

Musterlösung zur Probeklausur

Diese Probeklausur dient zur Klausurvorbereitung und soll einen Eindruck von möglichen Aufgabentypen vermitteln. Bitte beachten Sie dass alle Inhalte aus den Vorlesungen, Übungsaufgaben und Lese-Aufgaben (siehe Übungsblätter) zum Prüfungsstoff gehören. Die Klausur kann also durchaus Aufgaben zu Themen enthalten, die in der Musterlösung nicht angesprochen werden.

1. Sie haben zu Beginn der Vorlesung die beiden "klassischen" Sprachverarbeitungssysteme ELIZA und SHRDLU kennengelernt. Der wesentliche Unterschied zwischen beiden Systemen ist, dass SHRDLU im Gegensatz zu ELIZA wissensbasiert arbeitet. Erläutern Sie diese Feststellung an je einem Beispiel (d.h. einer Dialogsequenz mit System- und Benutzeräußerung) zu den beiden Systemen (Sie können die Beispiele frei, und falls Ihnen das lieber ist, auch in deutscher Sprache, wählen).

Beispiel 1 (ELIZA):

ELIZA: Hallo, ich bin Eliza. Erzähle mir von Deinem Problem...

Benutzer: Die CL Probleklausur war zu schwierig.

ELIZA: Bist du nun weiter gekommen?

(ELIZA erkennt 'war' als Vergangenheitsform und evtl. 'schwierig' und wählt eine entsprechende Antwort aus.)

Benutzer: Etwas, aber ich konnte noch nicht alle Fragen über Automaten beantworten.

ELIZA: Die Antwort auf alle Fragen zu Automaten kennst du doch selber.

(ELIZA erkennt 'nicht X beantworten' und setzt im entsprechenden Template für X 'alle Fragen zu Automaten' ein.)

Die Reaktionen des Systems bestehen nur aus vorgegebenen Antworten die durch entsprechende Schlüsselwörter/Muster in der Benutzereingabe ausgelöst werden. Gegebenenfalls werden einzelne Teile in Antwort-Templates durch Teile der Benutzereingaben ersetzt. Es wird dabei kein linguistisches -, Kontext-, oder Weltwissen verwendet.

Im Gegensatz dazu zeigen die Antworten von SHRDLU dass es sich um ein wissensbasiertes System handelt.

Beispiel 2 (SHRDLU):

Benutzer: Nimm die rote Pyramide und lege sie in die Kiste.

SHRDLU: Ich nehme an mit 'sie' meinst du die rote Pyramide. Ok.

(Um aufzulösen dass 'sie' auf die Pyramide referenziert ist linguistisches Wissen notwendig)

Benutzer: Was ist in der Kiste?

SHRDLU: Die rote Pyramide.

(SHRDLU hat offensichtlich Kenntnis über den aktuellen Zustand der Blockwelt).

Benutzer: Lege den grünen Würfel auf die rote Pyramide.

SHRDLU: Sorry, die rote Pyramide kann den grünen Würfel nicht tragen.

(SHRDLU weiß, dass sich auf Pyramiden keine weiteren Objekte legen lassen.)

2. Beantworten sie kurz die folgenden Fragen zu Korrekturprogrammen:

- Warum ist eine lexikonbasierte Rechtschreibkorrektur (d.h. es wird angenommen dass jedes Wort das nicht genau so im Lexikon steht ein Nicht-Wort ist) für Sprachen wie z.B. Deutsch nicht ausreichend?

Deutsch besitzt eine sehr produktive Morphologie (z.B. Komposition), sodass nie alle möglichen Wörter im Lexikon erfasst werden können.

- Nachdem in einem Korrekturprogramm ein Nicht-Wort erkannt wurde, soll dieses durch das korrigierte Wort ersetzt werden. Welche Methode kennen sie um das richtige Wort auszuwählen oder dem Benutzer mögliche Ersetzungen vorzuschlagen.

Das Korrekturprogramm könnte dem Benutzer n Lexikonwörter mit der geringsten Levenshtein-Distanz zum Nicht-Wort vorschlagen.

