

Lösungsvorschlag für das Übungsblatt 4.

Aufgabe 1.

Im CISLEX sind für das deutsche Kernlexikon 206.000 Lemmata, 1.300.000 Wortformen und 2.350.000 Lesarten kodiert. Wichtiger ist aber die Herangehensweise, um eine solche Aufgabe zu lösen.

aus Ihren Beiträgen:

Das Verhältnis Wortformen : Stämme könnte im Deutschen ungefähr 7:1 sein, auf jeden Fall ist es aber deutlich höher als im Englischen wo größtenteils mittels Präpositionen und Hilfsverben flektiert wird. Geschätzt wurde nach folgender Überlegung:

Nennenswerte Paradigmen werden im Deutschen von Substantiven (geschätzt: im Schnitt 3-4 Wortformen), vor allem aber Adjektiven (ca. 14) und Verben (ebenso ca. 14) gebildet.

Demonstrativ- oder Personalpronomina können zwar dekliniert werden (ca. 5-6 Formen), dürften aber aufgrund ihrer relativ geringen Anzahl nicht ins Gewicht fallen.

Ebenso stellen nicht komparierbare Adjektive (z.B. tot) mit jeweils um die 5 Formen eher die Ausnahme dar.

Adverbien können zwar nicht flektiert werden und könnten statistisch wohl eine Rolle spielen; da im Deutschen aber häufig adverbial gebrauchte Adjektive deren syntaktische Funktion übernehmen, wurden sie in der Schätzung eher schwach gewichtet.

noch ein Beispiel (dort teilweise die Zahl der möglichen Wortformen unterschätzt:

Um das Verhältnis Wortformen : Stämme ungefähr bestimmen zu können, habe ich mir zunächst Gedanken darüber gemacht, wie es bei den einzelnen Wortarten aussieht. Ich gelang zu folgendem Ergebnis:

1) Substantive

- der: 3 Formen

- die: 1 Form

- das: 3 Formen

2) Verben

jeweils 10 Formen

Bsp: lern- {en, e, st, t, te, test, ten, tet, end, ge-lern-t}

3) Adjektive

- insgesamt jeweils 4 Formen

4) Pronomen (mein, dein etc.)

- jeweils 5 Formen

Addiert man die verschiedenen Formen der Wortarten, kommt man auf 26. Dieses Ergebnis muss man dann durch 6 dividieren und man erhält als Ergebnis 4,33. Daraus folgere ich, dass das Verhältnis zwischen Wortformen und ihren Stämmen irgendwo zwischen 4 und 5 sein muss.

Aufgabe 2.

Das komplette Paradigma von "mager":

Steigerung: mager-magerer-am magersten

starke Deklination

	Maskulinum	Femininum	Neutrum	
	<u>Singular Positiv</u>			<u>Plural Positiv</u>
Nominativ	<i>magerer</i>	<i>magere</i>	<i>mageres</i>	<i>magere</i>
Genitiv	<i>mageren</i>	<i>magerer</i>	<i>mageren</i>	<i>magerer</i>
Dativ	<i>magerem</i>	<i>magerer</i>	<i>magerem</i>	<i>mageren</i>
Akkusativ	<i>mageren</i>	<i>magere</i>	<i>mageres</i>	<i>magere</i>

schwache Deklination

Nominativ	<i>der magere</i>	<i>die magere</i>	<i>das magere</i>	<i>die mageren</i>
Genitiv	<i>des mageren</i>	<i>der mageren</i>	<i>des mageren</i>	<i>der mageren</i>
Dativ	<i>dem mageren</i>	<i>der mageren</i>	<i>dem mageren</i>	<i>den mageren</i>
Akkusativ	<i>den mageren</i>	<i>die magere</i>	<i>das magere</i>	<i>die mageren</i>

gemischte Deklination

Nominativ	<i>ein magerer</i>	<i>eine magere</i>	<i>ein mageres</i>
Genitiv	<i>eines mageren</i>	<i>einer mageren</i>	<i>eines mageren</i>
Dativ	<i>einem mageren</i>	<i>einer mageren</i>	<i>einem mageren</i>
Akkusativ	<i>einen mageren</i>	<i>eine magere</i>	<i>ein mageres</i>

starke Deklination

	Maskulinum	Femininum	Neutrum	
	<u>Singular Komparativ</u>			<u>Plural Komparativ</u>
Nominativ	<i>magererer</i>	<i>magerere</i>	<i>magereres</i>	<i>magerere</i>
Genitiv	<i>magereren</i>	<i>magererer</i>	<i>magereren</i>	<i>magererer</i>
Dativ	<i>magererem</i>	<i>magererer</i>	<i>magererem</i>	<i>magereren</i>

