

ЯЗЫКОЗНАНИЕ

В.Ф. Выдрин

КОРПУСНЫЕ СЛОВАРИ ЯЗЫКОВ МАНДЕН¹

В течение последнего десятилетия были созданы миллионные корпуса текстов манинка и бамана. Параллельно с этим ведется работа по созданию и совершенствованию словарей этих языков. Складываются условия для вывода лексикографической работы на принципиально новый уровень, который имел бы целью создание «корпусных словарей» (*corpus-driven dictionaries*). Это позволило бы гораздо детальнее и достовернее, по сравнению с «традиционными» словарями, представлять полисемию лексем, их сочетаемостные свойства, стилистические и регионально-диалектные характеристики каждого значения. В статье дается пробная разработка структуры словарной статьи глагола *bánban* на основании Справочного корпуса бамана.

Ключевые слова: язык бамана, язык манинка, манден, корпусная лексикография, полисемия.

Примерно десять лет назад была осторожно высказана мысль о необходимости создания больших аннотированных корпусов текстов для языков манден [Выдрин 2008]. Сейчас можно сказать, что эта задача уже решена, по крайней мере в первом приближении: объем Справочного корпуса бамана на начало лета 2017 г. превысил 4,1 млн слов, при этом подкорпус со снятой омонимией перевалил за 900 тыс. слов. Справочный корпус манинка составил более 3,5 млн слов (3,1 млн в подкорпусе нко и почти 0,4 млн — в латинском подкорпусе).

Эти корпуса аннотированы при помощи словарей *Vamadaba* (баманский корпус) и *Malidaba* (манинканский корпус). Эти сло-

¹ This work is related to the research strand 3 «Typology and dynamics of linguistic systems» of the Labex EFL (financed by the ANR/CGI).

вари в настоящий момент очень различаются по степени проработанности: *Vamadaba* представляет собой полноценный бамана-французский и французско-баманский словарь, находящийся в открытом доступе на сайте баманского корпуса (<http://cormand.huma-num.fr/bamadaba.html>), он активно используется студентами и всеми теми, кто интересуется языком бамана, постоянно пополняется и уточняется в ходе работы по снятию омонимии. Словарь *Malidaba* находится на значительно менее продвинутой стадии: он был скомпилирован полуавтоматически из различных лексикографических источников и сейчас находится на этапе предварительной «чистки», которая состоит в устранении дублирующих словарных статей, подборе трехязычных глосс, и, менее последовательно, введения определенного минимума грамматической информации и первичной проработки полисемии (подробнее см.: [Vydrin, Rovenchak, Maslinsky 2016]). Первичная проработка словаря *Malidaba* к началу лета 2017 г. была проведена примерно на 65 %, после завершения первого этапа чистки планируется помещение *Malidaba* на сайте Справочного корпуса манинка (<http://cormand.huma-num.fr/cormani/>).

Создание этих словарей само по себе — значительный шаг в развитии лексикографии языков манден. *Vamadaba*, хотя и отстает по объему от словаря Жерара Дюестра [Dumestre 2011]², существенно превосходит его по уровню структурированности, а также обладает преимуществами онлайн-словаря: постоянно совершенствуется (в том числе пополняется новыми пластами лексики, хорошо представленной в текстах, но оставшейся в силу разных причин вне «традиционных» словарей), имеет прямой доступ к корпусу (каждая лексема имеет ссылку, по которой мгновенно создается полный конкорданс из подкорпуса со снятой омонимией), дает возможность поиска как в бамана-французском, так и во французско-баманском режиме. *Malidaba*, хотя и значительно

² Словарь Ж. Дюестра насчитывает 15 850 словарных статей, *Vamadaba* на начало сентября 2017 г. — почти 12 500. Впрочем, эти цифры нужно принимать с оговорками, поскольку принципы выделения лексем в этих словарях различаются в некоторых важных моментах.

