

Einführung in die Computerlinguistik

Syntaktische Funktionen & Abhängigkeiten

Hinrich Schütze

Center for Information and Language Processing

2018-10-29

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth erstellt. Fehler und Mängel sind ausschließlich meine Verantwortung.

- 1 Intro
- 2 Syntaktische Funktionen
- 3 Dependenzgrammatik

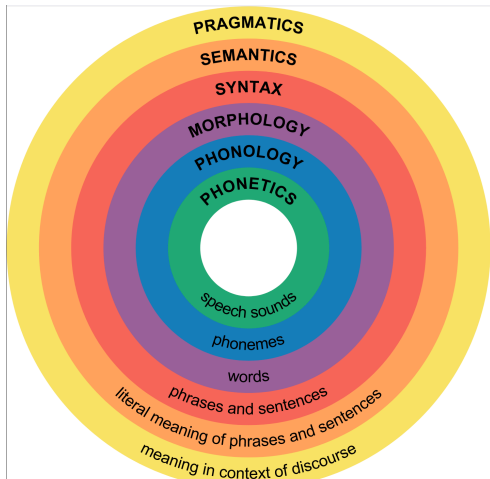
- 1 Intro
- 2 Syntaktische Funktionen
- 3 Dependenzgrammatik

- griech.: *sýn* (zusammen) + *táxis* (Ordnung)
- wörtlich:
Anordnung, Regelung, Organisation, Verhältnis
- (der Wörter in einem Satz)

- **Wortstellung**, “Regeln” der Wortstellung
- **Zusammensetzungen von Wörtern**:
Wortverbindungen, Syntagmen, Phrasen
- **Rekursiv**:
Zusammensetzungen von Phrasen zu größeren Phrasen
- Der **Satz** als Zusammensetzung von Phrasen
- **Funktion** der Wortarten und Phrasenarten beim Aufbau von größeren Phrasen und beim Aufbau des Satzes
- **Wechselseitige Beziehungen** der Glieder des Satzes und ihre Beziehung zum Satzganzen

- 5-Teilung: Phonetik/Phonologie – Morphologie – Syntax – Semantik – Pragmatik
- Interaktionen zwischen Phonetik/Phonologie und Syntax: “Gestern hat Santa Barbara Miller zum Stadtrat gewählt.”
- Syntax erfordert bestimmte Intonation: “Musst Du gehen?”
- Interaktionen zwischen Semantik und Syntax: Ein Satz hat oft viele mögliche syntaktische Analysen, aber nur wenige sind semantisch sinnvoll.
- “The post office will hold out discounts and service concessions as incentives.” (nächste Übung)

Levels of Language



- (linguistisch/theoretisch) Die Syntax beschreibt welche Sätze **syntaktisch wohlgeformt** (möglich) sind.
- (computerlinguistisch/praktisch) Die syntaktische Beschreibung eines Satzes ist das Gerüst für seine **semantische** Interpretation.

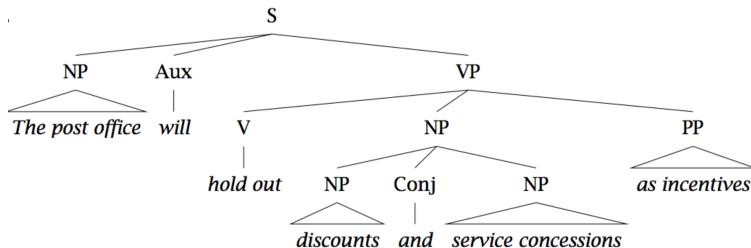
- Linguistisch/theoretisch
 - Welche sprachlichen Konstrukte sind wohlgeformt?
 - Welche Analysen sind kognitiv motiviert?
 - Welche Beschreibungsmechanismen sind auf alle Sprachen anwendbar?
 - ⇒ Von besonderem Interesse: **Grenz- und Sonderfälle**
- Computerlinguistisch/praktisch
 - Welche linguistisch motivierte Repräsentation beschreibt die wesentlichen Zusammenhänge einer sprachlichen Äußerung (wohlgeformt oder nicht) zur weiteren Verarbeitung?
 - Tweets, Tippfehler, Nicht-Muttersprachler, Korrekturen in gesprochener Sprache, ...
 - ⇒ Von besonderem Interesse: **Robustheit, Abdeckung der häufigsten Fälle, Konsistenz**

