Approximating Learning Curves for Active-Learning-Driven Annotation

Katrin Tomanek and Udo Hahn

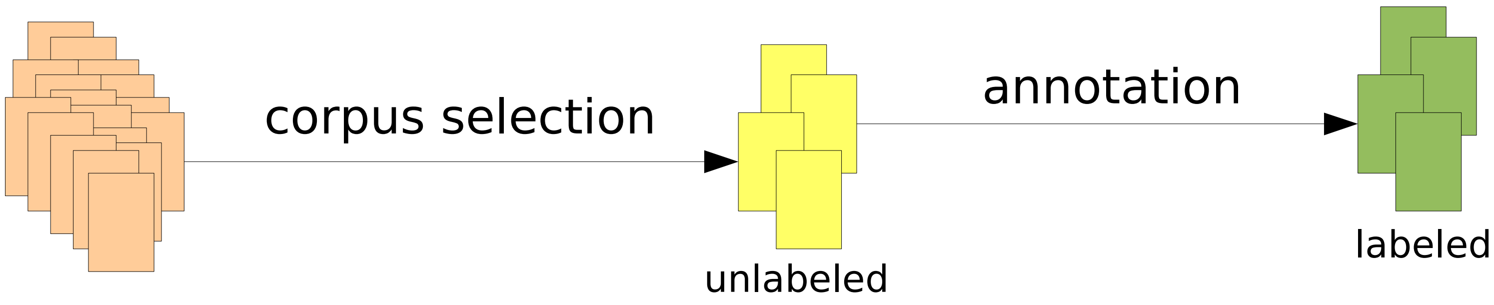
Jena University Language and Information Engineering (JULIE) Lab

Agenda

- Introduction to Active Learning
- Stopping Conditions
- Experiments & Results

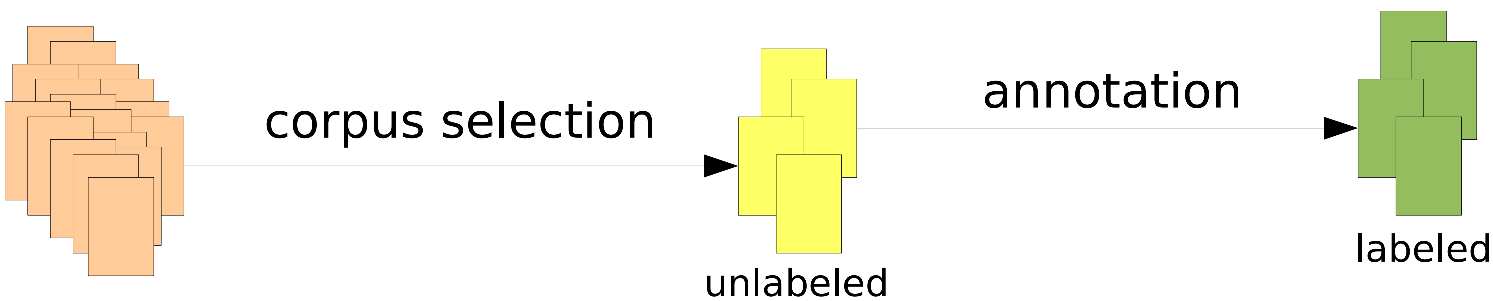
Passive versus Active Selection

passive annotation scenario (aka Random Sampling)

