



HAL
open science

New Electronic Resources for Texts in Manding Languages.

Valentin Vydrin

► **To cite this version:**

Valentin Vydrin. New Electronic Resources for Texts in Manding Languages.. Searching For Sharing: Heritage and Multimedia in Africa. , 2017, 978-1-78374-318-6, 978-1-78374-321-6, 978-1-78374-320-9. halshs-01584112

HAL Id: halshs-01584112

<https://shs.hal.science/halshs-01584112>

Submitted on 8 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



<https://www.openbookpublishers.com>

© 2017 Daniela Merolla and Mark Turin.

Copyright of each chapter is maintained by the author.



This work is licensed under a Creative Commons Attribution 4.0 International license (CC BY 4.0). This license allows you to share, copy, distribute and transmit the work; to adapt the work and to make commercial use of the work providing attribution is made to the authors (but not in any way that suggests that they endorse you or your use of the work). Attribution should include the following information:

Daniela Merolla and Mark Turin (eds.), *Searching For Sharing: Heritage and Multimedia in Africa*. Cambridge, UK: Open Book Publishers, 2017. <https://doi.org/10.11647/OBP.0111>

In order to access detailed and updated information on the license, please visit <https://www.openbookpublishers.com/product/590#copyright>

Further details about CC BY licenses are available at <http://creativecommons.org/licenses/by/4.0/>

All external links were active at the time of publication unless otherwise stated and have been archived via the Internet Archive Wayback Machine at <https://archive.org/web>

Digital material and resources associated with this volume are available at <https://www.openbookpublishers.com/product/590#resources>

Every effort has been made to identify and contact copyright holders and any omission or error will be corrected if notification is made to the publisher.

World Oral Literature Series, vol. 7 | ISSN: 2050-7933 (Print); 2054-362X (Online)

ISBN Paperback: 978-1-78374-318-6

ISBN Hardback: 978-1-78374-319-3

ISBN Digital (PDF): 978-1-78374-320-9

ISBN Digital ebook (epub): 978-1-78374-321-6

ISBN Digital ebook (mobi): 978-1-78374-322-3

DOI: 10.11647/OBP.0111

Cover image: *Interview* (2008) by ICT4D.at, CC BY-SA 2.0. Image from Flickr, <https://www.flickr.com/photos/ict4d/3068125986>

All paper used by Open Book Publishers is SFI (Sustainable Forestry Initiative), PEFC (Programme for the Endorsement of Forest Certification Schemes) and Forest Stewardship Council(r)(FSC(r) certified.

Printed in the United Kingdom, United States, and Australia
by Lightning Source for Open Book Publishers (Cambridge, UK)

5. New Electronic Resources for Texts in Manding Languages

Valentin Vydrin

Introduction¹

Since 2009, there has been continuous work conducted on electronic resources for Manding languages in Western Africa. In 2011, a Bamana Reference Corpus (Corpus Bambara de Référence) was published online (Vydrin, Maslinsky, Méric *et al.* 2011–2017), and as of March 2017 it accounts for 3,846,094 words, about 700,000 of these being within the disambiguated subcorpus.² In November 2014, a Bamana Electronic Library was made available online.³ A Maninka Corpus was opened to the public in April 2016 (Vydrin, Maslinsky, Rovenchak *et al.*, 2016–2017), and a Maninka Electronic Library was open to the public at the end of 2016.⁴

1 This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (reference: ANR-10-LABX-0083).

2 All the texts in the Corpus are automatically annotated for part of speech tags and for French glosses; the automatic annotation is based on a lexical database (electronic dictionary), Bamada, and on a formalized set of morphological rules. Originally, in the automatically analyzed texts, more than 70% of all words were annotated ambiguously (i.e. more than one variant of analysis was produced by the program); recently, thanks to the improvement of our electronic tools, this rate was reduced to 60%. At the next stage, the texts are treated (disambiguated) by human operators who should have a good knowledge of Bamana language and Bamana grammar; they select the correct analyses among those suggested by the automatic analyser, or produce them (if no correct analysis is suggested by the program).