Übung

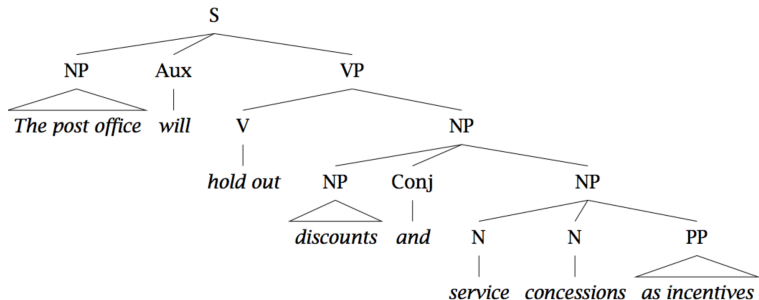
“The post office will hold out discounts and service concessions as incentives.”

Semantisch sinnvolle vs. semantische nichtsinnvolle syntaktische Analysen.

Semantically plausible reading

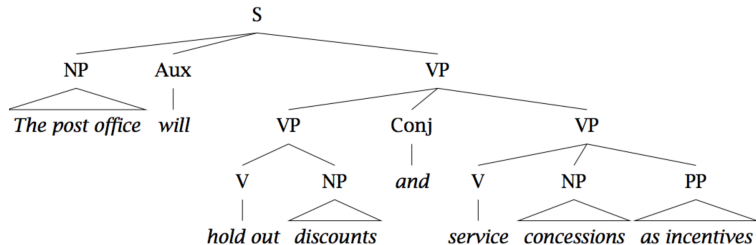


Implausible reading



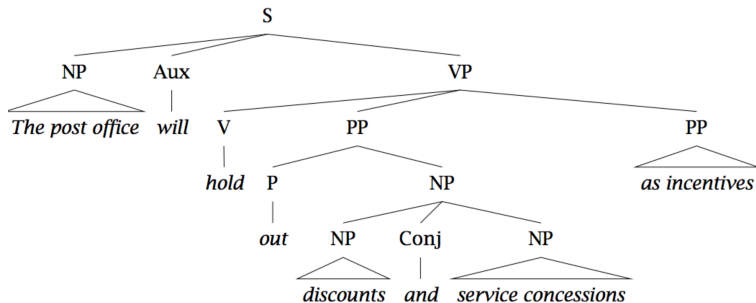
(discounts are not incentives)

Implausible reading



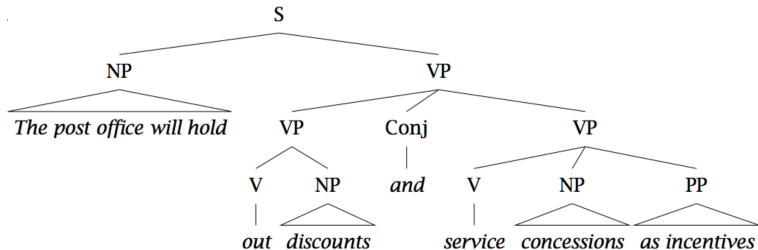
(the post office does not service concessions)

Implausible reading



("out discounts" is interpreted in analogy to "out the window")

Implausible reading



(“out” is interpreted as a verb)

- 1 Intro
- 2 Syntaktische Funktionen
- 3 Dependenzgrammatik

Grammatische Relation zwischen zwei Ausdrücken

- bestimmt durch die morphologische Markierung
- und / oder die strukturelle Relation der Ausdrücke zueinander

Funktionen sind durch satzsyntaktische syntagmatische Relationen zwischen Wörtern / Phrasen definiert.

