

## **Report Short Time Scientific Mission at ICLTT Vienna**

Collaborators: Karlien Franco, Mag. Eveline Wandl-Vogt

The STSM was carried out with two main goals. On the one hand, we wanted to come up with a way to link the DBÖ and the WLD and WBD to each other. This linking will be beneficial for both the COST Action and for my PhD research. Second, we aimed to discuss ways in which the methodology of my PhD research could be applied to the DBÖ.

More specifically, the following objectives were achieved during the STSM:

### **Objective 1: To exchange expertise about three European dialect dictionaries (The dictionary of Limburgian, dialects (WLD), The dictionary of Brabantian Dialects (WBD), The dictionary and database of Bavarian dialects in Austria (WBÖ, DBÖ)).**

Exchanging this expertise is useful in two ways: it helps us in coming up with a model to link the three dialect dictionaries and it provides us with information that is necessary for Karlien's dialectometrical PhD project.

During the STSM, we thoroughly discussed the structure of the DBÖ, we exchanged data, and we thought of ways in which we could use this structure to our benefit. By doing so, we noted a number of issues that can be important when we compare the Bavarian and Dutch data:

- We are working on a script to transform the Bavarian data to a table format, so it can be compared more easily to the Dutch data (which is in a csv-table format). This way we can use a larger proportion of the data available at the ICLTT.
- The Bavarian questionnaires are organized in broad semantic fields, but per field multiple questions are grouped.<sup>1</sup> This makes it harder to link the dictionaries to each other. The solution we came up with is to first link the dictionaries in an automatic way; afterwards, we will need to map problematic concepts manually onto each other.

### **Objective 2:**

#### **To construct a basic model for comparison of the Bavarian and Dutch data and to apply this to two concepts.**

In collaboration with Thierry Declerck, we devised an extension of the Ontolex scheme to be applied to the Dutch and Bavarian data (see appendix). We aim to apply this scheme to the concepts TARAXACUM, BODY and DEATH as soon as it is finalized.

---

<sup>1</sup> In the Dutch dialect data, different sources are available (questionnaires, local dictionaries ...), but we are currently focusing on a selection of the questionnaires. By doing so, we get a one-to-one mapping between a concept and the questions that were used to obtain lexicalizations for the concept.

### **Objective 3:**

**Preliminary steps towards the application of the methodological part of the PhD-project will be taken. In particular, the exchange will contribute to finding a solid methodology for establishing the particular features of the concepts in the dictionary, that is stable across languages.**

By thoroughly discussing the layout of the DBÖ, we were able to take preliminary steps towards applying Karlien's methodology to the Bavarian data. Furthermore, we detected a number of issues, which we have to keep in mind during the analysis of the data.

- In the Bavarian data, there may not be a lot of observations for each place, so we have to check that and take it into account for the comparison with the Dutch data.
- We have to take into account the time at which the questionnaire was distributed, because it may influence the framework of the concept.
- The location of the respondent of the questionnaire is available in the DBÖ in several degrees of granularity. We would like to focus on one particular place (rather than on a broader region), so we can use this information for our dialectometric analysis.
- For the PhD research, we aim to determine the influence of conceptual features, such as salience, vagueness and negative affect on lexica variation across dialects. A pilot study of this hypothesis has already been carried out for the semantic field 'the human body' in the Limburgish dialect data. In this study, a survey was carried out among native speakers of Dutch to establish the negative connotation and salience of the concepts in the semantic field. However, for the Bavarian data, we need to find a way to operationalize these measures as well. We thought of some possibilities:
  - we could replicate the survey of the pilot study among native speakers of Austrian German
  - we would also like to think about more objective measures: for salience in the semantic field of 'plants', we could, for instance, determine how useful plants are (by determining how poisonous they are, and by taking into account how often they are used for medical purposes)

Finally, we discussed a number of additional things we can do with the data:

- We can do a dialectometric analysis of the Bavarian data and compare the results to the classification of geographical space that is provided in the dictionary

- We can have a look at the influence of associative lists: do concepts that can be conceptually associated to one another behave in the same way when it comes to lexical variation? (cf. Michael Zock)
- There are other data as well for Austria that don't encompass the entire Austrian geographical area, but if we include them into our analysis, we may be able to use them as well.



