Improving African Languages Classification: initial investigation and proposal

Edgardo Civallero
UDC Editorial Team
Email: edgardo.civallero@udcc.org

Background

The importance of languages in the UDC is consistent with the significance of linguistic facets for knowledge organization in general. Languages are the main facet category implicated in processes as crucial as the development of the Linguistics class, the organization of national and regional literatures, the categorization of human ancestries, ethnic groupings and nationalities, and the description of the language in which a document is written. Language numbers are extensively used across the entire UDC scheme, and form the basis for a faceted approach in class structuring and number building. For this reason, Common Auxiliaries of Languages (Table 1c) deserve special attention.

In the 1980s, these auxiliary numbers underwent a major overhaul. The aim was to replace an obsolete, Western-biased and scientifically unacceptable language classification inherited from Dewey DC, which had attracted much criticism from experts through the history of the UDC. An improved and more detailed schedule was finally introduced in Extensions & Corrections to the UDC, 14 (3), 1989-1991, published in September 1992. The range of changes required to improve and update the old structure is evident from the outline of the old and the new classification shown in Figure 1.

Pre 1992 UDC Languages		Post 1992 UDC Languages	
=2	Western languages	=1/=2	Indo-European languages
=20	English	=1	Indo-European languages of Europe
=3	Germanic languages	=2	Indo-Iranian, Nuristani and dead
=4	Romance or Neo-Latin languages		Indo-European languages
=50	Italian	=3	Caucasian & other languages. Basque
=60	Spanish	=4	Afro-Asiatic, Nilo-Saharan, Congo-
=690	Portuguese		Kordofanian, Khoisan languages
=7	Classic languages. Latin and Greek	=5	Ural-Altaic, Japanese, Korean, Ainu,
=81	Slavonic languages		Palaeo-Siberian, Eskimo-Aleut, Dravidian, Sino-
=88	Baltic languages		Tibetan languages
=9	Oriental, African and other languages	=6	Austro-Asiatic. Austronesian languages
=91	Various Indo-European languages	=7	Indo-Pacific, Australian languages
=92	Semitic languages	=8	American indigenous languages
=94	Hamitic languages	=9	Artificial languages

Figure 1. *Top level structures of the old and new language classification in UDC*

Considering that the original classification structure had remained unchanged and in use for over 80 years, it was expected that a more representative structure would work better for a longer period without further revision. Indeed, we can see from Figure 1 a more consistent application of geographical organisation of the origins of the listed languages and a more inclusive listing. However, a number of theories, especially with respect to, previously less studied, languages of indigenous peoples[1] were revisited and re-examined during the last two decades of the 20th century, and there was a need to look into the classification of these languages again. On the one hand, linguists were able to reach

---

[1] These are frequently termed "indigenous languages" in the sense of languages spoken by indigenous peoples, i.e., all human groups which self-identify themselves as such (see also UN Permanent Forum on Indigenous Issues, http://www.un.org/esa/socdev/unpfii/index.html).

further into relatively unexplored linguistic territories; advanced tools assisted them in collecting comprehensive data on indigenous languages; latest findings and enhanced scientific methods allowed them to redefine the relationships between languages, linguistic groups and families. On the other hand, associated with the subsequent development of innovative theories supported by empirical data, a substantial, critical revision of the traditional phylogenetic classification was carried out.

As a result, previously ignored (dis-)connections between languages became apparent, leading to modifications in their ordering, grouping and interlinking. In addition, scholars adopted the vernacular names (and spelling) of languages; developed and applied a revised orthography; and included a great number of previously "marginalized" languages to represent better the global linguistic reality.

Revision work of the UDC common auxiliaries of languages (Table 1c), 2003-2010

Taking the above mentioned developments into consideration, it became necessary to examine some indigenous language families and make several changes in the arrays of UDC Language Table. The first area looked at was the area of South American indigenous languages. The research into these families was conducted from 2003-2007 by the author and resulted in a proposal for a revised schedule published in *Extensions & Corrections to the UDC*, 29 (Civallero, 2007). Upon the feedback received, minor corrections were brought in and the revised classification of South American languages was available in the UDC in 2008 (*Extensions & Corrections to the UDC*, 30, 2008: pp 54-63). Based on this experience and following the same requirements, the author revised North American language families, which were introduced in UDC in 2009 (*Extensions & Corrections to the UDC*, 31, 2009: pp. 54-63).