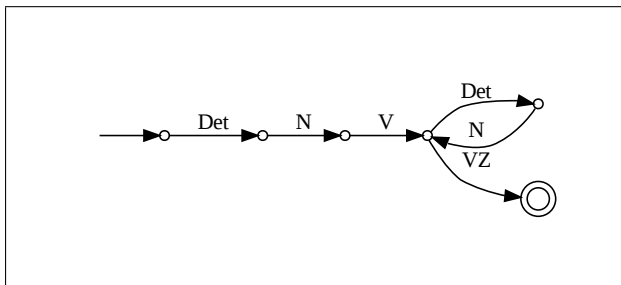
- Warum ist im allgemeinen eine kontextabhängige Korrektur wünschenswert?

Es kann vorkommen dass einige Schreibweisen zwar prinzipiell im Lexikon vorkommen, aber nicht der richtigen Schreibweise des Wortes entsprechen das im Kontext eigentlich beabsichtigt war. Beispiel: "...ein Kleid auf den Laib schneiden."

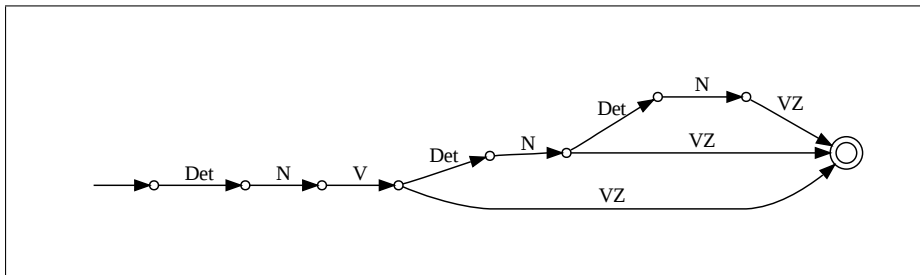
3. In folgenden Beispielen sind abgetrennte Verbzusätze (VZ) hervorgehoben:

Det N V VZ	Der Zug fährt ab
Det N V Det N VZ	Der Kellner sammelt die Teller ein
Det N V Det N Det N VZ	Der Trainer teilt den Spielern die Positionen zu

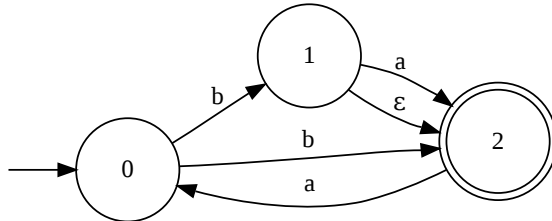
- a) Entwerfen Sie ein Zustandsdiagramm, das einfache Sätze mit abgetrennten Verbzusätzen erkennt. Akzeptiert werden sollen Folgen der Art:
Nominalausdruck - Verb - beliebig viele Nominalausdrücke - Verbzusatz
 Nominalausdrücke sollen immer aus genau einem Artikel gefolgt von einem Nomen bestehen.



- b) Geben Sie ein zweites Zustandsdiagramm an, das zwischen Verb und Verbzusatz maximal zwei Nominalausdrücke akzeptiert, ansonsten aber wie das erste Diagramm ist.



4. Betrachten Sie den folgenden NEA A_1 über dem Alphabet $\Sigma = \{a, b\}$:



- a) Konstruieren Sie mit den in der Vorlesung vorgestellten Verfahren:
 i. einen zu A_1 äquivalenten buchstabierenden Automaten.

Schritt 1: Keine Mehrsymbolkanten.
 Schritt 2:

p	$V_\epsilon(p)$	$N_\epsilon(p)$
0	{0}	{0}
1	{1}	{1,2}
2	{1,2}	{2}

Vorbereich und Nachbereich:

- ii. durch Potenzautomatenkonstruktion einen DEA, der die gleiche Sprache wie A_1 erkennt.

$A'_1 = \langle K, \Sigma, \delta, s, F \rangle$ mit
 $K = \{ \{0\}, \{1,2\}, \{0,2\} \}$, $\Sigma = \{a, b\}$, $s = \{0\}$, $F = \{ \{0\}, \{0,2\} \}$,

	a	b
$\delta =$	\emptyset	{1,2}
{1,2}	{2,0}	\emptyset
{2,0}	{0}	{1,2}

- b) Beschreiben Sie in Worten die Sprache, die der Automat erkennt.

Folge aus a 's und b 's, in der Ketten von b 's nicht länger als eins, Ketten von a 's nicht länger als zwei sind. Die Folge beginnt mit b und kann nicht auf aa enden.

