Sara-Bagirmi Languages Project Database[1]

The database is currently stored in dBase format. It consists of:

- a single *Languages Table* (tables.dbf), which general information about each project language together with the names of the dBase tables used to store the words, translations, sample sentences and expressions. For example, for the language Mbay, the value for *EntryTable* will contain the 'Mbay.dbf", the value for *TranslationTable* will be 'Mbay_Eng.dbf', the value for *SampleSentenceTable* will be 'MbaySamp.dbf', and the value for *ExpressionsTable* will be 'MbayExpr.dbf'.
- a set of related tables for each language that contain the words, translations, sample sentences, and expressions, as well as lookup tables.

For each table in the database there is a dBase .mdx file which contains the indexes for that table. In dBase, relationships between tables are not stored in the database, and need to be set up using code. [2]

The *Language Table*, which will contain one row for each of languages included in the database. This table will include the following information:

1. The *Language* string field contains the name of the language.
2. The *TranslationLanguage* string field contains the name of the language into which the base language is to be translated. For this project, the translation language is French.
3. The *EntryTable* string field contains the name of the table containing the dictionary entries.
4. The *TranslationTable* string field contains the name of the table containing the translations for the entries.
5. The *SampleSentenceTable* string field contains the name of the table containing sample sentences for the translations of the entries.
6. The *ExpressionsTable* string field contains the name of the table containing expressions and idioms for the translations of the entries.
7. The *FontName* string field contains the name of the font to be used when displaying the language data. This gives the dictionary author the ability to display language in a font appropriate to the language. The format of the font is irrelevant to the workings of the program.
8. The *LanguageNotes* string field contains the name of a .TXT or .PDF file containing information about the language: its geographical location, number of speakers, position in the Sara Family, and a summary of the status of the work completed (the number of words, sample sentences, and expressions, the status of recording progress, and information about sources, informants and process.
9. The SpellcheckTable string field contains the name of a table which is to be used to hold alternative spellings of words so that a spell check can be run on the language's sentences and expressions.

The database will then contain 5 tables for each of the languages. The "Entry Table" will contain:

1. The *EntryCode* field contains a string representation of a unique identifier for each entry.
2. The *Entry* field is a string field containing the word.
3. The *Phonetics* field is a string field containing the actual pronunciation of the word.
4. The *Loan* field is a single character field containing a loan language code. Values

that can be used in this field are taken from a "Loan Language" lookup table.

5. The *ToneSequence* field is a string field containing the sequence of tones found in the word. The value for this field will be generated programmatically from the word. This information can be used for statistical analysis of tone sequences.
6. The *SoundFile* field is a string field containing the name of the sound file containing the word spoken by a native speaker.
7. The *Discrepancy* field is a Boolean field that allows me to mark cases where the sound file recording is different from the transcription.
8. The *Source* field is a string field containing the source for the entry.
9. The *SaraWordListCode* field is a string field containing the code for the Sara Word List item to which the entry is associated. If it is not associated with an item from the Sara Word List, this field will be Null.
10. The *Recorders* string field will contain a comma-separated list of codes that indicate the Chadian collaborator(s) who made the recording.

There is a one-to-many link between the "Entry Table" and a "Translations Table" based on the unique *EntryCode* field, thereby capturing the fact that there can be multiple translations for a single word. The fields of the "Translation Table" will include:

1. The *EntryCode* field is used to identify the entry for which the translation is being provided.
2. The *TranslationCode* field contains a string representation of an identifier for the translation. When used with the EntryCode it forms a unique value used as a composite key for links to other tables (e.g. the "SampleSentence" table).
3. The *BriefTranslation* field is a string value containing an abbreviated form of the translation to be used in the generation of the "Target language" lexicon (e.g. English-Mbay or French-Ngambay).
4. The *Translation* field is a string field containing the full translation into French or English.
5. The *Category* field is a string field containing the category, or part of speech, of the translation. The values that can be used in this field are taken from a the configuration file for each language.
6. The *Grammatical Notes* string field contains any additional grammatical or semantic information about this use of the word with this translation.
7. The *Synonym* string field contains any synonyms for this translation of the word.
8. The *TextFile* string field contains the name of a .PDF file with a longer descriptive text when available.

There is a one-to-many link between the "Translations Table" and a "Sample Sentences Table", thereby capturing the fact that there can be multiple sample sentences for a single meaning of an entry. The fields of the Sample Sentence table include the following information:

1. The *EntryCode* field and *TranslationCode* field combine to form a unique value representing the specific translation of a specific entry for which a sample sentence is being provided.
2. The *SampleSentence* string field contains the sample sentence for a particular meaning of an entry.
3. The *SampleTranslation* string field contains the translation of the sample sentence.
4. The *Source* field is a string field containing the source for the Sample.
5. The *SoundFile* string field contains the name of the sound file containing a sound recording of the sample sentence.
6. The *SampleDiscrepancy* field is a Boolean field that allows me to mark cases where the sound file recording is different from the transcription of the sample

sentence.
7. The *Recorders* string field will contain a comma-separated list of codes that indicate the Chadian collaborator(s) who made the recording.

