

Syntax natürlicher Sprachen

1: Einführung

A. Wisiorek

Centrum für Informations- und Sprachverarbeitung,
Ludwig-Maximilians-Universität München

17.10.2023

Themen der heutigen Vorlesung

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

1. Kursübersicht

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

- Die Vorlesung behandelt **Grundbegriffe der Grammatik** und wesentliche syntaktische Konstruktionen des Deutschen im Hinblick auf eine **Verwendung in der maschinellen Sprachverarbeitung**.
- Dazu werden die in neueren Grammatiktheorien verwendeten **Klassifizierungen von Phrasen**, deren **innere Struktur** sowie deren **relationale Abhängigkeiten** erklärt.
- Außerdem werden in der Computerlinguistik übliche **Grammatikformalismen** (wie Kontextfreie Grammatiken, Unifikationsgrammatiken, PCFGs, Datenbasierte Abhängigkeitsgrammatiken, Partielle Parsingmodelle) besprochen.
- ebenso wie **syntaktische Annotationsstandards** (z.B. Penn Treebank, Universal Dependencies) vorgestellt und verwendet, um typische oder schwierige syntaktische Konstruktionen genau zu beschreiben.

- Kenntnis und Anwendung **funktionaler und struktureller Begriffe der grammatischen Beschreibung**:
 - Kongruenz
 - Rektion
 - Subkategorisierung
 - Valenz
- Kenntnis **Grammatikformalismus** und Anwendung für Analysen natürlichsprachlicher Sätze:
 - CFG (Kontextfreie Grammatik: *Konstituentenstruktur*)
 - FCFG (Feature-based CFG: *Modellierung grammatischer Merkmale*)
 - PCFG (Probabilistische CFG: *Gewichtung von CFG-Regeln*)
 - DG (Dependency Grammar: *Dependenzstruktur*)
 - **Chunk-Parser** (u.a. RegExParser: *partielles Parsing, 'Parsing as Tagging'*)
- Kenntnis und Bedienung eines **Programmsystems**, das einen Grammatikformalismus verwendet

2. Einführung

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

2.1. Syntax als linguistische Disziplin

- 1 Kursübersicht
- 2 Einführung
 - **Syntax als linguistische Disziplin**
 - Syntaktische Ausdrucksmittel
 - Konstituenz- und Dependenzrelation
 - Repräsentation syntaktischer Strukturen
 - Parsing als Syntaxanalyse
- 3 Analyse syntaktischer Einheiten
- 4 Konstituententests
- 5 Komplement/Adjunkt-Tests

Syntax (von gr. *syntaksis*, 'Zusammensetzung')

Lexikon der Sprachwissenschaft (Bußmann):

„**Teilbereich der Grammatik natürlicher Sprachen** (auch: Satzlehre).“

„**System von Regeln**, die beschreiben wie **aus einem Inventar von Grundelementen** (Morphemen, Wörtern, Satzgliedern) **durch spezifische syntaktische Mittel** (Morphologische Markierung, Wort- und Satzgliedstellung, Intonation u.a.) **alle wohlgeformten Sätze einer Sprache abgeleitet** werden können.“

<https://mediensprache.net/de/lexikon:>

„**Teilgebiet der Linguistik**, das sich mit der **Kombination von Wörtern zu komplexen Einheiten** (Analyse des Aufbaus von Satzstrukturen und der Zusammenfügung von Wörtern zu größeren Einheiten) beschäftigt, ohne sich für den internen strukturellen Aufbau der Wörter zu interessieren.“

„Der Begriff kann auch benutzt werden, um **den strukturellen Aufbau eines Satzes zu bezeichnen** ('Syntax eines Satzes' und so weiter).“

Syntax als Satzstruktur-Analyse

Allgemeiner Syntaxbegriff der Semiotik

- Syntax als **Struktur einer Zeichenfolge**
→ *Regeln der Kombination elementarer Zeichen zu komplexen Zeichen*

Syntax natürlicher Sprachen

- Syntax als **Struktur von Wortfolgen**
→ *Regeln der Kombination von Wörtern zu größeren Einheiten*
→ *Satz als zentrale sprachliche Einheit: Sprechen in Äußerungen*

Syntax als linguistische Disziplin

- Teilgebiet der **Grammatik** als Sprachstruktur-Analyse

Grammatik (von gr. (*technē*) *grammatikē*, 'Schreibkunst')