Akkusativ	<i>magereren</i>	<i>magerere</i>	<i>magereres</i>	<i>magerere</i>
-----------	------------------	-----------------	------------------	-----------------

schwache Deklination

Nominativ	<i>der magerere</i>	<i>die magerere</i>	<i>das magerere</i>	<i>die magereren</i>
Genitiv	<i>des magereren</i>	<i>der magereren</i>	<i>des magereren</i>	<i>der magereren</i>
Dativ	<i>dem magereren</i>	<i>der magereren</i>	<i>dem magereren</i>	<i>der magereren</i>
Akkusativ	<i>den magereren</i>	<i>die magerere</i>	<i>das magerere</i>	<i>die magereren</i>

gemischte Deklination

Nominativ	<i>ein magererer</i>	<i>eine magerere</i>	<i>ein magereres</i>
Genitiv	<i>eines magereren</i>	<i>einer magereren</i>	<i>eines magereren</i>
Dativ	<i>einem magereren</i>	<i>einer magereren</i>	<i>einem magereren</i>
Akkusativ	<i>einen magereren</i>	<i>eine magerere</i>	<i>ein magereres</i>

Singular Superlativ

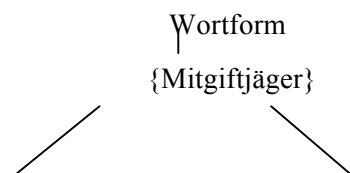
