Workshop title:
Quantitative Analysis of Language Variation in Time and Space

Definitive title of your presentation:
Aggregating Syntactic Variation: The Forest behind the Trees

Authors and affiliations:
Marco René Spruit, Meertens Instituut & John Nerbonne, Groningen

Abstract:
Dialectometry on the basis of categorical distinctions was invented by Seguy and has been further developed extensively by Goebl. The primary motivation for dialectometry has been its ability to abstract from the recalcitrant detail of dialect data to general characterizations of entire linguistic varieties and their relations to one another. The successful deployment of dialectometric techniques presupposes a large amount of comparable data, however, which has not been available for syntax until recently. Syntactic data may be collected with respect to categorical decisions, and an early, large collection of such data is the Syntactic Atlas of Netherlandic Dialects (SAND, Barbiers et al. 2005), which contains information on the left and right clause peripheries, negation and the pronominal system of Dutch, including information on word order, rection, and morphosyntactic form as well as the range of variants possible at all of the collection sites.
We subject SAND to a dialectometric analysis in this presentation, examining in particular the degree to which syntactic phenomena are geographically coherent, and the consistency of the measurements we are obtaining. We also examine where geographic conditioning of syntax appears to differ from indications we have for pronunciation and lexical data.

Reference
Barbiers, S., H. Bennis, M. Devos, G. de Vogelaer, M. van der Ham, eds., (2005). *Syntactic Atlas of Netherlandic Dialects*, volume 1. Amsterdam University Press, Amsterdam