

Einführung in die Computerlinguistik

Syntaktische Funktionen & Dependenz

Alex Fraser / Robert Zangenfeind

Center for Information and Language Processing

2019-11-04

Die Grundfassung dieses Foliensatzes wurde von Dr. Benjamin Roth unter Zuhilfenahme von Materialien aus Vorlesungen von Prof. Dr. Tania Avgustinova erstellt. Fehler und Mängel sind ausschließlich meine Verantwortung.

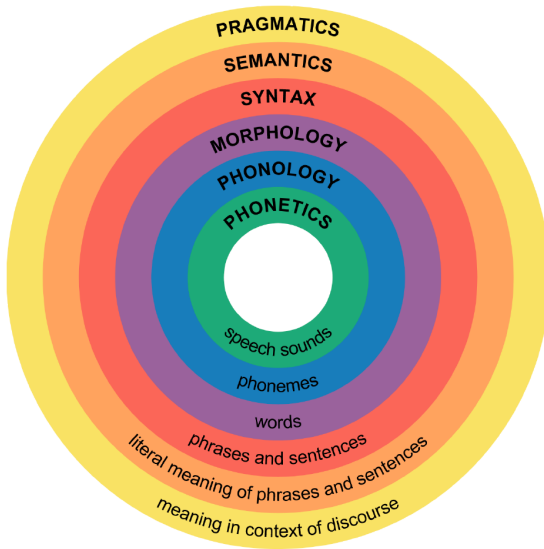
- 1 Intro
- 2 Syntaktische Funktionen
- 3 Dependenzsyntax

- griech.: *sýn* (zusammen) + *táxis* (Ordnung)
- \Rightarrow 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: *Du musst 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.* (s.u.)

Levels of Language



Syntax: Zwei Sichtweisen (1)

- (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.

Syntax: Zwei Sichtweisen (2)

- 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**

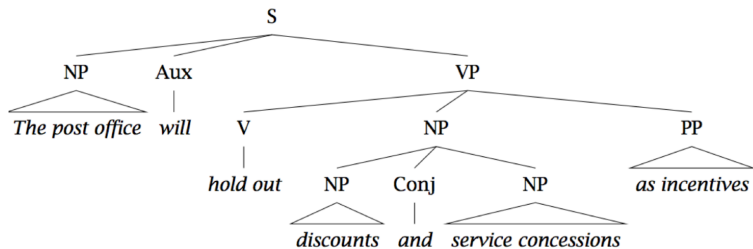
Übung

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

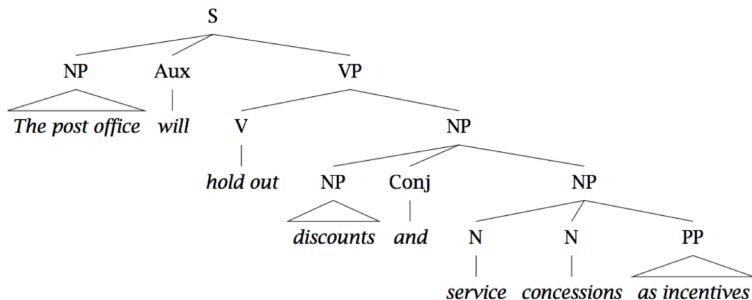
(Bsp. aus Manning, Schütze: Foundations of statistical natural language processing. Cambridge, Mass. u.a. 2001:409.)

Semantisch sinnvolle vs. semantisch nichtsinnvolle syntaktische Analysen:

Semantically plausible reading

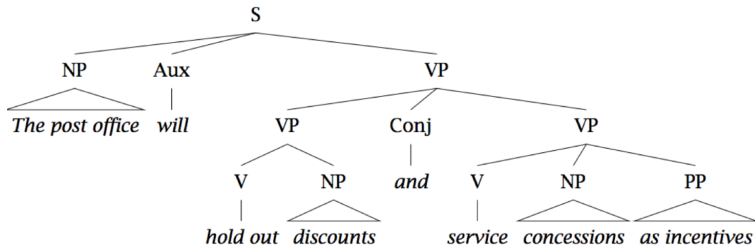


Implausible reading (1)



