TULQuest best practices guide for Questionnaire development, adaptation and sharing

Aimée Lahaussois, Sylvain Loiseau, Claire Moyse-Faurie, Stéphane Robert, Alice Vittrant, Sylvie Voisin, Marine Vuillermet

TULQuest best practices guide for
Questionnaire development, adaptation and sharing
3 August 2020 version
TULQuest work group

Please cite as:
Lahaussois, A., Loiseau, S., Moyse-Faurie, C., Robert, S., Vittrant, A., Voisin, S., Vuillermet, M. (TULQuest work group), 3 August 2020 version, *TULQuest best practices guide for Questionnaire development, adaptation and sharing*, http://tulquest.huma-num.fr/en/node/164.

1. Introduction

In order to promote the development and sharing of elicitation tools for linguistic analysis, the TULQuest project (http://tulquest.huma-num.fr/en) has developed an interactive space for linguistic questionnaires and elicitation stimuli. Linguists are encouraged to use this website by downloading available materials, and to contribute by uploading new materials or by providing feedback on existing tools.

In the same spirit of making materials freely available and building on existing methods and tools to create new ones, a group of TULQuest project members has developed the present guide to best practices. The list of suggestions presented herein comes out of a review of a hundred Questionnaires by the TULQuest work group in October 2018, in Aussois, France. Our comments are based on a specific corpus and therefore are not representative of all different elicitation tools. This working document is presented to help Questionnaire designers/users, but it is of course a work in progress. As such, we welcome feedback, which can be sent to: http://tulquest.huma-num.fr/contact.

We use the word Questionnaire (with a capital Q) in a broad sense, encompassing all types of elicitation materials for linguistic investigation, as per Lahaussois 2019.

There are a number of relevant dichotomies for Questionnaires. In many cases, there is overlap, but the main lines of opposition fall along the following distinctions: type of data generated, on the one hand; audience targeted, on the other. With regards to the latter, note that a distinction must of course be made between the audience the questionnaire is used by (typically the researcher) and the audience used on (typically the consultant); in some cases, these are one and the same.

The following list suggests some useful opposing pairs to consider when designing/adapting/sharing a Questionnaire:

- turnkey Questionnaires (word lists, video, stimuli) *vs.* investigation guides (incl. checklists, grammar-writing templates, theoretically-oriented questionnaires)

- analysis- *vs.* data-generating Questionnaires: the former will be theoretically-oriented, with questions aimed at collecting abstract, metalinguistic answers (which can be supported by data), whereas the latter will primarily generate linguistic forms from the language of investigation

- comprehensive *vs.* restrictive investigation: with the former, the investigator aims to collect an entire domain (*i.e.* all possible constructions used for it); with the latter, the investigator may use the Questionnaire to check whether a particular construction exists

- primary *vs.* secondary data: does the Questionnaire collect primary linguistic data, or rather an analysis of a given concept/domain through responses to analytical questions?

This can also be expressed as heuristic vs metalinguistic data. Note that many Questionnaires will in fact elicit both data types at once: this is the case for analytical Questionnaires, that ask questions of linguist investigators about the expressions of functions in a given language and request that supporting primary data be provided.

- semasiological *vs*. onomasiological data: are the data collected form-based (semasiological) or concept-based (onomasiological)?

- linguist- *vs*. speaker-oriented Questionnaires: is the Questionnaire designed to be used by a linguist as investigator, or by a (non-linguist) speaker?

- aggregated *vs*. descriptive data: the former might be data that are compiled into a comparative database or project, whereas the latter might be from a single (or limited number of) language.

2. General guidelines for all types of Questionnaires

This section presents some general guidelines for Questionnaire designers which apply to all types of tools.

2.1 Form/formatting recommendations

This section concerns questions of formatting which can facilitate the use of your Questionnaire by others.

2.1.1 Stand-alone capacity

Make sure the Questionnaire has stand-alone capacity.

To avoid losing important metadata and other relevant information when Questionnaires are downloaded/printed for use in the field, ensure that this information is integrated (e.g. via a header) into the Questionnaire document.
The following metadata should be included:
-Name of author
-Name of questionnaire (full name and abbreviation)
-Date/version number
For recommendations on citation form, see 2.2.1.

Include a section where you provide a summary of the protocol, goals, context for the research. Name the document explicitly so it can easily be found when downloaded or shared.

If the Questionnaire was originally published within the framework of a volume, and refers to definitions provided therein, make sure these are available (as well as relevant references) within the document if a stand-alone version is developed and disseminated.

2.1.2 Format for materials

Provide Questionnaires in Word format whenever possible, as PDFs cannot be used to enter data and are thus less convenient.

