

# Simullda

---

Structured Interlingua MultiLingual  
Lexical Database Application

Sonja Weber

# Gliederung

---

- Background
- Begriffsklärung
- Multilinguale Datenbanken
- WordNet
- Maschinelle Übersetzung
- Formale Begriffsanalyse und Simulda
- Interlingua
- Probleme und Schwierigkeiten

# Background

---

- Simulda ist ein Projekt der Universität Utrecht und des Opleideninstituut CKI
- Teilgebiete aus der Philosophie (Logik), Linguistik, Psychologie und Informatik
- → Multidisziplinärer Charakter

# Background

---

- ❑ 5000-6000 Sprachen: Für viele Sprachpaare gibt es kein zweisprachiges Lexikon
- ❑ Eine Datenbank, die Sprachpaare direkt übersetzt ist von Vorteil
- ❑ Deskriptive Übersetzung soll ermöglicht werden, wo die Ausgangssprache keine Bedeutung in der Zielsprache hat.

# Begriffsklärung

---

- Homonymie → Wortform hat verschiedene Bedeutungen, die nicht miteinander in Beziehung stehen
- Polysemie → Wort hat verschiedene Bedeutungen, zwei Worte müssen aber miteinander verwandt sein → gleiche etymologische Wurzel oder semantisch miteinander in Verbindung stehen/ Homonymie und Polysemie sind schwierig auseinander zu halten
- Synonym: Wort, dass die gleiche Bedeutung wie ein anderes Wort hat

# Begriffsklärung

---

- Hyponym: Unterbegriff eines Begriffs
- Hyperonym: Oberbegriff eines Begriffs
- Meronym: Teil- von- Beziehung
- Antonymie: Gegenwörter --> Wörter die Gegensätze beschreiben

# Multilinguale Datenbanken

---

- Interlingua: neutrale „Zwischensprache“, die grammatische Informationen des Quelltextes beinhaltet. Grammatische Informationen der Zielsprache wird hieraus erzeugt → Hilfreich bei komplexen Ausdrücken

# Multilinguale Datenbanken

---

- Repräsentation von verschiedenen Spezifika → Formale Begriffsanalyse
- Ziel war es ein System zu entwickeln, dass die Möglichkeit hat für jedes Sprachpaar eine Übersetzung bereitzustellen, jedoch ohne zweisprachige Lexika zu entwickeln



# Multilinguale Datenbanken

## Parallele Wortlisten

---

- Eine Ausgangssprache wird in eine Zielsprache übersetzt, bzw. in mehrere Zielsprachen
- Problem: Lexical gaps → Es gibt nicht alle Bedeutungen in der Zielsprache
- Bank → banque (fr) / rive(fr).

# Multilinguale Datenbanken

---

- Verlinkung über synonyme Wörter kann nur funktionieren, wenn es eins zu eins Übersetzungen gibt
- Anforderungen: Es sollten Wörter und Bedeutungen auseinander gehalten werden können

