Assessing the benefits of partial automatic pre-labelling for frame-semantic annotation

Ines Rehbein, Josef Ruppenhofer & Caroline Sporleder

FEAST 2009

Outline

Outline

Outline

Outline

Outline

Motivation

- Linguistic resources with high-quality manual annotations are a backbone of many supervised NLP scenarios
- Manual annotation of linguistic resources is time-consuming and costly
- How can we annotate a large amount of data and still get good quality?

Can partial automatic pre-labelling speed up the annotation process without sacrificing annotation quality?

Outline

Frame Semantics (Fillmore 1976, 1977, ...)

- **Semantic frames**
  - are **schematic representations of situations** involving various participants, propositions, and other conceptual roles, each of which is called a frame element (FE)
  - The situations include events, states, and relations
  - Some frames also focus on entities/things
- Frames are connected to each other via **frame-to-frame relations** (e.g. Inheritance (is-a), Perspective on, Subframe, Using, ...)

Frame Semantics (Fillmore 1976, 1977, ...)

Example: Self_motion Frame

- Frame Evoking Elements:
  advance.v, climb.v, crawl.v, hike.v, hike.n, swim.n, ...
- Core Frame Elements:
  Area, Direction, Goal, Path, Self_mover, Source
- Non-core Frame Elements:
  Co-theme, Depictive, Duration, Manner, Time, ...

[Many others $^{Self\_mover}$] RUSHED [back ^{Goal}] [Wednesday morning ^{Time}]

Frame Semantic Annotation

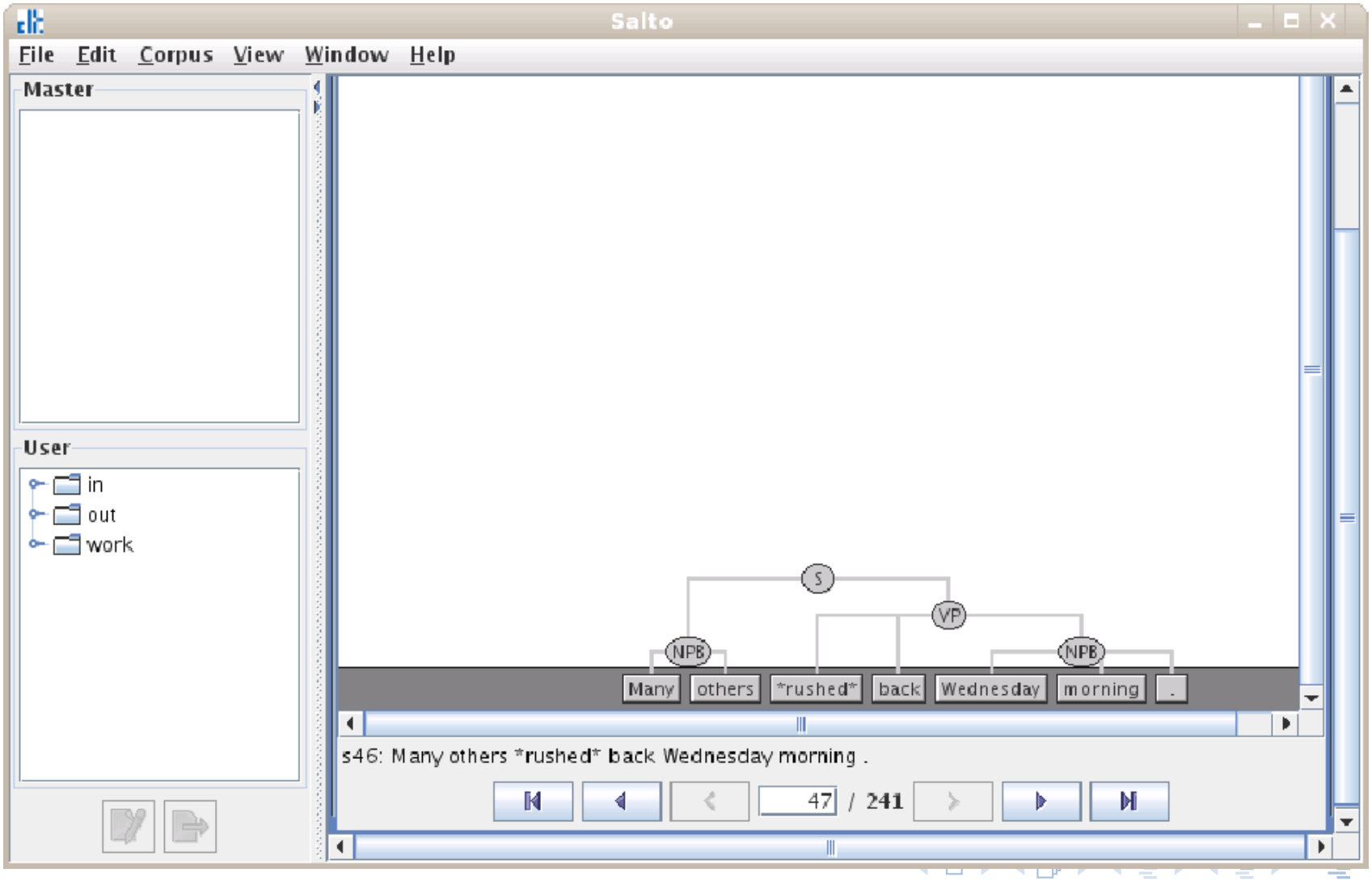
- Full-text
  - exhaustive annotation of running text with all different frames and roles that occur in the document
- Lexicographic annotation
  - annotation of instances of particular target words used in particular frames

