

# ZIELE

## der Computerlinguistik

(50er Jahre)

- Erfassung der Sprache (?)  
(Sprache = mathematische Formeln)
- Generieren der Sprache (?)
- maschinelle Übersetzung (?)  
(Nürnberger Prozess – großes Aufgebot von Dolmetschern)

# COMPUTERLINGUISTIK

- beschreibt sprachliche Strukturen und Prozesse mithilfe der Datenverarbeitung
- unterstützt die Korpuslinguistik
- erleichtert die Erkundung des Phänomens „Sprache“

# COMPUTER- UND KORPUSLINGUISTIK

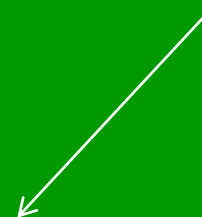
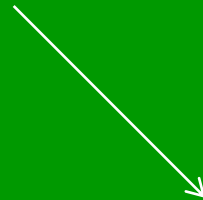
Computer+Linguistik

Korpus+Linguistik



Computerlinguistik

Korpuslinguistik



(Computer)Korpuslinguistik

# KORPUSTYPOLOGIE

- Welche Kriterien würden Sie erwarten?
- Charakterisieren Sie diejenigen Korpora, die Sie kennen.
- Welche wird es (kaum jemals) geben?
- Welche Korpora würden Sie begrüßen?
- Ist das World-Wide-Web auch eine Textdatenbank?

# KORPUSTYPOLOGIE

1. Stadium der gespeicherten Sprache
2. Medium (Ursprungsmedium)
3. Repräsentativität
4. Größe
5. Sprache

# KORPUSTYPOLOGIE

## 1. Stadium der gespeicherten Sprache

2. Medium
3. Repräsentativität
4. Größe
5. Sprache

### 1.1 Synchrone Korpora

DeReKo, Wortschatz-Portal, DWDS, AAC

### 1.2 Diachrone Korpora

DWDS (20. Jh.), DeReKo (18. – 21. Jh.)

### 1.3 Historische Korpora

MHDBDB (MHD), Bonner FnhdC (FNHD), AHD??

# KORPUSTYPOLOGIE

1. Stadium der gespeicherten Sprache

## 2. Medium

3. Repräsentativität

4. Größe

5. Sprache

### 2.1 Korpora der geschriebenen Sprache

DeReKo, DWDS, Leipzig...

### 2.2 Korpora der gesprochenen Sprache

Bayerisches Archiv für Sprachsignale, DGD (auch Ex Freiburger Korpus)

# KORPUSTYPOLOGIE

1. Stadium der gespeicherten Sprache

2. Medium

## 3. Repräsentativität

4. Größe

5. Sprache

### 3.1 Stilistisches Kriterium

#### 3.1.1 allgemeine

DeReKo, DWDS

#### 3.1.2 spezifische

MULTEXT-East, FALKO

### 3.2 Kriterium der Ausgewogenheit

#### 3.2.1 ausgewogene Korpora

DeReKo

#### 3.2.2 opportunistisch gebildete Korpora

Wortschatz-Portal



# KORPUSTYPOLOGIE

1. Stadium der gespeicherten Sprache
2. Medium
3. Repräsentativität
4. Größe
5. Sprache

## 4.1 große

DeReKo, BNC, ČNK...

## 4.2 kleine

DWDS (nicht-authorisierter Teil), Deu-Cze

# KORPUSTYPOLOGIE

1. Stadium der gespeicherten Sprache
2. Medium
3. Repräsentativität
4. Größe

## 5. Sprache

### 5.1 monolinguale Korpora

eine Sprache: (Nationalkorpora)

### 5.2 Bilinguale und Multilinguale Korpora

zwei oder mehrere Sprachen

#### 5.2.1 Vergleichskorpora

#### 5.2.2 Parallelkorpora

OPUS, ČNPK, InterCorp...

# VERGLEICHSKORPUS

ULDaF /28296816/ x DWDS | Suchergebnisse x COSMAS II - Corpus Manage x Tochter bei Raststätte verg

wien.orf.at/news/stories/2548393/

Veřejné služby Inform... Park Uni-Vie Katalog-MUNI MZK Brno(BIB) UniTime 3.4| Log In jednojazyčný Park

## Tochter bei Raststätte vergessen

**Eine Wienerin hat ihre elfjährige Tochter bei der Heimfahrt aus dem Urlaub bei einer Raststätte in Ungarn vergessen. Das Mädchen wurde von der ungarischen Polizei betreut, die Mutter hatte ihr Fehlen erst in Wien bemerkt.**

Laut Bundeskriminalamt kam es durch einen Irrtum bei der Heimfahrt aus dem Urlaub zu dem Zwischenfall. Die Mutter ließ das elfjährige Mädchen demnach in der Nacht auf Samstag bei einer Raststätte in der Nähe von Budapest aussteigen und die Toilette aufsuchen. Als die Frau kurz darauf eine Autotür zuschlagen hörte, glaubte sie, dass die Tochter wieder eingestiegen sei und fuhr los, so das Bundeskriminalamt in einer Aussendung.

