

# Einführung: Syntax in der Computerlinguistik

## Grundfragen:

1. Wie dürfen sprachliche Ausdrücke zu größeren Ausdrücken zusammengesetzt bzw. in kleinere zerlegt werden?
2. Was sind die „Einheiten“, die zusammengesetzt werden? Welche Kriterien entscheiden darüber, ob mehrere Ausdrücke eine syntaktische (grammatische) Einheit bilden?

## Abgrenzung von der Semantik:

- ▶ Syntax betrachtet den Aufbau von Ausdrücken nach „formalen“ Kriterien, „abgesehen“ von der Bedeutung.

## Abgrenzung von der Pragmatik:

- ▶ Syntax betrachtet hauptsächlich, ob eine Zusammensetzung (im geeigneten Kontext) überhaupt „möglich“ ist, nicht, wie üblich sie ist.

# Grammatik

Ziel der Grammatik ist, die Regelmäßigkeiten (und Ausnahmen) des Aufbaus von Ausdrücken genau zu beschreiben

*Grammatik =*

$\left\{ \begin{array}{l} \text{Gesamtheit der Regeln zum Aufbau von Ausdrücken } \oplus \\ \text{Gesamtheit der Regeln zur Umwandlung von Ausdrücken} \end{array} \right.$

Nutzen einer Grammatik:

1. Angabe eines Verfahrens, den Aufbau von Ausdrücken zu bestimmen
2. am Aufbau eines Ausdrucks soll (mit Zusatzinformation aus dem „Kontext“) seine Bedeutung berechnet werden

In der Computerlinguistik:

*Benutzereingabe* → *Syntaktische Analyse*  
→ *Fehlermeldung*  
| *Semantische Auswertung* → *Bedeutung*

# Wörter

Eine (deskriptive vs. normative) Grammatik beschreibt die (vorkommende vs. erlaubte) Zusammenstellung (*Syntax*) von Wörtern zu Ausdrücken (oder „syntaktischen Einheiten“).

*Wörter* sind die kleinsten in der Syntax betrachteten „Einheiten“.

Aber: Wörter können selbst aus Wörtern zusammengesetzt sein (Komposita), und daher ist die Grenze etwas willkürlich:

*Autofahrer = Fahrer eines Autos / Fahrer des Autos*

Wörter (Lexeme) sind Abstrakta: dasselbe Wort kann (endlich viele) verschiedene *Wortformen* haben, und eine Wortform hat beliebig viele *Vorkommen* (Token).

Wörter sind i.a. *bedeutungstragende* Einheiten. (Ist *aha* ein Wort?)

## Wortformen und Wortarten

Man unterscheidet einige wenige *Wortarten* (oder *Redeteile*, *parts of speech* (POS)). Die Einteilung erfolgt i.a. sowohl nach den Formen als auch nach der Bedeutung und dem Verwendungszweck.

Jede Wortart hat eine endliche Menge von *abstrakten Formen*. Jedes Wort der Wortart hat entsprechende *konkrete Formen*, die das *Paradigma* des Worts bilden:

*Paradigma* : *abstrakte Formen* → *konkrete Formen*

Die abstrakten Formen sind oft nach *Formdimensionen* gegliedert.

**Beispiel:** Formen der Nomen im Deutschen = Kasus × Numerus

abstrakte Formen			sg	pl
(nom,sg)	(nom,pl)	nom	Haus	Häuser
(gen,sg)	(gen,pl)	gen	Hauses	Häuser
(dat,sg)	(dat,pl)	dat	Haus	Häusern
(akk,sg)	(akk,pl)	akk	Haus	Häuser

# Aristoteles: logische Analyse von Ausdrücken

Aristoteles (384 -322 v.Chr.) unternimmt in den Kategorien (κατηγορίαι) und der Hermeneutik (περι ἑρμηνείας) [1] eine Untersuchung der sprachlichen Ausdrücke nach logischen Gesichtspunkten.

Er unterscheidet explizit zwischen Nomina und Verben, wobei nur der Nominativ des Nomens und das Präsens des Verbs als Benennung (ὄνομα) und Tätigkeitswort (ῥήμα) gelten.

Von der *bedeutungsvollen Rede* (λόγος, φωνή σημαντική) wird die (wahre) *Aussage* (λόγος ἀποφαντικός) ausgesondert und allein untersucht. Behauptung und Verneinung (κατάφασις und ἀπόφασις) werden unterschieden und als einfachste Aussagen von den zusammengesetzten unterschieden.

(Kap.5) Der Hauptbestandteil einer Aussage ist das Verb; Namen allein bilden keine aussagende Rede.