- Subjekt_von_X
- Objekt_von_X
- Prädikativ_zu_X
- Attribut_von_X
- Adverbiale_von_X

Funktion des (Haupt-)Verbes: **Prädikat**

- Das Hauptverb – der bestimmende Satzteil, der eine Aussage über das Subjekt macht
- “Der Bauer **pflügt** den Acker.”
- “Mich **wundert**, dass das **funktioniert**.”

- Wer? Was?
 - Kasus: Nominativ
 - Kongruenz mit dem finiten Verb
 - Realisierung durch verschiedene Kategorien:

Der Kater lässt das Mäusen nicht.

Er wittert Gefahr.

Dass das nicht funktioniert, wundert mich nicht.

Es überrascht niemanden, dass er schweigt.

Syntaktische Funktion: Subjekt

- In einer Nominalisierung entspricht dem Subjekt normalerweise das Possessivum:
“Paul reiste nach Rom.”
“Pauls Reis nach Rom”
- Im Deutschen muss immer ein syntaktisches Subjekt realisiert werden, auch wenn es semantisch leer ist.
- Dies geschieht z.B. durch das **Wetter-es**, ein sog. **Expletivum**.
“**Es** regnet.”
“**Es** graut mir vor der Prüfung”

- Verbergänzungen im Akkusativ, Dativ oder Genitiv:
*Peter isst **einen Apfel*** (Akkusativobjekt)
*Peter hilft **seinem Freund*** (Dativobjekt)
*Peter gedachte **seiner Mutter*** (Genitivobjekt)
- Objektsatz: vom Verb geforderte Satzwertige Ergänzung:
*Ich habe versprochen, **dass ich mich beeile.***

- Vom Verb geforderte obligatorische Ergänzung, die mit einer semantisch leeren Präposition realisiert wird:
Peter denkt nur an sich.
- Wird die Semantik der Präposition beibehalten, spricht man dagegen von einer adverbialen Ergänzung.
Peter arbeitet/schläft/tanzt im Garten.
- Obligatorische adverbiale Ergänzungen werden manchmal auch zu den Präpositionalobjekten gezählt:
Peter wohnt in Hamburg

Syntaktische Funktion: Prädikativ

- Prädikative ordnen Satzgliedern Eigenschaften zu.
- Subjektprädikativ bei Kopulaverben:
 - Kerstin ist **Informatikerin**.
 - Anna wird **reich und glücklich**.
- Objektprädikative bei Verben wie:
finden, nennen, heißen, schimpfen
 - Sie fand das Buch **recht teuer**.
 - Sie hieß ihn **einen Versager**.
 - Er nannte sie **eine Lügnerin**.
- **Prädikativ** nicht mit **Prädikat** (Verb oder Verbphrase) verwechseln.

- Bei Adverbialen wird die Hauptunterscheidung semantisch getroffen. In einem Satz können verschiedene Kategorien vertreten sein:

Sie liegt vor lauter Langeweile (kausal) den ganzen Tag (temporal) dösend (Art und Weise) im Bett (lokal).

- Je nach Valenz des Verbs unterscheidet man:
 - ① **Obligatorische** adverbiale Ergänzungen sind in der Valenz des Verbs fest angelegt.
*Ich fühle mich **gut**.*
*Hans wohnt **in Berlin**.*
 - ② **Fakultative** adverbiale Ergänzungen sind nur bei bestimmten Verben möglich, wenn auch nicht notwendig.
*Anna schwimmt **schnell***
 - ③ **Freie Angaben** können zu allen Verben ohne Beschränkungen hinzutreten.
Er arbeitet (am Wochenende)(gern)(in aller Ruhe)(im Garten).

