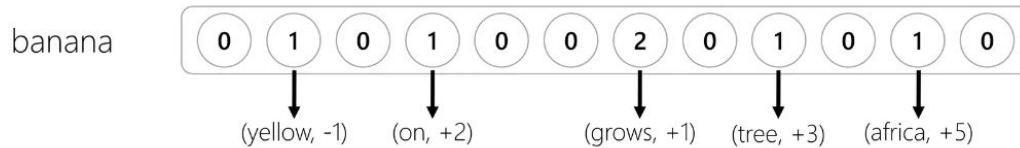


# Computerlinguistik

E07: Bedeutung (2)

# Ausblick: Embeddings – Vektorrepräsentation

- John Firth (1957): “You shall know a word by the company it keeps.”
- Zusammenstellung von Kontexten, in denen Wörter stehen
- Numerische Repräsentation in einem hochdimensionalen Vektorraum



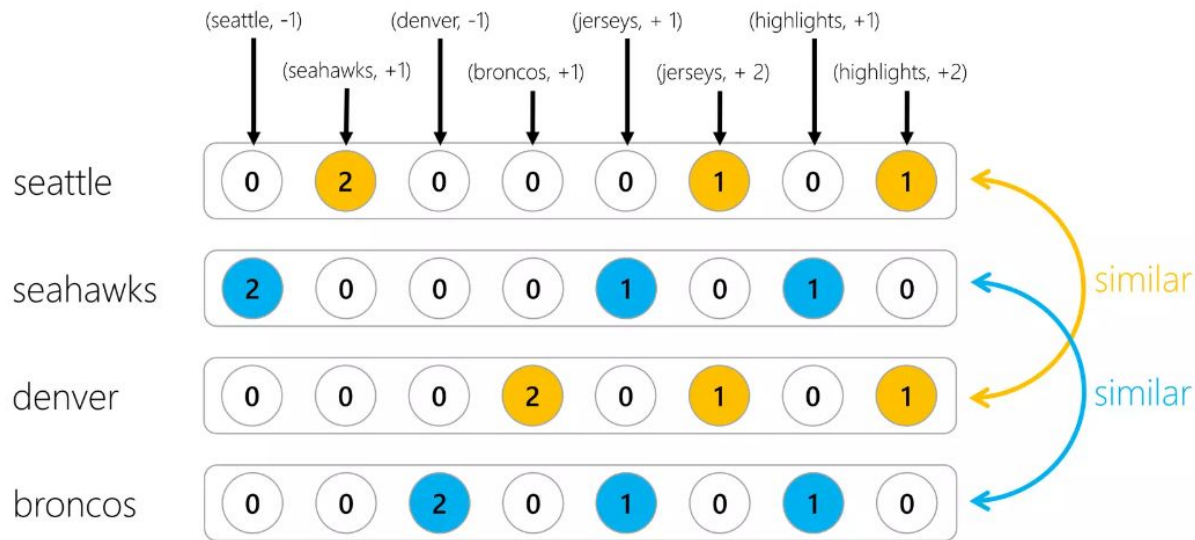
The vector can correspond to **neighboring word context**.

*e.g., "yellow banana grows on trees in africa"*

-1      0      +1    +2    +3    +4    +5

Aus (der insgesamt sehr übersichtlichen Präsentation von Bhaskar Mitra):  
<https://www.slideshare.net/BhaskarMitra3/a-simple-introduction-to-word-embeddings>

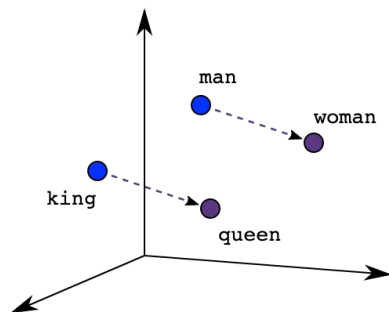
# Ausblick: Embeddings – Vektorähnlichkeit



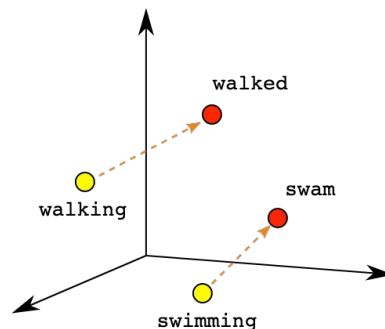
Auch aus (der insgesamt sehr übersichtlichen Präsentation von Bhaskar Mitra):  
<https://www.slideshare.net/BhaskarMitra3/a-simple-introduction-to-word-embeddings>

# Ausblick: Embeddings – Relationen

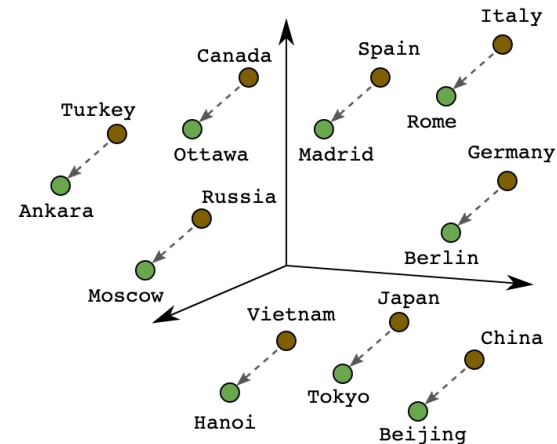
- Relationen zwischen Bedeutungen von Wörtern lassen sich über Vektoroperationen “errechnen”



