

Where corpus methods hit their limits: the case of separable adjectives in Bambara

Valentin Vydrin

► **To cite this version:**

Valentin Vydrin. Where corpus methods hit their limits: the case of separable adjectives in Bambara. Rhema, Moscow State University of Education, 2018, 10.31862/2500-2953-2018-4-34-49 . halshs-02011349

HAL Id: halshs-02011349

<https://halshs.archives-ouvertes.fr/halshs-02011349>

Submitted on 13 Feb 2019

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.

DOI: 10.31862/2500-2953-2018-4-34-49

В.Ф. Выдрин

Национальный институт восточных языков и цивилизаций,
75214 г. Париж, Франция;

Национальный центр научных исследований Франции,
94800, г. Вильжюиф, Франция;

Санкт-Петербургский государственный университет,
19034 г. Санкт-Петербург, Российская Федерация

Там, где корпусные методы теряют эффективность: пример разделяемых прилагательных в бамана

Разделяемые прилагательные в языке бамана (< группа манден < семья манде < макросемья нигер-конго, Мали) представляют собой морфологический подкласс в составе части речи прилагательное. Разделяемое прилагательное – это сложное слово, состоящее из именного корня, чаще всего со значением части тела, качественного глагола и соединительного элемента *-la- ~ -lan-* или *-ma- ~ -man-*. В предикативном употреблении финальный компонент (т.е. качественный глагол) отделяется от левой части вспомогательным словом (предикативным показателем) *ka* или *man*. Разделяемые прилагательные обозначают в основном человеческие качества (моральные или физические), их семантика нередко идиоматична. Продуктивность этого морфологического подкласса ограничена.

Для выявления инвентаря разделяемых прилагательных были применены два подхода: элицитация и поиск в Справочном корпусе бамана (который на момент исследования включал около 4 110 000 слов). Теоретически возможное количество разделяемых прилагательных – 570 (15 зафиксированных именных корней × 19 качественных глаголов × 2 коннектора). Элицитация позволила обнаружить 75 разделяемых прилагательных, а корпусное исследование – 25, три из которых отсутствовали в списке, полученном путем элицитации.

Этот эксперимент доказывает, что при изучении деривационной морфологии, когда лингвист должен заполнять матрицу, корпусное исследование не может заменить элицитацию. С другой стороны, корпусной подход дает ценные дополнительные данные, которые невозможно получить путем элицитации.

Ключевые слова: элицитация, корпусные исследования, прилагательные, качественные глаголы, язык бамана, языки манде.

DOI: 10.31862/2500-2953-2018-4-34-49

V. Vydrin

National Institute for Oriental Languages and Civilizations,
Paris, 75214, France;

French National Centre for Scientific Research,
Villejuif, 94800, France;

St. Petersburg University,
St. Petersburg, 199034, Russian Federation

Where corpus methods hit their limits: The case of separable adjectives in Bambara

Separable adjectives represent a morphosyntactic subcategory of the part of speech of adjectives in Bambara (< Manding < Mande < Niger-Congo, Mali, West Africa). A separable adjective is a compound lexeme consisting of a noun root designating most often a body part, a qualitative verb root and a connector *-la-* ~ *-lan-* or *-ma-* ~ *-man-*. When used predicatively, the final component of a separable adjective (the qualitative verb root) is split from the rest of the form by the auxiliary word *ka* or *man*. Separable adjectives express mainly human qualities (moral or physical), and their semantics are very often idiomatic. The productivity of this subclass is limited.

In order to establish an inventory of the separable adjectives, two approaches have been followed: elicitation and a search in the Bambara Reference Corpus (which included roughly 4,110,000 words at the time of this study). The potentially imaginable number of lexemes of this type equals 570 (15 noun roots × 19 qualitative verb roots × 2 connectors). Elicitation provided 75 separable adjectives, and the corpus study, 25, 3 of which are absent from the elicited list.

This experiment proves that in studies of derivative morphology, when a linguist needs to fill out a matrix, elicitation cannot simply be replaced by a corpus study. On the other hand, the corpus data provides invaluable supplementary data that cannot be obtained through elicitation

Key words: elicitation, corpus study, adjective, qualitative verbs, Bambara, Mande languages.

1. Introduction

In works on field linguistics methodology and in the current linguistic practice, elicitation is sometimes regarded as a rather inappropriate way of acquiring language data, and elicited data is viewed as second-rate. Some colleagues tend to reject the elicitation altogether, in a more or less explicit way, and the following quotation is representative of this trend:¹

Interview fieldwork is justified if there is nothing else to be done.

