The TULQuest
linguistic questionnaire archive

Aimée Lahaussois

Histoire des Théories Linguistiques UMR 7597
(CNRS – Université Paris Diderot)

This article describes the development and structure of an online interactive archive for linguistic questionnaires developed by the Fédération de Typologie et Universaux Linguistiques (CNRS) program on Questionnaires. The archive allows users to both retrieve and deposit material, with questionnaires categorized according to a taxonomy of features. Questionnaires, defined by our project as any methodological tool designed to collect linguistic data, and written with a capital to highlight this special use of the term, are accompanied by additional materials beyond basic metadata, ranging from a summary of usage protocol, development context, reviews and user tips, as well as the possibility of linking together questionnaires that have been adapted from an original, reflecting the dynamic nature of questionnaire use.

Keywords: Questionnaires, archive, taxonomy, tools for linguistic description

1. Context for archive

The Fédération de Typologie et Universaux Linguistiques (TUL) is a CNRS-funded research infrastructure involving 11 CNRS research groups from around France. It sponsors collaborative research projects on typology for 5-year periods. One of the

1 The website that makes up the archive was built by Alexandre Roulois, LLF/CNRS. I take this opportunity to thank him for the intuitive work. I also thank two Masters student interns for their work on the archive: Marie Benzerrak (summer 2017) and Célia Richy (summer 2018).

2 CNRS leadership has unfortunately decided to replace TUL with a different research infrastructure, and TUL will not be renewed after 2018.
projects for 2014-2018 is on Questionnaires (and the project is thus named TULQuest), with the aim to develop an online archive for linguistic Questionnaires. The TULQuest project involved members from a number of CNRS research groups (CRLAO, DDL, HTL, LACITO, LLACAN, LLF, SeDyL), and the archive in its current state is the result of active collaboration.

The basis for the TULQuest project was the recognition of a paradox: on the one hand, Questionnaires are a recognized part of a fieldlinguist’s toolkit – they are ubiquitously referred to in field manuals (Samarin 1967, Bowern 1998, Chelliah and de Reuse 2011, etc) – and developing them appears to be a useful heuristic tool for linguists working through a theoretical or analytical issue; on the other hand, there are few centralized places where linguists can go for Questionnaires. The Max Planck Leipzig Typological Tools for Fieldlinguists, developed by Jeff Good and Peter Cole, has been a popular resource, as has the L&C Field Manuals and Stimulus Materials website of the Language and Cognition department of Max Planck in Nijmegen. Apart from these two websites, Questionnaire projects tend to have their own websites and are thus difficult to find unless one knows about them beforehand (one example is Koptjevskaja-Tamm’s Temperature survey and questionnaire) or to be relegated to uncentralized locations (desk drawers, appendices of articles or books, etc) by researchers who do not consider them to be actual research products. It should be noted that in this regard there is a notable difference between visual stimuli, which have fairly wide distribution, perhaps due to the significant investment required for their development, and writing-based questionnaires.

The centralization of Questionnaires to facilitate access to them was not the sole purpose of the project. The goal was also to build an archive that reflected significant features of Questionnaires, such as, among others, their dynamicity, their design context and their history. Questionnaires can and are of course often used to investigate what they were designed for, but they are also frequently adapted by users, to reflect different linguistic or cultural contexts from that intended for the original questionnaire, theoretical evolutions of the field, and even a different topic of investigation (in cases where a linguist draws inspiration from the questionnaire protocol, methodology, design in order to study something different). They are thus tools that are far from stable, and we wanted the archive to make it possible to mark the connection between revised or adapted Questionnaires and their source Questionnaires, linking them together and documenting the changes they had undergone. Another important aspect of Questionnaires, and of all methodological tools, is that they are contextually anchored: they reflect the state-of-the-art of linguistic knowledge at the time of their development, and

---

3 By Questionnaire (with a capital Q to highlight the special usage of the term) we mean any methodological tool designed to collect linguistic data, thus include both written questionnaires and visual stimuli.
that historical context is crucial to understanding their impact and effectiveness. The archive thus provides a space to associate information of a historical nature to each Questionnaire.