Plural Superlativ

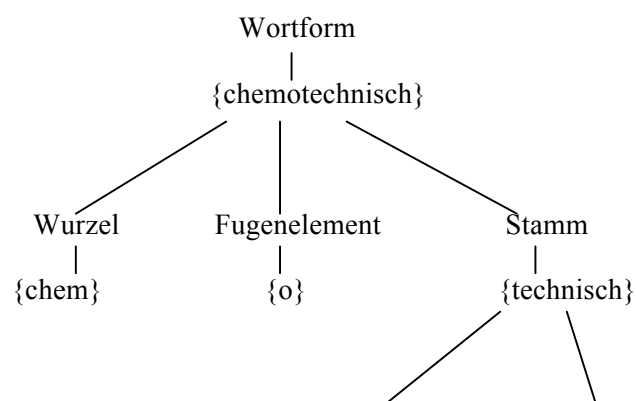
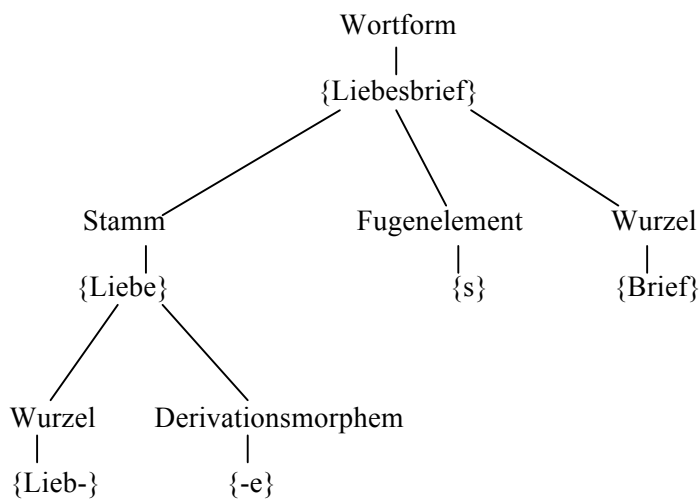
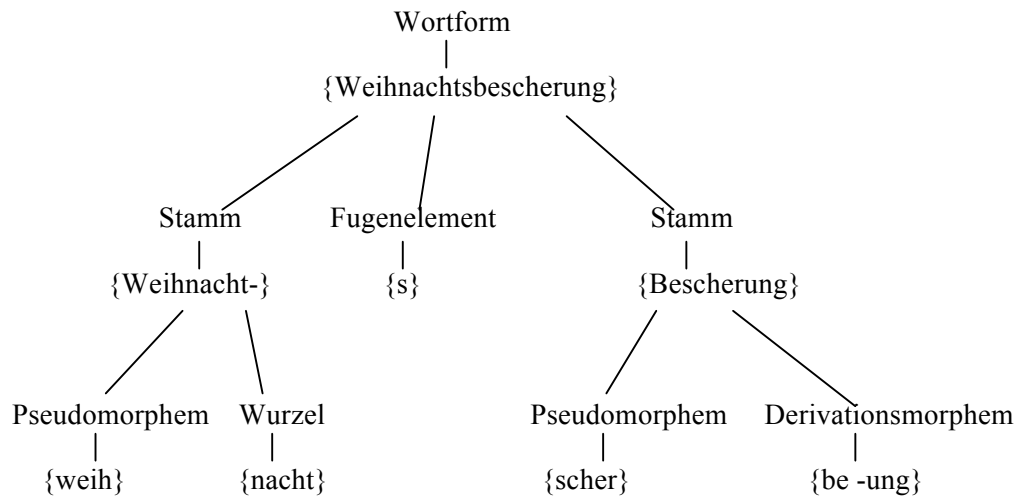
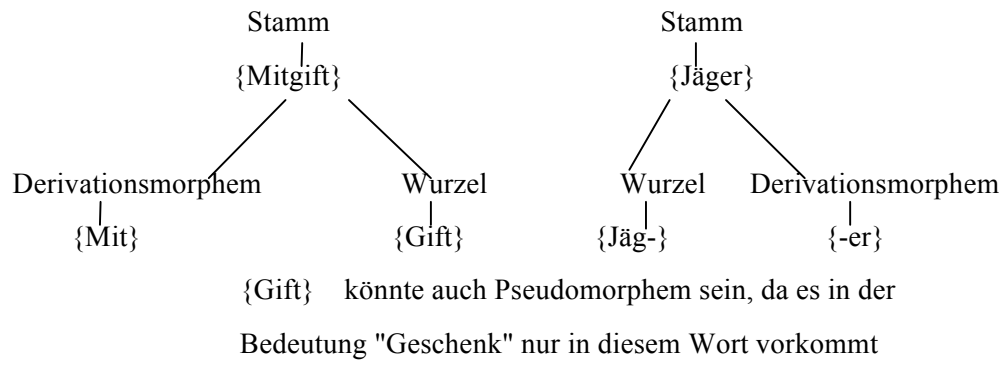
	Maskulinum	Femininum	Neutrum	
Nominativ	<i>der magerste</i>	<i>die magerste</i>	<i>das magerste</i>	<i>die magersten</i>
Genitiv	<i>des magersten</i>	<i>der magersten</i>	<i>des magersten</i>	<i>der magersten</i>
Dativ	<i>dem magersten</i>	<i>der magersten</i>	<i>dem magersten</i>	<i>der magersten</i>
Akkusativ	<i>den magersten</i>	<i>die magerste</i>	<i>das magerste</i>	<i>die magersten</i>

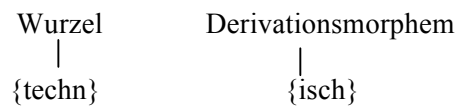
Pronomen "mein"

Nominativ	<i>mein</i>	<i>meine</i>	<i>mein</i>	<i>meine</i>
Genitiv	<i>meines</i>	<i>meiner</i>	<i>meines</i>	<i>meiner</i>
Dativ	<i>meinem</i>	<i>meiner</i>	<i>meinem</i>	<i>meinen</i>
Akkusativ	<i>meinen</i>	<i>meine</i>	<i>mein</i>	<i>meine</i>

Aufgabe 3.