3 <http://cormand.huma-num.fr/biblio/>

4 <http://cormand.huma-num.fr/maninkabiblio/index.jsp>

Manding is a large language/dialect continuum in Western Sub-Saharan Africa (see Map 5.1). The entire Manding speaking population is close to forty million, placing it among the most important languages of Africa. Manding (in some publications, also stylised as Mandingo) is a generic name for a great number of language varieties, among which the biggest ones are Bamana/Bamanakan (also Bambara) in Mali, Maninka (also Malinke) in Guinea, Mali, Senegal, and Sierra Leone, Mandinka in Gambia, Senegal and Guinea-Bissau, and Jula in Côte d'Ivoire and Burkina Faso. These varieties are usually regarded as individual languages, and separate written norms are emerging in spite of certain harmonisation efforts by linguists.

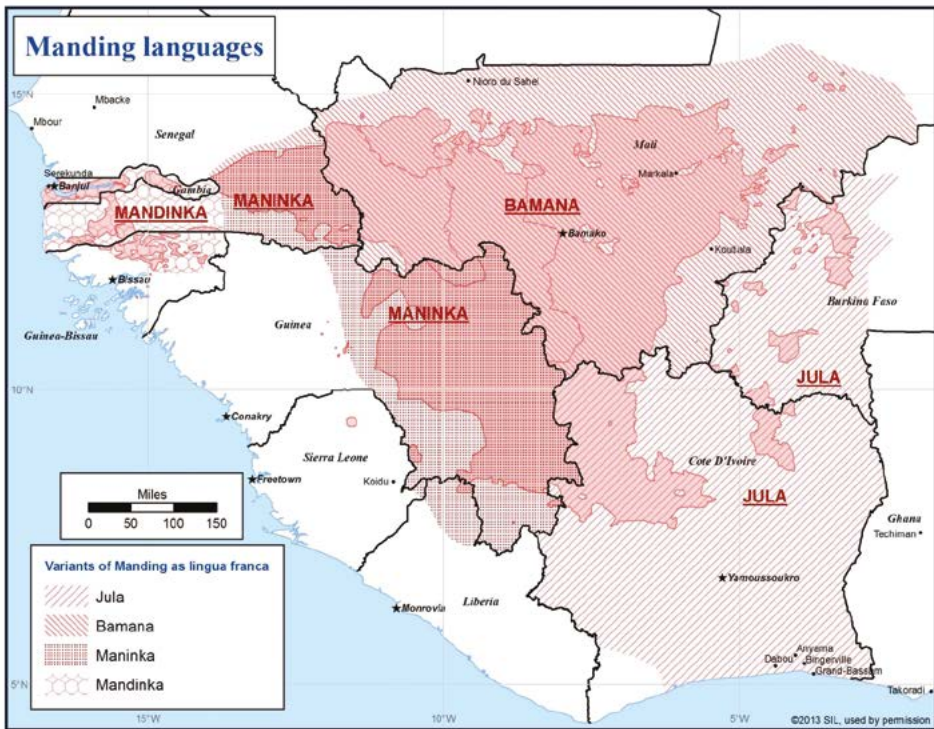


Fig. 5.1 Major Manding varieties. The light hatching covers the areas where Manding varieties are used as *lingua franca*. © 2004 SIL International, used with permission, redistribution not permitted.

In this chapter I will discuss the current situation regarding the collection and presentation of texts in the main Manding varieties in Mali (Bambara) and Guinea (Maninka). The resources in question deal mainly with written texts, and only marginally with audio texts. While this diverges from the main topic of this volume, it may still be of interest to the audience.

The Bambara Library

The availability of texts in Manding languages in Mali is ambiguous. On one hand, Bambara is the country's most prominent language, spoken by 80% of the population, and is well represented in electronic mass media. There is a written press in Bambara: the *Kibaru* monthly has existed since 1971, the *Jekabaara* monthly has existed since 1984, and a number of other periodicals have been launched (but disappeared more or less quickly). The number of books published in Bambara in Mali is considerable: in my bibliographic database, which is not exhaustive, there are five hundred Malian publications. While this may not be very impressive for a language with more than twelve million speakers (and even ridiculous if compared with the written output in any small language of Europe, such as Estonian or Latvian), it is still significant if we take into account the fact that Mali is firmly entrenched in the bottom twenty countries of the Human Development Index. Bambara does not have official language status in the Republic of Mali, is not used in administration, and is used only marginally in the education system. Among these books, booklets predominate; only around fifty books in my database exceed one hundred pages. Included in these are textbooks and religious books, but also a number of fiction books and collections of oral literature texts. While it is too early to discuss a fully-fledged written literature, an embryo literature does exist.⁵

Despite the hundreds of publications in Bambara, it is not an exaggeration to say that the language remains almost exclusively an oral one. Bambara publications are practically invisible in bookshops

5 For a survey of literatures in Manding languages, see the page *Littérature en mandingue* by Jean Derive on the ELLAF website, <http://ellaf.huma-num.fr/langues-et-litteratures/mandingue-2> (Derive 2016). For some reasons, in the first version of the site of ELLAF, there is no mention of the literature in Nko.