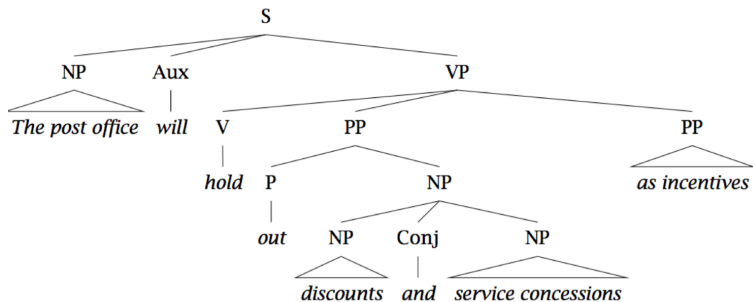
(discounts are not incentives)

Implausible reading (2)



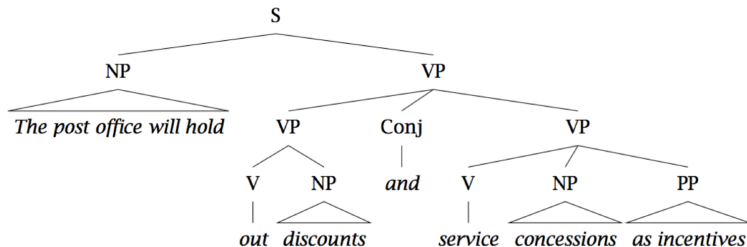
(the post office services concessions)

Implausible reading (3)



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

Implausible reading (4)



("out" is interpreted as a verb)

Grammatikalische Relation zwischen zwei Ausdrücken

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

Funktionen sind durch syntaktische 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.*

Syntaktische Funktion: Subjekt (1)

- **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 (2)

- Im Deutschen wird – mit wenigen Ausnahmen – ein syntaktisches Subjekt realisiert, auch wenn es semantisch leer ist.
- Dies geschieht z.B. durch das “Wetter-es”, ein sog. **Expletivum**, vgl.:
Es regnet.
Es geht mir gut.
- Bei einer Nominalisierung des Prädikats wird das Subjekt üblicherweise zu einem Genitivattribut:
Paul reiste nach Rom.
Pauls Reise nach Rom.

- 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**.*

- Ergänzung, die einen Aktanten (wesentlichen Mitspieler) des Verbs darstellt, und mit einer semantisch leeren Präposition an das Verb angeschlossen wird:
*Das Weltall besteht **aus vielen Galaxien**.*
- darunter sind auch obligatorische adverbiale Ergänzungen:
*Peter wohnt **in Hamburg***
- vgl. dagegen fakultative adverbiale Ergänzung (Semantik der Präposition wird beibehalten) (s.u.):
*Peter schläft/tanzt **im Garten**.*

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.*
- Achtung: **Prädikativ** nicht mit **Prädikat** (Verb oder Verbphrase) verwechseln!

Syntaktische Funktion: Adverbiale (1)

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

Sie döste den ganzen Nachmittag (temporal) vor lauter Langeweile (kausal) antriebslos (Art und Weise) auf dem Sofa (lokal).

Syntaktische Funktion: Adverbiale (2)

- Je nach Valenz des Verbs können unterschieden werden:
 - ① **Obligatorische** adverbiale Ergänzungen sind in der Valenz des Verbs fest angelegt.
*Ich fühle mich **gut**.*
 - ② **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 realisiert werden:

Lena spielt, während Mama arbeitet. (Temporalsatz)

Unglückliche Menschen, wohin man schaut. (Lokalsatz)

Syntaktische Funktion: Attribut (1)

- Beifügungen zur besonderen Bestimmung eines Substantivs (bzw. einer 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 (2)

- Adjektivattribut:
ein neues Buch
- Partizipialattribut:
schlafende Hunde
- Präpositionalattribut:
Der Mann im Mond
- Genitivattribut:
die Zerstörung der Welt
- Adverbialattribut:
Der Unterricht gestern war interessant.
- Apposition:
Heiner, der Chef der Firma
- Attributsatz:
*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 - - - >	

- 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 vom Computer verarbeitet werden zu können?

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

- (i) Kopf verlangt **obligatorisch** einen oder mehr Dependents als Argument:
Sie besucht Freunde.
besucht [Kopf] – *sie* [Dep.]; *besucht* [Kopf] – *Freunde* [Dep.]
- (ii) Kopf hat einen **optionalen** Dependenten als Modifizierer:
Sie besucht oft alte Freunde.
besucht [Kopf] – *oft* [Dep.], *Freunde* [Kopf] – *alte* [Dep.]

