1   *What is language death?*

The phrase 'language death' sounds as stark and final as any other in which that word makes its unwelcome appearance. And it has similar implications and resonances. To say that a language is dead is like saying that a person is dead. It could be no other way – for languages have no existence without people.

A language dies when nobody speaks it any more. For native speakers of the language in which this book is written, or any other thriving language, it is difficult to envision such a possibility. But the reality is easy to illustrate. Take this instance, reported by Bruce Connell in the pages of the newsletter of the UK Foundation for Endangered Languages (FEL), under the heading 'Obituaries':[1]

> During fieldwork in the Mambila region of Cameroon's Adamawa province in 1994–95, I came across a number of moribund languages . . . For one of these languages, Kasabe (called Luo by speakers of neighbouring languages and in my earlier reports), only one remaining speaker, Bogon, was found. (He himself knew of no others.) In November 1996 I returned to the Mambila region, with part of my agenda being to collect further data on Kasabe. Bogon, however, died on 5th Nov. 1995, taking Kasabe with him. He is survived by a sister, who reportedly could understand Kasabe but not speak it, and several children and grandchildren, none of whom know the language.

There we have it, simply reported, as we might find in any obituary column. And the reality is unequivocal. On 4 November 1995, Kasabe existed; on 5 November, it did not.

Here is another story, reported at the Second FEL Conference in

[1] Connell (1977: 27). The newsletters of this organization changed their name in early issues. The name was *Iatiku* for Numbers 2–4, and *Ogmios* for No. 6 on. Issues 1 and 5 had no distinctive name, and in this book these are referred to as *FEL Newsletter*.

1

Edinburgh in 1998 by Ole Stig Andersen.[2] This time, 8 October 1992 is the critical day:

> The West Caucasian language Ubuh . . . died at daybreak, October 8th 1992, when the Last Speaker, Tevfik Esenç, passed away. I happened to arrive in his village that very same day, without appointment, to interview this famous Last Speaker, only to learn that he had died just a couple of hours earlier. He was buried later the same day.

In actual fact, Kasabe and Ubykh (a widely used alternative spelling) had effectively died long before Bogon and Tevfik Esenç passed away. If you are the last speaker of a language, your language – viewed as a tool of communication – is already dead. For a language is really alive only as long as there is someone to speak it to. When you are the only one left, your knowledge of your language is like a repository, or archive, of your people's spoken linguistic past. If the language has never been written down, or recorded on tape – and there are still many which have not – it is all there is. But, unlike the normal idea of an archive, which continues to exist long after the archivist is dead, the moment the last speaker of an unwritten or unrecorded language dies, the archive disappears for ever. When a language dies which has never been recorded in some way, it is as if it has never been.[3]

The language pool

How many languages are at the point of death? How many are endangered? Before we can arrive at an estimate of the scale of the problem, we need to develop a sense of perspective. Widely quoted

[2] Andersen (1998: 3).
[3] There is, of course, always the possibility that other speakers of the same dialect will be found. In the Ubykh case, for instance, there were at the time rumours of two or three other speakers in other villages. Such rumours are sometimes found to be valid; often they are false, with the speakers being found to use a different dialect or language. But even if true, the existence of a further speaker or two usually only postpones the real obituary by a short time. For some Aboriginal Australian examples, see Wurm (1998: 193). Evans (forthcoming) provides an excellent account of the social and linguistic issues which arise when working with last speakers, and especially of the problem of deciding who actually counts as being a 'last speaker'.

figures about the percentage of languages dying only begin to make sense if they can be related to a reliable figure about the total number of languages alive in the world today. So how many languages are there? Most reference books published since the 1980s give a figure of between 6,000 and 7,000, but estimates have varied in recent decades between 3,000 and 10,000. It is important to understand the reasons for such enormous variation.

The most obvious reason is an empirical one. Until the second half of the twentieth century, there had been few surveys of any breadth, and the estimates which were around previously were based largely on guesswork, and were usually far too low. William Dwight Whitney, plucking a figure out of the air for a lecture in 1874, suggested 1,000.[4] One language popularizer, Frederick Bodmer, proposed 1,500; another, Mario Pei, opted for 2,796.[5] Most early twentieth-century linguists avoided putting any figure at all on it. One of the exceptions, Joshua Whatmough, writing in 1956, thought there were 3,000.[6] As a result, without professional guidance, figures in popular estimation see-sawed wildly, from several hundred to tens of thousands. It took some time for systematic surveys to be established. *Ethnologue*, the largest present-day survey, first attempted a world-wide review only in 1974, an edition containing 5,687 languages.[7] The Voegelins' survey, published in 1977, included around 4,500 living languages.[8] Since the 1980s, the situation has changed dramatically, with the improvement of information-gathering techniques. The thirteenth edition of *Ethnologue* (1996) contains 6,703 language headings, and about 6,300 living languages are classified in the *International encyclopedia of linguistics* (1992).[9] There are 6,796 names listed in the index

