Why we need better language maps,
and what they could look like

SEBASTIAN DRUDE

Why we need better language maps, and what they could look like

Sebastian Drude

Vigdís International Centre for Multilingualism and Intercultural Understanding
The University of Iceland • Brynjólfsgata 1 • 107 Reykjavík • Iceland
[sdrude@hi.is]

Abstract

Current language maps (maps aiming at showing the localization of languages) usually display either mostly non-overlapping areas or single dots for language localizations. Both are unsatisfactory given the much more complex geographical distribution patterns of speakers and language use. Future maps should take diglossia and diverse areas into account as well as multilingual individuals. The conceptual basis are patterns of language domains of use and significance. Each language needs a large set of maps for different domains, which then can be combined (through overlay or more advanced techniques) for one language or several languages. Interactive maps can give access to external additional information or allow visualization of developments over time.

Introduction: why language maps?

Linguistic maps in general show the geographical distribution of language-related phenomena with cartographic means. I focus here on the core type of ***language maps***, which show the geographical distribution of languages (or dialects, or language groups) themselves, leaving aside maps dedicated to specific phenomena related to linguistic form or meaning, such as typological maps (which represent the existence or shape of phonological, morphological or syntactic items, categories, systems or constructions). Note that this paper does not focus specifically on *dialect maps* (i.e. maps where dialects are in focus than rather the distribution of different languages, showing phonetic / phonological isoglosses, or the distribution of words for a certain concept) of one language or group of closely related languages.

Many language maps have a national scope, showing dialects of the majority language(s) together with minority languages in a country or similar (e.g. administrative) area.[1] While these are not outside the scope of this paper, I am here especially interested in maps that are supra-national up to global in scope, potentially covering all languages of a region.

As languages are a key factor for ethnic and national identity, there is a large overlap between language maps and maps showing the distribution of ethnic groups, nations or peoples; where these can be clearly distinguished, the focus is again not on the latter.

I believe that language maps are a relevant topic for linguistics, especially 'diversity linguistics',[2] because maps in general are an important tool of visual communication, allowing a visual grasp of sometimes complex phenomena – if these phenomena have a significant distribution over the geographical space at all. Languages certainly do have such a distribution, and in
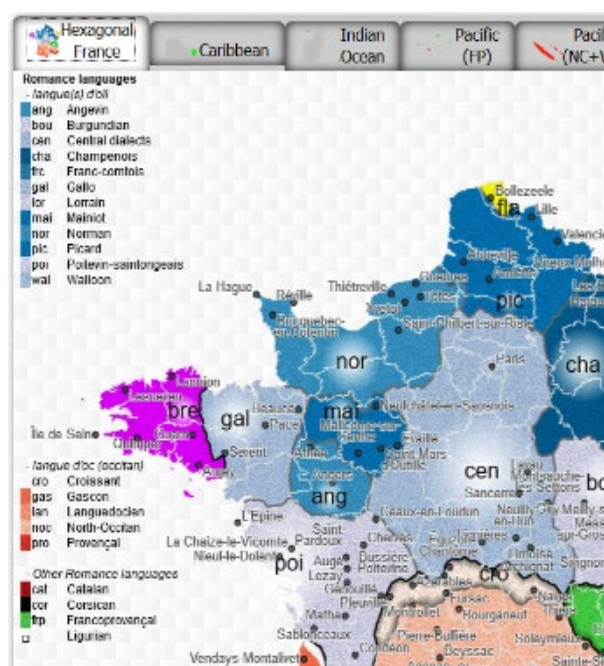
Figure 1: Atlas sonore des langues régionales de France

particular with the recent concern for (threatened) linguistic diversity and endangered languages, maps have proven to be an excellent communicative vehicle: and therefore, many journalistic reports on the topic have included, or sought to include, cartographic illustrations.

Furthermore, maps even stand out among other visualization techniques such as diagrams or pictures, because they seem to inherently attract the interest of the audience. Everybody can relate to maps by focusing on places that they are familiar with, and we all (at least western educated people) are very much used to reading and digesting cartographic information.

---

[1] If offered digitally, in particular online, such dialect and national maps can be enhanced with multimedia samples. An excellent example is the "Atlas sonore des langues régionales de France" (Mareüil, Rilliard & Vernier n.d.).

