understanding of Switzerland's position on language statistics. It should be pointed out that only language statistics compiled by official entities – at the national or regional level – were taken into consideration. These typically include censuses of population for whom language is one of the demographic criteria, along with age, gender or nationality.

The scope of the literature review covers research from a variety of fields: sociolinguistics, sociology, political science, demography and geography are the main areas which have addressed censuses and language statistics. Taking the most widely cited publications and authors as a basis, we systematically consulted all bibliographies and have redoubled our efforts to provide readers with an international overview. In fact, all authors consulted have worked with linguistic data from population censuses conducted at various levels – national or regional – in various parts of the world. In addition to this initial, global immersion in the subject, we conducted a thorough review of some 30 journals in sociolinguistics, political science and demography in French, English and German, and to some extent in Russian, Spanish, Italian, Romansh and Serbo-Croatian. Using a range of multilingual keywords, we further expanded our

enquiry via specialised search engines. In some cases, language appears not as the principal research topic but rather as a so-called “ethnic”, “indigenous” or even “racial” component. Thus, while language played a major role in the compilation of ethnic statistics and in the implementation of political projects, authors tended to treat it as a non-linguistic feature. As such, the selection criteria reflect the extent of the research we have conducted – as well as its limits.

Some countries and regions appear more often than others in the literature review. This is not only due to the languages used to conduct our investigation but also to the importance of surveys on language in a given society. Indeed, not all countries collect data on language with the same diligence as Switzerland or Canada. In some states, population censuses never place questions regarding language, even though some of them, such as France or China, feature a wide range of linguistic varieties. Nevertheless, language statistics collected through assessments or sample surveys are occasionally available in such countries. We observed that in most cases, the existence – or absence – of questions regarding the respondents' language or languages was often motivated by political reasons: depending on the context, language statistics – or an absence thereof – are a potential source of conflict.

To begin, the literature review and this summary address the subject from a historical point of view in order to distinguish the initial functions and the evolution of language surveys in various contexts. Then, the methodologies of gathering lan-

guage statistics are addressed in order to illustrate the full range of approaches used to categorise and narrow down the linguistic varieties of speakers. In a further step, we analyse the reaction to and appropriation of language statistics results, that occasionally generate highly politicised debates – as in the Basque Country, Belgium or Canada – and that are based on scholarly and political arguments. Finally, we synthesise these three aspects of language questions in censuses – the history, the methodology and the reaction to results – in order to identify both the main issues and the limits of official language statistics.

2 Language censuses through history

As early as the 19th century, language statistics were part of governance projects that combined scholarship and policy in the measurement process. Although quantification methods varied according to the goals of the surveys and the censuses, data on languages have consistently played a part in the foundation of nation-states, the definition of a common national identity and the distribution of resources.

The first comprehensive statistical surveys on language were carried out in Europe in the middle of the 19th century. Before that, France had already compiled estimates of dialects and patois existing on its territory with the purpose of identifying those regions in which French was not yet used as the predominant language. The aim was both to build an encyclopaedic knowledge of language practices in France and to identify regions in which dialects and patois were to be eradicated through schooling and administrative structures.

During the sessions of the *International statistical congress* (1853-1876) two census methods were discussed: one that sought to explore the respondents' "mother tongue" and one that considered the "spoken language". These two approaches were a matter of debate particularly between Austrian and Prussian statisticians. The Prussians argued that surveys considering the "mother tongue" were a more objective way of revealing respond-

ents' nationality; by contrast, the Austrians believed that language and nationality should not be mixed up, as they do not necessarily coincide. Indeed, several issues that are still a matter of debate today were discussed during the Congress. The main concerns were how to distinguish between various linguistic levels (languages, dialects, patois) and how to collect data about bilingual individuals' languages.

When compiling their data, Austria concentrated on identifying the "Umgangssprache" (language spoken on a daily basis) and Prussia on the "Muttersprache" (mother tongue). By collecting data on the language spoken on a daily basis, the Austrian government angered many Slavic minorities, who accused the government of trying to erase them from the results, claiming that this practice emphasised the use of German – a language mainly used in more formal contacts – and thus obscured the presence of other languages within the Austrian Empire. In Prussia, statisticians hoped to use their results on "Muttersprache" to demonstrate the supremacy of the German language throughout the territory, while some Slavic minorities used the same figures to make opposing claims in the east of the country. In both instances, statisticians put forward an "objective" approach of collecting data on language; and in both instances, the results were a source of political conflict.

During the 20th century, many censuses asked one or more questions about languages. An individual's language was considered not only as statistical data on language but also as a substitute for nationality, ethnicity and even race. As such, languages were given special attention in colonial censuses. British practices in India bear witness to this: recording languages was a colossal anthropological undertaking, the result of which constituted an encyclopaedic reference work for both scholars and the (post-) colonial administration.