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

Ich habe getrunken.

getrunken [Kopf] – *habe* [Dep.]

- Determinierer ist ein Dependent des Kopfes:

Das Lama trinkt Wasser.

Lama [Kopf] – *das* [Dep.]

- Kopf und Dependent bestehen jeweils unmittelbar immer nur aus einem Wort.
- Wenn eine Phrase aus mehreren Wörtern z.B. als Modifikator (bzw. Argument etc.) auftritt:
 - Bestimme zunächst den Kopf innerhalb der Phrase.
 - Der phraseninterne Kopf ist dann Dependent des modifizierten Wortes.
- Das finite Verb nimmt die zentrale Position im Satz ein.

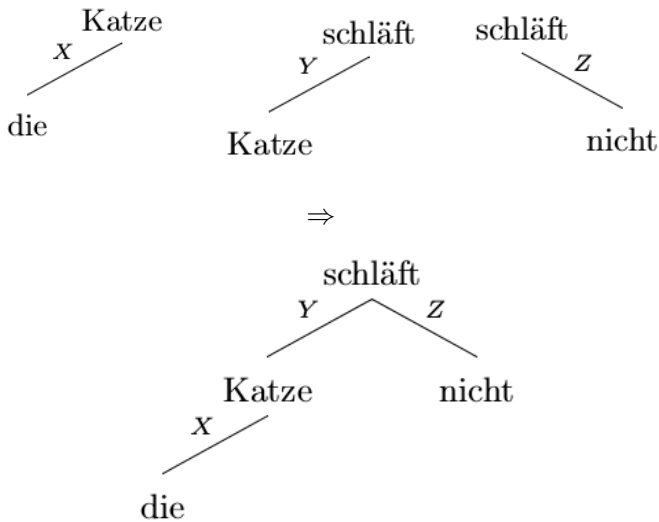
Dependenzanalyse (Beispiel): *Die Katze schläft nicht.* (1)

- Kopf und Dependent bestehen jeweils unmittelbar immer nur aus einem Wort.
- Wenn eine Phrase aus mehreren Wörtern z.B. als Modifikator (bzw. Argument etc.) auftritt:
 - Bestimme zunächst den Kopf innerhalb der Phrase.
 - Der phraseninterne Kopf ist dann Dependent des modifizierten Wortes.
- *Katze* [Kopf] – *die* [Dep.]
- *schläft* [Kopf] – *Katze* [Dep.]

Dependenzanalyse (Beispiel): *Die Katze schläft nicht.* (2)

- Das finite Verb nimmt die zentrale Position im Satz ein.
- root: *schläft*

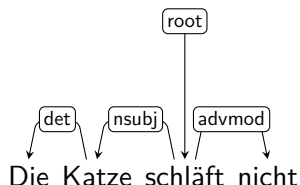
Dependenzanalyse (Beispiel): *Die Katze schläft nicht.* (3)



Benennung der einzelnen Dependenzrelationen

- Zur Benennung von Dependenzrelationen werden z.B. die **Wortart** des Dependents und seine **syntaktische Funktion** herangezogen.
- vgl. z.B. **Stanford Universal Dependencies** (UD), <http://universaldependencies.org/>:
- **Wortart** wird bei UD als Benennung verwendet z.B. bei Dependents der folgenden Wortarten:
 - Artikel/Determinierer (det), Hilfsverb (aux)sie dient als Teil der Benennung z.B. bei diesen Dependents:
 - Nomen (n), Adjektiv (a), Adverb (adv)
- **Syntaktische Funktion** als Benennung z.B. bei diesen Dep.:
 - indirektes Objekt (iobj), Apposition (appos)sie dient als Teil der Benennung z.B. bei diesen Dependents:
 - Subjekt (subj), Modifikator (mod)
- Kombination aus beidem z.B. bei diesen Dependents:
 - nominales Subjekt (nsubj), adverbialer Modifizierer (advmod)

- Kanten (Pfeile) enthalten Namen der Dependenzrelationen.
- Pfeile zeigen vom Kopf zum Dependenten.
- Das nicht-abhängige (finite) Verb des Satzes wird als Wurzel des Dependenzbaums (**root**) gekennzeichnet (bei UD u.a.).