[2] This term has been coined by Martin Haspelmath and others at the Max-Planck-Institute for Evolutionary Anthropology in Leipzig. See also Drude (to appear).

Linguistic interest in global geographical language distribution had a first high with an increased interest in global and regional language diversity since the first decade of this century.[3] There is now potential for a new impulse with the current focus on migration and the rise of 'hyper-diverse' multilingual cities (King & Carson 2016). Many actors, public and private, now want or need to understand better which languages are spoken by which, especially new, members of the society, how we can communicate with people visiting or settling at new places, and which linguistic competences they bring with them, often cultivating them over generations, with positive or negative consequences for their success in the matrix society.

In this paper I show which kind of language maps are currently generally available and produced (section 0), and why this is unsatisfactory given the kind of language-related geographical information we need (section 0). Then I list desiderata for more adequate and relevant language maps (section 0), and finally I offer some first thoughts on how these could be implemented and what would be needed to make progress in the field of language mapping (section 0). In general, the purpose of this paper is to provide some conceptual foundations, and to function as a call for action, trying to join interested individuals, institutions and initiatives to work together on advancing language mapping as a relevant tool for knowledge representation and communication.[4]

Current major language map types

The most common type of language map[5] divides a geographical region (often a country, a continent, or even the world) into areas, indicating which language is spoken in each area. What exactly "is spoken" means is often not made explicit; in many cases it can be interpreted as "is predominantly used" or "has an official status", in other cases it may mean "is spoken at all" (implying minimally: "people who know this language are/were living there"). The former often holds for national and official languages, the latter for minority & indigenous languages – they may thus appear on language maps even if they are hardly actively used by anyone.

Often the division of a geographical region in such areas is exhaustive – every (land) area is assigned to some language area. These areas are usually represented by 'polygon shapes' filled with a particular colour or pattern. They are almost always monochrome with no internal variation, which implies that no further distinctions can be made, such as, for instance, how densely populated the areas are in the first place,

Figure 2: Languages in Europe (detail); Source: user '*Andrei nacu*' at English Wikipedia

assigning even deserts to one or another language, possibly according to administrative borders.

If the division is not exhaustive, then unmarked (colourless) areas designate either uninhabited geographical regions, or the map focuses on only certain languages (such as the indigenous languages of a country), and does not mark the areas of other languages (such as the dominant 'default' languages, e.g. an official national
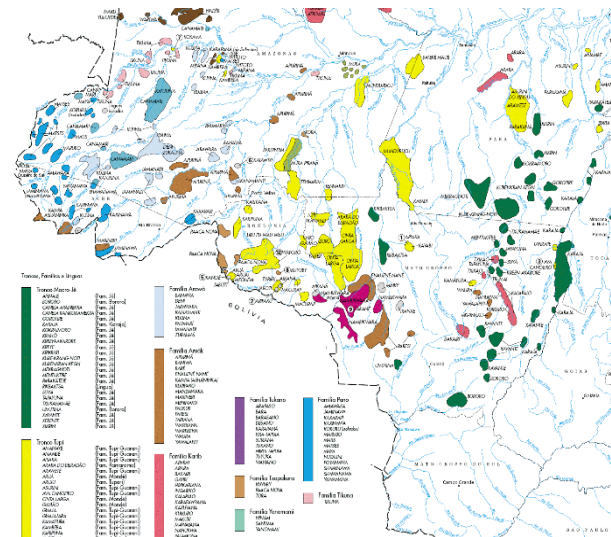

Figure 3: Linguas indígenas na Amazonia brasileira (detail)

language) with any specific colour. The example Figure 3 is from Queixalós and Renault-Lescure (2000).

In this first type of map, 'polygon maps', sometimes minority languages are shown together with dialects of major languages, forming a mixed language *and* dialect map type (see Figure 1, above).

Only sometimes do such polygon maps try to represent diglossic areas (where two languages or major varieties exist side by side, with different social functional roles & prestige), for instance by displaying hatched areas (extremely simple in Figure 4).

