



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

Computerlinguistik

Lehreinheit 1-2: Computerlinguistik

Dr. Christina Alexandris

Nationale Universität Athen

Deutsche Sprache und Literatur

Περιεχόμενα ενότητας

- Γενική Εισαγωγή
- Γενική Εισαγωγή στην Υπολογιστική Γλωσσολογία
- Εισαγωγή στην Επεξεργασία Φυσικής Γλώσσας
- Μορφο-συντακτική και Σημασιολογική Ανάλυση, Συντακτικοί Αναλυτές, Ηλεκτρονικές Γραμματικές και Φορμαλισμοί



Einführung Computerlinguistik und Künstliche Intelligenz

Je näher die Computer und mit ihr die Informatik an den Menschen herangehen, umso mehr haben sie es mit der Sprache zu tun“

(v. Hahn., 2001)



Die Computerlinguistik ist eine relativ neue Disziplin (1/2)

Sie ist zwar in den **80er** Jahren entstanden, aber ihre erste große Entwicklung war in den **90er** Jahren erkennbar.

Die Disziplin der Computerlinguistik ist aus den Bedürfnissen des **Zeitalters der Mikrocomputer** entstanden, in der zunehmend mehr Industrien, Organisationen, Behörden und private Benutzer den Computer für eine zunehmend größere Zahl und Varietät von Tätigkeiten.



Die Computerlinguistik ist eine relativ neue Disziplin (2/2)

Der Computer war **nicht mehr ein Werkzeug für Informatiker** und für das Labor, sondern er wurde zu einem unentbehrlichen Teil der Büroaustattung, aber auch in der alltäglichen Praxis, in den Bereichen des Verkehrs und der Kommunikation, in der Technik und Industrie, in dem finanziellen- und wirtschaftlichen Bereich, im Bereich der Medizin, der Didaktik aber auch in den Bereichen der theoretischen Wissenschaften, der Juristik und der Kunst und Kultur.



Die zunehmende Zahl der Anwendungsbereiche der Computer

- ist mit der zunehmenden Auswahl und Varietät der vom Computer zu verarbeitenden **Daten** verbunden: Texte, Bilder, Videos, Audio und dazu noch die natürliche Sprache als Mittel der Kommunikation zwischen Mensch und Computer.
- Für die Menschen ist die Kommunikation zum größten Teil mit der Faktor „**Sprache**“ verbunden. Das bedeutet, dass in vielen Fällen das Bedürfnis entstanden ist, dass der Computer die natürliche Sprache verstehen und gegebenenfalls auch erzeugen soll.



Aus den vorher erwähnten Bedürfnissen

- ist in dem Bereich der **Künstlichen Intelligenz** ein neues Anwendungsgebiet entstanden: die Computerlinguistik.
- Die Computerlinguistik ist ein Anwendungsgebiet der Künstlichen Intelligenz
- mit Modellen aus der **Theoretischen Sprachwissenschaft**
- und mit Methoden aus der **Informatik**

(v. Hahn, 2001)



In der Computerlinguistik...

- werden die **Modelle der Theoretischen Sprachwissenschaft** und deren Disziplinen
- wie z.B die Modelle der Syntax, der Semantik, der Morphologie aber auch die Modelle der Pragmatik und der Phonetik und Phonologie
- in der Form von **Regeln** umgesetzt, die der Computer später als **Programm** erkennen und verarbeiten kann
- Diese Modelle aus der Theoretischen Sprachwissenschaft versuchen das **Sprachliche Wissen** zu beschreiben



Einer der Ziele der Computerlinguistik ist,

- dass der Computer dieses sprachliche Wissen **erwirbt**.
- Dieser Prozess wird nach Hanneforth (2001) auf folgende Weise beschrieben:
- “Das **sprachliche Wissen** wird von den Computerlinguisten in die **Praxis** umgesetzt, um die größte **Barriere** zwischen Mensch und Maschine zu überwinden.” (Hanneforth, 2001)



Die Anwendungsbereiche der Computerlinguistik (1/2)

- können in verschiedenen **Kategorien** aufgeteilt werden.
- Ein bekanntes Gebiet, in dem die Computerlinguistik angewendet wird, ist die **Maschinelle Übersetzung** geschriebener und/oder gesprochener Sprache von Texten.