It is a very poor option if a speech community is available – but some researchers opt to concentrate on interview fieldwork with a few speakers conveniently placed in a city or in a township. A grammar of a language spoken by a few million people which is based on the work with one consultant in an urban environment could be interesting, but is unlikely to be comprehensive and fully reliable [Aikhenvald, 2007, p. 5].

According to this approach, only natural texts, only spontaneous data can be regarded as reliable, and any use of elicitation is just a desecration of linguistic fieldwork.

This negativity is a natural reaction to an inappropriate and maladroit application of elicitation methods. The position of other authors may be less categorical. Most often, they recognize the usefulness of elicitation for obtaining certain types of language data; cf. a detailed analysis of various elicitation methods in [Chelliah, De Reuse, 2011].

The value of the data from natural texts cannot be contested (although here too, one should beware of ungrammatical forms spontaneously used by speakers). There are however some aspects of language structure where one can hardly attain satisfactory results without elicitation. Among such spheres are word formation and derivation, verbal lability, fine-grained syntactic studies; in fact, any research topic where an exhaustive checking of numerous options in a matrix is necessary.

In this sense, I recollect Vladimir Nedjalkov's (p.c.) position on the utility of elicitation in the verbal derivation studies: "If you check a matrix with your language consultant, and if even in 20% of cases he produces wrong answers or fails to answer, you obtain 80% of correct data. And if you work exclusively with natural texts, you are lucky if you obtain data for 30 or 40% of positions in the matrix".

I heard this argument from Vladimir Nedjalkov some 25 years ago, when big electronic text corpora were rare. Since that time, the progress in language documentation has been impressive. After the world-biggest languages, many mid-size languages were provided with multimillion text corpora accessible

¹ In informal discussions, much more categorical judgements are often expressed.

on line, and now, more and more new corpora for minor languages and those without official status in their countries, are becoming available.

Could it be that the easy access to great amounts of searchable natural texts has made Vladimir Nedjalkov's stance obsolete?

In this study, I attempt to answer this question drawing on Bambara data for the averagely-productive word-compounding model for adjectives.

2. The Bambara language

Bambara (< Manding < Western Mande < Mande < Niger-Congo) is the biggest language of Mali (West Africa). It is spoken, mainly in Mali (but also in its diaspora), by some 4 million L1 speakers and at least 10 million L2 speakers. Bambara has some written literature and periodicals; it is widely used in the literacy programs and, to some extent, in the primary and secondary education. Bambara is a relatively well-described language: there is a reference grammar [Dumestre, 2003], a number of university courses and textbooks [Bird, Kante, 1976; Bird, Hutchison, Kante, 1977; Kastenholz, 1989; Bailleul, 2000; Vydrin, 2008], big dictionaries [Vydrine, 1999a; Bailleul, 2007; Bailleul et al., 2011; Dumestre, 2011]; many dozens (or even hundreds) of research articles have been published.

Since 2011, there exists an electronic annotated corpus of Bambara texts freely accessible on line [Vydrin, Maslinsky, Méric, 2011]. In 2011, it contained about 1,100,000 tokens (of these, about 28,000 tokens in the disambiguated sub-corpus); in November 2018, it reached the size of 9,146,875 tokens (of these, 1,122,416 tokens in the disambiguated sub-corpus), and it continues to grow. A language corpus of some 9 million tokens may seem unimpressive when compared with corpora of big and even mid-sized European or Asian languages which comprise hundreds of millions words; however, for an African language almost unrepresented on the Internet, this amount of data represents a revolutionary breakthrough and opens bright perspectives for language studies. Since its publication in 2011, the Bambara Reference Corpus is broadly used in Bambara grammar studies, lexicographic research and language teaching.

Bambara is a tonal language with two tones at the underlying level, low and high.² The basic word order is S AUX (O) V X, where S is a subject,

² The Bambara tonal system has been subject of numerous studies; for its compact presentation, see [Vydrin, 2016]. In the present paper, the tonal notation is phonological, it follows the principles formulated for the Bambara Reference Corpus [Vydrin, Maslinsky, Méric, 2011b], in short: tone markers appear only on the initial syllables of the tonal domain; absence of tonal mark on a vowel means that the syllables belongs to the same tonal domain as the preceding one.

Otherwise, all Bambara examples are transcribed according to the official Bambara orthography of Mali.

V is a verb, AUX is an auxiliary word expressing grammatical semantics of tense, aspect, mode and polarity (in the Mandeist tradition, AUX are named “predicative markers”), O is direct a object (whose absence makes the verb intransitive), X is an oblique (indirect object or adjunct), most often represented by a postpositional phrase. The word order in NP is $N_2 - N_1$ (N_1 is head noun, N_2 is dependent noun), N – Adj (the adjectival modifier follows the head noun).

