

Towards a Multilingual Corpus of Translations Annotated Parallel with Compositional Meaning Representations



Lasha Abzianidze¹, Johannes Bjerva¹, Kilian Evang¹, Hessel Haagsma
Rik van Noord¹, Pierre Ludmann², Duc-Duy Nguyen³ and Johan Bos¹

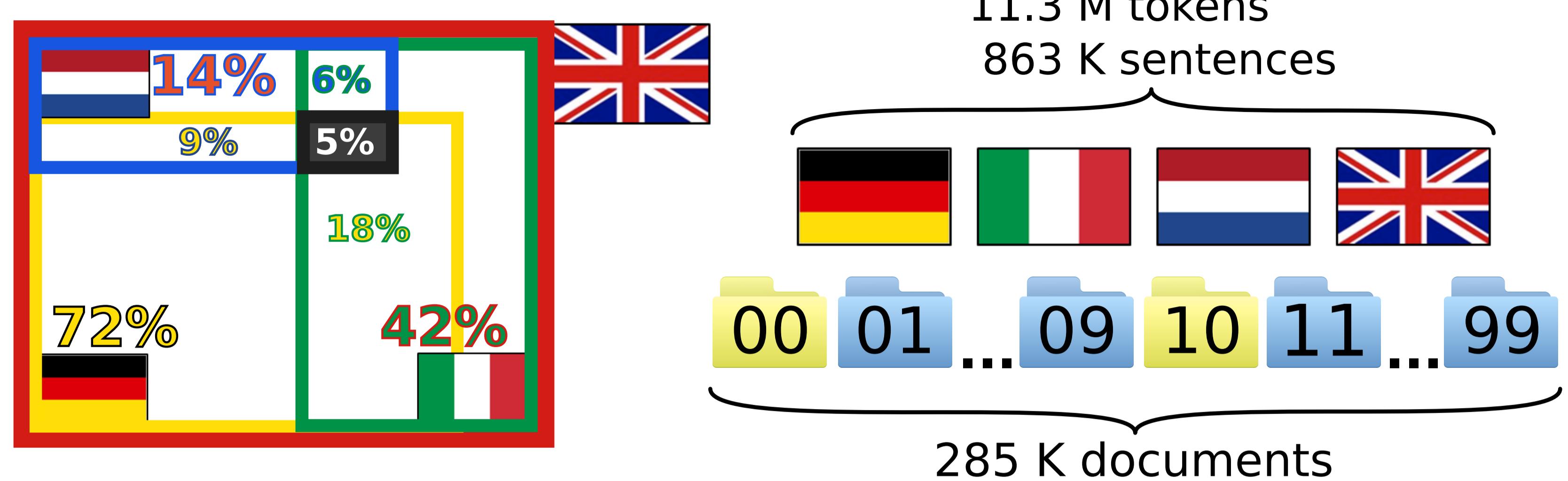
<http://pmb.let.rug.nl> BANK

¹CLCG, University of Groningen; ²Ecole Normale Supérieure de Cachan; ³University of Trento