Male-Female

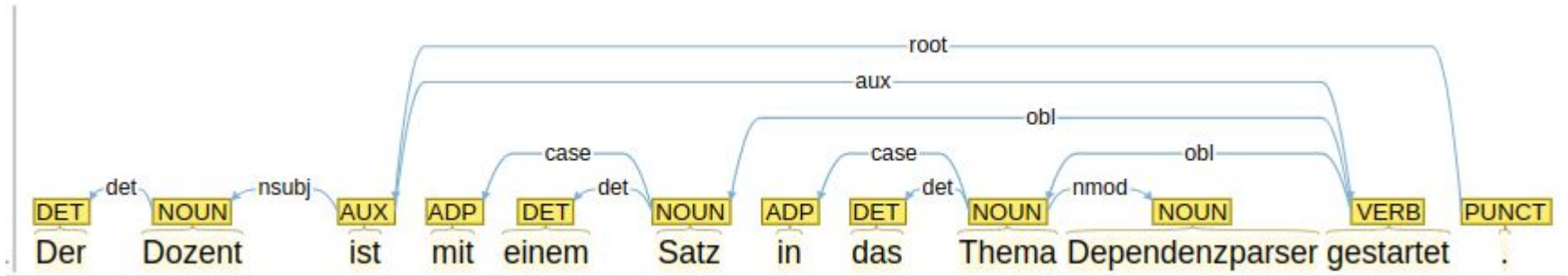


Verb Tense

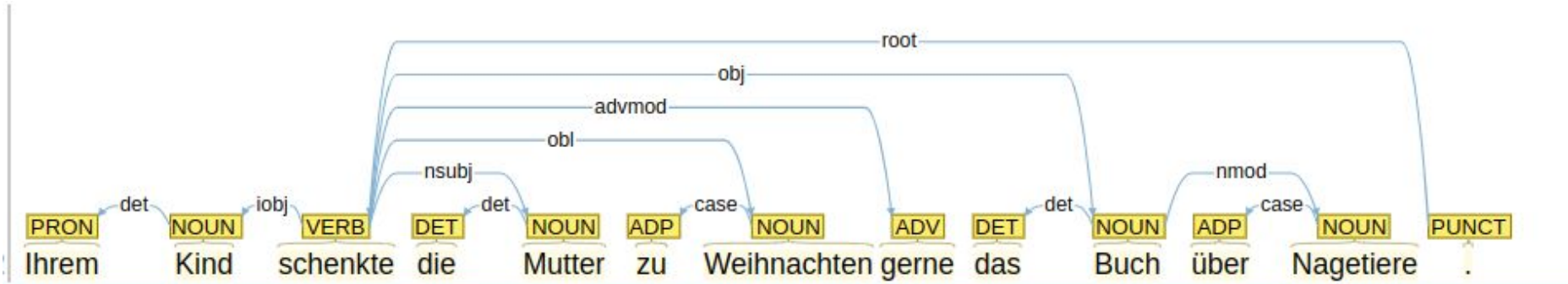


Country-Capital

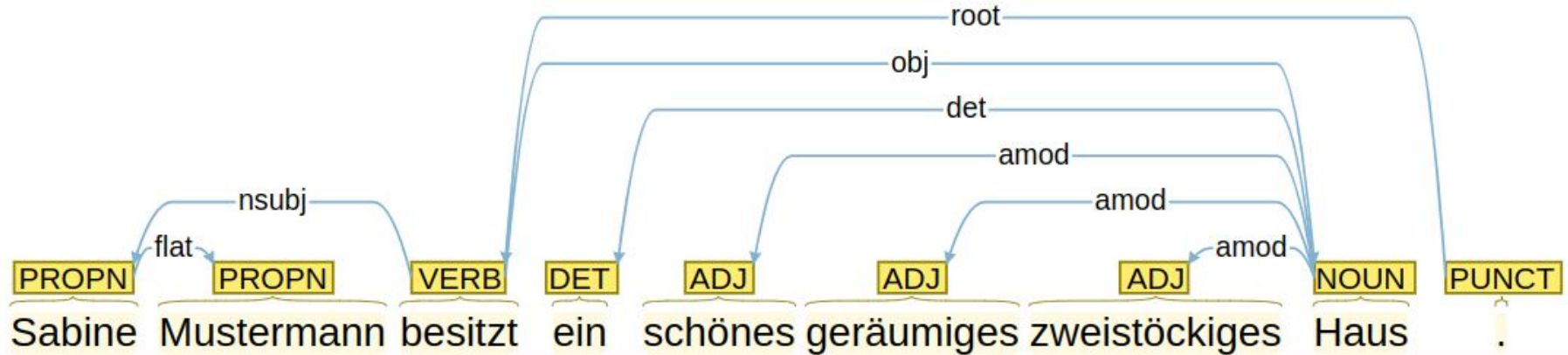
# Musterlösung Dependenzrelationen (I)



# Musterlösung Dependenzrelationen (II)



# Musterlösung Dependenzrelationen (III)



# Literatur / Hausaufgabe

## ➤ Zur Nachbereitung:

- Vervollständigen Sie die Annotationen der Stellenanzeige in Inception (Aufgabenstellung siehe ILIAS).
- Die nächste Stunde ist die letzte vor der Teilklausur. Diese besteht aus einem theoretischen Teil, in dem Konzepte der Computerlinguistik thematisiert werden und einem praktischen Teil, in der Sie einen Text auf unterschiedlichen Ebenen in Inception annotieren sollen. Beide Teile werden in etwa gleich gewichtet. Versuchen Sie zu identifizieren, wo noch Unklarheiten bezüglich des Seminar-/Klausurstoffes bestehen und bereiten Sie Fragen für die nächste Stunde vor!