Guidelines to the *APiCS* Questionnaire for Contributors

December 2006

1. Introduction

Managing the vast amounts of data for the *Atlas of Pidgin and Creole Language Structures* (*APiCS*) is a demanding task for both the editors and the contributors. In order to facilitate data collection and organization, we used the computer programme FileMaker Pro 7 (a widely used commercial database application) to prepare a computer-based questionnaire into which the data can be entered directly. If you have never worked with FileMaker Pro (or a similar application) before, it may take you a while to get used to the questionnaire. But we are confident that, after a short familiarization phase, you will discover that it is not so complicated after all, and quickly start to enjoy its benefits. (FileMaker runs both on Macs and Windows computers.) If you are familiar with FileMaker Pro or similar applications, you can probably skip sections 2-4 and continue reading with section 5.

This guide is intended to help you get started, but is not meant as a full reference manual. Introductions to Filemaker are available from various websites. We will make every effort to help you and solve your problems as quickly as possible.

A technical note for this manual: Text in `this typeface` refers to a command you can access through the menu bar at the top of the screen. So, `File → Print...`, for instance, would mean: Click on the `File` menu, then select `Print...`

2. Getting your questionnaire

Please download your version of **the *APiCS* Questionnaire** from the web site http://email.eva.mpg.de/~michaels/apics/catalogue.html. If you do not have FileMaker installed on your computer, please use the FileMaker runtime version (19 MB/17MB) of the questionnaire. This gives you all the relevant functionality of the database programme (searching, sorting, duplicating records, etc.), but you cannot modify the questionnaire.

To work with the FileMaker runtime version, you need to download the zipped file, unzip it (don't forget this!), and open the file called "APiCS [.app]" (for Mac) or "APiCS.exe" (for Windows). When you enter data into the fields, they will automatically be saved into a file called "APiCS.USR". It is this file that we ask you to return to us when you have finished.

If you have the programme Filemaker Pro (7 or 8) installed on your computer, it is more convenient for you to work with the regular (non-runtime) FileMaker file. Again, you need to unzip it before you can use it. Please send the entire file back to us when you're done. (In this case you can in principle modify the questionnaire, e.g. modify the layouts to make them suit your needs better, or even create additional fields. But please do not delete or rename any fields, because otherwise we will not be able to use your data.)

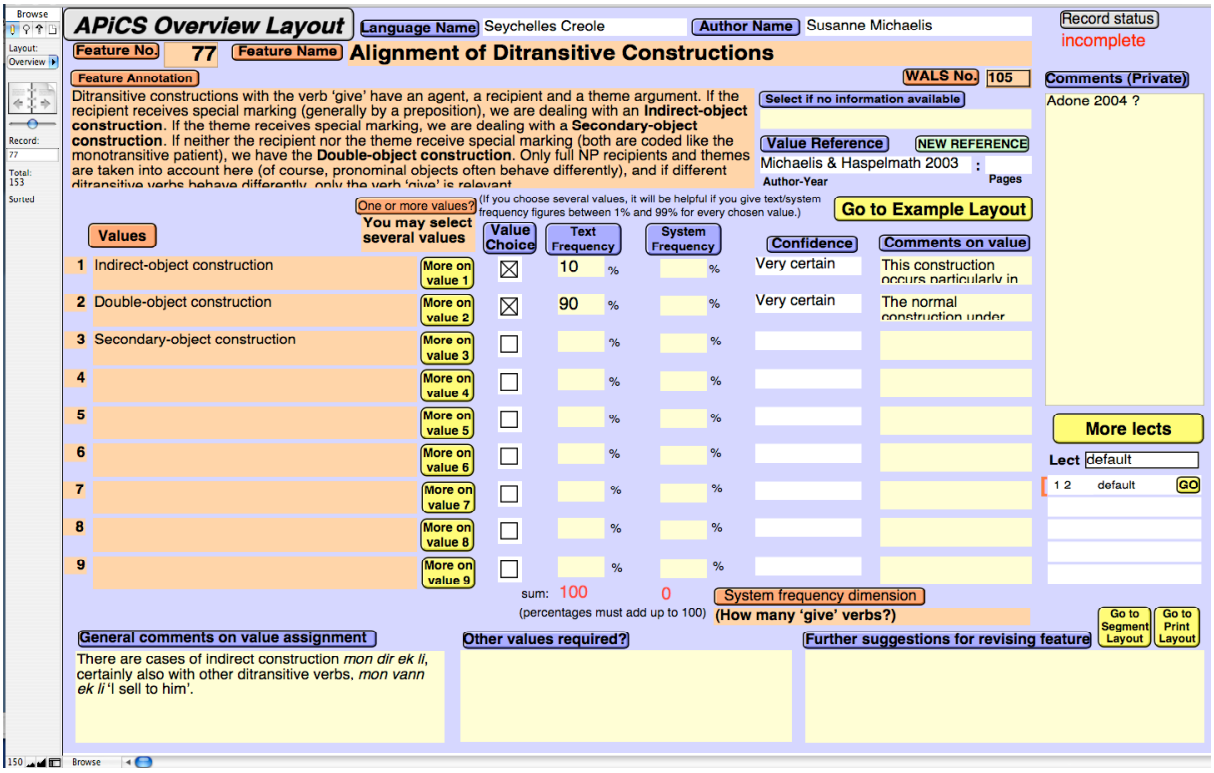
Note:

Please be aware that any **changes** that you make will **take effect immediately**, i.e. (a) you do not need to save anything manually; unlike in the best known word processing programmes, there is no `File → Save` command, and (b) it is recommended that you make regular backups of the file "APiCS.USR" before you start working, because you cannot automatically undo any changes that you have made.

3. Orientation

When you first open the questionnaire, it should look structurally similar to the following picture. (Screenshots are taken from the Mac version of FileMaker Pro, but Windows users will recognize the elements.)

Do not be confused if your screen displays more empty fields – it is a questionnaire, after all. The picture shows one into which some data has already been entered.



3.1. Fields

The **white** and **light yellow** fields await your input: While the white fields are for required information, input in the yellow ones is optional.

The **light orange** fields contain information entered by the editors. If you encounter a field where the content could not be fitted into the box in the layout, you can click on it to have it all displayed – but please do not alter the content of light orange fields, as alterations of these fields would be lost when you return the questionnaire to the editors. If you feel that a feature should be revised, please leave the editors a comment in the separate "Further suggestions" field.

3.2. Buttons

Accompanying the fields for storing information are buttons that perform a function when you click on them. We encourage you to do so freely; the buttons cannot do any harm to your data.

Most of the buttons serve as captions for a field or set of fields. This is true in particular of the **orange** and **blue** buttons. Clicking on such a button will open a window that tells you more about the field associated with it.

Clicking on a **yellow** button will take you to a different "layout" (= view of the data, see below). The title of the current layout is displayed in the **grey** button.

The **green** button lets you create a new reference as a source for examples that you cite.

3.3. Layouts

Because displaying all the fields of the database at once would be very confusing, they are arranged into different "layouts", each of which contains only selected fields. In order to fill out the whole questionnaire, you will have to switch between layouts. Some fields, however, appear in more than one layout. In these cases, information you have entered into a field in one layout will automatically be displayed in other layouts containing this field; you do not need to enter anything twice.

The central layout is the **Overview Layout** depicted above, in which you can enter your name, the language for which you are providing information (these fields need to be filled only once in the questionnaire), and assign values (together with a level of confidence that your assignment is correct) to each of the 153 features, and cite a source for your value assignment.

In the **Example Layout**, you are asked to provide at least one glossed example from your language that illustrates the current feature, and give the type of the example together with a source from which the example is taken.