Die Anwendungsbereiche der Computerlinguistik (2/2)

- aber auch mit dem Gebrauch der Maschinellen Übersetzung in **Frage-Antwort-Systemen** und **Dialogsystemen**.
- Anwendungsbereiche der Computerlinguistik in Bezug auf **Texte** sind auch die Textverstehende und Texterzeugende Systeme, die Automatische Textverfassung und die Automatische Informationsextraktion.



Anwendungsbereiche der Computerlinguistik als Anwendungsbereiche der Linguistik (1/2)

- **Morphologie, Lexikologie, Syntax, Semantik:**
Alle Anwendungsbereiche der Computerlinguistik
- **Phonetik-Phonologie:** Erkennung (ASR) und/oder Erzeugung gesprochener Sprache, Frage-Antwort-Systeme und Dialogsysteme und maschinelle Übersetzung gesprochener Sprache



Anwendungsbereiche der Computerlinguistik als Anwendungsbereiche der Linguistik (2/2)

- **Pragmatik:** Frage-Antwort-Systeme, Dialogsysteme (geschriebener und gesprochener Sprache)
- **Soziolinguistik:** Frage-Antwort-Systeme, Dialogsysteme (geschriebener und gesprochener Sprache)
- **Textlinguistik:** Textverstehende und -erzeugende Systeme, automatische Textverfassung



Zu den Anwendungsbereichen der Computerlinguistik

- **bezüglich der gesprochenen Sprache** gehört die Erkennung (Automatic Speech Recognition- ASR) und Erzeugung der gesprochenen Sprache für **Ferngespräche** und sonstige Dienstleistungen.
- Weitere Anwendungsbereiche bezüglich der gesprochenen Sprache sind die Anwendungen für **Behinderte**, Blinde, Schwerhörige und für Dialogsysteme gesprochener Sprache,
- sowie die Anwendungen in der **Raumfahrt** (Shriberg et al., 2003 (NASA-SRI)).



Multimediale Anwendungen oder Multimodale Anwendungsbereiche der Computerlinguistik

- sind die Anwendungen in **elektronischen Lehrmitteln**,
- zum Beispiel, für den Fernunterricht,
- für den Unterricht für Behinderte
- und für den Fremdsprachenerwerb.
- Zu den Multimedialen Anwendungen oder Multimodalen Anwendungsbereichen der Computerlinguistik gehört
 - die Sprachliche **Bild- und Szenenbeschreibung**,
 - die Sprachliche **Informationsvermittlung**
 - aber auch die **Automatische Informationsextraktion**.



Was macht
ein System für die Verarbeitung natürlicher Sprache?

Wie „versteht“ der Computer die
Sprache?

Kern aller Anwendungsbereiche der Computerlinguistik (1/2)

- ist ein System für die Verarbeitung der natürlichen Sprache
- **(Natural Language Processing system – NLP System).**
- So ein System für die Verarbeitung natürlicher Sprache kann die natürliche, bzw. „menschliche“ Sprache „verstehen“ und sie, anschließend, auf verschiedenen Weisen, je nach der Anwendung des Systems, verarbeiten.
- Der Vorgang der Verarbeitung kann je nach Struktur und Anwendungsbereich des Systems variieren.



Kern aller Anwendungsbereiche der Computerlinguistik (2/2)

- Jedoch kann in groben Zügen dieser Vorgang **in drei Phasen** beschrieben werden:
- Ein System für die Verarbeitung natürlicher Sprache analysiert den (geschriebenen oder gesprochenen) Text, der vom System erkannt wird (**Analyse**),
- es verarbeitet den analysierten Text, je nach Art der Anwendung (**Verarbeitung**)
- und generiert den Text in der gewünschten Form, je nach Art der Anwendung (**Generierung**).



