Question template for interviews

This interview template creates a framework for the interviews. The template should not be considered too restrictive. If an interview reveals information not covered by this template, the interviewer should add the questions he/she considers necessary and report about it in the answer sheet (in the cell called *Extra comments and notes*). When possible the below questions should be asked as "open questions" to ensure harvesting of as much information as possible.

The questions are divided into two groups; one group concerns the context of the organization and the interviewee and the setting they operate in – the other group concerns the users' expected requirements to the different elements of the LetsMT! platform.

All questions are not relevant for all types of interviewees. Some of the questions are quite technical and some require a deep insight into the organization's overall tasks and administrative affairs. The interviewer must select the relevant questions for each interview.

Background information about interviewee and his/her organization

1. Description of the interviewee's organization
    a. Localization Service Provider
    b. Translation agency (private or public)
    c. Organization with multilingual translation needs (private or public)
    d. Research institution (private or public)
    e. News/information agency
    f. Other (private or public)

2. Description of interviewee's job profile
    a. Administrator
    b. Translator
    c. Business user
    d. Communications professional
    e. Academic user (student, teacher, researcher)
    f. Non-professional translator
    g. Translation/communication manager
    h. Other

3. Specification of the interviewee's job tasks. Does the interviewee's job include one or more of the following tasks (the list is for inspiration)
    a. Technical tasks, such as upload and download of files and programmes, file conversion, file concatenation, installation of new programmes
    b. Use of CAT tools
    c. Use of online translation services such as Google translate
    d. Use of linguistic corpora
    e. Communication/translation management

4. If CAT tools are employed in the organization it should also be specified
   a. What translation memory system(s)
   b. What machine translation system(s)
   c. What terminology tool(s)
   d. Do the CAT tools have a web service API enabling interaction with an MT system
   e. Does the organization use online language resources
   f. Other relevant tools

5. The organization's experiences with CAT tools
   a. What are the main advantages/disadvantages/problems in connection with the use of CAT tools – as you see it
   b. Does your organization experience technical problems with the tools (what kinds)

6. Specification of software available in the organization (besides CAT tools mentioned above)
   a. What kinds of software are used (which are directly or indirectly part of the translation pipeline)
   b. What browsers are used in the organization, versions
   c. If Microsoft internet Explorer and Firefox are not typically used; does the staff have access to one of these browsers

7. Description of the organization's translation tasks
   a. What kind of translation domains (such as chemistry, legislation, finance)
   b. What are the file formats of translation texts
   c. How many words are translated per day/week/month/year in your organization (or another measurement which makes sense for the interviewee and gives us an idea of the translation volume)
   d. What language pairs does your organization work with
   e. If TM and/or MT is used in the organization: How large is the percentage translated by TM, by MT and by humans alone
   f. What language pairs is MT used for

8. Description of the organization's text resources
   a. Size of existing text resources (monolingual, parallel text resources)
   b. To what extent are existing text resources revised
      (this can be expressed in frequency, e.g. 1) text resources are often revised, 2) text resources are regularly revised or 3) text resources are rarely revised. The revision extent can also be expressed as some sort of quantity, e.g. in no. of words/sentences/man hours per week. The interviewer must select the type of measurement which makes sense for the interviewee and us)
   c. What are the language (pairs) of text resources
   d. Does the organization have parallel text resources not stored in a TM system; how much text is not stored in a TM system; is this text aligned

Let's MT!

e. What kind of segmentation in TM (typically): sentence level, more sentences, paragraph level etc.

f. What information types are used to structure texts (in the TM system or elsewhere). Examples are domain, customer, product line, production date

g. What are the file formats of stored text resources (corpora), e.g. TMX

9. Description of localization/translation workflow

a. Text receipt (who receives it, what is done to the text: e.g. file conversion, insertion/removal of tags, upload to a tool)

b. Does the organization use data management tools, e.g. to detect corrupted data, what kind(s) of tool(s)

c. Localization/translation task (one/several translators, type of translation management system etc.)

d. How is data administration carried out (storage of files and terminology, distribution of translation tasks etc.): who (one person, selected persons, everyone), manually administrated or administrated by translation management system

e. Validation (manual proofreading, automatic data validation)

10. IPR (intellectual property right) of data resources

a. Who has the IPR to the parallel text resources (usually stored in TM system) used in the organization

LetsMT! platform requirements

LetsMT! will provide a platform that supports uploading and sharing of the users' parallel texts. Users select the parallel texts to be used for automated training of an SMT system so that the system will be customized for the specific use.