Figure 4: Languages in Belgium

Another type of language map shows not polygons (areas), but rather one dot (a geolocation, usually defined by a pair of latitude and longitude coordinates) for each language. This type seems to be newer and possibly became popular with the wide availability of GIS (Geographic Information System) technology, promoted by tools such as *Google Earth* where "placemarks" are such precise dots and are used to locate phenomena (even if these are not restricted to one specific geolocation). For languages, this type has predominantly been applied when the aim of the map was to be exhaustive as to the languages under focus, especially on a global scale, perhaps first and most prominently in the UNESCO atlas of endangered languages.

What exactly the point stands for, and why this specific location has been chosen to represent a language, is often not made explicit. One can imply that the location should represent either the 'geographical centre' (for instance in the sense of being the 'centroid' or 'geometric centre'). In other cases, especially for languages with a wide distribution spanning major parts of a continent or several continents, the location chosen appears to be the historical origin of the language (for instance, for English, in the centre of England). Examples below.

There are also mixed types of these two types of maps, combining areas for larger languages (or language groups) with exact locations for smaller languages.

Major current global language map initiatives follow one of the two approaches:

One of the first complete set of language maps aiming at showing the distribution of all languages organized in families available online was provided by Jost Gippert on the *TITUS* pages (Gippert 1993) (Figure 5).

Asher & Moseley (2007) usually show non-overlapping areal maps (no sample maps available).

SIL International (formerly the Summer Institute of Linguistics) maintains the Ethnologue (Simons & Fennig 2017; for a critical review, see Hammarström 2015), the best probably known catalogue of the languages of the world. Since its fourth edition (1953), the Ethnologue also provides language maps for continents and countries

(Simons & Gordon 2004), and in recent editions also more detailed maps – but these maps are now behind a paywall. Steve Huffman has produced maps based on earlier SIL maps, for the "World Language Mapping
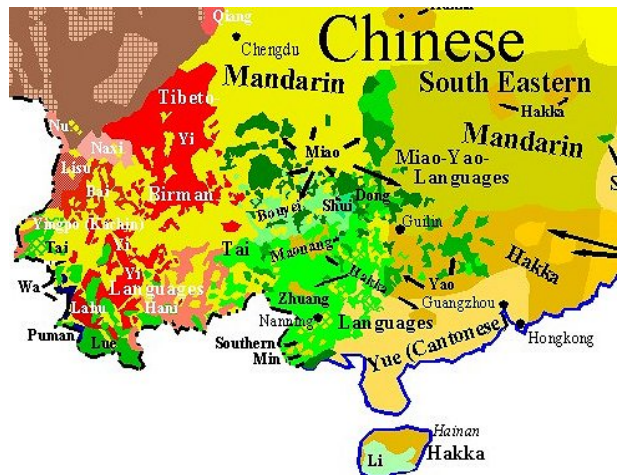

Figure 5: Detail of a TITUS map

System" (World GeoData Sets, which has in turn been bought by SIL international in 2017) (Figure 6).
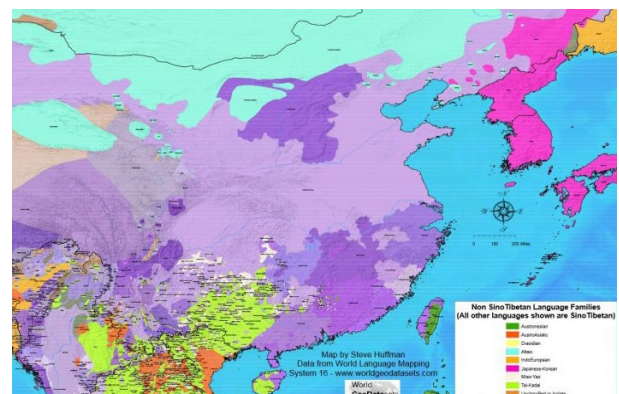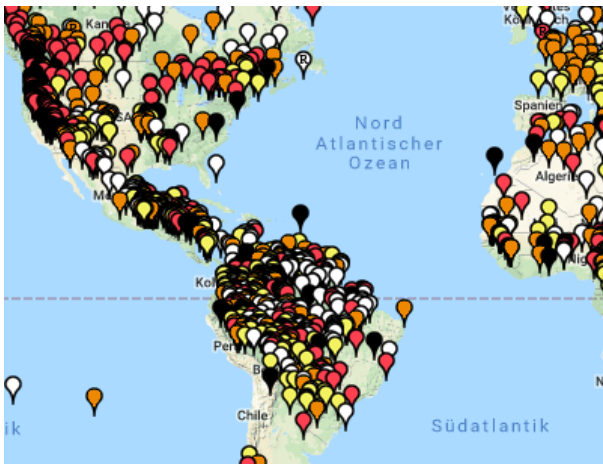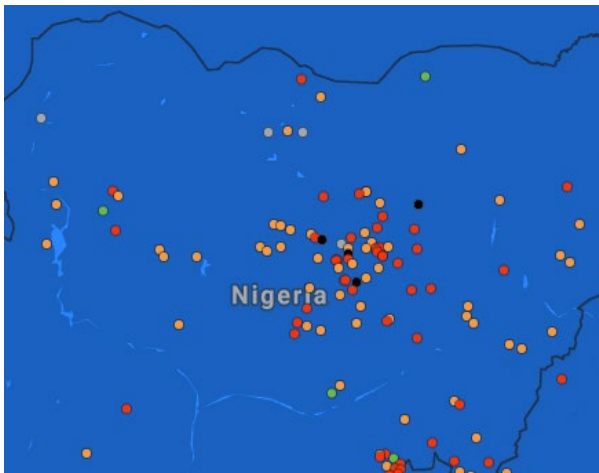

