

Linguistic Projects Some Guidelines

A linguistic project (resulting in a linguistic paper or leading to the submission of a graduate project or even when serving in grant writing) should be based on a well defined corpus of speech and/or text data. The data needs to exist independently of the paper or proposal (more about that below). The corpus can consist of recorded and transcribed sessions (dialogs or monologues), recorded and transcribed interviews, or published or unpublished texts in the target language. The material needs to be provided together with source information. All data should be presented in electronic form and exist in an accessible repository. To support a linguistic project primary data needs to be annotated appropriately, in addition translations into English need to be given.

A linguistic project is based on linguistic data and consists of an in depth description of the material or some aspects of the material using standard linguistic terminology. A formal linguistic analysis of some aspects of the language is desirable, yet, there needs to be a clear focus on the analysis of the primary linguistic data presented in an accessible corpus.

What is required in a Linguistics Project?

The data chosen for the linguistics project may be originally spoken or written language as it naturally occurs. The form the data can take could include the following:

- an appropriate transcription of a sample of continuous speech (for example, an interview or a radio program or a phone call);
- a written text or part of a text which may be published or unpublished;

Primary data needs to be sorted and annotated using linguistic methods. The appropriate form of the annotation (speech and/or text) will vary depending on the type of study undertaken. If the formal linguistic analysis concerns phonological variation for instance, it is desirable to work with signal annotation and a fine-grained phonological transcription (phoneme level) of selected words. For a morpho-syntactic analysis morpheme level annotation is adequate. Interlinear glossed samples consist of a set of examples taken from the primary data with a focus on the phenomenon to be treated in the project. The linguistic context of the samples should be made clear and the relevance of the sample for the topic of the project should be discussed.

Notice, to undertake detailed phonetic analysis using signal annotation or to provide adequate morpheme level grammatical annotation requires linguistic training. The annotation of long stretches of speech or text should therefore be avoided in an early stage of your linguistic studies.

Part of the linguistic project is to discuss your choice of primary data and to discuss your method of transcription. A linguistic project should demonstrate awareness of the ethical and legal implications of the linguistic work.

The size of the linguistic data chosen for your project will vary. In-depth annotation is a time consuming task. Therefore it is important to select a set of paradigms from the primary material that will feature as the core material for the project. When creating a project it is important to discuss your primary resources as well as the sample you will be using in detail.

Your core data should be subjected to a phoneme- or morpheme- level annotation and should always

come with a translation. Use an Interlinear Glosser to annotate your text. TypeCraft is one such tool, there are others available. Linguistic tools are to be preferred over annotating glosses in a standard word processor like Microsoft WORD. The relation between the annotated material and the primary material needs to be made clear so that your annotation sample can be traced back to the primary sources. Metadata about the sources needs to be made accessible.

Your description of your project must contain a description of the goal of your project, your methodology and if applicable an outline of the formal linguistic analysis that you intend to perused.

The project text is basically a discussion of the research question that is the focus of your project and an outline of the analysis that you suggest in order to arrive at a more satisfactory representation of the data. The phenomenon chosen must be placed it in the context of studies of the language(s) in question and the linguistic work that has been done on this language.

Factors that determine the success of your proposal

Assessment of the linguistics project will take into account the quality and the presentation of your primary data as well as the quality of your structured data, that is the data selected for an in-depth analysis. (How motivated data selection is in the light of the topic chosen, quality of transcriptions, and quality of the annotation will be taken into consideration) Important is a solid command of linguistic terminology. The description of the proposed linguistic analysis must show an awareness of present days linguistic theories. The quality and how representative the data seems will be taken into consideration. If only secondary data is used it must be clear that a more adequate or original approach to the data will be attempted.

Presentation of linguistics projects must be word-processed. The project should be typed in 1.5 spacing, using A4 paper and a 12pt standard font. Headings appear in bold. A link to your primary data must be provided. Preferably your data should be accessible using one of the standard linguistic tools, such as Praat, Toolbox, FLEx or TypeCraft. You are responsible for providing the readers of the proposal with the information necessary to access your research corpus. However, the project text should be self-contained and include the relevant core data in an annotated format.

The naming of your text sections is not crucial. Important is that your project has a telling title and that you provide a short summary of your project at the beginning of your text.

The following sections may, but must not occur in your project text:

- Introduction
- Background or Goal and Background
- Statement of the problem
- Research objective
- Significance of the study
- Literature Review

- Methodology
- Hypothesis
- Conclusion

Not all sections need to occur; they might not even be relevant in some cases. It is more important to make sure that your discussion is coherent, and that your methodology is appropriate given your research goal. Avoid repetition, the sections *Introduction* and *Background* should not contain the same information. In some cases it will be sufficient to write a *Literature Review*. The sections *Statement of the problem*, *Research objective*, *Methodology* and *Significance of the study* are the central sections of your proposal. This should also be reflected in their length.