Outline

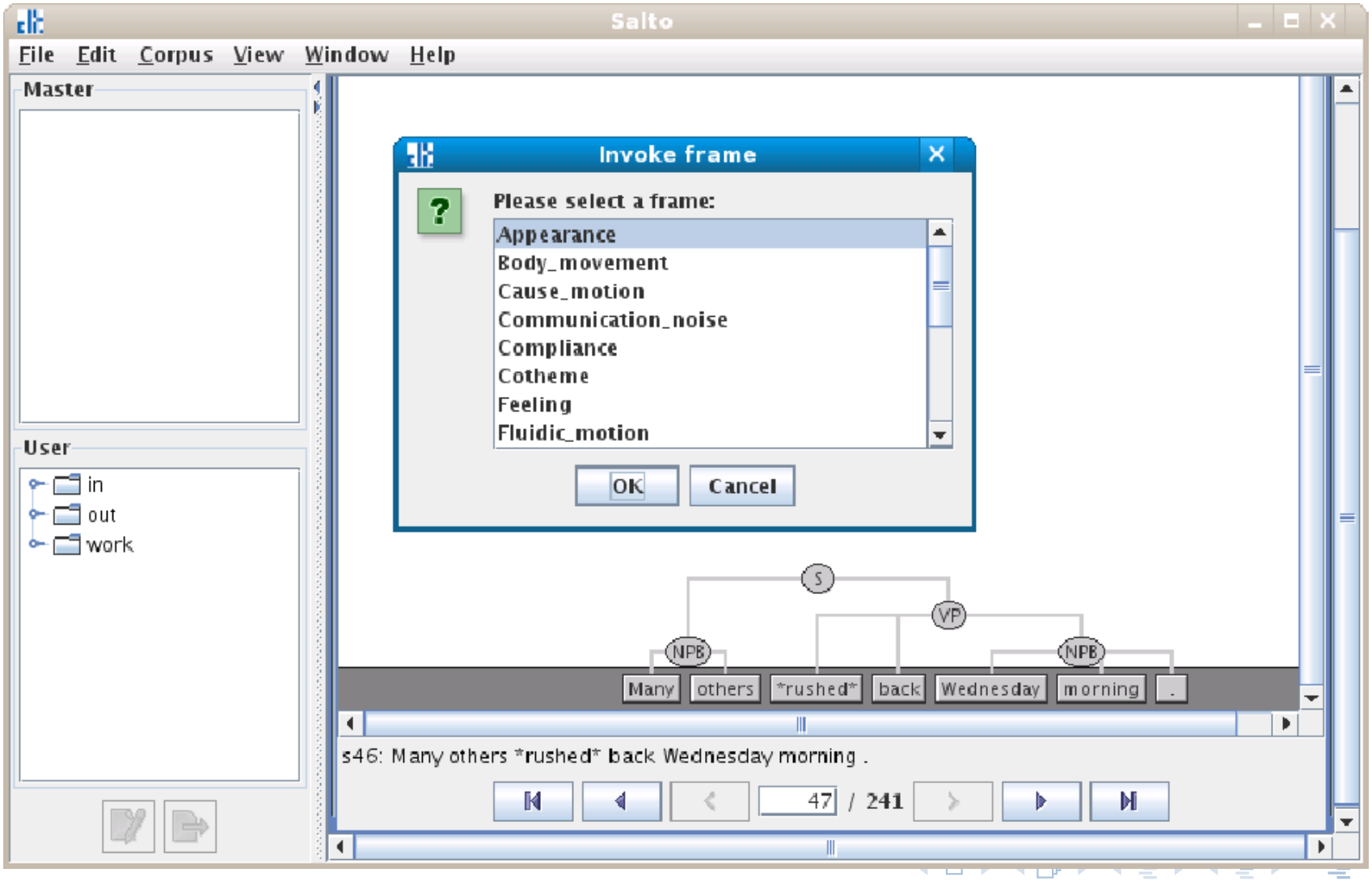
Annotation Set-Up

- **Lexicographic annotation** of FrameNet data
- **6 Annotators** (authors + 3 computational linguistics undergraduates with at least 1 year experience in frame-semantic annotation)
- **Annotation process**: decorating automatically derived syntactic constituency trees with semantic role labels using Salto *(Burchardt et al., 2006)*
  1. **Frame assignment**: choosing the correct frame for a target lemma from a pull down menu
  2. **Role assignment**: draw the available frame element links to the appropriate syntactic constituent(s)