Figure 6: Sample map "World Language Mapping System"

Figure 7: Map by Steve Huffman (detail)

Figure 8: Detail of the online map of UNESCO's atlas

UNESCO's Atlas of the World's Languages in Danger (3rd edition: Moseley 2010) has geolocations for 2464 endangered languages. The information of how the coordinates have been obtained and compiled is currently

Figure 9: Detail of the Endangeredlanguages.com map

not available online on UNESCO's pages. UNESCO plans for a much more comprehensive atlas of all the world's languages. (see Figure 8)

The Endangeredlanguages.com map (Alliance for Linguistic Diversity 2013ff, see Figure 9) also shows coordi-

Figure 10: Sample LLmap (detail):
Australia at the Time of Contact: Non-Pama-Nyungan

nate locations for "more than 3000" endangered languages, based on the Endangered Languages Catalog (ELCat: Campbell, Aristar & Aristar-Dry 2013), which relies of other sources, especially the UNESCO atlas.

The LinguistList maintains a collection of language maps (**LLmaps** currently ~680) compiled from the literature (LinguistList 2018), displaying them as an overlay to a base map in a specialized viewer. The maps contain areas (see example), dots and sometimes lines and arrows, depending on the original map (Figure 10).

Glottolog (Hammarström, Forkel & Haspelmath 2017), arguably the most comprehensive and accurate catalogue of the world's languages (Drude to appear), offers one geolocation for each language. This dataset is freely available and has been used in a number of language mapping projects, for instance the **Glottolog Data Explorer** (Caines et al. 2016) (Figure 11).

Figure 11: Map for the Glottolog Data Explorer (detail)

Also created at the Max-Planck-Institute EVA in Leipzig, **WALS**, the World Atlas of Language Structures (Dryer & Haspelmath 2013) offers typological maps which show the distribution of structural features (each map for on average 400 languages). The locations are given as a dot, based on previous published work, in particular (Asher & Moseley 2007).

The **Langscape project** at the Maryland Language Science Center works on more informative language maps using advanced visualization, apparently much in line with the thinking of this paper. Currently it uses the Glottolog database. "Earlier development work was done by [the University of Maryland's] Center for Advanced Study of Language, which also contributed a great deal of data." ('About Langscape' 2015).

There are, of course, many maps for countries, continents and the world produced by individual linguists, geographers and language-aficionados; too many to even start to list them here. However, one comprehensive and very attractive project deserves an honorary mention: the beautiful language area maps produced by 'Muturzikin' (2011). Unfortunately, the project has not been continued since 2011 (Figure 12, next page).