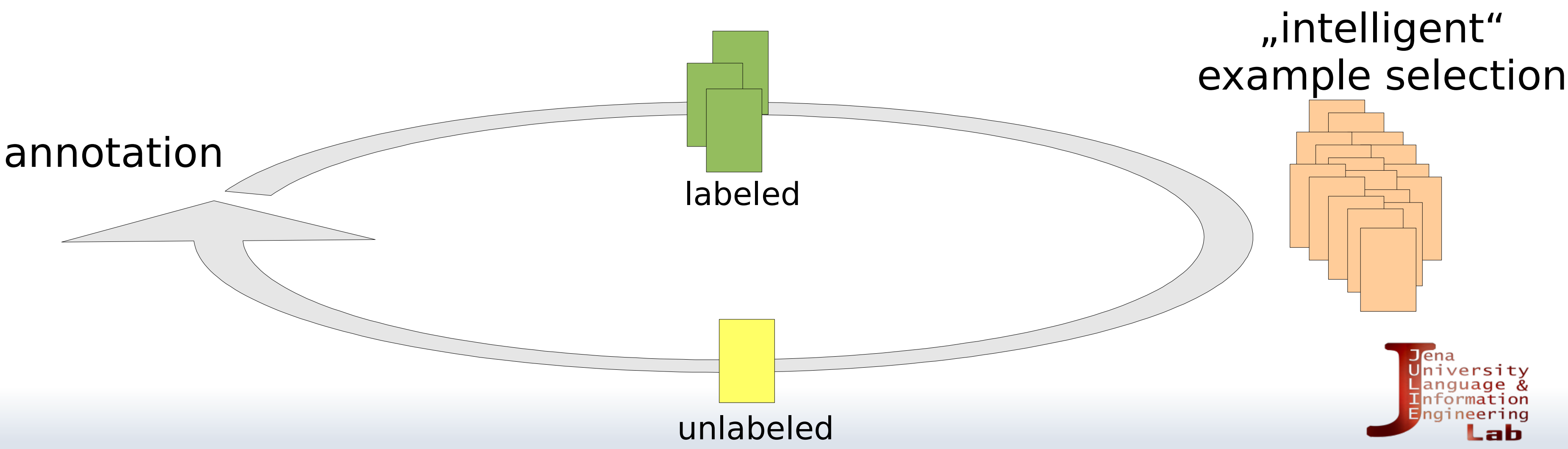


Passive versus Active Selection

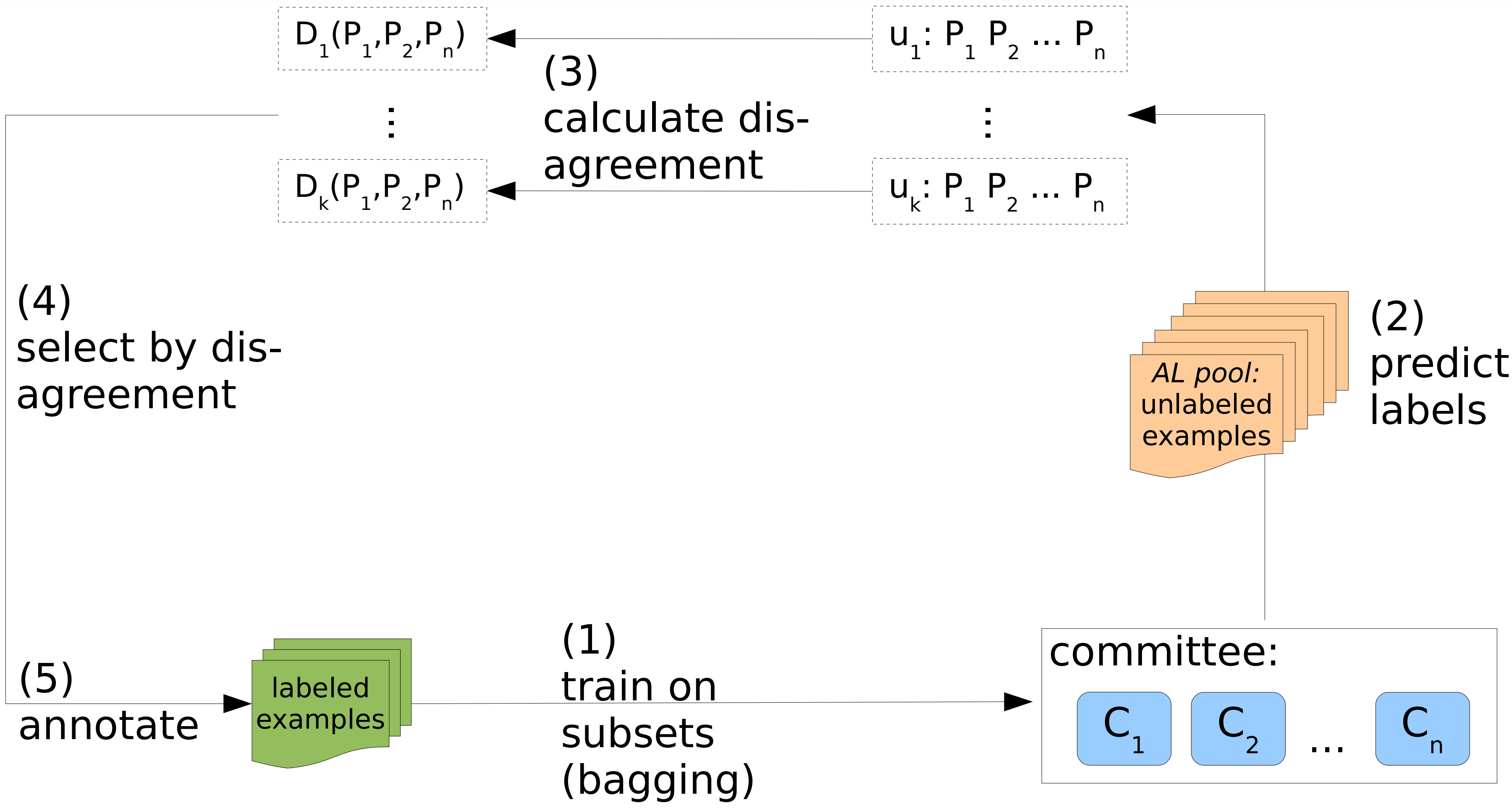
passive annotation scenario (aka Random Sampling)