## Wortarten bei Aristoteles

**Def** (16a,19) Ein (*einfaches* (ἁπλως)) *Nennwort* (ὄνομα, Nomen) ist ein Ausdruck, der

1. eine (durch Konvention festgelegte) Bedeutung hat,
2. keine Zeitmarkierung trägt,
3. in Verbindung mit „ist“, „war“ und „wird sein“ etwas Wahres oder Falsches bedeutet, und
4. dessen Teile keine Bedeutung haben.

**Def** (16b,6) Ein *Aussagewort* (ῥῆμα, Verb) ist ein Ausdruck, der

1. eine Bedeutung hat,
2. eine Zeitmarkierung trägt, die das Jetzt-Vorliegen angibt,
3. ein Zeichen für das ist, was von etwas anderem ausgesagt wird, und
4. dessen Teile keine Bedeutung haben.

Nur der Nominativ bei Nomina und nur das Präsens bei (finiten) Verben gelten als Nennwort bzw. Aussagewort; jede andere Form ist eine Abwandlung (πτῶσις) davon.

Für andere Wortarten gibt Aristoteles keine Definitionen an.

Aus 10.20 geht hervor, daß **Adjektive** in Verbindung mit ἔστω ein Verb bilden, in Sätzen wie (ἔστι) δίκαιος ἄνθρωπος.

In Kap. 12 (21b,5-10) wird sogar die finite Form des Verbs mit Hilfe des Partizips auf diesen Fall zurückgeführt: οὐδὲν γὰρ διαφέρει εἰπεῖν ἄνθρωπον βαδίζειν ἢ ἄνθρωπον βαδίζοντα εἶναι.  
(Es macht keinen Unterschied zu sagen, „der Mensch geht“ und „der Mensch ist ein Gehender“.)

Ein *unbestimmtes* (ἀόριστον) Nenn- bzw. Aussagewort ist ein Nenn- bzw. Aussagewort mit vorangestelltem „nicht“ (οὐκ).

**Notation:** Ich schreibe  $N$  für ein Nennwort,  $V$  für ein Aussagewort,  $\overline{N}$  für unbestimmte Nennwörter,  $\overline{V}$  für unbestimmte Aussagewörter,  $e = \xi\sigma\tau\iota\nu$ .

Beispiel: Gesundheit :  $N$ , gesundet :  $V$ , ist weiß :  $V$

**Def** (16b,26) Eine *Phrase* (λόγος) ist ein Ausdruck,

1. der etwas bedeutet (besagt, nicht: behauptet), und
2. dessen Teile auch getrennt etwas bedeuten.

**Def** (17a,20) Ein (*Aussage-*)*Satz* (λόγος ἀποφαντικός) ist eine Phrase, die wahr oder falsch sein kann.

Nicht jede Phrase kann wahr oder falsch sein, z.B. eine Bitte kann es nicht.



## Einfache Aussagen (Kap 5, Kap 8, Kap 11)

Ein *Basissatz* ( $\acute{\alpha}\pi\lambda\eta\acute{\alpha}\pi\acute{o}\phi\alpha\nu\sigma\iota\varsigma$ ) ist ein atomarer oder verneinter atomarer Aussagesatz.

**Def** (17a,15) Ein *einfacher* ( $\acute{\epsilon}\iota\varsigma$ , einheitlicher) Satz ist

1. ein atomarer Aussagesatz ( $\kappa\alpha\tau\acute{\alpha}\phi\alpha\sigma\iota\varsigma$ ), oder
2. ein verneinter atomarer Aussagesatz ( $\acute{\alpha}\pi\acute{o}\phi\alpha\sigma\iota\varsigma$ ), oder
3. ein in bestimmter Weise (s.u.) zusammengesetzter Aussagesatz.

**Bem.** (18a,12) Eine Einfachaussage macht *eine* Aussage über *einen* Gegenstand (der allerdings allgemein, d.h.  $N$  oder  $\bar{N}$ , sein kann).

Eine Mehrfachaussage macht eine Aussage über mehrere Gegenstände, oder besteht aus mehreren Einfachaussagen.

Was ist hier *ein* Gegenstand und *eine* Aussage?

