Tutorium der DGfS-Sektion Computerlinguistik

KORPUSLINGUISTIK MIT ONLINE-RESSOURCEN

EINE INTERAKTIVE EINFÜHRUNG FÜR LINGUISTEN

Annotation

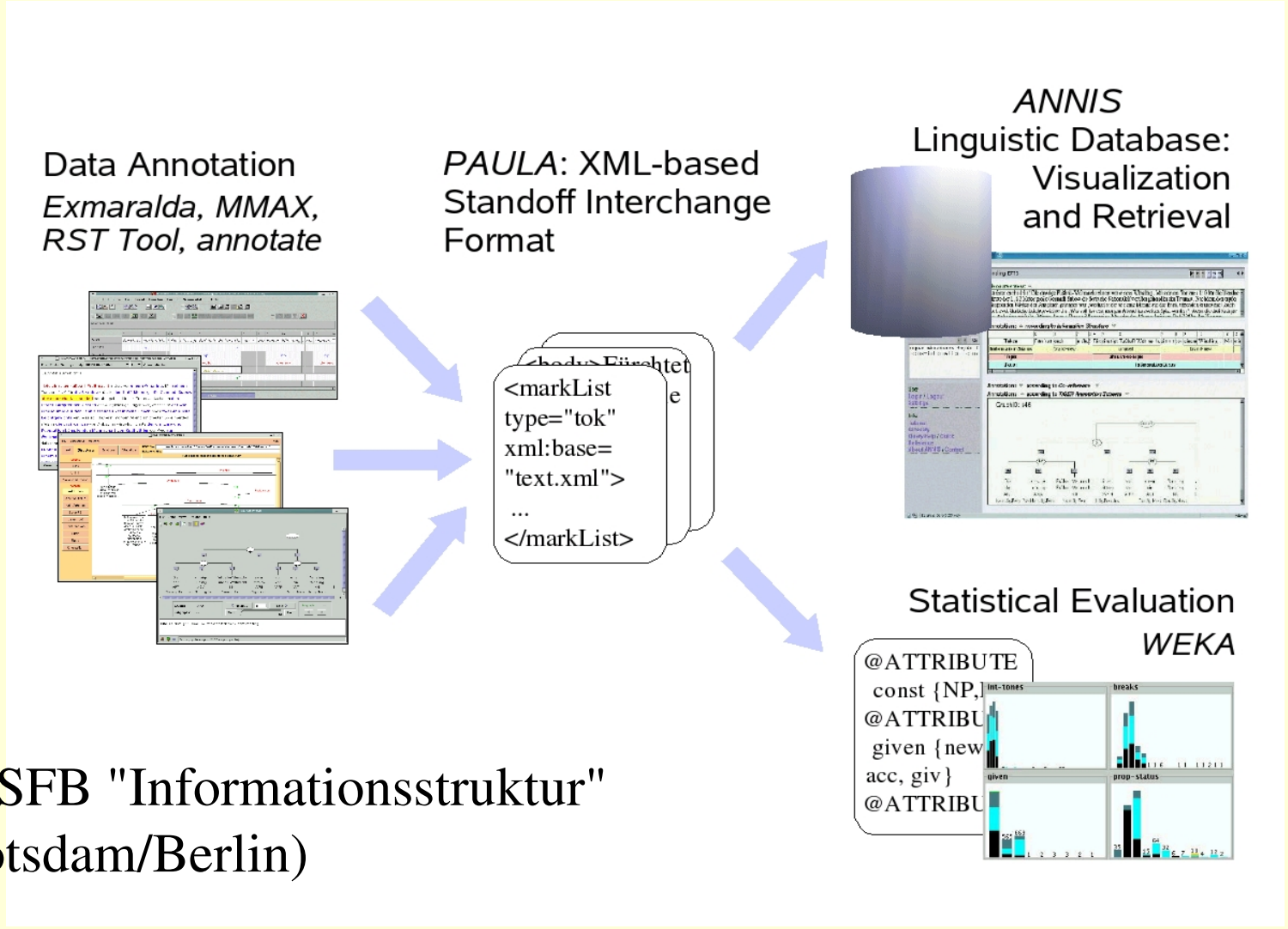
Stefanie Dipper
Stefan Evert
Heike Zinsmeister