Problems with current language maps

The current practice with language maps is not satisfactory for a number of reasons.
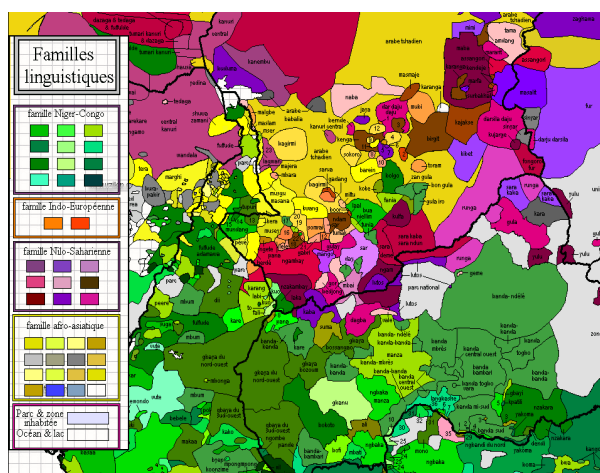
Figure 12: Muturzikin's W.-Central African map (detail)

Polygon / area maps often cover only a selection of languages, usually focusing on official or major languages only, leaving out smaller and minority languages. They also tend to group languages together; the wider the coverage, the less differentiation is possible. On a worldwide scale, usually only a (few) dozen languages or language families/groups are shown. This may be useful for a first orientation, but it is an unduly gross simplification given the around 6500 languages spoken worldwide.

Punctual / geolocation maps (showing only dots), in turn, make it easier to include all languages of a region without omissions, but they obviously do not show the geographical extension of any language area, however defined. On a continental or global scale, this may be appropriate for local languages spoken in a small region which on the scale of the map appear as a dot rather than an area. Still, when one is interested in the details ('zooming in'), this is not adequate, and not only for larger languages with a wider distribution.

There are other aspects where most or all current maps fall short. In particular, they only exceptionally take the consequences of recent migration of (people bringing with them their) languages into account; rather, they represent only the traditional 'homeland' of the languages, not their diaspora. Closely related to this is the shortcoming that areas where several languages co-exist are not well represented in language maps – even when maps provide overlapping or hatched areas, they usually do not show more than two languages in use in one location, but in many places there are many languages, in particular in modern cities it can be dozens and even hundreds of languages (King & Carson 2016).

Equally importantly, language maps fail to show (individual) multilingualism. The existing language maps suggest an unreal picture as, in its extreme, idealized by the nation-state ideology: linguistically homogenous communities live in geographically confined areas, neighbouring with similar areas and communities, but with no overlap and shared allegiances; linguistically speaking: with only monolingual members.

Bilinguals and more so multilinguals who are able to speak (with some degree of fluency) more than one language are often considered to be an exception, however, multilingualism (either by growing up with more than one native language, or by acquiring them later through learning, e.g. in schools) is, currently and historically, not the exception, but rather the rule (Edwards 2012). Multilingualism exists in every country of the world, in every class of society, in all age groups (Grosjean 2012). The different consequences of the ongoing increased integration of remote areas and countries (globalization) – such as, for instance, the current migrations as well as the learning of languages at school – generate ever more multilinguals.

In any situation, multilinguals have to opt for using one of the languages. Often the choice is obvious, if the institutional setting requires the use of one particular language, or if only one language provides a common ground for all interlocutors. In other situations, such as informal conversations between friends and family members, more than one language could serve the bare communicative needs. In these situations, individual attitudes as well as societal and official language policies guide the choice for one or another language, often unconsciously. This is part of the linguistic landscape, but is even less well represented on language maps.

In sum, current maps are not helpful in getting a clear picture of the linguistic situation in a region: Where do which groups use which language? Which competencies do exist there? The whole 'linguasphere' (a term created by D. Dalby (cf. Dalby, Barrett & Mann 1999)) is much more complex and intricate than the simplistic polygon or point maps show: there are languages with a global distribution (in particular English, which is an official language in numerous countries and is the preferred first foreign language to be learned/taught at school in many more) alongside with local languages; in many communities, more than one language is used in different circumstances, and multilinguals adapt to these circumstances and chose which languages to use, to maintain and to pass on to the next generation. No existing language map (that I am aware of) on any larger scale is capable of making these complex patterns visible.

Desiderata for future language maps

In early stages of project development, I believe it is important to brainstorm without restrictions, instead of censoring the ideas for limitations which later may well turn out to be surmountable. Therefore, in this section I provide some general thoughts without considering the feasibility nor the technical implementation; these will be briefly addressed in the next section.