уступает баманскому словарю и в количественном³, и в качественном отношении, станет первым манинка-французским словарем, отвечающим хотя бы минимальным лексикографическим требованиям (обозначение тонов, разработка полисемии, наличие грамматических помет и т.д.).

Таким образом, создание и совершенствование миллионных корпусов манден уже оказало ощутимое влияние на лексикографию этих языков. Тем не менее, не будет преувеличением сказать, что лексикографический потенциал корпусов пока задействован лишь незначительно. Настало время всерьез задуматься о создании корпусных словарей (*corpus-driven dictionaries*) для языков манден с учетом накопленного опыта корпусной лексикографии для других языков. Такие словари, по сравнению с «традиционными», позволяют представить лексику языка гораздо более объемно, максимально приближая представление лексемы в словаре к ее реальному употреблению в языке.

Поскольку литература о корпусной лексикографии достаточно богата (см., в частности: [Kilgarriff 1997; Kilgarriff, Rundell, Dhonnchadha 2006]), вместо изложения известных методологических подходов я попробую показать на примере одной статьи (глагол *bánban*), что дает корпусной подход к решению лексикографических задач для бамана.

В словаре Дюместра [Dumestre 2011] (далее DUM) указано, что эта лексема имеет также фонетический вариант *báman*, и даются (простым списком) следующие переводные эквиваленты: *s'évertuer, persévérer, faire de son mieux, prendre au sérieux; prendre appui; enfoncer, fixer solidement* BA⁴; *centrer, concentrer; tendre, déployer, hisser; effort tendu* КТ, *attention soutenue, sérieux* (стараться, упорно делать, делать изо всех сил, относиться серьезно; опираться; втыкать, прочно закреплять ВА; сосредотачивать, концен-

³ По предварительной оценке, после завершения первого этапа чистки *Malidaba* будет содержать примерно 6300 лексем.

⁴ Двухбуквенные индексы у Дюместра отсылают к источникам, список которых дается в конце его словаря.

трировать; натягивать, развертывать, поднимать (парус, флаг); напряженное усилие КТ, напряженное внимание, серьезность)⁵.

Далее в словаре приводятся 16 примеров, иллюстрирующих употребление *bánban*.

В Манден-русском словаре [Выдрин, Томчина 1999] (далее ВТО) даны три формы: *bánban* (общеупотребительная), *bánba* (только бамана), *báman* (неуточненный диалект бамана). Значения глагола представлены с иерархизированной рубрикацией, даны пометы, отражающие актантную структуру глагола в каждом значении (переходный/непереходный/рефлексивный глагол; при наличии сильной валентности указывается послелог, вводящий косвенный член), в случаях, когда какие-то значения характерны не для всего ареала, вводятся соответствующие диалектные пометы. Таким образом, полисемия глагола *bánban* здесь выглядит следующим образом (иллюстративные примеры и идиоматические выражения опущены):

1. (m) *ng* натягивать, туго наматывать

2.1. *ng* прикреплять, укреплять, вбивать; прочно ставить, устанавливать

pg прочно встать

2.2. (m) *ng* делать прочным

3. *ng* упираться чем-л. (в — *lá*), надавливать чем-л.

pg опираться, упираться (в — *lá*)

4. *pg* усердствовать (в — *lá*), очень стараться (делать что — inf.)

5. (m) *ng* поддерживать, помогать

6. (bSegu) *pg* выпутаться, легко отделаться.

Сравнение двух словарей⁶ показывает, что они дополняют друг друга:

⁵ Последние три значения именные. Дюместр рассматривает конверсию «глагол → существительное» как полисемию. Далее именные значения не анализируются.

⁶ В словаре Шарля Байоля [Bailleul 2007] структура полисемии этого глагола представлена более сжато, она полностью перекрывается словарями DUM и ВТО и отдельно рассматриваться не будет.