München, 28.1.2011

Manuelle Annotation

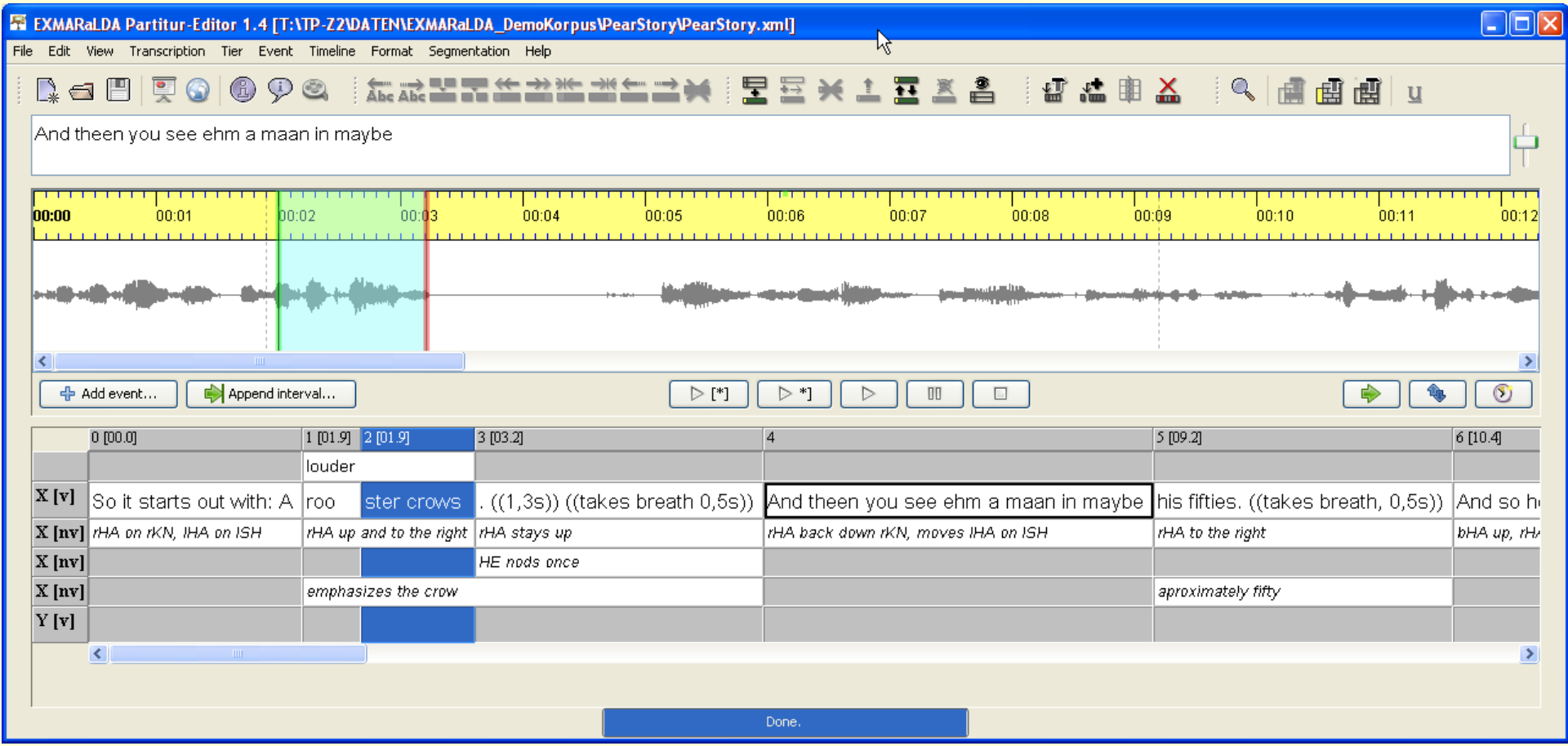
- komplexe Phänomene: manuelle Annotation
- Beispiel Informationsstruktur
  - komplexes Zusammenspiel vieler Faktoren
  - z.B. Informationsstatur, Topik, Fokus
- Idee: Annotation der einzelnen Faktoren erlaubt Untersuchung der Interaktion

PAULA und ANNIS



SFB "Informationsstruktur" (Potsdam/Berlin)