There is also a one-to-many relationship between the "Translations Table" and the "Expressions Table", capturing the fact that there can be multiple expressions and idioms for a single meaning of an entry. The fields of the "Expression Table" will include the following information:

1. The *EntryCode* field and *TranslationCode* field combine to form a unique value representing the specific translation of a specific entry for which an expression is being provided.
2. The *Expression* string field contains the expression.
3. The *ExpressionTranslation* string field contains the translation of the expression.
4. The *ExpressionType* string field contains the type of expression (e.g. a Verbal Expression, Serial Verb, etc.).
5. The *SampleSentence* string field contains the sample sentence using the expression.
6. The *SampleTranslation* string field contains the translation of the sample sentence.
7. The *Source* field is a string field containing the source for the Expression.
8. The *SoundFile* string field contains the name of a sound file containing a sound recording of the sample sentence spoken by a native speaker.
9. The *ExprDiscrepancy* field is a Boolean field that allows me to mark cases where the sound file recording is different from the transcription of the expression.
10. The *SampleDiscrepancy* field is a Boolean field that allows me to mark cases where the sound file recording is different from the transcription of the sample sentence. The *Recorders* string field will contain a comma-separated list of codes that indicate the Chadian collaborator(s) who made the recording.

There is also a “Spell Check” table for each language. This table is used to include conjugated words and other words that do not appear in the dictionary, and to link them to words in the dictionary so that their meaning can be found programmatically. For example, the word `àl̄` ‘to go up’ is found in the Gor entry table, but `āl`, the form used for the 1st person singular and 2nd person, is not found. When the language software detects a word in a sample sentence that is not in the dictionary, the user is allowed to add that word to the spell check table and to associate it with a word that is in the dictionary (e.g. `āl` with `àl̄`). Subsequently, the word does not show up as an error, and users are able to see its origin and its meaning. This table is a lookup table with the following field structure:

1. The *Entry* field is a string field containing the word which is not found in the dictionary.
2. The *RealEntry* field is a string field containing the word in the dictionary that it is derivied from.
3. The *EntryCode* field is the unique value for the word in the dictionary.
4. The *Translate* field is the translation (currently in French) for the word that was not found in the dictionary.

In the next version of the project database, three new tables will be added. First, there will be a new “Language Dialect” lookup table which will contain the names of dialects for a given language:

1. The LanguageName field is used to identify the name of the language.
2. The DialectName field indicates the name of a dialect for that language.

Then there will be a one-to-many relationship between the "Entry" table and a new "Dialectal Variant" table for each language.

1. The *EntryCode* field is used to identify the entry for which the translation is being provided.
2. The *AltEntry* field contains a string value indicating the alternative pronunciation for a word.
3. The *DialectName* field will indicate the name of the dialect in cases where a specific dialect name exists. Only Dialect Names found in the "Language Dialect" table are permitted.

In cases where there is no dialectal variant for a word, there will be no entry in the "Dialectal Variant" table.

There will also be a new "Recorders" table which will contain information about the person who recorded each word or sample:

1. The *RecorderCode* field is a string field containing a unique code identifying the person who made the recording.
2. The *RecorderName* field is a string field containing the full name of the person who made the recording.
3. The *Language* field is a string field identifying the language for which he/she makes recordings.
4. The *FathersLanguage* field identifies the native tongue of the recorder's father.
5. The *MothersLanguage* field identifies the native tongue of the recorder's mother.
6. The *SpouseLanguage* field contains the native tongue of the recorder's spouse.
7. The *HomeLanguage* field primary language spoken at the recorders's home.
8. The *OtherLanguages* field contains a comma separated list of other languages the record speaks.
9. The *BriefBio* field contains text information providing some biographical information on the recording, including place of birth, education and where he/she has lived.

[1] For historic reasons, the actual field names with tables are not the same as used in this document, as the tool used to create tables only permit fieldnames of 10 characters. Users of these tables should be aware of the following mappings:

Entry Table:	*ToneSequence* is actually named *Tone*
	SaraWordListCode is actually named *SaraWdCode*
Translations Table:	*Translation* is actually named *Translate*
	BriefTranslation is actually named *BriefForm*
	GrammaticalNotes is actually named *GramNote*
Samples Table:	*SampleSentence* is actually named *SampleSent*
	SentenceTranslation is actually named *Trans_Sent*
Expressions Table:	*Expression* is actually named *Idiom_Expt*
	ExpressionTranslation is actually named *Trans_Exp*
	SampleSentence is actually named *SampleSent*
	SentenceTranslation is actually named *Trans_Sent*

[2] There is a major bug in the dBase Borland Database Engine which prevents its use with blob, as in time the data becomes invalid and unusable. For this reason, the database contains the names of the sound files rather than the sound data itself.