3. The case study: Separable adjectives in Bambara

3.1. Adjectives and qualitative verbs

Adjectives in Bambara represent a part of speech of its own [Vydrine, 1999b; Tröbs, 2008].³ An adjective follows the modified noun (1a); the tonal article (designated by a suspended acute diacritic) and the plural marker /-u/ (in the standard orthography represented by -w) follows the adjective (1b).⁴

- (1) a. *só* *jé*
 house white\ART⁵
 ‘white house’
- b. *só* *jé-w*
 house white\ART-PL
 ‘white houses’

There are several morphological subclasses of adjectives; some of these are tonally compact with the modified nouns (i.e., they are prosodically non-autonomous), other subclasses are non-compact.

There is another class of lexemes in Bambara specialized in the expression of property values: qualitative verbs.⁶ They can be regarded as a separate part of speech or as a subclass of verbs. Their syntactic behavior is similar to that of “dynamic verbs” (i.e. all the other verbs of the language), but

³ The part-of-speech status of adjectives in Bambara has been amply discussed in the special literature, and it is hardly appropriate to resume this discussion here.

⁴ The inverse word order, adjective-noun, appears in the inersive construction [Dumestre, 1987, p. 249–259; Vydrin, in press, leçon 30]. The article and the plural marker follow the noun immediately when the adjective is used as secondary predicate [Vydrin, in press, leçon 31]. Both these constructions are relatively rare in texts and represent no special interest for the topic of the current study.

⁵ Abbreviations and glosses: ADR – addressative postposition; ART – tonal article; INF – infinitive marker; IPFV – imperfective; NMLZ – nominalization suffix; NP – noun phrase; PL – plural marker; QUAL.NEG – negative predicative marker of qualitative verbs; RECP – reciprocal pronoun; REL – relative marker (pronoun or determinative).

⁶ Also referred to as “stative verbs” [Creissels, 1985] or “predicative adjectives” [Vydrine, 1990; Baillieu, 2007]. In the terminology of Gérard Dumestre (1987; 2003), these lexemes are just “adjectives”.

- they can be only intransitive (while dynamic verbs are very often labile, and intransitiva tantum are relatively rare among the dynamic verbs),
- they can be accompanied by only two predicative markers, *ka* affirmative and *man* negative, which express no tense, aspect or modal semantics. To the contrary, the dynamic verbs can appear with a whole set of predicative markers expressing various TAM meanings, but not with *ka* and *man*.

The qualitative verbs represent an unproductive closed class of about 60 lexemes (of these some 40 are frequently used, the others are more or less rare). Many qualitative verbs produce adjectives by the means of conversion.

3.2. Morphosyntax of separable adjectives

Separable adjectives represent one of the tonally non-compact classes of the adjectives. They are formed according to the following formula: N + Conn + QV, where:

- N is a nominal stem, most often the name of a body part (or a term belonging to a semantically adjacent lexicon);
- Conn is a connector *la* (variants: *lan*, *na*, *nan*) or *ma* (a variant: *man*), both connectors stem from locative postpositions;
- QV is a stem of a qualitative verb.

The forms of connectors *-ma-* and *-man-*, *-la-* and *-lan-* are, in fact, phonetic variants (*-na-* and *-nan-* are in complementary distribution with *-la-* and *-lan-*: they appear when the preceding component ends on a nasal vowel); there are sometimes duplicates, as for example *bólomandəgə* and *bólomadəgə* ‘poor, weak’. There seems to be no evident semantic difference either between the connector *-la-* and *-ma-*, their distribution is lexical.

These adjectives express mainly human properties (physical or moral), their meanings are often (but not always) idiomatic (2a). They are very easily convertible into nouns (2b) designating persons imparted with the quality in question.

- (2) a. *məgə* *dá-la-fəgən*
 human mouth-in-light
 ‘indiscreet person’
- b. *dálafəgən*
 ‘indiscreet person’

A peculiar morphosyntactic feature of the separable adjectives (not attested for other word classes) is that, when used predicatively, their final component is separated from the rest by the predicative marker of qualitative verbs *ka* or *man*, so that the initial bicomponent constituent appears as a part

of the subject NP, but it carries no tonal article (even in the contexts where, normally, one would expect the article). Cf. an adjectival attributive use (3a) and a predicative use (3b).