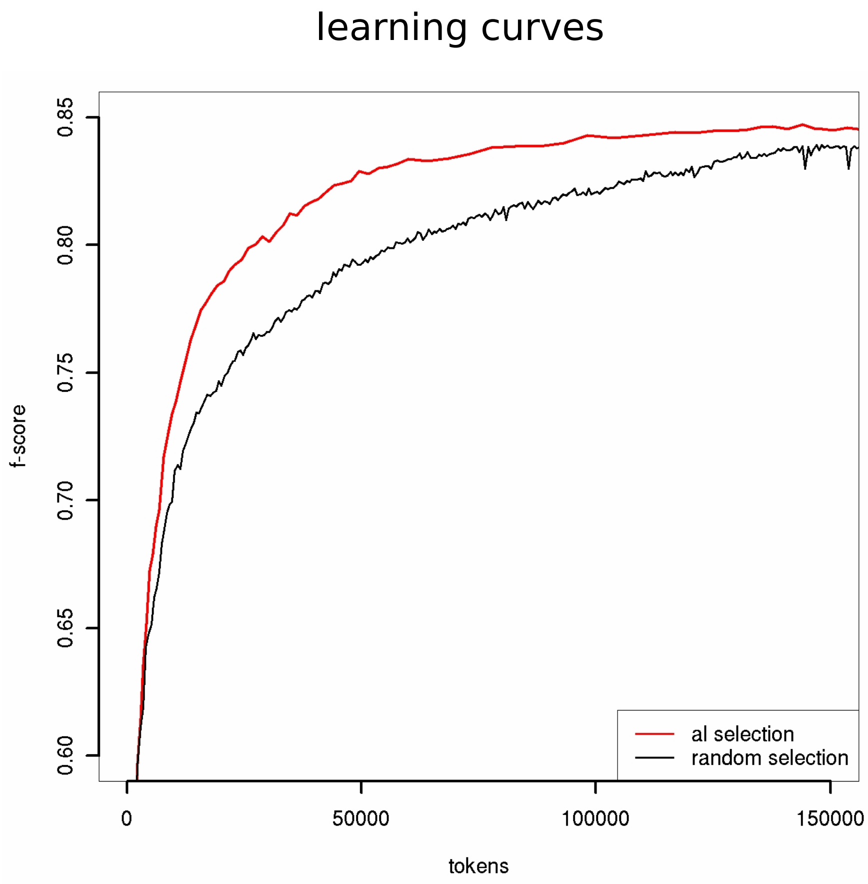
active annotation scenario (aka Active Learning)