— в DUM есть значения ‘сосредотачивать, концентрировать’, ‘поднимать (флаг, парус)’, ‘относиться серьезно’, отсутствующие в ВТО;

— ВТО, но не DUM, дает значения ‘*pg* прочно встать’, ‘упираться чем-л., надавливать’, ‘поддерживать, помогать’, ‘*pg* выпутаться, легко отделаться’ (диалектное значение); для манинка, но не для бамана указано также значение ‘*ng* делать прочным’ (которое, впрочем, можно объединить со значением ‘укреплять’, отделив последнее от 2.1 ‘прикреплять’ и т.д.).

Кроме того, ВТО указывает значение ‘*ng* натягивать, туго наматывать’ с пометой «манинка». DUM подтверждает значение ‘натягивать’, но не ‘туго наматывать’ (которое, по-видимому, следует рассматривать не вместе с ‘натягивать’, а как отдельное значение или даже как вариант значения 2.1).

Можно также предложить «приблизить» манинканское значение (5) ‘поддерживать, помогать’ к семантически смежному значению 2.2. ‘делать прочным, укреплять’.

По результатам этого сравнения полисемия глагола может быть представлена следующим образом:

1.1. *ng* натягивать, развертывать

1.2. *ng* поднимать (парус, флаг)

2.1. *ng* прикреплять, туго наматывать; вбивать; прочно ставить, устанавливать

pg прочно вставать

2.2. (m) *ng* делать прочным, укреплять

2.3. (m) *ng* поддерживать, помогать

3. *ng* упираться чем-л. (в — *lá*), надавливать чем-л.

pg опираться, упираться (в — *lá*)

4.1. *pg* усердствовать (в — *lá*), очень стараться (делать что — inf.)

4.2. *pg* относиться серьезно (к — *lá*)

4.3. *ng* концентрировать, сосредотачивать (на — *kân*)

5. (bSegu) *pg* выпутаться, легко отделаться.

В мае 2013 г., когда Справочный корпус бамана насчитывал около 1,5 млн слов, я провел корпусную проверку полисемии это-

го глагола. На тот момент в корпусе обнаружилось 123 употребления формы *banban*, из которых 12 (именные употребления, омонимы) были отсеяны. Таким образом, к интересующему нас глаголу на тот момент относились 111 употреблений (в это число вошли и регулярные дериваты, значения которых автоматически выводимы из глагольных, и случаи контекстной номинализации глагола в составе именных композитов). Почти половину корпуса (точнее, около 700 тыс. слов из 1,5 млн) на тот момент составляли тексты Ветхого Завета. Первым интересным результатом стало то, что из 111 употреблений *bánban*, 76 (т.е. 68,5 %) происходили из Ветхого Завета, иначе говоря, в этом источнике глагол *bánban* оказался вдвое более частотным, чем в остальных текстах Корпуса. Интересно также, что фонетический вариант *baman*, отмеченный в словарях ВТО и DUM, не встретился ни разу.

В сентябре 2017 г. я снова проанализировал употребление *bánban* по действующей версии Справочного корпуса бамана, которая превосходила по объему (4,1 млн слов) версию 2013 г. в 2,73 раза; соответственно, доля текстов Ветхого Завета в Корпусе снизилась до 17 %.

Общее число употреблений формы *banban* с учетом дериватов и композитов составило 253. После отсева омонимов, нерегулярных дериватов (каузатив *lábanban*) и именных употреблений это число сократилось до 226. Таким образом, прежняя тенденция сохранилась — частотность *bánban* в Ветхом Завете и сейчас в два раза выше, чем в остальной части корпуса.

Несмотря на увеличение выборки, вариант *baman* так и не появился (встретившаяся пять раз форма *bamanna* оказалась во всех случаях не перфективной интранзитивной формой интересующего нас глагола, а неправильным написанием слова *bámàna* ‘бамана’); очевидно, его следует вынести за пределы современного письменного узуса бамана.

Ниже будет проанализирована полисемия глагола *bánban* на основании корпусных данных. Последовательность значений в целом соответствует той, что приведена выше; новые (под)значения вводятся (в соответствии с логикой, принятой в ВТО) рядом с теми значениями, которые наиболее близки к ним семантически.