1. (Kap. 8) Eine Koordination von Benennungen bezeichnet keinen *einheitlichen* Gegenstand, sondern dient zur Abkürzung einer Koordination von Aussagen: man expandiere gemäß

$$(NP_1 \wedge NP_2)VP \implies (NP_1VP) \wedge (NP_2VP).$$

Aussagen der Form  $(NP_1 \wedge NP_2)VP$  sind also *nicht* einfach: sie haben implizite Teilsätze. Beisp. „Mensch und Pferd sind weiß“

2. (Kap 11, 21a,6) Eine Koordination von Aussagewörtern bzw. prädikativ gebrauchtem Adjektiv und Nomen ergibt nur manchmal *ein* Prädikat und im Satz *eine* Aussage: z.B. bei „gut  $\wedge$  Schuster“, aber nicht bei „zweifüßig  $\wedge$  Lebewesen“, wo sie in zwei implizite Prädikate zerfällt:

$$\begin{aligned} NP e (AP_1 \wedge AP_2) &\implies (NP e AP_1) \wedge (NP e AP_2) \\ NP e (AP \wedge N) &\implies? (NP e AP) \wedge (NP e N) \end{aligned}$$

## 10 Kategorien von Ausdrücken im Satz

In der Kategorienschrift unterscheidet A. zwei Aussageformen (als Anwendung von Seiendem  $X$  auf Seiendes  $Y$  verstanden):

1.  $X$  καθ' ὑποκειμένου τινός ( $Y$ ) λέγεται, „ $X$  wird von einem Zugrundeliegenden ( $Y$ ) ausgesagt“:  
ἄνθρωπος ζῶον ἐστίν = der Mensch ist ein Lebewesen =  
( $Y:N$  ist ein  $X:N$ ) =:  $X(Y)$ .
2.  $X$  ἐν ὑποκειμένῳ ( $Y$ ) ἐστίν, „ $X$  kommt an einem Zugrundeliegenden  $Y$  vor“:  
σῶμα λευκός = der Körper ist weiß = ( $Y:N$  ist  $X:AP$ ) =:  $X[Y]$   
bzw. ( $Y:N$  hat  $X:N$ ) der Körper hat die Farbe Weiß

Die semantischen Werte der so (atomar) *verbundenen* Ausdrücke sind die *Wahrheitswerte*  $\{W, F\} = \{0, 1\}$ .

Innerhalb des Satzes werden zehn *Kategorien* von *unverbundenen* Ausdrücken bzw. ihren semantischen Wertebereichen unterschieden:

			Wertebereich
1.	τί ἐστί·	Was?	Substanz, Wesen (ουσία), Objekte
2.	ποσὸν	Wieviel?	Grade, Skalenwerte, Quantitäten
3.	ποιὸν	Wie?	Eigenschaften, Qualitäten
4.	πρὸς τί	bezüglich was?	Relationen
5.	ποῦ	Wo?	Orte
6.	ποτέ	Wann?	Zeiten
7.	κεῖσθαι	Liegen	Lagen (im Raum)
8.	ἔχειν	Haben	Zustände (Medium)
9.	ποιεῖν	Tun	Tätigkeiten, Handlungen (Aktiv)
10.	πάσχειν	Erleiden	Einwirkungen (Passiv)

Ausdrücke zu 1.- 6. sind Antworten zu bestimmten (Test-) Fragen

# Wortarten bei Dionysos Thrax

Die Grammatik des Dionysos Thrax (170 - 90 v.Chr.) unterscheidet für das Griechische acht Wortarten, die weiter untergliedert werden.

Die Wortarten und ihre Formen (Parameter, Akzidenzien) sind:

## 1. **Nomen** (Genus, Art, Figur, Numerus, Kasus):

Genus: Maskulinum, Femininum, Neutrum

Art: primär, abgeleitet

Figur: Simplex, Kompositum, Dekompositum

Numerus: Singular, Dual, Plural

Kasus: Nominativ, Genitiv, Dativ, Akkusativ, Vokativ

An **abgeleiteten Nomen** werden 7 Unterarten genannt:

patronymische, possessivische, komparativische,  
superlativische, diminutivische, denominale, deverbale.

Dann werden 24 **Unterklassen** genannt, darunter

Eigenname (Nomen Proprium), Gattungsname (Nomen Appellativum),  
Adjektiv (oder Epitheon, Beiwort),  
Relationsnomen, Quasirelationsnomen (z.B. Tag, Nacht),  
Homonym, Synonym,  
Pheronym (Benennung nach einem Ereignis, z.B. der Rächer),  
Eponym (oder Dionym, Beiname: Phoibos = der Strahlende),  
Ethnikon (Volksname),  
Interrogativum, Indefinitum (wer auch immer),  
Anaphorisches Nomen (Gleichsetzungsnomen, z.B. ein solcher),  
Kollektivum, Distributivum (jeder Einzelne),  
Komprehensivum (nach dem Inhalt benanntes Umfassendes, Lorbeerhain  
Artnomen (was in Arten eingeteilt wird, z.B. Lebewesen),  
Spezialnomen (das, was der Art nach unterschieden ist),  
Ordnungsnomen (*n*-ter), Numeral (was eine Zahl bezeichnet),  
Absolutes Nomen (was aus sich selbst begriffen wird, z.B. Gott),  
Partizipativum (was an einem Stoff teilhat, z.B. aus Weizen)

2. **Verb** (Modus, Diathese, Art, Figur, Numerus, Person, Tempus, Konjugation):

Modus: Indikativ, Imperativ, Optativ, Konjunktiv, Infinitiv

Diathese: Aktiv, Passiv, Medium

Art: primär, abgeleitet

Figur: Simplex, Kompositum, Dekompositum

Numerus: Singular, Dual, Plural

Person: erste, zweite, dritte

Tempus: Präsens, Präteritum, Futur.