# Multilinguale Datenbanken

## Beispiel OMBI →



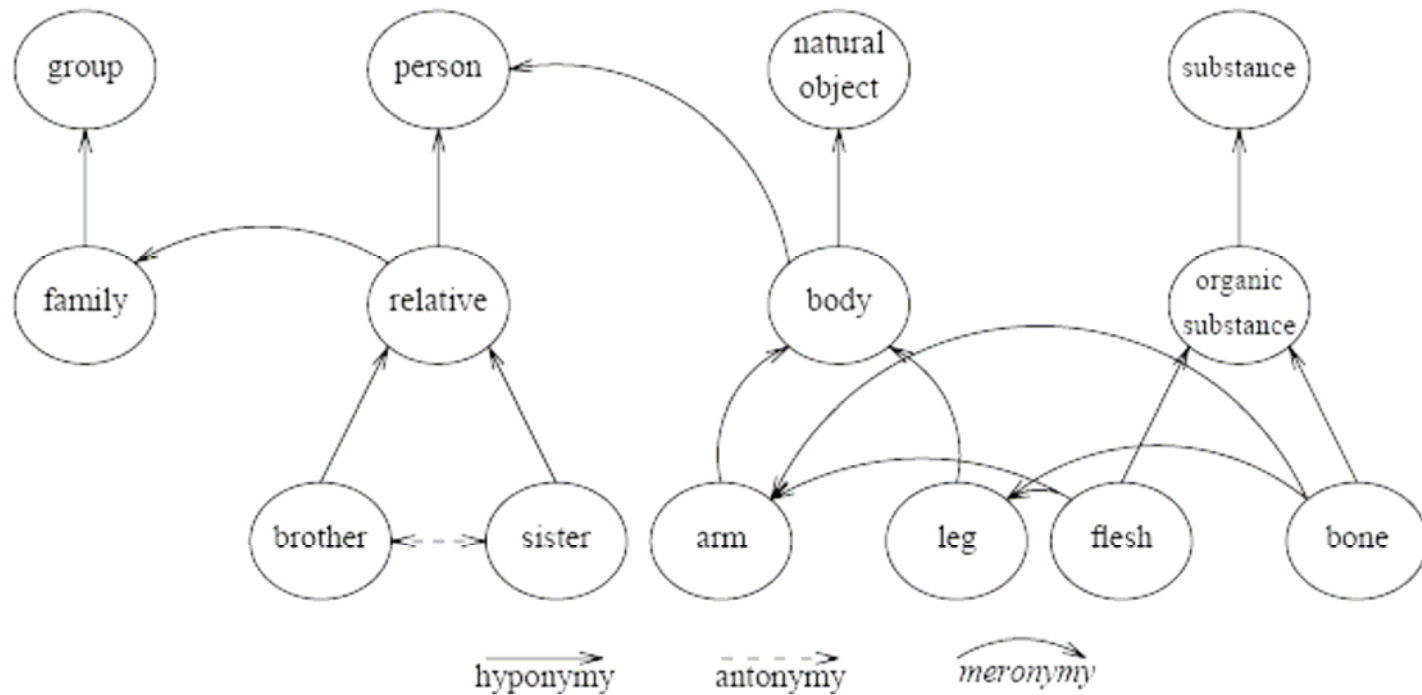
Figure 1.1: OMBI structure for *bank*

# WordNet

---

- ❑ Organisiert mittels Bedeutungen
- ❑ Bedeutung eines Wortes entsteht im Zusammenhang mit einem anderen Wort → Beziehung der Wörter untereinander
- ❑ Arbeit mit Synsets ( ein oder mehrere Synonyme)

# WordNet



# Maschinelle Übersetzung

---

- Simullda : Human consumption approaches → lexikographische Informationen für den Benutzer
- Andere Systeme arbeiten mittels maschinellem Gebrauch
- Auch bei Simullda sind Elemente der maschinellen Übersetzung zu finden, jedoch nicht ausschließlich

# Maschinelle Übersetzung

---

- Formale Repräsentation von Wortbedeutungen, keine Definitionen → Informationen die die Problematik des Übersetzens betreffen
- Keine Ober- und Unterbegriffe → Beschreibung der Wortart

# Simulda

---

- → Simulda soll lexikographische Information besser organisieren
- Mehr Informationen zur Bedeutung eines Wortes sollen angegeben werden , verschiedene Bedeutungsunterschiede zwischen den Sprachen sollen klargestellt werden
- Mittels Formaler Begriffsanalyse



# Formale Begriffsanalyse und Simulda

---

- Kontext besteht aus Gegenständen und Merkmalen und einer Relation zwischen diesen
- Gegenstände → Lemmata
- Merkmale können beliebig festgelegt werden → nach Vorgabe der Definitionen einsprachiger Wörterbücher
- In der Regel von einem Lexikographen festgelegt