При каждом значении сразу после валентностной пометы (*пг*, *нг*, *рг*) указано в скобках число употреблений глагола *bánban* в данном (под)значении.

При количественной оценке веса каждого значения следует иметь в виду, что учитывались не только собственно глагольные употребления, но и регулярные дериваты, а также композиты, в которых глагольное значение просматривается достаточно отчетливо; для некоторых значений доля таких дериватов и композитов оказалась весьма высокой. Нужно отметить, что в значениях *bánban*, для которых характерна лабильность, у дериватов и композитов противопоставление по переходности/непереходности/рефлексивности (укреплять — укрепляться, прочно ставить — прочно вставлять) нейтрализуется, что создает трудность для статистического подсчета. Другая трудность — вопрос об отнесении пассивных употреблений к переходному или непереходному подзначению (этот вопрос я решил в пользу переходных значений, т.е. ‘быть установленным’ считается вместе с ‘прочно устанавливать’, а не ‘прочно вставлять’).

Другой важный момент — это синкретизм значений. Зачастую оказывается не так просто определить, в каком именно из смежных значений употреблен глагол в том или ином предложении, например: 2.1 ‘прочно ставить’ или 2.2 ‘делать прочным, укреплять’, 4.1 ‘усердствовать в чем-л.’ или 4.2 ‘относиться серьезно к’.

Таким образом, статистику распределения употреблений глагола по значениям следует рассматривать как приближительную.

1.1. *пг* (4) натягивать, растягивать.

Это значение в ВТО помечено как «манинканское»; впрочем, значение ‘*tendre*’ упоминается и в DUM. Из четырех примеров один — из газеты *Jekabaara*, в абстрактном контексте описания отношений между людьми; три — из художественной литературы. Таким образом, «манинканская» помета может быть убрана, однако эти значения в бамана оказываются несчастными.

1.2. *пг* (0) поднимать (парус, флаг).

Встает вопрос о том, нужно ли выделять это значение, или считать его вариантом значения 1.1.

2.1. *ng* (22) прикреплять, туго наматывать, вбивать; прочно ставить, устанавливать;

ng (3) прочно вставать;

rg (0) прочно вставать.

Примеров на рефлексивный вариант этого значения в корпусе не найдено, зато обнаружилось (пусть и не очень многочисленными) непереходные употребления, не предусмотренные в имеющихся словарях. Впрочем, из трех таких употреблений бесспорным является только одно, еще в двух случаях мы имеем форму результативного причастия *bánbannen*, употребленную в идентичных контекстах (1). Таким образом, непереходный вариант значения также можно считать маргинальным.

(1)	<i>Dùgukolo`</i>	<i>bánban-nen</i>	<i>bé</i>	<i>kójùman</i>	<i>àní</i>
	земля\ART	натягивать-PTCP.RES	быть	хорошо	и
	<i>à</i>	<i>tɛ</i>	<i>sé</i>	<i>kà</i>	<i>lámàga.</i>
	3SG	IPFV.NEG	МОЧЬ	INF	двигать

‘Земля стоит/установлена прочно, ее нельзя пошевелить’.

Интересно также, что из 22 переходных употреблений *bánban* в этом значении 19 приходится на Ветхий Завет (почти всегда — в описании создания мира и его компонентов Богом), два — на газету *Jekabaara*, один — на газету *Kibaru*. Получается, что довольно высокая цифра для данного значения (около 10 % от всех корпусных употреблений глагола *bánban*) — следствие описанного Адамом Килгарифом «эффекта морской улитки»⁷, тогда как за пределами Ветхого Завета это значение оказывается редким.

2.2. *ng* (55) делать прочным, укреплять;

ng (31) укрепляться, закрепляться, укореняться, становиться устойчивым;

rg (3) укрепляться, закрепляться.