Ein System für die Verarbeitung natürlicher Sprache enthält

- (1) Datenbanken mit **lexischen Einheiten** und
- (2) Computerprogramme mit **semantischen und morphosyntaktischen Regeln**.
- Das System enthält eine Datenbank aller von ihm **erkennbaren** Wörter (lexische Einheiten).
- Diese **Datenbank** bildet die lexischen Regeln der Sprache also eine Art „Liste“ oder „Wörterbuch“ mit allen erkennbaren Wörtern einer Sprache (oder mehreren Sprachen, je nach der Anwendung).
- Diese Datenbank, d.h. alle lexische Regeln, **„kommuniziert“** mit den semantischen und morphosyntaktischen Regeln.



Die Programme mit den syntaktischen Regeln und den morphologischen Regeln,

- **sorgen dafür** daß jeder Satz und jedes individuelle Wort des Satzes von dem System **richtig verstanden**, bzw. analysiert wird (**Analyse**).
- In einigen Anwendungen der Computerlinguistik zum Beispiel, in der automatischen Textverfassung oder in der maschinellen Übersetzung sorgen **weitere syntaktische und morphologische Regeln** dafür, daß jeder Satz und jedes individuelle Wort des Satzes von dem System richtig **umgesetzt (Transfer)** und **erzeugt (Generierung)** wird,
- oder in der Syntax und der Morphologie der **Zielsprache** richtig umgesetzt und erzeugt wird (je nach der Art der Anwendung und der Phase des Prozesses bzw. Modul des Systems, in der diese morphosyntaktischen Programme aktiviert werden).



Die Programme mit den semantischen Regeln,

- die in manchen Systemen mit den syntaktischen Regeln verknüpft sind, enthalten Informationen über die **semantische** Bedeutung jedes Wortes,
- oder in manchen Fällen jeder Äußerung.
- Grammatische Regeln die sowohl morphosyntaktische als auch semantische Informationen enthalten, haben die Form von **Grammatikformalismen** mit morphosyntaktischen und semantischen **Merkmalsstrukturen**.



Ohne die semantischen und morphosyntaktischen Regeln ...

- ...ist das System kein System sprachlicher Verarbeitung sondern ein elektronisches Wörterbuch.
- Die semantischen und morphosyntaktischen Regeln werden "**Grammatiken**" genannt.
- Je nach der **Phase** der Verarbeitung natürlicher Sprache werden die entsprechenden Grammatiken aktiviert -
 - die **Analyse-Grammatik**, die **Transfer-Grammatik** und die **Generierungs-Grammatik**.



Was macht
ein System für die Verarbeitung natürlicher Sprache?

Syntaktische Analyse

Ein System künstlicher Intelligenz,

- also z.B. ein Computer, kann hierarchische Strukturen verstehen.
- Wenn man die natürliche Sprache als eine **hierarchische Struktur** beschreibt, dann kann der Computer die natürliche Sprache "verstehen" und "bearbeiten".



Zwei Hauptrichtungen der syntaktischen Analyse sind in der linguistischen Tradition erkennbar:

- (1) die **Dependenz- und Determinations- Syntax** und
- (2) die **Konstituentenstruktursyntax**.
- In beiden Traditionen werden Baumgraphen bzw. Strukturbäume zur Notation von syntaktischen Strukturen verwendet. **syntaktische Beschreibungsmodelle**
- Beide werden in der Computerlinguistik benutzt



Die Dependenz- und Determinationsyntax

- fasst syntaktische Strukturen als Relationen zwischen Wörtern auf.
- In der **Dependenzgrammatik** wird die Struktur des Satzes vom Verb festgelegt und alle anderen Wörter sind unmittelbar oder mittelbar vom Verb abhängig.
- Diese Abhängigkeit ("dependency") wird als Relation in den Baumgraphen durch Kanten wiedergegeben.



In der Konstituentenstruktursyntax

- werden neben Wörtern auch komplexere Einheiten, die sogenannten **Konstituenten** oder **Phrasen** angenommen.
- Die konstituentenorientierte Tradition hatte – vor allem aufgrund der richtungssweisenden Arbeiten von Noam Chomsky (Chomsky, 1968, Chomsky, 1965) – einen starken Einfluss auf die Computerlinguistik der vergangenen Jahrzehnte.



