Informatics 1 CG – Tutorial 4
Solutions
Mirella Lapata, Carina Silberer
Week 5

In class we discussed the problem of syntactic category acquisition and how syntactic category learning can be modelled by means of agglomerative hierarchical clustering that groups words together on the basis of their distributional information. We also discussed about categorization and how conceptual knowledge is organized. The aim of this tutorial is to (a) understand the agglomerative clustering algorithm and the three different linkage criteria and (b) to examine how concepts are organized in adult speakers.

1 Learning Syntactic Categories: Clustering

You are given a data set of five words characterised by a single feature and the corresponding distance matrix:
1.2 Solution for Complete-link Clustering

\[
\begin{array}{c|cc}
 d & k & K \\
\hline
 0 & 5 & \{\text{neat}, \{\text{empty}\}, \{\text{occupy}\}, \{\text{finish}\}, \{\text{race}\}\} \\
 1 & 3 & \{\text{neat, empty}, \{\text{occupy, finish}\}, \{\text{race}\}\} \\
 2 & 2 & \{\text{neat, empty}, \{\text{occupy, finish, race}\}\} \\
 5 & 1 & \{\text{neat, empty, occupy, finish, race}\} \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 & \text{neat} & \text{occupy} & \text{finish} \\
\hline
\text{occupy} & 3 & - & - \\
\text{finish} & 4 & 1 & - \\
\text{race} & 5 & 2 & 1 \\
\end{array}
\]

1.3 Solution for Average-link Clustering

\[
\begin{array}{c|cc}
 d & k & K \\
\hline
 0 & 5 & \{\text{neat}, \{\text{empty}\}, \{\text{occupy}\}, \{\text{finish}\}, \{\text{race}\}\} \\
 1 & 3 & \{\text{neat, empty}, \{\text{occupy, finish}\}, \{\text{race}\}\} \\
 1.5 & 2 & \{\text{neat, empty}, \{\text{occupy, finish, race}\}\} \\
 3.75 & 1 & \{\text{neat, empty, occupy, finish, race}\} \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 & \text{neat} & \text{occupy} & \text{finish} \\
\hline
\text{occupy, finish} & 3 & - & - \\
\text{race} & 4.5 & 1.5 & - \\
\end{array}
\]

\[
\begin{array}{c|ccc}
 & \text{neat} & \text{occupy} & \text{finish} \\
\hline
\text{occupy} & 2.5 & - & 1 \\
\text{finish} & 3.5 & 1 & - \\
\text{race} & 4.5 & 2 & 1 \\
\end{array}
\]
2 Semantic Categories

This is mostly a discussion-based exercise, ask students to form groups and discuss potential difficulties in grouping the given concepts.

3 Concepts

Again, each person should take a piece of paper and pen and do the free association task for the words ocean, fish, toothbrush. Then discuss questions (a)–(e) in class. Point out that different individuals come up with different solutions, but also that there are some associates that everyone will come up with.