For each value, there is a separate **Value Layout** for more detailed information on the individual value. Most of this layout's fields appear either in the Overview Layout or in the Example Layout as well, except for (a) the Annotation field, which gives criteria for assigning this value or examples for languages that have this value, and (b) a field where you are asked to specify which of the examples you have entered in the Examples layout are relevant to this feature value.

In the **Segments Layout**, we would like to know which of 96 phonological segments that are frequent in the languages of the world occur in your language as well. Each phone comprises a separate record (see below). The **References Layout** stores bibliographical information on the sources from which you take your data. As in the Segments Layout, each reference forms a separate record.

Finally, the **Print Layout** presents the data in such a way that you can see the feature annotation, the value names and the value annotations together. This is the best layout for printing the questionnaire, e.g. if you want to work on it away from your computer. (However, you will not be able to print the full content of all fields in this way.)

3.4. Records

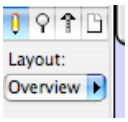
All the fields from all layouts belonging together in one *table* of the database form one record. Thus, there is a *data table* containing the information on each of the 153 features (comprising the fields from the Overview, Example, and Value Layouts), with one record for each feature/language pair, totaling 153 records per language. As the phonological and bibliographical data are globally relevant to the entire language and not just to a particular feature, they are stored in separate tables (the *segments table* and the *references table*, respectively) and therefore have different record counts (96 records in the segments table, and initially 3 ### records in the references table).

4. How do I…?

4.1. enter data

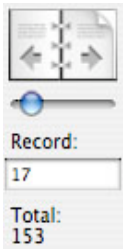
Data input starts with a click in the appropriate field. Most fields take text entry, so you will see a blinking cursor upon your click and can just start typing. In some fields your click will open a drop-down menu from which you can select a value. For the Value Choice, your click will simply check the box (meaning "yes, my language has this value") or uncheck it.

4.2. switch between layouts



The easiest (and probably most convenient) way of switching to a different layout is by clicking on the corresponding yellow button. Alternatively, you can use the drop-down list under “Layout:” on the sidebar.

4.3. switch between records



To switch from one record to another, e.g. if you want to proceed to a different feature, use the navigation elements on the sidebar. Clicking on the right page of the book symbol will take you to the next record, clicking on the left page to the previous record. If you want to skip more than one record at a time, you can drag the slider below the booklet symbol to move through the database. The field under “Record:” displays the number of the record at the current slider position. In case you know the number of the record that you want to jump to, entering it into this field and pressing Enter will take you directly to that record.

4.4. search the questionnaire database

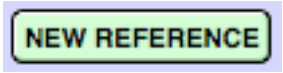
To search the database for some particular content of a field, you will need to enter FileMaker’s Find Mode by clicking on the magnifying glass symbol on the sidebar, or by selecting View → Find Mode from the menu bar. You will then see an empty questionnaire form. There you can enter the data for which you would like to search into the appropriate field in the appropriate layout. This may both be data that you entered yourself or data in the fields that the editors provided. Then click the “Find” button on the left (or simply press Enter). The program will return to Browse Mode (the ordinary view in which you can enter data into the records; you can go back there manually, e.g. if you want to abort your search, by clicking on the pencil symbol in sidebar or by selecting View → Browse Mode from the menu bar) and display only those records that match your search criteria. The number of such records is then indicated under “Found:” on the sidebar. To view **all records** again, select Records → Show All Records from the menu bar.

A powerful tool for searching the database are so-called **wildcards**. These are special characters that serve as placeholders and allow you to search for unspecified elements. For example: Entering only an asterisk (*) into a text field in Find Mode will return all the records in which this field is *filled with anything*. Entering only an equals sign (=) will return all the records in which this field is (still) *empty*.

4.5. sort records

After you performed a search, the records are arranged in a different order, and the last line in the sidebar will say "Unsorted". To sort the records, ctrl-click (or click with right mouse button) the field by which you want to sort the records, and select Sort Ascending from the menu. (Alternatively, you can select Records → Sort Records from the menu bar and then specify a field by which the records should be sorted.) You will need the sorting function primarily to sort the records by the feature number.

