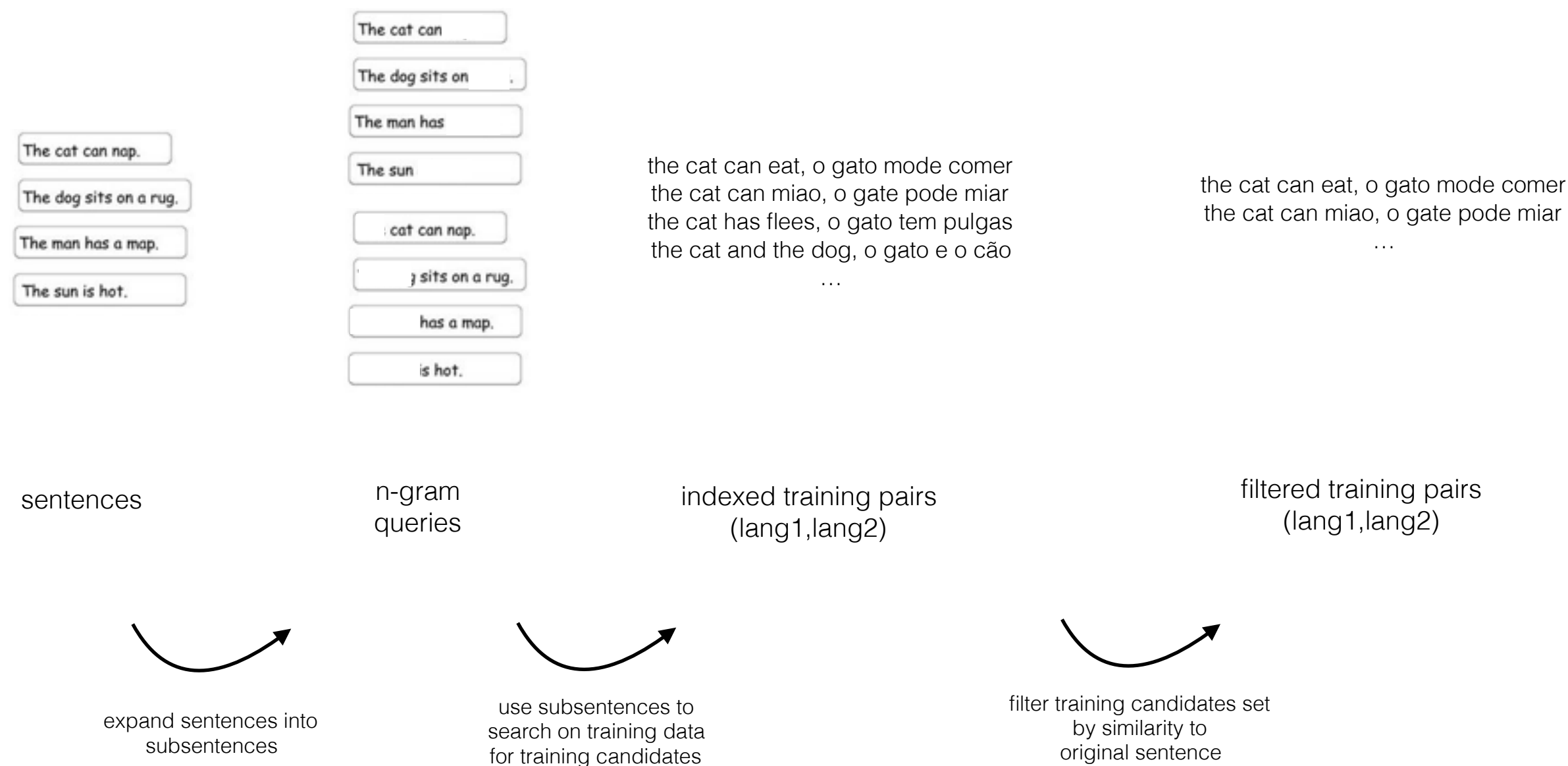


Continuous learning training set selection



Similarity measure

$$\textit{sim}(q_1, q_2) = \lambda * \textit{struc}(q_1, q_2) + (1 - \lambda) * \textit{grammar}(q_1, q_2)$$

$$\textit{struc}(Q_1, Q_2) = \frac{|Q_1 \cap Q_2|}{\max(|Q_1|, |Q_2|)}$$

Normalised overlap of *meaningful* words between set Q1 and Q2

$$\textit{grammar}(q_1, q_2) = \text{longest common subsequence}(\hat{q}_1, \hat{q}_2)$$

Normalised longest common subsequence of the *grammatical* representation of q1 and q2.

q1: "The cat Mike ate a plum"

p1: "ART NOUN PNOUN VERB ART NOUN"

q2: "John ate a banana"

p2: "PNOUN VERB ART NOUN"

p1: "ART NOUN PNOUN VERB ART NOUN"

.....

p2

Examples (on real data)

Query	Top Candidates on training data	Scores
ich weiß	<p>Ich weiß .</p> <p>Ja , ich weiß .</p> <p>Ich weiß , ich weiß , ich weiß .</p> <p>Ich weiß es nicht . Vielleicht . Wer weiß ?</p>	<p>1</p> <p>0.88</p> <p>0.88</p> <p>0.8125</p>
Ja, guten Tag meine Damen, meine Herren.	<p>Meine Damen und Herren , das Scottish Ensemble .</p> <p>Willkommen in Oxford , meine Damen und Herren .</p> <p>Aicha el @-@ Wafi : Guten Tag meine Damen und Herren .</p> <p>Meine Damen und Herren , kommen Sie näher .</p>	<p>0.517</p> <p>0.483</p> <p>0.475</p> <p>0.45</p>
mein Name ist Christian Kromberg und ich bin, seltsamer Begriff, Beigeordneter bei der Stadt Essen.	<p>Ihr Name ist ?</p> <p>Mein Name ist Arvind Gupta und ich bin Spielzeugmacher</p> <p>Mein Name ist Ezra Gelehun</p> <p>Mein Name ist Kate Hartman .</p> <p>Mein Name ist Taylor Wilson .</p>	<p>0.41</p> <p>0.355</p> <p>0.3333</p> <p>0.3333</p> <p>0.3333</p>