The LetsMT! system will maintain a high level of security. The system's home page will give anonymous users information about the service and will give access only to public resources and publicly available MT systems. All other functions will be available only to authenticated users with respective access rights. Different groups of authenticated users with different levels of access rights and different rights to access particular resources will be supported.

Whenever users will upload texts or create an SMT system on the LetsMT! platform, they will be asked to specify the level of publicity. There will be several levels of publicity such as "public for any purpose", "accessible only to a certain user group for SMT training", "public for research purposes only, "accessible only for me/the organization" etc.

1. What criteria/functionalities define a good MT system

2. Upload of parallel corpora. What kind of system facilities would you like to have when uploading your organization's data to the LetsMT! platform:

a. Will it be manageable if all files to be uploaded must have a specific format, e.g. TMX

b. Would the organization appreciate if LetsMT! included data management tools to detect corrupted data

c. Would your organization be prepared to share data with other companies

d. Would your organization appreciate if LetsMT! included an on-line alignment tool (for those organizations having parallel texts which are not already aligned)

3. Upload of parallel corpora. How large text resources do you think your organization could provide for the following categories

a. public for any purpose

b. accessible only to a certain user group for SMT training

c. public for research purposes only

d. accessible only for me/the organization

4. Metadata. In connection with selection of parallel corpora for SMT creation: What kind of information types (metadata) would be useful for the organization in order to select the most appropriate corpora:

a. Language pair

b. Source language identifier

c. Domains (is it necessary to distinguish subdomains?)

d. Text type (e.g. sales and marketing material, patents, manuals)

e. Name of data owner

f. Name of data provider

g. Upload date

h. Year of text production

i. Alignment (automatic/manual alignment; alignment on word/sentence/paragraph level)

5. Feedback. Should the LetsMT! system accommodate feedback/rating facilities - and if so in what context(s)?

a. Should it be possible to give feedback/comments to specific text resources - or on a higher level, for example to entire domains, specific data owners etc.

b. Should it be possible for users to rate resources in LetsMT!, e.g. to rate

I. specific data owners, entire text resources etc.

II. only their own data (and not data provided by others)

III. rating of resources is not interesting

6. Configuration. The LetsMT system will provide functionalities for building customized SMT systems. This means that each SMT system is trained automatically on the basis of user selected texts. Would it be helpful if:

a. The user can get information about the text volume required for training of a reliable SMT system

b. The user has access to precise information about the system configurations used in the development of the available SMT systems?

c. The user could specify parameters in order to optimize the training of text data. Such configuration could for example include selection of specific language models and phrase lengths (in other words: will the linguistic/technical staff of your organization be able to utilize such configuration facilities)

7. Website for translation. The system will have a website where users can translate their texts. This website will give access to public SMT systems and to the user's own SMT system trained by him/herself on selected texts

    a. Which domains will your organization typically be interested in

    b. Which language pairs

    c. What are the typical file formats to be translated via this website(e.g. txt, doc, html, pdf)

8. Translation widget. This translation service will be integrated in websites of business and financial news.

    a. Would the organization be interested in a translation widget provided for inclusion into websites to translate website contents

    b. Why or why not

    c. Language pairs

    d. What domains

9. Browser plug-in. LetsMT! will provide a browser plug-in for instant translation of web pages. The plug-in can be downloaded by the user and installed in his/her web browser.

    a. How important is the translation speed

    b. Typical file formats to be translated (html, pdf etc)

    c. Language pairs

10. Integration in CAT tools and other applications

    a. From what kinds of tools would you appreciate direct access to an SMT system

11. (Especially for researchers, students and teachers).
The LetsMT! platform will provide access to text resources within various domains.

    a. How would you wish to work with these text resources

    b. Would you be interested in working with the trained SMT systems for research purposes

    c. Would you like to have access to exactly the same sets of training data that were used to develop the language models stored in the SMT Multi-Modal repository?

Let's MT!