Trends in Linguistics
Studies and Monographs 181

Editors
Walter Bisang
Hans Henrich Hock
Werner Winter

Mouton de Gruyter
Berlin · New York

Language Diversity Endangered

edited by
Matthias Brenzinger

Mouton de Gruyter
Berlin · New York

Chapter 1
Classification and Terminology for Degrees of Language Endangerment

Michael Krauss

1. The following is a suggested framework or schema for classifying languages according to degree of viability, from *'safe'* to *extinct*, with terminology and designators.

'safe'			*a+*	
endangered	*stable*		*a–*	all speak, children & up
	incipiently	*instable; eroded*	*a*	some children speak; all children speak in some places
		definitively endangered	*b*	spoken only by parental generation and up
		severely endangered	*c*	spoken only by grandparental generation and up
		critically endangered	*d*	spoken only by very few, of great-grandparental generation
extinct			*e*	no speakers

The schema and discussion do not address definition of "language" vs. "dialect" for example, or the type of rapid linguistic evolution or "decay" (such as loss of inflection, incorporation of loan words) which is considered by some also as "endangerment".

The three basic categories are *'safe'* and *extinct*, with everything in between *endangered*, by far the largest category, to be taken up last.

2. The term *'safe'*, designated *a+*, I have adopted as a technical term, so keep that in single quotes, to be used with caution, though perhaps that is inconsistent with the rest, also proposed as technical terms. *'Safe'* are those languages which are not only being learned as mother-tongue by children as the norm, but which we predict will still be being so learned

for the foreseeable future, i. e. throughout this new century, still having at least a viable community, critical mass, of children speakers in the year 2100. Such would be a very large proportion of languages which are now spoken by a million or more, including children, and/or are firmly supported by the power of a state or regional government, this including also, though the population may be well below a million, e. g. *Icelandic* or *Faroese*, 250,000 and 40,000 respectively. The total number of *'safe'* languages may thus currently be about 300 out of 6,000 or 5 %, the majority of those being both spoken by over a million and supported by state power, used in educational and media domains. Sometimes, however, state support does not suffice, as in the notorious case of *Irish*, already severely endangered before it gained that support, or, more often, a million does not suffice without that support, as in the case of *Breton*, or *Quechua*. Even so, other factors may prevail, as in the case of *Yiddish* in conservative or Hasidic communities, which might classify *Yiddish*, perhaps the most famously "dying" language, in the elite class of *'safe'* — to put the enormity of the *endangered* class in perspective — where *Yiddish* may well be at the 95th percentile for 'safety.' Probably no language with fewer than 10,000 speakers could anywhere be classified as *'safe'*, and 10,000 is probably at least at the 65th percentile for language speakership size, the median size being closer to 5,000. (It remains a major study, not addressed here, to consider factors detracting from language *'safety,'* such as genocidal violence, industrial development, environmental degradation, demographic intrusion or upheaval, urbanization, indifference, television exclusively in the dominant language, along with the whole spectrum of attitudes, both of the minority and dominant language speakers).

3. *Extinct*, designated *e*, are languages no longer spoken or even potentially spoken (remembered) by anyone, so for which no new documentation can be obtained. Questionable cases of recent extinction of course are common, perhaps more common than cases where it is certain that not a single speaker anywhere survives. Borderline cases exist too, where a few words or phrases are remembered, which could be quite valuable in determining at least the genetic position of an otherwise undocumented language, though there is no one able to generate sentences in it. Such cases might be designated *e+*. Perhaps still more difficult to classify, but probably *e+*, are languages for which there are no fluent speakers or persons able to generate new sentences, but for which there are persons who may remember extensive rote ritual or epic text, such as *Ainu yukar*. (Similarly, however, there are or have been languages with extensive written literature but with no native speakers, such as *Hebrew* until the late 19th century, *Coptic*, or clerical *Latin*, regularly used in ritual. Commonly these may have new text generated in writing may even be the only written language for the community, and may have persons potentially capable of conversation in them under special conditions. Along with other types, such as *Esperanto* or pidgins, these may perhaps be designated by the proposed scale, or something parallel to that, some perhaps even so high as *'safe,'* however restricted their domain may remain. These, or any language sufficiently documented, have also the potential to be revived and will be considered at the end of this discussion.).

