

Fortgeschrittene Methoden der statistische maschinelle Übersetzung

Übung 4.1

Abgabe (pdf, Zip) an kaeshammer@phil.uni-duesseldorf.de
bis einschließlich Sonntag, 21.12.2014

1

Gegeben ist folgende Phrasenübersetzungstabelle (Format: $f \ ||| \ e \ ||| \ \phi(f|e)$):

```
la ||| the ||| t1
bleue ||| blue ||| t2
maison ||| house ||| t3
la maison ||| the house ||| t4
la maison bleue ||| the blue house ||| t5
```

Stellen Sie den Suchgraphen dar, den ein phrasenbasierter Dekodierer für die Eingabe $\vec{f} = \text{la maison bleue}$ erzeugt. Der Anfang ist in Fig. 1 bereits gegeben. Wir nehmen ein simples Übersetzungsmodell mit Phrasenübersetzungswahrscheinlichkeiten, Trigramm-Sprachmodell und distanzbasierter Umordnungsfunktion an. Der Dekodierer rekombiniert Hypothesen (falls möglich), wendet aber kein Pruning an.

Jede (Teil-) Hypothese speichert also die bereits übersetzten f Wörter, die letzten beiden e Wörter (nötig für die Berechnung der Sprachmodellwahrscheinlichkeit aller Erweiterungen), den End-Index der zuletzt übersetzten f -Phrase (nötig für die Berechnung der Umordnungsfunktion aller Erweiterungen) und die Wahrscheinlichkeit.

2

In Bezug auf Aufgabe 1: Wieviele Pfade zu vollständigen Hypothese gibt es im Suchgraph, wenn der Dekodierer nicht rekombiniert? Wieviele verschiedene Übersetzungen gibt es? Begründen Sie kurz.

