KLIP Bericht

Working on an FNS grant proposal to systematize our knowledge on agreement phenomena in individual Italo-Romance non standard varieties into a database, we asked the ZüKL for 2’000 CHF funding in order to hire a database manager.

The funding given allowed us to commission the development of a custom database and software annotation tool that is carefully designed to fit the needs and purposes of the project. The software was developed by Taras Zakharko and is a derivative of the software he has developed for corpus annotation within the scope of his PhD thesis (SNF project # 140881).

In its current form, the preliminary annotation system fits every type of data we aim at recording (grammaticality judgements, translation and connected speech) and allows us to annotate items and relations with respect to part of speech and selected morphological, syntactic and semantic features. This is illustrated in Figure 1, where the analysis of *fi romaftu a rromu* ‘thou. M stayed in Rome’ in the dialect of Ripatransone (central Italy) is displayed.

The database structure is devised to accommodate for the maximum complexity we expect to find in the data. Thus, under the noun ‘Rome’ one finds a specification of both (inherent) gender (the normal case) and contextual gender agreement, which is highly unusual for nouns but is found in Ripano. The architecture of the data model allows us to annotate complex relations between linguistic items. Furthermore, the annotation system is designed in a flexible, modular way, which makes it easy for us to refine the parameters of the annotation. The annotation tool produces standoff annotations, in which each annotation item is linked to the original text via unique identifiers and to the relevant metadata. Moreover, the annotation tool allows the user to quickly navigate between the tokens and create annotations on individual words and phrases and also to specify agreement and government relations between them, as seen in Figure 1. The annotation data is entered in dynamic annotation boxes via a convenient table view. To make the annotation procedure faster and less error-prone, the annotation boxes adjust to the context by displaying only relevant information.
We applied for the grant with the SNF in April 2014. The proposal has been accepted and our project "The Zurich database of agreement in Italo-Romance" will be funded by the SNF (total 666'064 SFr). The reviews were extremely positive: all three of the referees considered the project to be excellent. The project will commence in autumn 2015.