Aufgabe 4.

aus Ihren Beiträgen:

Bei der maschinellen Sprachverarbeitung ist es wichtig sich das Prinzip der Beziehung zwischen Syntax eines Wortes oder einer Kombination von Wörtern und Semantik bewusst zu sein. So entsteht ein Prozess, in dem eine Äußerung und eine Antwort erfolgt, die in einem bestimmten Kontext stehen müssen. Darauf folgt, die Darstellung des Textes als Satz, der in einer logischen Form wiedergegeben wird und eine sinnvolle und vor allem bekannte Zeichenfolge berücksichtigt. Die Äußerung muss entsprechend verstanden werden, so dass eine logische Antwort im Rückzug mit Hilfe der gleichen Variablen der Darstellung des Satzes, erfolgen kann. Letztendlich erfolgen weitere Prozesse zur Ausführung des Textes und Erleichterung der Variablenersetzung.

Die einzelnen Teilgebiete der Computerlinguistik spielen gerade in der maschinellen Sprachverarbeitung eine wichtige Rolle. Dazu zählen nicht nur die Theorien der Syntax und der Semantik, welche in der Linguistik eine Anwendung finden, sondern auch die mathematische Logik und deren logische Formeln. So können die Resultate zu bestimmten Antworten auf gestellte Äußerungen besser erzielt werden und gewisse Antwortmöglichkeiten unterliegen einer Prüfung, so dass sie in einer semantisch und syntaktisch korrekten Form wiedergegeben werden können.

Eine Veränderung scheint mir inzwischen in der Nutzung von Systemen, die nicht mehr auf genaue Syntax und Semantik in der Wiedergabe bestimmter Äußerungen Wert legen. So kann das System bzw. die Maschine auch kleine Fehler bei der Eingabe erkennen und dennoch richtig ausführen.

Weitere Anmerkungen:

Entscheidend ist für den Author, dass für ein isoliertes Problem sowohl die Formulierung einer Theorie als auch deren praktische Implementantation erfolgt vgl. AI Zitat.

Die nach wie vor offene Realisation der Beziehung von Syntax und Semantik muss unbedingt durch pragmatische Eigenschaften ergänzt werden um eine Äußerung geeignet zu interpretieren.

Wenn man das process Prädikat und dessen Auflösung ansieht gibt es allerdings auch fast 20 Jahre nach Erscheinen des Artikels keine Theorie, die das represent-Prädikat implementierbar macht.

Wenn man sich fragt, was sich geändert hat: für mehrere Jahre hatte sich vor allem die Reichweite des CL Programms geändert. Der Mainstream hatte sich davon entfernt etwas so Umfassendes, wie das im Artikel Dargestellte zu versuchen. Nach wie vor ist klar, dass nur stark parallele Ansätze eine volle Repräsentation von logischer Form und pragmatischem Kontext wenn sie denn verfügbar wäre, auch praktisch umsetzen könnten.

Auch im vorliegenden Artikel wird – trotzdem – die Linguistik als einzig möglicher Lieferant von Basistheorien genannt – vgl. hierzu auch jetzt wieder, Shuly Winters Artikel.

Ziele wie: automatische Übersetzung, natürlichsprachliche Datenbankabfrage Typ I Anwendungen im Artikel sind jedenfalls z. T. erreicht worden.

Die Extraktion von Wissensbasen aus Daten natürlicher Sprache, z.B. aus dem www wird im Moment vorangetrieben. Vgl. auch Prof. Günthners Veranstaltung Corpus Calculus und Arbeiten zur Faktenextraktion.

Prolog; Natürlich hat sich hier der Hype gelegt. Ein Grund ist sicherlich, dass Prolog Programme sehr schwer zu übersehen sind und dass es kaum möglich ist zu modularisieren, d.h. Programme mit vielen durchschnittlichen Programmieren zu erstellen. Dies frisst den Vorteil auch komplexe Dinge „einfach hinschreiben“ zu können in fast allen Fällen wieder auf. Trotzdem hat die Sprache unserer Meinung nach als Prototypensprache für die Implementation linguistischer Probleme nach wie vor große Bedeutung.

Aufgabe 5.

(b) $L = \{ab^n c \mid n \geq 0\}$.

Der Automat fängt mit "a" an, gefolgt von "b"s, die beliebig oft (auch 0 Mal) wiederholt werden können, und endet mit "c".

(c) Im Programm muss folgende Zeile hinzugefügt werden:

transition(q0,a,q0).

Aufruf nach dem Einfügen:

?- accept([a,b,b,c]). %Eingabestring als Liste in Eckigen Klammern, kommagetrennt

true.

?- accept([a,a,b,b,c]).

true.