

## Data entry with the BURS Database Server

Alexis Dimitriadis

This version: 6 October 2008

### Contents

1. Background
2. System requirements
3. The data entry process
4. Important points (Make sure you read this before editing)
5. Other editing tips
6. Linking to sentences
7. Authentication
8. Things to watch out for
9. Directions for glossing

### 1. Background

The software for this server was created by Floris van Vugt (programmer) and Alexis Dimitriadis, for the Berlin-Utrecht Reciprocals Survey.

The system is designed to house information about languages in a flexible and highly configurable fashion. It is based on a general framework that should allow other data collections to be easily supported. The database design is built around the following sorts of entities:

- An **answerset** represents the results of one completed questionnaire. In the typical case there will be one answerset per language, but the system allows multiple answersets to be completed for the same language, representing different speakers or dialects.
- A **marker** is (for our purposes) a reciprocal. A language may have any number of markers, which are the object of study (e.g., *sich* and *einander* for German). In some error messages you may see markers referred to as *strategies*.
- A **sentence** is a fragment of text in the language under study, with an associated gloss and translation. A sentence is always associated with a language, but does not necessarily contain one of the markers under study (in many cases it is useful to include such sentences in the database).

These three types are the primary **Entities** of the database. They represent the objects of the target language that can be described. The database defines **questions** (attributes), like *Language Name* or *Compatible situation types*, which represent all the properties of an entity for which a value can be given. In addition to the *value* of a question (a number, a choice from a fixed list, or free text), many questions allow the answer to be linked to one or more example sentences. Usually there is also a Comments field for each question.

The system is unusual, among databases of this type, in that most questions are not hard-coded into the definition of the database: they are maintained as a separate set of tables that list the questions applying to each entity, and the type of answer(s) that will be accepted. Knowledge of this component should not be necessary for data entry. Just be aware that it is there, and that it is very easy to add new questions, or add values to a list of choices (“enumerated values”) if none of the existing choices fits.

## 2. System Requirements

Server entry point: <http://language.link.let.uu.nl/burs/>.

The system is accessible with an ordinary **web browser** (and requires an internet connection, of course). It has been tested with both Firefox and Internet Explorer. You should have javascript turned on (it is probably turned on already). No other plug-ins are required.

If you will be working with data that uses IPA characters or other non-English characters, you will need a suitable **Unicode font** with the necessary characters. The database cannot supply such fonts; they must be installed on your browser. If you already use such a font for other work, simply set it as the default font from your browser (from the preferences panel). Otherwise, download a wide-coverage Unicode font and set it as the default font. You can use SIL's Gentium font ([scripts.sil.org/Gentium\\_download](http://scripts.sil.org/Gentium_download)).

## 3. The data entry process

To enter data, you must be logged in and select an answerset (see above). You then start from the *Data Entry* link on the sidebar; you can edit the properties of the answerset by clicking on the *Modify* link, add a new reciprocal marker to the answerset, or answer questions about one of the existing markers. You can also enter new sentences (which you can later link to, as you answer questions about the markers).

The server is not meant to fully guide data entry: it should be used in conjunction with the printed questionnaire. The “prompt” sentences, for example, are not shown by the server.

Questions about the markers are divided into sections, corresponding to the sections of the printed questionnaire. Select a section, then click on one of the links *Modify Morphology*, *Modify Binding*, etc., to edit the relevant section. Be sure to always use a button, not a hyperlink, to submit an edited form to the server.

The navigation is a bit more involved than it absolutely needs to be, but it does not take very long to get used to. I hope to streamline it further in the future.

## 4. Important points: Make sure you read this before editing

Editing and navigation rely on the facilities and **keyboard shortcuts** provided by your browser and the windowing system. E.g., you can **copy** and **paste** text (with control-C and control-V) between applications or form elements, and open additional browser windows by right-clicking on a link and choosing “Open in new window”. You can also use the browser's **Back** button if you find yourself somewhere you don't want to be. Finally, I recommend ensuring that your browser's **form completion** features are activated. When entering sentences, form completion can help you avoid a lot of typing by pasting and modifying recently typed sentences.