Das Mädchen wurde von einem Mann gefunden, der die Polizei in Ungarn informierte. Da das Handy des Mädchens im Auto der Mutter lag, wurde die Polizei in Wien verständigt.


### Strafrechtliche Prüfung eingeleitet

Die Mutter war in der Zwischenzeit in Wien angekommen und hatte bemerkt, dass ihre Tochter nicht im Auto war. Sie suchte die Polizeiinspektion in Wien-Margareten auf, zugleich kam die Mitteilung, dass das Bundeskriminalamt bereits Kontakt zur Tochter hatte.

Die Elfjährige verbrachte die restliche Wartezeit auf der Polizeistation in Budaörs und wurde von ihren Eltern in der Früh abgeholt. Trotz des glücklichen Endes muss der Sachverhalt laut Bundeskriminalamt zur strafrechtlichen Prüfung an die Staatsanwaltschaft Wien übermittelt werden.

**Link:**

- Bundeskriminalamt

Ganz Österreich 

- Wien-News
- Radio Wien  
Expertinnen,  
Veranstaltungen,  
Musikwünsche, Podcast,  
Sendungen
- Wien heute  
On demand & live
- Volksgruppen  
Vielfalt in Wien
- Landesstudio  
Kontakt

<http://ORF.at/>

DWDS | Suchergebnisse x COSMAS II - Corpus Manage x Tochter bei Raststätte verg

ahranicni/evropa/277762-rakusanka-ujela-200-km-nez-zjistila-ze-na-odpočivadle-nea

Uni-Vie Katalog-MUNI MZK Brno(BIB) UniTime 3.4| Log In jednojazyčný Park

## Novinky.cz

Hlavní stránka » Zahraniční » Evropa Podrubriky: **Evropa** Blízký a Střední východ

### Rakušanka ujela 200 km, než zjistila, že na odpočívadle nechala dítě

Rakušanka ujela 240 kilometrů z Budapešti do Vídně, než si všimla, že její 11letá dcera není ve voze. Zapomněla ji na dálničním odpočívadle u Budapešti. O případu, který se stal v pátek, informovala v pondělí agentura MTI.



[Zvětšit obrázek](#)

**pondělí 3. září 2012, 19:43** FOTO: tisková Profimedia

Maďarský řidič našel brzo ráno 11letou dívku na parkovišti u dálnice, nedaleko obce Budaörs za Budapešti. Dítě mu prozradilo, že maminka ji tam zapomněla a odjela. Muž okamžitě telefonoval rakouské policii. Dívka řekla policistům po telefonu své jméno a bydliště. Číslo matčina mobilního telefonu však nevěděla.

Matka zjistila, že dítě není v autě až po příjezdu do Vídně. Telefonovala na policii, kde ji uklidnili, že

<http://Novinky.cz>

# KORPUSTYPOLOGIE

- synchron – diachron – historisch
- gesprochene – geschriebene
- allgemein – spezifisch /
- ausgewogen – opportunistisch
- ❖ groß - klein
- ✓ monolinguale - multilinguale

# Wo sind die Korpora keine Hilfe?

- Kontrastive Forschung der gesprochenen Sprache
- Untersuchung der im Korpus nicht vorhandenen Texte (z.B. Aufschriften)
- Literaturwissenschaftliche Arbeiten im engeren Sinne (Interpretationen, Komparatistik)
- Aufstellung absoluter Urteile über die Sprache

Jedes Korpus ist durch seine Größe bzw. Ausgewogenheit der Texte begrenzt!

# KORPORA DES DEUTSCHEN

DWDS

<http://www.dwds.de/>

Wortschatz-Portal

<http://wortschatz.uni-leipzig.de/>

IDS-Mannheim (DeReKo, DGD u.a.)

<http://www.ids-mannheim.de/>

InterCorp

[Startseite](#)

Vor der Recherche bitte anmelden unter **Login** (rechts oben).

## **Korpora geschriebener Gegenwartssprache des IDS**

<http://www.ids-mannheim.de/kt/projekte/korpora/>

Bestand: über 4 Milliarden Token

## **C4**

<http://chtk.unibas.ch/korpus-c4/search>

### **DWDS-Korpus**

<http://www.dwds.de/>

Bestand: 1.2 Milliarden Token (intern), 100 Millionen Token (öffentlich)

### **Schweizer Textkorpus**

<http://www.schweizer-textkorpus.ch/>

Im Testbetrieb besteht das Korpus nun aus etwa 10 Mio. Token. Im Endausbau sind 20 Mio. Token geplant.

### **Austrian Academy Corpus:**

<http://www.aac.ac.at/index.html>

<http://corpus1.aac.ac.at/brenner/>

### **Südtirol**

<http://search.korpus-suedtirol.it:8089/cqpWeb/mycgi.pl>

## **Dortmunder Chatkorpus**

<http://www.chatkorpus.uni-dortmund.de/>

Bestand: 1.1 Mio. Token aus 150'000 Chat-Beiträgen.

## **British National Corpus (BNC)**

<http://www.natcorp.ox.ac.uk/>

Bestand: über 100 Millionen Token