# Beispiel eines Kontexts

---

	horse	male	female	adult	young
horse	×				
stallion	×	×		×	
mare	×		×	×	
foal	×				×
filly	×		×		×
colt	×	×			×

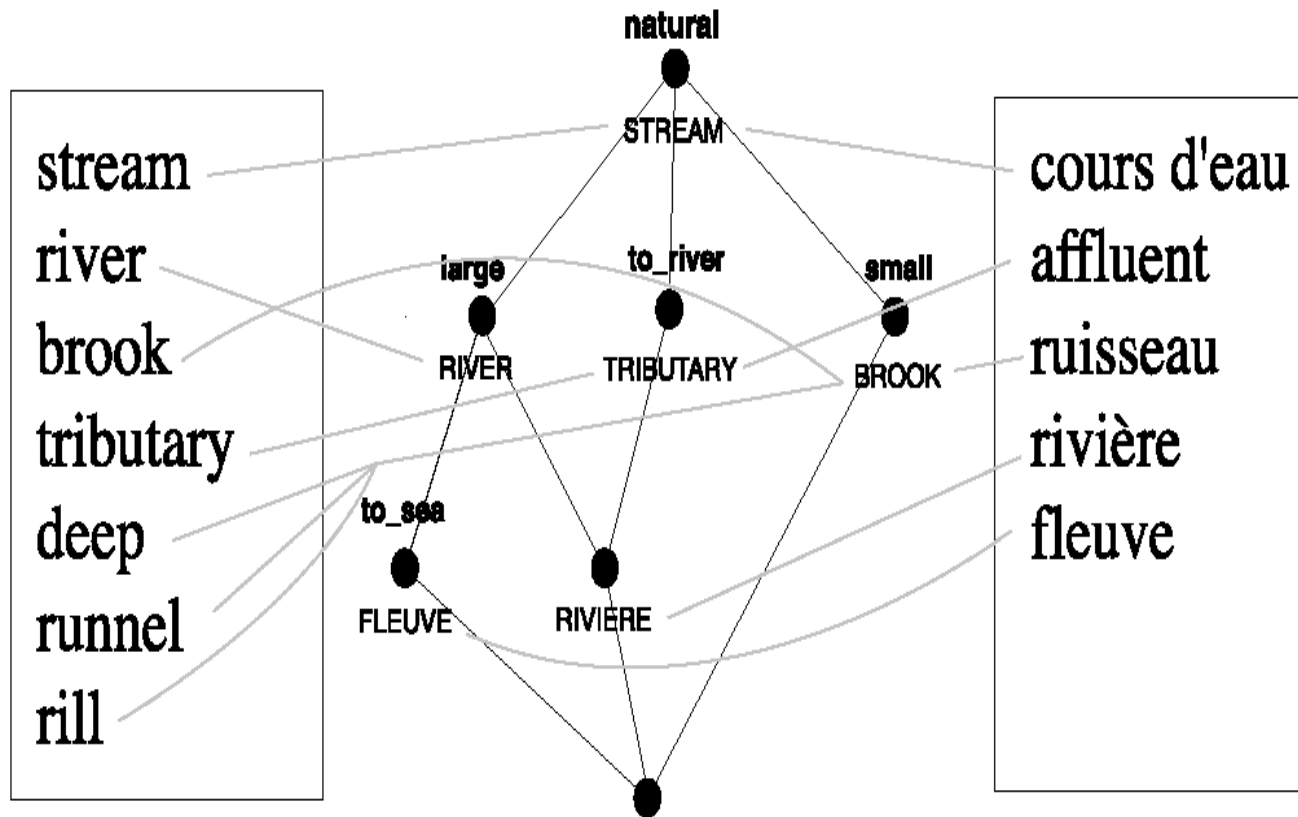
# Formale Begriffsanalyse und Simulda

## Hasse Diagramme

---

- Vorteil: informativere Struktur gegenüber einer Kreuztabelle
- Übersichtlicher als Tabellen
- Übersetzungen befinden sich außerhalb des Hassediagramms und sind über Linien mit den jeweiligen Gegenständen bzw. Merkmalen verbunden