The USSR also undertook extensive anthropological research using population censuses. While language statistics primarily served to identify the ethnic groups individuals belonged to, they were also used to demonstrate the progression of Russian as the language used for interethnic communication. After the collapse of the USSR, the majority of censuses in the post-Soviet era reversed priorities: the goal was to play down the presence of Russian and to show that the local language, which had been reduced to the status of second language, had become the dominant language in the former Soviet satellites undergoing a transition. In the US, census questions on language also sought to gain more information about ethnicity, race and language. If one of the goals was to identify groups of non-English-speaking individuals to determine areas where voting material should be provided in other languages than English, the American statistics also led to discriminatory practices through the concept of "linguistically isolated households".

While governments may have considered language statistics as a way to impose lan-

guage policies, linguistic minorities saw them as a means to defend their interests. After gaining independence from the United Kingdom, Ireland reappropriated the issue of knowledge of the Irish language for the purpose of compiling statistics to help determine which regions should receive government support to revitalise the Irish language. The British, by contrast, had seen these same statistics as a means to evaluate the progression of English on the territories under their governance. Latin America provides other such examples, for instance in Mexico, where language statistics were used first to confirm the supremacy of the Spanish language, and then to determine which so-called "indigenous" minorities required government support.

From the 1960s onwards, data on linguistic minorities became an international phenomenon, both in the scientific community and on the political stage. Specific scholarly disciplines that rely on language statistics were growing in importance in the scientific world, especially geolinguistics and demolinguistics (otherwise known as language geography and demography). The historical contextualisation of the emergence of such fields, in particular demolinguistics in Canada, sheds light on the political ideology underlying the development of these scholarly disciplines.

3

Methods to collect data on language

Official statistics on language rely on a number of methodological criteria. A comparison of figures published in several encyclopaediae reveals great variations in the number of languages and speakers accounted for. These differences are mainly due to the methods used to document and quantify languages. Indeed, depending on the source, some linguistic varieties may be grouped under a single category, appear under different names or not be mentioned at all. Such divergences are partly due to the nature of the processes used to document languages since they embed the scientific component into political governance projects. Whether or not a language is considered as such in a census is also induced by the method used to categorise languages and to determine the constituencies of speakers prior to or while compiling the statistics. This is highlighted by the extent of the work carried out by the Summer Institute of Linguistics (SIL), which seeks to document all languages in the world and whose original mission is to translate the Bible into as many languages as possible for its missionary work.

Concerning official censuses and statistical surveys, there may be other methodological differences in the collection of language data. It is important to take into account aspects such as the size and features of the samples studied in order to understand the limits of what can be analysed

based on the given data. Indeed, a sample survey does not allow to draw the same conclusions as an extensive population census, especially if the sample excludes some sections of the population. Furthermore, not all statistical data are collected at the same frequency. For example, some population censuses are carried out every five years, while others take place every ten years. Whether the questions are asked orally or in writing also leads to differing results, as does the type of question (open or closed). The minimum age of respondents and the language(s) used to collect data belong to a multitude of criteria that potentially impact the interpretative scope of results. In the scientific literature, it appears that the respondents' multilingualism is not systematically taken into account. And when surveys allow for more than one language, multilingualism is not necessarily quantified in the same manner.

The wording of questions on language largely determines which language(s) within an individual's repertoire will appear in the results. Scholarly literature distinguishes between questions which focus on: the respondent's origin; the respondent's language practices at the time the statistical data were collected (sometimes distinguishing between specific fields of use); the respondent's linguistic competence or skills; the respondent's attitude and opinions regarding certain language varieties.

Some statistical surveys contain only one question on language while others include several. It should also be noted that some language statistics are compiled on the basis of non-linguistic data such as nationality, ethnicity or race, or they may be estimates calculated according to the number of inhabitants domiciled per region. The terminology used to refer to questions regarding languages – “mother tongue”, “native language” or “home language” – is inconsistent. For example, the data recorded under “mother tongue” could refer to: the first language learnt during childhood and still understood by the respondent at the time the census was conducted (in Canada); the language spoken by the respondent's mother to her/him (in India, Czech Republic and Slovakia); the language spoken at home (in the US); the language in which the respondent thinks and says he or she is most proficient (in Switzerland).

4

Reaction to and appropriation of statistical results on languages

Official language statistics are often used by the scientific and political communities, with no clear boundary between the two domains. Indeed, in many cases dialogue is established between the two communities. Scholarly literature abounds with examples of situations in which language statistics are a major issue for political recognition within a society, for instance, debates on linguistic territoriality and/or language rights (access to official services, translations, promotion of minority languages as a language of instruction, etc.). Political debates frequently concern the definition of thresholds or quotas for a percentage, or for absolute numbers, that must be met before specific rights or subsidies are granted to a linguistic minority. Setting such thresholds or quotas generally gives rise to hotly contested debates. Moreover, no accepted convention for determining these thresholds exists, and every state or region has their own decision making process. Language statistics are also used to identify populations, although the absence of such data may reveal a government's desire not to grant visibility to certain minorities, as illustrated by French and Turkish statistics, which do not collect data on languages to avoid giving their minorities a voice and to strengthen the sense of national unity provided by a single language.