(which are quite rare even in Bamako, let alone in provincial centers), in the small number of libraries, and in schools. The periodicals in Bambara are absent from the newsstands. Households where books in Bambara can be found are few and far between. For the great majority of Malians, Bambara remains an unwritten language.

In rare instances where Manding is involved in a commercial sphere,⁶ namely in the transcription of song lyrics on the CDs of popular singers, Bambara is usually treated as a language without any written norm. The texts are written using a French-based “orthography”. Some examples of this are: *gn* is used for *n*, there is no distinction between the semi-closed and semi-open vowels *e* and *ɛ*, the vowel length is ignored, there is no standard regarding the rules of word segmentation, etc. Furthermore, when my students contacted a popular singer to offer help with the transcription of song texts, the reaction was negative and even hostile.

Evidently, the written variant of Bambara is trapped in a vicious cycle. Written documents in Bambara are (almost) unavailable, causing people not to consider Bambara as a written language. As a result, publications in Bambara have no audience, appearing rarely in tiny print runs and disappearing soon afterwards almost without a trace. Very little impact and no cumulative effect is observed.

The “Bibliothèque Électronique Bambara” project was conceived in June 2014 by scholars from Langage, Langues et Cultures d’Afrique Noire (LLACAN), a CNRS laboratory where I work, and the Académie Malienne des Langues (AMALAN) as an attempt to curb this negative tendency. The idea was very simple: to make as many Bambara documents as possible, of any genre, available in Open Access mode. In November 2014, the Bibliothèque Électronique Bambara website was launched.⁷

From the very beginning, the project was intended as not-for-profit; it had a non-existent budget. So far, I have scanned almost all of the documents posted on the site; a small number of the documents (in PDF format) have been contributed by their authors and editors. The

6 Otherwise, publications in Bambara in Mali are most often sponsored by official structures or by non-profit NGOs. Publishing houses producing Bambara books, as a rule, do it with grant money, therefore these publications do not really represent a profit-oriented sector.

7 See <http://cormand.huma-num.fr/biblio/index.jsp>. The site has been developed and is maintained by Tahar Meddour, a computer engineer from LLACAN.

website consists of three sections: “Ouvrages”, “Publications bilingues”, and “Périodiques”. At the site’s launch, 167 books were posted in the “Ouvrages” section, and among these 68 were downloadable with the others being in “partial display” mode (a couple of initial pages are shown in a non-downloadable format). These numbers are constantly growing.

In 2015, I obtained a “promotion grant” from the LabEx Empirical Foundations of Linguistics which allowed me to delegate the job of scanning Bambara (and also Nko) books to a student. When the extent of my personal library comes to an end, I plan to scan the personal library of Gérard Dumestre.⁸ I count on the help of other colleagues, as well as the public and university libraries in European countries, where Bambara books tend to remain safer than in the harsh Sahelian environment. Our AMALAN partners assist in the task of making publications (and other documents) available in the Library, and in handling the copyright questions with (mostly) Malian authors and publishers.

Most of the downloadable books are published by the Direction Nationale de l’Alphabétisation Fonctionnelle et Linguistique Appliquée (DNAFLA), which was the name of Mali’s national literacy agency until the end of 1990s, and by other governmental bodies whose work can be considered copyright-free (AMALAN is an heir of DNAFLA). For the other publications (about 60% of the collection), negotiations are ongoing with authors and publishers, and the “partial display” mode is a temporary solution that enables the books to be listed on the website. It is natural for authors and publishers to want to earn a profit through book sales, and for them to refrain from posting their material to access freely on the internet. However, the lack of dynamism within the Bambara book market makes profit-oriented publishing very difficult: Malians, with some rare exceptions, do not tend to buy books (apart from religious books, textbooks, and other publications necessary for their careers or another practical purpose). Therefore, after a relatively short period following publication, the copyright holder would be unlikely to lose any potential profit if the book is available to access for free on the internet. In any case, I try to convince the copyright holders to allow online access once a book runs out of print, in order to save it from oblivion.

