Counting the losses: numbers as the language of language endangerment

Robert E. Moore, Sari Pietikäinen and Jan Blommaert

Abstract
This paper provides a set of critical reflections on the use(s) of numbers to communicate facts about the changing dynamics of speech communities undergoing language shift. Such numerical representations are widespread, and they are of importance in segments of language expertise. After a literature survey of counting practices, the paper focuses on three language-ideological decisions underpinning language counting: First, decisions need to be made as to who counts as a speaker. In discussions of language endangerment, speaker counts are the most important single index of the endangered character of the language. Secondly, in order to count the number of distinct languages in a region, country, or world area, decisions must be made which privilege a notion of languages as bounded, closed, and geographically fixed entities. Finally, decisions need to be made with respect to the domains in which ‘small’, endangered, or minority languages continue to be used. From the discussion of domains we develop an alternative vision that centres not on distinct, named, countable languages, but on speakers and repertoires, and on the actual resources that speakers deploy in actual contexts. The contemporary situation of speakers of indigenous Sami, African and Native American languages will be drawn upon for examples.

Keywords: language endangerment and shift; minority and indigenous languages; Sámi; African and Native American languages; globalization

Affiliations
Robert E. Moore: Dublin City University, Ireland
email: rem10us@yahoo.com
Sari Pietikäinen: University of Jyväskylä, Finland
email: sari.p.pietikainen@jyu.fi
Jan Blommaert: University of Jyväskylä, Finland and the University of Tilburg, Netherlands
email: j.blommaert@uvt.nl
1 Introduction

The focus of this paper is on the uses of numbers to represent complex sociolinguistic states of affairs in local speech communities where processes of language shift, replacement, and endangerment are clearly taking place. This examination of practices of enumeration forms the basis for a broader methodological critique of the rhetoric of language endangerment. We summarise the practices whereby (a) the endangerment of named languages is represented by numerical counts of speakers (speaker counting); (b) the loss of linguistic diversity on a planetary scale is represented by estimates of the number of (named) languages on earth at various dates in the past, in the present, and in the (projected) future (language counting); (c) the relationship between actually existing Endangered Languages (ELs) and the other languages present in the local speech community is represented by the use of a simplified version of the concept of speech domains (Fishman 1991) – a familiar and very useful heuristic, but one that cannot be applied in a mechanistic fashion to language shift and death situations.

We argue not only that ‘speakerhood’ and ‘language-hood’ are matters whose complexity poorly suits them for numerical representations – a fact already acknowledged by many who publicise these same numbers – but that the use of such numbers, which continues unabated, privileges a conception of ‘languages’ as neatly-bounded, abstract, autonomous grammatical systems (each of which corresponds to a neatly-bounded ‘worldview’). This, we argue, inevitably draws attention away from the speech-community dynamics of language contact and change that we know to have been central to virtually every documented case of language shift or replacement. Such conceptions of language, we argue, tend to obscure the complex pragmatic and metapragmatic dimensions of actual language-in-use, and the totalising effects of such conceptions also draw the attention and energy of scholars and policymakers away from detailed considerations of the actual linguistic and communicative resources that people have and deploy in social life. Our critique, thus, calls for a fundamental shift in analytical gaze, away from abstract ‘languages’ and ‘speakers’ to practices and actual resources.

However faithful or not these numbers may be as representations of what is actually going on in speech communities, they have become one of the most significant and authoritative ways that expert knowledge about sociolinguistic change gets reported to policymakers and planners in State and other realms. They also underlie what we will call in this paper the Endangered Languages (EL) discourse – a discourse in which languages are often compared to endangered species on the road to extinction and are hence in need of expert intervention, protection and revitalisation (Mühlhäusler 1996, 1996).

Such counting practices, we argue, disregard the fact that all linguistic communities, including especially minoritised ones, are discursively and ideologically complex – hence the striking disconnection between this discursive apparatus and the concerns of researchers in linguistic anthropology, sociolinguistics, and related fields over the past thirty or so years (cf. Silverstein 1979, Gal 1979, Schieffelin, Woolard and Kroskrity 1998, Kroskrity 2000, Gal 2006, Nevins 2004, Kroskrity 2009). Yet the counting of languages and their speakers continues, and the numbers are continually being presented and (re)circulated even when the problems in counting are recognised.

The practice of counting languages and speakers permeates the way that complex sociolinguistic issues and situations are reported and discussed in public sphere institutions and in the mass media – a milieu increasingly obsessed with equating ‘figures’ and ‘facts’ – and continues to have a significant presence in scholarly literature, notwithstanding the more noticeable presence there of all manner of caveats, disclaimers, and ‘health warnings’ associated with such numerical presentations. The fact is that the counting practices discussed in this article have become a powerful discourse of truth on languages in society. The literature documenting the colonial ‘discovery’ and cataloguing of languages, for example, is replete with (sometimes heated) debates on the number of languages and the numbers of their speakers (see Fabian 1986, Van de Velde 1999; also Errington 2001 and Makoni and Pennycook 2007 for discussions). Such debates continue, most noticeably in EL discourses, but also in scholarly and political discourses on ‘minority’ languages and cultures, and, finally, in attempts to use econometric models to calculate the costs and benefits of societal multilingualism, for instance in the EU (Grin 2003).

Our goal is to call into question the ‘faith in numbers’ as neutral, non-ideological, precise, and objective (Porter 1995). It is precisely because such numericalised representations of languages and speakers carry implicit but nonetheless strong claims to objectivity and factuality that the ways in which such figures are obtained and established urgently need to be re-evaluated.

