After several years of successful bootcamps on corpus linguistics as well as statistics for linguists with R, this year offers participants the possibility to

− either take a **30 contact hours corpus linguistics bootcamp**, in which they learn how to write R scripts to extract and process data from differently-annotated corpora to generate different types of frequency lists, concordance displays, etc.;
− or take a **30 contact hours statistics for linguistics bootcamp**, in which they learn how to use R to perform many different monofactorial significance tests, visualize results, and perform regressions as well as hierarchical cluster analysis;
− or take **both in a row** to learn, first, how to get data out of corpora and then, second, how to evaluate them statistically.

I. **Bootcamp contents**

1. **Bootcamp 'Corpus linguistics with R'**
The corpus bootcamp is a 30-hours hands-on introduction to quantitative corpus linguistics for both graduate students and seasoned researchers. Using the open source software and programming language R, we will learn

− how to generate frequency lists and search for words and patterns;
− how to process corpora and perform corpus-linguistic searches in ways that typical corpus software does not support;
− how to write small functions for recurrent corpus-linguistic tasks.

Data to be dealt with include plain text corpora, corpora with SGML or XML annotation, ICE-GB files, and others. The participants will also get small functions and scripts they can use for their own corpus-linguistic tasks (concordancing, generating n-grams of words or characters, and others).

The content of this corpus linguistics bootcamp is based on Gries (2009e, [http://tinyurl.com/QuantCorpLingWithR](http://tinyurl.com/QuantCorpLingWithR)) but (i) structured differently to accommodate the workshop format of the bootcamp and (ii) provides functions and examples not discussed in it.

2. **Bootcamp 'Statistics for linguistics with R'**
The statistics bootcamp is a 30-hours hands-on introduction to statistical methods for both graduate students and seasoned researchers. Using the open source software and programming language R, we will

− briefly recap basic aspects of statistical evaluation as well as several descriptive statistics;
— discuss monofactorial statistical tests for frequencies, means, dispersions, correlations;
— explore different kinds of multifactorial and multivariate methods, in particular different kinds of regression approaches as well as hierarchical cluster analysis.

For all statistical methods to be explored, we will discuss how to test their assumptions and visualize their results with nice and annotated statistical graphs, and sometimes we will reanalyze published data from corpus-linguistic studies. The participants will also get small functions they can use for their own statistical applications. Also, there will be a small section on how to write small statistical/visualization functions yourself.

The content of this statistics bootcamp is based on Gries (2013c, <http://tinyurl.com/StatForLingWithR>).

II. Schedule

1. Bootcamp 'Corpus linguistics with R'

<table>
<thead>
<tr>
<th>Date</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>11 August 2014</td>
<td>9.00-10.00 arrival/admin</td>
</tr>
<tr>
<td></td>
<td>10.00-1.00 class</td>
</tr>
<tr>
<td></td>
<td>1.45-5.00 class</td>
</tr>
<tr>
<td>12 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>13 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>14 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>15 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>16 August 2014</td>
<td>departure</td>
</tr>
</tbody>
</table>

Class sessions of more than two hours include a 15-minute break.

2. Bootcamp 'Statistics for linguistics with R'

<table>
<thead>
<tr>
<th>Date</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 August 2014</td>
<td>9.00-10.00 arrival/admin</td>
</tr>
<tr>
<td></td>
<td>10.00-1.00 class</td>
</tr>
<tr>
<td></td>
<td>1.45-5.00 class</td>
</tr>
<tr>
<td>19 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>20 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>21 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>22 August 2014</td>
<td>9.00-12.15 class</td>
</tr>
<tr>
<td>23 August 2014</td>
<td>departure</td>
</tr>
</tbody>
</table>

Class sessions of more than two hours include a 15-minute break.
III. Accommodation

We reserved accommodation for up to 20 participants in ten two-bedroom apartments in Santa Ynez Apartments, a housing complex next to the UCSB campus. Each apartment has

- two separate bedrooms;
- a shared living area and a shared kitchen area;
- 1.5 bathrooms;
- bed linens (sheets, a blanket, a pillow plus pillowcase);
- laundry facilities and wireless internet access.

Cf. <http://www.conferences.ucsb.edu/santa-ynez-apartments> and our bootcamp map in the list of links below. The campus and neighboring Isla Vista offer many eateries close by.

IV. Pricing, registration, and payment

Registration is on a first-pay-first-served basis, beginning immediately, with priority given to participants signing up for option 3, i.e., both bootcamps. If you would like to register, send an email to Stefanie (<swulff@ufl.edu>) with "Bootcamps 2014" in the subject line and state which of the three registrations options you would like to book. Provided that there are still open seats, Stefanie will send you a registration form with instructions for payment. The following three registration options are available:

1. Only the corpus bootcamp

US$650. This price includes 5 days of workshop instruction plus materials, a reception dinner, and accommodation in a Santa Ynez apartment (check-in: 10 Aug 2014, check-out: 16 Aug 2014).

