Basics

- **Optimal Transport (OT).** In the context of style shift detection, we treat each sentence as a **cloud** of token embeddings. OT measures the minimal "work" to transform one cloud into the other. This way, we obtain a permutation-invariant style distance.
- **Sinkhorn solver.** This algorithm provides an efficient approximation of the entropy-regularized OT cost by iterative matrix scaling. It makes OT computation feasible by reducing the complexity from $O(n^3)$ to $\approx O(k\,n^2)$, where k is the number of Sinkhorn iterations. Since the updates are basic matrix operations, there exist particularly **fast GPU** implementations.