- Adverbiale Ergänzungen sind typischerweise Adjektive (Adverbien) oder Präpositionalphrasen.
- Die adverbiale Funktion kann auch durch Sätze wahrgenommen werden:
 - “Lena spielt, **während Mama arbeitet.**” (Temporalsatz)
 - “Unglückliche Menschen, **wohin man schaut.**” (Lokalsatz)

Syntaktische Funktion: Attribut

- Beifügungen zur besonderen Bestimmung eines Substantivs (bzw. Nominalphrase)
- Können nur in Abhängigkeit dieses Substantivs im Satz auftreten.
- Syntaktisch können sie entweder als Teil der näher bestimmten Nominalphrase auftreten, oder als eigenes Satzglied.
- Als **Satzgliedteil** nur zusammen mit dem Bezugselement verschiebbar:

Er beantwortet [den Brief [des Freundes]] heute.

** [Des Freundes] beantwortet er [den Brief] heute.*

- Attribut als **umstellbares Satzglied**:

*Sie trinkt den Tee **mit Milch**.*

*Den Tee trinkt sie **mit Milch**.*

***Mit Milch** trinkt sie den Tee.*

Syntaktische Funktion: Attribut

- Adjektivattribut:
ein neues Buch
- Partizipialattribut:
schlafende Hunde
- Präpositionalattribut:
Der Mann vom Mond
- Genitivattribut:
die Zerstörung der Welt
- Adverbattribut:
Der Unterricht gestern war langweilig
- Appositionen:
Heiner, der Chef der Firma
- Attributsätze:
die Dänen, die Bier trinken, ...
die Hoffnung, dass alles gut wird

Beispiel: Analyse eines Satzes Anhand von Wortarten und Syntaktischen Funktionen

	Opa	erzählte	gestern	eine	lange	Geschichte
Wortart	Nomen	Verb	Adverb	Artikel	Adjektiv	Nomen
Funktion	Subjekt	Prädikat	Temporal- adver- biale		Adj.attribut ← - - - Objekt - - - →	

Übung

“Im Deutschen muss immer ein syntaktisches Subjekt realisiert werden, auch wenn es semantisch leer ist.”

Denken Sie sich einen Beispielsatz aus, auf den dies **nicht** zutrifft, d.h. einen (relativ) normalen Satz des (gesprochenen?) Deutschen, der kein Subjekt hat.

Einzelwörter (“Ja.”, “Danke.”) und Einzelphrasen (“Wieso denn das?”, “Voll gut!”) zählen nicht.

- 1 Intro
- 2 Syntaktische Funktionen
- 3** **Dependenzgrammatik**

- Wir haben gesehen, welche syntaktischen Funktionen zwischen Satzgliedern existieren können.
- Wie kann eine syntaktische Beschreibung eines Satzes aussehen, die
 - linguistisch motiviert ist, und gleichzeitig
 - formal genug ist, um von einem Computer verarbeitet werden zu können?

- Beschreibt die Abhängigkeiten im Satz anhand **binärer Relationen zwischen Wörtern**.
- Grundprinzip bei der Verbindung von Wortpaaren:
Ein Wort ist der **Kopf** (engl. **head**), das andere Wort ist das **Dependent**.

- Kopf verlangt Dependent als **obligatorische** Ergänzung (Argument):
“Ich trinke Wasser”
trinke[Kopf]-ich[Dep.], trinke[Kopf]-Wasser[Dep.]
- Dependent modifiziert Kopf als **optionale** Ergänzung:
“Ich trinke gerne viel Wasser”
trinke[Kopf]-gerne[Dep.], Wasser[Kopf]-viel[Dep.]

- In syntaktischen Konstruktionen aus Hilfswörtern (Partikel, Hilfsverben, ...) und Inhaltswörtern (Vollverben, Nomen, ...) ist das Hilfswort Dependent zum Inhaltswort.

“Ich habe getrunken”

getrunken[Kopf]-habe[Dep.]

- Determinierer ist ein Dependent zum Kopf:

“Das Llama trinkt Wasser”

Llama[Kopf]-das[Dep.]