**Important:** Many data entry forms contain both action buttons and hyperlinks (in the form and in the sidebar). If you press a button, your changes to the form will be committed to the database (and any other action requested by the button, e.g., “Delete”, will be carried out). But if you click on a regular hyperlink, the form is *not* submitted. Any information you have entered is discarded. This is a limitation of the http form protocol; it's easy to live with, as long as you are aware of it.

While editing in a form, many questions are answered by choosing from a list of choices instead of typing free text. There are two kinds of “controls” for doing this: a **drop-down menu**, which allows you to choose *one* option (or none, if you select the empty line); and a **list box**, which displays several options and allows you to select as many options as appropriate (“choose all that apply”). To **select multiple values** in list box, hold down the Control key while clicking on

the additional values. Control-clicking on a selected value (even the only selected value) will de-select it. This is how you **cancel an answer** to a question.

Most forms have a **Reset** button. This tells the browser to restore the form to its state when the page was first loaded, discarding any typing or selections you have made. This is entirely local to your browser: the form is not re-fetched from languagelink.

If you open multiple browser windows they will function independently, and should work as expected (unlike those of many e-commerce servers).

Be aware that a browser window displays a *copy* of the data that is in the database; this is similar to opening a file in a text editor. If someone changes the data in the database while you're editing your copy, you will overwrite their changes when you submit your form. Thus the same piece of data should not be edited simultaneously in two separate windows (by two different users or by the same person). The database does not enforce this; if each answer set is the responsibility of only one person at a time, there should never be a problem.

## 5. Other editing tips

To cancel an answer in a drop-down menu (which does not allow multiple (de)selections), select the blank line that is the top element of the list. If there is no blank line, you *must* choose an answer.

The size of the text boxes is for convenience only: The database is very generous in the length of the answers it accepts. If you want to put a long passage in a small text box, just keep writing and the text will scroll. (It will be displayed properly in the read-only views).

If you create an object by mistake, you can delete it. This is always safe if the object has not been used yet. Deleting a marker or sentence will also delete all the answers that have been entered about it, but it is an error to delete an answer set if it contains any markers or sentences: Delete them first, then delete the empty answer set.

There are additional considerations involved in deleting objects in the Questionnaire Management section; they are not discussed in this guide, so avoid deleting (or creating) anything in that section unless you have an understanding of its workings.

## 6. Linking to sentences

Many questions allow you to link a sentence to the answer. There are two types of answers: One is a list of options, allowing a separate example link for each selected option. The other type, which is more common, associates one or more example links with the entire question (not with each part of the answer).

The linking process is not particularly intuitive at the moment. You will see a search form, meant to help you link to existing sentences. If you want to enter a new sentence to link to, use the link [New sentence](#) at the top. After you create a new sentence, you will find yourself in a screen with the hyperlink [Link to the answer](#) near the top of the page. Click on it to make the link.

To link to an existing sentence, fill in any search conditions and click on the *Search* button. To view all existing sentences, click on the *Search* button without entering any search terms (but you'll probably want to check the box next to your language). When you find the sentence you want, open it with [Details](#) and you'll again see the hyperlink [Link to the answer](#) at the top of the page. Click on it to make the link. Removing a sentence link (with the [Remove link](#) button) does not delete the sentence itself.

## 7. Authentication

To enter data into the database, you must log on. Volker or myself can create user accounts. There are currently two versions of the database; one allows browsing and searching without a password, and contains only those languages that we consider ready to show; the other version is only accessible through the **Login** button, and contains the complete data collection.

When you log on, select an Answerset that you will be working on (choose one of the existing ones, or click on “New” to start working on a new language. Once you log on, you have a **session** identifier (visible as a long string of nonsense characters in the browser’s address bar), which keeps track of your permissions. You can log on as multiple sessions, or open new windows from the same session, etc. The session will expire after a long period of inactivity. You’ll then need to log on again.