- (Theorie der) **Sprachstruktur**

Abgrenzung Syntax zu anderen Disziplinen

Abgrenzung zur Phonologie und Morphologie

- Syntax: Analyse des *Strukturaufbaus* sprachlicher Einheiten *oberhalb der Wortebene*

Abgrenzung zu Semantik und Pragmatik

- Syntax: unabhängig von semantischer Interpretation (vgl. Chomsky 1957, 'Syntactic Structures')

Chomsky 1957: *colorless green ideas sleep furiously*

- Satz erfüllt **Wohlgeformtheitsbedingungen**, ist also grammatisch
- hat aber keine sinnvolle semantische Interpretation

Wortartenklassifikation

- **Zusammensetzung** syntaktischer Einheiten
aus **Klassen von Wörtern** (POS = lexikalische Kategorien)
→ *lexikalische Regeln*: ADJ → *groß* / *klein*

Flexionsmorphologie

- Analyse von **Wortformen**, insofern sie für die syntaktische Strukturanalyse relevant sind (**Morphosyntax**)
- **Kasus** und **Agreement** als morphologische Ausdrucksmittel syntaktischer Funktionen
- formale Repräsentation als **Merkmalstrukturen**

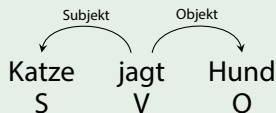
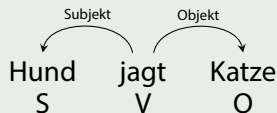
2.2. Syntaktische Ausdrucksmittel

- 1 Kursübersicht
- 2 Einführung
 - Syntax als linguistische Disziplin
 - **Syntaktische Ausdrucksmittel**
 - Konstituenz- und Dependenzrelation
 - Repräsentation syntaktischer Strukturen
 - Parsing als Syntaxanalyse
- 3 Analyse syntaktischer Einheiten
- 4 Konstituententests
- 5 Komplement/Adjunkt-Tests

Wortstellung

- Markierung syntaktischer Funktion durch *lineare Anordnung*

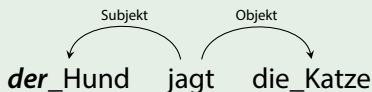
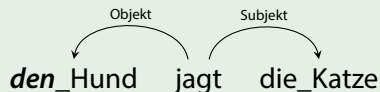
SVO-Wortstellung



Kasus

- *morphologische Markierung* der syntaktischen Funktion am *abhängigen Element*
- Rektion (*dependent-marking*)

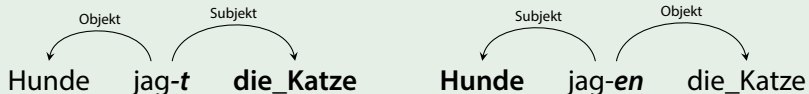
Objekt-Kasus-Markierung



Kongruenz

- Übereinstimmung von Merkmalen zwischen abhängigen Elementen
- *morphologische Markierung* der syntaktischen Funktion des abhängigen Elements am *Kopf* (*head-marking*)

Subjekt-Kongruenz (Numerus und Person)



→ *Verb kongruiert in nominalen Kategorien (Numerus, Person) mit Subjekt-NP*

- (Argument für Regel: $S \rightarrow NP VP$ (Subjekt > Objekt))

2.3. Konstituenz- und Dependenzrelation

- 1 Kursübersicht
- 2 Einführung
 - Syntax als linguistische Disziplin
 - Syntaktische Ausdrucksmittel
 - **Konstituenz- und Dependenzrelation**
 - Repräsentation syntaktischer Strukturen
 - Parsing als Syntaxanalyse
- 3 Analyse syntaktischer Einheiten
- 4 Konstituententests
- 5 Komplement/Adjunkt-Tests

Syntax: Analyse der hierarchischen Struktur von Sätzen

1 Struktur einfacher Sätze (*Clause*):

- **hierarchischer Aufbau** aus Wortgruppen (Konstituenten, insbesondere Phrasen)
- **funktionalen Abhängigkeiten** zwischen diesen syntaktischen Einheiten
→ insbesondere zwischen sog. **Satzgliedern** mit zentralen syntaktischen Funktionen (Subjekt, Objekt, Adverbial, Prädikat)