4. Between *'safe'* and *extinct* is the entire spectrum of *endangered* languages, probably 95 % of the 6,000. The term *endangered* is clearly adopted from its use in the field of biology, where "endangered species" are defined as "in danger of extinction throughout all or a significant portion of their range", as distinct from "threatened species", defined as "likely to become endangered within the foreseeable future throughout all or a significant portion of their range". For language we have agreed to use the term *endangered* much more broadly, to include a large category, perhaps already a third of "endangered" languages, which are no longer spoken by children; those would correspond in biology to species which have lost the capacity to reproduce. Rather than use terms like "dying", "doomed", "terminal", or even "moribund" or "non-viable", which might well have a discouraging or negative effect, we extend the term *endangered* to include those too, however euphemistic that use may seem, trusting that the term *endangered* may itself be sufficiently alarming. For our purposes, I do not see either that a distinction between *threatened* and *endangered* for language would be very useful; since we are using "endangered" already as a euphemism for a status much graver than what "endangered" means in biology, it seems wrong to compound the obfuscation by calling merely "threatened" (not yet endangered) what does in most cases correspond to "endangered" in biology. Perhaps in further elaboration of a separate study of factors endangering a language, "threatened" might be defined to designate an upper minority of the category of stable but unsafe languages, to double or triple the number of non-endangered languages (*'safe'* 5 % + *threatened* 5–10 % more).

4.1. At the top of the scale in the *endangered* category is the class designated *a, stable*. So long as a language is being learned as mother-tongue by the children, it remains classified as *stable*. This would generally require that virtually all children are so learning the language, in the family,

and actually speaking it not only to their elders but to each other. Home is the essential domain, and so long as home use remains stable, though another language may be used increasingly in school, work, religion, etc., the language remains *stable*, however threatened it may be by factors external to the home. Failure of a language to expand into new technological domains may indeed increase the threat to it, but need not necessarily reclassify it as less than *stable*. A language might remain *stable* if it is merely the "norm" that children learn and speak it in the home, so long as cases where that is not so are truly exceptional, or are common only in a diaspora, e. g. permanently urbanized families, at some remove from the core area. The term *stable* seems detachedly realistic and not leading to complacency, so long as one remembers that that is merely the top category, still the majority worldwide, though probably for not much longer, of endangered languages. The use of the term *stable* is partly inspired by the phrase "Stabilizing Indigenous Languages" in the title for a series of North American conferences.

4.2. For the next subclass of *endangered* languages, incipiently in decline, I see no qualifier better than *instable* or *partly stable*, designated *a–*. (The designators, it is time to explain, are partly derived from the American public school grading system, where *a* is 'excellent', *b* 'good', *c* 'fair, average', *d* 'poor', and *f* 'failing', and *a–* is less than *a* but closer to *a* than to *b*, *b+* better than *b* but closer to *b* than to *a*, etc. They have the advantage of being more iconic than a numeral system, where it is not immediately clear whether the lower number 1 or I as opposed to the higher number 4 or IV is better.) There are clearly two different types of *a–* situations, both defined as where some of the children speak the language. The first type, *instable*, is where "some" of the children speak the language, e. g. some of the children in a single village, or scattered through a wider area, perhaps still a majority, but not constituting a stable or critical mass. *Instable* includes also a situation where the children speak the language some of the time, i. e. to elders, but amongst each other speak the replacing language, so seem destined to speak that to their spouses and children. The other subtype of *a–* might be *partly stable* or *eroding*, where for a more complex situation of several communities where the children all speak the language in one or more parts of it but there also is part or parts where only some children speak the language, especially where a clear geographical distinction is not easy to make. If it is easy to make such a distinction, then a further dimension of the designator system should be used, more than one designator separated by a comma, here namely *a, a–*, meaning that in a part of the language area all children are learning the

language, but in another part only some children are learning it. The language in that case as a whole would be *a–*, *partly stable* or *eroding*, but could be further designated as *stable* plus *instable*. The term *instable* (c. f. instability) is used as a technical term instead of "unstable", in order to avoid the connotation of "mentally unstable".