8 Gérard Dumestre was the professor of Manding languages at INALCO until 2010; his personal library of publications in Bambara is probably the richest in the world.

Documents posted in the section “Publications bilingues” are academic publications which appeared mainly in Europe. For a while, this section has been very modest, but we hope to augment it with contributions from colleagues willing to make their works available for a wider audience. In the section “Périodiques”, almost all of the Bambara periodicals I know of are represented: *Kibaru*, *Jekabaara*, *Kalamene*, *Jama*, *Kolonkise*, *Nafarinma*, *Ntuloma*, *Netaa*, *Saheli*, *Sankore*, and *Faso Kumakan* (a weekly Bambara supplement to the official newspaper *l’Essor* during the 1980s). In November 2014, when the Bibliothèque was launched, about 180 issues were posted; this number is constantly growing. We are also planning to create a section called “First Publication”, where previously unpublished Bambara texts will be posted (for example, recordings of various genres of oral tradition).

The Cross-Fertilisation of the Bibliothèque Électronique Bambara and the Corpus Bambara de Référence

Who will use the Bibliothèque Électronique Bambara? The main visitors will be students of the Bambara languages in European universities (Paris, St. Petersburg, Bayreuth, Mainz, Köln, Wien, and others) and in North America (Bloomington, Chicago, Boston, for example), as well as fully-fledged Manding specialists. The big question is: will the Bibliothèque remain an “expatriate club” unnoticed by Malians, or will it also be consulted by at least some members of the Malian intellectual elite? The latter is not impossible; I have already received emails from Malians studying in France asking for various publications in Bambara, and I have redirected them, with pleasure, to the Bibliothèque website. But will it become popular among the Malians who visit internet cafés in Bamako and Ségou? Or among village dwellers who connect to the internet by the means of a GPRS or 3G modem? I hope that the Bibliothèque will play some role in the preservation of the Manding cultural heritage, and that at least some Malians will read the books and newspapers downloaded from our site.⁹

⁹ We began using Google Analytics to track the number of visitors to the Bambara Electronic Library in February 2016. During one month, more than 3,500 documents

The Bibliothèque Électronique Bambara is a small part of a global project of Bambara language documentation. The Corpus Bambara de Référence is the larger component, an annotated electronic corpus of texts: <http://cormand.huma-num.fr>. So far, to my knowledge, it is the only open-access mid-size corpus of a language from Sub-Saharan Africa. I hope, and expect, that the Bibliothèque Électronique Bambara will profit from a cross-fertilisation with the Corpus Bambara de Référence.

The Corpus is a much more time-consuming project than the Electronic Library, and it is also a powerful analytical tool. All of the texts included in the Corpus contain detailed metatextual information, and every word and morpheme is annotated to identify the tone¹⁰ and the part of speech characteristics, and provided with a gloss (a simplified French translation equivalent). For the major part of the Corpus (about 85%), the annotation is automatically performed by a morphological analyser based on the electronic Bamana-French dictionary "Bamadaba"¹¹ and a set of morphological rules. About 15% of the Corpus has been disambiguated semi-manually; the disambiguated subcorpus can be used for much more subtle searches.¹² In March 2017, the entire Bambara Corpus contained 3,846,094 words. At the same time, the disambiguated subcorpus consisted of about 700,000 words (these figures are constantly growing).

The Bambara Reference Corpus is, first and foremost, a powerful tool for linguistic studies. It allows for much more fine-grained research than in the pre-corpus era. In particular, the Corpus was used intensively to elaborate the proposals for the Bambara orthography (Konta and Vydrin 2014), while many questions which would otherwise require

were downloaded or consulted; there were 110 visits by 83 visitors. The distribution of visits was as follows: France 48, Mali 10, Russia 10, US 6, UK 5, Germany 3, Guinea 3, Sudan 3, etc. The total number of visits from African countries was 22, i.e. 20%.