(3) a. *Nin báara` in man kán kà ké kó*
 this work\ART this QUAL.NEG equal INF do matter
sèn-na-teli yé.
 foot-in-quick like
 ‘This work should not be done hastily’ (lit.: ... should not be transformed into a fast matter’) [Kibaru 391]⁷

b. *Bènbaliya` sèn-na ka teli mɔ̀gɔ̀-ʷ*
 disagreement\ART foot-in QUAL.AFF quick person\ART
kála-li` mà jɔ̀gɔ̀n ná.
 sew-NMLZ\ART ADR RECP at
 ‘A discord is more rapid than a reconciliation among the people’
 [Jekabaara 136].

The element *sènna* cannot be used in any other syntactic context as an autonomous word; it is a quasi-lexeme.

3.3. Inventorization of separable adjectives

The separable adjectives were described for the first time by Gérard Dumestre who tried to inventory the lexemes of this class [Dumestre, 1987, p. 235–248]. According to him, 15 nominal roots and 11 qualitative verbs roots take part in the word formation of this type. Since then, more items have been found (mainly through elicitation), and currently [Vydrin, in press, Leçon 29], we have 15 nominal roots and 19 qualitative verbs in play.

The nouns in question are:

bólo ‘hand/arm’, *dá* ‘mouth’, *dùsu* ‘heart’, *jà* ‘shadow’, *jòli* ‘blood’, *jù* ‘bottom, buttocks’, *kólo* ‘bone’, *kɔ̀nɔ̀* ‘stomach’, *kùn* ‘head’, *ní* ‘soul’, *nún* ‘nouse’, *né* ‘eye’, *sèn* ‘foot/leg’, *téɛɛ* ‘palm (of hand)’, *túlo* ‘ear’.

The qualitative verbs are:

cá ‘numerous’ (in the separable adjective, the derived adjectival form appears, *cáman*), *dí* ‘pleasant’, *dɔ̀gɔ̀* ‘few’, *jé* ‘white’, *júgu* ‘bad, evil’, *fárin* ‘courageous, aggressive, strong’, *fěgen* ‘light (non-heavy)’, *fin* ‘black’, *gèlen* ‘hard, difficult’, *gírín* ‘heavy’, *gó* ‘unpleasant, unpalatable’, *gòni* ‘hot’, *kálan* ‘hot’, *kègun* ‘sly’, *kólon* ‘empty’, *kúnán* ‘bitter’, *misèn* ‘small’, *súma* ‘calm; cool’, *téli(n)* ‘rapid, fast’.

⁷ In square brackets, references to the sources of language examples are indicated. *Kibaru* and *Jekabaara* are periodicals published in Bambara in Mali.

The number of theoretically imaginable combinations of these components is 285. This number should be multiplied by 2 (the number of connectors), which brings us to the figure of 570. However, in reality, most of theoretically imaginable separable adjectives do not exist. The task is to establish the inventory of really existing items.

Elicitation [Dumestre, 1987, p. 239; Vydrin, in press, leçon 29] has produced 75 separable adjectives. This study was carried out with several native speakers during different periods of time; different informants produced similar results, therefore, they can be regarded as reliable. Here is the full list of the lexemes.

- bólolandi* <hand-in-pleasant> ‘careful, skilful’,
bólomadɔgɔ <hand-on-small> ‘poor; weak’,
bólolanjugu <hand-in-evil> ‘disorderly, unkempt; bad worker’,
bólomanjugu <hand-on-evil> ‘disorderly, unkempt; bad worker’,
bólolafegen, *bólolanfegen* <hand-in-light> ‘thievish; one who touches everything’,
bólomagelen <hand-on-difficult> ‘stingy’,
bólolango <hand-in-unpleasant> ‘miserly’,
bólolankolon <hand-in-empty> ‘poor’,
bólolamisen <hand-in-small> ‘thievish, light-fingered; who touches everything; meticulous’,
bólolasuma <hand-in-slow> ‘slow, who works slowly’,
bólolateli <hand-in-quick> ‘very quick, prompt; skilful; who touches everything; who comes to hit people easily’,
dálacaman <mouth-in-numerous> ‘talkative’,
dálandi <mouth-in-pleasant> ‘talkative; indiscreet’,
dámandi <mouth-on-pleasant> ‘talkative; indiscreet’,
dálafegen, *dálafjɛn* <mouth-in-light> ‘indiscreet’,
dálagelen <mouth-in-difficult> ‘obstinate and cheeky (who fails to recognise one's fault)’,
dámagelen <mouth-on-difficult> ‘obstinate and cheeky (who fails to recognise one's fault)’,
dálagirin <mouth-in-heavy> ‘discreet’,
dálajugu <mouth-in-evil> ‘uncouth (in speech)’,
dálango <mouth-in-unpleasant> ‘venomous (in speech); who answers in an unpleasant way or does not answer’,
dámango <mouth-in-unpleasant> ‘venomous (in speech); who answers in an unpleasant way or does not answer’,
dálankolon <mouth-in-empty> ‘who has nothing to say’; ‘toothless’ (the literal meaning),
dálakunan <mouth-in-bitter> ‘malicious, venomous, evil-tongued’,