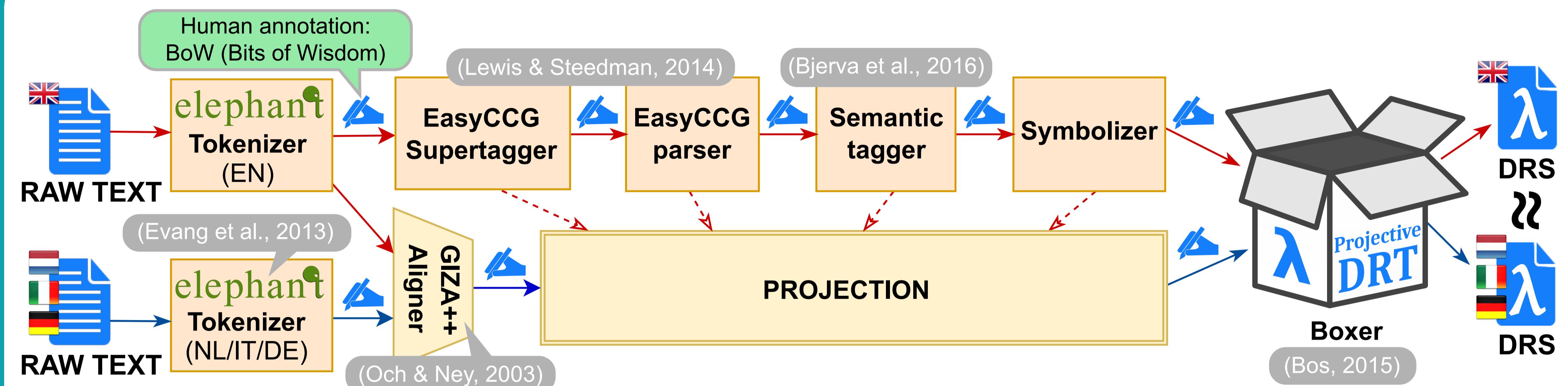
Motivation

- We propose a framework for producing meaning banks for several languages
 - Basic assumption: translations have similar meanings
 - Method: projection of meaning representations from one language to another
 - Driven by a word alignment, syntactic derivations and compositional semantics
 - Result: the Parallel Meaning Bank (PMB)
 - A multilingual parallel corpus with shared meaning representations
 - The PMB provides data for building semantic parsers for languages other than English
 - A test bed for cross-lingual compositional semantics

Languages & Corpora



Automatic Annotation Pipeline



Universal Semantic Tagging

- Language neutral (80 tags)
 - Generalises over part-of-speech & named entity classes

PER Name of a person

NOT Negation

EXS Untensed simple event

PST Past tense

SYM CAT SEM

tom N PER

Projective DRT

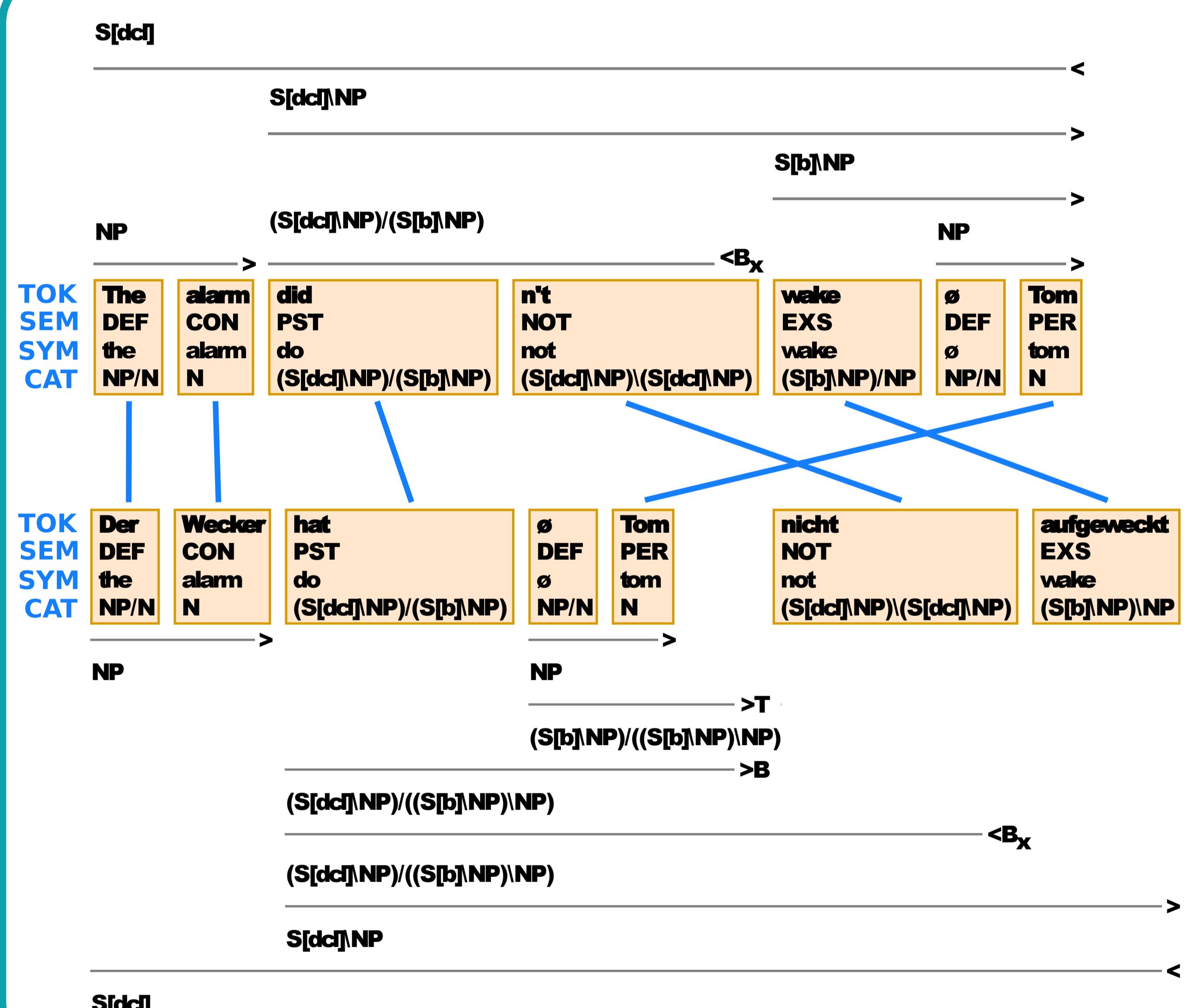
Boxer

Av1.