We are not the first to raise questions about the ideological underpinnings of quantification in discussions of language contraction and shift (see, e.g., Collins 1998, Meek 2007, Nevins 2004, Silverstein 1998, Duchene and Heller 2007, Kroskrity 2009). In her powerful article, Jane Hill (2002) identifies three themes that are ‘ubiquitous in expert rhetoric on language endangerment’, and that she claims ‘may inadvertently undermine its goals of advocacy’. The first two of these, which she calls ‘universal ownership’
and ‘hyperbolic valorization’, will be familiar to anyone who has surveyed the literature of language endangerment: ‘universal ownership’ asserts that ‘endangered languages in some sense “belong” to everyone in the world’; ‘hyperbolic valorization’ can be found in ‘locutions like “Endangered languages are priceless treasures”’ (Hill 2002:120). Hill’s third theme is ‘the theme of enumeration, which attempts to create a sense of crisis by the compilation and recitation of alarming statistics, such as those that show that over half of the world’s languages are endangered, or that some language has only three remaining speakers’ (120). According to Hill, enumeration ‘expresses a form of power that may amplify the alienation of endangered languages from the domain of quotidian practice of those who use them to the domain of esoteric expert knowledge’ (Hill 2002:120–121; see also Errington 2003, Whiteley 2003, Silverstein 1998, Kroskrity 2009).

On the other hand there is likewise nothing particularly new about the idea that complex sociolinguistic situations involving language change, language contact, multilingualism, and language shift can be rendered in numerical form. Greenberg, for example, made one notable attempt in the era of glottochronology and lexicostatistics (Greenberg 1956). Leaving Greenberg’s (many) numbers aside, we note his fundamental assumptions (and the caveats attached thereto):

Our general expectation, subject to significant qualifications, is that areas of high linguistic diversity will be those in which communication is poor, and that the increase of communication that goes with greater economic productivity and more extensive political organization will lead typically to the spread of a lingua franca, whether indigenous or imported, resulting in widespread bilingualism and the ultimate disappearance of all except a single dominant language. Measurements of language diversity may therefore be expected to show significant correlations with economic levels and with degrees of acculturation. (Greenberg 1956:110)

The assumption that linguistic diversity always represents a barrier to effective communication has been shown to animate much contemporary language policy in the EU (Gal 2006, Moore in press). More proximately to Greenberg, the same assumption, of course, was prominent in the thinking of Leonard Bloomfield (1933:47), who argued that boundaries between different language communities coincided with ‘lines of weakness in communication’. But the whole later enterprise of variationist sociolinguistics has showed that this is by no means necessarily true – that it is in situations of intimate contact between speakers of different linguistic varieties that minor variations (e.g., of pronunciation) can become heavily weighted with indexical value, as in Labov’s classic study of Martha’s Vineyard (Labov 1963). The deeper assumption here,
of course, is that the ‘normal’ speaker of any language is a monolingual. The fact that one can draw a straight line from Greenberg’s (1956) proposals to the contemporary discourse of language endangerment – without passing through either the literature of variationist sociolinguistics or the literature of the ‘ethnography of speaking’ (e.g., Hymes 1962, Bauman and Sherzer 1974) – exemplifies the curious disconnection of the EL literature from contemporary concerns in linguistic anthropology and related fields.

The moment we call into question the counting practices that provide an ‘objective’ grounding for EL discourses, various critical questions emerge, which we hope will be addressed by future researchers:

- How does the EL discourse – in which numbers play a central role in a representational apparatus designed to frame the phenomenon of language endangerment as a global crisis – fit into broader epistemic frameworks of ‘reflexive modernisation’ (Beck 1994) in which expert discourses of quantifiable risk loom large (e.g., global climate change, infectious diseases like HIV/AIDS, global finance capital, genetically modified crops, etc.)?

- How have new institutions and social networks emerged around the issue of ‘endangered languages’, so defined, in local speech communities, in academic linguistics, in the NGO sphere, and elsewhere (see e.g. Collins 1998, Nevins 2004, Duchene and Heller 2007, Hill 2002, Kroskrity 2009)?

- How does the discourse of endangerment summon or interpellate these differently-positioned constituencies or ‘publics’ (Warner 2002) in particular ways?

- How does the EL discourse – its claims (and/or aspirations) to be ‘non-political’ notwithstanding – fit into a politics of ‘indigeneity’ and minoritisation within regions, States, and larger institutional formations (e.g., the European Union) (cf. Nevins 2004, Silverstein 1998)?

- Are the sociolinguistic processes that give rise to ‘new’ languages (e.g., pidgins, creoles, and so-called ‘mixed languages’) connected in any way to those that bring about the ‘extinction’ of ‘old’ ones (Silverstein 1998: 409; cf. DeGraff 2003, Mufwene 2001, 2003, 2007, Chaudenson 2001)?

In the discussion below we examine the language-ideological assumptions that underlie and license a diverse set of counting practices, upon which the discourse of minority languages and language endangerment is based. We argue that this curious calculus of ‘languages’ and ‘speakers’ revolves around three critical issues: one about the nature of the speaker, one about the nature of language, and one about the nature of domains. Having laid out the fundamental issues, in Section 2 we survey the different practices of counting, before
moving on to discuss the three critical issues identified above: the counting of speakers (Section 3), of languages (Section 4) and of domains (Section 5). In the conclusion (Section 6), we briefly explore the implications of this for ongoing relationships between linguistic experts and professionals in the mass media, and with the governmental and other policy professionals who use the numbers provided to them by these experts as evidence (‘facts on the ground’) to support specific policy objectives and programmes.

2 The curious calculus of language counting

It’s a fact: language endangerment has been a growth industry in journalism and in book publishing over the past ten or fifteen years. A Lexis-Nexis search of the English-language press (carried out on 15 June 2009) shows that the phrase ‘endangered languages’ has appeared in 1,552 separate articles, most of them published since the turn of the last century – results below in Figure 1.

![Figure 1: Articles in English-language press on 'endangered languages'](source: Nexis UK 15 June 2009)
If trends for 2009 continue, the frequency of such articles in the English-language press will approach one per day. Not surprisingly, a kind of template has emerged in feature stories on endangered languages, and no doubt its robustness has helped to support this remarkable rate of production:

When [name of last speaker] dies, an entire (ancient, etc.) universe will disappear forever from the face of the earth.