dálamisen <mouth-in-small> ‘talkative (who cannot hold one's tongue), indiscreet; quick-tempered; glutton’,
dálasuma <mouth-in-slow> ‘discreet (that is, one who keeps secrets); reserved (one who is not talkative)’,
dálateli <mouth-in-quick> ‘who speaks hastily; who insults easily’,
dúsumandi <heart-on-pleasant> ‘affable, kind; in a good mood’,
dúsumango <heart-on-unpleasant> ‘unpleasant; irritable, short-tempered’,
jàlafarin <shadow-in-courageous> ‘reckless, courageous’,
jàlagelen <shadow-in-difficult> ‘brave’,
jàlamisen <shadow-in-small> ‘fearful’,
jòlimandi <blood-on-pleasant> ‘nice, likeable; preferred, preferable’,
jòlimango <blood-on-unpleasant> ‘unpleasant’,
jùlajugu <buttocks-in-evil> ‘shameless; nymphomaniac’,
jùlankolon <buttock-in-empty> ‘naked’,
kòlomadɔgɔ <bone-on-small> ‘puny, undeveloped’,
kòlomagelen <bone-on-difficult> ‘resistant, tough, vigorous’,
kòlomamisen <bone-on-small> ‘small, thin; puny, feeble; frail’,
kɔ̀nɔ̀nandi <stomach-in-pleasant> ‘well-intentioned, helpful; kind’,
kɔ̀nɔ̀naje <stomach-in-white> ‘sincere, nice’,
kɔ̀nɔ̀nafin <stomach-in-black> ‘dishonest, bad, wicked’,
kɔ̀nɔ̀nango <stomach-in-unpleasant> ‘insincere; ill-intentioned; nervous’,
kɔ̀nɔ̀najugu ~ *kɔ̀nɔ̀nanjugu* <stomach-in-bad> ‘ill-intentioned, wicked’,
kùnmadɔgɔ <head-on-small> ‘puny, feeble; shameless’,
kùnnagelen <head-in-difficult> ‘recalcitrant’,
kùnnandi <head-in-pleasant> ‘lucky’,
kùnnango <head-in-unpleasant> ‘unlucky’,
nìlango <soul-in-unpleasant> ‘ill-tempered, sad, sullen’,
nùnnango <nose-in-unpleasant> ‘quick-tempered, aggressive; teasing’,
jànandi <eye-in-pleasant> ‘happy, cheerful’,
jànagelen <eye-in-difficult> ‘rash, foolhardy’,
jànajugu <eye-in-evil> ‘sponger, parasitic, beggar’,
jànango <eye-in-unpleasant> ‘clumsy, tactless’,
jànakegun <eye-in-sly> ‘dexterous’,
jànakunan <eye-in-bitter> ‘cheeky; jealous’,
jànamisen <eye-in-small> ‘meticulous, delicate’,
sènnandi <foot-in-pleasant> ‘quick, light-footed’,
sènnagoni <foot-in-hot> ‘rapid’,
sènnakalan <foot-in-hot> ‘rapid’,
sènnasuma <foot-in-slow> ‘slow’,
sènnateli <foot-in-quick> ‘rapid, light-footed’,

tégelandi <palm.of.hand-in-pleasant> ‘nimble, skillful’,
tégelango <palm.of.hand-in-unpleasant> ‘clumsy’,
tégelankolon <palm.of.hand-in-empty> ‘poor’,
tégelamisen <palm.of.hand-in-small> ‘thievish; who touches everything’,
tégenasuma <palm.of.hand-in-slow> ‘slow (of hands)’,
tégelatelin, tégelantelin <palm.of.hand-in-quick> ‘nimble (of hands)’,
tégemagelen <palm.of.hand-on-hard> ‘stingy, miserly’,
túlomagelen <ear-on-hard> ‘obstinate, stubborn’.

In October 2017, I carried out an alternative study in which all imaginable combinations (every noun stem with every qualitative verb stem, with the 4 allomorphs of the 2 connectors) were searched for in the Bamana Reference Corpus which then comprised 4,113,006 tokens.

