## The *Multidimensional Analysis Tagger*, or how I stopped worrying and created a tagger using regular expressions

Andrea Nini, University of Manchester

andrea.nini@manchester.ac.uk

The aims of the talk are to introduce the Multidimensional Analysis Tagger (MAT), a program that replicates Biber’s (1988) analysis, and to demonstrate that regular expressions can also be employed for more sophisticated uses, such as creating a tagger. Although MAT uses the Stanford Tagger to generate an initial tagged version of the input, the main tagging system behind MAT is an application of regular expressions to find and then count Biber’s (1988) linguistic features. In addition to tagging texts, MAT also plots the input on to Biber’s (1988) Dimensions and assigns it a text type as proposed by Biber (1989). By exploring the code of MAT, the talk aims at revealing that the mechanics of non-probabilistic taggers based on recognising specific linguistic features are essentially founded on simple regular expressions and substitutions. Additionally, MAT will be demonstrated and evidence will be presented on its reliability and usefulness to analyse new data in the context of Biber’s (1988) research.

**References**

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1989). A typology of English texts. *Linguistics*, 27(1), 3–43.