4.3. The next subclass is *b, definitively endangered* (and definitively in decline) for lack of a better term, meaning that the language has passed the crucial basic threshold of viability, is no longer being learned as mother-tongue by children in the home, that the youngest speakers are of the parental generation, or more precisely that the youngest generation of which all are speakers is the parental generation. (That age could of course vary widely in different parts of the world, minimum probably from 15 to 20.). This might include also situations where the parents not only can but do speak the language to their children, yet permit the children regularly to respond in the replacing language, so that the children hardly become active speakers of the endangered language. Designated with variants of *b*, such as *b–*, might be cases where some of the parents speak the language, or where more uniformly the youngest speaker age is 25 or 30, again meaning in a different way that some parents speak the language. Another type of finer designation might be *b+*, for where the youngest speakers may be 5 or 15, but the intergenerational transmission is definitively interrupted. If the youngest speakers are already even five, the language has probably been definitively abandoned. Though numerically more children may still be able to speak it than cannot, the language should probably be designated *b+* rather than *a–*, because of the dynamic, which is always more important than sheer numbers. The system allows also for two more types of complexity. The first is as shown above, two designators divided by comma, e. g. here *a,b* for two communities or distinct geographical areas, in one of which all speak the language, in the other only parents and up. The second type, which I have very often used, is joining two designators with a hyphen, ambiguous in American notation as the same symbol as a minus sign, e. g. *a–b*, for a complex or continuum which ranges from all children speaking the language to only parents and older speaking it. Perhaps the joining symbol should not be used so ambiguously, so that a situation ranging from where some of the children speak the language in one part to where only parents and up speak it would be *a–b*, thus allowing very finely also, e. g. for *a–b+*, *a–b–*.

4.4. The next subclass is *severely endangered*, *c*, where the youngest speakers are of grandparental generation, middle aged (mutatis mutandis age span of 35–60 or even wider) where parents cannot teach the language

to their children. I have picked "severely" over "seriously endangered", which implies that it is not "serious", or "gravely endangered", too funereal. This is by far the most common basic category for indigenous North American languages, for example, both because of historical timing and the breadth of the age-span. More complex distinctions such as *a–c*, *c–*, can of course be used. For example, the long *Inuit* continuum I have designated *a–c*, for *a* in Greenland and Eastern Canada, but not Labrador, *b* in Central Canada and a few parts of Alaska, *c* in Western Canada and most of Alaska. Another type of fine designation that I occasionally used in my circumpolar report was *–c*, for where the youngest speakers were in the range 35–40, i. e. youngish for grandparent but closer to *c* than to *b*; this should, according to the above, be better symbolized *c+*.

4.5. The last subclass before *extinct* is *critically endangered*, *d*, for languages of which the youngest speakers are in the great-grandparental generation, and are also very few, often fewer than 10 for most American languages, and constituting the second largest class, after *c*, for the U.S. Languages very close to extinction, with all speakers at the very end of life expectancy, and fewer than 10, could be designated *d–*. In this class too the numbers could be problematical. e. g. in the case of *Hawaiian* (not counting *Ni'ihau*, the one isolated island which is *a*), where the youngest speakers throughout are over 70, but with a large enough population that there may be still a thousand such elders, some of whom also could become centenarians; such a language might well resist extinction longer than small languages designated *c* where the number of speakers is 10, some aged only 55, but of whom none might reach 80. Sheer numbers should only secondarily be a factor in the classification however, the *Hawaiian* case being far less frequent than the smaller language populations. For languages in class *d* it should also be noted that the further toward extinction a language moves, the more the actual language ability of the last speakers may become an issue. Often neither academic linguistics nor community language interest can afford to discount speakers with less than eloquent or complete command of a language in this category; a designation of *d–* might well include or consist only of a very few speakers with less than complete competence, or very rusty speakers or semi-speakers. Some languages, e. g. *Sayan Samoyed*, resisted extinction for a generation or more on that basis, possibly also *Ubykh* now.