2 Struktur komplexer Sätze (*Sentence*):

- Aufbau aus einfachen Sätzen
- funktionale Abhängigkeiten zwischen Teilsätzen

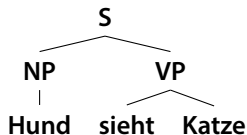
Syntaktische Struktur

- **Menge von Relationen** zwischen Elementen des Lexikons einer natürlichen Sprache (Wörtern) und/oder daraus gebildeten **syntaktischen Einheiten**

Zwei syntaktische Relationstypen

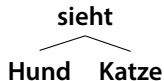
Konstituenz

- **Teil-Ganzes-Beziehung** zwischen Wörtern und aus diesen bestehende syntaktische Einheiten (Konstituenten / Syntagmen)



Dependenz

- **Abhängigkeitsbeziehungen** zwischen Wörtern
- Kopf (*Head*) bestimmt Vorkommen und/oder Form von Dependent



Konstituenten-Struktur

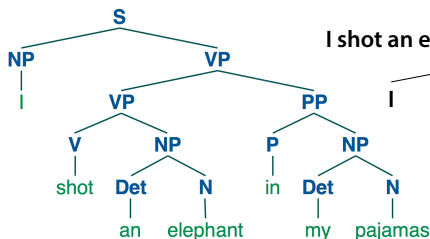
- Aus welchen **syntaktischen Einheiten** besteht ein Satz?
- Wie läßt sich eine Satz **hierarchisch zerlegen**?
- *Hintergrund: Aristotelische Logik (binäre Subjekt-Prädikat-Struktur)*

Syntagmatische Satzanalyse

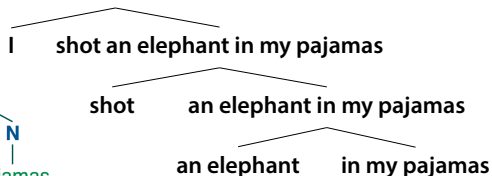
- Feststellung von syntaktischen Einheiten oberhalb Wortebene durch **syntaktische Tests**
- hierarchische Strukturregeln zur Erklärung des **Aufbaus von Sätzen** als lineare Wortfolgen
- Einfacher Satz (*Clause*) besteht aus:
 - **NP**
 - **VP**

I shot an elephant in my pajamas

⇓ *Analysen der Konstituentenstruktur* ⇓



I shot an elephant in my pajamas



Dependenz-Struktur

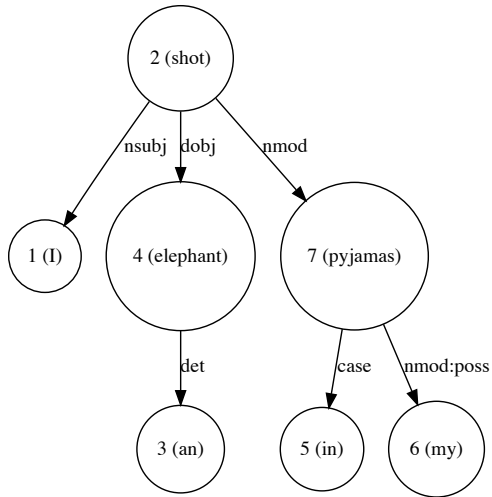
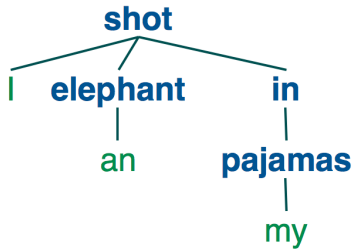
- In welcher **syntaktische Beziehung** stehen Wörter?
- Welche **Funktion** haben sie im Satz?
- *Hintergrund: Frege/Prädikatenlogik (mehrstellige Prädikate)*

Funktionale Satzanalyse

- notwendige und nicht-notwendige Einheiten im Satz
- **Abhängigkeitsverhältnisse** zwischen Wörtern
- Einfacher Satz (*Clause*) benötigt:
 - **Prädikat**
 - **Komplemente** (notwendige Erweiterungen)
 - **Adjunkte** (nicht notwendige Erweiterungen)

I shot an elephant in my pajamas

↓ *Analysen der Dependenzstruktur* ↓



2.4. Repräsentation syntaktischer Strukturen

- 1 Kursübersicht
- 2 Einführung
 - Syntax als linguistische Disziplin
 - Syntaktische Ausdrucksmittel
 - Konstituenz- und Dependenzrelation
 - **Repräsentation syntaktischer Strukturen**
 - Parsing als Syntaxanalyse
- 3 Analyse syntaktischer Einheiten
- 4 Konstituententests
- 5 Komplement/Adjunkt-Tests