4.6. add a new reference



The most convenient way to add a new source is via the “New Reference” button in the Overview and Example Layouts. Once you

have entered the required information, it will be available from the drop-down lists in the Overview and Example Layouts. Clicking on the “Back” button will take you back to where you were before.

Alternatively, you may switch to the References Layout and create a new record (Records → New Record) there, but then you will have to return to the other layouts manually. The References Layout is handy, however, if you need to modify a reference you already entered (e.g. to complete biographical information).

5. The contributor-filled fields

5.1. Outline

Every *APiCS* contributor is asked to provide data on the 153 structural parameters ("features") that were selected by the editors, in the pidgin or creole language of their expertise. There are 12 obligatory fields per feature, and you can enter more data into about 40 further optional fields. Most importantly, you are asked to choose one (or more) values for each structural feature, and to provide at least one example for each feature.

5.2. The obligatory fields

With each value, you are requested to give information on the degree of reliability of the information ("confidence", with five values: very certain, certain, intermediate, uncertain, very uncertain). This allows you to provide also information that you are not totally sure of. You are also asked to provide a source for the information. This source may be "own knowledge" of the language, but it may also be published work on the language (by yourself or by someone else).

5.3. The optional *fields*

You may provide prose comments in a number of fields. First, there is a general field for explanation ("General comments on value assignment") that is intended to be read by the *APiCS* users, so this should be formulated clearly and without abbreviations. Next, there are two fields for comments that are intended for the editors, to help them improve the database in the second round. The field "Other values required" can be filled with a comment if the values provided do not seem sufficient to capture the language's structure. The field "Further suggestions for revising feature" is for any other comments that might help the editors improve the questionnaire. There is also a "Comments (private)" field for notes written by you for yourself or for your assistant, concerning the nature of the data entry process (e.g. "Need to check with informant", etc.). The content of this field will not be read by the editors.

5.4. The example fields

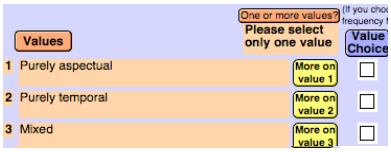
You may provide up to five examples for each feature, and you should provide at least one example. For each example there are nine fields, of which six are normally obligatory: (i) The primary text field, (ii) the gloss field (which gives a word-by-word gloss, or a morpheme-by-morpheme gloss), (iii) the translation field, (iv) the "example type" field, which gives information on whether the example is elicited, spoken, written, etc., (v-vi) two source fields, with author-year and page information. In addition, there is an optional field for an analyzed version of the text (especially a version with hyphenation for morpheme breaks, if this is seen as useful), an optional field for a translation into a language other than English (especially the main contact language, or the lexifier language), and an optional fields for any kinds of comments.

6. Choosing a value

The most important part of the questionnaire is the choice of a value for every feature in the Overview Layout.

6.1. Single-value selection

In the simplest case, only one value may be chosen. For example, in feature no. 47 ("Tense-Aspect Systems"), three values are proposed:

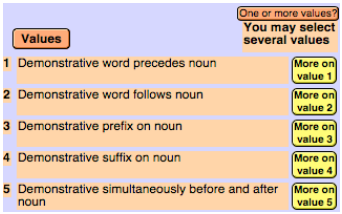


1. Purely aspectual (system)
2. Purely temporal (system)
3. Mixed (system)

These three values are exhaustive, and selecting several simultaneously would make no sense. The orange field next to the Value Choice button ("One or more values") therefore tells you: "Please select only one value". So here all you need to do is to select one of the three values by clicking the Value Choice field next to it. The frequency fields are irrelevant. About two fifths of all features are of this type.

