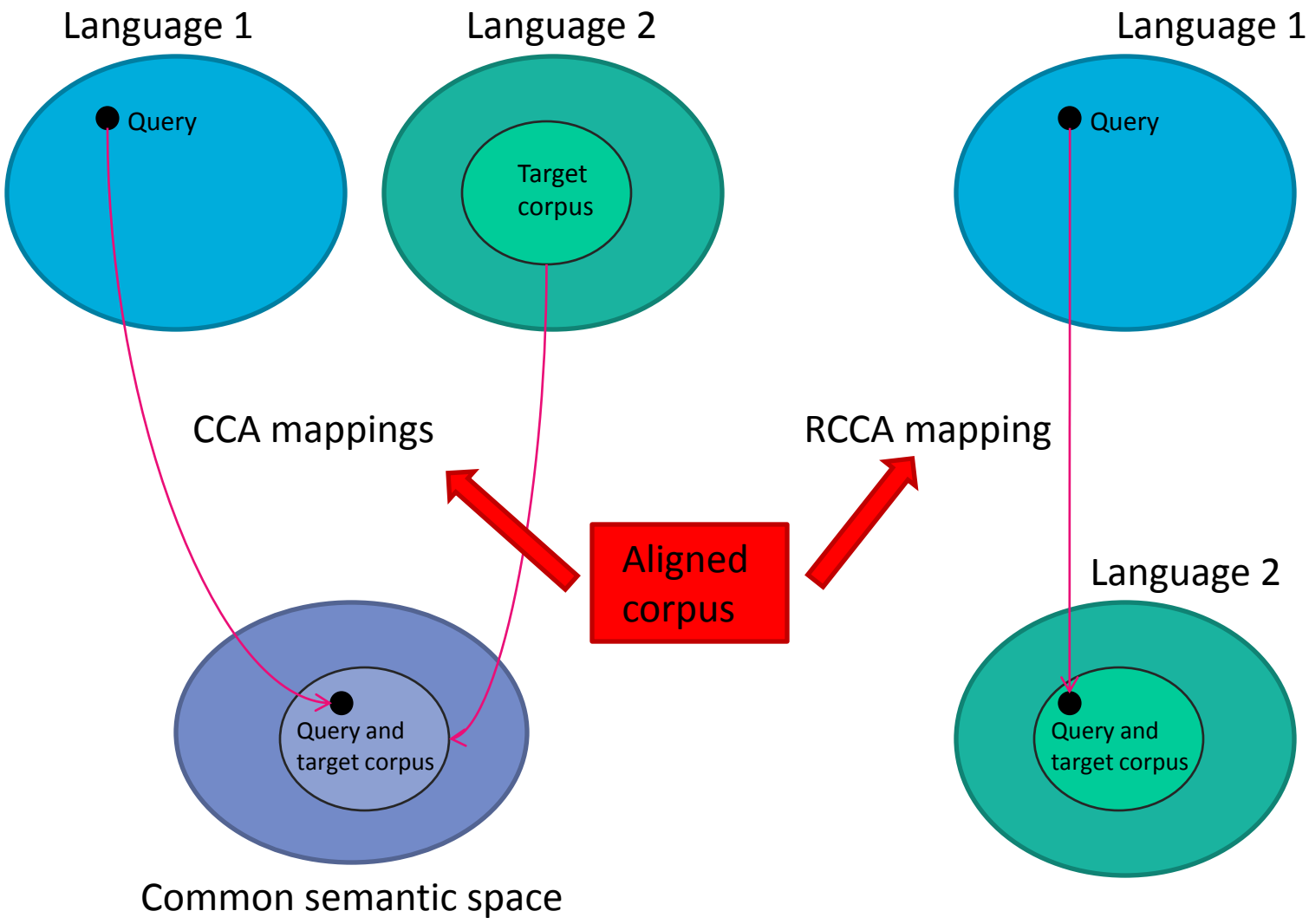
Regression CCA – information retrieval



Optimization problem

$\boldsymbol{X} \in R^{m,d}$, $\boldsymbol{Y} \in R^{n,d}$ (aligned corpus), $\boldsymbol{q} \in R^{m}$ (query)

Optimization problem:

$$max_{\alpha}\ q'XY'\alpha$$
$$\text{s.t.}\quad \alpha' Q\, \alpha = 1$$

- Regularized variance: $Q := (1-\kappa)YY' + \kappa I$
- Lagrangian → system of linear equations.

Solution:

$$\boldsymbol{\alpha} := \boldsymbol{Q^{-1}(YX'q)} \quad \textit{(without rescaling)}$$

Conjugate Gradient method suitable, take advantage of sparsity

Can be done online in query time!