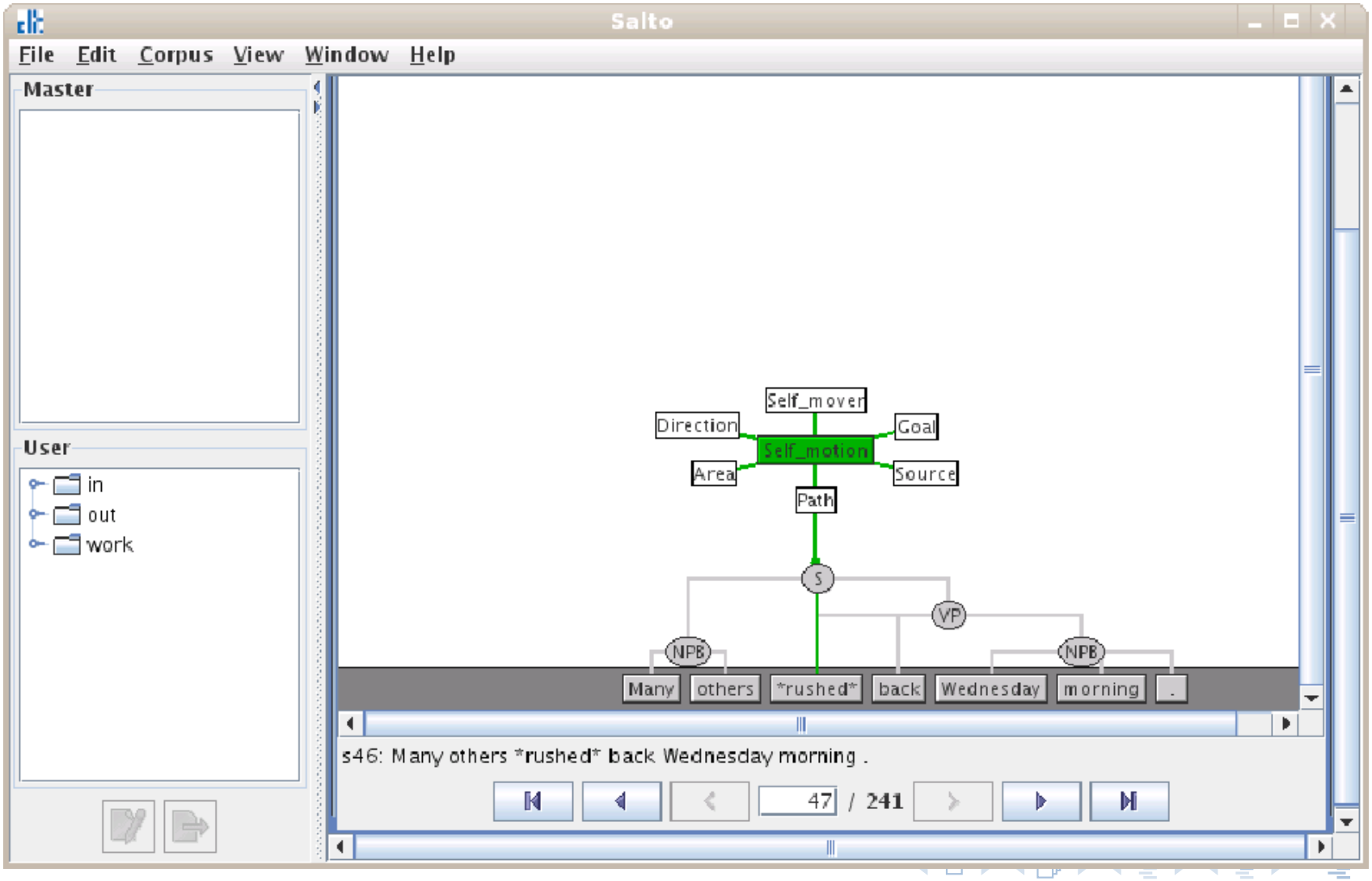
Frame Semantic Annotation with Salto (1)



