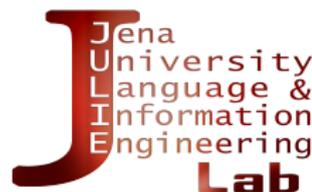


Annotation Time Stamps — Temporal Metadata from the Linguistic Annotation Process

Katrin Tomanek Udo Hahn

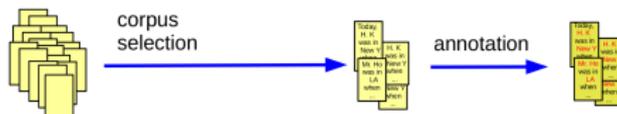
Jena Language & Information Engineering (JULIE) Lab
Friedrich-Schiller-Universität Jena, Germany

<http://www.julielab.de>



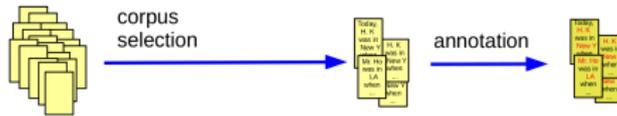
Economizing the Creation of Training Material

Standard Procedure

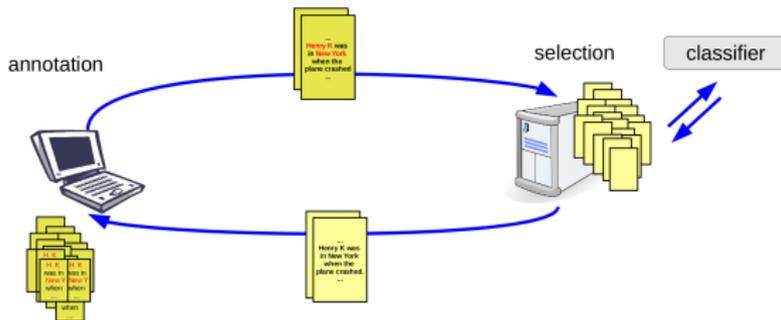


Economizing the Creation of Training Material

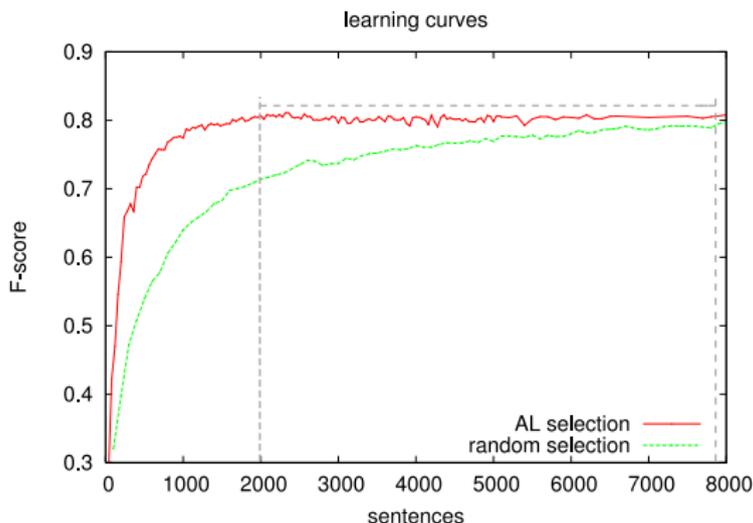
Standard Procedure



Active Learning



Evaluation of Active Learning



- “Does Active Learning really reduce annotation time ?”
- requires cost-sensitive evaluation of Active Learning
- but: how to **simulate** AL with true annotation cost?
→ corpus with annotation time stamps

The MUC7_T Annotation Project

- re-annotation of well-known corpus
 - MUC7 corpus (news-wire)
 - ENAMEX types (PER, LOC, ORG)
 - reproducible annotation guidelines
 - (hopefully) reasonably large for AL simulations
- store annotation time information for each annotation unit