Strukturbäume aus der Tradition der Konstituentenstrukturgrammatik

- enthalten folgende Informationen (Langer, 2004):
 - **(1) Segmentierung: Zerlegung eines komplexen Ausdrucks in Teile** (die wiederum komplex sein können)
 - **(2) Kategorisierung der komplexen Teilausdrücke** (Subjekt, Prädikat)
 - **(3) Lineare Abfolge: Die lineare Abfolge der Wörter** (kleinste Bestandteile der Satzkonstituenten, Blätter) entspricht der Wortstellung



Nehmen wir an, dass unser Computer, z.B., den folgenden Satz "verstehen,, soll:

- "*Ein dicker Kater sitzt auf dem Stuhl*"
- Der Satz "*Ein dicker Kater sitzt auf dem Stuhl*" wird von dem Computer (dem System) nur als eine **Reihe von nicht-mathematischen Zeichen**, nämlich als "**alphanumerische Zeichen**" (**Strings**) und (leere Lücken) **Leerzeichen** "verstanden".
- Für den Computer ist das **Wort** eine Einheit aus **alphanumerischen Zeichen**, die rechts und links durch **Leerraumzeichen** (engl. "white space") oder durch **Interpunktion** begrenzt werden.
- Diese Reihe von Elementen bildet die "**Eingabedaten**" (**Input**) des Computers (des Systems).



Die Reihe von Elementen

- ist für den Computer nur dann von Bedeutung,
- wenn man dieser Eingabe **eine Bedeutung zuweist**.
- Das kann in Form von **lexikalischen** und **grammatischen Regeln** realisiert werden.



Die Reihe von alphanumerischen Zeichen, die rechts und links durch Leerraumzeichen begrenzt werden, (1/2)

- **werden als individuelle Einheiten definiert** so wird die Eingabe der Reihe von alphanumerischen Zeichen "***Ein dicker Kater sitzt auf dem Stuhl***" in **einzelne Stücke** oder "**Segmente**„ geteilt.
- Es handelt sich um einen Prozess, der als "**Segmentierungsverfahren**„ (**Segmentation**) bezeichnet wird.



Die Reihe von alphanumerischen Zeichen, die rechts und links durch Leerraumzeichen begrenzt werden,(2/2)

- werden als **individuelle Einheiten** definiert. Diese **Stücke bzw. Segmente**, die eigentlich den einzelnen **Wörtern** in dem Satz "*Ein dicker Kater sitzt auf dem Stuhl*" entsprechen, werden **Tokens** genannt (**tokens**).
- Die Art des Segmentierungsverfahrens wird **Tokenisierung (tokenization)** genannt.
- Mit der **Tokenisierung** wird jedes Wort ("**Token**") eines Textes erfasst.



Mit einem anderen Segmentierungsverfahren werden Wörter gemäß ihrer Wortarten

- (linguistische Einheiten wie z.B. Wörter, Phrasen, Sätze) ausgezeichnet.
- Somit findet ein Prozess der **Kategorisierung** statt.
- Dieser Prozess wird **Tagging (tagging)** genannt und mit Hilfe eines speziellen Programs, eines **Taggers (tagger)**, durchgeführt.
- Ein Tagger, der eine Analyse und Generierung jedes Tokens gemäß seiner **Wortart** leistet, wird **Part-of-Speech Tagger (POS tagger)** genannt.
- Das **Inventar an Wortarten (tags)**, das für den Tagging-Prozess benutzt wird, wird als "**Tagset**" (**tagset**) bezeichnet.



In einigen Anwendungen

- kann der Tagger nicht nur Wörter und Phrasen sondern auch Ansätze oder Diskursabschnitte auszeichnen.
- Das Tagset kann außer einem Inventar an Wortarten auch Klassifikationen für Interpunktion, numerische Angaben oder Daten enthalten.



Was macht
ein System für die Verarbeitung natürlicher Sprache?