If possible, consider providing an Excel spreadsheet of examples to be collected, each assigned a numerical ID, so they can be easily copied and sorted.

2.1.3 Presentation

Style

Ensure that the document is well-formatted and styled, page-numbered, without typos or numbering inconsistencies, with up-to-date URLs.

Additional material

Make it clear when supplementary material is available (such as a distinct booklet with lengthier instructions, a website, a spreadsheet) and where to find it.

Contextual material

Make sure to include an introduction, a protocol, goals.

Consider adding a summary, at the top of the Questionnaire, in the form of table with information about the protocol, context, history, even when all the information is present below. Interesting and relevant material can sometimes get lost in the text of longer Questionnaires.

Examples

Examples are very helpful in clarifying notions: include them if at all possible. Additionally, they will serve to illustrate the format that should be used to annotate collected data.

If examples are provided in a language other than the Questionnaire metalanguage, make sure they are glossed using a standard interlinearization format.

2.2 Metadata

This section provides types of metadata which could be mentioned in the Questionnaire to ensure their collection.

2.2.1 Questionnaire metadata

In addition to the most essential metadata on the Questionnaire, which should be integrated into the file, as per 2.1.1 above, we recommend adding these additional metadata, when relevant:

- Complete information on authorship of Questionnaire (individual, team, affiliation...)
- Contact information for Questionnaire author
- Citation form for Questionnaire, so that users know how to reference it (the citation form should refer to a specific version of the Questionnaire)

2.2.2 Metadata on consultants

It can be useful to remind Questionnaire users to collect metadata on the consultants and to include it in their responses.

Classic metadata may include the following:

Name
Age
Gender
Language/dialect
Other languages spoken and levels thereof
Educational level
Place of residence

Other relevant data may include the date and location where the investigation was carried out, the name of investigator, etc.

2.3 Design recommendations

This section presents a few recommendations for Questionnaire design, essentially laying out sections that should be included to facilitate use.

2.3.1 Questionnaire introduction

Questionnaires should be accompanied by an introduction situating them in terms of overall goal, type of Questionnaire (database-building, ...; see Introduction for suggestions), theoretical framework, type of audience it is designed for (laypeople, trained linguists, etc.).

We also recommend adding to the top of the Questionnaire document a short summary of important information, before the introduction. This will allow potential users to quickly evaluate whether or not the Questionnaire is appropriate for their purposes and means.

The summary should minimally include the following:

- Topic of investigation
- Intended audience for Questionnaire (who will use it and who it will be used on: students, native speakers, linguists, ...)
- Data type that can be collected (lexical, phrasal, paradigmatic, ...)
- Type of task(s) involved
- Estimate for time to run the Questionnaire
- Material required (computer, objects/tools, etc)
- Prerequisites for using the Questionnaire (theoretical, level of knowledge of language, metalanguage, number of consultants/linguists required to use the Questionnaire, ...)
- Areality: whether directed at one language or a group, or conceived of as being more widely applicable

The introduction should make clear the intended goals (in terms of data/constructions) and the relevant parameters/variables which need to be distinguished for proper data collection on the topic at hand. (In some cases, the parameters of the investigation are not defined clearly enough, making it difficult to transpose to a different language context from that originally intended).

Add information about the development context and history for the Questionnaire (this will be useful information for future linguists using the Questionnaire--and difficult to recover if you do not provide it).

Be clear about the type of data being elicited by the Questionnaire (e.g. metalinguistic vs. linguistic data).

Describe the structure of the Questionnaire in the introduction: this will situate investigators and allow them to envision how using the Questionnaire will unfold.

If references are cited in the Questionnaire, be sure to include a full bibliography. Ideally, sources could also be commented, so it is clear what their relevance is to the investigation at hand. This results in a literature review, in effect, and will be helpful to linguists using your Questionnaire for their own investigation.

If you hope for feedback, be sure to provide long-term contact information.

2.3.2 Protocol

A protocol, *i.e.* a description of all the methodological steps of the investigation, for using the Questionnaire is very important: it helps ensure comparability of data collected with the Questionnaire, even across a relatively long time span.

If possible, assemble a protocol that does not require specific theoretical knowledge.

Make the protocol as explicit as possible, making no assumptions about a user's ability to intuitively understand what is to be done. (Even translation Questionnaires can be unclear if what is expected is not clearly laid out.)

2.3.3 Data format

Explain how to annotate and to organize the collected data (with glossed examples of how to do this.)