10 Bambara is a tonal language, but in the published texts tones are usually not marked.

11 <http://cormand.huma-num.fr/bamadaba.html>

12 To illustrate the difference between the disambiguated and non-disambiguated subcorpora, it is enough to say that about 70% of all the words in a non-disambiguated Bambara text (originally without tonal marks) have two or more variants of translation. This is due to a set of factors, such as: abundant homonymy and (for toneless texts) quasi-homonymy; extremely scarce inflectional morphology and productive parts of speech conversion; and very productive word compounding.

tedious studies and long discussions found more or less quick and convincing solutions.¹³ It is used heavily in Bambara language teaching in European universities, with great effectiveness. A path is open for a corpus-driven Bambara dictionary, where every word and meaning will include statistical information about its representation in the Corpus. A corpus-driven dictionary would be labour-intensive, but at the same time a very gratifying project.¹⁴

It would be wrong to think that the Bambara Corpus is of interest only to linguists. It can be used by specialists in all adjacent fields dealing with Bambara texts and terms: scholars from cultural, social and even political anthropology who participate in the study of written and oral literature all need to know the preferred spelling of a term from time to time. The Corpus can also be of great value since the meaning of a word may change. Various written sources are given, beginning in the 1970s (even earlier documents could be included, although such documents are rare), and special efforts are put into making the Corpus representative, as much as possible, for different periods. One can easily imagine a corpus study of the evolution of the use and meaning of key words in various fields, e.g. *f̄roba* (which in the March 2016 version of the Corpus had 208 occurrences) evolved from connoting “collective farm” to “public (property)”; or *j̄ahadi*, *j̄ihadi* (45 occurrences) which changed from meaning “effort; holy war” (in Islam) to “catastrophe,

13 For example, the spelling of the names of deciles (the first number in each group of ten) from thirty to ninety is illustrative. These names are formed by adding the element *b̄i* in preposition to the respective names of the units. This element is tonally autonomous, which is mainly a characteristic of separate words, but no other word can be inserted between *b̄i* and the name of the unit (which is an argument in favour of the single-word interpretation). Among linguists, opinions diverge and discussions are sometimes quite fierce; in the texts, both spellings (one word or two) can be found. A corpus study in March 2016 (which took me about ten minutes) produced the following result: “30” *bi saba* – 215 occurrences, *bisaba* – 23; “40” *bi naani* – 235 occurrences, *binaani* – 32; “50” *bi duuru* – 189 occurrences, *biduuru* – 15; “60” *bi w̄ɔɔɔ* – 83 occurrences, *biw̄ɔɔɔ* – 7; “70” *bi wolonwula* – 114 occurrences, *biwolonwula* – 0; “80” *bi segin* and *bi seegin* – 66, *bisegin* and *biseegin* – 4; “90” *bi k̄ɔɔntɔn* and *bi k̄ɔɔntɔ* – 26, *bik̄ɔɔntɔn* – 1. The total number of split forms in the Corpus is 928, while one word spelling occurred only 82 times. Therefore, in practice, Bambara authors overwhelmingly prefer to use two words, which was also recommended by Konta and Vydrin (2014).

14 A three-million-word corpus is big enough for the study of grammar, but it is rather small for research on lexical semantics. For a medium-sized corpus-driven dictionary, at least a six- or seven-million-word corpus is usually deemed necessary.

disaster”; or yet *sáraka* (182 occurrences), which developed from “alms” to “sacrifice” or “to cast a spell”.

The statistics that track visits to the Bambara Corpus show its growing popularity. In 2015, the counter recorded 861 individual visitors,¹⁵ and 1591 visits in total. The peak of its popularity was reached in December (176 individual visitors, 404 total visits), and during the year, we observed a steady growth (with a predictable lull during the summer months). During March 2016, the number of unique visitors reached 500, and the total number of visits was close to 800. The vast majority of visitors are from Europe (mainly France, Germany and Russia), and visits from African countries are not frequent (unless they are hidden in the large category “Unknown”). Evidently, the Bambara Corpus is not only visited by specialists in the Bambara language, as the number of students of Bambara in the countries of Europe and North America is certainly far below the number of visitors to the website. We can conclude that this tool is being used more and more by specialists from adjacent disciplines and by a wide audience interested in Mali and the Bambara language in particular.