The most basic shift I propose is that the enhanced next generation of language maps should be based on *speech events* rather than communities of speakers. That is, the maps should primarily show not only where but also when a language is used, and for what it is relevant. This crucially depends on the different *contexts* for these

speech events or language relevance. It may be common that the majority of a population has a certain language as mother tongue, and/or that a language is official in the region where that population lives, but this is just one of many settings.

Wherever some individuals live (or are frequently present) who speak more than one language (and that may be pretty much everywhere), we have a multilingual setting, which often manifests itself as diglossia (where one [or few] of the languages/varieties has a high status and is dominant in public and official situations of language use, while the other[s] is/are seen to be inferior, often used more in informal contexts). What is fundamental in multilingual settings is that the division of use between the different languages involved typically goes according to different domains of language use. Therefore, I propose that these domains or areas of relevance need to be captured and visualized in language maps.

In what follows, I list some domains and areas of relevance. For each of them, a map could show where this kind of language use/relevance setting is present. Some of the closely related items have been conflated together in one entry below; they contain some variables, indicated by slashes "/" or "and/or" formulations. Each different selection or setting for such variables implies a slightly different setting with different results (maps) which has not been separately listed for reasons of space.

Language maps showing where a language X …

1. … is used by the local inhabitants who have X as first language or strong main language. (This is the main aspect covered in many current maps, see section 1. above.)
2. … is used daily and (almost) exclusively by the whole population of that area. (For English, there would be few places, most probably in rural areas of the UK or US, where this holds.)
3. … is used daily and exclusively by a certain percentage P or more of the individuals of the area. (There will be different results, and hence possibly different maps, for different values of P.)
4. … is used as a lingua franca by (native and) non-native speakers.
5. … is the main official language of a state / a regional administrative unit. (This is another main aspect covered in many current maps, see sec. 1. above.)
6. … has an official status as secondary / minority / heritage language in an administrative unit.
7. … is used by highly fluent speakers frequently but alongside other languages.
8. … is used by and with non-resident visitors (e.g. tourists; merchants, people across a border visiting for shopping etc.).
9. … is used in homes of families and possibly among friends.
10. … is used in the public area (market, cultural events).
11. … is actively used in university teaching classes.
12. … is used for publishing scientific work.
13. … is required for university learning (e.g., reading of academic literature in the language is presupposed).
14. … is used in local print and/or broadcasting media.
15. … is taught to children (obligatorily/optionally) in public or general schools (from which age on).
16. There are second language learners in private language schools for adults.
17. … is learned remotely using online courses or resources.
18. … is used in digital media (webpages, online resources…). (This may be difficult to localize geographically below larger administrative areas.)
19. … is used in social media (by all / some percentage of the population which uses social media at all).
20. … is the source and/or target of translation activities for official documents / for literature.
21. … has recordings and similar (in particular digital) data stored/archived by some institution or individual.
22. … has material that is accessed by interested users from the area in question.

Furthermore, each of these settings of language use or language relevance can be observed over time, showing different developments which may be at least as meaningful as understanding the current situation. So time adds a whole additional important dimension to these envisaged future language maps.

Implementation

Each of the domains in the previous section, and potentially more, could be represented by a separate base map for each of many languages. Only for very restricted local languages may the areas for several of these domains coincide (have the same extension), or some of the domains may not exist or not be relevant at all.

But generally, all these different domains are significant and tell us something about the geographical extension where the language is used or is relevant. This means that the full picture, for each language, is best to be expressed by a *set* of maps, one for each of the domains (for which data are available).

A challenge is then to combine the different maps for one language in a way that makes them meaningful and interpretable. With digital technology this should be possible to achieve in collaboration with specialists in digital cartography and modern visualization techniques (which have rapidly developed in recent years).

