Comments received for
ISO 639-3 Change Request
2007-064

Outcome: Approved in part
Effective date: 2008-04-18
Registration Authority decision on Change Request no. 2007-064: requested addition of Lyngngam [lyg] and Khasi (individual language) [xix], and change scope of Khasi [kha] to macrolanguage. Outcome:

Lyngngam [lyg]: adopted
Khasi (individual language) [xix]: rejected
Khasi [kha] change scope of to macrolanguage: rejected

The code element Khasi [kha] is in the code sets of both ISO 639-2 and 693-3. The proposed actions of this change request were motivated by the reluctance to cause disruption to code elements in ISO 639-2, in view of their wide existing use base, in addition to the following statement made in the ISO 639-3 standard document:

> Given sufficient reasons, new code elements may be added, or the denotation for a given code element may be broadened. The denotation of a code element shall not be narrowed, however, as this can result in an unknown proportion of the existing uses of a code element becoming invalid. If it is determined that the denotation of a code element was too broad, such as if a linguistic variety that was thought to be a dialect of a given language was later determined to be a distinct language, the existing code element would be deprecated and superseded by one or more new code elements with narrower denotations.

This statement, without further evaluation, leads to the following possible three options:

1. Retire the existing code element and create two (or as many as are needed) new code elements (the most disruptive with regard to the existing code element).
2. Change the scope of the existing code element to Macrolanguage, and create as many new code elements as are needed.
3. Change the scope of the existing code element to Collection, and create as many new code elements as are needed.

The risk of "an unknown proportion of the existing uses of a code element becoming invalid" is being given very high consideration in the standard, while pragmatically being at odds with another statement in the standard: "To ensure continuity and stability, the identifier for any given language shall not be changed, though the names listed in relation to an identifier may change" (speaking here of the three letter identifier being changed while the denotation intended stays essentially the same).

Arguing that the denotation is not staying the same belies common usage in at least one use-case, that of a request for a code element for a minority, highly divergent “dialect” of a major language. In such a case where the requested variety qualifies for a code element on intelligibility and other linguistic grounds, and is not taken to be "badly spoken standard X," the "unknown proportion of existing uses" where the code element has actually been applied to represent the divergent dialect, rather than the standard form of the language (or much closer to it) is likely to be very low indeed, such that the risk of error is also low, and likely far less disruptive than either a true split (as in 1 above) or a change of scope (2 or 3).

All three effectively result in a change of identifier for the standard language with almost no significant change in denotation as it is applied in common use, unless preexisting applications of the code element that really mean the standard form remain unchanged (to the new identifier for the standard form) by system administrators. In that case, the change in scope engenders its own confusion regarding when the standard form is meant, and when the macrolanguage or collection is meant in any given application of the existing code element. There is really then little difference between the new individual code element for the standard language and the macrolanguage/previously individual code element. However, the confusion would increase if the membership of the macrolanguage (or collection) is expanded to incorporate the other varieties that
already had their own code elements but should also be included by virtue of being at least as closely associated with the larger group as the variety instigating the split. If this last step is not taken, then the code set potentially will have two active code elements that have virtually the same denotation, one of them being a badly defined macro/collective code element.

A fourth option, one that appears less disruptive to the stability of the code set, has therefore been considered and approved by the ISO 639 Joint Advisory Committee: to allow the introduction of a new code element for a variety that was widely considered a dialect of another major language, without splitting (and retiring or altering) the code element for that major language. The following criteria were imposed for this situation to apply:

1. Recognition as a separate language is justifiable on linguistic sociolinguistic grounds, just as would apply to any proposed new code element;
2. Ethno- and sociolinguistic evidence demonstrates a distinct identity and a linguistic awareness of speakers such that the "dialect" or low variety is not generally a result of attempting to use the standard form with insufficient ability (though there may be some speakers of the low form who do not have a good command of the high form);
3. The proportion of known instances of use of the existing code element to actually be intended to mean the low variety in question can be reasonably estimated and is very low (on the order of 4%, or no more than 2 in 50).
4. Sources dependent on and informing the use of ISO 639-2 (e.g., the MARC Code List for Languages, the IANA Language Subtag Registry) do not explicitly link the variety in question to an existing code element.

Linguistically, the author who has written most directly on the question, K.S. Nagaraja, did not directly state whether he thought Lyngngam ought to be considered a dialect of Khasi or a separate language, preferring to allow the evidence to speak for itself. Other linguists more distant from the situation, when asked after reading his paper, were quite prepared to accept the linguistic evidence as strong enough to separate them. The situation of Lyngngam in relation to Khasi fits the second and third criteria, both in the research itself and in the metadata records of documented research. Finally there is no evidence from the IANA language subtag registry that anything other than standard Khasi has been presumed with [kha]. Likewise, the MARC Code List of Languages associates no other variety with [kha].

REFERENCES


