

On counting languages, diversity-wise

MAURO TOSCO
University of Turin

ABSTRACT

The article discusses the problem of counting languages with an eye to assessing the loss of language diversity. It opposes internal and external definitions of language. The article rejects current literature which rests on sociological definitions of language, based upon the conventional wisdom of the speakers and the use of languages in order to flag identities, and pleads for the necessity to embrace an internal definition based upon intelligibility and structural differences. In so doing, it harks back to a time-honored quest for a strictly structural definition of language (vs. dialect). In such a view, neither the speakers' attitudes to their or other languages nor recognition on the part of official bodies play no role, while the task of defining language is therefore handed back squarely from social studies and social-oriented analyses of language to linguistics.

KEYWORDS

Dialect; language diversity; counting languages; intellegibility; Ausbau; Abstand

1. COUNTING LANGUAGES: WHY?

This article is borne out of a long-standing interest in what constitutes a language and a nagging feeling that a discipline which cannot even define its subject matters is not much of a science.

What is a language (as opposed to capitalized Language as a human capacity)? And what is a dialect, if such a thing exists? What about “accents,” “varieties,” etc.? As linguists, we often shy away from such questions, preferring to resort to the old quip on a language being a dialect with an army and a navy (where the absence of an air force points at least to the relative antiquity of the quip itself),¹ or expressing more articulate opinions to the effect that everything boils down to matters of prestige, officialdom, a graphic norm, or still appealing to matters of “identity” and the like.

And yet, as linguists we are still eager to point out that we care about language diversity, language shift, and even language death.

Hagège (2000) calculates that, out of 5,000 existing languages, 25 die out every year. Ranka Bjeljac-Babic (quoted by Calvet 2002: 116) assumes the death of only 10 languages pro year, and a total number of 6,000 languages. Crystal (2000: 4), too, starts with 6,000 languages, but calculates that the mortality rate is one language every 15 days (Crystal 2000: 19), which makes for 24 languages a year — close to Hagège’s figure. Even worse is Krauss’s calculation: ‘between *safe*’ and *extinct* is the entire spectrum of *endangered* languages, probably 95% of the 6,000’ (Krauss 2007: 3; emphasis in the original).

Against this flood of widely divergent figures, Calvet (2002: 116) rightly points out that such divergent estimates simply emphasize the total lack of scientific accuracy in them.

Linguists should definitely strive to come up with better measures of language death. The obvious parallel with the (much more talked-about) issue of the decrease in biological diversity does not help much: first, many measure of biological diversity do not stand up to scientific scrutiny (as evidenced by the heated debates which accompanied Lomborg 2001). Second, counting biological species is not without its problems. For both species and languages, the problems arise out of the very definition of the entities to be counted: what counts as a species in biology? And what as a language? If, following Mayr’s (1942: 120) famous definition, ‘species are groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups,’ what about partial interbreeding and the “incipient species,” as in the famous case of the herring gulls? Moving westwards from Northern Europe along an imaginary ring around the North Pole, one finds

¹ Max Weinreich is often credited with having heard it between 1943 and 1944 and having first popularized it in an article in Yiddish in 1945. Other possible authors include Antoine Meillet, Viktor Shklovsky and the French general Hubert Lyautey.

herring gulls which look less and less “prototypical” herring gulls and more and more lesser black-backed gulls until one finds in Europe two species: the herring gull and the lesser black-backed gull, whose aspect is clearly different and do not interbreed (Dennett 1995: 45)? Then there is species merging, which gives rise to hybrid species stemming from two ancestral species: ‘In plants, pollen from one species commonly fertilizes ova from another species,’ and, occasionally, hybrid species result which stem from two ancestral species. Now, ‘[E]stimates of the proportion of plant species in general that are of hybrid origin run as high as 30 or 40 percent.’ (Hull 1988: 103). Finally, there is the problem of asexual organisms.

That linguists tend to avoid such messy problems is therefore understandable, even if not laudable. The results tend to be scientifically untenable, as the next section will try to prove.

