

## ***Visualizing forests, not trees: Aggregate varietal comparison***

online course on Thursday (all day) and Friday morning by  
**John Nerbonne**

**Short CV:** John Nerbonne (Ph.D, Ohio State 1984) worked in industry for 8 years before becoming professor of computational linguistics in Groningen (NL). He has worked broadly in applied and theoretical linguistics, always with a computational slant. He began the work on dialectology, for which he is best known, in 1997. He and a very talented group of collaborators extended computational dialectology to a measure of pronunciation distance for phonetic transcription, software for data review, statistical analysis and mapping. The web application Gabmap supports these analyses (<https://gabmap.let.rug.nl/>). He is a member of the Royal Netherlands Academy of Science, and past president of the Association for Computational Linguistics (2002) and of the European Association for Digital Humanities (2014).



### **Course description**

**Remark:** In order to make the course more interesting to specialists in English, and in view of Freiburg's prominence in the study of WORLD ENGLISHES, I will structure the course to enable an examination of this topic.

**Background:** A lot of contemporary work in dialectology attempts to measure differences between varieties at specific levels – lexical, phonetic, or syntactic – based on a sizeable sample of suitable material. These studies show how language variety is geographically distributed, and the characterizations may be compared to other population-level descriptions, e.g., in genetics (Manni & Nerbonne 2021), in socio-economics (Lameli, Falck et al. 2012), or with respect to linguistic structure (Huismann, Franco & van Hout 2021). The long-term goal is to understand how linguistic culture is related to other aspects of culture.

**Goals:** Participants should understand how and why aggregate analyses work, the course of development they've taken, and perhaps where they might be useful in studying English varieties.

**Material:** To do this, the course will develop the main ideas of DIALECTOMETRY—the use of exact techniques in dialectology. Nerbonne (2009) is a good introduction to this research line. The course will be structured not historically, but rather around the main ideas in the techniques, aggregation, distance metrics, the analysis of distance tables, reliability and validation, insights the new techniques have enabled, application to older dialectological questions such as the influence of regions as opposed to continua, the nature of the geographical distribution of variation. Where feasible, examples will be taken from English.

There will be five sessions (as I'd plan now, 7/2023)

1. The perspective from dialectology
  - a. Why aggregate? Advantages of the aggregate perspective
  - b. Simple categorical differences – lexical variation
  - c. How much data is needed?
  - d. Case study in lexical variation – Linguistic Atlas of the Middle and South Atlantic States (LAMSAS; American)
2. How to measure differences to obtain distances, enabling subsequent analysis
  - a. Avoiding simple analyses of matrices
  - b. Categorical analyses of syntax, morphology
3. How to analyze distance tables
  - a. Clustering
  - b. The bootstrap and noisy clustering
  - c. Multidimensional Scaling
  - d. Gabmap tutorial, with exercises
4. Pronunciation analyses (edit distance)
  - a. Exercise in edit distance
  - b. Case study in pronunciation analysis

5. Une perspective mondiale – World Englishes
  - a. Putting world Englishes on the map
  - b. A basement inventor's view of information visualization
    - i. Tufte's measure of quality in informational graphics
    - ii. Mapping aggregate variation

## References

- Falck, O., Heblich, S., Lameli, A., & Südekum, J. (2012). Dialects, cultural identity, and economic exchange. *Journal of urban economics*, 72(2-3), 225-239.
- Huisman, J. L., Franco, K., & van Hout, R. (2021). Linking linguistic and geographic distance in four semantic domains: computational geo-analyses of internal and external factors in a dialect continuum. *Frontiers in Artificial Intelligence*, 4, 668035.
- Manni, F., & Nerbonne, J. (2021). Linguistic Diversity and Human Migrations in Gabon. In: M. de Lourdes Muñoz-Moreno & M. Crawford (eds.) *Human Migration: Biocultural Perspectives*, 99-114.
- Nerbonne, J. (2009). Data-driven dialectology. *Language and linguistics compass*, 3(1), 175-198.