5. Die Potenzautomatenkonstruktion erzeugt aus einem buchstabierenden NEA mit n Zuständen

einen DEA mit 2^n Zuständen. Der Potenzautomat für einen NEA mit 10 Zuständen hat also 1024 Zustände. Trotzdem kommt man beim Aufschreiben des DEA meist mit einer weit kleineren Zustandsmenge aus. Wieso?

Die meisten der 2^n Zustände im Potenzautomaten sind nicht erreichbar (d.h. vom Startzustand aus gibt es keinen Pfad in dem diese Zustände vorkommen) und können beim Aufschreiben ignoriert werden. Der Algorithmus zur Potenzautomatenkonstruktion berücksichtigt ohnehin immer nur erreichbare Zustände.

6. Geben Sie zwei Zerlegungen des Wortes "Abteilungen" an, die ein automatenbasiertes Morphologiesystem erkennen könnte. Beschreiben sie die morphologischen Prozesse denen die Zerlegungen jeweils entsprechen (Flexions-Morphologie, Ableitungs-/Derivations-Morphologie oder Kompositazerlegung).

Ab|₁teil|₂ung|₃en

- 1: Ableitungsmorphologie (Verbalpräfix mit Verbstamm)
- 2: Ableitungsmorphologie (Verbstamm mit Nominalisierungssuffix)
- 3: Flexionsmorphologie (Nominalisierung mit Pluralsuffix)

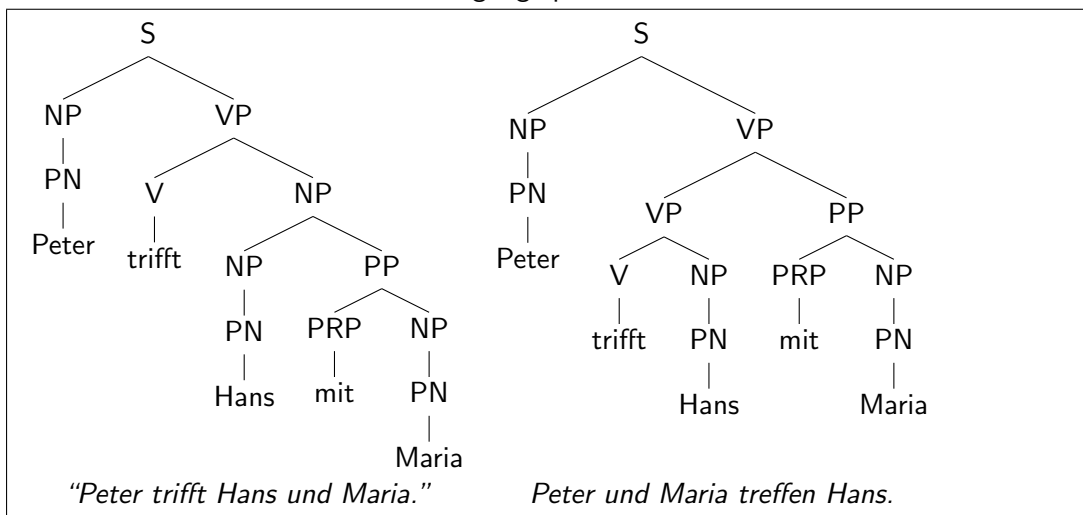
Abt|₁ei|₂lung|₃en

- 1: Kompositazerlegung (Nomen mit Nomen)
- 2: Kompositazerlegung (Nomen mit Nomen)
- 3: Flexionsmorphologie (Nominalisierung mit Pluralsuffix)

7. Gegeben sei eine kontextfreie Grammatik mit den Produktionsregeln

S	→	NP VP	V	→	trifft
VP	→	V NP	PN	→	Maria Peter Hans
VP	→	VP PP	PRP	→	mit
NP	→	PN			
NP	→	NP PP			
PP	→	PRP NP			

- a) Leiten Sie den Satz "Peter trifft Hans mit Maria" auf zwei unterschiedliche Weisen ab. Geben Sie die beiden Strukturbäume an, die aus der Ableitung resultieren und charakterisieren Sie beide Lesarten umgangssprachlich.



b) Zeigen Sie wie der Earley-Algorithmus den Satz verarbeitet. Geben Sie die Chart an, sowie die Schritte, die angewendet wurden.