Annotation Units

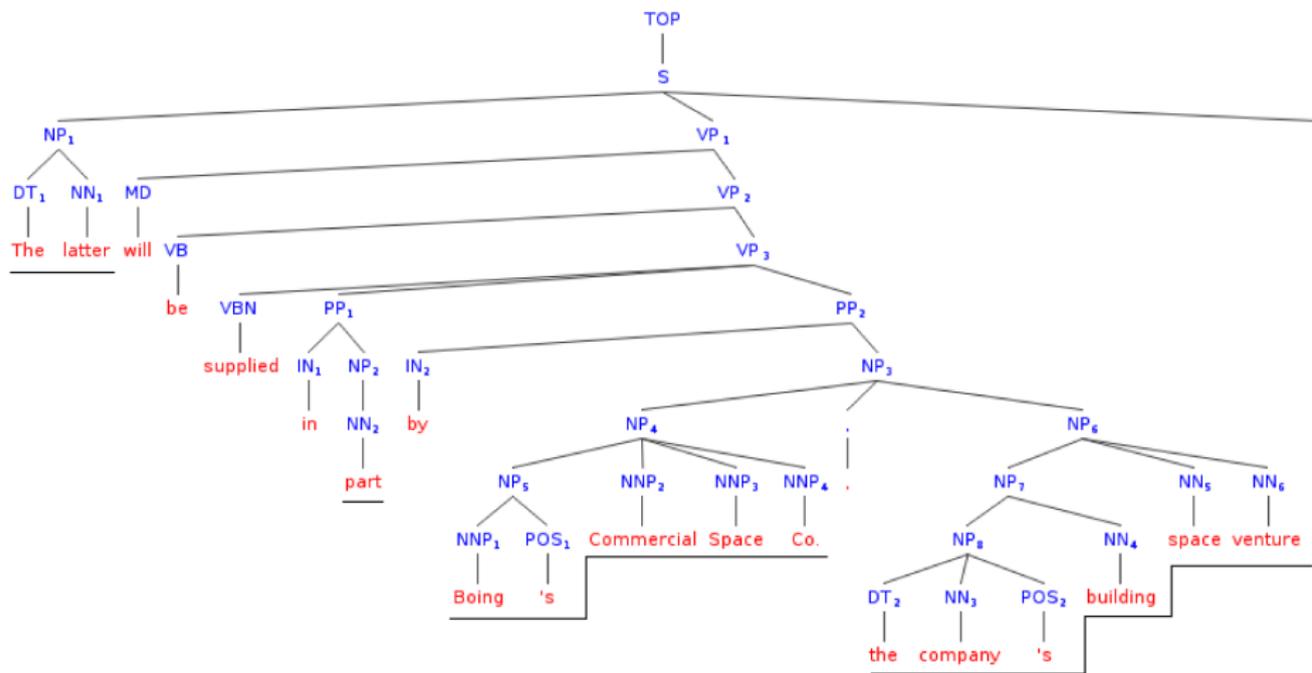
Sentences

- most natural linguistic unit
- might be too coarse for some applications

Complex Noun Phrases (CNPs)

- top-level NPs derived from sentence constituency structure
- by definition MUC7 entities occur within CNPs
- smallest syntactic unit completely covering entity mentions
 - 98.95% of MUC7's ENAMEX entities contained in CNPs
 - remaining 1.05% mostly due to parsing errors

Complex Noun Phrases



Annotation Principles

- one **annotation example** shown at a time
 - MUC7 document
 - single annotation unit (sentence or CNP) highlighted and annotatable
- annotation examples randomly shuffled
 - in order to guarantee independence of single annotations (avoid learning/synergy effects due to consecutive reading of a text)
- annotation in blocks of 500/100 annotation examples
 - to be annotated without breaks and under quiet noise conditions
 - to avoid exhaustion effects
- annotation GUI controlled by keyboard shortcuts
 - avoids “mechanical” annotation overhead
 - assumption: measured time reflects only cognitive process

Annotation Principles

4/14/10

File Extra

14 / 500

NEW YORK

the \$ 60,000 daily cost of rented salvage ships to the price of high-tech forensic equipment to police overtime , the investigation into the crash of Trans World Airlines Flight 800 is already the nation 's most expensive aircraft-accident investigation . The investigation 's full cost will not be known until it is completed , a process that is likely to take months . But federal and local officials said that expenses were rapidly approaching \$ 10 million , and a dispute has already erupted over whether TWA , its insurer and other private parties will help pay the bills . Earlier this week , National Transportation Safety Board officials sent a letter to TWA asking the airline for a voluntary payment of \$ 5 million toward salvage and recovery costs , according to Mark Abels , a company spokesman . But Abels said TWA 's position was that it bore no responsibility to help pay the government 's bills . " We think this is a government expenditure , " Abels said . Peter Goetz , a safety board spokesman , said that the agency expected that private parties would resist paying for the inquiry if it was determined that the crash of Flight 800 was caused by a criminal act . If not , the decision of how much , if any , individual companies will pay is largely up to them .