It is interesting to observe that among the most common changes carried out during the revision of American indigenous languages were corrections to classes that previously considered each linguistic entity as a single language instead of as a family or group of languages, as it is recognized today. In addition, a number of amendments in the naming of languages needed to be made: for instance, designations that are nowadays widely regarded as pejorative or offensive (often a legacy of colonial times) had to be replaced by appropriate and correct terminology. Occasionally there were also requirements to modify class structure in order to place language families and groups correctly.

Upon the completion of the revision of American indigenous languages, it was evident that other language families would benefit from the same careful examination. The next class we are looking to improve is =4, Afro-Asiatic, Nilo-Saharan, Congo-Kordofanian, Khoisan languages. Thus, in 2009 the author has started a research into indigenous languages of Africa. In this issue of the *Extensions & Corrections* we present the initial research findings and proposal for the improvement of the class =4.

The UDC classification of African languages seems to be largely based on Joseph Greenberg's work *The languages of Africa* (1963), still one of the most cited studies in the area. Despite of its popularity, in recent years Greenberg's scheme has received much criticism from scholars on its conceptual and structural approach, e.g. Newman (1977), Bendor-Samuel (1989), Hayward (1997), Hetzron (1997), Bender (2000), Williamson & Blench (2000) and, more recently, by Campbell and Poser (2008). The significant contribution of these scholars is reflected in some of the corrections implemented in this proposal:

a) Changes in hierarchical structure

The author proposes to rename a number of language arrays, and moved them from one group to another in order to reflect a more generally accepted linguistic view. These corrections also led to the change of the =4 caption, which is now intended to be more accurate: "Afro-Asiatic, Nilo-Saharan, *Niger-Congo*, Khoisan languages". Examples of the most important modifications required are listed below, arranged by class number.

a.1. Subdivisions within =41, "Afro-Asiatic languages" have been re-organized using span or range classes and updating class names, especially in =411, "Semitic languages" (Hetzron, 1997). Within this array, Hebrew has been placed among the Canaanite languages; the situation of extinct languages such as Amorite has been revised; and the labyrinth of Aramaic variants (evolving through both space and time) has been clarified following Beyer (1986). The inner structure of =413, "Berber languages", has been improved and updated according to Kossmann (1999) and Aikhenvald & Militarev (1991); a similar process has been carried out in =414, "Chadic languages", after Newman (1977). "Cushitic languages", =415, has been one of the most serious challenges, for it is a highly controversial group. Some authors argue that it should not be considered as a group, and many others continue discussing its internal composition. Since no clear evidence for radical corrections has been found yet, the existing classification has not been modified; some information has been updated and several minor changes introduced, e.g. in =415.2, "Central Cushitic (Agaw) languages" (according to Appleyard, 2006) and in =415.3, "East Cushitic languages" (according to Hayward, 1997). "Omotic languages", =416, still cause a lot of disagreement; whether it belongs or not to "Afro-Asiatic languages" has been questioned in recent years, and many influential authors (e.g. Newman, 1980, and Theil, 2006) believe that it should be regarded as an independent family. Until new evidence is available, "Omotic languages" will remain within "Afro-Asiatic languages", and has been organized according to Hayward's scheme (2003).

a.2. In general terms, Greenberg's 1963 classification of "Nilo-Saharan languages", =42, is still accepted as valid (Dimmendaal, 2008), though a set of languages (Kuliak, Kunama, Koman, Gumuz, Kadu, Meroitic) are under discussion. Groups' placement within the family has been arranged following Bender (2000). "Eastern Sudanic languages", =426, has undergone serious transformations, and its subdivisions have been organized according to Bechhaus-Gerst (1989, for Nubian), Fleming (1983, for Surmic), Thelwall (1981, for Daju), Vossen (1982, for Eastern Nilotic), Rilly (2007, for Meroitic), and Rottland (1982, for Southern Nilotic). "Kadu languages", thought to belong to current "Niger-Congo languages", has been placed in this class agreeing with Schadeberg (1981).