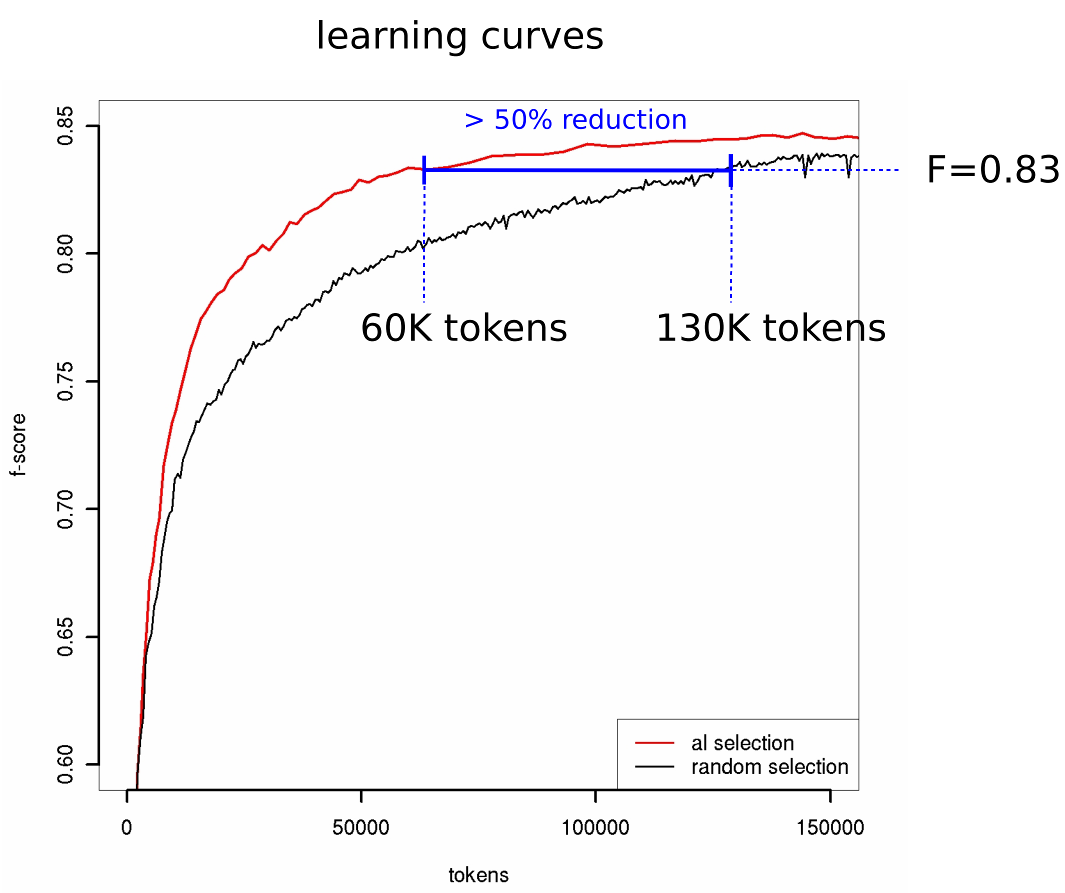
Committee-based AL Framework



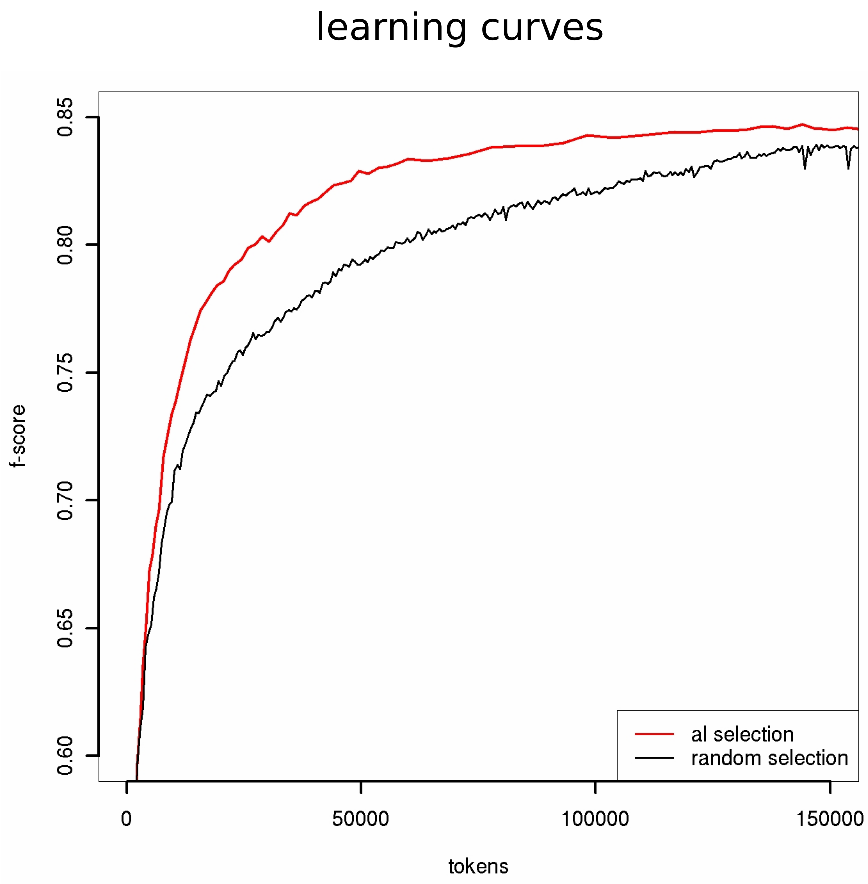