⁷ «The whelks problem» [Kilgarriff 1997: 138–139]. Имеется в виду ситуация, когда в выборке оказывается текст, специально посвященный какому-то в целом редко упоминаемому объекту или явлению (например, морской улитке, *whelk*), в результате чего суммарная частотность этого слова в выборке оказывается завышенной.

Это значение (‘делать прочным’) в ВТО помечено как «манинканское», однако оказалось, что в баманском корпусе оно (особенно его абстрактный вариант ‘укреплять’) — самое частотное. Обнаружилось немало употреблений и непереходного варианта этого значения, не отмеченного в ВТО⁸, а также рефлексивного варианта, более редкого. Распределение по источникам для этого значения оказывается относительно равномерным (точнее, перевес в пользу Ветхого Завета наблюдается и здесь, но он значительно меньший, чем для значения 2.1).

Таблица 1. Распределение употребления 2.2 по источникам

	Ветхий Завет	Jɛkabaara	Kibaru	другие
<i>ng</i>	25	21	4	5
<i>ng</i>	7	24	0	0
<i>pg</i>	2	0	1	0

2.3. *ng* (3) поддерживать кого-л., помогать кому-л.

Это значение в ВТО также отнесено к манинка, однако оно отмечено также в баманском Ветхом Завете, из которого и происходят все три обнаруженных употребления.

3.1. *ng* (0) упираться чем-л. (в — *lá*), надавливать чем-л.;

pg (0) упираться (в — *lá*).

Это значение, упомянутое в ВТО как вариант значения ‘опираться’ (однако в действительности достаточно дистантное от него), в корпусе не обнаружено; по-видимому, имеет смысл исключить его из словаря.

3.2. *pg* (22) опираться (на, о — *kàn*);

ng (12) опираться (на, о — *kàn*).

Отделение этого значения от 3.1 (с которым оно объединено в ВТО) подкрепляется и тем, что его косвенное дополнение требует послелога *kàn*, а не *lá* (вопреки тому, что указано в ВТО; впро-

⁸ Отметим, впрочем, что из 31 употребления, отнесенного при подсчете к непереходным, лишь 20 приходится на финитные, остальные 11 — причастия и отглагольные имена, которые соотносены с непереходным вариантом этого значения лишь предположительно.

чем, вполне возможно, что тут речь не о разном управлении при разных значениях, а об ошибке в ВТО или различии между манинка и бамана по этому параметру).

Практически во всех отмеченных случаях употребления (как непереходных, так и рефлексивных) речь идет об абстрактном варианте значения ('опираться на кого-то в своих действиях', 'опираться на правила' и т.п.), физическое действие имеется в виду лишь в одном случае.

Различие между рефлексивным и непереходным употреблением обусловлено, за редкими исключениями, типом подлежащего: при одушевленном подлежащем глагол употребляется рефлексивно, при неодушевленном — непереходно, что соответствует общему правилу [Vydrine, Coulibaly 1995: 71]. Нужно оговориться, что приведенная здесь статистика условна, поскольку более половины употреблений *bánban* в этом значении — в составе композитов, в таких случаях рефлексивный или непереходный характер исходной глагольной конструкции может быть определен только по типу смыслового подлежащего.

Что касается распределения этих употреблений по источникам, то здесь отмечено лишь одно рефлексивное употребление в *Kibaru* и четыре рефлексивных употребления в Ветхом Завете, все остальные происходят из *Jekabaara*⁹, причем почти все они принадлежат перу одного автора — Тумани Ялама Сидибэ¹⁰, главного редактора и основного автора этой газеты. Таким образом, частое употребление *bánban* в этом значении можно считать особенностью авторского стиля Т.Я. Сидибэ.

4.1. *pg* (20) усердствовать (в — *fê*, *kàn*), стараться (делать что — *fê*, *kàn*, Inf., Sbjv).

Вопреки ВТО, косвенный член (объект приложения стараний) вводится послелогом *fê* (при обозначении сферы деятельности) или *kàn* (если обозначается цель деятельности), а не *lá*.