a.3. Old "Congo-Kordofanian languages" classification has been re-structured, its top class corresponding to =432, "Niger-Congo languages". The whole family has been revised following the guidelines provided by Bendor-Samuel (1989). Kordofanian, Mande and Dogon languages have been removed from this array and placed in new and independent ones (see a.4. and a.5. below). Old "Senegalo-Guinean (West Atlantic) group" (currently known as "(West) Atlantic languages", =432.1) is not regarded any longer as a genetically-related family (Wilson, 1989): until new evidence is brought forward, the family will be described as a group of three geographically-related branches with new names (proposed by Sapir, 1971). Since slight structural adjustments (e.g. grouping Fula and Wolof into a Senegambian branch) would have necessarily involved major changes, they were ultimately disregarded. "Senufo languages", =432.3 has been removed from within "Gur languages", taking the opinion of Naden (1989) and Williamson & Blench (2000) into account. Old "Kwa languages", =432.4/.5, has been drastically reduced according to Williamson & Blench (2000: 11-42): many of its families have been grouped in a new one ("Volta-Niger languages", =432.53/.58) and a few others have become independent groups on their own (e.g. =432.51, "Kru languages", classified as suggested by Marchese, 1989). "Akan languages", =432.413, has been internally re-organized. Old "Ewe languages" group (now "Gbe languages", =432.54) has been arranged (and renamed) after Capo (1988). "Adamawa-Eastern languages", =432.6, has been re-named and completely re-structured according to Boyd (1989) and Williamson & Blench (2000), and such subdivisions as Daka and Boa groups have been moved. "Benue-Congo languages", =432.7/.9, has also been remodelled after Williamson (1989). Finally, "Bantu languages", =432.8/.9, has not suffered any structural change: the geographical model of Guthrie (1948, 1971, updated by Maho, 2009) used in its classification is still recognized as a valuable tool.

a.4. "Mande languages", =433, are now considered to be outside Niger-Congo linguistic group. Therefore, a number of important changes have been implemented in order to reflect this new trend in the tables. Mande internal classification has been corrected according to Kastenholz (1996), and some details on "East Mande languages", =433.4/.5, were based on Williamson & Blench (2000).

a.5. The placement of Dogon languages within the Niger-Congo linguistic scheme is far from being clear. Although many different theories have been proposed, the one most supported today —and included in the UDC— considers them as an independent family. The internal description was developed following Hochstetler (2004).

a.6. "Khoisan" languages (African "click languages" not belonging to any other linguistic family) constitute a class suggested by Greenberg and soon rejected by many of his colleagues (cf. Sands, 1998). At present, "Khoisan" is used as a convenient, practical "container term" without any inner relationship. The families included are considered to be independent groups. Former "Northern Khoisan" or Juu languages have become Juu-Hoan or, agreeing with Heine & Honken (2010), Kx'a languages. "Central Khoisan" languages are now known as Khoe and divided into two main branches, following Güldemann & Elderkin (2003).

When the current trend in linguistic classification is taken on board and compared to the current UDC schedules, there are a number of challenges that have to be faced. This is especially so when it comes to classes based on geographical grouping of languages. For instance, languages in UDC currently classed as "Eastern Omotic languages" should be properly called "South Omotic languages" and "Western Omotic languages", "North Omotic languages". It is even more complex to replace the old class "East Chadic languages" with "Central Chadic languages" and "Sahel group" with "East Chadic languages". And probably the most complex are changes that are required in the case of reclassing "North-western Mande languages" into "Central Mande languages", "Southern Mande languages" into "South-east Mande languages" and "South-eastern Mande languages" into "Bisa-Busa" languages". Occasionally, the redirection note of the cancelled class to a new class, that we normally provide in UDC database, could appear unclear or hard to follow. If this was the case, a "superfluous" message appears in the comment next to the cancelled class. The care was taken that the relocation of individual languages and specific language groups is always expressed with an accurate redirection and 'replaced by' note.

b) Improvement of the class content presentation and vocabulary

"Including" notes have been extensively used to supply terminology for language groups and languages that had not been previously included in the scheme or were hidden under the (main) class name. In order to avoid over-expanding of the schedules, these notes are in principle used to identify place for languages with less than 10,000-50,000 speakers (depending on the group or family). Languages with more than 50,000-100,000 speakers (again, depending on the group or family) where provided with a class number of their own, e.g. =414.113.1, Kamwe; =415.21, Awngi; =415.314, Kambaata; =415.321.1, Afar; =415.323.1, Somali; =415.323.2, Maay; and =427.41, Lendu.