Reduction of Annotation Effort



Reduction of Annotation Effort



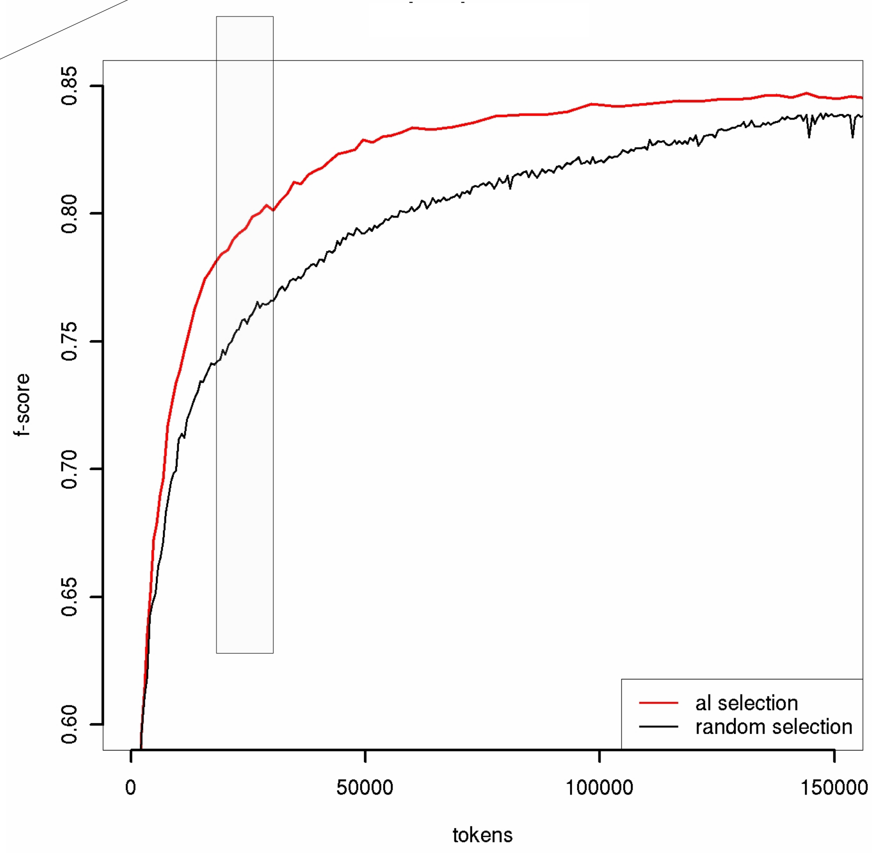
When to Stop the Annotation ?