- Kopf und Dependent bestehen jeweils immer nur aus einem Wort.
- Wenn ein komplettes Satzglied (Phrase aus mehreren Wörtern) Argument oder Modifikator ist:
 - Bestimme zunächst Kopf innerhalb des Satzgliedes.
 - Der phraseninterne Kopf ist dann Dependent zum einbettenden (modifizierten) Wort.
- Das Verb nimmt die zentrale Position im Satz ein.

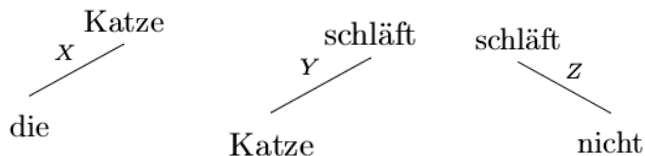
Dependenzanalyse (Beispiel): “Die Katze schläft nicht.”

- Kopf und Dependent bestehen jeweils immer nur aus einem Wort.
- Wenn ein komplettes Satzglied (Phrase aus mehreren Wörtern) Argument oder Modifikator ist:
 - Bestimme zunächst Kopf innerhalb des Satzgliedes.
 - Der phraseninterne Kopf ist dann Dependent zum einbettenden (modifizierten) Wort.
- schläft[Kopf]-Katze[Dep.]
- Katze[Kopf]-die[Dep.]

Dependenzanalyse (Beispiel): “Die Katze schläft nicht.”

- Das Verb nimmt die zentrale Position im Satz ein.
- root[Kopf]-schläft[Dep.]

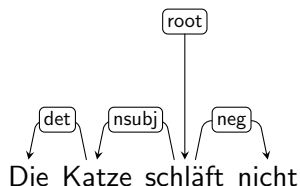
Dependenzanalyse (Beispiel): "Die Katze schläft nicht."



Benennung der Abhängigkeitsrelationen

- Abhängigkeitsrelationen werden entweder durch die **Wortart** des Dependents oder seine **syntaktische Funktion** gekennzeichnet.
- Siehe z.B. **Stanford Universal Dependencies**:
<http://universaldependencies.org/>
- **Wortart oder phrasale Kategorie** z.B. wenn Dependent: Nominal (n), Verb/Satzeinbettung (c – **clause**), Adjektiv (a), Adverb (adv), Artikel/Determinierer (det), Negation (neg), Hilfsverb (aux), Partikel/Konjunktion (mark)
- **Syntaktische Funktion** z.B. wenn Dependent: Subjekt (subj), Objekt (obj), Modifikator/Adverbiale/Attribut (mod), Prädikativ (pred), Apposition (appos)
- Kategorisierungen kombinierbar: “Kalt duschen erfrischt.”
erfrischt[Kopf]–*duschen*[csubj] (**clausal subject**)

- Die Kanten enthalten die Namen der Dependenzrelationen.
- Pfeile werden immer vom Kopf zum Dependenten gezeichnet.
- Das nicht-abhängige Verb des Satzes wird als Wurzel des Dependenzbaums (**root**) gekennzeichnet.



Präpositionalobjekte: Zwei Konventionen

- ① **Rein Syntaktisch.** Die Präposition ist das Dependent zum Verb, und das Nomen ist Dependent zur Präposition.
 - Analyse gemäß der ursprünglichen Variante der Stanford Dependencies. Widerspricht dem Prinzip “Primacy of content”.
 - Wird von vielen automatischen Analyse-Tools so ausgegeben.
- ② **Teils semantisch.** Das Nomen der Präpositionalphrase ist das direkte Dependent zum Verb, und die Präposition fungiert als Kasus-Markierer (**case**).
 - Analyse nach neueren Stanford **Universal** Dependencies.
 - oblique (**obl**): Argumente mit Präposition
 - Es wird keine Unterscheidung zwischen Argumenten und Adjunkten, getroffen, da der Unterschied oft fließend ist.