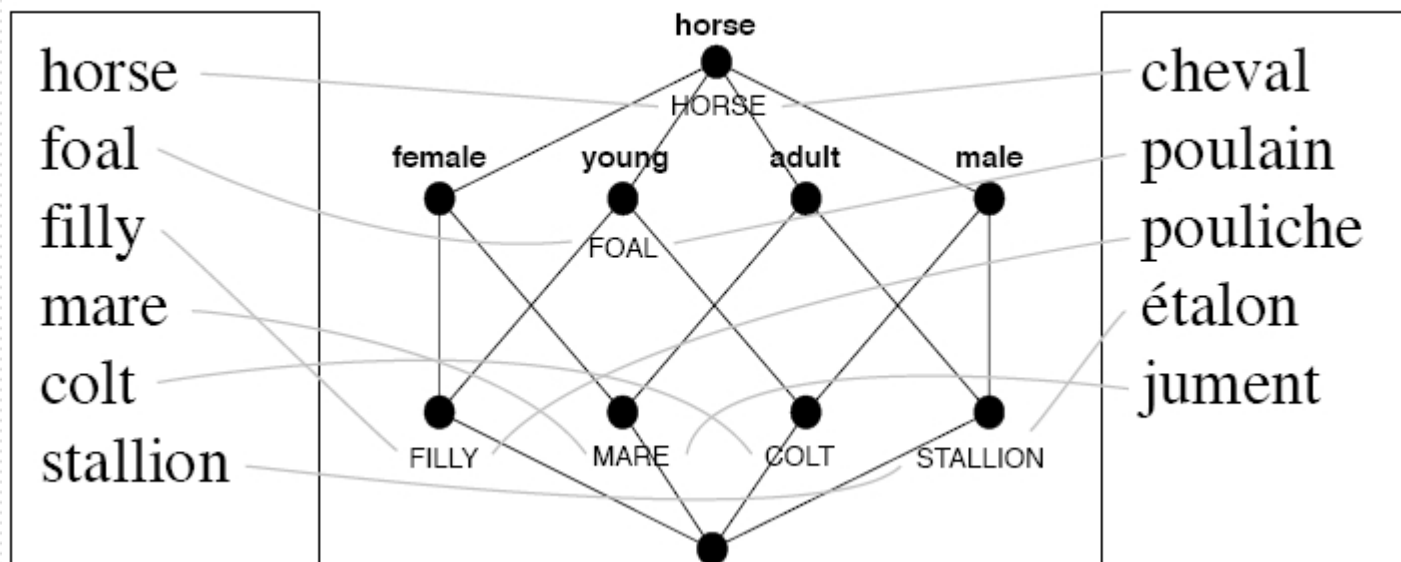
# Beispieldiagramm



# Formale Begriffsanalyse und Simulda

---

- Merkmale von Simulda sollen auf semantischer Ebene angesiedelt sein und nah an Wörterbuchdefinitionen sein → lexikographischer Kontext
- Im Idealfall wird die Bedeutung der Zielsprache parallel zur Ausgangssprache gefunden