In addition, information notes (represented in the proposal with the abbreviation IN:) and scope notes (abbreviation SN:) have been introduced to supply general, relevant data on the most important African languages, including a brief geographical report, their "official" or "national" status, the use of indigenous alphabets (when available), and the number of speakers quoted by Ethnologue report (Lewis, 2009).

All these notes, which might be helpful at an early stage of the revision process, will have to be examined at a later stage to decide whether they should be retained or removed as superfluous.

c) Correction of language names

Many language groups have had their names adapted to new linguistic classification patterns, e.g. =411.3, "South-east Semitic languages", has been changed into "Old South Arabian languages"; =411.5, "Mahri-Sokotri group", into "Eastern South Semitic languages"; =426.72, "Didinga-Murle", into "Surmic languages"; =432.1, "Senegalo-Guinean (West Atlantic) group", into "(West) Atlantic languages"; =433.3, "Gur (Voltaic) languages", into "Gur languages"; =432.34, "Moré-Gurma group", into "Oti-Volta

languages"; =432.54, "Ewe group", into "Gbe languages"; and =432.413, "Akan (Twi-Fanté; Volta-Comoe) group", into "Akan-Bia (Central Tano) languages".

The "–oid" suffix has been added to the names of several groups. That is the case for =432.48, Ijoid; =432.531/.534, Nupoid; =432.535, Idomoid; =432.56, Yoruboid; =432.57, Igboid; and =432.58, Edoid.

Also, other language names have been modified to suit modern language naming standards, as it is the case with =45, "Khoisan languages" (to which have been applied new orthographic rules and patterns) or with =413, "Berber languages" (which have been written relying on modern spellings, e.g. Tamasheq, Taqbaylit, and Tarifit).

Derogatory names have been substituted, e.g. Hottentot or Bushmen in "Khoisan languages", or Teuso (actually, Ik, =428.5). Duplicate classes have been unified (e.g. Maban/Mabaan, connected to the same entity, =423) and other particular cases have been eliminated after careful analysis, e.g. "Mandekan", an artificial creation which refers to the core group of Mande family — Maninka, Jula and Mandinka.

Finally, some classes have been proposed for cancellation as concepts they represented were ambiguous, outdated and lacked sufficient literary warrant, e.g. =414.273. Tuburi or =414.28, Warjawa-Gesawa.

d) Addition of new languages and language variants

A number of the existing classes in the schedule have been expanded, e.g. =411.2, Arabic (turned into "Arabic languages" to include the many standard language variants); =413, "Berber languages" (a set of languages whose presence in northern Africa is particularly strong); =414, "Chadic languages"; or =432.37, "Senufo languages" (expanded according to Mensah, 1983, and Mills, 1984).

In addition, new languages and previously not mentioned groups have been introduced, e.g. recently discovered Dahalik, =411.47 (described by Simeone-Senelle, 2005), and =434, "Dogon languages". Finally, when it was thought to allow better identification, variants of names were added to captions, e.g. components of class =432.962, "Nguni languages".

Considering that the Table of Language Auxiliaries plays a very important role in the scheme besides being connected to many other areas (e.g. the above mentioned human ancestries, ethnic groupings and nationalities), our first priority when we made the decision to revise Table 1c was to alter the existing classification as little as possible, and avoid re-using any of the cancelled numbers in the Table 1c. However, in view of the importance of changes made in the field of linguistics, this purpose has not always been easy to achieve.

It was also our intention to respect the policy of taking care of languages and language groups, while leaving dialects and variants to be resolved with A/Z extensions. Making this distinction has not always been straightforward, and provision had to be made for some well justified exceptions. For instance, in case of variants in the process of distinguishing themselves through standardized forms independent of neighbouring standard languages (a goal achieved by being taught in schools or by being used as written language in a wide variety of official functions). Such is the case of Songhai dialects, which are used as *lingua franca* and in education; their increasing importance has led to provide them a class number of their own, =421.1/.3.

