



**Österreichische  
Akademie der Wissenschaften**



# **„Texts and Files“**

*Prof. Dr. Dmitrij Dobrovol'skij*

**WS 2004**



# **„Texts and Files“**

*Stephen Miller, M. A.*

**WS 2004**



# **„Texts and Files“**

*Prof. Dr. Peter Ďurčo*

**WS 2004**



## **Corpus-Anwendungen in der Sprachwissenschaft**

- 1. Wozu brauchen die Linguisten Textcorpora?**
- 2. Zum Forschungsstand**
- 3. Linguistische Forderungen an Corpora**
- 4. Semantische Fragestellungen**
- 5. Anwendung paralleler Textcorpora in semantischen Untersuchungen**



## **Corpus-Anwendungen in der Sprachwissenschaft**

- 1. Wozu brauchen die Linguisten Textcorpora?**
  - 1.1. Corpora und Theoriebildung**
  - 1.2. Corpora und lexikalische Semantik**
  - 1.3. Corpora und Lexikographie**
  - 1.4. Sprache als System vs. Sprache in Aktion**



# **Corpus-Anwendungen in der Sprachwissenschaft**

## **2. Zum Forschungsstand**

## **3. Linguistische Forderungen an Corpora**

– **3.1. Werkzeuge**

– **3.2. Was ist Lemmatisierung?**

– **3.3. Part-of-Speech-Tagging: Pro und Contra**

– **3.4. Problem des „optimalen Umfangs“**



## **Corpus-Anwendungen in der Sprachwissenschaft**

- 4. Semantische Fragestellungen**
- 5. Anwendung paralleler Textcorpora in semantischen Untersuchungen**



## **POS-Tagging (minimal):**

**„Sie braucht einen Narren, einen  
Narren, wie du einer bist, sie hat dich  
lange nicht gesehen, darum lädt sie  
dich ein!“**



# AUSTRIAN ACADEMY CORPUS

**<s>**

Sie PPER  
braucht VVFIN  
einen ART  
Narren NN  
, \$,  
einen ART  
Narren NN  
, \$,  
wie KOUS KON KOKOM  
du PPER  
einer PIS  
bist VAFIN  
, \$,

sie PPER  
hat VAFIN  
dich PPER PRF  
lange ADV ADJA  
nicht PTKNEG  
gesehen VVPP  
, \$,  
darum PAV  
lädt VVFIN  
sie PPER  
dich PPER PRF  
ein PTKVZ  
! \$.

**</s>**



## **POS-Tagging (maximal):**

**„Sie braucht einen Narren, einen  
Narren, wie du einer bist, sie hat dich  
lange nicht gesehen, darum lädt sie  
dich ein!“**



**<s>**

<b>Sie</b>	<b>PPER</b>	<b>Sie sie</b>	<b>f PPER</b>
<b>braucht</b>	<b>VVFIN</b>	<b>brauchen</b>	<b>f VVFIN VVIMP</b>
<b>einen</b>	<b>ART</b>	<b>ein</b>	<b>f ADJA ART PIS VVFIN VVINF</b>
<b>Narren</b>	<b>NN</b>	<b>Narr Narren</b>	<b>f NN</b>
<b>,</b>	<b>\$,</b>	<b>,</b>	<b>f \$,</b>
<b>einen</b>	<b>ART</b>	<b>ein</b>	<b>f ADJA ART PIS VVFIN VVINF</b>
<b>Narren</b>	<b>NN</b>	<b>Narr Narren</b>	<b>f NN</b>
<b>,</b>	<b>\$,</b>	<b>,</b>	<b>f \$,</b>
<b>wie</b>	<b>KOUS</b>	<b>wie</b>	<b>f KOKOM KON KOUS PWAV</b>
<b>du</b>	<b>PPER</b>	<b>du</b>	<b>f PPER</b>
<b>einer</b>	<b>PIS</b>	<b>ein</b>	<b>f ART PIS</b>
<b>bist</b>	<b>VAFIN</b>	<b>sein</b>	<b>f VAFIN</b>
<b>,</b>	<b>\$,</b>	<b>,</b>	<b>f \$,</b>



## Online Corpora:

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

<http://www.ids-mannheim.de/cosmas2/>