For a start, each map can be a transparent layer, where different colours and/or patterns symbolize the respective domains. The overall result would be a colourful overlaid language domain landscape, where more intense colours (through overlaying of several layers) indicate that the language has relevance for several domains. On the other hand, one may also consider using shading or varying

colour intensity on *single* layers (versions of choropleth maps) for gradual phenomena, such as percentages of speakers in a larger population (points 2 and 3 in the list
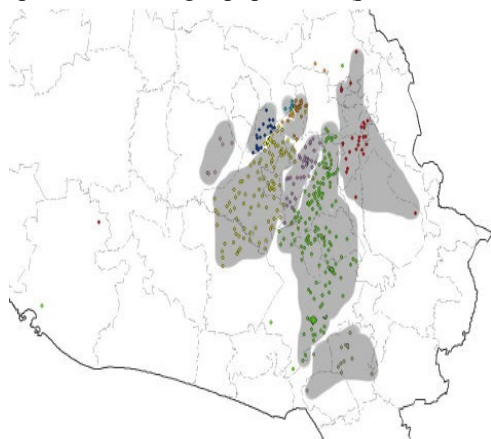

Figure 13: Map showing areas combined with locations of villages.

in section 4), or density of speakers per area, or similar. Therefore, the combination of these layers is far from trivial, and in many cases some kind of animation may help to get the complex combined information across.

First steps are to combine areas with individual points, as has been done in Figure 13 (Cline, Marlett & Paterson III 2012: 8, next page). The *Lingtypology* package (Moroz 2017) now allows the easy production of such maps with the R program, using the Glottolog database. By changing the underlying database, the model can easily be applied to other data sets.

As digital technology allows a much more interactive way of displaying such maps, questions of user needs and dynamic procedures of selecting and combining layers come into the picture. One obvious advantage of current digital visualization technology is that elements in the maps can be interlinked with further information, as has been done in the case of the maps shown in Figures 1 (with a sound sample for each dialect / language), 8, 9, and 11.

In this context, if the data allow, one would also aim to display the development of such areas over time, possibly again as animations. In combination with interactive layer selection and combination, and if the technology allows jumping to certain points on a timeline, or accelerating or slowing down playing a connected sequence of images, such interactive maps can develop into real research tools in addition to mere attractions on a webpage or in an exposition.

The bigger challenge is, it seems, to obtain the needed data to create each individual map in the first place. The methods evidently differ considerably depending on the size and status of the language. For major and official languages, certain information may be readily available while for many local and endangered languages only specialists and the speakers themselves will be able to provide information about many of the questions. To fill in a matrix for all (and more) domains in each known

language is obviously only possible with wide collaboration and many contributions from very different people and institutions. It may be a good approach for different projects each to focus on a smaller region with a limited group of heritage languages, getting as much information for as many of the domains as possible, and only cover selected domains for recently immigrated languages, for instance.

For major languages, some relevant information can possibly be extracted from existing and easily accessible sources such as almanacs or official government or United Nations data (for instance data underlying points 5, 6, 11, or 15, above). In some countries, language related data is obtained in general census data (potentially providing a basis for points 1, 2, and perhaps 3, above). For other domains, research of online sources combined with consulting experts and the scientific literature as well as commercial information sources may be the best approach (in particular 10, 12, 13, 14, 16, 17, 18, 20, or 21). Targeted empirical research (collecting many observations, or using questionnaires) would be needed for other points (e.g., 4, 7, 8, or 9). The use of advanced digital methods may be useful for some points (such as 19, if combined with automatic language recognition; or 22, in collaboration with specialized archives and data centres).

For medium-sized (around 10.000–500.000 speakers) languages where internet access is available, one could think of crowd-sourcing approaches, mobilizing the speakers themselves.

For data obtained from online or official sources or scientific literature, the sources would ideally be shared in a central location, so that later projects can make use of sources and results of earlier projects where relevant. Often the same source may contain relevant information about more than one language.

References

About Langscape. 2015. (langscape.umd.edu/about/) (Accessed 2018-08-12)

Alliance for Linguistic Diversity. 2013ff. Endangered Languages Project. (www.endangeredlanguages.com) (Accessed 2018-05-01)

Asher, R.E. & Christopher Moseley (eds.). 2007. *Atlas of the World's Languages* 2nd ed. London: Routledge.

Auer, Peter. 2013. The geography of language: Steps toward a new approach. *Freiburger Arbeitspapiere Zur Germanistischen Linguistik. FRAGL* 16.

Caines, Andrew, Christian Bentz, Dimitrios Alikaniotis, Fridah Katushemererwe & Paula Buttery. 2016. The Glottolog Data Explorer: Mapping the world's languages. In *VisLR II Workshop*,.