6.2. Multi-value selection

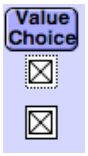
In the other three fifths, you may select several values, because a given language may have several of the proposed possibilities. For example, feature No. 66 ("Order of Demonstrative and Noun") has the following five values:



1. Demonstrative word precedes noun
2. Demonstrative word follows noun
3. Demonstrative prefix on noun
4. Demonstrative suffix on noun
5. Demonstrative simultaneously before and after noun

Here the orange "One or more values" field tells you: "You may select several values".

If your language has just demonstrative words that always precede the noun (e.g. English), then you select just one value (the first one). (Choosing several values is never obligatory.)



However, some languages have flexible order of demonstrative and noun -- sometimes the demonstrative word precedes, and sometimes it follows the noun. If your language is like this, then you should select the Value Choice fields for both 1 and 2. Likewise, if different demonstratives in your language behave differently (say, 'this' preceding, and 'that' following the noun), choose both values.

6.3. The frequency fields

If you choose several values, then the frequency fields come into play, because the different possibilities may not occur equally commonly in language use or in the language system. For example, where word order is flexible, it may still be the case that one order occurs far more often than another order (i.e. the two orders may have a different **text frequency**). And where different demonstratives behave differently, it may still be the case that the great majority of them precede, while only a single demonstrative follows the noun (i.e. the two orders may have a different **system frequency**).

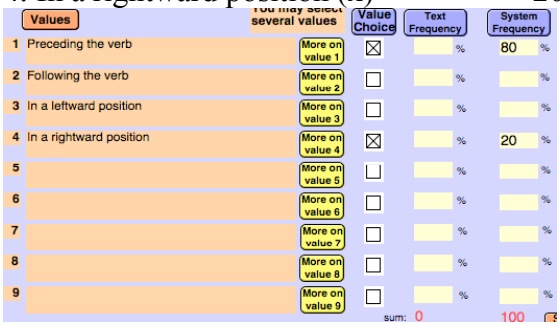
The frequency information that the questionnaire requests is the **relative frequency** of the various options. It is given in percentages, i.e. as a number between 1 and 99 (only integers should be used). The frequency figures of the chosen values must always add up to exactly 100, for example:

Feature 33 "Numeral classifiers"

	Text frequency:
1. Numeral classifiers are used (x)	30%
2. Numeral classifiers are not used (x)	70%

Feature 45 "Position of Tense-Aspect Markers"

	System frequency:
1. Preceding the verb (x)	80%
2. Following the verb	
3. In a leftward position	
4. In a rightward position (x)	20%



The database programme automatically sums up the frequencies that you enter and writes the sum at the bottom of the column with the percentages. This sum must be 100.

The questionnaire distinguishes between **text frequency** and **system frequency**:

6.3.A. Text frequency is relevant if *the same element of the language system* (marker, lexical item or construction) occurs with several of the values in texts. Thus, when the same demonstrative occurs before or after the noun, or when the same negative marker occurs before or after the verb, or when the same numeral occurs with or without classifiers, it will be very helpful to give text frequency information (if the frequency of the different values is not roughly equal).

Precise text frequency figures are of course hard to come by, and we are not asking you to do original corpus research. The purpose of this field is to allow you to provide information that you have anyway. We expect that there will be many cases where two values are possible, but it is very clear that one of them is a major option and the other one is a minor option. In such cases it is helpful if you give us even a very impressionistic assessment of the quantitative relationships (e.g. 90% vs. 10%, or 75% vs. 25%).

6.3.B. System frequency is relevant if the different values are due to *different elements of the language system*. Thus, when different demonstratives behave differently, with e.g. one preceding and one following the noun, or when different tense-aspect markers occur in different position with respect to the verb, or when some numerals occur with classifiers and some do not, then we ask you to give system frequency information. For example, if a language has five different tense-aspect markers, and four precede the while one occurs in a rightward position, then the figures 80% and 20% are given, as in the example immediately above.

Like the text frequencies, the system frequencies can be based on rough impressions. For example, if most adjectives follow, but only a handful precede the noun, then a 95%/5% system frequency assessment is extremely useful, even if it is based on impressions rather than actual counting of system elements.