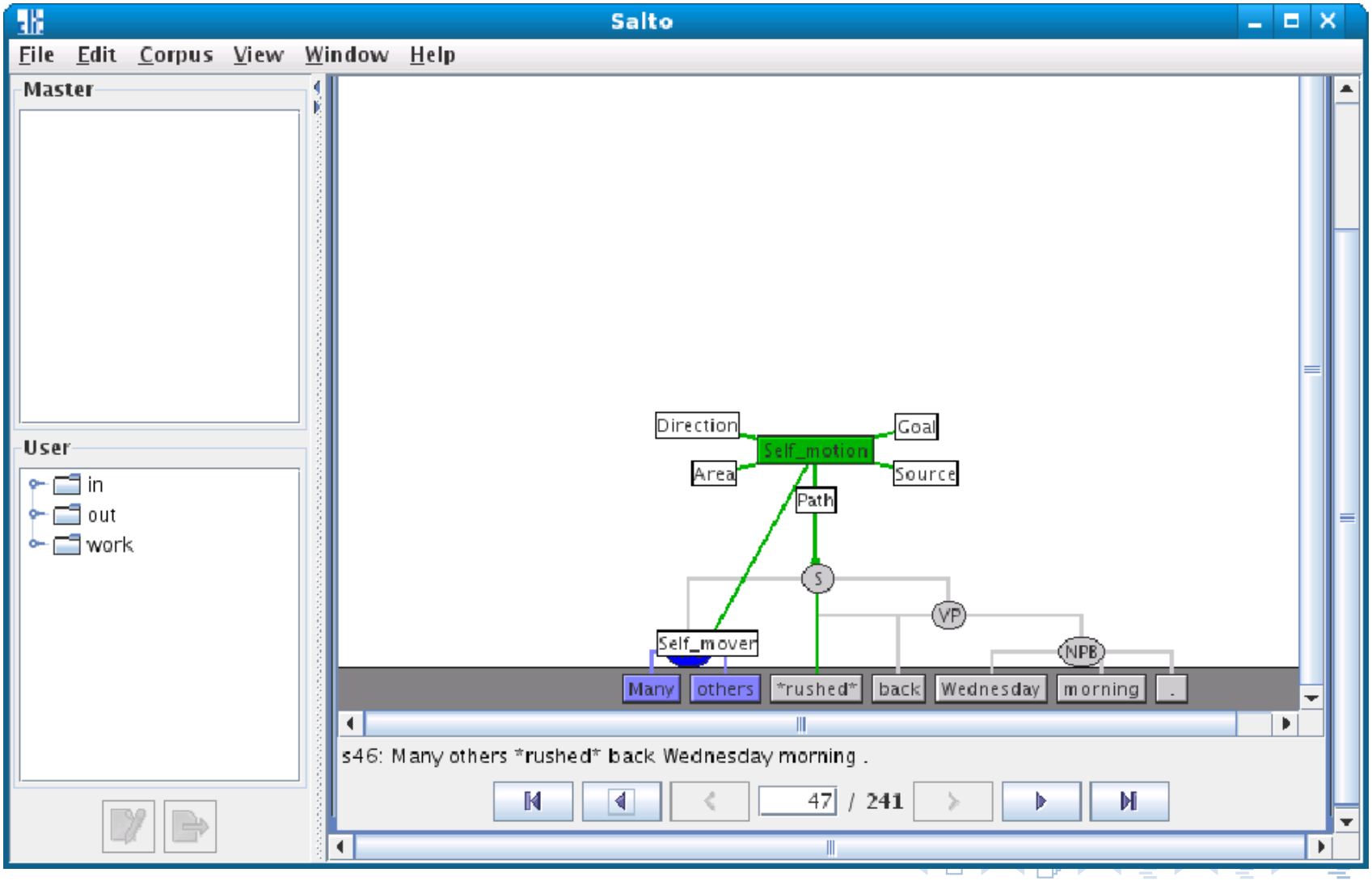
Frame Semantic Annotation with Salto (2)



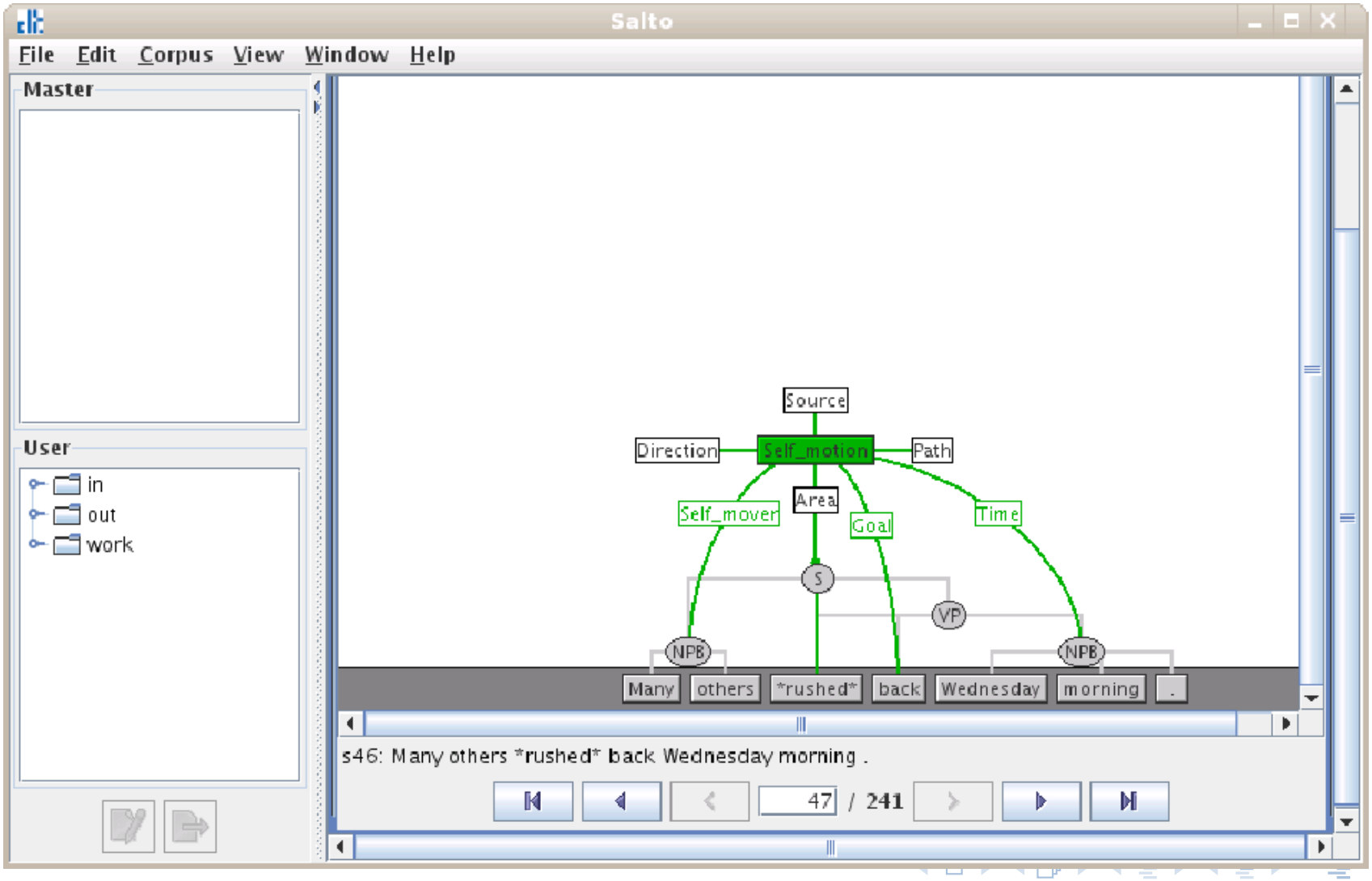
Frame Semantic Annotation with Salto (3)