⁹ При том что объемы текстов из обеих газет, *Jekabaara* и *Kibaru*, в корпусе примерно одного порядка, см.: http://cormand.huma-num.fr/sc_non_desambig.html, http://cormand.huma-num.fr/sc_desambig.html.

¹⁰ Т.Я. Сидибэ является выходцем из г. Кита, т.е. из зоны распространения центрально-западного варианта языка манинка.

(2) *Án ka kán k' án b́anban*
 1PL QUAL.AFF равный INF 1PL натягивать
kóori-sene in fè...
 хлопок-возделывать DEF с

‘Мы должны усердно растить хлопок...’ [jekabaara116_03sidibe-kita_makoci].

(3) *An ka dá án yèrê lá; k' án*
 1PL SBJV ложиться 1PL сам в INF 1PL
b́anban án ka ké-ta-w ké-cogo
 натягивать 1PL POSS делать-PTCP.POT-PL делать-способ
juman` kán.
 хороший\ART на

‘Мы должны верить себе и стараться как следует выполнять намеченное’ [jekabaara113_01sidibe-kalo_laadilikan].

Достаточно часто глагол в этом значении употребляется и без косвенного члена.

Что касается распределения употреблений *b́anban* в этом значении, то и здесь газета *Jekabaara* лидирует с большим отрывом (13 употреблений из 20, в том числе девять примеров принадлежат перу Т.Я. Сидибе), остальные случаи распределены равномерно.

4.2. *pg* (14) относиться серьезно (к — *kán, mà*).

Это значение отмечено в DUM и отсутствует в ВТО; корпусная проверка показала, что оно может быть отнесено к семантическому ядру глагола *b́anban*. Чаще всего косвенный член вводится послелогом *kán*, встретился лишь один случай, когда в этой функции употреблен послелог *mà* (возможно, речь идет о каком-то особом нюансе значения, но для проверки этого предположения пока недостаточно данных); еще один пример с послелогом *mà* приводится в DUM.

По распределению употреблений здесь также абсолютно доминирует *Jekabaara* (11 употреблений).

4.3. *ng* (11) концентрировать, сосредотачивать, нацеливать (на — *kán*);

pg (6) сосредотачиваться, концентрироваться (на — *kán*).

Это значение также есть в DUM и отсутствует в ВТО. Надо признать, что в некоторых случаях довольно трудно отграничить его (особенно в рефлексивном варианте) от значений 4.1 и 4.2.

Все рефлексивные употребления и десять (из 11) переходных употреблений происходят из газеты *Jekabaara*, причем авторство подавляющего большинства из них принадлежит Т.Я. Сидибэ.

4.4. *pg* (9) упорствовать (в — *kàn*), (упорно) следовать (чему — *kàn*), настаивать (на — *kàn*).

В обоих словарях это значение отсутствует, однако оно очевидным образом не сводится ни к 4.1, ни к 4.2, ни к 4.3, хотя семантическая связь с ними несомненна. Пять употреблений происходят из *Jekabaara* (однако на этот раз лишь одно из них принадлежит Т.Я. Сидибэ), остальные четыре — из Ветхого Завета.

5. (bSegu) *pg* (0) выпутаться, легко отделаться.

Значение приводится в ВТО (без иллюстративных примеров), корпусными данными пока что не подтверждается.

Еще два значения могут быть выделены в предварительном порядке: для каждого из них найдено по одному примеру (в обоих случаях — из газеты *Jekabaara*, авторства Т.Я. Сидибэ). Необходимость включения этих значений в словарь подлежит проверке.

6. *ng* (1) довлекать, лежать тяжелым грузом (на — *kàn*).

(4) *Sarili Tayilori ka fànga` gèleya`*
Чарлз Тэйлор POSS сила\ART тяжесть\ART

bánban-na kósebe liberiya-ka-w kàn...
натягивать-PFV.INTR очень либериец-GENT-PL на

‘Власть Чарлза Тэйлора подавляла либерийцев...’
[jekabaara281_8sidibe_wale].