[4] See Silverstein (1971: 113).
[5] Bodmer (1944: 405). Pei (1952: 285); in a later book (1954: 127), this total decreased by 1.
[6] Whatmough (1956: 51).
[7] See now the 13th edition, Grimes (1996); also www.sil.org/ethnologue. The first edition in fact dates from 1951, when Richard S. Pittman produced a mimeographed issue of ten pages, based on interviews with people attending the Summer Institute of Linguistics.
[8] Voegelin and Voegelin (1977). I used their total in the first (1987) edition of my *Cambridge encyclopedia of language* (Crystal 1997a).
[9] Bright (1992); the files of *Ethnologue* (then in its 11th edition) were made available for this project, hence the similarity between the totals.

to the *Atlas of the world's languages.*[10] The off-the-cuff figure most
often heard these days is 6,000, with the variance sometimes going
below, sometimes above.[11] An exceptionally high estimate is
referred to below.

 A second reason for the uncertainty is that commentators know
that these surveys are incomplete, and compensate for the lack of
hard facts – sometimes by overestimating, sometimes by underes-
timating. The issue of language loss is itself a source of confusion.
People may be aware that languages are dying, but have no idea at
what rate. Depending on how they estimate that rate, so their
current global guess will be affected: some take a conservative view
about the matter; some are radical. (The point is considered
further below.) Then there is the opposite situation – the fact that
not all languages on earth have yet been 'discovered', thus allowing
an element of growth into the situation. The ongoing exploration
of a country's interior is not likely to produce many fresh encoun-
ters, of course, given the rate at which interiors have already been
opened up by developers in recent years; but in such regions as the
islands of Indonesia and Papua New Guinea, or the South
American or Central African rainforests, reports do come in from
time to time of a previously unknown community and language.[12]
For example, in June 1998 two such nomadic tribes (the Vahudate
and the Aukedate, comprising 20 and 33 families, respectively)
were found living near the Mamberamo River area, 2,400 miles
east of Jakarta in Irian Jaya. This is a part of the world where the
high mountains and deep valleys can easily hide a community, and

---

[10] This is my count of Mosely and Asher (1994).

[11] Dixon (1997: 143) cites 5,000–6,000, as do Grenoble and Whaley (1998a), in their
preface; Wardhaugh (1987: 1) cites 4,000–8,000, and settles on 5,000; Ruhlen (1987) goes
for 5,000; Wurm (1991: 1) says 'well over 5,000'; Krauss consulted a number of linguists
in writing his article on 'The world's languages in crisis' (1992: 5), and found widespread
agreement that 6,000 was a reasonable estimate; Crystal (1997a: 287) also cites 6,000.
Other major surveys are in progress: a 'World Languages Report', supported by UNESCO
and Linguapax, and financed by the Basque Country, is scheduled for publication in
2001; see also the Global Language Register below.

[12] The world's languages have a highly uneven distribution: *c.* 4% are in Europe; *c.* 15% in
the Americas; *c.* 31% in Africa; *c.* 50% in Asia and the Pacific. The countries mentioned
have the highest distributions: Papua New Guinea and Indonesia alone have 25% (1,529
languages) between them (according to the 1996 edition of *Ethnologue*).

it is likely that their speech will be sufficiently different from that of other groups to count as a new language. The social affairs office in the region in fact reports that its field officers encounter new groups almost every year.[13]

Even in parts of the world which have been explored, however, a proper linguistic survey may not have been carried out. As many as half the languages of the world are in this position. Of the 6,703 languages listed in the thirteenth edition of *Ethnologue*, 3,074 have the appended comment – 'survey needed'. And what a survey chiefly does is determine whether the speakers found in a given region do indeed all use the same language, or whether there are differences between them. If the latter, it then tries to decide whether these differences amount only to dialect variations, or whether they are sufficiently great to justify assigning the speakers to different languages. Sometimes, a brief preliminary visit assigns everybody to a single language, and an in-depth follow-up survey shows that this was wrong, with several languages spoken. Sometimes, the opposite happens: the initial visit focuses on differences between speakers which turn out not to be so important. In the first case, the number of languages goes up; in the second case, it goes down. When decisions of this kind are being made all over the world, the effect on language counts can be quite marked.