Conclusion

The work on class =4 revision will continue in 2011 and 2012. We plan to share and discuss these proposals with linguists and those documentalists who are familiar with the literature dealing with these languages as well as the literature written in these languages. The goal is to discover a workable consensus on the

way African indigenous languages should be organised. With this paper, we take the opportunity to invite comments and suggestions in relation to Table =4 presented below, either by contacting UDC Editorial Team (editorial@udcc.org) or the author of the proposal (edgardo.civallero@udcc.org) or our associate editor responsible for common auxiliaries of langauges Sylvie Davies (s.davies@rgu.ac.uk)

References

Aikhenvald, A.Y.; Militarev, A.Y. (1991). Livijsko-guanchskie jazyki [Lybic-Guanche languages]. In: *Jazyki Azii i Afriki* [Languages of Asia and Africa], vol. 4, pp. 148-266. Moscow: Nauka

Appleyard, D. L. (2006). A Comparative Dictionary of the Agaw Languages. In: *Kuschitische Sprachstudien* [Cushitic Language Studies], vol. 24. Köln: Rüdiger Köppe Verlag

Bechhaus-Gerst, M. (1989). Nile-Nubian Reconsidered. In: L. Bender (ed.). *Topics in Nilo-Saharan Linguistics.* Hamburg: Heinrich Buske

Bender, L. (2000). Nilo-Saharan. In: B. Heine; D. Nurse (eds.). *African Languages: An Introduction.* Cambridge: University Press

Bendor-Samuel, J.; Hartell, R. L. (eds.) (1989). *The Niger-Congo Languages - A classification and description of Africa's largest language family.* Lanham, Maryland: University Press of America

Beyer, K. (1986). *The Aramaic Language: its distribution and subdivisions.* Göttingen: Vandenhoeck und Ruprecht

Boyd, R. (1989). Adamawa-Ubangi. In: J. Bendor-Samuel; R. L. Hartell (eds.). *The Niger-Congo Languages - A classification and description of Africa's largest language family.* Lanham, Maryland: University Press of America, pp. 178-215

Campbell, L.; Poser, W. J. (2008). *Language Classification: History and Method.* Cambridge: Cambridge University Press, pp. 203-245

Capo, H. B. C. (1988). *Renaissance du Gbe: Réflexions critiques et constructives sur L'EVE, le FON, le GEN, l'AJA, le GUN, etc.* Hamburg: Helmut Buske Verlag

Civallero, E. (2007). South American indigenous languages: a proposal. In: *Extensions & Corrections to the UDC*, 29, pp. 45-58

Dimmendaal, G. (2008). Language Ecology and Linguistic Diversity on the African Continent. *Language and Linguistics Compass* 2 (5), p. 842

Fleming, H. (1983). Surmic etymologies. In: R. Vossen; M. Bechhaus-Gerst (eds.). *Nilotic Studies: Proceedings of the International Symposium on Languages and History of the Nilotic Peoples.* Berlin: Dietrich Reimer, pp. 524-555

Güldemann, T.; Elderkin, E. D. (forthcoming). On external genealogical relationships of the Khoe family. In: M. Brenzinger; C. König (eds.). *Khoisan languages and linguistics: the Riezlern symposium 2003. Quellen zur Khoisan-Forschung 17.* Köln: Rüdiger Köppe. Available at: http://email.eva.mpg. de/~gueldema/pdf/Gueldemann_Elderkin.pdf

Hayward, R. J. (1997). Review of Grover Hudson, Highland East Cushitic Dictionary, 1989. *Journal of African Languages and Linguistics*, 18, pp. 106-120

Hayward, R. J. (2003). Omotic. The 'empty quarter' of Afroasiatic Linguistics. In: J. Lecarme (ed.). *Research in Afroasiatic Grammar II. Selected Papers from Fifth Conference on Afroasiatic Language, Paris, 2000.* Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 241-261

Heine, B.; Honken, H. (2010). The Kx'a Family: A New Khoisan Genealogy. *Journal of Asian and African Studies. 79*, pp. 5-36

Hetzron, R. (ed.) (1997). *The Semitic Languages.* Routledge: London

Hochstetler, J. L.; Durieux, J. A.; Durieux-Boon, E.I.K. (2004). *Sociolinguistic Survey of the Dogon Language Area*. SIL International, online version. Available at: http://www.sil.org/silesr/2004/silesr2004-004.pdf