So please don't regard the frequency fields as a burden, but as an opportunity for you to provide readily available information that would otherwise be lost. If you don't have clear

intuitions that one of the values is more frequent than another, just leave the frequency fields blank – in this case, the different values will count as equally widespread, which is probably close to the truth, because otherwise you would have noticed and would have developed clear intuitions.

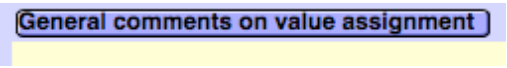
When is system frequency relevant, and along which dimension is it assessed?

While different values can always be assessed for text frequency (=token frequency), system frequency (= type frequency) is often irrelevant. That is, often there will be no (known) quantifiable system factors that lead speakers to choose between different options. For example, in feature 69 we ask about the order of the adverb 'often', the verb and the object. When a language allows several orders, it is unlikely that these will depend on certain verb classes or object NP classes, so in this case only text frequency will be relevant.

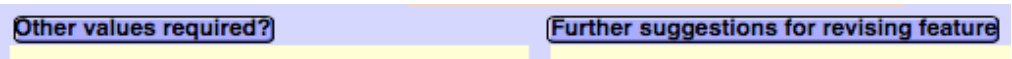
System frequency can be assessed along a variety of dimensions. To help you with assessing the system frequency, we have suggested the most straightforward options in the field "system frequency dimension". (E.g. for feature 41, Head or Dependent Marking of the Patient, we suggest as the most likely system frequency dimension is the number of Patient NP types.) When we can imagine a system frequency dimension, but it is actually not so likely that it will play a role, this is given here in parentheses.

7. Commenting on a feature

The Overview Layout contains four comments fields for each feature, plus comments fields for the individual values. The first ("General comments on value assignment") is for comments that are intended for the APiCS users. The text in this field should eventually be carefully edited (no typos etc.). The other comments fields are internal to the project and may have typos and all sorts of other imperfections.

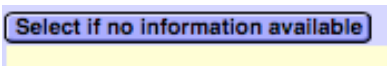


The second comments field ("Other values required?") and the third comments field ("Further suggestions for revising feature") are for the editors. The comments provided by the contributors in the first round will be the basis for revising the questionnaire for the later definitive version. The second field is for suggestions for adding values (in case the proposed values are insufficient for the author's language or are otherwise problematic), and the third field is for suggestions of any other sort. (Of course, such suggestions can also be communicated in other ways, e.g. by e-mail.)



The fourth comments field, "Comments (private)", is for any notes from contributors to themselves or their assistants. The contents of this field will not be read by the editors.

If you cannot give a value for a given feature, please specify one of the reasons for missing information in the field "Select if no information available". If the language is extinct and the information is unobtainable because of the limited corpus, select "Information unobtainable".



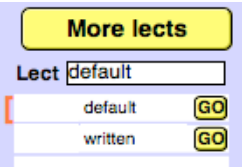
If the information could in principle be obtained but you happen not to have it, select "information lacking". (This is acceptable as long as not more than a handful of features are affected.)

8. More lects

In the normal case, you should provide information on a single, reasonably homogeneous variety (= lect) of your language, consistently for all the features.

However, in some exceptional cases, you may want to provide information on more than one lect, e.g. if you happen to know that for one feature the older generation uses a different construction, or rural speakers differ from urpan speakers, etc.

(If you want to give information on two different lects systematically for all features, you should provide several complete questionnaires (which we'll be very happy to accept! – but of course we don't expect this)).



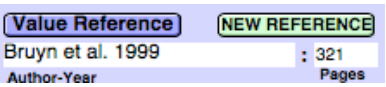
What you need to do if you want to give information on another lect is to duplicate the record for the feature by selecting Records → **Duplicate** Record from the menu bar. This will copy all the information contained in the current record to a new record and switch to that record, which you can then fill in in the usual way. Please specify the kind of lect in the field "Lect" in the lower right corner of the Overview Layout. The lect that you normally described is called "default". The additional lect can be given any name (the names "written" and "older generation" are just examples). The drop-down menu is editable and you can add any number of further lect names to it.