Exmaralda

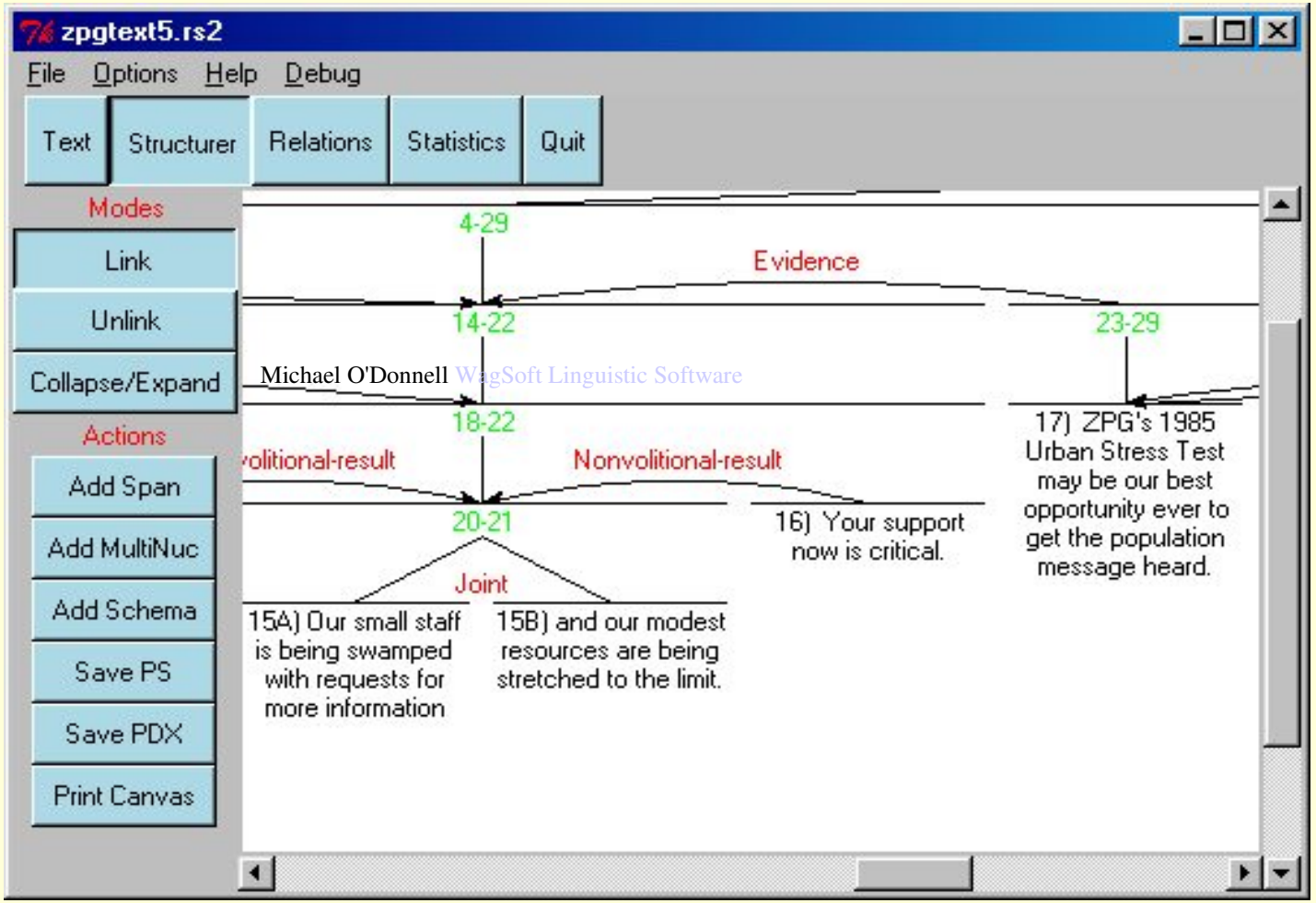


SFB "Mehrsprachigkeit" (Hamburg)

MMAX2

HITS (Heidelberg; früher EML)

RST Tool



Michael O'Donnell, WagSoft

Annotations-Richtlinien

Für die Interpretation und Wieder-Verwendbarkeit von annotierten Daten:

- Bedeutung der Annotation muss klar sein
  - Welche Symbole gibt es?
  - Wie sind sie definiert?
    - z.B. NN = normales Nomen
  - Kriterien für die Annotation
    - (linguistische) Tests für die Abgrenzung NN vs. NE

Inter-annotator agreement

- Problem: Annotierer können sich irren
- Methode zur Messung der Annotationsqualität:
  - Übereinstimmung zweier unabhäniger Annotatoren
  - *kappa*: Maß, das die Zufallsübereinstimmung mit berücksichtigt