Kastenholz, R. (1996). *Sprachgeschichte im West-Mande: Methoden und Rekonstruktionen*. Köln: Rüdiger Köppe Verlag

Kossmann, M. (1999). *Essai sur la phonologie du proto-berbère*. Köln: Rüdiger Köppe

Lewis, M. P. (ed.) (2009). *Ethnologue: Languages of the World, Sixteenth edition*. Dallas: SIL International. Available at: http://www.ethnologue.com/

Maho, J. (2009). *A classification of the Bantu languages: an update of Guthrie's referential system*. Available at: http://goto.glocalnet.net/mahopapers/nuglonline.pdf

Marchese, L. (1989). Kru. In: J. Bendor-Samuel; R. L. Hartell (eds.). *The Niger-Congo Languages - A classification and description of Africa's largest language family*. Lanham, Maryland: University Press of America

Mensah, E. N. A.; Tchagbale, Z. (1983). *Atlas des langues Gur de Côte d'Ivoire*. Abidjan: Institut de Linguistique Applique

Mills, E. (1984). *Senoufo phonology, discourse to syllabe (a prosodic approach)*. SIL publications in linguistics, 72

Naden, A. J. (1989). Gur. In: J. Bendor-Samuel; R. L. Hartell (eds.). *The Niger-Congo Languages - A classification and description of Africa's largest language family. Lanham,* Maryland: University Press of America, pp. 140-168

Newman, P. (1977). Chadic classification and reconstructions. *Afroasiatic Linguistics, 5 (1),* pp. 1-42

Newman, P. (1980). *The classification of Chadic within Afroasiatic*. Leiden: Universitaire Pers

Rilly, C. (2007). *La langue du Royaume de Meroe*. Paris: Champion

Rottland, F. (1982). Die Südnilotischen Sprachen: Beschreibung, Vergleichung und Rekonstruktion. In: *Kölner Beiträge zur Afrikanistik,* vol. 7. Berlin: Dietrich Reimer, pp. 26, 138-139

Sands, B. (1998). *Eastern and Southern African Khoisan: Evaluating Claims of Distant Linguistic Relationships*. Cologne: Rüdiger Köppe

Sapir, D. (1971). West Atlantic: An inventory of the languages, their noun class systems and consonant alternations. *Current Trends in Linguistics, 7*, pp. 45-112

Schadeberg, T. C. (1981). The classification of the Kadugli language group. In: T. C. Schadeberg; L. Bender (eds.). *Nilo-Saharan: Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980*. Dordrecht: Foris Publications, pp. 291-305

Simeone-Senelle, M.-C. (2005). *Dahālík, a newly discovered Afro-Semitic language spoken exclusively in Eritrea*. Available at: http://llacan.vjf.cnrs.fr/fichiers/Senelle/Dahalik_shaebia05.pdf

Theil, R. (2006). *Is Omotic Afroasiatic? A Critical Discusion*. Available at: http://www.uio.no/studier/emner/hf/iln/LING2110/v07/THEIL%20Is%20Omotic%20Afroasiatic.pdf

Thelwall, R. (1981). Lexicostatistical Subgrouping and Reconstruction of the Daju Group. In: T. C. Schadeberg; L. Bender (eds.). *Nilo-Saharan: Proceedings of the First Nilo-Saharan Linguistics Colloquium, Leiden, September 8-10, 1980*. Dordrecht: Foris Publications

Vossen, R. (1982). *The Eastern Nilotes: Linguistic and Historical Reconstructions*. Berlin: Dietrich Reimer Verlag

Williamson, K. (1989). Benue-Congo Overview. In: J. Bendor-Samuel; R. L. Hartell (eds.). *The Niger-Congo Languages - A classification and description of Africa's largest language family*. Lanham, Maryland: University Press of America, pp. 248-274

Williamson, K.; Blench, R. (2000). Niger-Congo. In: B. Heine; D. Nurse (eds.). *African Languages: An Introduction.* Cambridge: University Press, pp. 11-42

Wilson, W. A. A. (1989). Atlantic. In: J. Bendor-Samuel; R. L. Hartell (eds.). *The Niger-Congo Languages - A classification and description of Africa's largest language family.* Lanham, Maryland: University Press of America