Campbell, Lyle, Anthony Aristar & Helen Aristar-Dry. 2013. About the Catalogue of the Endangered Languages. (drive.google.com/file/d/0B7vQLUpU_2qcV EVqNzFKWFA2cUk/view) (Accessed 2018-05-01)

Cline, Kevin L., Stephen A. Marlett & Hugh Paterson III. 2012. Las Conexiones Externas e Internas. In Stephen A Marlett (ed.), *Los Archivos Lingüísticos Me'phaa*, 15. Retrieved from www.mexico.sil.org/es/publicaciones/e-wpindex/ensayos_preliminares_-_archivos_linguisticos_mephaa (Accessed 2019-03-22)

Dalby, David, David Barrett & Michael Mann. 1999. *The linguasphere register of the world's languages and speech communities* 1st ed., Vols 1–2. Hebron, Wales, UK: Published for Observatoire Linguistique by Linguasphere Press / Gwasg y Byd Iaith.

Drude, Sebastian. to appear. Reflections on Diversity Linguistics. In Bradley McDonnell, Andrea Berez-Kroeker, & Gary Holton (eds.), *Reflections on Language Documentation on the 20 year anniversary of Himmelmann 1998*. Honolulu, Hawaii: Univ. of Hawaii Press.

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from wals.info

Edwards, John. 2012. *Multilingualism: Understanding Linguistic Diversity* 1st ed. London; New York: Continuum.

Gippert, Jost. 1993, 2010. TITUS: Language maps. (titus.uni-frankfurt.de/indexd.htm?/didact/didact2.htm#karten) (Accessed 2018-08-11)

Grosjean, François. 2012. *Bilingual: life and reality*. Cambridge, Mass.: Harvard University Press.

Hammarström, Harald. 2015. Ethnologue 16/17/18th editions: A comprehensive review. *Language* 91(3). 723–37. DOI: doi.org/10.1353/lan.2015.0038 (Accessed 2018-06-01)

Hammarström, Harald, Robert Forkel & Martin Haspelmath (eds.). 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. Retrieved from glottolog.org

King, Lid & Lorna Carson (eds.). 2016. *The multilingual city: vitality, conflict and change*. Bristol, UK; Buffalo, NY: Multilingual Matters.

LinguistList. 2018. *LLMap: LL-MAP: Language and Location–A Map Annotation Project*. Retrieved from www.llmap.org

Luebbering, Candice R., Korine N. Kolivras & Stephen P. Prisley. 2013. The lay of the language: surveying the cartographic characteristics of language maps. *Cartography and Geographic Information Science* 40(5). 383–400. DOI: doi.org/10.1080/15230406.2013.809892 (Accessed at 2018-08-09)

Mareüil, Philippe Boula de, Albert Rilliard & Frédéric Vernier. n.d. Atlas sonore des langues régionales de France. (atlas.limsi.fr) (Accessed 2018-08-10)

Moroz, George. 2017, May 16. rOpenSci | Easy linguistic mapping with lingtypology. (ropensci.org/blog/2017/05/16/lingtypology) (Accessed 2018-08-12)

Moseley, Christopher (ed.). 2010. *UNESCO Atlas of the World's Languages in danger* 3rd ed. Paris: UNESCO. Retrieved from www.unesco.org/languages-atlas (Accessed 2018-07-05)

Muturzikin. 2011. MuturZikin language maps. (www.muturzikin.com/countries.htm) (Accessed 2018-08-11)

Paterson III, Hugh. 2012, March 23. Types of Linguistic Maps: The Mapping of linguistic Features and Researcher Interactivity. Retrieved from hugh.thejourneyler.org/2012/types-of-linguistic-maps-the-mapping-of-linguistic-features (Accessed 2018-08-11)

Queixalós, Francisco & Odile Renault-Lescure (eds.). 2000. *As línguas amazônicas hoje*. São Paulo: Instituto Socio-Ambiental.

Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the World* 20th ed. Dallas, Texas: SIL International. Retrieved from www.ethnologue.com

Simons, Gary F. & Raymond G. Jr. Gordon. 2004. Ethnologue. In *Encyclopedia of Language and Linguistics*, 2nd ed. Elsevier. Retrieved from www-01.sil.org/~simonsg/preprint/ELL2%20Ethnologue.pdf