[Name] is the last speaker of [language name].

In [(past) date], there were [number] languages spoken (in [community > region > nation > continent > planet]).

Today only [number] remain. Linguists estimate that by [(future) date], there will only be [number] left.

Linguists are scrambling to record (preserve, document) [language name] before it finally disappears.

   [a. value to science (‘us’)]
   [b. value to ethnolinguistic community (‘them’)]

Several different kinds of counting activity are taking place; the most significant of these are summarised below.

Modes of enumeration in the EL discourse

The forms of enumeration catalogued below are all present in the EL discourse and in the policy and planning discourses that accompany the politics and business of minority recognition and rights.

1) The counting of (named) languages, using absolute or cardinal numbers to represent sum totals of enumerable individuals of the same kind (e.g., there are 6,912 languages in the world, according to the current [Fifteenth] edition of the Ethnologue)

   a. within some delimited territory

      i. regions (e.g., languages of the Balkans, of the Circumpolar zone)

      ii. nation-states, cities, polities, and other communities (e.g., languages of China, of Dublin, of the EU, of the Islamic World)

      iii. continents, hemispheres, planets (e.g., languages of Africa, of the Western Hemisphere, of the World)
b. and/or at some time-period or point(s) in time, figured calendrically (e.g., the number of languages in North America in 1492 AD [versus 2000 AD]; one [language] ‘dies’ every two weeks)

c. and/or within larger groupings based on phylogenetic (Stammbaum) models of language classification (e.g., there are 4 living languages that belong to the Caddoan language family, according to Ethnologue).

2) For each named language, the counting of (fluent? native?) speakers
   a. using absolute or cardinal numbers to represent sum totals
   b. using relative or ordinal numbers (only as last, e.g., the last speaker of Catawba).

3) For each plurilingual speaker, the enumeration or listing of the (finite) inventory of distinct languages or linguistic varieties under the control of the speaker
   a. using absolute or cardinal numbers to represent sum totals (e.g., So-and-so speaks five languages)
   b. using relative or ordinal numbers to represent the chronological order in which linguistic varieties are acquired over the life-course of the individual speaker (L1, L2, etc. – i.e., ‘first language’, ‘second language’, etc.).

4) For each multilingual speech community, the enumeration or quantitative assessment of the (finite) inventory of domains of usage where the variety in question is present or absent (e.g., at home, at school, in public places, etc.; cf. Fishman 1991).

5) The use of (cardinal) numbers to symbolise scalar values along some independently determined parameters – including the numbers generated from 1–4 above, as thresholds – used as criteria for assessing states of ‘vitality’ and ‘endangerment’ of (named) languages. The UNESCO Ad Hoc Expert Group on Endangered Languages, for example, assigns values from 0–5, in the manner of a Likert scale, for each of five such criteria or Factors; hence, a score of 5 for Factor 1, ‘Intergenerational Language Transmission’, is assigned when ‘The language is spoken by all generations’ with ‘no sign of linguistic threat from any other language’; any language scoring 5 is assessed as ‘Safe’ on this criterion; at the other end of the scale, a score of 0 – ‘there is no one who can speak or remember the language’ – qualifies a language as ‘Extinct’ (UNESCO 2003: 7–8; emphasis in original). Readers are sternly enjoined not to add these
numbers together to arrive at an overall ‘score’ for a language across domains, but it is impossible not to notice that the only languages that score 5 out of a possible 5 on all domain indicators are the major ‘world languages’ (English, Spanish, Chinese, Hindi, etc.).

The listing provided – summarising a truly massive amount of literature – suggests how deeply saturated with numbers is the mainstream discourse of Endangered Languages (ELs). The different counting practices listed above cross-cut and can either undermine or buttress each other in complex ways. There are also multiply interlocking and interdependent relationships between these diverse practices of (e)numeration in the construction of linguistic varieties as EL, the way that one practice of enumeration creates (or destroys) the conditions of applicability of another.

But there are also hidden dependencies: in order to be counted in the first place (whether as ‘a language’ or ‘a dialect’), a linguistic variety must first have a name. The mere fact that a linguistic variety has a name – usually more than one, of course, with distinct (if overlapping) groups of users (speakers as well as non-speakers of the variety in question) using distinct names for ‘the same’ language – is itself obviously and irreducibly a matter of the sociocultural and political dynamics of speech communities, social groups, and populations. The identification of a linguistic variety with a name (a glossonym, whether ‘found’ in the community or imposed through re-christening by the analyst) is what creates the conditions under which the counting of languages becomes possible, and often takes place simultaneously with status and language planning. An important question to explore is the degree to which linguistic varieties, once named and counted, are brought within the conceptual orbit of standardisation, and if so, how.

Once (re-) christened, Language X becomes countable (along with Languages W, Y, and Z); and once Language X has been counted, its speakers become countable, precisely as speakers of Language X, regardless of whatever other varieties they may speak and/or understand. Once counted as ‘speakers of X’, the group can now be internally differentiated, again on a numerical basis, as monolingual speakers of X, or as speakers bilingual or plurilingual in X and other varieties. Now the group of bi- and plurilingual speakers of X can itself be internally differentiated, again on a numerical basis, but in a different way and using a different type of enumeration: bi- and plurilingual speakers can be distinguished (and hierarchically ranked) according to whether they speak X as a first language (L1), a second language (L2), or an nth language. In this way, the use of ordinal numbers allows for the elements of a multilingual speaker’s repertoire to be hierarchically ranked (putatively according to the chronological order in which they were acquired by the speaker); this in turn produces the...
basis for a hierarchical ranking of the speakers themselves, as ‘First-language’ (L1) speakers of X, ‘Second-language (L2) speakers of X’, and so forth.