5. Finally, the designations so far have dealt only with unidirectional movement from *stable* downward to *extinct*. One type of exception would be languages traditionally learned in adulthood, e. g. as I have heard

about *Tarascan* in Mexico, where children learn only *Spanish*, but are expected to learn in young adulthood *Tarascan* and henceforth to speak that with other adults. It would be important to note other such cases. Possibly, *Esperanto* and pidgins could have similar special designations, or secret or ritual languages, such as *Demiin*, *Coptic* or clerical *Latin*. Before considering and designating those, however, we need to consider another type more frequent and important for our purposes, increasingly the result of community programs for reversal of language loss. One example is *Hawaiian*, which might be designated *a*, *d–a*; the *a* is for *Ni'ihau*, the small isolated island where the children all still speak the language, and *d–a* is for the rest of the islands, where until recently the only speakers were the generation of those now past 70 or 75, but where some younger adults began learning the language about 20 years ago, instituted the Punana Leo (Language Nest) movement now spanning kindergarten through high school (all basically taught through the medium of *Hawaiian*, graduates of which are now raising native *Hawaiian*-speaking children). *Maori* in New Zealand had probably reached a designation of *b–c* (maybe *a–c*), but the Kohanga Reo movement (parent to its Hawaiian counterpart) has now produced many child speakers; *Maori* thus might be designated *b–c–a* (or *a–c–a*). *Cornish* was indeed extinct for about a century, and insofar as revivalist claims are correct, that there are now some native-speaking *Cornish* children, could be designated *e–a*, or if not, then *b–a*. *Irish* is still *a* in some of the Gaeltachtaí (*Irish*-speaking districts), but has many more speakers who actually do speak it to each other as a second language and whose children are native speakers of it e. g. in Dublin, so *Irish* generally might be designated *a, b–a*. A pidgin in the process of becoming a creole might also be designated *b–a*.

Possibly, using *x+* instead of the *–y* device used occasionally in my circumpolar paper, the designation *–b* might be reserved for languages learned only in adulthood or school, such as *Tarascan*, *Esperanto*, pidgins, or clerical *Latin*, where the state of decline is not relevant. *Sanskrit* however, reportedly has children speakers, so should accordingly be designated *–a*. Also, until the late 19th century, *Hebrew*, which is now of course, *a+*, and still may have more second-language than native speakers, might be designated as *–a+*! Finally, cases like *Ainu* or successful results of the California-type master-apprentice program, insofar as one or very few adults have successfully learned the language from a last aged speaker, might be designated *d–b*, noting that the hyphen in those loss-reversal cases does not signify a range of speakers throughout the intermediate generations. Presumably terms could be assigned to various types of loss

reversal, e. g. *revived* (from extinction) for *Cornish*, *e–a,* or even *e–b; revitalized* or *restabilized* or *restored*, variously, for cases like c–a, *d–a*, *d–b*; *renativized* for *Sanskrit* or *Hebrew, –b–a*, *nativized* or *creolized* for pidgins, also *–b–a*.

Chapter 2
Threatened Languages in Hispanic South America

Willem F. H. Adelaar

1. Introduction

In the western part of South America language endangerment is not a recent phenomenon. The process of linguistic reduction may have started during the 15th century with the conquest wars conducted by the Incas of Cuzco. At the height of their power the Incas dominated the Andean region from southern Colombia to the centre of Chile with the inclusion of all the coastal areas. The Incas introduced the habit of relocating entire populations from newly conquered areas to places in the centre of the empire, where they could be controlled more easily. Conversely, loyal populations from the centre were taken to the borders for reasons of defense. This practice, known as *mitma*, may have favored the use of the imperial language (*Quechua*) to the detriment of the original languages of some of the affected populations.

2. The linguistic consequences of conquest and colonization

The spread of epidemic diseases during the 16th century, as well as the actions of the Spanish conquerors, who had introduced them, had a devastating effect upon the ethnic and linguistic diversity existing in the area under discussion. Several coastal populations disappeared during the 16th and 17th centuries. If they had languages of their own, these fell into oblivion before they could be described or documented. An example is the *Quingnam* language, which was spoken along the Peruvian coast near Trujillo and further south. The survivors assimilated with the newcomers and turned to *Spanish*. Coastal cities such as Lima and Trujillo became predominantly European in culture and in language, as well as in the physiognomy of their inhabitants.

In the highlands of what are today Bolivia, Ecuador and Peru, the presence of the indigenous population remained strong. A language of communication and administration, known as 'the general language of the Inca', was widely used in the Inca Empire. If not as a mother tongue, it was used as a second language by most of the Inca's subjects, who often main-