2. Only the statistics bootcamp

US$650. This price includes 5 days of workshop instruction plus materials, a reception dinner, and accommodation in a Santa Ynez apartment (check-in: 17 Aug 2014, check-out: 23 Aug 2014).

3. Both bootcamps

US$1360. This price includes 10 days of workshop instruction plus materials, two reception dinners, and accommodation in a Santa Ynez apartment including the weekend between the bootcamps; this is why this price is a bit higher than both bootcamps separately (check-in: 10 Aug 2014, check-out: 23 Aug 2014)
Note: your registration and your payment must be received by 23 April 2014 because that is when we must commit financially to a minimum number of places and accommodation at UCSB (otherwise we have to pay for unused slots out of our own pockets). After that, registration may still be available but probably more restricted.

Note: if you decide to cancel your registration by 9 April 2014, you will be refunded 50% of your payment; after that, we cannot refund your registration fees.

V. Recreation

Workshop participants can take advantage of the UCSB Recreation Center (<http://www.recreation.ucsb.edu/>) for a daily charge of approximately US$12.-.

VI. The instructor (based on Wikipedia)

Stefan Th. Gries (STG) is Full Professor of Linguistics in the Department of Linguistics at the University of California, Santa Barbara (UCSB), Honorary Liebig-Professor of the Justus-Liebig-Universität Giessen, and Visiting Chair (Linguistics and English Language) at the ESRC Centre for Corpus Approaches to Social Science.

Methodologically, STG is a quantitative corpus linguist at the intersection of corpus linguistics, cognitive linguistics, and computational linguistics, who uses a variety of different statistical methods to investigate linguistic topics such as morpho-phonology, syntax, the syntax-lexis interface, semantics, and corpus-linguistic methodology, as well as first and second language acquisition. Occasionally, he also uses experimental methods (acceptability judgments, sentence completion, priming, self-paced reading times, and sorting tasks).

Theoretically, he is a cognitively-oriented linguist (with an interest in Construction Grammar) in the wider sense of seeking explanations in terms of cognitive processes without being a cognitive linguist in the narrower sense of following any one particular cognitive-linguistic theory. The researchers who have influenced his work most are (in alphabetical order) R. Harald Baayen, Douglas Biber, Nick C. Ellis, Adele E. Goldberg, and Michael Tomasello.

STG has authored three books – one research monograph, an introduction to statistics with R for linguists (two editions), and a book on corpus linguistics with R. He has also co-edited six volumes – two on corpora in cognitive linguistics, two on frequency effects in language, one on corpus linguistics, and one on cognitive linguistics. Since 1999, he has (co-)authored more than three dozen articles in the leading peer-reviewed journals of his fields (Cognitive Linguistics and International Journal of Corpus Linguistics) as well as in many other peer-reviewed journals, plus another 50 or so articles in edited volumes, proceedings, etc. He is founding editor-in-chief of the international peer-reviewed journal Corpus Linguistics and Linguistic Theory, associate editor of Cognitive Linguistics, and performs editorial functions for more than ten international peer-reviewed journals and book series. Since the beginning of 2007, he has given more than 150 talks and workshops at national and international venues, more than half of them invited.
VII. Contacts

For further information, please contact

- Stefan Th. Gries (<stgries@linguistics.ucsb.edu>) for all questions having to do with the content of the bootcamps, or
- Stefanie Wulff (<swulff@ufl.edu>) for all questions having to do with organizational matters

by email; please use "Bootcamps 2014" as the subject header in all correspondence!

VIII. Links

Stefan Th. Gries (instructor)   http://tinyurl.com/stgries
Stefanie Wulff (organizer)    http://tinyurl.com/stwulff
Bootcamp overview map:        http://g.co/maps/gp5fe
UC Santa Barbara               http://www.ucsb.edu/
                              http://www.aw.id.ucsb.edu/maps/
Department of Linguistics, UC Santa Barbara  http://www.linguistics.ucsb.edu/
Santa Barbara, CA             http://www.santabarbaraca.com/
                              http://www.santabarbaraca.gov/
Santa Barbara airport         http://www.flysba.com/
                                  http://groups.google.com/group/statforling-with-r/topics
Quantitative Corpus Linguistics with R  
                                  http://tinyurl.com/QuantCorpLingWithR  
                                  http://groups.google.com/group/corpling-with-r/topics
R                                http://www.r-project.org/  
                                  http://cran.r-project.org/index.html
RStudio                          http://www.rstudio.org