Let us note one final point with respect to this dense numerical apparatus – a kind of balance-sheet representation of Babel. The languages that are subjected to these regimes of numericalisation are often seen as tied inextricably to a particular region. Mobility of people (i.e., speakers) can make them and their languages more difficult to count, and indeed, diasporic groups often are seen as outside the concerns of language endangerment. In another context, they may still be counted and included in the group of speakers. More will be said about this below. Here the point is that counting the speakers of endangered languages is often about counting people in a particular region. The remainder of this paper will address the three critical issues that underlie this calculus. In order to arrive at decisions about numbers of languages, decisions need to be made with regard to ‘speakers’, to ‘language’ itself, and to the domains in which (a) language is effectively used.

3 Who counts as a speaker?

Theoretically inclined sociolinguists and linguistic anthropologists may believe that they have long since moved well beyond the ‘native speaker’ as a primordial given in analysis (Davies 2003, Singh 1998; Fraurud and Boyd 2006). The fact is, however, that such commonsense notions are alive and well in the policy and planning discourses that accompany the politics and business of minority recognition and rights. Many in minority language communities, moreover, appear to subscribe to them as well – often for good strategic and instrumental reasons, even though the ambiguities and tensions bound up in speakerhood are part of their daily experience.

If we briefly look at the case of the Finnish Sami, for instance, we see that officially, three languages are identified: Northern Sami, Inari Sami and Skolt Sami. One question to ask could be: how many monolingual speakers of the three Sami languages are there? The answer, it seems to be, is none, even if figures in the Saami Cultural Encyclopaedia (Kulonen, Seurujärvi-Kari and Pulkkinen 2005) state approximately 35,000, 400 and 400 speakers for the three languages, respectively. The languages exist only as parts of a multilingual repertoires in which Finnish (and increasingly English) has an important place. Does this mean, then, that these Sami languages are dead? Obviously not – not from a sociolinguistic perspective, and even less from the perspective of people in contemporary, multilingual Sami communities. The insertion of the Sami languages in multilingual repertoires of course guarantees them a continued existence (and a raison d’être), but it also makes decisions about ‘native speakerhood’ rather problematic, and evidently causes all sorts of
difficulties when it comes to calculating the number of speakers. In addition, describing such people as users of multilingual repertoires begs the question as to what exactly these repertoires are composed of, and of whether ‘language’ as traditionally understood in terms of (countable) L1, L2, … Ln is the best possible denominator for this.

The hegemonic character of speaker enumerability is such that practices of ‘undercounting’ and ‘overcounting’ almost inevitably become key resources for instrumental and strategic language planning. Linguistic varieties whose speakers seek recognition within state or other encompassing units may benefit from stating as high a number of speakers as possible, so as to demonstrate the continued presence and vitality of a ‘community’; for a linguistic variety that is already recognised as endangered, however, lower numbers may at times be more beneficial in making the case for the urgency of support (cf. Friedman 2003: 6–7, Urla 1993: 820). A dramatic decrease in the number of speakers over a relatively brief and recent period is the argument for qualifying a language as ‘endangered’; an increase in the number of speakers can be seen as evidence that revitalisation efforts have caught on. In any event, the deployment of numbers of speakers inevitably conjures up an image of the speaker as a stable – hence, countable – entity; it obscures the actual elasticity of speakerhood in real sociolinguistic life, even if that elasticity may come back with a vengeance whenever new counts have to be made. Then rather predictable questions and controversies arise: Do we count people who speak the language sometimes, or only people who use it all the time? Do people who are learning the language at a later age, including as adults, count? What do we do with people who grew up bilingually and can be said to belong to language group X as well as language group Y?

Along with this countable notion of the speaker comes a package of suggested competences and skills that is presented as ‘ideal’ and ‘complete’, and against which people’s actual language skills can be perpetually measured. A ‘native’ or ‘mother-tongue’ (L1) speaker is assumed to be someone who possesses all the resources of the language. Unsurprisingly, most actual performances fall short of that ideal. In communities where an ‘endangered’ or ‘minority’ variety is spoken, being a ‘speaker’ is almost always a matter of degree: some speak ‘more’ of the language, others ‘less’ (cf. the notion of ‘semispeaker’ developed by Dorian, e.g. 1977, 1981). In fact, observations like these could be made for all languages, language groups and speech communities. Such rather obvious sociolinguistic observations, however, do not sit comfortably with the demands of counting practices, and difficult compromises have often been made.

The use of surveys and questionnaires to arrive at speaker numbers only adds another layer of mediated complexity to an already ambiguous and
fraught situation. Anyone who has worked in a community with one or more obsolescent languages has frequently encountered widespread insecurity in self-reporting on language skills and ‘native speakerhood’. For interesting – and relevant – reasons, people are often loath to put themselves forward as ‘good’ or ‘native’ speakers of the languages, often preferring to attach caveats and disclaimers to their attestations, especially by way of answering the importunings of a visiting researcher: ‘I don’t speak the language well’, ‘I can’t write it very well’, ‘I am not a real speaker of the language’, ‘I’m not good at storytelling’, and so forth (Moore 1988). Such disclaimers and other hedges on one’s own willingness or comfort in inhabiting the authority of a (full, fluent, native, expert) speaker are not mere ‘noise’ to be filtered out of our calculations; they, together (minimally) with relevant facts about the usages displayed by such speakers, are essential data of the most perishable sort (Hill 1983, Moore 1988; see Hymes 1981 [1971] for a classic treatment).

The model for the ideal (full, fluent, perhaps monolingual) speaker in the business of counting, then, seems increasingly to be a pre-language-contact speaker – just the kind of person who, unsurprisingly, is very hard to find in actually existing speech communities today. It is not an identity that is readily volunteered in self-reporting practices; it is very much a category imposed by observers and, at other times, by state or supranational organisations and language planners. Speakers who fit the description are very rare, just at a time when their symbolic presence would be most needed. To the extent that the speech of previous generations – e.g., of those who were first consulted by missionaries, anthropologists, or other observers – is taken to be the ‘authentic’ variety of a language, and is museologically recorded and preserved; it becomes an archaic variety, lifted to a level somewhat above that of a national Standard language, but (re-)made in its image (Gal 2006: 170–171; Moore in press).