2. AGAINST EXTERNAL DEFINITIONS OF LANGUAGENESS²

I share with much sociologically-oriented literature the assumption that language entities are social constructs; they are constructs because they are mental artefacts of members of *Homo Sapiens Sapiens*, and they are social because members of this species live in social groups.

Still, the result of these social constructs is so diverse across members of the species that it is scientifically suitable to consider them different entities. In other words, I consider it possible to count languages on the basis of strictly internal criteria. We may want to count languages, *inter alia*, in order to measure their degree of diversity, and the interaction of the latter with time and space.

Further, we can adopt either an *internal* or an *external* definition of language: by the former I mean a definition based upon characters and features of the languages themselves and upon mutual comprehensibility — to a great extent a consequence of the characteristics themselves. By *external* definition of language I mean a social one, which takes into account the speakers’ perception and judgements. External definitions of language are of concern to sociolinguists but also to general linguists; e.g., Croft (2000), following Hull’s (1988) work in biology, has proposed a definition of language as ‘the population of utterances in a speech community’ (Croft 2000: 26). “Population” is used here in its biological meaning as a spatiotemporally bounded set of actual individuals, such ‘that every speaker perceives every other speaker as someone he or she should be able to communicate with by using what they *perceive* as the same language’ (Croft 2000: 18; emphasis mine).

² This section expands on ideas originally put forward in Tosco (2011).

As Croft's definition is meant to be evolutionary and not static, structural features, genetic relationship, or intelligibility play no role in it. It is essentially a social definition of language, based upon one population's (here in its common-sensical meaning, i.e., the speakers) *view* of its language(s) ('what they *perceive* as the same language') rather than of the language itself. This seems highlighted by Croft himself when he defines as *sibling languages* 'two linguistic varieties that are structurally so similar that they are considered to be "dialects of the same language", yet are perceived by the speakers — or at least by one group of speakers — as distinct languages.' Examples of sibling languages would be, among others, Macedonian and Bulgarian, Serbian and Croatian, Hindi and Urdu, etc. Immediately afterwards, Croft is forced to notice that opinion must not be unanimous across the speaking community: e.g., 'many Bulgarians tend to see Macedonian as a dialect of Bulgarian, but the reverse does not hold. Of course, this reflects different perceptions about the social and political separateness of the communities that speak these linguistic varieties' (Croft 2000: 16). Again, this is not a statement about languages, but about the communities speaking them. How much powerful, influential, and vociferous must an opinion be in order for *siblingness* to be established?

As variation is continuous and not discrete, within an internal definition we may want to define a minimum threshold which may distinguish what counts as a different language from what will be regarded as separate instances of one and the same entity.

But is there a *minimum* threshold of diversity which a linguistic object must cross in order to be perceived by its speakers as *different enough*? Apparently not: a modicum of lexicon variation seems sometimes to be sufficient to mark identity. The point can be illustrated by Francanglais (or Camfranglais), a French-based variety spoken in particular by the urban youth in Cameroun. As convincingly shown by Féral (2009b), it is entirely French in grammar and in the great majority of its lexicon. Its Cameroonian "flavor" is given by 1. a limited number of loans from African languages and the English-based Cameroonian Pidgin, and/or just "plain" English, 2. the use of many French colloquialisms (some of them obsolete in France), and 3. a number of semantic shifts and phonological and morphological manipulations, such as truncations and metatheses. All this is sufficient for Francanglais to be *perceived* by at least a subset of the speakers as a separate entity.

Finally, what about the well known cases when attitudes change and two varieties which were considered by the speakers (even *all* of them) as dialects of the same language come to be considered (even by *all*) as two separate languages? Is this a statement about the linguistic objects (dialects, languages), or about the *perception* of such objects? What about the whole problem of *naming* an object (in our case, a *linguistic* object) as an essential element of its perceived existence (on which cf. Féral 2009a and the articles in the volume)? The very denomination Camfranglais/Francanglais (very possibly originally

exogenous, but soon adopted by the speakers) transformed what had until then been called *français makro* (“roughneck French” – obviously another external denomination) into something else: a language, which moreover, by its very name, well accords to the official Cameroonian ideal (actually: *ideology*) of French-English bilingualism (*Cam-Fr-Anglais*). A variety of French could then become an identity marker (cf. Féral 2009b: 144 and Féral 2011).