The evolution of the Corpus Bambara de Référence (besides a further growth of its size) is planned in three steps as follows:

- Building an audio-corpus of Bambara texts. Initial steps in this direction have been taken recently: Jean Jacques Méric is working on the software conversion necessary to synchronise the audio and the written texts. Building an audio-corpus is considerably more time-consuming than a written corpus, but these extra efforts allow sound, if some computer difficulties are solved, and graphs provided by speech analysis programs.
- Building a parallel subcorpus of texts (mainly Bambara-French). Currently all words and morphemes in the Bambara Corpus are annotated with part of speech tags and with French glosses), but no free translations of phrases are available. This means the use of the Corpus by those who have no command of Bambara is

¹⁵ This figure was noted down on 27 December 2015. However, there were two gaps, in April-May and September-October, when the counter did not work. The actual number of individual visitors might therefore have been well above 1,000 (the number of total visits should be increased accordingly).

difficult. In a parallel corpus, every Bambara sentence will be provided with its French equivalent. Even for those who speak Bambara perfectly, such a subcorpus will facilitate the possibility of searching for French idiomatic expressions, for example. A parallel corpus has great potential for language technologies, and in particular, for the development of automatic translation programs.

- Development of an automatic statistic-based disambiguation tool. Such a tool will reduce the rate of ambiguous annotations in the main subcorpus. The first steps for this were taken in 2015.

The Maninka Library for Guinea

We also intend to expand the existing model of electronic resources to closely related languages, starting with Guinean Maninka. The situation in Guinea regarding literature and other sources in the Manding languages is rather peculiar. During the period of the First Republic (1958–1984), an attempt was made to introduce local languages in education as much as possible, as well as in all other spheres of life. However, this initiative was not properly prepared. In my collection, there are only seven books in Maninka published during this period: school textbooks, primers for adults, and functional literacy books about stockbreeding; no fiction or oral literature is available. Most likely, other publications did exist, but disappeared without a trace. After the death of the first president Sékou Touré, education in the national languages was abandoned altogether; it was maintained only for small-scale adult literacy programs, and only two to three dozen Maninka books have appeared in Guinea since 1984 using the reformed Latin script.¹⁶

However, there is a written tradition in Nko writing, in which an original alphabet for Manding has existed since 1949 when it was introduced by Solomana Kantè. The alphabet began to flourish after 1984, mainly in Guinea, but also in neighbouring countries.¹⁷ Today, the number of people literate in Nko can be counted in the hundreds of

16 Letters for semi-open vowels *è, ô* were replaced by *ε, ο* respectively; the digraphs *ty, dy, ny* by single letters *c, j, n*.

17 The history and practice of Nko has been dealt with in numerous publications, for instance Amselle (2001); Oyler (2001); Vydrine (2001, 2011); Wyrod (2008).

thousands; there are several periodicals in Nko,¹⁸ and there are more than four hundred Nko book titles in my database.

In co-operation with Guinean colleagues from the IRLA (Institut de Recherches Linguistiques Appliquées) and *Ŋkó Dúnbu* (the Nko Academy), a Maninka Electronic Library project which includes books and periodicals was open for public at the end of 2016 (<http://cormand.huma-num.fr/maninkabiblio/index.jsp>). It follows the same model as the Bambara Library. In contrast to Bambara Latin-based writing, Nko enjoys broad popularity. The publication of books in Nko is often sponsored by individuals and the books are distributed as photocopies all over Guinea. They are often available in provincial Nko bookshops years after publication, and many people keep them in their own collections. Undoubtedly, the preservation of Nko literature is much more advanced than that of Bambara literature, although this does not mean there is no need for an electronic library.

An important question in relation to such an open-access resource is the attitude of the Nko authors. Among Nko authors and publishers, the question of copyright is significant, and cases of copyright-related conflicts concerning the intellectual heritage of Solomana Kantè have been attested. Taking into account the often mistrustful attitude of some Nko militants towards the Western world, a reserved attitude of certain authors toward the project is understandable. However, the Nko movement is far from homogeneous. Side by side with hardened Third World activists who distrust external initiatives, one can find inviting people — often from a western academic background — who have a positive experience of co-operation with outsiders. Fortunately, the attitude of *Ŋkó Dúnbu* (the Nko Academy) has been extremely positive towards our projects from the very beginning, and its Academic Secretary, Ibrahima Sory II Condé, has contact authors, collected electronic versions of their books, and requested permission for the inclusion of their works in the Electronic Library. It is also important that from the very beginning, the Electronic Library has been conceived as a joint project with *Ŋkó Dúnbu*, which helped to convince those who were initially reluctant and feared a wicked scheme designed to rob the Nko people of their intellectual wealth.

18 In 2012, *Dálilu Kénde* became the first weekly publication in Maninka, and for a while it managed to surpass all other periodicals in Guinea (appearing in French) with its print-run of 3,000 copies.