Under these circumstances it is understandable that people in minority and ‘endangered’ language communities today often rely on collaborative efforts in accomplishing demanding linguistic tasks. Thus, a Sami journalist had a small network of older speakers of the language in which the broadcasting was done, and whenever the journalist was in doubt about the accuracy and adequacy of certain expressions, she would call one or more of these speakers for language advice. The journalist’s public virtuosity in the language (an effect of this collaborative work), however, soon placed her in a position in which other people would call on her for language support – a task which she felt exceeded her own competences (cf. Pietikäinen 2008).

In the gap between majoritarian views of full linguistic competence and the mixed and heterogeneous quality of actual communicative practice in multilingual communities, speakers’ own usages can often emerge as defective,
especially in ‘minority language’ communities located in nation-states with an official standard language and a vigorous culture of standardisation (as the US, UK, and parts of Europe and Asia; Silverstein 1987). The presence in the community of well-intentioned outsiders working as researchers can have the effect of throwing an awkward spotlight on the different fluencies of remaining speakers and semispeakers, some of whom may even come to view themselves as Bloomfield (1927) memorably characterised the Menomini man named White Thunder: as someone ‘who could be said to speak no language tolerably’. In a context where being ‘a speaker’ of a particular language has profound sociopolitical and economic impact, such feelings of deficiency often can be turned into feelings of being an outsider to a particular language community, of failing in the task of being a ‘good’ minority member, or of a lack of voice, agency and ownership in that language.

We begin to see a wide range of implications here: practices of counting affect issues of group membership, legitimacy of identities, distribution of resources, questions of competence, ownership and voice, of inequality and variation in language. Reconceptualising the issues in the way suggested here cuts across such familiar questions as how many people of such-and-such ethnic – even linguistic – self-identification turn up in census counts administered at ten-year intervals; it also cross-cuts the kinds of questions asked by a traditional sociology of language (e.g., Fishman 1991), where the number (and circulation) of minority-language newspapers, or the number of radio and TV broadcast hours per week in the language become diagnostic indicators of the relative ‘health’ or ‘vitality’ of minority languages (Extra and Gorter 2001). In the end these, too, are lenses that majoritarian societies use to search for – and usually not find – the spaces where erstwhile ‘minority’ linguistic forms and practices actually ‘live’ (and ‘die’).

As mentioned earlier, social and geographic mobility makes ‘native speakers’ more difficult to count. Diasporic groups, then, pose problems. Often enough, diasporised speakers are simply absent from the discourse of language endangerment (as, with a few exceptions, are lingua francas, pidgins, creoles, and so called ‘mixed languages’). Diasporic speakers may still be counted and included in the group of speakers, though, as when indigenous groups try to follow the frequency of hits of their websites and arrive at numbers that exceed the population count in their ethnolinguistic homeland. In the case of the Finnish Sami, for instance, a recent official study, using figures provided by the Finnish Sami Parliament (particularly their voting register) and by the Finland National Census, arrived at a number of 9,350 Sami community members, 3,577 of whom lived in the Sami domicile area in Northern Finland (Länsman 2008). Even if only the Sami who live in the Sami domicile area have full rights for their languages, all of the 9,350 may potentially be seen as
targets of language revitalisation efforts, the number of Sami language speakers presently in Finland having been estimated at around 2,300.

The fact is, of course, that within any given speech community – especially a plurilingual speech community in which one or more varieties is obsolescent or falling out of use – the identification of speakers is always politically and sociolinguistically fraught. The fact that speakerhood is usually a matter of degree makes a digital ‘yes/no’ computation simply inappropriate. Indeed, researchers with long-term knowledge of local speech communities often make fine-grained distinctions between different types of speakers (see Dorian 1977, 1980, 1981, Moore 1988, Hill 1993). Not surprisingly, such fine distinctions rarely survive intact in more public formulations of the EL phenomenon. How such more ‘nuanced’ understandings of speech community dynamics might effectively be communicated to media professionals and policymakers is addressed briefly in the conclusion below.

4 What counts as a language? Purity, authenticity and (im-)mobility

The second critical issue in the arithmetic of languages is that of language itself: which utterances count as samples of a language, and which do not?

Numericalisation is a particular discourse of truth in the context of endangered and minority languages, and, as noted earlier, this occurs at two levels: (i) the counting of (named) languages on the one hand, and (ii) the counting of the speakers of each named language on the other. Through numericalisation, we enter a world of ‘objective’ figures and percentages and of statements in which relative ‘more’ and ‘less’ clauses are vital. The whole concept of a ‘minority’ language, of course, hinges on its numerical position vis-à-vis ‘majority’ languages; it rests on assumptions of enumerability and numerical ranking (Makoni and Pennycook 2007). Consequently, counting is a crucial activity in defining the languages of a community, region, state, or continent into existence or disappearance, particularly to the outside world. Counting is what renders these linguistic varieties and their speakers visible to encompassing sociopolitical and economic orders and powers, and hence it is also what renders them tractable to various administrative, language-planning, and other expert interventions.

Thus, counting three Finnish Sami languages creates an ‘essential’ existence for three communities of speakers (whose ethnonym and glossonym coincide: the languages generate the Northern Sami, Inari Sami and Skolt Sami). Different orthographies corroborate the separateness of the languages and their communities of speakers, and institutional language regimes further cement the existence of these units: the three groups have an official status in the Sami domicile area in Finland, their speakers derive rights from that status,
and the three languages are official languages of business in the Sami media, education and political institutions (cf. Kulonen et al. 2005).