In Croft’s view, the mirror case of the sibling languages is provided by the *polytypic languages*, i.e., ‘linguistic varieties that are structurally so diverse that linguists would characterize them as different languages, yet their speakers perceive them as dialects of the same language’ (Croft 2000: 16): examples are the Chinese “dialects”, the speech of diglossic communities (as in the Arab world), and the postcreole continua. Also the ‘traditional dialects of English, German, Italian and other western European languages may be instances of a lower degree of polytypy, depending on the degree to which their speakers identify themselves as speakers of English, German, etc., albeit non-standard speakers.’ (Croft 2000: 17). Let us imagine a particularly “aggressive” and demographically powerful community of X-speakers which, any linguistic (structural) difference notwithstanding, considers the neighboring, demographically weaker variety Y as a “dialect of the same language” (X, obviously). Would we still have polytypic languages? Probably yes. Would as a consequence linguistic research be led astray in its quest for linguistic diversity across the globe? Certainly yes.

In short, there seems to be a gross oversimplification here: speakers may still identify themselves as speakers of X while being well aware that communication between theirs and another, “standard” variety not only practically does *not* occur for social reasons (because certain topics or speech contexts ban the use of one of the varieties), but it is also *impossible* for strictly linguistic (structural) reasons, because there is no mutual intelligibility. All this of course has not even addressed the quite common case in which speakers simply cannot or do not want to make up their mind about what is what, a dialect, a language, or whatever.

Croft’s definition of language closely resembles Connor’s (1978) classical definition of a nation: while an ethnic group can be objectively defined from the outside by an external observer, a nation, Connor argues, is nothing more than an ethnic group which “has discovered itself” and defines as such. In short, it amounts to “seeing oneself as X.” Both Croft’s definition of language and Connor’s view of nation cannot escape an obvious paradox: while biological populations are defined externally (by the biologist), for linguistic/national populations the observer should be content with registering the — often volatile, always inconsistent — opinion of the community itself, i.e., the mutually contradictory opinions of its members. What counts as a language becomes then a *statistical* truth.

Let us go back to the case of Camfranglais: we could of course list Camfranglais among the languages of Cameroon, on the basis of its role as an identity marker, and at the same time *not* list other varieties of French which are at least as different from “normal” French as Camfranglais but which do not mark an identity. Maybe they even do act as identity markers (whatever this means) and we do not know. Maybe they will, in due time social scientists will take duly note and this knowledge will trickle down to sociolinguistics. We can even describe this process in its unfolding, and the difficult, painful and always contested ways through which a language is socially construed as part of a new identity (another case from Africa being Juba Arabic, an Arabic-based pidgincreole of South Sudan, which is gradually evolving as an identity marker, as studied by Manfredi forth. and Manfredi and Tosco forth.). We can do all these things and much else, but I doubt this will help us much if our task is to identify languages, rather than identities and social groups.

To list languages which are used as identity markers would probably be Croft’s solution (although he wisely sticks to “easy” cases such as Macedonian and Bulgarian, Urdu and Hindi, etc.), and seems to be a very widespread choice. Certainly, this solution means to give up any serious attempt at defining, naming and counting languages: identities only will be counted (and the problems in defining them left to specialists in the field).

We could also list as separate languages Camfranglais as well as any other French (or French-based) variety: this would certainly make for a nice catalogue of language variation across the globe. But where to stop? In principle, we should count any register of any language, and maybe even single idiolects. Maybe we should even make ones step further and take into account the fact that everyone’s idiolect constantly changes (even dramatically) during one’s life. Certainly, reference to the individual’s language (the idiolect) would at least give away with the problem of defining languages on the basis of the intuitions and feelings of an ill-defined community.

It is not accidental that external views of what counts as a language often end up belittling the whole problem of language diversity and its loss. If everything can count as a language, then the concept of language loses any interest: there are no longer languages, nor, *a fortiori*, language shift and language death.