Frame Semantic Annotation with Salto (4)



Frame Semantic Annotation with Salto (5)



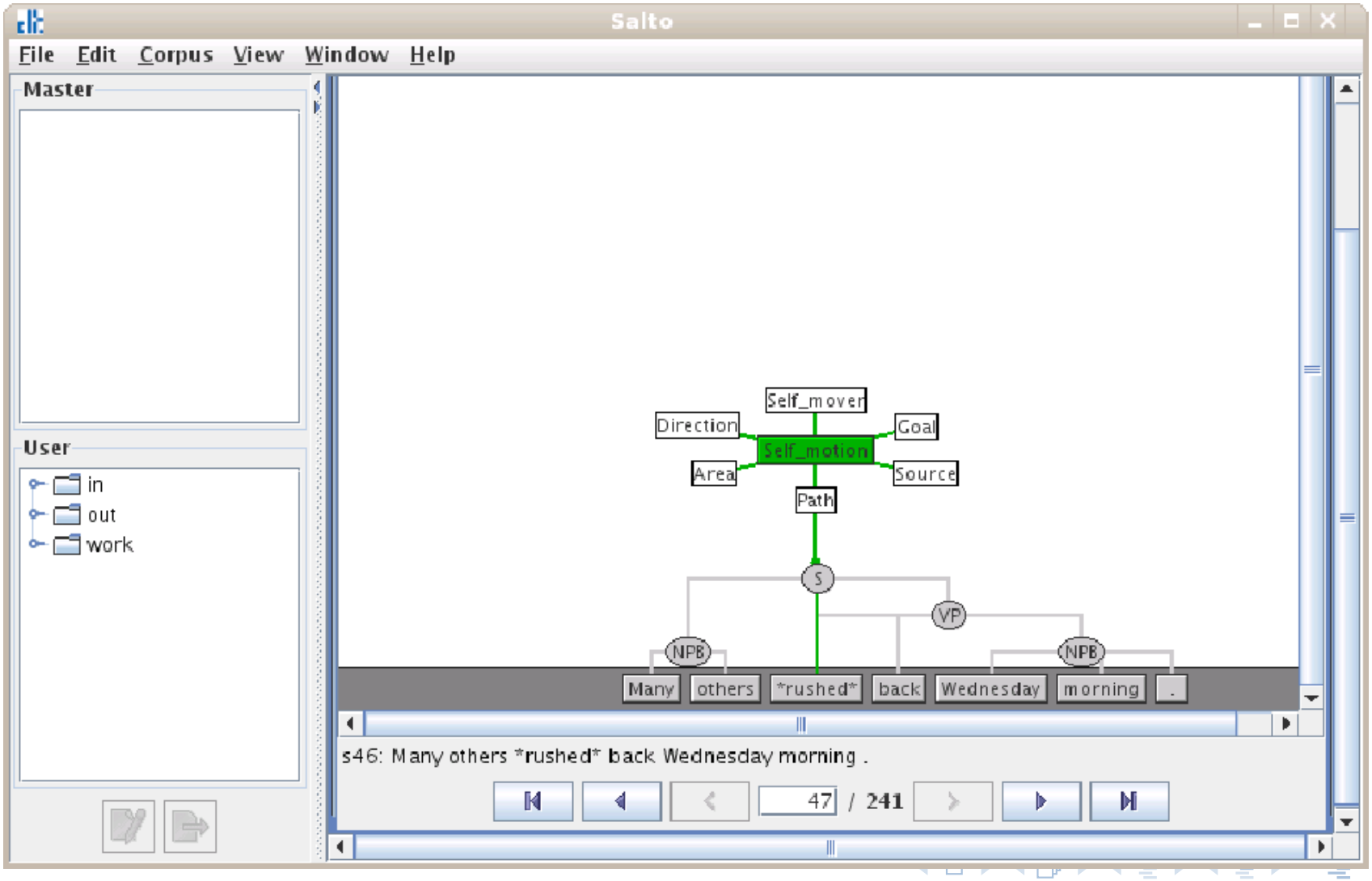
Data

- 360 FrameNet sentences (BNC) exemplifying all the senses defined for 6 different lemmas in FrameNet 1.3

	Instances	Senses
feel	134	6
follow	113	3
look	185	4
rush	168	2
scream	148	2
throw	155	2

- 3 random sets of equal size (120 sentences each)
- 3 versions of each set:
  No pre-annotation, **S**tate-of-the-art, **E**nhanced

Automatic Pre-Annotation of Frame Assignment



Study design

- Assignment of the 6 annotators to 3 groups of 2 (Group I-III)
- Each annotator experiences all 3 annotation conditions (**N**o pre-annotation, **S**tate-of-the-art, **E**nhanced)
- Order of annotation condition varies between Groups I-III