7. *ng* (1) иметь причиной (что — *lá*), происходить (из — *lá*).

(5) *Fóyì wére ma ké jámana-den-w*
ничто другой PFV.NEG делать страна-ребенок-PL

ka híne` bɔ́-ɲɔ́ɲɔ́-na` kɔ́, n’
POSS жалость\ART ВЫХОДИТЬ-RECP-B\ART после если

<i>ò</i>	<i>bánban-na</i>	<i>jàmana`</i>	<i>lè-maa-w</i>
ЭТО	натягивать-PFV.INTR	страна\ART	глаз-человек-PL
<i>ka</i>	<i>jò-yɔ̀rɔ̀`</i>	<i>fá-bali-ya`</i>	<i>lá</i>
POSS	стоять-место\ART	наполнять-PTCP.PRIV-ABSTR\ART	в
<i>kà</i>	<i>táa</i>	<i>súgu-kɔ̀nɔ̀-fen-w</i>	<i>sòngɔ̀`</i>
INF	идти	рынок-внутри-вещь-pl	цены\ART
<i>dàn-sigi-li`</i>		<i>mà.</i>	
граница-садиться-NMLZ\ART		по	

‘Ничто другое не облегчило участь жителей страны, и причиной этому стало то, что руководители страны не сделали то, что они должны были бы сделать для ограничения цен на товары на рынках’ [jekabaara276_6sidibe_geleya].

За рамками анализа остались пять предложений, значение глагола *bánban* в которых неясно. Учитывая, что это составляет всего 2 % от выборки, этими примерами можно пренебречь.

Как я попытался показать, проведенный на основе корпусных данных анализ позволяет существенно уточнить и детализировать представление семантики, грамматических и стилистических особенностей лексемы в словаре:

— выявить значения, не попавшие в «традиционные» словари, а также отсеять «фантомные» значения;

— уточнить данные о сочетаемости и глагольном управлении;

— ранжировать значения по их частотности. Так, для *bánban* очевидно, что этимологически (по-видимому) первичное значение ‘натягивать’ в современном бамана стало достаточно редким; если выстраивать значения по принципу «от самых частотных к самым редким» (а не по принципу логической выводимости, как это сделано в ВТО), то, несомненно, оно должно уйти на второй план;

— учесть особенности источников. Так, употребление *bánban* в значениях 2.1, 2.2, 2.3 оказалось характерным в основном для библейских текстов, да и в целом львиная доля употреблений этого глагола приходится на Ветхий Завет и тексты Т.Я. Сидибе. Такого рода ограничения, по-видимому, также важны для характери-

стики этой лексемы и представляют интерес для мандингской лексикографии.

* * *

Вряд ли у кого-то есть сомнения по поводу того, что корпусный словарь существенно превышает по надежности и проработанности данных словарь «традиционного» типа. Вопрос в том, насколько реалистично создание такого словаря если не в ближней, то хотя бы в среднесрочной перспективе?

Одним из препятствий является то, что на данный момент Справочный корпус бамана еще недостаточно велик и сбалансирован. Можно надеяться, что в ближайшие год или два его объем вырастет до 7–8 млн словоупотреблений (т.е. достигнет примерно того объема, на котором был создан первый в мире основанный на электронном корпусе текстов Кобилдовский словарь английского языка), при этом будет обеспечено большее разнообразие источников и жанров текстов.

Другой (и главной) трудностью является то, что создание корпусного словаря оказывается значительно более трудоемким делом, чем словаря «традиционного». Так, предварительная обработка данных для словарной статьи глагола *bánban* заняла у меня примерно шесть часов; создание такой (средней по объему) словарной статьи — это целый рабочий день. Учитывая, что число лексем в баманском словаре — порядка полутора десятков тысяч, становится очевидным, что один человек вряд ли сможет завершить такой словарь за всю свою жизнь. Иначе говоря, такой проект может быть только коллективным.