0	$\langle S \rightarrow \bullet NP VP \quad [0, 0] \rangle$	init	$\langle NP \rightarrow NP \bullet PP \quad [2, 3] \rangle$ $\langle S \rightarrow NP VP \bullet \quad [0, 3] \rangle$ $\langle VP \rightarrow VP \bullet PP \quad [1, 3] \rangle$ $\langle PP \rightarrow \bullet PRP NP \quad [3, 3] \rangle$ $\langle PRP \rightarrow \bullet mit \quad [3, 3] \rangle$	complete complete complete predict predict
	$\langle NP \rightarrow \bullet PN \quad [0, 0] \rangle$	predict		
	$\langle NP \rightarrow \bullet NP PP \quad [0, 0] \rangle$	predict		
	$\langle PN \rightarrow \bullet Maria \quad [0, 0] \rangle$	predict		
	$\langle PN \rightarrow \bullet Peter \quad [0, 0] \rangle$	predict		
	$\langle PN \rightarrow \bullet Hans \quad [0, 0] \rangle$	predict		
1	$\langle PN \rightarrow Peter \bullet \quad [0, 1] \rangle$	scan	$\langle PRP \rightarrow mit \bullet \quad [3, 4] \rangle$ $\langle PP \rightarrow PRP \bullet NP \quad [3, 4] \rangle$ $\langle NP \rightarrow \bullet PN \quad [4, 4] \rangle$ $\langle NP \rightarrow \bullet NP PP \quad [4, 4] \rangle$ $\langle PN \rightarrow \bullet Maria \quad [4, 4] \rangle$ $\langle PN \rightarrow \bullet Peter \quad [4, 4] \rangle$ $\langle PN \rightarrow \bullet Hans \quad [4, 4] \rangle$	scan complete predict predict predict predict predict
	$\langle NP \rightarrow PN \bullet \quad [0, 1] \rangle$	complete		
	$\langle NP \rightarrow NP \bullet PP \quad [0, 1] \rangle$	complete		
	$\langle S \rightarrow NP \bullet VP \quad [0, 1] \rangle$	complete		
	$\langle VP \rightarrow \bullet V NP \quad [1, 1] \rangle$	predict		
	$\langle VP \rightarrow \bullet VP PP \quad [1, 1] \rangle$	predict		
	$\langle V \rightarrow \bullet trifft \quad [1, 1] \rangle$	predict		
	$\langle PP \rightarrow \bullet PRP NP \quad [1, 1] \rangle$	predict		
	$\langle PRP \rightarrow \bullet mit \quad [1, 1] \rangle$	predict		
	2	$\langle V \rightarrow trifft \bullet \quad [1, 2] \rangle$		
$\langle VP \rightarrow V \bullet NP \quad [1, 2] \rangle$		complete		
$\langle NP \rightarrow \bullet PN \quad [2, 2] \rangle$		predict		
$\langle NP \rightarrow \bullet NP PP \quad [2, 2] \rangle$		predict		
$\langle PN \rightarrow \bullet Maria \quad [2, 2] \rangle$		predict		
$\langle PN \rightarrow \bullet Peter \quad [2, 2] \rangle$		predict		
$\langle PN \rightarrow \bullet Hans \quad [2, 2] \rangle$		predict		
$\langle S \rightarrow NP VP \bullet \quad [0, 5] \rangle$		complete		
3	$\langle PN \rightarrow Hans \bullet \quad [2, 3] \rangle$	scan	$\langle PP \rightarrow \bullet PRP NP \quad [5, 5] \rangle$ $\langle PRP \rightarrow \bullet mit \quad [5, 5] \rangle$	complete predict predict
	$\langle NP \rightarrow PN \bullet \quad [2, 3] \rangle$	complete		
	$\langle VP \rightarrow V NP \bullet \quad [1, 3] \rangle$	complete		

8. a) Was ist der Unterschied zwischen syntaktischer Kategorie und Funktion?

Syntaktische Kategorien ordnen Ausdrücke anhand ihres *distribuellen* Verhaltens. Ausdrücke, die in gleichen syntaktischen Umgebungen gegeneinander ausgetauscht werden können, ohne dass der Satz ungrammatikalisch wird, sollten idealerweise die selbe Kategorie besitzen. Auch Ausdrücke, die *intern den selben Aufbau* besitzen, sollten die selbe Kategorie besitzen. Beispiele sind lexikalische Kategorien (Adjektiv, Verb, ...) und phrasale Kategorien (Nominalphrase, Verbalphrase, ...)