It is our hope that the TULQuest archive will make it possible to centralize and share the vast but sometimes unseen production of Questionnaires by linguists, making them available to other scholars and providing a documentational framework that brings to the fore important features of the Questionnaires.

2. Definition of Questionnaire

One of the first steps in the project was to define the scope of our investigation and attempt to set ourselves a usable definition for what we considered a Questionnaire to be. We decided from the start to be broadly inclusive and to extend our definition to any tool used in the elicitation of linguistic data for a typological or descriptive activity: wordlists, questionnaires (e.g. analytical, translation), stimuli kits, checklists (such as one might find as an appendix to an article describing a phenomenon), even templates for structuring grammars (such as Comrie and Smith Lingua Descriptive series questionnaire, Comrie and Smith 1977).

While this definition is very broad, we feel that from an epistemological point of view, a more inclusive perspective allows us to capture some realities that might be lost if we exclude certain types of tool from consideration. In order to capture the fact that our use of the term is specific to our research group, we capitalize it (as per Haspelmath 2010) to Questionnaire. The term is henceforth used as a cover for the various tools included in the archive.

3. Organization of site

The TULQuest archive, which can be found at http://tulquest.huma-num.fr/en, is a bilingual English/French website. The architecture of the website is by design very simple. The top menu features five rubrics: Home, Presentation (information about project members), Categories (the taxonomy used to access Questionnaires), References (a list of relevant books on field linguistics methodology and grammar writing) and a Contact interface. The website can be used without registration; registration is only necessary for users who wish to add or modify content. These features of the home page can be seen in Figure 1 below.
There are two main actions that one may take using the website: searching for a Questionnaire, or adding (or modifying) content. These are discussed in sections 3.1 and 3.2 respectively.

3.1 Searching for and downloading a Questionnaire

Questionnaires can be searched for by using either a keyword in a Search box or by selecting a category in the proposed taxonomy, which can be found in the Categories rubric. The taxonomy was the subject of lengthy discussions, attempting to take into account various aspects of Questionnaires that might be relevant to their selection for a given project. The basic categories found are the following:

- TUL Questionnaires
- Areal Questionnaires
- Questionnaires by metalanguage
- Questionnaires by linguistic subfield
- Questionnaires by data type
- Questionnaires by medium

*One of the reasons for creating this archive was as a place for TUL projects to share the results of their work in the form of Questionnaires, thus explaining the decision to include TUL as a “category”.*
The TULQuest linguistic questionnaire archive

The category *areal questionnaires* was deemed an important one to the extent that questionnaires developed for a particular linguistic or cultural area are very much marked by that factor, and need to be adapted in order to be successfully used elsewhere. The Swadesh list (Swadesh 1955), which suffers from a clear cultural bias rendering it difficult to use in some linguistic areas (see, for instance, Chelliah and de Reuse 2011: 229), has undergone numerous adaptations (see for example the Matisoff (1978) “Culturally Appropriate Lexicostatistical Model for SouthEast Asia (CALMSEA)” list).

The inclusion of a category *metalanguage* was in recognition of the fact that Questionnaires are developed in languages other than English (or adapted into other languages). While many such Questionnaires have been translated into English or other languages, this category of the taxonomy allows us to recognize other Questionnaire creating “traditions” beyond the anglophone.

The category *linguistic subfield*, with subcategories for different subfields of linguistics, is an obvious one. Nonetheless, sorting Questionnaires into subfields is not always simple, insofar as many of them cover more than a single subfield—hence the inclusion, for example, of ‘morphosyntax’ in our taxonomy—but categorization by subfield, even if it is not fine-grained, is an important and useful categorizational tool.