In the case of the US and Australia, the foremost goal is to identify the percentage of non-English speakers in order to determine where language support should be provided with regard to official contacts. In both countries, however, the process of interpreting language statistics has also given rise to discriminatory practices.

In the scientific community, official language statistics are received differently, depending on the researchers who consult or exploit them. For instance, most language statistics are criticised for their lack of data on respondents' bilingualism or multilingualism, as most statistical surveys limit their questions to only one language. The shortcomings of censuses have prompted many researchers to address these weaknesses and to verify the validity and relevance of statistical results by comparing them with other statistical data, such as school registers.

Researchers cite further limitations, depending on the statistics they consult. For example, authors point out that the data collected on languages spoken at home are not representative of all language practices of affected individuals, as they may speak other languages outside of their home. This is particularly the case in populations with a migration background. Statistics on respondents' linguistic competence and skills

should also be handled with caution, as without precise descriptors of what is meant by competence or skills in the language surveyed, respondents' statements are likely to be inconsistent and unreliable.

Official language statistics are used in numerous studies on language shift and language maintenance that address whether an individual will continue using the same language throughout his or her life or, on the contrary, whether an individual will switch to another language or learn other languages. The purpose is to examine how groups of speakers evolve at a macrosociolinguistic level. These studies sometimes distinguish between domains of use and are compared with other variables, for example, fertility and mortality rates. The analyses can have political consequences, as is the case in Canada, where some researchers turn their scholarly findings into a political argument.

Results can also reveal attitudes towards languages and speakers. In the case of post-Yugoslav censuses for example, answering "Serbo-Croatian", "Serbian" or "Croatian" to questions about the mother tongue can be viewed as an evolution in the appropriation of language denominations in relation to specific socio-political developments. At an international level, the measurement of endangered languages, i.e. languages that are used less and less, is a topic that divides researchers. Some consider the disappearance of these languages to be a threat to human biodiversity, while others believe that these calculations ignore issues that are far more important for the populations concerned.

5 Conclusions

In all the censuses and language surveys we have considered in our literature review, the statistical argument combines both scientific and political issues. Some states have invested considerably in developing appropriate statistical tools, some of which even include various types of surveys. In this regard, the history of language statistics such as those developed in Canada, demonstrates how methodologies have been refined following arguments that combine scholarly and political reasoning. As a result of constant criticism directed at the precision of language data in Canadian censuses, additional questions are gradually being added to meet scholarly requirements. On a smaller scale, statistics on the Basque language are gaining momentum through official socio-linguistic surveys that rely on methods that are gradually moving away from traditional surveys with the aim of better quantifying the use of the Basque language on the streets of Basque-speaking territories.

Despite the obvious diversity in how features of a more or less linguistic nature are collected, most censuses and surveys have a relatively similar goal: the quantification of language as the expression of a social feature which may have consequences for political processes and demands that go beyond linguistic issues. The statistics may indeed become a warhorse for minorities who seek recognition in a given territory or who

wish to gain access to specific resources and rights such as the opportunity for their children to attend school in one language rather than another – or to define the official administrative language(s) at a regional and/or national level. Depending on language regimes, language statistics can have a more or less direct impact on the allocation of resources – not only linguistic resources – in multilingual contexts, in particular when it comes to defining minimum thresholds for minorities to win their case.

Depending on the impact of statistics on the organisation of policies, the very methods for collecting language data and setting minimum thresholds may give rise to heated debates. In Belgium for example, census results had a direct impact on determining the official language in municipalities. The publication of language statistics caused so much tension between the Walloons and the Flemish that the authorities decided to remove language questions from questionnaires and to set language borders once and for all. In other countries, such as Switzerland, the principle of linguistic territoriality prevails, meaning that official statistics are of little importance: a territory's inhabitants are required to adapt to the official language(s) in their commune or canton of residence. The canton of Graubünden is the only Swiss canton to use statistical data to implement language policies that are drawn up according to thresh-

olds, thus opting for a dynamic territoriality principle, which to this day continues to feed political debate.

Several scholars have considered ways to verify the validity of statistical data on languages or have developed other types of in-depth surveys to qualify census results. Numerous areas of application were also developed, the most obvious being the studies on language shift and language maintenance on a demographic scale, which can equally be found under other headings, for instance, in attempts to measure language vitality. In international comparisons, studies on language vitality or language shift give rise to numerous epistemological debates that reflect methodological and ideological differences.

Debates generated by language statistics are not limited to the world of science or to high political spheres. Indeed, these statistics have real consequences on the distribution of resources in areas where multilingualism is a political issue. Taking into account demands of linguistic minorities for the creation of financial funds for implementing translations and administrative services is not always straightforward: in some cases, official language statistics are used to settle disputes when problematic issues arise. Nevertheless, although statistics are presented as an objective solution to solve political tensions, the literature review shows that science, when used for such purposes, remains fundamentally influenced by governance issues.