You can fill in the feature for up to four additional lects in this way. These are listed in the box in the lower right corner, and you can switch between them by clicking on the "GO" button to the right of the lect name.

(Please note that new records are always appended to the end of the data set, so in order to return to the record that you started out from, you may have to switch several records back. Alternatively, just search for the record you desire.)

9. References

You should provide at least one source for each feature. This can be your own knowledge of the language, or an informant (e.g. "Mary Jones, p.c."), or a published work. This should be entered into the two value reference fields, the Author-Year field and the pages field. The Author-Year field should consist of a single author surname plus year (e.g. Keesing 1988), or two author surnames plus year (e.g. Thomason & Kaufman 1988), or one surname plus "et al." plus year (e.g. Bruyn et al. 1999) in case there are more than two authors. The pages field is important and is obligatory (unless the work is a paper with a very narrow focus that only deals with the feature in question).

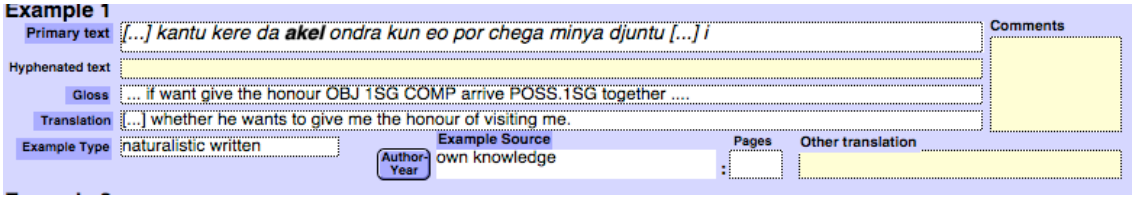


In order to avoid typing the same author-year information again and again, the author-year field is filled through a menu which includes values from the References table. This you can access via the button "New reference", and then you can enter a new reference. For each source that you enter there, please fill both the fields "Author-Year" and "Full Reference". (The precise format of the latter is not important at this point.)

If you feel that you want to give more than one reference, please include these in the "General comments on value assignment" field.

10. Examples

The Examples Layout has space for five examples, with nine fields for each example (six of which are are obligatory). Please give at least one layout for each example. Since giving examples is perhaps the most laborious aspect of filling in the *APiCS* questionnaire, we do not expect you to give more than one example in the normal case. But space for more examples was provided because some contributors may have easy access to examples, and if examples are available, they will eventually be very useful to the users.



The **Primary text** field contains the example text, as it would be written in any grammatical description, or as it was written in the example source. The transcription should follow the usual conventions for the language (standard spelling, or most widely used spelling, or most widely used transcription among linguists), or if you have good reasons for choosing a less widespread transcription, you may use this as well. Phonetic transcription is not necessary, not even for the phonological features.

For many of the features, the example will not be a complete sentence, but just a phrase or a single word. These cases are treated just like example sentences, i.e. one example goes into one field.

The optional **Hyphenated text** field may contain a version of the example text with morphemes separated by hyphens, thus allowing a morpheme-by-morpheme translation in the Gloss field.

The **Gloss** field contains a word-by-word or morpheme-by-morpheme translation of the example text. There should be an exact match between the number of words/spaces (or morphemes/hyphens) between the example text and the gloss text. If there is no exact match, you may get an "alignment error" message below the example fields.

Please use CAPITALS for morpheme abbreviations such as FUT or PROG. A list of stadnard abbreviations is included in the Leipzig Glossing Rules (###). These conventions should also be observed with regard to morpheme break issues.

The **Translation** field contains an idiomatic translation into English. In case the main contact language is a different one, it may sometimes be useful to give a translation into this language as well. The optional **Other translation** field is intended for such translations into another language (another possibility would be a translation into the main lexifier language, which may be useful for comparison).