To put some flesh on these statistics, let us take just one of those languages where it is said a survey is needed: Tapshin, according to *Ethnologue* also called Tapshinawa, Suru, and Myet, a language spoken by 'a few' in the Kadun district of Plateau State, Nigeria. It is said to be unclassified within the Benue-Congo broad grouping of languages. Roger Blench, of the Overseas Development Institute in London, visited the community in March 1998, and sent in a short report to the Foundation for Endangered Languages.[14] He stressed the difficulty of reaching the settlement: Tapshin village is a widely dispersed settlement about 25 km north of the

---

[13] The report is reproduced in *Ogmios* 9. 6. For similar discoveries in South America, see Adelaar (1998: 12); Kaufman (1994: 47) reports that about 40 languages have been discovered in South America during the past century.   [14] Blench (1998).

Pankshin–Amper road, reached by a track which can be traversed only by a four-wheel drive, and which is often closed during the rainy season. The Tapshin people call themselves Ns'r, and from this derives Blench's name for them, Nsur, and presumably also the name Suru in *Ethnologue*; but they are called Dishili by the Ngas people (referred to as the Angas in *Ethnologue*). The name Myet derives from a settlement, Met, some distance west of Tapshin. The Tapshin people claim that the Met people speak 'the same' language as they do, but Blench is cautious about taking this information at face value (for such judgements may be no more than a reflection of some kind of social or historical relationship between the communities). No data seems previously to have been recorded on Nsur. From his initial wordlists, he concludes that there has been substantial mutual influence with the Ngas language. He estimates that there are some 3–4,000 speakers, though that total depends on whether Met is included along with Nsur or not.

This small example illustrates something of the problem facing the linguistic analyst. There is a confusion of names which must be sorted out, in addition to the observable similarities and differences between the speakers.[15] The Nsur situation seems fairly manageable, with just a few alternatives to be considered. Often, the problem of names is much greater. Another Plateau State language, listed as *Berom* in *Ethnologue*, has 12 alternative names: *Birom, Berum, Gbang, Kibo, Kibbo, Kibbun, Kibyen, Aboro, Boro-Aboro, Afango, Chenberom,* and *Shosho.* The task then is to establish whether these are alternative names for the same entity, or whether they refer to different entities – the name of the people, the name of an individual speaker, or the name of the language as known by its speakers (a European analogy would be *Irish, Irishman/woman,* and *Gaelic/Irish/Erse,* respectively). Then there is the question of what the language is called by outsiders. There could of course be several 'outsider' names (*exonyms*), depending on how many other groups the language is in contact with (cf.

[15] For a discussion of the problem of naming, with particular reference to China, see Bradley (1998: 56 ff.).

deutsch being equivalent to *allemand, German, Tedesco*, etc.), and
these might range from friendly names through neutral names to
offensive names (cf. 'He speaks French' vs 'He speaks Frog').
Shosho, in the above list, is apparently an offensive name. But all
this has to be discovered by the investigator. There is no way of
knowing in advance how many or what kind of answers will be
given to the question 'What is the name of your language?', or
whether a list of names such as the above represents 1, 2, 6, or 12
languages. And the scale of this problem must be appreciated: the
6,703 language headings in the *Ethnologue* index generate as many
as 39,304 different names.

Many of these names, of course, will refer to the dialects of a lan-
guage. But this distinction raises a different type of difficulty: does
a name refer to the whole of a language or to a dialect? The ques-
tion of whether two speech systems should be considered as separ-
ate languages or as dialects of the same language has been a focus
of discussion within linguistics for over a century. It is crucial to
have criteria for deciding the question, as the decisions made can
have major repercussions, when it comes to language counting.
Take, for example, the Global Language Register (GLR), in the
process of compilation by the Observatoire Linguistique:[16] in a
1997 formulation by David Dalby, this project proposed a three-
fold nomenclature – of *tongue* (or *outer language*), *language* (or
inner language – or *idiom*, in a further proposal), and *dialect* – to
avoid what it considered to be the oversimplified dichotomy of *lan-
guage* and *dialect*. Early reports related to this project suggested
that, using these criteria, an order of magnitude of 10,000 lan-
guages was to be expected – a surprisingly large total, when com-
pared with the totals suggested above. The explanation is all to do
with methodology. The GLR total is derived from the *tongues* and
idioms of their system, and includes as languages many varieties
which other approaches would consider to be dialects. One

---

[16] The following details are taken from a Logosphere Workshop held at the School of
Oriental and African Studies, London, September 1997, specifically from Dalby (1997),
and his follow-up paper subsequently circulated.

example will illustrate the 'inflationary' effect of this approach. The orthodox approach to modern Welsh is to consider it as a single language, with the notable differences between (in particular) north and south Welsh referred to as dialects. On grounds of mutual intelligibility and sociolinguistic identity (of Wales as a nation-principality), this approach seems plausible. The GLR analysis, however, treats the differences between north and south Welsh as justifying the recognition of different languages (each with their own dialects), and makes further distinctions between Old Welsh, Book Welsh, Bible Welsh, Literary Welsh, Modern Standard Welsh, and Learners' Normalized Welsh (a pedagogical model of the 1960s known as 'Cymraeg Byw'). Excluding Old Welsh, in their terms a total of six 'inner languages' can be recognized within the 'outer language' known as modern Welsh. One can see immediately how, when similar cases are taken into account around the world, an overall figure of 10,000 could be achieved.