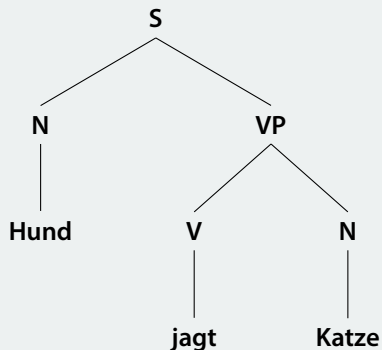
Baumstruktur = gerichteter Graph

- mathematische Repräsentation hierarchischer Struktur

Komponenten einer Baumstruktur

- **Knoten** = Elemente der Struktur
- **Kanten** = geordnete Paare von Knoten (ggf. gelabelt)
 - *Repräsentation der Relation zwischen zwei Knoten*
 - *bei Konstituentenstruktur: Teil-Ganzes-Relation*

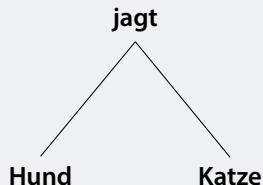
Baumdarstellung:



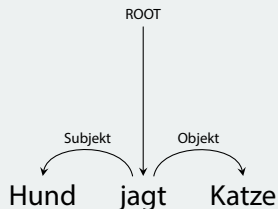
Klammerausdruck:

[S [N Hund] [VP [V jagt] [N Katze]]]

Baumdarstellung (Stemma):



mit gelabelten Kanten:



Klammerausdruck:

[jagt [Hund] [Katze]]

Notation als Tripel:

(jagt, Subjekt, Hund), (jagt, Objekt, Katze)

2.5. Parsing als Syntaxanalyse

- 1 Kursübersicht
- 2 Einführung
 - Syntax als linguistische Disziplin
 - Syntaktische Ausdrucksmittel
 - Konstituenz- und Dependenzrelation
 - Repräsentation syntaktischer Strukturen
 - **Parsing als Syntaxanalyse**
- 3 Analyse syntaktischer Einheiten
- 4 Konstituententests
- 5 Komplement/Adjunkt-Tests

Aufgaben eines Parsingalgorithmus

- 1 **Überprüfung** der grammatischen Struktur einer Eingabe auf Wohlgeformtheit
→ ***Suche einer Ableitung** gemäß der Regeln einer formalen Grammatik*
- 2 **Wiedergabe** der gefundenen grammatischen Struktur
→ ***Ableitungsbaum** (auch: Parsebaum, Syntaxbaum)*

Verarbeitung formaler Grammatiken

- Formale Grammatik = Syntaktisches Strukturmodell
→ *dessen Regeln sind aber nicht mehr als **Sammlung von Strings***
- **Verfahren** notwendig, um zu entscheiden, ob eine Eingabe gemäß einer gegebenen formalen Grammatik **wohlgeformt** ist

Parsing-Algorithmen

- Verfahren zur **Verarbeitung** von formalen Grammatiken zur **Strukturerkennung und -Analyse** der **Eingabe** (Satz als Tokensequenz)

Strukturerkennung

- **Überprüfung** der grammatischen Struktur einer Eingabe als Suche einer Ableitung aus den Regeln einer formalen Grammatik (ob Satz in formaler Sprache enthalten ist)

Strukturzuweisung

- gleichzeitig **Wiedergabe** der in der Suche aufgebauten grammatischen Struktur der Eingabe (Syntaxbaum)

Haupttypen

- 1 **CFGs** (kontextfreie Grammatiken)
 - Konstituentenstruktur (Phrasenstrukturgrammatik)
 - Strukturinformationen in den Knoten des Syntaxbaums
- 2 **Dependenzgrammatiken**
 - Abhängigkeitsrelationen zwischen Wörtern
 - Strukturinformationen in den Kanten des Syntaxbaums (Relationslabel)
- 3 **Chunk-Parser** (u.a. reguläre Grammatiken)
 - partielle Syntaxanalyse (nur wichtigste Konstituenten: NPs, VPs, PPs)
 - flache, nicht-hierarchische Strukturanalyse