Dependenzformalismen: Rein syntaktisch vs teils semantisch

- **Rein syntaktisch** motivierte Formalismen.
Entscheidungskriterien sind typischerweise:
 - Welches Wort bestimmt die möglichen Kontexte in denen die Verbindung aus Kopf und Dependent auftreten kann?
 - Zwischen welchen Wörtern besteht Kongruenz in bestimmten syntaktischen Merkmalen?
 - Welches Wort fordert bestimmte syntaktischen Eigenschaften eines anderen Wortes (z.B. Kasus)?
- **Teils semantisch** motivierte Formalismen.
Entscheidungskriterien sind typischerweise:
 - Welches sind die Inhaltswörter und ihre Argumente?
 - Welche Elemente haben nur syntaktische Funktionen?

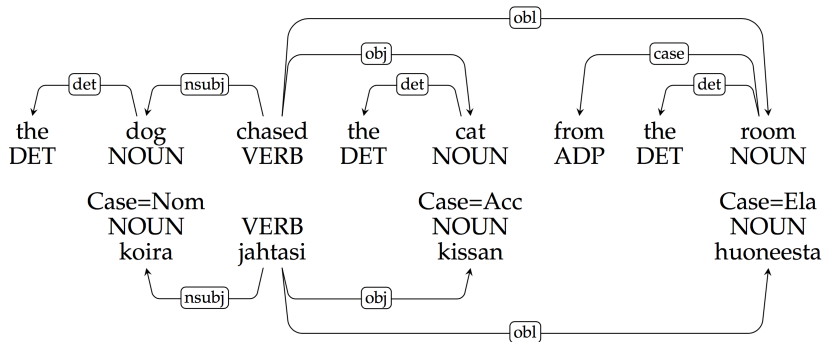
Dependenzformalismen: Rein syntaktisch vs teils semantisch

- Unterschiede bei der Analyse von:
 - Konstruktionen mit Hilfsverben
 - eingebetteten Nebensätzen
 - Präpositionen
- Weitere Argumente für verschiedene Analysen:
 - Argumentstruktur / Semantische Rollen (z.B. bei Hilfsverben, Präpositionalobjekte)
 - Phrasenstatus (z.B. X-bar Theorie)
 - Parallelität zu morphologischer Markierung in anderen Sprachen (Präpositionen)
- Wichtig: Analysen innerhalb des Formalismus konsistent
- Nützlich ist auch eine Darstellung, die eine semantische Verarbeitung erleichtert.
- In dieser VL: [Stanford Dependency Formalismus](#), der als Köpfe vorrangig die [Inhaltswörter](#) wählt.

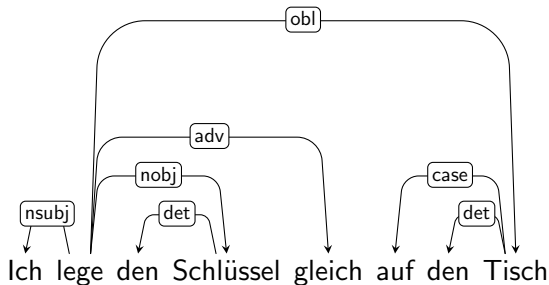
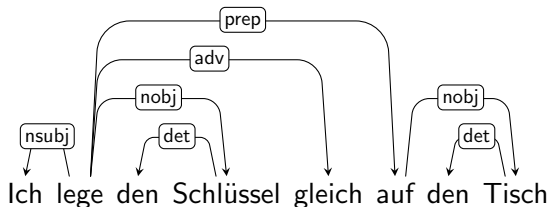
Primacy of Content Words (UD guidelines)

- Dependency relations hold primarily between content words, rather than being indirect relations mediated by function words.
- Preferring content words as heads maximizes parallelism between languages because content words vary less than function words between languages.
- ... one commonly finds the same grammatical relation being expressed by morphology in some languages or constructions and by function words in other languages ... , while some languages may not mark the information at all (e.g., tense, definiteness).
- ... view the relations between content words and function words [...] as operations that modify the grammatical category of the content word so that it can participate in different dependency relations with other content words.

