

1. Probabilistisches Sprachmodell - Verständnisfragen

- a. Definieren Sie den Begriff eines *probabilistischen Sprachmodells* für eine Sprache.
- b. Bewerten Sie die folgende Aussage als richtig oder falsch: „In einem probabilistischen Sprachmodell ist die berechnete Wahrscheinlichkeit ein Maß für die syntaktische und semantische Korrektheit eines Satzes.“
- c. Für die Berechnung eines Sprachmodells sollen *Wort-Trigramme* verwendet werden. Erläutern Sie, wie das probabilistische Sprachmodell einer Sprache mit Hilfe von Trigrammen berechnet werden kann.

2. Probabilistisches Sprachmodell - Beispiele

Ein Trainingskorpus für das Deutsche bestehe aus den Sätzen

- (1) Johann gibt Marie das Manuskript.
- (2) Peter sieht, dass Johann Marie das Buch gibt.

- (a) Berechnen Sie die Wahrscheinlichkeit des Satzes: „Johann gibt Marie das Buch.“ in einem Trigramm-Modell, das auf den Sätzen (1)-(2) trainiert wurde.
- (b) Berechnen Sie die Wahrscheinlichkeit des Satzes „Johann gibt Marie das Buch.“ unter Berücksichtigung folgender Gewichtungsfaktoren beim *smoothing* mittels *einfacher Interpolation*:

$$\lambda_3 = 0,8 \text{ (Trigramme)}$$

$$\lambda_2 = 0,15 \text{ (Bigramme)}$$

$$\lambda_1 = 0,05 \text{ (Unigramme)}$$

- (c) Wie verändert sich das Ergebnis von (b), wenn Satzankfangsmarker angewendet werden?

3. Probabilistisches Sprachmodell - Anwendungen

- (a) Nennen Sie drei Anwendungen des probabilistischen Sprachmodells im Bereich der Automatischen Sprachverarbeitung.
- (b) Was versteht man unter POS-Tagging? Wo ist der Unterschied zum Parsen?
- (c) Erläutern Sie die grundlegenden Konzepte, Annahmen und Verfahrensschritte beim POS-Tagging.