When to Stop the Annotation ?

bad

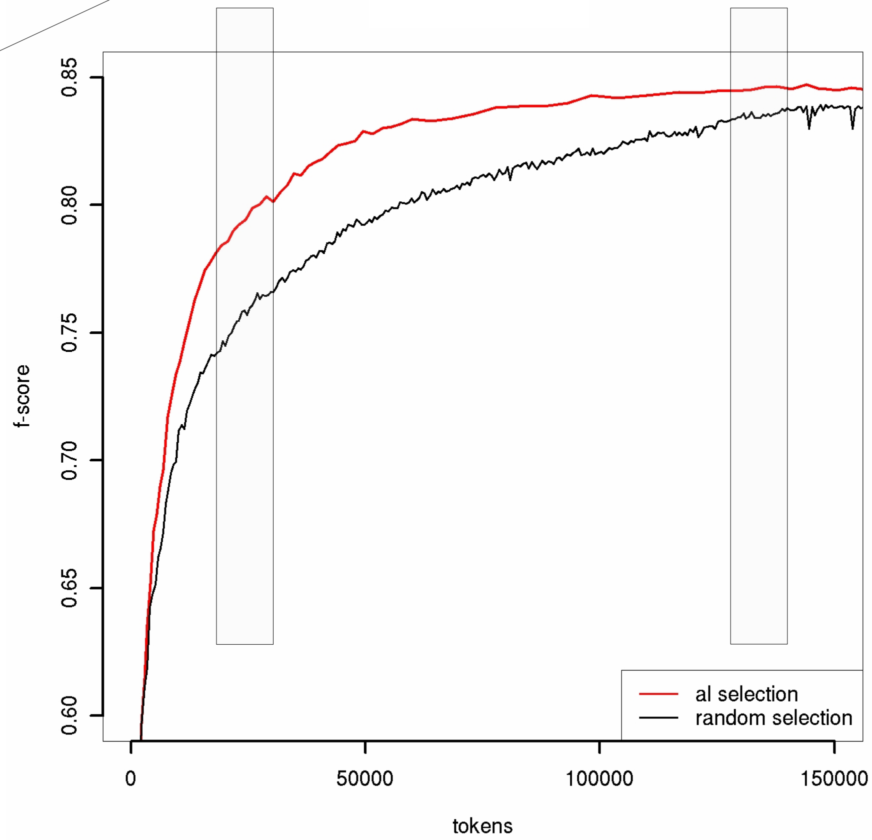
Could gain a lot
by further annotation



When to Stop the Annotation ?