Conclusion

When this chapter was first composed, I wrote: “A Maninka Corpus can also be developed in the future”. At the time of publishing, this corpus is available online (since April 2016) and Open Access at <http://cormand.huma-num.fr/cormani>. It includes two subcorpora, “Corpus Nko”, which includes texts originally created in Nko writing, and “Corpus Manika”, originally in the Latin alphabet. The Nko subcorpus has reached 3,122,178 words, and the Latin-based Maninka subcorpus contains 396,389 words.¹⁹ Both Nko and Maninka subcorpora are available in Nko and in Latin-based alphabets (the converters have been developed by Andrij Rovenchak). The software package Daba described earlier, devised by Kirill Maslinsky for the Bambara Corpus, has been adapted by him for the Mandinka Corpus. Kirill Maslinsky and Andrij Rovenchak, who has done excellent work in converting Nko publications into electronic texts, have combined three sources for the development of the Maninka electronic database: an enlarged version of Vydrin’s dictionary (Vydrine 1999); a word frequency list generated on the basis of the preliminary version of the Nko Corpus containing two million words; and an electronic version of the French-Nko Dictionary (Kantè 2012). The resulting electronic dictionary, Malidaba, is used for the morphological analysis of the Maninka/Nko Corpus. A “cleansing” of Malidaba is being carried out, which consists mainly of the elimination of duplicates and providing all the entries with French, English and Russian glosses. In April 2016, when the first version of the Maninka Corpus was published online, some 15% of the Malidaba had been cleansed; in March 2017, when the second update of the Corpus took place, this rate reached 54%.

Thanks to computer technology, written versions of African languages have been exposed to a rapid progress. The Bambara and Maninka projects may serve as models for other languages of the region.

19 A large part of the collection and conversion (first into UNICODE fonts, then from Nko to Latin transliteration) was done by Andrij Rovenchak. A great number of texts were contributed in different electronic formats by members of the Nko Academy (*Ŋkó Dúnbu*) and other authors; this collection was organised by Ibrahima Sory II Condé.

References

- Amselle, J. L. (2001) *Branchements. Anthropologie de l'universalité des cultures* (Paris, Flammarion).
- Derive, J. (2016) "Littérature en mandingue", <http://ellaf.huma-num.fr/langues-et-litteratures/mandingue-2>
- Kantè, S. (2012) *Dictionnaire bilingue français-n'ko/Kodofolan Faransi-N'ko* (Conakry, Académie N'ko).
- Konta, M. and Vydrin, V. (2014) "Propositions pour l'orthographe du bamanankan", *Mandenkan* 52: 22–54.
- Maslinsky, K. (2014) "Daba: A Model and Tools for Manding Corpora", in: Mangeot, M. and Sadat, F. (eds.) *Actes de l'atelier sur le traitement automatique des langues africaines TALAf 2014*, <http://talaf.imag.fr/2014/Actes/MASLINSKY - Daba%3B a model and tools for Manding corpora.pdf>
- Oyler, D. (2001) "The Era of Mande Enlightenment", *Mande Studies* 3: 75–94.
- Vydrin(e), V. (1999) *Manding-English Dictionary (Maninka, Bamana)*, vol. 1 (St. Petersburg, Dmitry Bulanin Publishing House).
- (2001) "Soulemane Kantè, un philosophe-innovateur traditionaliste maninka, vu à travers ses écrits en Nko", *Mande Studies* 3: 99–131.
- (2011) "L'alternative du N'ko: une langue écrite mandingue commune, est-elle possible?", in: Lexander, K. V., Lyche, C. and Moseng Knutsen, A. (eds.) *Pluralité des langues pluralité des cultures: regards sur l'Afrique et au-delà* (Oslo, The Institute for Comparative Research in Human Culture): 195–204.
- (2012) "Une bibliographie préliminaire des publications maninka en écriture N'ko", *Mandenkan* 48: 59–121.
- Vydrin V., Maslinsky K., Méric J. J. *et al.* (2011–2017) Corpus Bambara de Référence, <http://cormand.huma-num.fr>
- Vydrin, Maslinsky, Rovenchak *et al.* (2016–2017) Corpus Maninka de Référence et Corpus N'ko, <http://cormand.huma-num.fr/cormani/projet.html>
- Wyrod, C. T. (2008) "A Social Orthography of Identity: The N'ko Literacy Movement in West Africa", *International Journal of the Sociology of Language* 192: 27–44.