The investigation has produced 27 separable adjectives. Among these, two appear only as components of further derivatives, and such occurrences can hardly be seen as sufficient to include the forms in the list of separable adjectives: *dá-la-fégen* <mouth-in-light> ‘indiscreet’ was found as a component of *dálafégen-ya* ~ *dálaŋfiyen-ya* ‘indiscreetness’, and *jòli-man-di* <blood-on-pleasant> ‘nice, sympathetic’ in *jòlimandi-ya* ‘likability, fascination’, e.g. (4).

(4) *Ñka* *sára*, *jòlimandiya* *àni* *cèŋe* *mîn*
 but charm\ART fascination\ART and beauty\ART REL
bé *ntòlatan* *ná*, *ò* *bé* *jîni* *kà* *túnun*.
 be football\ART at that IPFV search INF disappear
 ‘But the charm, the fascination and the beauty of football
 is on the verge of disappearance’ [Kibaru 390].

8 separable adjectives have single occurrences:

bólolandi <hand-in-pleasant> ‘careful, skillful’,
kónɔmandi <stomach-on-pleasant> ‘well-wishing, kind’,
kónɔnaje <stomach-in-white> ‘sincere, nice’,
kùnnagelen <head-in-difficult> ‘recalcitrant’,
jénandi <eye-in-pleasant> ‘happy, cheerful’,
jénagelen <eye-in-difficult> ‘rash, foolhardy’,
jénakegun <eye-in-sly> ‘dexterous’,
jénakunan <eye-in-bitter> ‘cheeky; jealous’.

According to the statistical standards, such potential lexemes should be regarded as unreliable. This leaves us with 19 lexemes; of these only 14 occur more than 5 times which can be regarded as a reliable sample size for our Corpus:

Table 1

Separable adjectives

	<i>cáman</i>	<i>dí</i>	<i>dǒgɔ</i>	<i>jé</i>	<i>júgu</i>	<i>fárin</i>	<i>fěgen</i>	<i>fin</i>	<i>gèlen</i>
<i>bólo</i>		<i>lan</i> <i>lan 1</i>	<i>lan,</i> <i>man</i> <i>ma 18</i>		<i>la,</i> <i>man</i>		<i>la, lan</i>		<i>ma</i>
<i>dá</i>	<i>la</i>	<i>lan,</i> <i>man</i>			<i>la</i>		<i>la</i> <i>(la 1)</i>		<i>la, ma,</i> <i>rɔ</i> <i>la 2</i>
<i>dùsu</i>		<i>man</i>							
<i>jà</i>						<i>la</i>	<i>la 2</i>		<i>la</i>
<i>jòli</i>		<i>man</i> <i>(man 2)</i>							
<i>jù</i>					<i>la</i>				
<i>kólo</i>			<i>ma</i>						<i>ma</i>
<i>kɔnɔ</i>		<i>nan</i> <i>man 1</i>		<i>na,</i> <i>nan</i> <i>na 1</i>	<i>la, na,</i> <i>lan,</i> <i>nan</i> <i>na 1,</i> <i>nan 9</i>			<i>na</i>	
<i>kùn</i>		<i>nan</i> <i>na 3,</i> <i>nan 48</i>	<i>ma</i>						<i>na</i> <i>na 1</i>
<i>ní</i>		<i>(sɔŋ)</i>							
<i>nún</i>									
<i>ɲé</i>		<i>nan</i> <i>nan 1</i>			<i>na</i>				<i>na</i> <i>na 1</i>
<i>sèn</i>		<i>na</i>			<i>na 1</i>				
<i>tége</i>		<i>lan</i>							<i>ma</i> <i>ma 6</i>
<i>túlo</i>									<i>ma</i> <i>ma 6</i>

in Bamana through elicitation

<i>gírin</i>	<i>gó</i>	<i>gòni</i>	<i>kálan</i>	<i>kègun</i>	<i>kólon</i>	<i>kúnan</i>	<i>misen</i>	<i>súma</i>	<i>téli(n)</i>
	<i>lan</i>				<i>lan</i> <i>lan 74</i>		<i>la</i>	<i>la</i>	<i>la</i>
<i>la</i>	<i>lan,</i> <i>man</i>				<i>lan</i> <i>la 1,</i> <i>lan 6</i>	<i>la</i>	<i>la</i>	<i>la</i>	<i>la</i>
	<i>man</i> <i>man 12</i>								
	<i>man</i> <i>man 6</i>								
					<i>lan</i> <i>la 3,</i> <i>lan 8</i>				
							<i>ma</i>		
	<i>nan</i> <i>nan 12</i>								
	<i>nan</i> <i>nan 6</i>				<i>na 1,</i> <i>nan 11</i>				
	<i>lan,</i> <i>sɔŋ</i>								
	<i>nan</i>								
	<i>nan</i>			<i>na</i> <i>na 1</i>		<i>na</i> <i>na 1</i>	<i>na</i>		
		<i>na</i>	<i>na</i>		<i>na, nan</i> <i>64</i>			<i>na</i>	<i>na</i> <i>na 2</i>
	<i>(lan)</i>			<i>lan</i>	<i>lan</i> <i>lan 9</i>			<i>na</i>	<i>la, lan</i>