Erweiterungen von CFGs

- 1 **PCFGs** = Probabilistische CFGs
- 2 **FCFGs** = Feature-basierte CFGs

Voraussetzungsschritte für automatische Syntaxanalyse

- Sentence Segmentation
- Tokenisierung
- Part-of-Speech-Tagging
- Stemming
- morphologisches Parsing (Kasus, Agreement)

Mögliche Folgeanwendungen

- **Entity Extraction**
- **Relation Extraction**
- **Semantic Parsing** (Natural Language Understanding)

3. Analyse syntaktischer Einheiten

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

Segmentierung: Feststellung syntaktischer Einheiten

- **Zerlegung** Satz in **Konstituenten**
→ *Konstituenten = sprachliche Einheiten, die Teile einer größeren Einheit sind (insbes. Wörter, Phrasen oder Teilsätze)*
- Analyse **syntagmatischer** Beziehung zwischen Einheiten
- zusammengehörende Elemente feststellbar über **Konstituententests**

Klassifizierung: Feststellung syntaktischer Kategorien

- **syntaktische Kategorien = Mengen syntaktischer Einheiten** mit gleichen Eigenschaften (**Klassen**)
- morphologische, syntaktische und semantische **Klassifizierungskriterien**
 - syntaktisches Kriterium: **Austauschbarkeit im gleichen Kontext**
- Analyse **paradigmatischer** Beziehung zwischen Einheiten

Syntagmatische vs. paradigmatische Dimension

syntagmatisch

lineare Kombinierbarkeit

der { große
kleine } Hund jagt { die Katze.
Ferdinand. }

Ein großer { Hund
*Ferdinand } jagt die Katze.

{ Sie
die sich im Gehen nach dem Hund umschauende Frau } stürzt.

Er { sieht einen Hund
geht } auf dem Weg. Er { sieht
*geht } einen Hund.

paradigmatisch

vertikale Austauschbarkeit

4. Konstituententests

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

Konstituententests

- zur Feststellung syntagmatischer Einheiten (Konstituenten)
- Konstituenten-Hierarchie: **unmittelbare vs. mittelbare Konstituenten**

Hauptkriterien

Ersetzbarkeit und Verschiebbarkeit unter Erhalt der **Grammatikalität**

auch: Ersatzprobe

- **eine unter Erhalt der Grammatikalität des Satzes durch eine andere Wortfolge ersetzbare Wortfolge ist Konstituente**
- **Feststellung von Einheiten** (Austauschbarkeit im gleichen Kontext; allgemeiner: auch für Wortarten, Flexionsparadigmen)
- **Pronominalisierung**
→ *Eine durch Pronomen ersetzbare Wortfolge ist Konstituente*
- **für Feststellung Verb mit Erweiterungen (komplexe VP) als Konstituente**
→ *Ersatz komplexer VP durch (intransitives) Verb*

Der Junge verkauft die Äpfel des Bauern.

Identifizierung der Wortfolgen *der Junge* und *die Äpfel des Bauern* als Konstituenten durch **Pronominalisierung**:

Er verkauft sie.

Identifizierung der Wortfolge *verkauft die Äpfel des Bauern* als **komplexe VP-Konstituente** durch Ersatz mit intransitivem Verb:

Der Junge arbeitet.

auch: Verschiebeprobe

- eine im Satz ohne Zerstörung der Grammatikalität verschiebbare Wortfolge ist Konstituente
- im Deutschen: Verschiebung vor finites Verb
→ **Topikalisierung**: rhetorisch-pragmatische Funktion
- Verwendung zum **Testen von Wortstellungsmöglichkeiten**

auch: Auflösung von Ambiguität

- *Der Junge beobachtete das Mädchen mit dem Fernglas.* (ambig)
- *Das Mädchen mit dem Fernglas beobachtete der Junge.* (NP-Attr.)
- *Mit dem Fernglas beobachtete der Junge das Mädchen.* (VP-Adj.)

Der Junge verkauft die Äpfel des Bauern.

Identifizierung der Wortfolgen *der Junge* und *die Äpfel des Bauern* als Konstituenten durch **Permutation**:

Die Äpfel des Bauern verkauft der Junge.

auch: Weglassprobe

- eine ohne Zerstörung der Grammatikalität eines Satzes weglassbare Wortfolge ist Konstituente
- Feststellung syntaktisch notwendiger bzw. optionaler Einheiten (Dependenzbeziehungen)

Der Junge verkauft die Äpfel des Bauern.