Список глосс и сокращений

1, 2, 3 — 1, 2, 3 лицо

ng — непереходное употребление глагола

ng — переходное употребление глагола

rg — рефлексивное употребление глагола

ABSTR — суффикс абстрактного имени

AFF — утвердительный показатель

ART — тоновый артикль

(b) — бамана

- (bSegu) — бамана Сегу
 DEF — «новый определенный артикль»
 GENT — гентильный суффикс (имя жителя какого-л. места или выхода из этого места)
 INF — инфинитив
 INTR — непереходный
 IPFV — имперфектив
 (m) — манинка
 NEG — отрицание, отрицательный
 NMLZ — суффикс имени действия
 PFV — перфектив
 PL — множественное число
 POSS — посессивная связка
 PTCP.POT — потенциальное причастие
 PTCP.PRIV — отрицательное причастие
 PTCP.RES — результативно-стативное причастие
 QUAL — предикативный показатель качественных глаголов
 RECP — взаимное местоимение
 SBJV — конъюнктив
 SG — единственное число

Библиография

- Выдрин В.Ф.* Электронные корпуса африканских языков: завтра или послезавтра? // Петербургская африканистика. Памяти Андрея Алексеевича Жукова / Отв. ред. А. Ю. Желтов. СПб.: СПбГУ, 2008. С. 29–39.
- Выдрин В.Ф.* Электронный глоссированный корпус текстов языка бамана: первый этап // Mandeica Petropolitana II / Отв. ред. В. Ф. Выдрин. СПб.: Наука, 2011. С. 343–380. (ACTA LINGUISTICA PETROPOLITANA. Труды Института лингвистических исследований РАН / Отв. ред. Н. Н. Казанский. Т. VII. Ч. 2).
- Выдрин В.Ф., Томчина С.И.* Манден-русский словарь. СПб.: Дмитрий Буланин, 1999. Т. 1. 342 с.
- Bailleul Ch.* Dictionnaire Bambara-Français. 3^e éd. Bamako: Donniya, 2007. 476 p.
- Dumestre G.* Dictionnaire bambara-français suivi d'un index abrégé français-bambara. P.: Karthala, 2011. 1187 p.
- Kilgariff A.* Putting frequencies in the dictionary // International journal of lexicography. 1997. Vol. 10, No. 2. P. 135–155.

Kilgarriff A., Rundell M., Dhonnchadha E. Úi. Efficient corpus development for lexicography: building the New Corpus for Ireland // *Language Resources and Evaluation Journal*. 2006. No. 40 (2). P. 127–152.

Vydrine V. avec la collaboration d'Adama Diokolo Coulibaly. Classification des verbes réfléchis // *Mandenkan*. 1995. No. 29. P. 1–114.

Vydrin V., Rovenchak A., Maslinsky K. Maninka Reference Corpus: A presentation // *TALAf 2016: Traitement automatique des langues africaines (écrit et parole) Atelier JEP-TALN-RECITAL 2016*. Paris, le 4 juillet 2016. Actes. URL: http://talaf.imag.fr/2016/Actes/VYDRIN_ET_AL%20-%20Maninka%20Reference%20Corpus:%20A%20Presentation.pdf (дата обращения: 06.11.2017).

TOWARDS CORPUS-DRIVEN DICTIONARIES FOR MANDING LANGUAGES

V.F. Vydrin

Since last decade, big text corpora have been built for Maninka and Bamana languages. While working on the corpora, electronic dictionaries for these languages are being built too. Time is coming for launching projects for building dictionaries of a new generation, i. e. corpus-driven dictionaries. In such a dictionary, each lexeme could be represented in a much more elaborated way than in “traditional” dictionaries. This would concern the polysemy of a lexeme, its collocations and valencies, stylistic and regional characteristics of each sense, and other lexicographical aspects. As a sample, the structure of an entry for the Bamana verb *bánban* is presented.

Keywords: Bamana language, Bambara language, Maninka language, Malinke language, Manding, corpus-driven lexicography, polysemy.