bólomadɔɔ <hand-on-small> ‘poor; weak’,
bólolankolon <hand-in-empty> ‘poor’,
dálakolon <mouth-in-empty> ‘who has nothing to say’; ‘toothless’ (the literal meaning),
dúsumango <heart-on-unpleasant> ‘unpleasant; irritable, short-tempered’,
jòlimango <blood-on-unpleasant> ‘unpleasant’,
jùlankolon <buttock-in-empty> ‘naked’,
kɔ̀nɔ̀najugu <stomach-in-bad> ‘ill-intentioned, wicked’,
kɔ̀nɔ̀nango <stomach-in-unpleasant> ‘insincere; ill-intentioned; nervous’,
kùnnandi <head-in-pleasant> ‘lucky’,
kùnnango <head-in-unpleasant> ‘unlucky’,
kùnnankolon <head-in-empty> ‘uncovered’,
sènnankolon <foot-in-empty> ‘barefooted’,
tégemagelen <palm.of.hand-on-hard> ‘stingy, miserly’,
tégelankolon <palm.of.hand-in-empty> ‘poor’,
tùlomagelen <ear-on-hard> ‘obstinate, stubborn’.

Among the separable adjectives found in the Corpus, four are missing from the list obtained through elicitation. Two are reliable: *sènnakolon* ~ *sènnankolon* ‘barefooted’ (64 occurrences) and *kùnnankolon* ~ *kùnnakolon* ‘bareheaded’ (12 occurrences). One more lexeme has single occurrence: *kɔ̀nɔ̀mandi* ‘well-wishing, kind’.

3.4. Elicitation and corpus study: Comparison of the results

Let us compare the results of both approaches, cf. Table 1. The noun roots are in vertical columns, the qualitative verbs stems are in horizontal lines; the connectors are in the cases where the combinations of N and QV are attested. The connector is black if the form has been obtained through elicitation, and it is red if the form has been found in the Corpus; in the latter case, the number of occurrences is indicated.

In this study, elicitation has proved to be three times more effective (with respect to the inventorizing the lexemes in question) than the corpus study: 75 elicited lexemes vs. 25 ones found in the Corpus; and if we take into account only reliable lexemes (5 occurrences or more), the elicitation turns to be five times more effective.

Another advantage is the fact that during an elicitation session, semantic information about the lexeme is normally produced at the same time as the form. In the case of a corpus study, the semantics of a form (especially if we have just a single occurrence) may be obscure. This may be especially true for the separable adjectives whose sense is often idiomatic and not directly derivable from the meanings of the components.

Nonetheless, it would be wrong to think that a corpus study were useless. Although it yields to the elicitation in terms of productivity, it allows us to discover lexemes which, for some reason or another, may have been skipped during an elicitation session, despite their frequency.

Another strong point of the corpus study, even more important than the previous one, is the availability of statistical data for each lexeme. In our particular case, even superficial analysis of frequencies shows that among the stems of qualitative verbs, the most productive ones are *kólon* ‘empty’, followed by *dí* ‘pleasant’, *gó* ‘unpleasant’ and *gèlen* ‘difficult’. Based on the frequencies, it is easy to single out the nucleus of this category from the periphery. The nucleus consists of the most frequent lexemes *bólolankolon* (74 occurrences) ‘poor’, *sènnakolon ~ sènnankolon* (64) ‘barefooted’, *kùnnandi ~ kùnnadi* (51) ‘fortunate’, *bólomadɔgɔ* (18) ‘poor; weak’.⁸

4. Conclusions

There is hardly any serious linguist who would contest the importance of data obtained from natural texts in a language description. However, the Bambara separable adjectives case study has clearly shown that the use of only natural texts would fail to provide a near complete inventory of lexemes of the subclass. Evidently, the same situation can be expected in other studies of derivative or word-compounding models. It should be underlined that Bambara is in a privileged position: out of some 2000 of languages spoken in Africa, less than a dozen possess text corpora of more than 1,000,000 words, and among these, few are in open access. If a language has only a limited amount of natural texts available (in the range of 20,000 to 100,000 words, which is most typically of languages studied by a single linguist), rejection of elicitation may lead to very thin description of derivation (or any other study where a matrix-like examination is necessary).