Eine *grammatische Funktion* gibt die Relation an, in der ein Ausdruck zu anderen Teilen des Satzes steht. Auf oberster Ebene werden die Funktionen aufgeteilt in *Köpfe*, die Merkmale wie z.B. Subkategorisierung einer Phrase bestimmen, in *Argumente* die von einem Kopf subkategorisiert werden und in *Modifikatoren*, die freie, optionale Ergänzungen eines Ausdrucks darstellen.

b) Wählen Sie aus dem folgenden Satz zwei Konstituenten und geben Sie jeweils deren Kategorie und Funktion an.

“Der Forscher erfindet ein neues Mittel gegen Müdigkeit”

Konstituente	Kategorie	Funktion
<i>neues</i>	Adjektiv	Modifikator zu <i>Mittel</i>
<i>gegen Müdigkeit</i>	Präpositionalphrase	präpositionales Objekt von <i>Mittel</i>
<i>der Forscher</i>	Nominalphrase	Subjekt von <i>erfindet</i>

9. Betrachten Sie alle möglichen Paarungen der folgenden Merkmalsstrukturen:

a) $\left[\begin{array}{c} A \left[\begin{array}{c} B \ a \\ C \ \boxed{1} \ b \end{array} \right] \\ B \ \boxed{1} \end{array} \right]$ b) $\left[\begin{array}{c} B \ b \\ A \ \left[\begin{array}{c} C \ a \end{array} \right] \end{array} \right]$ c) $\left[\begin{array}{c} A \ \left[\begin{array}{c} B \ \boxed{1} \end{array} \right] \\ C \ \boxed{1} \ a \end{array} \right]$ d) $\left[A \ \left[B \ a \right] \right]$

a) Welche Paare sind konsistent/unifizierbar, welche Paare nicht (✓ oder X)?

	a)	b)	c)	d)
a)	✓	X	✓	✓
b)		✓	✓	✓
c)			✓	✓
d)				✓

b) Zwischen welchen Paaren besteht die Subsumptionsrelation \sqsubseteq und in welcher Richtung?

$d) \sqsubseteq a), d) \sqsubseteq c)$

10. Gegeben sei eine Grammatik mit Merkmalsconstraints, sowie ein paar lexikalische Einträge.

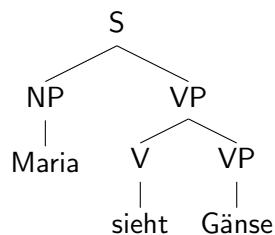
$$\begin{array}{l}
 S \rightarrow NP VP \\
 \quad < S \text{ HEAD} = VP \text{ HEAD} > \\
 \quad < S \text{ HEAD SUBJ} = NP \text{ HEAD} > \\
 VP \rightarrow V NP \\
 \quad < VP \text{ HEAD} = V \text{ HEAD} > \\
 \quad < V \text{ SUBCAT} = NP \text{ HEAD} >
 \end{array}$$

$$\begin{array}{l}
 \text{Gänse:} \left[\begin{array}{c} \text{CAT} \quad \text{NP} \\ \text{HEAD} \left[\begin{array}{c} \text{AGR} \left[\text{NUM} \text{ pl} \right] \\ \text{SEM} \text{ gänse} \end{array} \right] \end{array} \right] \\
 \text{Maria:} \left[\begin{array}{c} \text{CAT} \quad \text{NP} \\ \text{HEAD} \left[\begin{array}{c} \text{AGR} \left[\text{NUM} \text{ sg} \right] \\ \text{SEM} \text{ maria} \end{array} \right] \end{array} \right]
 \end{array}$$

$$\begin{array}{c}
 \text{sieht:} \left[\begin{array}{c} \text{CAT} \quad \text{V} \\ \text{HEAD} \left[\begin{array}{c} \text{SUBJ} \left[\begin{array}{c} \text{AGR} \left[\text{NUM} \text{ sg} \right] \\ \text{SEM} \boxed{1} \end{array} \right] \\ \text{SEM} \left[\begin{array}{c} \text{PRED} \text{ sehen} \\ \text{ARG1} \boxed{1} \\ \text{ARG2} \boxed{2} \end{array} \right] \end{array} \right] \\ \text{SUBCAT} \left[\text{SEM} \boxed{2} \right] \end{array} \right]
 \end{array}$$