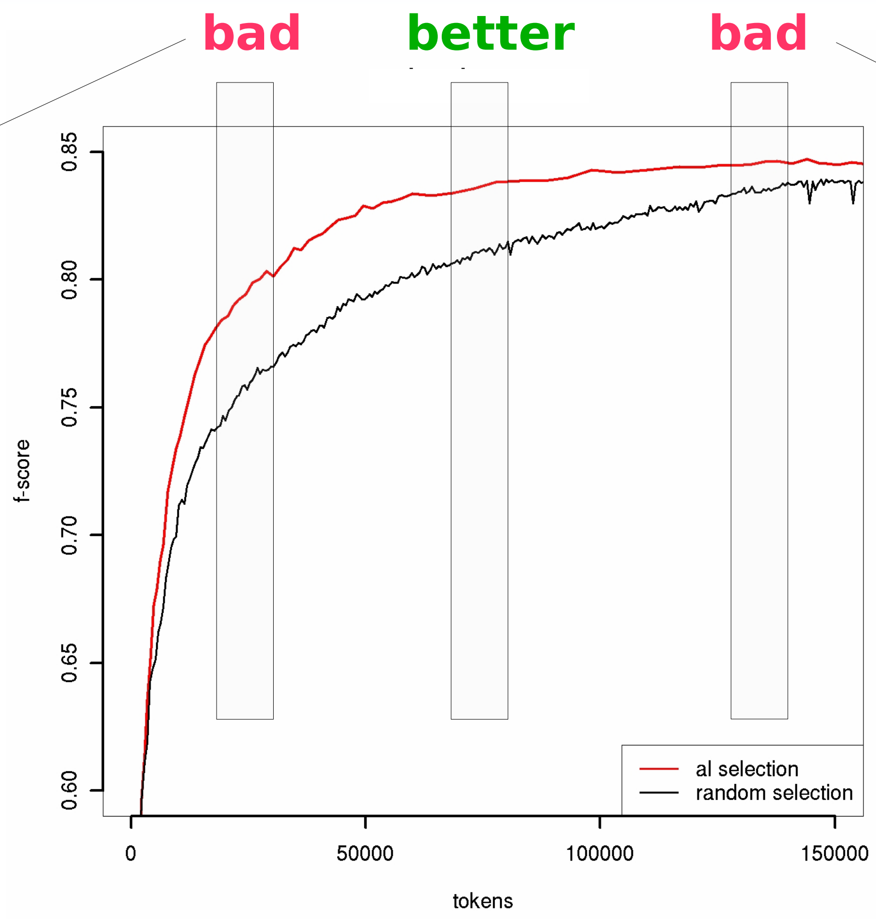
bad **bad**

Could gain a lot
by further annotation



Further annotation will not
increase classifier
performance significantly

When to Stop the Annotation ?



Could gain a lot
by further annotation

Further annotation will not increase classifier performance significantly

Jena University Language & Information Engineering Lab

Stopping Condition based on Learning Curve ?

- Pro: stopping condition directly based on classifier performance
- Contra: requires **labeled** gold standard
- ➔ not applicable in practice as gold standard not available

- Goal:
  - Estimate the (progression of) learning curve without need for gold standard

Approximating the Learning Curve

- Approach:
  - Based on agreement among committee members
  - Does not require extra labeling effort
  - Agreement curve approximates *progression* of learning curve
  - ➔ We can tell relative position in annotation process from it:
    - *relative* trade-off between annotation effort and gain in classifier performance from it
      - Steep slope ?
      - Convergence ?

Approximating the Learning Curve

- Intuition:
  - Agreement among committee:
    - Low in early AL iterations
    - High in later ones
    - ➔ When agreement among committee members converges, also learning curve does