	1st	2nd	3rd	Annotators
Group I	E	S	N	5, 6
Group II	S	N	E	2, 4
Group III	N	E	S	1, 3

Table: Annotation condition by order and group

- Training sequence to rule out difficulties with unfamiliar frames and frame elements:
  Total of 240 sentences exemplifying all 6 verbs in all their senses

Data Analysis

- Measures:
  - Precision, Recall, F-score for frame assignment against FrameNet gold standard
  - Annotation time for each text segment
- Analysis of Variance (ANOVA)
  - impact of automatic pre-annotation on annotation time
  - impact of automatic pre-annotation on annotation quality (f-score)

Outline

Can pre-annotation of frame assignment speed up the annotation process?

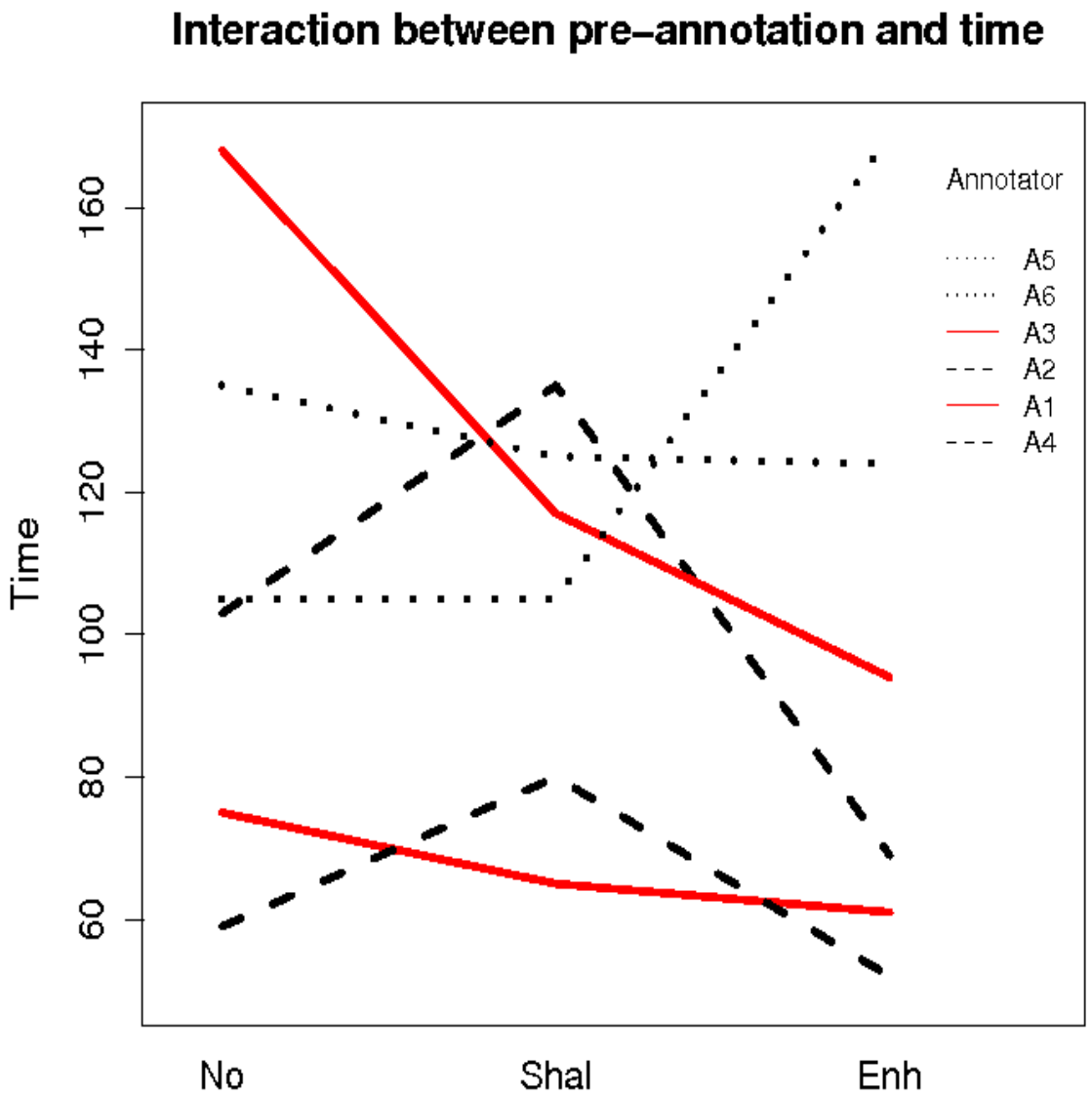
- 2-way ANOVA (within-subjects design), crossing the dependent variable (time) with the order of text segments and condition of pre-annotation

No significant influence of pre-annotation on annotation time (but 5 out of 6 annotators were faster on the text segment with **E**nhanced pre-annotation)

Can pre-annotation of frame assignment speed up the annotation process? (2)

- Order of text segments has significant influence on time requirement: all but 1 annotator needed most time for the text segment given to them first ($p \leq 0.05$)
  $\rightarrow$ **ongoing training effect**

Interaction between training effect and pre-annotation might prevent significant effect of pre-annotation on annotation time



Order	Annot.
N E S	1, 3 ———
S N E	2, 4 - - -
E S N	5, 6 .....