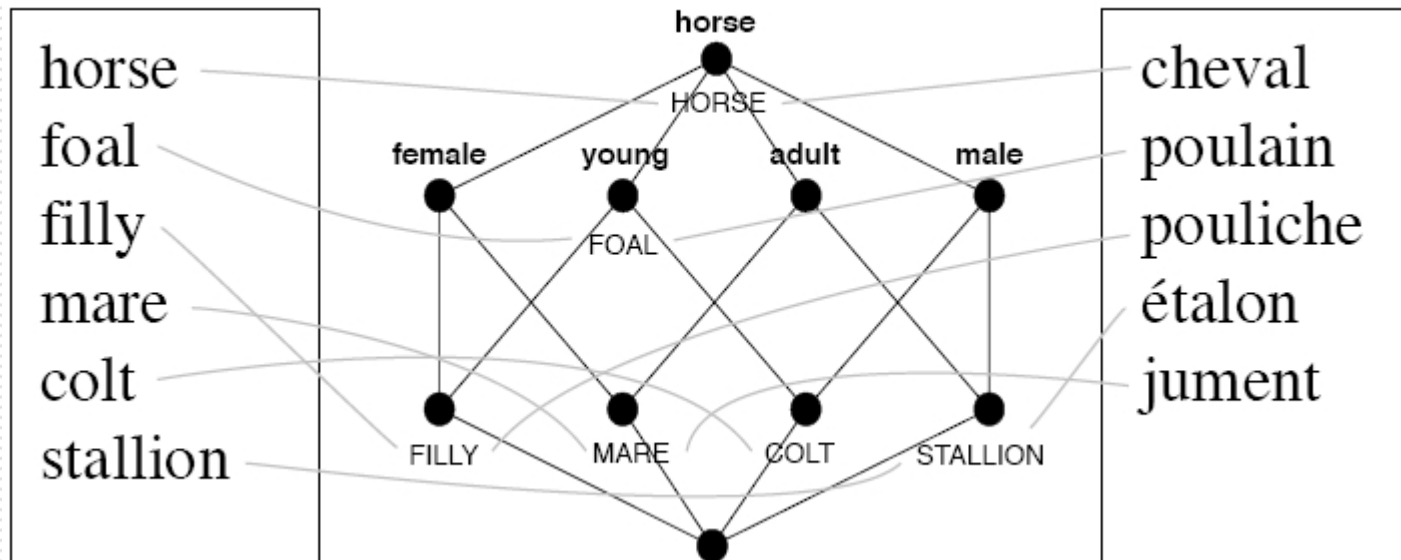
# Formale Begriffsanalyse und Simulda Language Boxen

---

- ❑ FBA Gitter soll sprachunabhängig gemacht werden → Wortformen sollen in language boxes geschrieben werden
- ❑ Bedeutungen sind mit den verschiedenen Sprachen verbunden, werden aber nicht mit ihnen identifiziert
- ❑ Wortformen sind nicht ins Diagramm integriert sondern befinden sich außerhalb
- ❑ Merkmale müssen sprachunabhängig sein → groß, klein, jung etc.

# Language Boxes

---





# Formale Begriffsanalyse und Simulda

## Lexical Gaps

---

- Lexical Gaps: Wenn es in der Zielsprache keine Übersetzung für einen bestimmten Ausdruck in der Ausgangssprache gibt
- Bsp.: Colt → englisch männliches Fohlen → im Französischen gibt es hierfür keine Entsprechung
- Ersatz: Colt ist Unterbegriff von Foal und Foal im Französischen → poulain
- Colt besitzt die Merkmale: young and male
- Male auf Französisch: mâle → colt = mâle poulain