The visibility is that of a series of neatly ordered and clearly labelled ‘languages’ that can be compared to one another in ‘size’ by virtue of counting the number of speakers of each; what is made invisible is the effective multilingual repertoire of people and the far more complex relationship between languages and their speakers than the dichotomies of L1–L2, dead–alive, native speaker–non (native) speaker suggest. This duality is the main principle organising the arithmetic we examine here. To the extent that these practices of enumeration effectively cause the multiplexity of people’s day-to-day communicative practices in such communities to become invisible, they hamper our ability to understand how it is that language change (a constant, and universal condition of linguistic existence) can become a ‘change of language’ (community-wide language shift). Given that many of the communicative practices that have continuing vitality and importance in minority language communities will tend to appear ‘marginal’ and ‘deviant’ when compared (invidiously) with those of a Standard language, it is not surprising that they have not always been visible to the state (or other ‘central’ authorities), precisely because they have thrived at the margins of institutional, geographic, and other orders.

Thus, for instance, the Sami Hip hop artist Amoc now produces music in Inari Sami, creating new registers and artistic representations of the language, albeit within the conventions of the globalised, ‘un-traditional’ and hence culturally hybrid speech genre of rap (Pietikäinen 2008). He thus effectively contributes to the vitality of the language, but he does so outside a framework of state support and in a framework of globalised commodification. Interestingly, Amoc’s status as a ‘native speaker’ would be unclear in traditional terms. He speaks some Inari Sami but his competence in the language is by no means ‘complete’, making his public performances invariably multilingual, with Inari Sami, Finnish and English being deployed. Similar to the Sami journalist mentioned earlier, he, too, relies on help of older speakers of the language for crafting his lyrics (Leppänen and Pietikäinen 2010). The case of Amoc raises complex issues of language ‘quality’ in the context of language revitalisation. Samiland, like many other minority communities, has always been plurilingual and diverse, and ‘pure’ Sami – or any other language – is and has always been an exceptional language variety, a register of the language but not ‘the whole’ language. As we have seen, crafting Inari Sami lyrics involves, in Amoc’s case, collaboration with other speakers whose repertoire in the language is more elaborate or more ‘complete’ in areas other than his. This work of language ‘purification’ is complex and demanding, and the result is, in terms of Amoc’s own repertoire, ‘special’ and ‘exceptional’
in its purity. The demands of Hiphop battling or freestyle, however, require Amoc to use Finnish along with Inari Sami, as well as, at times, English. Improvisation – a task for which he cannot call on the collaboration of other more virtuoso speakers – pushes him back, so to speak, to a mixed, hybrid and multilingual repertoire, probably closer to ‘normal’ speech and demanding less effort than the ‘purified’ Inari Sami lyrics of his recorded songs. The point is that, in Amoc’s repertoire of languages and styles, ‘pure’ Inari Sami appears as a specialised register, the highly stylised manifestation of a language which for Amoc and others manifests itself in daily life mostly in a ‘mixed’ form, alongside Finnish and English (Pietikäinen 2008).

Yet, in many language endangerment discourses, it is one or another variant of this idea of ‘pure’, ‘unmixed’ language that is central. Knowledge of it is seen as the key to revitalisation, and deviation from it (by using ‘hybrid’ and ‘mixed’ forms of it) is seen as evidence of endangerment. So the issue is not just about what counts as a speaker, but even more about what counts as a language. Fluent monolingual speech, preferably employing ‘classical’ grammatical features and talking about ‘traditional’ topics, is the iconic form of language in this respect. This explains why the researcher’s ideal informant is often taken to be an old man or woman, representing the pre-language shift generation and speaking what is believed to be the ‘original’ form of the language. It also explains why in language revitalisation efforts, people often try to involve grandparents as language teachers, believing that the teaching to the young of ‘old’ forms of language – by people who are themselves old – will mean that the ‘real’ language is taught. Hence also the emphasis in many language revitalisation efforts on highly specific cultural speech forms such as folk song, epic or traditional poetry – genres that are not only believed to articulate the deeper and more lasting cultural values of a group, but also its original and authentic language.

This model of language is an ideological construct, one that blends images of culture-as-immutable-authenticity with images of language as a pure denotational and structured – artefactualised – object (cf. e.g., Blommaert 2008, Collins 1998, Nevins 2004, Silverstein 1998, Kroskrity 2009). We saw above how a wide range of numerical metaphors is used to define Endangered Languages (ELs) – and the traditional ecological knowledge (TEK) seen as encoded in them – largely and essentially in terms of immobility and strict locality. We can now add that it also defines the ‘real’ language in terms that claim to be ahistorical, but in fact are rooted in particular moments of linguistic contact-history. The notion that the ‘ideal’ form of (a) language is a form from (a specific period in) the past, interestingly enough, can easily become the notion that re-imagines language use as the transmission of a timeless and essential cultural authenticity.
5 Language in action: domains and registers

In the actual practice of language counting, the question of what counts as a ‘domain’ (Fishman 1991) becomes very important, because the ‘presence’ of a language in certain social domains (e.g. schools, church services, administrative encounters) is the raw evidence for seeing that language ‘in use’ and, hence, alive. As we have shown, language counting almost never rests on evidence derived from the careful observation of actual language use in actual speech environments. Such evaluations of language vitality, then, invite the reader to imagine that the ideal speaker of an endangered language is someone who effectively uses a ‘pure’ form of the language in such-and-such a domain (or even better, across all domains).