It is best if you only work on one language at a time—there are some kinds of changes that you can make without switching languages, but other kinds are not entirely safe; so it is best to avoid doing so altogether. To switch languages, log out and log back in again.

Once you are logged on, the database will allow you to delete or modify any entity or value. This does not mean you should! The project participants should decide who is allowed to edit what, and hold to that policy without help from the server (this is not a bank!) Note that there is now a lot of valuable data in the database, so until you’re sure you know what you’re doing, **don’t delete or edit anything you did not create yourself.**

## 8. Things to watch out for

The system still suffers from some bugs we are aware of, and probably others that we do not know about. Be on the lookout for surprising behaviour, and **make sure to let me know about it** so I can try to address it.

Any pages mentioning “Server error”, or scribbling error messages over the sidebar, indicate a bug. This is never supposed to happen.

You should be able to enter text longer than the text area in a form; but be on the lookout for accidental truncations. If you notice that the tail end of something you type gets discarded, let me know.

*Known bugs:*

**Do not edit** any sentence from a language other than the one you have logged on for; if you do, its link to the reciprocal marker it uses will be lost.

It is currently possible to link an answer multiple times to the same sentence. This is counterintuitive, but does no harm. The extra link can simply be deleted.

*Important limitations*

The Sentence data entry form has a section on “Symmetry properties”. This section only needs to be completed if you are entering answers to the Discontinuous Reciprocals section, but there is at present no way to selectively suppress its display. Just ignore the section unless it’s relevant.

Linked examples are not displayed on the same page as the question. In lists of examples, the gloss and translation are displayed side by side, not below the text. It would be nice to fix this in the future.

There is no validation or support for the gloss abbreviations used. This is the next major feature that will be implemented if we continue development of the server.

## 9. Directions for glossing

The examples you provide should be accompanied with interlinear morphemic gloss. Please observe the following conditions in your glosses.

1. The sentence form provides three word-aligned lines (“tiers”), and a translation line. If you are working with a morphologically simple language, you can use just one of the “original” and “morphemic” tiers. If you use both tiers, there should be dashes on the morphemic tier only.
2. Use dashes (-) to separate the morphemes on both the text and the gloss tier. **There should always be the same number of dashes on the two lines.** If there are multiple gloss labels for the same morpheme, separate them with a period, not a dash.
3. On the gloss tier, capitalize the entire abbreviation for every word that is not a translation. For example: *eat-PST.PASS*.
4. You don’t need to align the words on the text and gloss tiers with extra whitespace—this is handled automatically by the webserver when it displays the data. But if you wish to insert extra whitespace anyway because it helps you edit the text, it’s ok to do so.
5. Do not use glosses that repeat the word being glossed (e.g., DE for *-de*). Your gloss can be accurate or vague depending on the situation. If you really don’t know what a morpheme does, you can use PRT (“particle”), XXX (“fix later”), or “?”.
6. Gloss a reflexivizing morpheme as REFL, a reciprocal as RCP. For a morpheme that is ambiguous between several interpretations (reflexive, reciprocal, middle, or some such), use the gloss RR (for “reflexive-reciprocal”). If you use RR, use it for *all* occurrences of the morpheme, not just in ambiguous sentences. If a sentence is ambiguous, this can be indicated in the translation (wherever possible) and/or the Comments field.
7. Person and number should not be separated by a period when they are expressed by the same morpheme (1PL, 2sg). In this case no capitalization is necessary.
8. On other matters of glossing protocol, please follow the *Leipzig glossing conventions*. They are documented at

<http://www.eva.mpg.de/lingua/files/morpheme.html>

The Leipzig rules include a list of abbreviations for common terms. If a term is included in the list, please use the recommended abbreviation.

9. The project policy is to retain Latin-based orthographies, but to transcribe text from other orthographies (e.g., Cyrillic, Hebrew, or Kanji), or from languages without a standard orthography. Use an appropriate Unicode font (see section 2 above) for any non-Latin symbols, including IPA transcription. If you need a way to enter the required symbols, you can use an on-screen keyboard such as the TDS IPA Console (<http://language.link.let.uu.nl/tds/ipa/>).