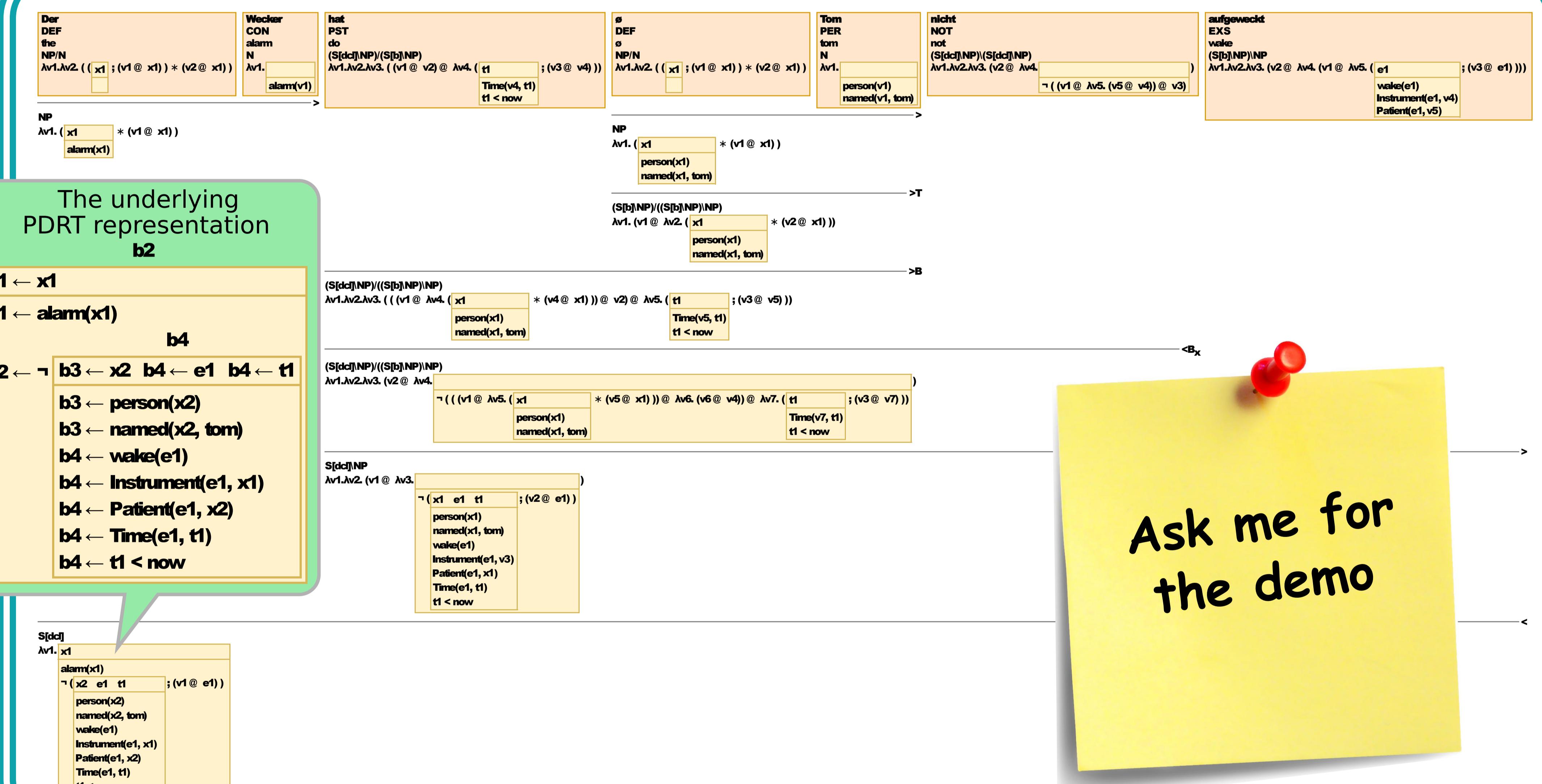
person(v1)

named(v1, tom)

Cross-Lingual Projection



Composed Meaning Representation



Annotated Documents

	Manually checked that all tags are correct	Non-gold with at least one BoW per layer		
Layer				
Tokens		7,784	2,550	274,953
		5,191	689	198,504
		3,182	416	117,396
		1,047	511	38,878
SemTags		3,104	41,227	241,872

Besides adding BoWs, we maintain annotation conflict resolution:
BoW vs BoW or BoW vs System