It is obvious, for example, that language revitalisation programmes are themselves creating new domains of use. The way that these new domains often engage with new media and technology has become an important area of ethnographic investigation (see Eisenlohr 2004; cf. Ahlers 2006, Nevins 2004). In many language revitalisation programmes children learn the target language as a register (and not as a language): what they learn is often a domain- and medium-specific set of resources. They might learn a quasi-standardised ‘school variety’ of the language, very often oriented towards the generative competence to reproduce grammatical correctness in utterances; they might also learn specific speech genres that are strongly tied up, on the one hand, with traditional classroom genres, and with genres of cultural reproduction on the other hand (folk songs, poetry, etc.). Others might participate in immersive and/or ‘native-language-only’ activities, as seen in Maori ‘language nests’ (Spolsky 2003) and in the Master-Apprentice method pioneered in California Native communities (Hinton and Hale 2001).

Given the absence of active language proficiency in their parents’ generation, the home environment often involves shifting between languages – the production of a non-normative, ‘sub-standard’ hybrid variety which may not be accepted as ‘the’ language. The full performance of the language resources learned at school is sometimes reserved for school practices and, outside it, for moments of cultural performance such as, e.g. cultural events in the community, festivals or other occasions organised around the performance and display of recognised and emblematic cultural forms (see Ahlers 2006, Dinwoodie 1998, Silverstein 2003).

This might not sound like good news for those concerned about endangered languages and the preservation of biolinguistic diversity in its pure, prelapsarian form. But, to the extent that this view imagines ELs as embedded in hybrid, plurilingual practices rather than in static cultures of unbroken transmission, it may be good news.
Yohani Msanjila (2004) examined the plurilingual patterns of language usage in a very marginal Tanzanian community of Kisafwa speakers. Kisafwa is a language which, according to the counting logic of *Ethnologue*, has 150,000 speakers, but only a few thousand in Tanzania. It is dominated by Kiswahili (the national language in Tanzania) and English, and several of its neighbouring languages would be significantly bigger and carry significantly more prestige than Kisafwa. Msanjila examined the way in which people across generations divided their language repertoires over specific domains, and he found that almost all speakers, including young members of the community, maintained Kisafwa as the medium for intimate in-group communication. In spite of the overwhelming dominance of Kiswahili in formal and prestige domains, Kisafwa was not endangered – at least when we accept that Kisafwa had been ‘shrunk’ (or just never expanded), so to speak, from an all-purpose language to a language deployed in specific registers only, those of intimate in-group communication (see also Dyers 2008). In the case of Kisafwa, the reduction of the language to a restricted set of registers, operating in a plurilingual repertoire, provided the conditions for the continued existence of the language; in this case, the ‘shrinking’ of the language (in terms of domains) and its insertion in multilingual practices safeguarded the survival of the language. The fact that the language became part of multilingual repertoires made sure the language continued to be used by its speakers. In a scenario such as this one, probably no one would be counted as a monolingual speaker of the language, because in order to use the language, other languages needed to be used as well (notably Kiswahili). So in order to be a Kisafwa speaker, one also needed to be a Kiswahili speaker, and Kisafwa speaker identity to a considerable extent dependent on being a speaker of Kiswahili too. Such findings may sound counter-intuitive, and they surely prompt a different vocabulary for talking about language endangerment and revitalisation. But they probably are a better reflection of actual sociolinguistic life than suggestions about L1 speakers speaking all of a language perfectly in all spheres of life – until, that is, they draw its last breath. They also mean that we need to give up the comfort of the countability of languages, because this is no longer about languages but about far more specifically defined linguistic resources.

6 Conclusions

We hope to have shown that practices of counting in the field of language endangerment rest on three critical assumptions – one about the nature of the speaker, one about the nature of language, and one about the nature of domains.
numbers as the language of language endangerment – not one of which holds up under scrutiny. We also hope to have shown how methodological reflection on each of these questions dislodges the stability of language counting practices and brings their results out of balance, and how specifically reflection on domains may lead us into new ways of thinking about language endangerment and revitalisation. To the extent that such issues are the object (or should be the object) of any sound sociolinguistic effort, such reflections need to be made. The fact that many linguistic anthropologists and sociolinguists fully recognise the complexity of the concepts of ‘speaker’ and ‘language’ – and do not subscribe to the more simplistic versions of these concepts discussed here – only underscores the urgency of the point being made.

Nonetheless, governments, international organisations and action groups are keen on figures and statistics, and if figures provide a strong argument in favour of the preservation of unique cultural heritage, then they must be used. This leaves scholars who seek to communicate their findings in terms that will be accessible to mass audiences, and understandable and relevant to the concerns of policymakers, in something of a bind.

How can such complex understandings be effectively conveyed to the several specialised ‘constituencies’ involved in the problem of language endangerment, from policymakers to minority language activists to researchers to governmental institutions – and to the mass media, where the bulk of the simplistic discussion of endangered languages appears? How can interlocutors in these constituencies make good use of the slightly more nuanced notions of ‘language’ and ‘speaker’ that scholars recognise and that we embrace in this paper? These are the kinds of questions we hope our reflections have prompted in the minds of readers. Our main findings here are three:

1) Numbers are poorly suited to represent the kinds of facts known to be of relevance to processes of language shift, replacement and endangerment – as everyone knows, and as we have shown again here. This is partly because the kinds of questions that numbers are the answer to – of how many ‘languages’ there are (in some place) and, for any given (named) language, of how many ‘speakers’ there are – are just not the interesting questions to ask if one wants to know how the endangerment and eventual loss of languages actually comes about. Despite this,

2) quantification has been used as a legitimating device by States as well as by activists working on behalf of minority language communities (see the now classic discussion of statistics in Basque language politics in Urla 1993) – though never without creating or exposing new social cleavages within the minority-language community, cleavages that are precisely those created by language standardisation projects everywhere: ‘Once
certain forms are chosen through language planning as the standard ones, some speakers/users of minority forms become doubly stigmatised’, writes Gal, since ‘the speech of minority speakers whose linguistic forms are not included in the new minority standard comes to seem inadequate, and perhaps even inauthentic’ (Gal 2006: 170–171). When it comes to languages and speakers, only the powerful can afford to imagine that counting is a ‘value-neutral’ rational-technical ‘procedure’ that can be detached from and reapplied to various situations at will, producing relatively more or relatively less ‘accurate’ results. Counting – of speakers or of languages – is itself a speech act, one whose performativity emerges over time, in the processes by which the numbers are produced, circulated, and, perhaps eventually taken as a Foucauldian discourse of truth – with very real consequences for communities of speakers.