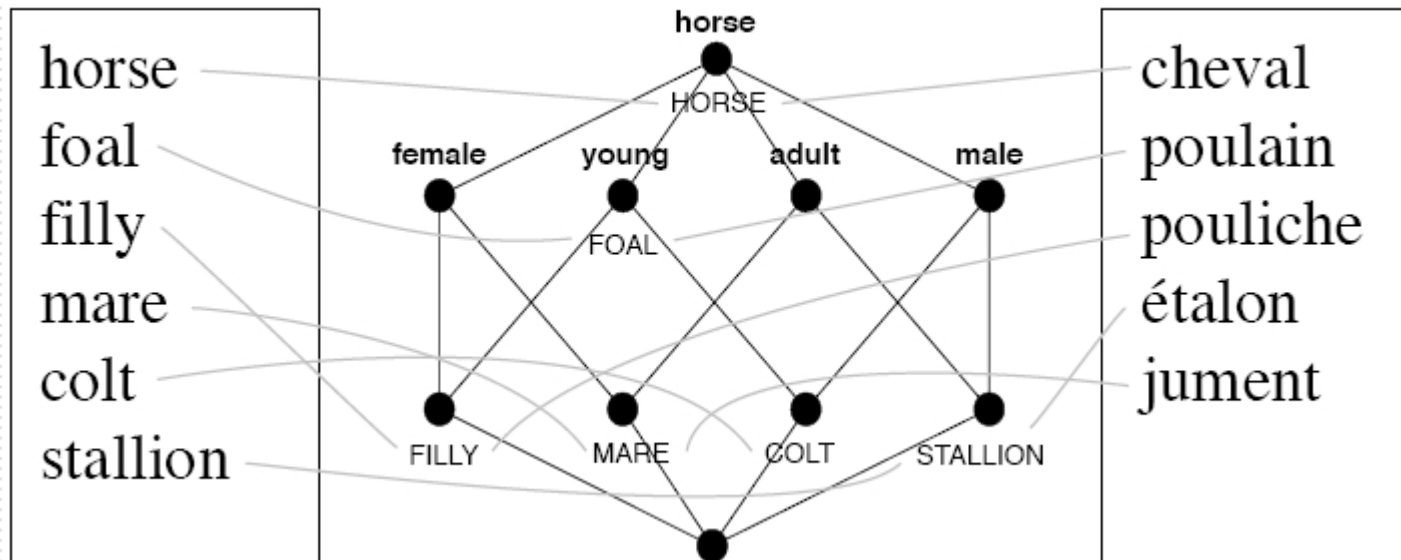
# Formale Begriffsanalyse und Simulda Lexical Gaps

---

- Um ein Lexical Gap zu füllen muss ein Oberbegriff gewählt werden und mit den Merkmalen des Begriffs der Ausgangssprache angereichert werden
- Es gibt nicht nur ein Ergebnis → Wichtig ist, dass das Ergebnis richtig ist