The category *data type* focuses on the type of the output of the Questionnaire, with the following subcategories: lexical, paradigmatic, narrative, conversational, phrasal/clausal. As with linguistic subfields, many Questionnaires cover more than a single subcategory, and in some cases, interpretations of what type of data the Questionnaire produces vary. We nonetheless feel it of practical interest for users to be able to select Questionnaires in terms of the type of data they were likely to produce.

In the category *Questionnaires by medium*, we arrive at the classic basis for categorization of Questionnaires, namely the crucial feature of their design which is their material form. The subcategorization here deserves discussion: the category is divided into visual stimuli, on the one hand, and writing-based questionnaires, on the other. Even though the decision for the TULQuest archive was to be inclusive of all types of Questionnaires, this subdivision into writing-based questionnaires and visual stimuli is ultimately bound to emerge in any classification scheme. The subcategorization is as follows:

---

5 An example, among many, of a Questionnaire being in many categories at once is Rose’s (2013) Questionnaire on genderlects, which is as much about sociolinguistics as it is about morphosyntax.

6 Note that these two types of tools are considered so different that there are, to my knowledge, no published taxonomies that take into account both types. The existing taxonomies tend to focus on one kind or another: Mosel (2014) proposes a taxonomy of written questionnaires, with categories translational questionnaires, scenario questionnaires, and grammatical structure questionnaires; taxonomies of (visual) stimuli can be quite detailed (see Lüpke 2009, Hellwig 2006, Majid 2014) but tend to refer to non-stimuli with expressions such as ‘traditional elicitation’ (Hellwig 2006: 330), ‘non-linguistic stimuli’ (Majid 2014: 55), or ‘verbal prompts’ (Lüpke 2009: 70).
• Visual stimuli
  • Pictures
    • Picture books
  • Videos
  • Constructed scenes
• Writing-based questionnaires
  • Checklists
  • Analytical
  • Translation-based
  • Word lists
  • Stage directions

Unsurprisingly, visual stimuli will include Questionnaires where the main medium for eliciting material is primarily visual: these consist of individual pictures (drawings, photographs) which are either discrete (such as Dotte’s (2012) photographs of areally relevant items for the elicitation of possessive classifiers) or can be arranged to form part of a larger sequence (see San Roque et al. 2012), as well as picture books, which can be designed to elicit specific grammatical categories (such as the Hunting Story by Vuillermet and Desnoyers (2013), originally designed to elicit Associated Motion) but also be used to collect “general” narrative data (such as Carroll et al.’s (2011) Jackal and Crow stimulus). Another medium is video, of which there are many examples and which can be used creatively to elicit both materials of a fundamentally dynamic nature (see the Trajectoire stimulus set by Ishibashi & al. (2006) designed to capture descriptions of Path of motion realized by various Figures in different Ground types; the protocol and materials are described in the article by Kopecka and Vuillermet in this issue) and constructions that are not necessarily dynamic (see the article by Lovick and Tuttle about video montages of taboo scenarios to elicit prohibitives and related in Yukon languages). Note that while visual stimuli are often the domain of psycholinguists—see the productivity of the scholars at and from the MPG Nijmegen—this is not an absolute.