3) Insofar as it merges on one side with the discourse of biodiversity (via the mistaken ‘language = species’ metaphor), and on the other with the discourse of minority rights, the ELs discourse actually depends on the use of numbers. Numbers are not incidental to the vast literature surveyed here; numbers are the very language of the ‘language endangerment’ discourse. The contradiction between this and the recognition that numbers cannot adequately represent the complex dynamics of plurilingual speech communities (see the first point above) at the very least needs to be acknowledged and addressed within the relevant disciplines of linguistic anthropology, sociolinguistics, and applied linguistics and language planning.

In the nineteenth century, Porter reports, ‘most statistical discussions proceeded on an assumption of transparency: that the numbers mean what they seem to mean, and can be analysed without worrying too much about the conditions of their production’ (Porter 2006: 1278). ‘Numbers’, Porter writes, are the quintessential form of information, which … ideally, requires little interpretation, and so presumes no special training, no extensive experience of a thing from the inside, but is immediately available to almost anyone for do-it-yourself use. Information presumes a world of self-similar objects, and of inhabitants who are familiar with them, as in those old American IQ tests which attributed feeblemindedness to anyone unfamiliar with Crisco, pocketknives, and golf. Information must travel readily. (Porter 2006: 1279)

Our argument here has been that languages, if that is even the best term, do not exist in a world of ‘self-similar objects’, and moreover that the people centrally
concerned with them are ‘familiar’ with them in radically different ways, and that it is to just this kind of disjunction and difference that scholars should seek to be accountable when communicating their findings to policymakers and the wider public. In any case, this is the political field into which sociolinguists, language planners, and others enter as players every time they furnish numbers to government officials, NGOs, or the media.

There is a wider dimension to this. Language counting practices belong to a language-ideological complex which we can call artefactualisation (Blommaert 2008). They sit there alongside other widespread and deep-rooted linguistic practices such as the production of grammars and dictionaries or the development of practical language pedagogies. All of these practices share the fundamental premise that language can be turned into an easy-to-handle, closed and finite artefact, something that can be neatly separated from other similarly artefactualised languages and can thus be listed, catalogued and ordered within such universalising constructs such as the Stammbaum model. Rendering languages and their speakers countable represents the completion of artefactualisation: when language is effectively artefactualised, it can also be quantified, counted, ranked according to ‘size’.

Of course, all processes of language artefactualisation share the same problem: the fact that a closed, finite and bounded image of language sits very uncomfortably with the actual, highly flexible and dynamic realities of language-in-use. Artefactual images of languages, like the rather crude calculus that underpins them, consequently inform us very little about the actual communicative practices of real people in a real world. To the extent that these numerical and other artefactual representations of language are widely used as baseline ‘facts’ by practitioners in the popular, scholarly, and policy realms, there is a problem with the foundations of sociolinguistics.

About the authors

Robert E. Moore has engaged in ethnographic and linguistic fieldwork in Indian reservation communities in the northwestern US, and is the author of articles about language shift and endangerment, the semiotics of branding, and the politics of accent in contemporary Irish English, among other topics.

Sari Pietikäinen is a professor in Discourse Studies at the University of Jyväskylä. Her research interest includes language endangerment, peripheral multilingualism and indigenous Sámi community. She has published on these topics and is currently leading an international research project on Northern Multilingualism.

Jan Blommaert is Finland Distinguished Professor of Linguistic Anthropology at the University of Jyväskylä, Finland, and Professor of Language, Culture and
Globalisation at Tilburg University, The Netherlands. He is also Professor of African Linguistics and Sociolinguistics at Ghent University. He is the author of *Discourse: A critical introduction* (Cambridge University Press 2005), *Grassroots Literacy* (Routledge 2008) and *The Sociolinguistics of Globalization* (Cambridge University Press 2010), and the editor of *Language Ideological Debates* (Mouton de Gruyter 1999).

Notes

1 This article was developed in the context of the Northern Multilingualism project of the University of Jyväskylä, directed by Sari Pietikäinen and funded by the Academy of Finland, whose support is hereby gratefully acknowledged. Jan Blommaert’s participation was also made possible by the FiDiPro project of the University of Jyväskylä, funded by the Academy of Finland. Robert Moore wishes to thank the Centre for Translation and Textual Studies at Dublin City University and its Director, Professor Michael Cronin, for research support and valuable interlocution. We are grateful to Helen Kelly-Holmes, Max Spotti, Thomas P. Strong, and Michael Silverstein for very useful comments on a preliminary version of this paper. The current version has benefited from the comments of two anonymous referees.

2 Ammon – to take only one example – after giving speaker numbers for ‘world languages’ (English, Chinese, German), points out that ‘All the figures have of course to be taken with caution in the face of the well-known methodological problems of counting the speakers of a language’ (Ammon, Dittmar, Mattheier and Trudgill 2005:1527). Interestingly, Ammon provides no bibliographic references to discussions of these methodological problems, perhaps because the problems are common knowledge.


4 And consider all the many cases, especially in Native North America since the 1970s, in which languages and/or groups have been re-named, so that now we speak not of Pima-Papago but of Tohono O’odham, not of Kwakiutl but of Kwakwaakaa’wakw; the name of the largest reservation in Washington State was recently revised by the Tribal Council so that its spelling (Yakama, formerly Yakima) would conform to the spelling used by the (non-Indian) drafters of the 1855 Treaty that established the reservation.

5 Recall the notions of elaborated and restricted codes (within British English) developed by Basil Bernstein, and introduced into 1970s cultural anthropology by Mary Douglas.
References