Das Präteritum hat vier Sorten: Imperfekt (sich erstreckend), Perfekt (vorliegend), Plusquamperfekt (mehr als vollendet), Aorist (nicht begrenzt).

Es gibt sechs Konjugationen der nicht endbetonten und drei Konjugationen der endbetonten Verben.

3. **Partizip** (Genus, Art, Figur, Numerus, Kasus, Diathese, Art, Figur, Tempus, Konjugation)
4. **Artikel** (Genus, Numerus, Kasus): wird den Nomen vorangestellt (ὁ ho=der) oder nachgestellt (τις tis=ein)
5. **Pronomen** (Person, Genus, Numerus, Kasus, Figur, Art): ein Wort, das statt eines Nomens gebraucht wird und das bestimmte Personen bezeichnet.
6. **Präposition**: ein Wort, das allen Redeteilen vorangestellt wird, sowohl synthetisch als auch parataktisch



7. **Adverb** (Figur): ein unflektierter Redeteil, der vom Verb her ausgesagt oder dem Verb hinzugefügt wird.

Figur: Simplex, Kompositum

Hiervon unterscheidet Thrax 26 Arten nach ihrer Bedeutung oder kommunikativen Funktion, z.B.

Zeitadverbien	(nun, damals, später, heute, solange),
Zahladverbien	( <i>n</i> -mal),
Ortsadverbien,	
Wunschadverbien	(wenn doch),
Adverbien der Ablehnung	(nicht [gr. οὐ], nirgends, keineswegs),
Adverbien des Verbotens	(nicht [gr. μή]),
Adverbien des Vergleichens	(wie),
Adverbien des Vermutens	(wahrscheinlich, vielleicht),
Adverbien des Fragens	(wer, warum, wann, wie)

8. **Konjunktion** (Art): ein Wort, das Gedanken in einer bestimmten Anordnung verknüpft und das verdeutlicht, was nur implizit geäußert wird.

Art: kopulativ (und, aber, ferner),  
disjunktiv (oder),  
kontinuativ (wenn),  
subkontinuativ (da, weil),  
kausal (damit, um-zu, weil, weshalb),  
dubitativ (wohl, etwa, doch-wohl-nicht),  
syllogistisch (folglich, nun-aber, also),  
expletiv (offenbar, ja, nämlich, durchaus, allerdings)  
(ggf. adversativ: (dennoch, trotzdem))

Beachte:

- ▶ die Gliederung ist relativ fein, trennt aber nicht zwischen Unterklassen- und Formparametern (z.B. beim Nomen: Genus, Art, Figur vs. Numerus, Kasus)
- ▶ die Gliederung berücksichtigt syntaktische Eigenschaften (bei Artikel, Pronomen, Präposition, Verb) und semantische oder kommunikative Eigenschaften (bei Adverb, Konjunktion)
- ▶ die heutigen Klassifizierungen weichen z.T. in den Wortarten ab, z.B. ist das Adjektiv bei Dionysos ein Nomen, das Partizip eine Wortart.

Die Grammatik von Dionysos [2] ist –neben Euklids *Elementen*– das langlebigste Lehrbuch des Abendlands; sie wurde über 1000 Jahre lang verwendet.

Sie enthält aber keine Grammatikregeln im heutigen Sinne.

Man erkennt einige allgemeine Probleme der Klassifizierung:

- ▶ Soll man nach reinen Formeigenschaften klassifizieren?
- ▶ Wo soll man abgeleitete Wortformen einordnen?
- ▶ Form- und Bedeutungsaspekte entsprechen einander nicht eindeutig: z.B. gibt es Nomina im Singular, die eine Vielheit bedeuten, wie Demos (Volk).

# Literatuur



Aristoteles.

*Kategorien. Hermeneutik oder vom sprachlichen Ausdruck.  
Organon Band 2.*

Felix Meiner, 1998.



Pierre Swiggers and Alfons Wouters.

*De Tékhne Grammatikē van Dionysius Thrax: De oudste  
Spraakunst in het Westen. Inleiding; Griekse tekst met  
Nederlandse vertaling en noten; Duitse vertaling (door Wilfred  
Kürschner); terminologisch apparaat en bibliografie.*

Peeters, Leuven - Paris, 1998.