ANALYSEN

Tokenisierung und Tagging der Eingabe: Ein dicker Kater sitzt auf dem Stuhl

Tokens

+ ein

+ dicker

+ Kater

+ sitzt

+ Auf

+ dem

+ Stuhl

Tagset (Tags):

+ ein =D

+ dicker =ADJ

+ Kater =N

+ sitzt =V

+ auf =P

+ dem =D

+ Stuhl =N



Griechisches Beispiel: Tokenisierung und Tagging der Eingabe: ένας χοντρός γάτος κάθεται πάνω στην καρέκλα

Tokens:

+ ένας

+ χοντρός

+ γάτος

+ κάθεται

+ πάνω

+ στην

+ καρέκλα

Tagset (Tags):

+ ένας =D

+ χοντρός =ADJ

+ γάτος =N

+ κάθεται =V

+ πάνω =P

+ στην =D2

+ καρέκλα =N



Die Wörter, die (anhand der Tags) gemäß ihrer Wortarten ausgezeichnet sind,

- können nun von einem Programm bearbeitet werden, das eine automatische syntaktische Analyse eines Satzes oder einer Phrase durchführt.
- Im Allgemeinen kann diese automatische syntaktische Analyse eines Satzes oder einer Phrase (Eingabe eines Systems) als der Prozess des **Parsings (parsing) definiert werden.**
- Nach Schneider (2002) wird Parsing als "automatisches Zerlegen eines komplexen Ausdrucks in seine Konstituenten" definiert (Definition nach Schneider, 2002).



Basisinstrument für solche syntaktische Analysen ist eine kontextfreie Grammatik

- in der die lineare Abfolge der Konstituenten bestimmt wird.
- Eine traditionelle kontextfreie Grammatik basiert auf dem Standard-Modell der Generativen Grammatik (Chomsky 1965).
- In diesen Grammatiken beginnt die Analyse, in der Regel bei der Ebene des Satzes, bis sie zu den kleinsten Einheiten gelangt.
- Typischerweise werden diese Elemente mit dem Konzept "Wörter" als lexikalische Einheiten identifiziert.
- Die Satzebene kann als Startpunkt der Analyse, mit Startsymbol "S", bezeichnet werden, während die lexikalischen Einheiten als "Terminalsymbole" bezeichnet werden können (Langer, 2004)



Mit den folgenden Regeln einer sehr einfachen kontextfreien Grammatik

- kann der Computer die Eingabe "*Ein dicker Kater sitzt auf dem Stuhl,*" als eine Folge von sieben (7) Strings wahrnehmen, die eine engere oder weniger enge Beziehung zueinander besitzen.
- Diese Strings-Elemente können nur bestimmte mögliche Reihenfolgen eingehen.
- Akzeptierbare Reihenfolgen werden von syntaktischen Regeln bestimmt.
- Die **syntaktischen Regeln sind zweidimensional**, denn sie werden von einer hierarchischen Struktur bestimmt -zum Beispiel von den Strukturen der Generativen Grammatik (X-bar Theorie).



Kontextfreie Grammatik

- Ohne diese hierarchische Struktur könnten die engeren Beziehungen, die manche Elemente (Strings) zueinander haben, nicht beschrieben werden, wie zum Beispiel die Beziehung "Verb – Verbalphrase".
- Mit diesen Regeln "weiß" der Computer an welchen Stellen er die Eingabe "Ein dicker Kater sitzt auf dem Stuhl" in weitere Stücke/Segmente teilen/segmentieren kann und sie in kleinere und noch kleinere Stücke segmentieren und analysieren kann.



Eine einfache kontextfreie Grammatik

- für die Generierung des Satzes
- "***Ein dicker Kater sitzt auf dem Stuhl***" (Analyse nach Jurafsky and Martin, 2008):
- Regel:
 - + S -> NP VP (S = Startsymbol)
 - + NP -> D N'
 - + VP -> V PP
 - + PP -> P NP
 - + N' -> ADJ N
 - + NP -> D N



Anhand einer kontextfreien Grammatik wird der Satz "Ein dicker Kater sitzt auf dem Stuhl" in verschiedenen Stufen allmählich geparst (Parsing) (1/3)

(i) 1. Ebene:

[ein dicker **Kater**] [sitzt auf dem Stuhl]

Griechisches Beispiel: [ένας χοντρός γάτος] [κάθεται πάνω στην καρέκλα]

Regel:

S -> NP VP (S = Startsymbol)



Parsing: "Ein dicker Kater sitzt auf dem Stuhl" (2/3)

(ii) 2. Ebene:

[ein [dicker Kater]] [sitzt [auf dem Stuhl]]

Griechisches Beispiel:

[ένας [χοντρός γάτος]] [κάθεται [πάνω στην καρέκλα]]

Regel:

NP -> D N'

VP -> V PP



Parsing: "Ein dicker Kater sitzt auf dem Stuhl" (3/3)

(iii) 3. Ebene:

[ein [dicker [Kater]]] [sitzt [auf [dem Stuhl]]]

Griechisches Beispiel:

[ένας [χοντρός [γάτος]]] [κάθεται [πάνω [στην
καρέκλα]]]

Regel:

PP -> P NP

N' -> ADJ N



In jeder Stufe werden bestimmte Regeln der kontextfreien Grammatik verwendet bzw. aktiviert, die mit der Erzeugung der entsprechenden Ebenen der hierarchische syntaktischen (Baum-) Struktur korrespondieren.

(iii) 3. Ebene:

[ein [dicker [Kater]]] [sitzt [auf [dem Stuhl]]]

Griechisches Beispiel:

[ένας [χοντρός [γάτος]]] [κάθεται [πάνω [στην καρέκλα]]]

Regel:

PP -> P NP

N' -> ADJ N

(iv) 4. Ebene:

[ein [dicker [Kater]]] [sitzt [auf [dem [Stuhl]]]]

Griechisches Beispiel:

[ένας [χοντρός [γάτος]]] [κάθεται [πάνω [στην [καρέκλα]]]]

Regel:

NP -> D N

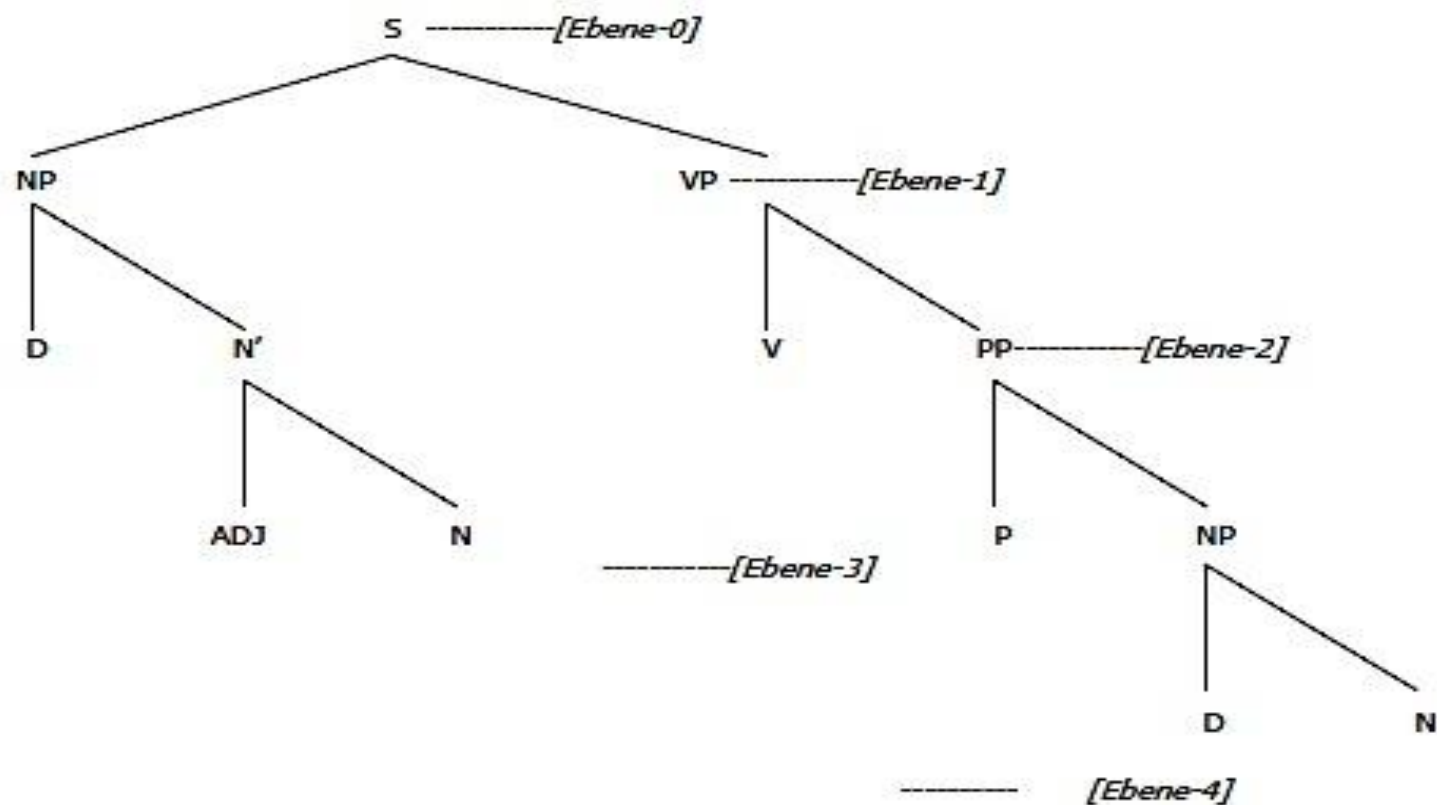


Syntaktische Struktur für die Eingaben:

[ein dicker **Kater** **sitzt** auf dem Stuhl]
[der große **Löwe** **springt** aus dem Käfig]

Griechisches Beispiel:

[ένας χοντρός **γάτος** **κάθεται** πάνω στην καρέκλα]



Literaturverzeichnis (1/2)

- Alexandris, C., Fotinea, S-E and Efthimiou, E. (2005). Emphasis as an Extra-Linguistic Marker for Resolving Spatial and Temporal Ambiguities in Machine Translation for a Speech-to-Speech System involving Greek. In: *Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction (UAHCI 2005)*, 22-27 July 2005, Las Vegas, Nevada, USA.
- Alexandris, C. (2003). Translational Issues in the Sublanguage of Written and Spoken Journalistic Texts in Modern Greek. In: *Proceedings of the International Conference on Choice and Difference in Translation*, Athens 2003, 287-307.
- Bateman, J, Paris, C. (2004). Benützermodellierung. In: *Computerlinguistik und Sprachtechnologie, Eine Einführung*, Carstensen, K.U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., Langer, H. (Hrsg.), 2te überarbeitete und erweiterte Auflage, München: Spektrum Akademischer Verlag.
- Cohen, P., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997). Quickset: Multimodal interaction for distributed applications. In *Proceedings of the 5th ACM International Multimedia Conference*, pages 31-40.
- Dorna, M., Jekat, S. (2004). Maschinelle und computergestützte Übersetzung. In: *Computerlinguistik und Sprachtechnologie, Eine Einführung*, Carstensen, K.U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., Langer, H. (Hrsg.), 2te überarbeitete und erweiterte Auflage, München: Spektrum Akademischer Verlag.
- Fairclough, N. (2001). *Language and Power*. Pearson Education, Upper Saddle River, NJ .
- Forrester, M. (1996). *Psychology of Language*. SAGE Publications, Thousand Oaks, CA, USA.
- Hatim, B. (1997). *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*, University of Exeter Press.
- Jurafsky, D., Martin, J. (2008). *Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, 2nd edition, Prentice Hall series in Artificial Intelligence, Pearson Education, Upper Saddle River, NJ, USA.