2.3.4 Identification of tasks/audiences

If various types of tasks or various goals are combined into a single Questionnaire, provide a clear system for the investigator to identify which parts relate to what: consider labelling sections and questions with a special code (alphabetical, numerical…) or using different typography for different levels or audiences, such as different fonts, colors, styles.

If the Questionnaire is to be used to collect data for a large database, and you are asking for the data to be returned to you, think of ways to formulate the Questionnaire so the investigator also gets something concrete out of it: an introduction that gives a very concise overview of a domain, a protocol that may lead to further ideas for investigation, etc. In some cases, the formulation of this type of Questionnaire can make the investigator feel used, and this is of course best avoided. Remember to inform the investigator how s/he will be cited for her/his contribution.

Provide suggestions/advice about data collection techniques that go with different phases of protocol: recording, filming, taking notes. Do not assume that the investigator will have the same recording practices as yours and consider that this may affect the usability of the data.

2.3.5 Alternatives/further investigation

Consider providing proxies for situations when the main object of investigation (lexeme, marker, construction, domain…) is not found. This can be done by encouraging the investigator to go beyond the examples in the Questionnaire.

List additional domains to explore in conjunction with the theme of investigation; this amounts to providing suggestions about how to develop the Questionnaire into further stages of investigation.

3. Special guidelines when designing visual Questionnaires

This section provides recommendations for designers of visual Questionnaires.

3.1 Clear instructions

Make sure to include a protocol explaining how to use the stimulus. This is particularly important with Questionnaires that are primarily visual, as users may assume they understand what to do, while they in fact need guidance in order to make the best use of the Questionnaire.

3.2 Naturality of images

In some cases, it can be difficult to identify images, especially with no background (which provides context for understanding). Take this into account when including images in your Questionnaire, and consider that even though it may be appealing to isolate the image from a background in order to put focus on it, it may be counterproductive.

3.3 Acculturation of images

Consider that certain scenes will be culture-specific. If this is the case, it may be helpful to indicate in the protocol that the stimulus was intended as a deliberately culturally-marked stimulus. You can suggest that users working with different cultures either adapt the stimulus by appropriating the protocol to the region they work in (by creating new images), or use the cultural markedness as a basis for collecting interesting narratives. (As an example, a Questionnaire featuring images of 'strange' animals can generate very interesting material.).

If possible, test visual Questionnaires with different speakers before distributing.

4. Special guidelines when designing written Questionnaires

For Questionnaires designed to build databases, it is important to collect information about the absence of a category and alternative ways it can be expressed.

For word lists, it is useful to insert hierarchical numbering by section: e.g. 1. body parts, 1.1 head, 1.1.1 head-hair. This will make it easier to link the original Questionnaire with adaptations, if it is revised for a different linguistic region.

For translation/elicitation tasks, consider the fact that organizing the stimulus sentences according to features of English (or any other metalanguage) may introduce a bias into the data that is collected. One solution is to have the sentences in a spreadsheet, so they can be sorted and rearranged easily.

If possible, provide contextual information for the interpretation of sentences to be translated/elicited.

Make it clear in the protocol that sentences are not to be translated per se (to avoid calques), but that the intent is to produce a natural sentence corresponding to the desired situation. To increase the naturalness of the produced data, one possibility is to have speakers act out the scenes and have another speaker describe the action.

For analytical Questionnaires/checklists, it can be helpful to illustrate possible constructions or structures, as this will make the questions less abstract, and might trigger new paths of investigation for the user.

5. Special guidelines for Questionnaire users

Make sure you read the entire protocol: it may seem clear what you are meant to do, but you may miss/omit something significant for the experiment if you forgot reading the instructions.

Make sure you cite the Questionnaire if you publish data collected using it. If you have adapted an existing Questionnaire, it should be cited as such.

If there is a cultural bias in the material of the Questionnaire, consider taking advantage of this feature when you use it in a different setting: e.g. record the discussions as speakers from different cultures attempt to interpret the images. This could result in precious naturalistic data.

If you use a Questionnaire which is strongly marked for a certain linguistic area, and you adapt it to another setting, consider sharing the result of your work; the new lists will most likely be useful to other researchers. Remember to indicate the reference to the original Questionnaire in your new version.

When adapting a word list, consider maintaining or keeping traces of the original's numbering scheme (even if you decide to add a new numbering system). This will allow for the possibility of comparing the collected data across regions.

References

Lahaussois, Aimée. 2019. The TulQuest linguistic questionnaire archive. In Lahaussois, Aimée & Marine Vuillermet (eds), *Methodological Tools for Linguistic Description and Typology*, Language Documentation & Conservation Special Publication No. 16. Honolulu: University of Hawai'i Press. http://hdl.handle.net/10125/24856