board officials estimated that the investigation was costing the agency more than \$ 100,000 a day , or \$ 3 million to date . The agency has an annual budget of \$ 38 million , but only about \$ 1 million of that is dedicated to accident investigations like the TWA case . The agency 's resources have also been taxed by a other recent disasters , including the crash of a ValuJet plane into the Florida Everglades in May . " Certainly , this is the most expensive accident investigation that we have encountered , " said Bernard Loeb , the director of the safety board 's office of aviation safety . A safety board spokesman said that before the TWA inquiry , the agency 's previous largest expenditure probably involved **the investigation of the 1994 crash of a USAir jet upon its approach to Pittsburgh** . He said that inquiry had cost \$ 1 million . Loeb said the agency often turned to private parties like an airline 's insurer or a plane 's manufacturer to contribute to an investigation 's salvage costs . But the chairman of TWA 's insurer disputed the notion that such contributions were routine and said that an airline 's insurer typically paid only for the cost of environmental cleanups at a crash site . " The government is obviously expending a lot of effort and cost ,

☰ 🔍 ↺ ↻

Annotation Principles

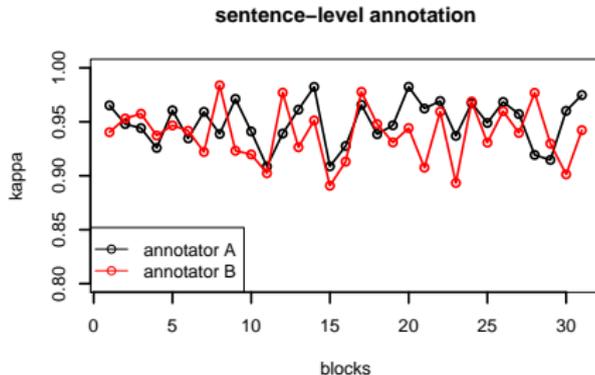
- one **annotation example** shown at a time
 - MUC7 document
 - single annotation unit (sentence or CNP) highlighted and annotatable
- annotation examples randomly shuffled
 - in order to guarantee independence of single annotations (avoid learning/synergy effects due to consecutive reading of a text)
- annotation in blocks of 500/100 annotation examples
 - to be annotated without breaks and under quiet noise conditions
 - to avoid exhaustion effects
- annotation GUI controlled by keyboard shortcuts
 - avoids “mechanical” annotation overhead
 - assumption: measured time reflects only cognitive process

Annotators

- 2 students with good general English skills
- original MUC7 guidelines
- extensive training on MUC7 test set
- final annotations: MUC7 set on airplane crashes completely annotated by both annotators

Annotation Performance

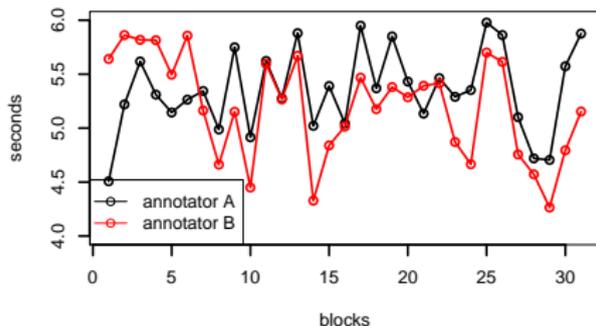
- agreement with Muc7 annotations
 - $\kappa_A = 0.95$ and $\kappa_B = 0.96$
 - $F_A = 0.92$ and $F_B = 0.94$
- both annotators perform similarly on different blocks
- annotation performance largely stationary



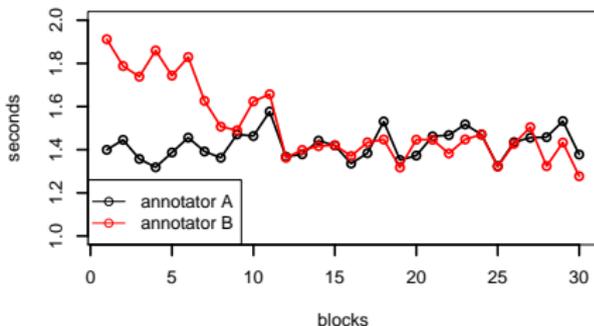
Annotation Time Measurements

- times similar for both annotators
- learning effect for annotator B up to block 9
- quite stationary annotation times