More has been written on the categorization of written questionnaires than of visual stimuli, a fact which is unsurprising considering the much longer history of written questionnaires. The oldest written questionnaire in TULQuest is from 1880: Powell’s elicitation schedules for American Indian languages, which are essentially wordlists. There is a period of intense questionnaire development around the EUROTyp (“Typology of the languages of Europe”) project, funded by the European Science Foundation from 1990-1994 and directed by E. König (see the General Preface to Dahl (2002) for details), which led to a vast number of questionnaires on various topics. These were invariably written questionnaires, but of different types, with Dahl’s “translation-based” questionnaire
appearing in this period. Questionnaires at this time were generally divided into analytical questionnaires—lists of questions about a topic, generally addressed to a trained linguist—and translation-based questionnaires—which could be used by anyone familiar with the metalanguage. As far as our taxonomy is concerned, we have kept the terminology: “word-list”, “analytical questionnaire” and “translation-based questionnaire” are all transparent. To these we have added checklists and stage directions. Checklists are any apparatus which helps ensure that a language documenter can get as complete a picture of the phenomenon under description as possible (within the context of a given time frame): as such, they appear to be an important part of the field linguist’s toolkit. An example of such a tool might be Comrie and Smith’s questionnaire: as a grammaticographical template, it serves as a checklist, helping ensure that no topic is omitted (of course, the problem is then that if a topic in the list does not exist in a language, it ends up listed in the resulting grammar as absent\textsuperscript{7}), even though this was not the intent of the authors, for whom the generation of a comparable table of contents across grammars was the main goal. A different example of a checklist is Jacques (2016) on relative clauses\textsuperscript{8}: this Questionnaire is derived from an appendix to an article on relativization in a particular language, provided as an aid to others writing on the same topic. It sums up general literature, suggests lines of research and questioning that can be of use to others—it essentially recycles the knowledge attained in exploring a category in one language to the description of a comparable category in another. Stage directions, finally, are meant to guide a narrative performance with the goal of ultimately being able to produce roughly comparable material across different languages.

The taxonomy does not take into account every existing type of Questionnaire: among visual stimuli, the very wide range of protocols means that they cannot be categorized simply according to the scheme above, which is unable to account for the creative elements and combinations of tasks that may make up a Questionnaire (see for example the Getting the Story Straight protocol, San Roque et al. 2012); this is also the case with writing-based Questionnaires, such as François’s “conversational questionnaire” (see this issue) which combines word lists, stage directions, translation of set dialogues, and thus crosses over types.

This raises the issue of whether a taxonomy for elicitation tools needs to account for every type in existence, or whether we must consider that the creativity and changing needs (and technologies) of linguists developing these tools make them inherently uncategorizable beyond the relatively basic taxonomy discussed above. As far as the archive is concerned, any number of categories can be selected when entering a new Questionnaire,

\textsuperscript{7} See, as an example among many, Section 2.1.3.2.1.4 Future in the grammar of Rapanui: “There is no exclusively future form.” (Du Feu 2010:158).

\textsuperscript{8} \url{http://tulquest.huma-num.fr/fr/node/28}
giving the Questionnaire developer some flexibility when archiving their tool. Entering new Questionnaires into the database is discussed in 3.2.1 below.

3.2 Adding content

The second action one can take with the TULQuest archive is to add content to the site. Selecting "Add content" on the site, after registering as a user, leads to four choices: Questionnaire; Questionnaire: history; Review; Revision. Each of these types of content will be addressed in turn.

3.2.1 Adding a Questionnaire

A new Questionnaire is entered using an online form, through which one enters metadata for the Questionnaire as well as the tool itself. The relevant categories of the taxonomy must also be selected from a menu, with any number of choices allowed. The person inputting the Questionnaire also enters short texts summing up what the goals the Questionnaire is meant to achieve, a summary of the usage protocol, and the development context.

The Questionnaire can be uploaded as an attached file. There are also fields available for a URL and bibliographical references, making it possible to include information about Questionnaires even when intellectual property rights forbid uploading the material directly onto TULQuest.

A final text field allows the deposito to enter preferred citation format for the Questionnaire, particularly useful in the case of as-yet unpublished Questionnaires as it makes it possible to cite them.

The result of the online form is a page containing metadata on the Questionnaire, in addition to information about its goals, protocol, development context (which may all be expanded upon in the actual document), a preferred citation form, and files for the Questionnaire. A URL may also be listed, if relevant. Note the field for User comments at the very bottom of the page. The interface is exemplified in Figure 2.