In fact, a study based on the data from natural texts (in particular, a corpus-based study) and a study through elicitation are complementary. Each of these approaches fits some particular tasks, and it is less appropriate for some others. It would be counterproductive to absolutize one of them and to reject the other one.

References

- Aikhenvald, 2007 – Aikhenvald A.Yu. Linguistic fieldwork: setting the scene. *Sprachtypologie und Universalienforschung*. 2007. № 60 (1). Pp. 3–11.
 Bailleul, 2000 – Bailleul C. Cours pratique de bambara. Bamako, 2000.
 Bailleul, 2007 – Bailleul C. Dictionnaire Bambara-Français. 3rd ed. Bamako, 2007.

⁸ A more sophisticated analysis is possible if metadata are taken into account.

Bailleul et al., 2011 – Bailleul C., Davydov A., Erman A., Maslinsky K., Méric J.-J., Vydrin V. Bamadaba: Dictionnaire électronique bambara-français, avec un index français-bambara. 2011. URL: <http://cormand.huma-num.fr/bamadaba.html>.

Bird et al., 1977 – Bird C., Hutchison J., Kante M. Beginning Bambara / An Ka Bamanankan Kalan: With Bambara-English Glossary. Bloomington, 1977.

Bird, Kante, 1976 – Bird C., Kante M. Intermediate Bambara / An Ka Bamanankan Kalan. Bloomington, 1976.

Chelliah, De Reuse, 2011 – Chelliah S.L., De Reuse W.J. 2011. Handbook of descriptive linguistic fieldwork. Springer, 2011.

Creissels, 1985 – Creissels D. Les verbes statifs dans les parlers manding. *Mandenkan*. 1985. № 10. Pp. 1–32.

Dumestre, 1987 – Dumestre G. Le bambara du Mali: Essai de description linguistique. Paris, 1987.

Dumestre, 2003 – Dumestre G. Grammaire fondamentale du bambara. Paris, 2003.

Dumestre, 2011 – Dumestre G. Dictionnaire bambara-français suivi d'un index abrégé français-bambara. Paris, 2011.

Kastenholz, 1989 – Kastenholz R. Grundkurs Bambara (Manding), mit Texten. Köln, 1989.

Tröbs, 2008 – Tröbs H. Bambara. *La qualification dans les langues africaines. Qualification in African Languages*. H. Tröbs, E. Rothmaler, K. Winkelmann (eds.). Köln, 2008. Pp. 13–28.

Vydrin, 1990 – Vydrine V. Les adjectifs prädicatifs en bamana. *Mandenkan*. 1990. № 20. Pp. 47–90.

Vydrin, 1999a – Vydrine V. Manding-English Dictionary (Maninka, Bamana). Vol. 1. St. Petersburg, 1999.

Vydrin, 1999b – Vydrine V. Les parties du discours en bambara : un essai de bilan. *Mandenkan*. 1999. № 35. Pp. 73–93.

Vydrin, 2008 – Выдрин В. Язык бамана: Учебное пособие. СПб., 2008. [Vydrin V. Jazyk bamana [Bamana language]. Textbook. St. Petersburg, 2008.]

Vydrin, 2016 – Vydrin V. Tonal inflection in Mande languages: The cases of Bamana and Dan-Gweetaa. *Tone and Inflection: New facts and new perspectives*. E.L. Palancar, J.L. Léonard (eds.). (Trends in Linguistics Studies and Monographs 296). De Gruyter – Mouton, 2016. Pp. 83–105.

Vydrin, in press – Vydrin V. Cours de grammaire bambara. Presses de l'INALCO. To appear in 2019.

Vydrin, Maslinsky, Méric et al., 2011 – Vydrin V., Maslinsky K., Méric J.-J. Corpus Bambara de Référence. 2011. URL: <http://cormand.huma-num.fr/index.html>.

Статья поступила в редакцию 21.10.2018

The article was received on 21.10.2018

Выдрин Валентин Феодосьевич – доктор филологических наук; профессор кафедры африканистики, Национальный институт восточных языков и цивилизаций, Париж, Франция; исследователь лаборатории языка, речи и культуры Чёрной Африки, Национальный центр научных исследований Франции; профессор кафедры африканистики восточного факультета, Санкт-Петербургский государственный университет

Vydrin Valentin F. – Dr. Phil. Hab.; professor of Manding at the African Department, National Institute for Oriental Languages and Civilizations, Paris, France; researcher at the laboratory “Languages and cultures of Sub-Saharan Africa”, French National Centre for Scientific Research; professor at the Department of African Studies of Faculty of Asian and African Studies, St. Petersburg University, Russian Federation

E-mail: vydrine@gmail.com