The language/dialect issue has been addressed so many times, in the linguistics literature, that it would be gratuitous to treat it in any detail here.[17] In brief, on purely linguistic grounds, two speech systems are considered to be dialects of the same language if they are (predominantly) mutually intelligible. This makes *Cockney* and *Scouse* dialects of English, and *Quechua* a cover-name for over a dozen languages. On the other hand, purely linguistic considerations can be 'outranked' by sociopolitical criteria, so that we often encounter speech systems which are mutually intelligible, but which have nonetheless been designated as separate languages. A well-recognized example is the status of Swedish, Danish, and Norwegian, which are counted as separate languages despite the fact that the members of these communities can understand each other to an appreciable extent. A more recent example is *Serbo-Croatian*, formerly widely used as a language name to encompass a set of varieties used within former Yugoslavia, but following the

[17] Standard accounts are to be found in Chambers and Trudgill (1980: ch. 1) and Crystal (1997a: ch. 47).

civil wars of the 1990s now largely replaced by the names *Serbian*, *Croatian*, and *Bosnian*. In 1990 there was a single language spoken in these countries; now there are three. The linguistic features involved have changed hardly at all; but the sociopolitical situation has changed irreversibly.

It is of course likely that the linguistic differences between these languages will increase, as their respective communities strive to maximize them as symbols of local identity. This process is already happening. If it continues, then one day it is conceivable that Serbian and Croatian could become mutually unintelligible – a further example of something that has happened repeatedly and normally in linguistic evolution. Indeed, it is possible that a significant increase in the world's languages may one day emerge as an evolutionary consequence of the contemporary trend to recognize ethnic identities. Even global languages could be affected in this way. The point has been noted most often in relation to English, where new varieties have begun to appear around the world, as a consequence of that language's emerging status as a world lingua franca. Although at present Singaporean, Ghanaian, Caribbean, and other 'New Englishes' continue to be seen as 'varieties of English', it is certainly possible for local sociopolitical movements to emerge which would 'upgrade' them to language status in due course. Books and articles are already appearing which (in their nomenclature, at least) anticipate such outcomes.[18] After all, if a community wished its way of speaking to be considered a 'language', and if they had the political power to support their decision, who would be able to stop them doing so? The present-day ethos is to allow communities to deal with their own internal policies themselves, as long as these are not perceived as being a threat to others. The scenario for the future of English is so complex and unpredictable, with many pidgins, creoles, and mixed varieties emerging and gradually acquiring prestige, that it is perfectly possible that in a few generations time the degree of local distinctiveness in a speech

[18] McArthur (1998), Rosen (1994), and the journal *World Englishes*. See also Crystal (1998).

system, and the extent of its mutual unintelligibility with other his-
torically related systems, will have developed to the extent that it
will be given a name other than 'English' (as has happened already
– though not yet with much success – in the case of Ebonics). At
such a time, a real evolutionary increase in the number of 'English
languages' would have taken place. A similar development could
affect any language that has an international presence, and where
situations of contact with other languages are fostering increased
structural diversity. The number of new pidgins and creoles is likely
to be relatively small, compared with the rate of language loss, but
they must not be discounted, as they provide evidence of fresh lin-
guistic life.

Estimates about the number of languages in the world, there-
fore, must be treated with caution. There is unlikely to be any
single, universally agreed total. As a result, it is always problematic
translating observations about percentages of endangered lan-
guages into absolute figures, or vice versa. If you believe that 'half
the languages in the world are dying', and you take one of the
middle-of-the-road totals above, your estimate will be some 3,000
languages. But if you then take this figure out of the air (as I have
seen some newspaper reporters do), and relate it to one of the
higher estimates (such as the Global Language Register's 10,000),
you would conclude that less than a third of the world's languages
are dying – and, as a consequence, that the situation is not as
serious as has been suggested. The fact that this reasoning is illegit-
imate – the criteria underlying the first total being very different
from those underlying the second – is disregarded. And, as I read
the popular press, I see all kinds of claims and counter-claims being
made, with the statistics used to hold a weight of argument they
cannot bear.

At the same time, despite the difficulties, we cannot ignore the
need for global measures. As so much of the situation to be
described below is bound up with matters of national and interna-
tional policy and planning, we have to arrive at the best estimates
we can, in order to persuade governments and funding bodies
about the urgency of the need. Accordingly, I will opt for the range