Identifizierung der Wortfolge *des Bauern* als **attributive Konstituente** durch Eliminierung:

Der Junge verkauft die Äpfel.

Koordination: Verbindung mit *und* / *aber*

- eine mit einer anderen Wortfolge unter Erhalt der Grammatikalität des Satzes koordinierbare Wortfolge ist Konstituente
- geeignet für die **Ermittlung von Phrasenteilen** (Attributen usw.)
- Analyse der Struktur von **komplexen Konstituenten**

Konstituenten müssen vom gleichen Typ sein

- *Er schrieb einen Brief und eine Karte.*
- *Er schrieb an dich und an mich.*
- **aber nicht:** **Er schrieb einen Brief und an mich.*

Anwendung Koordinationstest

Der Junge verkauft die Äpfel des Bauern.

Identifizierung der Wortfolge *des Bauern* als **Konstituente** durch Koordination:

Der Junge verkauft die Äpfel des Bauern und der Bäuerin

Konstituentenstruktur als Ergebnis aller Konstituententests

Der Junge verkauft die Äpfel des Bauern

Der Junge verkauft die Äpfel des Bauern

verkauft die Äpfel des Bauern

die Äpfel des Bauern

5. Komplement/Adjunkt-Tests

1 Kursübersicht

2 Einführung

- Syntax als linguistische Disziplin
- Syntaktische Ausdrucksmittel
- Konstituenz- und Dependenzrelation
- Repräsentation syntaktischer Strukturen
- Parsing als Syntaxanalyse

3 Analyse syntaktischer Einheiten

4 Konstituententests

5 Komplement/Adjunkt-Tests

- neben den Konstituententests gibt es auch Test auf Abhängigkeitsebene
- Unterscheidung notwendiger vs nicht-notwendiger Abhängigen
- **Komplement:**
 - notwendiges Element
 - Auftreten ist von Kopf gefordert (valenzgefordert)
 - z.B. Subjekt, Objekt als Verb-Komplemente: *die Katze sieht einen Hund*
- **Adjunkt:**
 - nicht-notwendiges Element
 - Auftreten ist nicht von Kopf gefordert
 - z.B. Adjektiv-Attribut als Adjunkt des Nomens: *rotes Auto*
 - z.B. PP-Adverbial als Adjunkt des Verbs: *sieht im Park*

- **Eliminierung der Konstituente** (Feststellung ihrer Notwendigkeit).
- Wenn Grammatikalität erhalten: Angabe/Adjunkt, sonst:
Ergänzung/Komplement

Er beantwortete einen Brief im Arbeitszimmer.

*Er beantwortete einen Brief **im Arbeitszimmer**.* (Adjunkt)

Er beantwortete **einen Brief im Arbeitszimmer.* (Komplement)

- **Auslagerung der Konstituente in einen Adverbialsatz** (funktioniert nicht bei Zeitangaben).
- Wenn Grammatikalität erhalten: Angabe/Adjunkt, sonst: Ergänzung/Komplement

Er wartete im Park auf die Katze.

*Er wartete auf die Katze, als er **im Park** war.* (Adjunkt)

Er wartete im Park, als er **auf die Katze war.* (Komplement)

- **Auslagerung der Konstituente in einen Satz mit dem Verb *geschehen*.**
- Wenn Grammatikalität erhalten: Angabe/Adjunkt, sonst: Ergänzung/Komplement

Er wartete im Park auf die Katze.

*Er wartete auf die Katze, und das geschah **im Park**.* (Adjunkt)

Er wartete im Park, und das geschah **auf die Katze.* (Komplement)

Obligatorische, fakultative und optionale Dependents

- **Obligatorischer** Dependent = **Komplement** (valenzgefordert):
**Er beantwortet einen Brief*
- **Fakultativer** Dependent = **Komplement, aber kontextabhängig eliminierbar**:
Er schreibt einen Brief
→ Ellipse eines an sich obligatorischen Dependents
→ oder Annahme von zwei Valenzrahmen
- **Optional** Dependent = **Adjunkt** (immer eliminierbar):

Er schreibt den ganzen Tag

- **Differenzierung fakultativer von optionalen Dependents**:
beide: eliminierbar → Differenzierung über *geschehen*-Test:

Er schreibt, und es geschieht **einen Brief (fakultativ)*

*Er schreibt, und es geschieht **den ganzen Tag** (optional)*