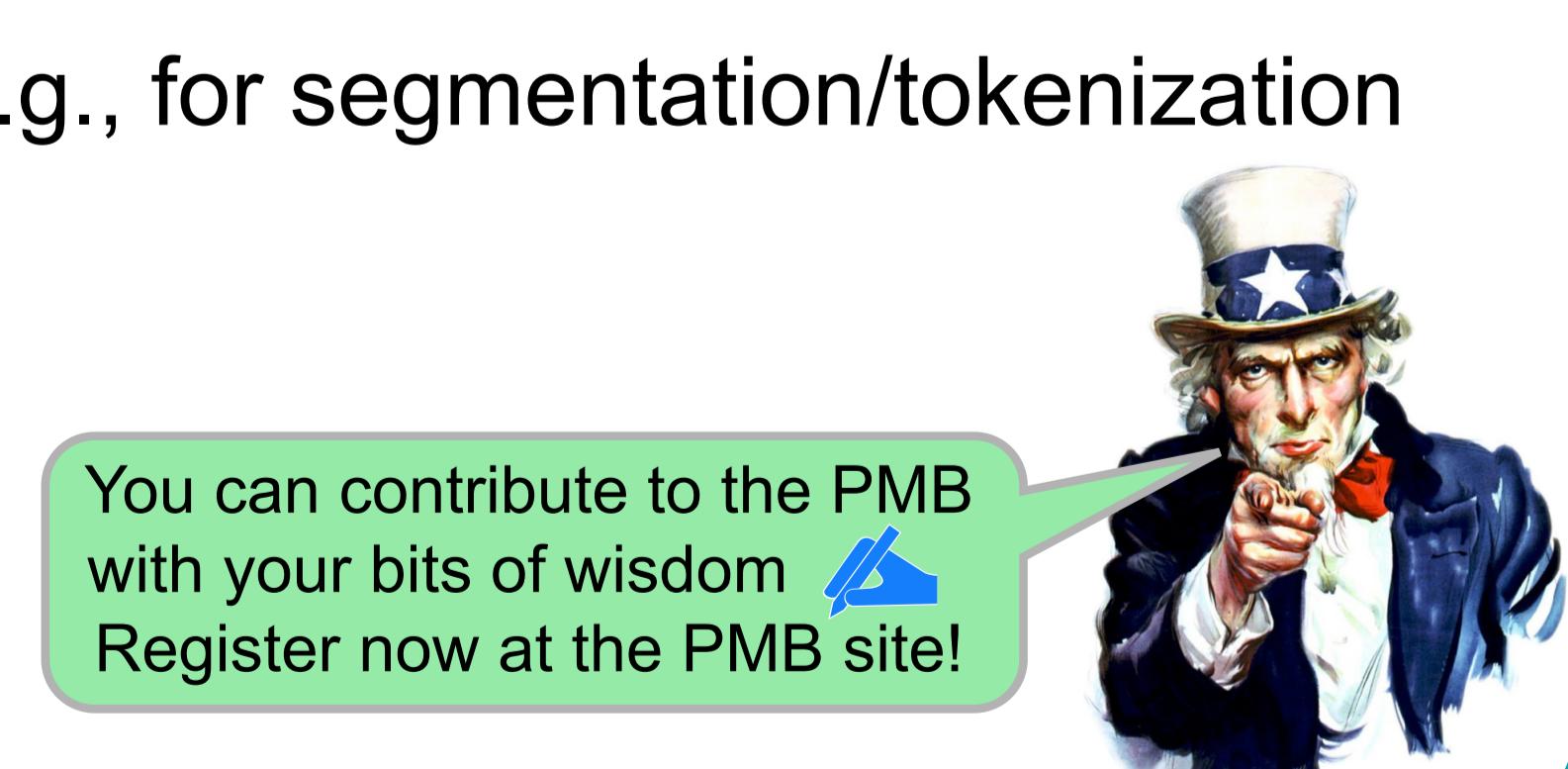
Stats as of
27.03.2017

Conclusion

- We proposed new language-neutral tagging schemes (e.g., for segmentation/tokenization and universal semantic tagging)
 - Promising results in meaning projection
 - An online PMB explorer allows collaborative annotation and automatic reprocessing of the documents



You can contribute to the PMB with your bits of wisdom  Register now at the PMB site!



Future Work

- Enabling systematic studies of the challenges arising from loose translations and different semantic granularities

 She removed the dishes from the table.
 Ze ruimde de tafel af.

Loose translation

 - Develop methods to automatically detect shifts in meaning that are deliberately made by human translators to obtain better translations

 Tom played the piano and Mary sang.
 Tom begleitete Maria's Gesang mit dem Piano.