# Colt als Lexical Gap

---



# Interlinguale Bedeutung

## Wortbedeutung

---

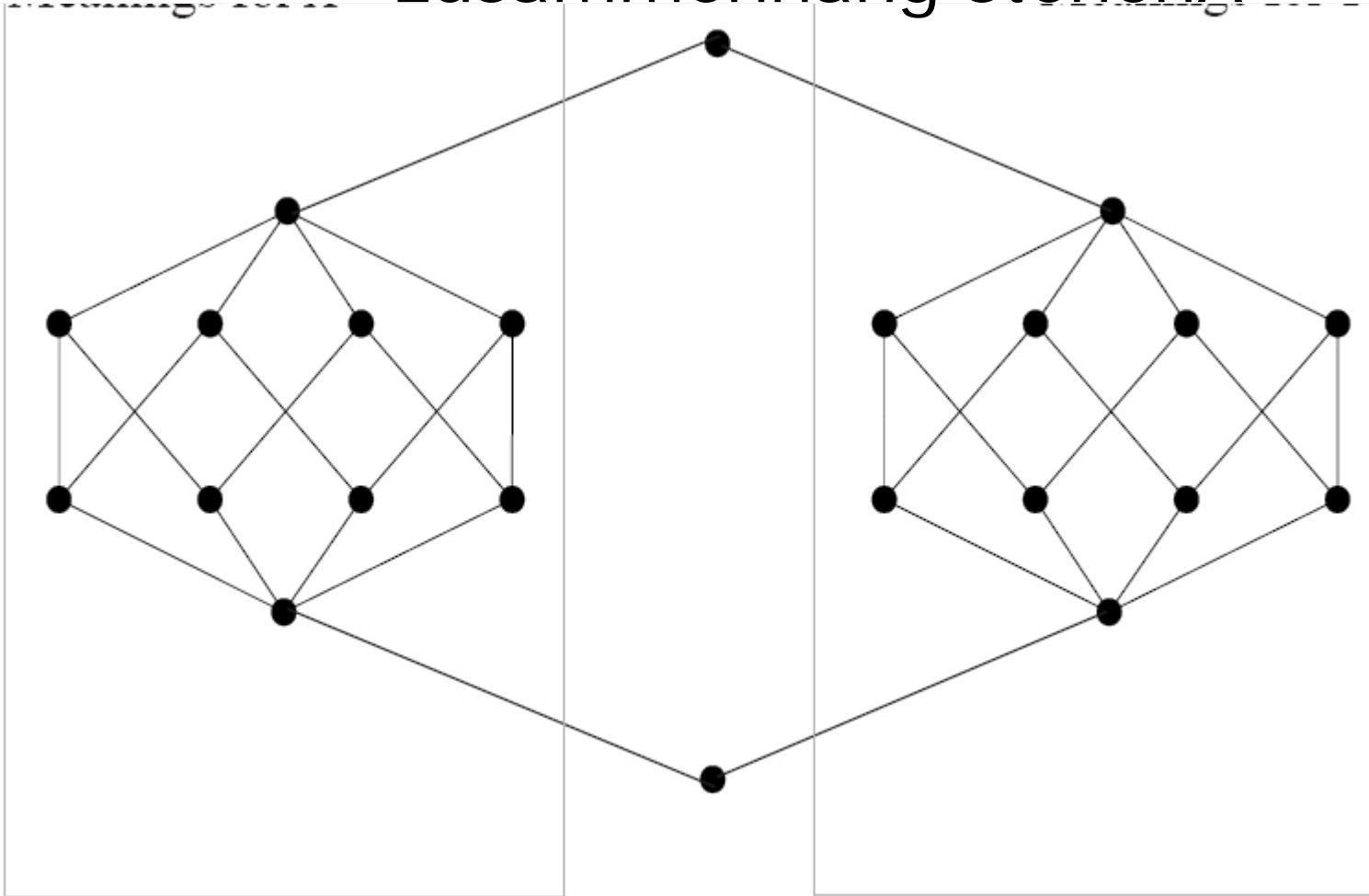
- Hinter jedem Wortsinn steckt eine Idee → Begriff
- Bei Simulda geht es um formale Begriffe nicht um Begriffe → Paare von interlingualer Bedeutung und definierten Merkmalen

# Interlingua

---

- Verschiedene Sprachen können die gleiche interlinguale Bedeutung ausdrücken → Bedeutungen sind interlingual und sprachenunabhängig
- Nicht alle interlingualen Bedeutungen müssen in allen Sprachen lexikalisiert werden ( Lexical gaps)
- Es gibt auch Sprachen, die nicht miteinander in Zusammenhang gebracht werden können

# Zwei Sprachen , die nicht miteinander in zusammenhang stehenA



# Interlingua

---

- ❑ Simulda ist nur effektiv, wenn es genügend Überlappungen zwischen Sprachen gibt
- ❑ Es sollten jedoch auch Lexical Gaps auftreten
- ❑ Interlingua im Bezug auf Simulda → Begriffsverband
- ❑ Lexikalisierung findet in den Word boxes statt

## Probleme und Schwierigkeiten

---

1. Durch kulturell bedingte Unterschiede kann eine Übersetzung schwierig werden
  - Lunch vs. Collazione
  - Simulda basiert auf der Analyse des Inhalts einsprachiger Wörterbücher
  - Kulturelle Unterschiede können müssen aber nicht zu verschiedenen Übersetzungen führen
2. Synonyme: können lexikalisch variieren
  - In Simulda handelt es sich um Synonyme, wenn sie dieselbe interlinguale Bedeutung beschreiben
  - Ein gutes Wörterbuch muss damit umgehen können



# Aufbau