Is annotation quality influenced by automatic pre-annotation?

- 2-way ANOVA (within-subjects design), crossing the dependent variable (f-score) with the order of text segments and condition of pre-annotation

Significant effect ($p \leq 0.05$) for impact of pre-annotation on annotation quality

- All annotators achieved higher quality on **E**nhanced pre-annotated text segments
- 4 out of 6 annotators achieved higher quality on **S**tate-of-the-art pre-annotated text segments

How good does pre-annotation need to be to have a positive effect?

- 4 out of 6 annotators achieved higher f-score on **S**tate-of-the-art pre-annotated texts
  $\longrightarrow$ not statistically significant
- State-of-the-art ASRL system is not yet good enough
  - to significantly speed up the annotation process
  - to improve annotation quality
- No evidence that the error-prone pre-annotation decreases annotation quality

How good does pre-annotation need to be to have a positive effect? (2)

- The 2 annotators who showed a decrease in f-score were in the same Group (**Group I**: E, S, N)
- Benefit from ongoing training, resulting in higher f-scores for the 3rd text segment (N)
- ANOVA for 4 annotators (**Groups II,III**):
  - all 4 annotators show decrease in annotation quality for **N** (compared to **S**)
  - both types of pre-annotation (S, E) increase f-scores for human annotation quality

Impact of pre-annotation on annotation quality is weakly significant ($p \leq 0.1$)

Do annotators make different types of errors on pre-annotated texts?

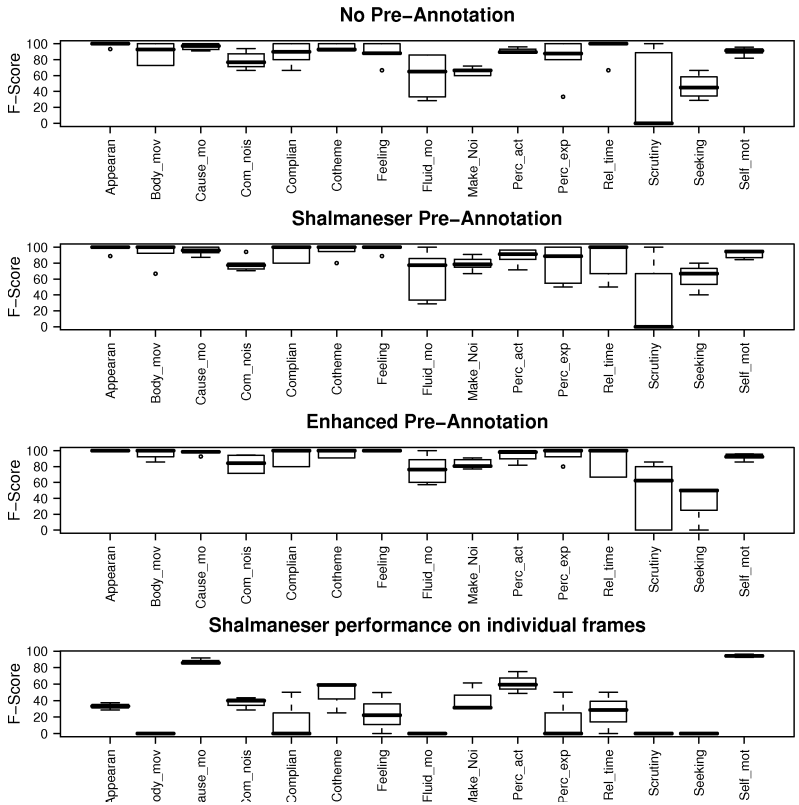


Figure: F-Scores per frame for human annotators on different levels of pre-annotation and for state-of-the-art ASRL

Outline

Conclusions and Future Work

- Assessing the benefits of partial automatic pre-annotation
  - Automatic pre-annotation has a **positive effect on quality** of human annotation
  - Error-prone automatic pre-annotation does **not decrease quality** of human annotation
  - **Strong interaction** between order of text segments ($\rightarrow$ ongoing training effect) and annotation condition, masking the benefits of automatic pre-annotation
- Future work: annotation experiment controlled for order of text segments

Thank You!

Questions?

