Statistical Analysis of Corpus Data with R
Exercise Sheet #3

Multi-dimensional scaling (MDS) is another popular dimensionality reduction technique (see, e.g., Cox and Cox, 2001, *Multidimensional Scaling*, Chapman & Hall).

Perform an MDS analysis of the Italian NN compound data, based on (scaled versions) of the cues described in the course slides.

1. MDS operates on a *distance* matrix, a symmetric matrix of distances between each point in the data-set and each other point. Thus, the first thing you will need to do is to generate a distance matrix from the cue matrix. Look at the documentation for the `dist()` function, and use it to generate distance matrices using two different methods to compute distance.

2. In order to perform MDS, you will use the `cmdscale()` function: take a look at its documentation, and run MDS on each of your distance matrices (if you want to perform some further exploration of MDS, consider also the `sammon()` and `isoMDS()` functions in the venerable `MASS` package).

3. Plot the compounds in the first two dimensions produced by the MDS analyses, using different colours for relational and attributive compounds.

4. Try *k-means* clustering on the MDS outputs, and look at performance by cross-tabulating the clusters and the relational/attributive labels.