QUALD2014: From theory to data and back  
Course syllabus and preparation

The aim of this course is to improve our reasoning with data. You can see it as an exercise in combining everything we know in a solid methodology of discovering new knowledge about language. Please be ready to take an active participation in this course. As with all other exercises, it does not help if you do not do it.

This sheet is a guide for your preparation for the course. It contains an overview of the topics which we will address, the list of related literature, and some tasks for you. Please read it carefully, fill it in, and bring it with you to the course.

1 Deductive vs. inductive prediction in science

Based on your general methodological background and the course readings, describe:

- The role of prediction in science:

  Example:

- The role of data in deductive prediction:

  Example:

- The role of data in inductive prediction:

  Example:

2 The role of data in linguistic theory

Based on your general linguistic background and the course readings, describe the kind of data used in:

- Comparative linguistics

  Example:

- Structuralism
Example:

- Generative grammar

Example:

- Typology

Example:

- Language use

3 Type-level vs. token-level observations

Where do we normally find:

- Type-level data?

- Token-level data?

4 Hypothesis formulation

- The "so-what" problem:

  Write down any thoughts you might have about what this problem is:

Think of a linguistic study. It can be something you know of or something you would like to do.
• What does the study prove?

• Why? ("Interesting" is not good enough.)

• What kind of data are used? How are these data described?

5 Counting: Never ignore what you do not see!

Write down any thoughts you might have about what these problems are:

• The problem of biased samples

• The problem of small samples

6 Statistical models and scientific proofs

Based on your general background and the course readings, describe:

• The two main components of a statistical model

• The difference between variation and a variable

• Probability distribution
Write down any thoughts you might have about the following question:

- What does a statistical model prove and what not?

7 Statistical analysis: In search of the right formula

Most analyses are based on the following two notions. Try to explain them in your own words:

- Correlation

- Regression

8 Interpreting statistical scores

Notes:
9 Example studies

Analyse the following studies:

1. On the dative alternation:

2. On semantic maps:

Try to fill in the following table:

<table>
<thead>
<tr>
<th></th>
<th>Dative alternation</th>
<th>Semantic maps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothes(i/e)s:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Samples:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The kind of data:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable(s):</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dative alternation</td>
<td>Semantic maps</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Formula(s):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalisations:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance of the conclusions:</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Write down any questions:
Readings:

http://elanguage.net/journals/lilt/article/view/2578


http://www.stanford.edu/~bresnan/Roots_05_Bresnan.pdf


Peter Norvig. "On Chomsky and the Two Cultures of Statistical Learning."

Chapter 3, Section 3.3 "Statistical analysis", pp. 101-112. (provided PDF)