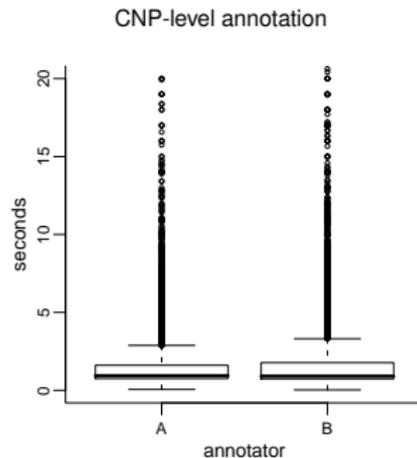
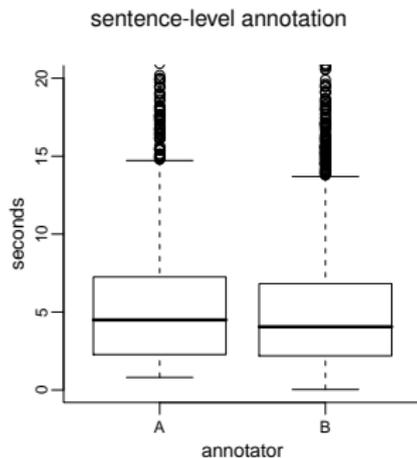
sentence-level annotation



CNP-level annotation



Variability of Annotation Times



- annotation times subject to high variance
- confirms findings of Settles et al. 2008 and Ringger et al. 2008

Application of MUC7_T

Cost-sensitive Evaluation of Active Learning

Experimental settings

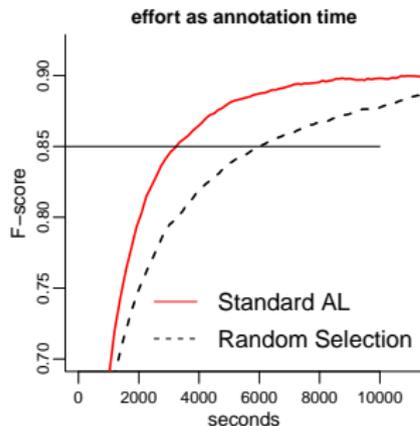
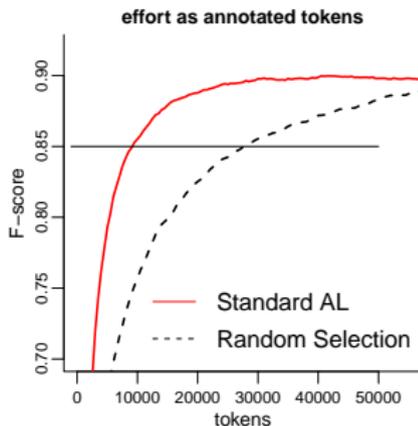
- Named Entity Recognition with Conditional Random Fields

$$P_{\theta}(\vec{y}|\vec{x}) = \frac{1}{Z_{\theta}(\vec{x})} \cdot \prod_{i=1}^n \exp \left(\sum_{j=1}^k \lambda_j f_j(y_{i-1}, y_i, \vec{x}, i) \right)$$

- straight-forward approach to Active Learning
 - Uncertainty Sampling
 - utility function

$$u(\theta, \vec{x}) = 1 - \max_{\vec{y}' \in \mathcal{Y}} P_{\theta}(\vec{y}'|\vec{x})$$

Evaluation



Savings of Active Learning over Random Selection

- number of tokens: 67.8 %
- annotation time: 44.0 %

Summary

- MUC7_T is a corpus with information on annotation time
 - annotation time stamps are new breed of linguistic metadata
 - coarse- and fine-grained time measurement
- allows for more realistic evaluation of selective sampling strategies (e.g., Active Learning)
- currently also used to learn predictive annotation cost models (ongoing research, see Tomanek et al. 2010)

<CA>LBC



<http://www.julielab.de>