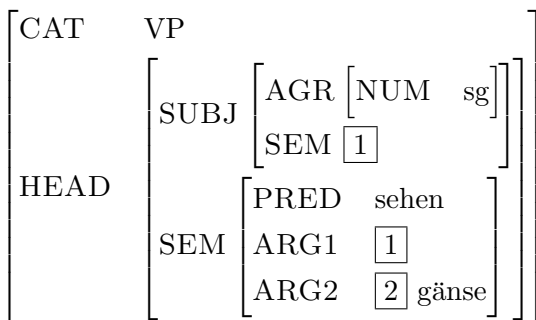
Leiten Sie mit dieser Grammatik den Satz "Maria sieht Gänse" ab, geben Sie insbesondere die Merkmalstruktur für den gesamten Satz an.

Man erhält folgenden Ableitungsbaum:

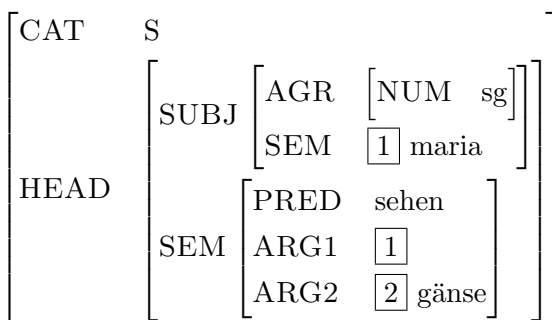


Die zu den Nichtterminalknoten gehörenden Merkmalstrukturen sind:

sieht Gänse:



Maria sieht Gänse:



11. Erklären Sie den Unterschied zwischen Transfer- und Interlingua-Modell in der maschinellen Übersetzung. Nennen Sie jeweils Vor- und Nachteile.

Im Interlingua-Modell wird die Quellsprache zunächst in eine univeselle Repräsentation in der 'Interlingua'-Sprache übersetzt. Diese Repräsentation kann dann in beliebige Zielsprachen übersetzt werden. Der Vorteil des Interlingua Modells ist, dass man für jede Sprache nur je einen Übersetzer von/nach Interlingua braucht, aber nicht zwischen den einzelnen Sprachen. Allerdings muss die Interlingua sehr feingranular sein um alle Unterschiede in allen Sprachen darstellen zu können. Der dadurch entstehende Übersetzungsaufwand lohnt sich insbesondere für sehr ähnliche Sprachpaare nicht.

Im Transfer-Modell wird (meist auf syntaktischer Ebene) direkt zwischen zwei Sprachen übersetzt. Dieses Modell ist normalerweise effizienter, hat aber den Nachteil dass sich Übersetzungssysteme nicht einfach auf neue Sprachen anpassen lassen und dass man $\frac{n(n-1)}{2}$ Übersetzer benötigt um zwischen jedem Sprachpaar übersetzen zu können. Einen Kompromiss zwischen Interlingua- und Transfermodell bilden Transfermodelle auf semantischer Ebene.

12. Geben Sie **jeweils zwei** Beispiele für folgende von WordNet erfasste Relationen. Machen

Sie (wo sinnvoll) deutlich, in welche Richtung die Relation besteht (Sie können dafür deutsche Bezeichnungen verwenden).

a) Hypernymie-Hyponymie

Möbel ist ein Hypernym zu *Tisch*

Guppy ist ein Hyponym zu *Fisch*

b) Holonymie-Meronymie

Tisch ist ein Holonym zu *Bein*

Kiemer ist ein Meronym zu *Fisch*

c) Synonymie

Mensch - *Homo Sapiens*

Ast - *Zweig*

13. Wörter können in einem Vektorraummodell durch die Verteilung benachbarter Wörter repräsentiert werden. Unter <http://clg.wlv.ac.uk/demos/similarity/> ist ein solches Modell online zu testen.

Lassen Sie die nächsten Nachbarn zu *fünf* Wörtern ausgeben (wählen Sie den Cosinus als Ähnlichkeitsmaß). Versuchen Sie, für *jeweils drei*¹ interessante ausgegebene Wörter die Relation zu finden, in der sie zum eingegebenen Wort stehen (z.B. *ist-ein*, *Oberbegriff-von*, *Teil-von*,...). In welchen Fällen ist es schwierig, eine passende Relation zu finden?

¹insgesamt werden also 15 Paare betrachtet