In the **Example type** field, we ask you to specify whether the example was (i) elicited from a speaker, was taken from a naturalistic (ii) spoken or (iii) written text, or (iv) was constructed by a linguist. (Since there may be more possibilities that we did not think of, the value list is editable, i.e. you can add more values.)

In the optional **Comments** field, you can enter any kind of comments on the example.

Finally, the two **Example source** fields should be filled by information on the source. If this is a written work, you should fill in the **Author-Year** field (e.g. "Keesing 1988") and give page numbers in the **Pages** field. If the source is your own knowledge of the language, select "Own knowledge". If the example is from a personal communication, write "Mary Jones, p.c." (or similar) The conventions are thus very similar to those concerning the general feature references. But it is important to keep the two separate: The general feature reference

gives the source of the value assignment, while the example references give the sources of the examples.

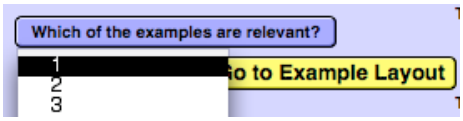
If you make use of special characters, it is essential that you use **Unicode** encoding. Two widely used special-character fonts with Unicode encoding are *Arial Unicode* (which is part of Microsoft's Office package and is therefore available on many computers) and Charis SIL (available for free from SIL International's website, http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=CharisSILfont)

11. The Value Layouts

A little more needs to be said on each value than has space on the Overview Layout, so there are nine additional Value Layouts, one for each value ("Value 1 Layout", "Value 2 Layout", etc.). These can be accessed by going to "More on value 1" etc. There are two crucial fields here:

(i) The orange **Value Annotation** field gives more explanation about what is meant by each value. In many cases, the value names together with the general feature annotation are sufficient to define the values, and in these cases the Value Annotation fields merely give a few examples (or sometimes just mention languages exemplifying the type). But in other cases, they provide further crucial information, so it is really important to study them before making a decision on the value.

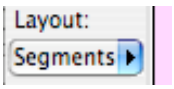
(ii) The **example relevance** field concerns the relation between the examples (if there are more than one) and the values. If several values are chosen, it will typically be very useful for the readers to have several examples, but it is possible that several examples illustrate the same value, or even that a single example illustrates more than one value. So here you should say which examples are relevant to the current value.



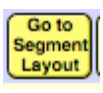
The Value Layouts also show the **Value Comments** field that is also included in the Overview Layout (but with more space), and the text and translation fields for each of the examples (to facilitate filling the example relevance field). In the lower left corner there is also an overview of the other values.

So apart from the Value Annotation field and the example relevance field, these layouts do not contain additional information, but just display the information from the Overview Layout and the Example Layout in a different way.

12. Data on phonological segments



In addition to the 153 primary features (which include some phonological features, but no segmental features), we ask you to provide information on the phonological segments (or phonemes) of your language. Since the phoneme inventory of a language is fairly limited, we thought it was not asking too much to request a complete list of all phonemes. This list should be entered into the special layout "Segments", which you can access from the yellow button in the lower-right corner of the Overview Layout, or from the layout drop-down menu.

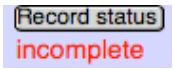


The Segments table contains a list of 96 segments, described verbally in the "Segment name" field and by its symbol in the "Segment symbol" field. We ask you, for each of these segments, to fill in the field "Presence in the language". Basically we want to know whether

the segment exists or not, but you can also say that it exists only in loanwords, or only as a minor allophone.

There is also a comments field, and two optional example fields.

13. Record completed?



In the Overview Layout, there is a field "Record Status" that shows you whether you have completed filling in the records, i.e. whether all the obligatory fields contain some information. When some of the obligatory fields are still unfilled, this field says "incomplete", while you get a "complete (thank you)" report once every obligatory field has some content. (Of course, you can still add more information to "complete" records, e.g. add more examples or fill the optional fields.)