3. LANGUAGE DIVERSITY UNDER THE CARPET

Croft’s (2000) evolutionary view of language as ‘the population of utterances in a speech community’ is the language-external solution devised by a great typologist in order to get rid of the problem of defining languages. But his polytypic languages are nothing more than a sophisticated way to say that

whatever traditional wisdom – backed up by modern governments – calls “a dialect of X” (be X Chinese, English, or whatever) is just a variant of X. End of the story. But linguistically it is not, because the members of a polytypic language rarely bear the same structural relationship with each others. On their part, polytypic languages are rather the linguistic expression of cultural areas, and can be likened to the concept of “macrolanguage” introduced in the most recent editions of *Ethnologue*.

The same, sad fate of the perception of language death around the world can be seen in sociolinguistic literature, too, and nurtured again by an external view of what count as a language. Joseph notes that, while business people and anthropologists are likely to have a more balanced view of identity loss, ‘[L]inguists, on the other hand, tend toward more extreme negative reactions’ (Joseph 2004: 182). He goes on criticizing (rightly, in my opinion) much current views against globalization, and claiming that the loss of language diversity in the contemporary world (although ‘real and lamentable;’ Joseph 2004: 186) is not at all unprecedented. Joseph prefers instead to stress the rise of new dialects (possibly tomorrow’s separate languages) as a result of the spread of international languages, and asks:

‘Was Europe more linguistically diverse before the spread of Latin and the retreat of various pre-Indo-European and Indo-European languages than it was after the break-up of Latin into Romance dialects which in part reflected the structure of those earlier substratum languages? The linguist is inclined to say as a knee-jerk reaction that the prior situation was one of more diversity because the languages involved showed a larger typological difference from one another. Yet degree of typological difference does not really mean much to ordinary speakers of the language’ (Joseph 2004: 187).

Calvet (2002) likewise notes that, while much talk is made about the death of languages, nothing is heard about the *birth* of them. Now, if we examine which languages are born, we note that they are often the new Englishes and Frenchs in, e.g., Africa: one could make a good case of these new languages as being genetically less diverse than the dead languages they come, so to speak, to replace, and, therefore, the net result of language shift being an impoverishment.

An external view of what counts as a language, and a cautious, if not overtly hostile attitude toward language loss are probably logically and inescapably linked: if speakers only are to be the judges of what counts for a language and what does not, language loss can no longer be assessed (and measured) objectively. Maybe, it does not even exist anymore: we will be faced instead with *identity* loss – something very different and certainly even more problematic to pinpoint.

In this perspective, linguistics is reduced to (or transformed into) socio-linguistics – a move Joseph seems to explicitly advocate when he stresses the

need for *rehumanising* linguistics (probably to be understood as *resocializing* linguistics).

Joseph (2004) is also definitely right on one point: speakers could not care less for typological difference. Linguists, instead, do.

It is with an eye to this crucial difference that we move to the *pars construens* in our argument, sketchy and tentative as it may be.

4. A PLEA FOR INTERNAL DEFINITIONS

Linguists have not used armies and navies to speak of languages and dialects all the time, and the quest for strictly structural definitions of language is of course far from new. As aptly reminded in the very first lines of Tamburelli (2014) ‘[D]uring the twentieth century, many linguists were preoccupied with identifying the criteria that would allow for a structural (i.e. purely linguistic) definition of ‘language’ and how this related to its ‘dialects’ [...] This preoccupation faded, however, as consensus grew that ‘languages’ and ‘dialects’ are social constructs definable only in terms of sociolinguistic status and breadth of use and are thus not independently identifiable structural entities’ (Tamburelli 2014: 252). In a way, out of the classical Klossian dichotomy of Ausbau vs. Abstand languages (Kloss 1967), it is the former which, after attracting the most interest, definitely got the upper hand. Abstand languages – languages which can be defined as such on the basis of their inherent linguistic features – seemed not only to be a rarity (and they were usually mentioned only in the relatively rare cases of linguistic isolates), but also to be scarcely interesting to the linguist. “Ausbau-centrism,” as Tamburelli calls it, implied that to measure intelligibility came to be considered impossible or leading to contradictory results. Dialect chains have often been mentioned as a litmus test proving the futility of measuring intelligibility (although Hammarström 2008 has, convincingly in my view, demonstrated the mathematical possibility of counting languages even in this case). I cannot even approach here the complex question of how intelligibility can and is actually measured. In his pioneering study of Lombard and Italian,³ Tamburelli (2014) has not only laid bare the logical pitfalls and paradoxes of “Ausbau-centrism,” but has also shown how intelligibility (and the lack thereof) may be tested and measured (in his case, using a version of the “Speech Perception in Noise,” or SPIN, test).