Präpositionalobjekte: Zwei Konventionen

- ① **Rein syntaktisch:** Die Präposition ist der Dependent des Verbs, und das Nomen ist Dependent der Präposition.
 - Analyse gemäß der ursprünglichen Variante der Stanford Dependencies. Widerspricht aber dem Prinzip “Primacy of content”.
 - Wird von vielen anderen automatischen Analyse-Tools so ausgegeben.
- ② **Teils semantisch:** Das Nomen der Präpositionalphrase ist der direkte Dependent des Verbs, und die Präposition fungiert als Kasus-Markierer (*case*).
 - Analyse nach neueren Stanford **Universal** Dependencies.

- **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 morphologischen Merkmalen?
 - Welches Wort fordert bestimmte morphologische 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:

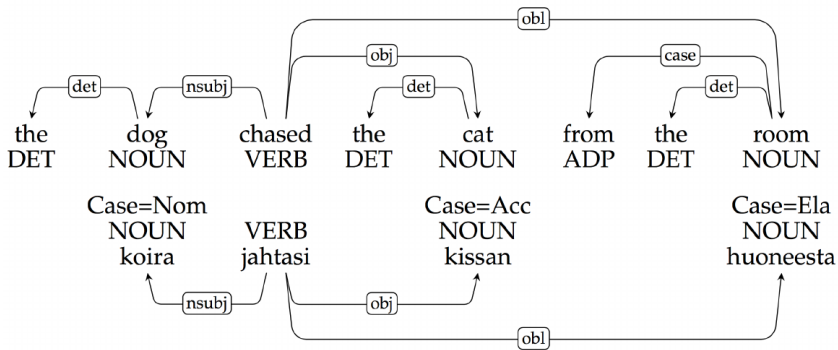
Rein syntaktisch vs. teils semantisch (2)

- Unterschiede bei der Analyse von:
 - Konstruktionen mit Hilfsverben
 - eingebetteten Nebensätzen
 - Präpositionen
- Weitere Argumente für unterschiedliche 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.

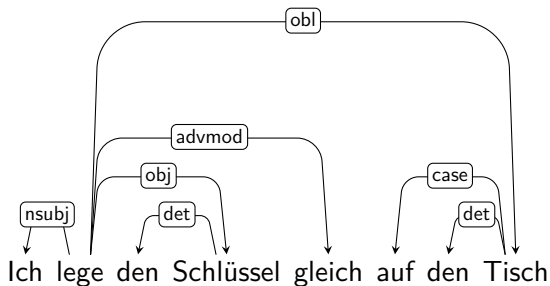
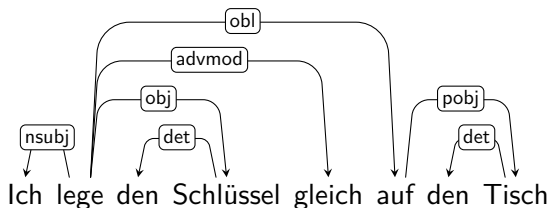
Primacy of Content Words (UD guidelines)

- 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.

Example: Finnishization of English



Primacy of content words: Vor-/Nachteile?



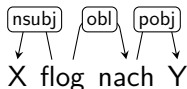
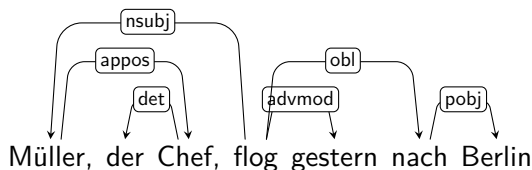
- Automatische Informationsextraktion

Vorgehen:

- ① automatische Dependenz-Analyse durchführen (**dependency parsing**).
- ② **Pfad** (kürzeste Verbindung) zwischen zwei Entitäten ermitteln.
- ③ Wörter und Dependenzrelationen auf dem Pfad mit einer Liste an **bekannten Mustern** vergleichen.

Praktische Verwendung von Dependenzbäumen (2)

Man beschränkt sich somit auf die wesentliche Information
(Achtung: hier syntaktisch motivierte Analyse, d.h keine primacy
of content!):



⇒ travelled_to(Müller, Berlin)

- Software zur automatischen Dependenz-Analyse: Dependenz-Parser (mit etwas Vorsicht zu verwenden)
 - <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.

- 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 zugrundeliegenden 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

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