---

9 http://tulquest.huma-num.fr/en/node/28#overlay=en/node/add/questionnaire
10 Currently accepted formats are .txt, .pdf, .doc, .docx.
A hunting story: eliciting associated motion through guided narrative

Auteur : Vuillermet, Marine & Desnoyers, Antoine

Date de publication : 2013

Réferences bibliographiques :
Vuillermet, Marine & Antoine Desnoyers. 2013. A hunting story — Yendo a cazar: A visual stimulus for eliciting constructions that associate motion with other events. Linguistic Department, UC Berkeley, ms.

Objets
Pictures and video stimuli are useful to investigate specific semantic domains and/or grammatical categories, especially for comparative work. The AM/Associated Motion; see Development context below for details) storybook was specifically conceived to elicit AM morphemes, i.e. morphemes which encode a back-grounded motion to a main action. This is the reason why all the pictures illustrate main actions that occur within a journey – e.g. checking the rifle before leaving, waving to the family while leaving, greeting friends while walking, etc.

Analyzing the data collected with the AM stimulus should:
(a) result in a better understanding of the semantic features of each AM morpheme and of the system as a whole (inventory)
(b) permit to investigate the discourse use of the morphemes;
(c) allow comparison across consultants – according to their age, gender and level of fluency;
(d) help cross/linguistic comparisons, also in terms of inventories and discourse use.” (Vuillermet & Desnoyers, 2013 : 1)

METHODOLOGICAL TOOLS FOR LINGUISTIC DESCRIPTION AND TYPOLOGY

Figure 2. Screenshot of a Questionnaire file.
3.2.2 Adding a Questionnaire history

Because of the importance of the historical context in which Questionnaires are produced (see for example Dollinger 2015), we have provided the means for an additional file, in the form of a full-scale, authored article on the history of the Questionnaire in question, to be added to any Questionnaire. Authors of such articles are encouraged to discuss both the historical context for the development of the Questionnaire and its impact on the field.

These articles can be written by scholars who were not involved in the Questionnaire’s development although ideally, for Questionnaires developed in the last few years, it would be particularly useful to benefit from the insights of the designers themselves on what they were trying to achieve, how their design was influenced by such goals, and on the quality of the data collected and the impact of the tool. In many cases, it is unreasonable to hope that complete data of this type will be collected from the authors of Questionnaires, and it is thus our hope that the short texts entered along with Questionnaire metadata, describing the goals, usage protocol, and development context, will provide some clues for historians of linguistics who may some day wish to study a particular Questionnaire in the archive.

3.2.3 Adding a Review

Another type of file that can be associated with a Questionnaire file is a Review, in other words a critical analysis of the Questionnaire. In some cases, this Review section is used to associate material that appeared independently of the archive on a particular Questionnaire, such as a review of the Questionnaire and associated work in the linguistics literature for example. In other cases, the review can be written directly into the appropriate file in the archive. In both cases, the review file provides a theoretical analysis of the Questionnaire, evaluating its success as a tool for collecting data from speakers or from linguists. In this sense, the review is related to User comments (see §3.2.5), but takes the form of a more analytical document.

3.2.4 Adding a Revision

One aspect of Questionnaires which we have attempted to address through the archive is their dynamicity: rather than being the stable tools they may appear to be at surface-level, in actual usage they are often re-adapted to the particular requirements of the linguist using them. In order to highlight this dynamic quality, the archive makes it possible to connect Questionnaires that are adapted from an original back to their source. This can be done in one of two ways, depending on the intellectual distance the reviser considers there to be between the adapted version and the original:
a) If the linguist creating an adaptation of an original Questionnaire feels that the new Questionnaire represents something independent enough to deserve its own unique file in the archive, it can be entered as a new Questionnaire, using the metadata file discussed in §3.2.1. In this case, instead of appearing within the file for the original Questionnaire, it will have its own independent file – with metadata, categories, goals, development context, and protocol – but it can be linked to the original Questionnaire that inspired it by entering the original’s ID11 in the metadata formula. When this option is selected, an additional line of text will appear in the Questionnaire file in the archive, alongside the metadata, with a link to the original Questionnaire’s file within TULQuest, thus ensuring that the connection between the two Questionnaires can be traced. An illustration of this is seen in Figure 3.