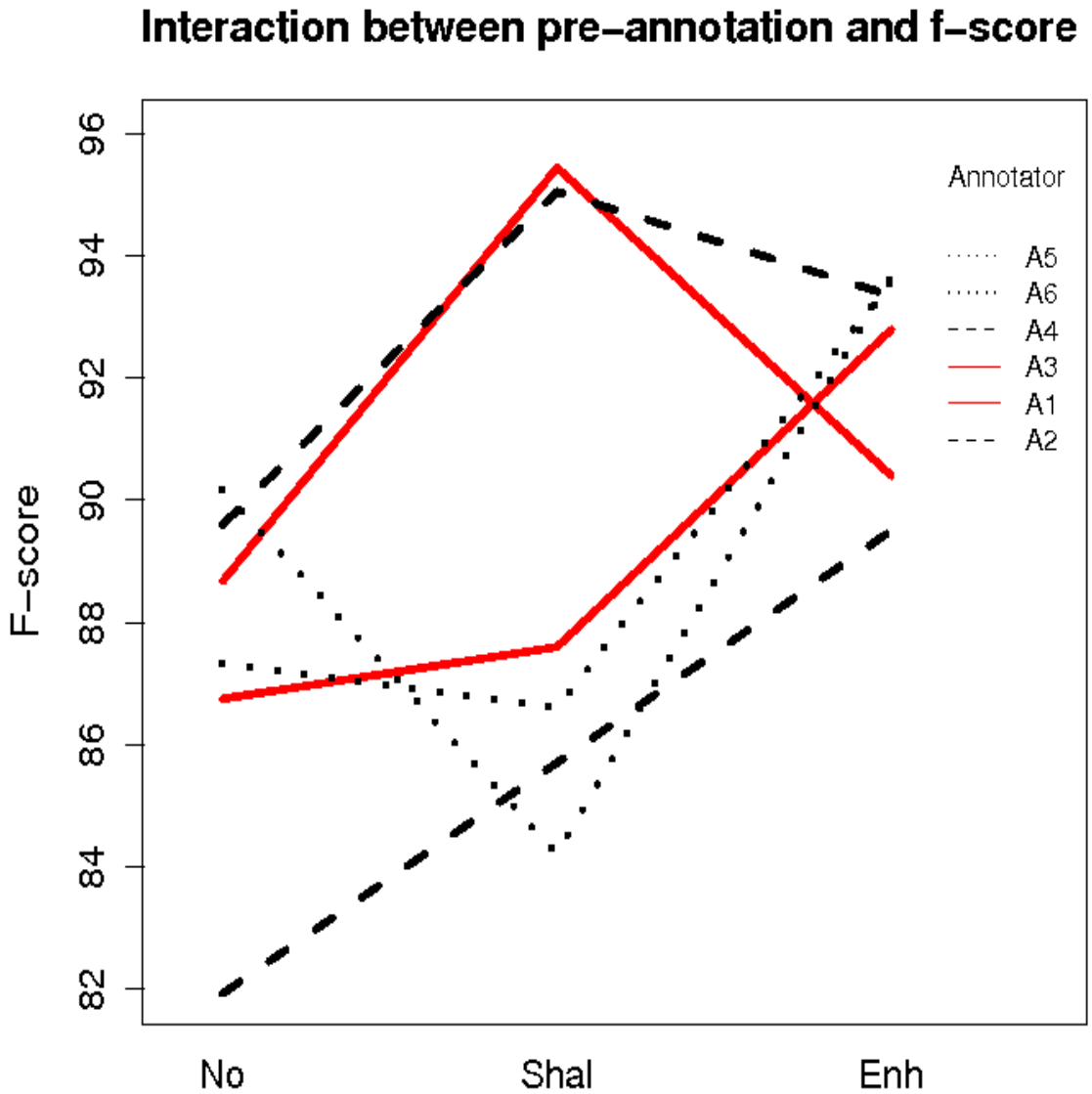
Baselines for automatic pre-annotation (Shalmaneser) and enhanced pre-annotation

Seg.	Precision		Recall		f-score
		Shalmaneser			
A	(70/112)	62.5	(70/96)	72.9	67.30
B	(75/113)	66.4	(75/101)	74.3	70.13
C	(66/113)	58.4	(66/98)	67.3	62.53
		Enhanced Pre-Annotation			
A	(104/112)	92.9	(104/111)	93.7	93.30
B	(103/112)	92.0	(103/112)	92.0	92.00
C	(99/113)	87.6	(99/113)	87.6	87.60

Table: Baselines for automatic pre-annotation (Shalmaneser) and enhanced pre-annotation

Annotator	Precision		Recall		F	t	p
	94/103	91.3	94/109	86.2	88.68	75	N
1	99/107	92.5	99/112	88.4	90.40	61	E
	105/111	94.6	105/109	96.3	95.44	65	S
	93/105	88.6	93/112	83.0	85.71	135	S
2	86/98	87.8	86/112	76.8	81.93	103	N
	98/106	92.5	98/113	86.7	89.51	69	E
	95/107	88.8	95/112	84.8	86.75	168	N
3	103/110	93.6	103/112	92.0	92.79	94	E
	99/113	87.6	99/113	87.6	87.60	117	S
	106/111	95.5	106/112	94.6	95.05	80	S
4	99/108	91.7	99/113	87.6	89.60	59	N
	105/112	93.8	105/113	92.9	93.35	52	E
	104/110	94.5	104/112	92.9	93.69	170	E
5	91/103	88.3	91/113	80.5	84.22	105	S
	96/100	96.0	96/113	85.0	90.17	105	N
	102/106	96.2	102/112	91.1	93.58	124	E
6	94/105	89.5	94/112	83.9	86.61	125	S
	93/100	93.0	93/113	82.3	87.32	135	N

Table: Results for frame assignment: precision, recall, f-score (F), time (t) (frame and role assignment), pre-annotation (p): **N**on, **E**nhanced, **S**halmaneser



Order	Annot.
N E S	1, 3 ———
S N E	2, 4 - - -
E S N	5, 6 .....

Semantic Role Assignment

Anot1	Anot2	Anot3	Anot4	Anot5	Anot6
85.2	80.1	87.7	89.2	82.5	84.3

Table: Average f-scores for the 6 annotators

Neither pre-annotation nor order of text segments has significant impact on Semantic Role Assignment