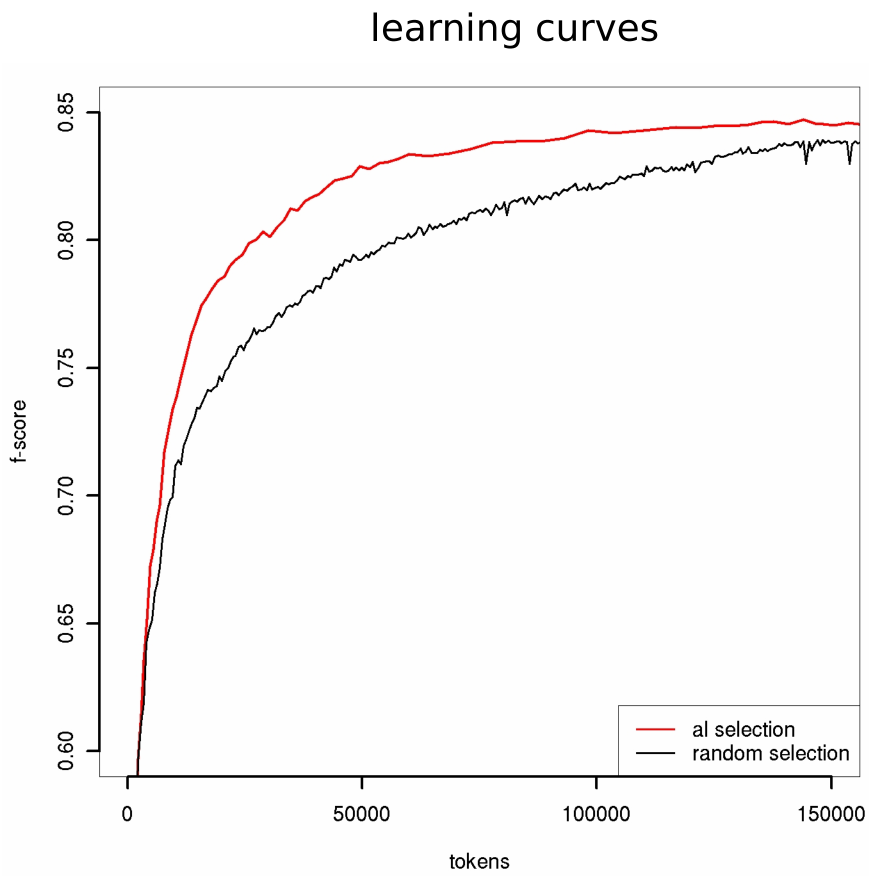
Approximating the Learning Curve

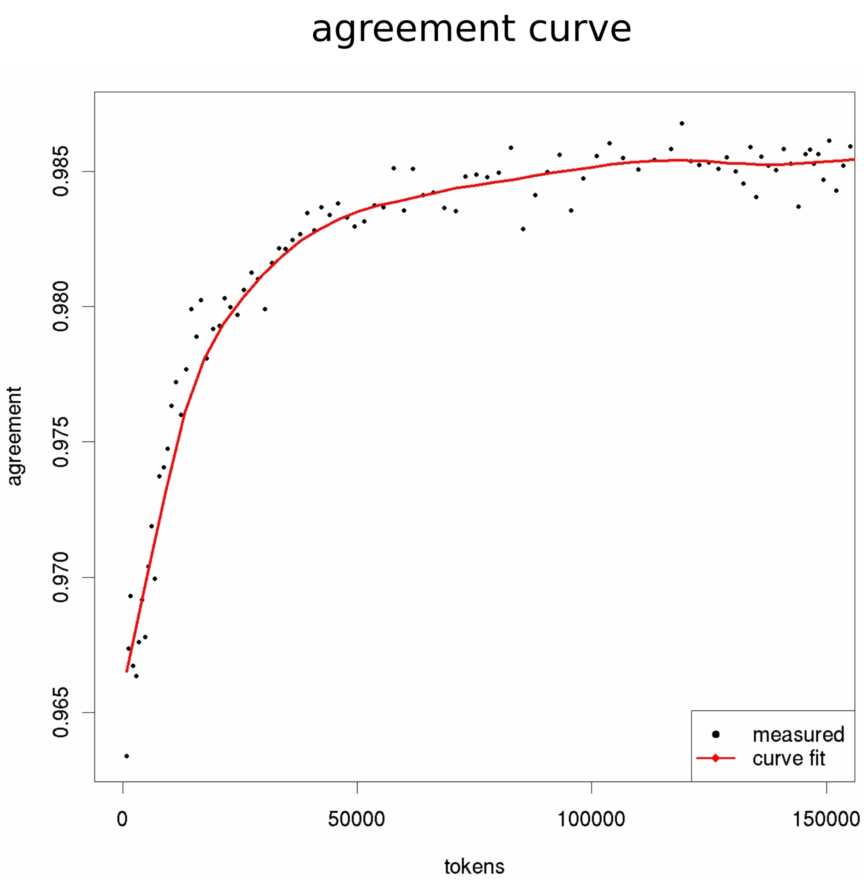
- Where to calculate the agreement:
  - On separate *validation set*
    - Not be involved in AL selection process itself
    - Agreement values comparable over different AL iteration
  - Otherwise agreement curve often not reliable approximation due to „simulation dilemma“
    - When e.g. agreement calculated on examples selected in each AL iteration:
      - Approximation of learning curve usually works well in simulation scenarios, because...
        - » few hard cases left in later AL iterations (perfect agreement)
      - But fails in real-world annotation scenarios, because...
        - » in practice AL will always find tricky cases...

Experiments

- For annotation of Named Entity mentions
- Whole sentences selected (20 each round)
- Simulation on CoNLL-2003 corpus
  - News-paper, MUC entities (PERS, LOC, ORG)
  - AL pool: ~ 14,000 sentences
  - Gold Standard: ~ 3,500 sentences
    - For learning curve
    - For agreement curve (labels ignored)

Results


learning curves
f-score
0.60
0.65
0.70
0.75
0.80
0.85
0
50000
100000
150000
tokens
al selection
random selection


agreement curve
agreement
0.965
0.970
0.975
0.980
0.985
0
50000
100000
150000
tokens
measured
curve fit

Summary & Conclusions

- AL has high potential to reduce annotation effort
- Proper stopping point necessary to profit from savings
  - → Method to monitor progress of annotation needed
- Agreement curve
  - Works well: good approximation of learning curve
  - No extra annotation effort: does not require labeled gold standard

Jena University Language & Information Engineering Lab

Approximating Learning Curves for Active-Learning-Driven Annotation

Thanks. Questions ?

http://www.julielab.de/