Finnishization of English

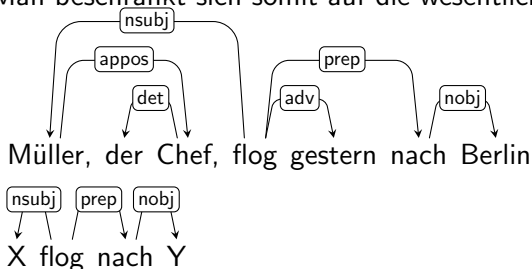


Übung: Primacy of content words: Vor-/Nachteile?



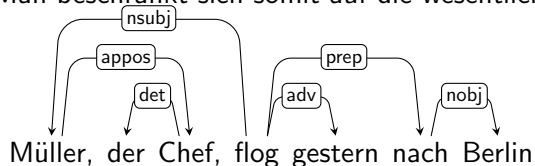
Praktische Verwendung von Dependenzbäumen

- Dependenzstrukturen spielen eine wichtige Rolle bei der automatischen Informationsextraktion. Vorgehen:
 - 1 Eine automatische Dependenz-Analyse durchführen (**dependency parsing**).
 - 2 Den **Pfad** (kürzeste Verbindung) zwischen zwei Entitäten ermittelt.
 - 3 Die Wörter und Dependenzrelationen auf dem Pfad mit einer Liste an **bekannten Mustern** vergleichen
- Man beschränkt sich somit auf die wesentliche Information:



Praktische Verwendung von Dependenzbäumen

Man beschränkt sich somit auf die wesentliche Information:



⇒ `travelled_to(Müller, Berlin)`

- Software zur automatischen Dependenz-Analyse: Dependenz-Parser
 - <http://demo.ark.cs.cmu.edu/parse>
 - <https://nlp.stanford.edu/software/lex-parser.shtml>
 - Demo: <http://corenlp.run/>
- Dependenz-Parsern liegt meist ein statistisches Modell zugrunde.
- Die Dependenzrelationen werden so gewählt, dass die Wahrscheinlichkeit der Analyse für den Satz maximiert wird. (Kombinatorisches Problem!)
- **Fehleranfällig:** Typischerweise 90% oder weniger Genauigkeit (pro Kante).
- Automatische Dependenz-Analysen sind oft der Ausgangspunkt für weitere regelbasierte oder statistische Verfahren.

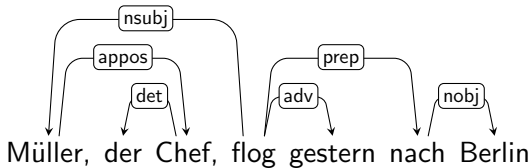
Dependenz-Parsing: Demo

- Die Dependenzanalyse zeigt die Beziehungen der Wörter in einem Satz auf
- Wenn man ein automatisches Dependenz-Analyse-Tool verwenden will, muss man
 - wissen, was die Grundzüge des unterliegenden Formalismus sind
 - beachten, dass automatische Analysen immer fehleranfällig sind
- Die von uns betrachtete Sichtweise ist an die Stanford Universal Dependencies angelehnt
 - Grundlage vieler computerlinguistischer Analyse-Tools
 - Inhaltswörter stehen im Zentrum: bessere Generalisierung zwischen verschiedenen Sprachen und Konstruktionen
 - weicht teils von anderen linguistischen Theorien ab

Übung

Erstellen Sie den Dependenzgraphen für:

“Im Juli schenkte die Großmutter Peter, dem glücklichen, gleich zweimal 100 Euro.”



- Interaktionen von Syntax mit Phonologie, Semantik
- Syntaktische Funktionen:
Prädikat, Subjekt, Objekt, Prädikativ, Adverbiale, Attribut
- Abhängigkeitsrelationen: Namen von Knoten & Kanten
- Abhängigkeitsbaumkonstruktion
- Sätze als Abhängigkeiten (Subjektsatz etc.)
- Zwei Hauptkonventionen der Abhängigkeitsgrammatik:
syntaktisch vs. primacy of content words
- Finnishization of English