Figure 3. Screenshot of a Questionnaire file with active link to original Questionnaire

b) If the distance between the two Questionnaires is considered to be closer, then one can opt to set up the revised Questionnaire as an adaptation of an original by using the Revision file. The Revision file will allow the inputter to assign a new title to the adaptation, to select the original via a drop-down menu listing all Questionnaires in the archive, to complete a text field listing the changes that were made, and to attach the adapted document, along with a revision date. The Revision to the Questionnaire:

\[\text{In the text box labeled "Original questionnaire", enter the number at the end of the URL for the Questionnaire; e.g. the number '39', found at the end of URL http://tulquest.humanum.fr/en/node/39 for Chevrier's Questionnaire on lexical elicitation for Costa Rica.}\]
The TULQuest linguistic questionnaire archive

The questionnaire will not have its own independent file in the archive, but rather will appear on the file for the original Questionnaire, along with the new title, the attached document, and associated with another document listing the changes that were made and the date.

We believe that this feature of the website, with two different ways of connecting revisions to source Questionnaires, is unique in making it possible to reflect the dynamic nature of Questionnaires in actual use.

3.2.5 Adding a User comment

For reviews of the Questionnaire of a more casual nature than those discussed in §3.2.3, we have included a user comment section directly on the Questionnaire file. The author of these comments is identified through their registration on the website (obligatory in order to be able to use this feature), and this section is intended as a space for sharing usage or adaptation tips by users who have tested the Questionnaire in real conditions. It can also be used to document any comments on the Questionnaire.

4. Conclusion

The TULQuest archive is a work in progress, this progress being entirely dependent on cooperation from linguists far and afield to enter material,12 and as new Questionnaires are entered, the archive is made to evolve to accommodate whatever specific and particular needs arise with the new material. The archive too is thus, like Questionnaires, somewhat more dynamic than static, and the description above is bound to change with future modifications to the input format and Questionnaire display.

The main things to be noted are that we have attempted to design and implement an architecture reflecting the basic dynamic nature of Questionnaires, allowing adaptations of original Questionnaires to be connected in order to trace their evolution. The taxonomy we have devised for the categorization of Questionnaires contains some expected types and others which are less so, making it possible to capture the main characteristics of all the Questionnaires we have entered into the archive thus far, from the type of information they generate, to the medium they use to do so.

Another innovative feature of the TULQuest archive is the possibility of associated peripheral materials with Questionnaires, surrounding them by critical reviews and information about the historical context for their development, and thus providing a richer, 12 Note that there is also a contact form on the site which can be used to provide feedback.
more complete picture of these tools that for so long were considered off-shoots of linguistics research without any serious scientific value. We hope that researchers developing Questionnaires are now able to cite them and to get feedback on their tools, and that this collection will also make it possible to carry out serious epistemological and historical studies of Questionnaires in the future.

Acknowledgements

I am most grateful to the members of the TULQuest project team (listed on the TUL-Quest site) for interesting discussions and brain-storming sessions during official meetings from 2014-2018, and for more informal conversations ever since. I also wish to thank Stéphane Robert, former director of TUL, and Isabelle Bril, the most recent director of TUL, for their stewardship of an infrastructure that made it possible to collaborate fruitfully with so many researchers across France.

References

Carroll, Alice & Kelly, Barbara & Gawne, Lauren. 2011. *The jackal and crow picture task. Designed for use by the Social Cognition and Language Project*. A collaboration of The Australian National University, Griffith University, University of Melbourne and the Max Planck Institute for Psycholinguistics.