In short, language diversity may only be vindicated if we stick to an internal definition of language, essentially resting upon the criterium of intelligibility. Intelligibility – all the well-known problems in defining and measuring it notwithstanding – will of course be tested on speakers, but it will be

³ Cf. also Grimes (1988) for an early defense of intelligibility as a valid criterium for counting languages.

defined by the observer only, and will not be based for our purposes on the sociolinguistic status of the linguistic objects under examination. As a result, the only ones *not* entitled to have a say in matters of what counts and what does not as a language are the speakers.

Such an approach will probably reduce the total number of languages in the world as given, e.g., by *Ethnologue* (which often lists what are sociolinguistic languages but linguistic dialects – a subset of Croft’s “sibling languages”). On the other hand, it will exclude mere accents and registers of languages: no confusing and unnecessary blow-up of the number of languages will be engendered. Much else will be gained: notwithstanding their social (and political) status as *sociolinguistic* dialects, many varieties will be defined as *linguistic* languages. It is these sociolinguistic-dialects-plus-linguistic-languages which make up much of the language diversity across the globe and it is them, and their fate, which we address when we speak of language diversity and its reduction.

5. UNINTENDED CONSEQUENCES: TURNING A BLIND EYE TO REALITY

Quite paradoxically, much discussion on minority languages, language rights and language death, is not really concerned with the problem of defining what counts as a language, nor, actually, of what is a minority, what are linguistic rights and who or what are the linguistic-right holders. Many social studies are content with what legislatures define (legitimately for their purposes) as languages, and reflect on the implementation of these choices and their (often negative) results.

We live in a “legislating world,” where laws are continuously enacted in all fields of human activity and are widely presumed to be the solution to most if not all problems. In so doing, law must define its object matters for its purposes.

Being concerned with the allocation of finite resources, legislatures will always de facto select a convenient number of languages and groups to which special consideration will be granted. This fact alone encourages the frame of mind whereby a minority exists only insofar as it is legislated to be such. Minorities left out of consideration, for whatever reason, are consequently aligned to the majority, forced to pay in consequence, and their linguistic assimilation comes to be taken for granted. These excluded minorities not only are non-existent as far as positive legislation is concerned, but very often cease to exist in the speakers’ mind too: an entity exists only insofar as power has decreed it to, and under its conditions.⁴

⁴ Other possible and more tragical outcomes of positive language legislation are not

The “Ausbau-centric” view of language tends to do much the same: resting as it does on the existent (e.g., laws 1, 2, 3), it is forced to work on the linguistic material defined by it (languages a, b, c) and passes under silence whatever portion of the linguistic reality is not taken into account by those same laws (languages x, y, z). From this to the negation of the un-legislated reality it is but an easy step.

It is also an easy step with many unintended consequences. In the end, only what law has defined as existing really does – much to the joy of the frame of mind which sees reality as a stipulated social convention and truth as conventional wisdom and nothing else. In such a paradoxical upshot, post-modernism welds with a legalistic frame of mind against both common sense and science – and good riddance to all the (self-styled) potential for social criticism which, we are often told, would be inherent in social studies.

really germane to the present discussion and must be left out of consideration; they involve the devaluation of bottom-up, grassroot activities for language preservation which, as Fishman (1991) argues at length, are possibly the most (or even the only) effective in matters of language diversity: when governments “take care” (usually inefficiently) individuals and local groups may well take a rest. Cf. Tosco (forth.) for a few preliminary observations.