Literaturverzeichnis (2/2)

- Kellner, A. (2004). Dialogsysteme. In: *Computerlinguistik und Sprachtechnologie, Eine Einführung*, Carstensen, K.U., Ebert, C., Endriss, C., Jekat, S., Klabunde, R., Langer, H. (Hrsg.), 2te überarbeitete und erweiterte Auflage, München: Spektrum Akademischer Verlag.
- Lehrndorfer A. (1996). *Kontrolliertes Deutsch: Linguistische und Sprachpsychologische Leitlinien für eine (maschiell) kontrollierte Sprache in der technischen Dokumentation*, Tübingen: Narr.
- Moegele, H., Moritz Kaiser, M., Schiely, F. (2006). SmartWeb UMTS Speech Data Collection, The SmartWeb Handheld Corpus. In: *Proceedings of LREC 2006*, Genova, Italy, pp. 2106-2111.
- Müller, S. (1998): Babel 1.50, Web-Interface, Universität Bremen.
- v. Hahn, W. (2001). Maschinelle Übersetzung, Proseminar der Fakultät für Informatik, Universität Hamburg.
- Hanneforth, T. (2001). Was ist Computerlinguistik?, Übersicht des Computerlinguistikprogramms, Institut für Linguistik, Universität Potsdam.
- Shriberg, E, Stolcke, A., Stone, L., Bratt, H., Ferrer, L. and Sömnez, K. (2003). Harnessing Speech Prosody for Robust Human-Computer Interaction, Active Research Task, Intelligent Systems Project, CICT, SRI-International, NASA-Ames Research Center.
- Tomita, M., Mitamura, T., Musha, H. and Kee, M. (1988). The Generalized LR Parser/Compiler Version 8.1, Center For Machine Translation, Carnegie Mellon University, Pittsburgh, PA, USA
- Wardhaugh, R. (1992). *Introduction to Sociolinguistics*. Oxford, Blackwell.
- Wodack, R (1996). *Disorders of Discourse*. Longman, New York.
- Vertan, C. (2001). Einführung in Grundprobleme der Maschinellen Übersetzung, Seminar der Fakultät für Informatik, Universität Hamburg.
- Έργο διαλογικού συστήματος. Verbmobil (Γερμανία): <http://verbmobil.dfki.de/>



Τέλος Ενότητας

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Σημειώματα

Σημείωμα Ιστορικού Εκδόσεων Έργου

Το παρόν έργο αποτελεί την έκδοση 1.0.

Έχουν προηγηθεί οι κάτωθι εκδόσεις:

- Έκδοση διαθέσιμη εδώ. <http://eclass.uoa.gr/courses/GS158/>



Σημείωμα Αναφοράς

Copyright Εθνικών και Καποδιστριακών Πανεπιστημίων Αθηνών, Χριστίνα Αλεξανδρή. «Υπολογιστική Γλωσσολογία. Computerlinguistik». Έκδοση: 1.0. Αθήνα 2014. Διαθέσιμο από τη δικτυακή διεύθυνση:

<http://opencourses.uoa.gr>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Μη Εμπορική Χρήση Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



[1] <http://creativecommons.org/licenses/by-nc-sa/4.0/>

Ως **Μη Εμπορική** ορίζεται η χρήση:

- που δεν περιλαμβάνει άμεσο ή έμμεσο οικονομικό όφελος από την χρήση του έργου, για το διανομέα του έργου και αδειοδόχο
- που δεν περιλαμβάνει οικονομική συναλλαγή ως προϋπόθεση για τη χρήση ή πρόσβαση στο έργο
- που δεν προσπορίζει στο διανομέα του έργου και αδειοδόχο έμμεσο οικονομικό όφελος (π.χ. διαφημίσεις) από την προβολή του έργου σε διαδικτυακό τόπο

Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.



Σημείωμα Χρήσης Έργων Τρίτων (1/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Εικόνες/Σχήματα/Διαγράμματα/Φωτογραφίες

Διάγραμμα 1: διαφάνεια 47/Syntaktische Struktur für die Eingaben, Alexandris, C. (2010): Linguistik und ihre Anwendungen in der Computerlinguistik: Ein Arbeitsbuch, Athens, Papasotiriou. (Students book, in German)



Σημείωμα Χρήσης Έργων Τρίτων (2/2)

Το Έργο αυτό κάνει χρήση των ακόλουθων έργων:

Πίνακες