REFERENCES

- CALVET, L.-J. (2002), *Le marché aux langues. Essai de politologie linguistique sur la mondialisation*. Paris: Plon.
- CONNOR, W. (1978), A Nation is a Nation, is a State, is an Ethnic Group, is a... *Ethnic and Racial Studies* 1/4: 377-400.
- CROFT, W. (2000), *Explaining Language Change*. London: Longman.
- CRYSTAL, D. (2000), *Language Death*. Cambridge: Cambridge University Press.
- DENNETT, D. C. (1995), *Darwin's Dangerous Idea*. Harmondsworth: Penguin Books.
- FÉRAL, C. de (2009a), "Introduction", in C. de Féral (ed.), *Le nom des langues III. Le nom des langues en Afrique sub-saharienne : pratiques, dénominations, catégorisations. Naming Languages in Sub-Saharan Africa: Practices, Names, Categorisations*. Louvain-la-Neuve: Peeters: 9-17.
- FÉRAL, C. de (2009b), "Nommer et catégoriser des pratiques urbaines : Pidgin et Francanglais au Cameroun", in C. de Féral (ed.), *Le nom des langues III. Le nom des langues en Afrique sub-saharienne : pratiques, dénominations, catégorisations. Naming Languages in Sub-Saharan Africa: Practices, Names, Categorisations*. Louvain-la-Neuve: Peeters: 119-152.
- (2011), "Les 'variétés' du français en Afrique. Stigmatisations, dénominations, réification: à qui la faute ?", *Cahiers de sociolinguistique*, PUR – Presse Universitaire de Rennes 15: 41-54.
- FISHMAN, J. A. (1991), *Reversing Language Shift*, Clevedon: Multilingual Matters.
- GRIMES, B. F. (1988), "Why Test Intelligibility?", *Notes on Linguistics*, 42: 39-64.
- HAGÈGE, C. (2000), *Halte à la mort des langues*, Paris: Odile Jacob.
- HAMMARSTRÖM, H. (2008), "Counting Languages in Dialect Continua Using the Criterion of Mutual Intelligibility", *Journal of Quantitative Linguistics* 15/1: 36-45.
- HULL, D. L. (1988), *Science as a process: an evolutionary account of the social and conceptual development of science*, Chicago: The University of Chicago Press.
- JOSEPH, J. E. (2004), *Language and Identity: National, Ethnic, Religious*, Houndmills, Basingstoke, Hampshire & New York: Palgrave Macmillan.

- KLOSS, H. (1967), "Abstand Languages" and "Ausbau Languages." *Anthropological Linguistics* 9/7: 29-41.
- KRAUSS, M. (2007), "Classification and Terminology for Degrees of Language Endangerment", in M. Brenzinger (ed.), *Language Diversity Endangered*, Berlin: Mouton de Gruyter: 1-8.
- LOMBORG, B. (2001), *The skeptical environmentalist*, Cambridge: Cambridge University Press.
- MANFREDI, S. (forth.), "The construction of linguistic borders and the rise of national identity in South Sudan. some insights into Juba Arabic (árabi júba), in R. Bassiouney (ed.), *Identity and Dialect Performance*, London: Routledge.
- MANFREDI, S. and TOSCO, M. (forth.), "A new state, an old language policy, and a pidgincreole: Juba Arabic in South Sudan", *Sociolinguistic Studies* 2018.
- MAYR, E. (1942), *Systematics and the origin of species*, Cambridge: Cambridge University Press.
- TAMBURELLI, M. (2014), "Uncovering the 'hidden' multilingualism of Europe: an Italian case study", *Journal of Multilingual and Multicultural Development* 35/3: 252-270.
- TOSCO, M. (2011), "The Nation-State and Language Diversity", in P. Valore (ed.), *Multilingualism. Language, Power, and Knowledge*, Pisa: Edistudio: 87-101.
- (forth.), "From language to dialect and back: the case of Piedmontese", in R. Bassiouney (ed.), *Identity and Dialect Performance*, London: Routledge.