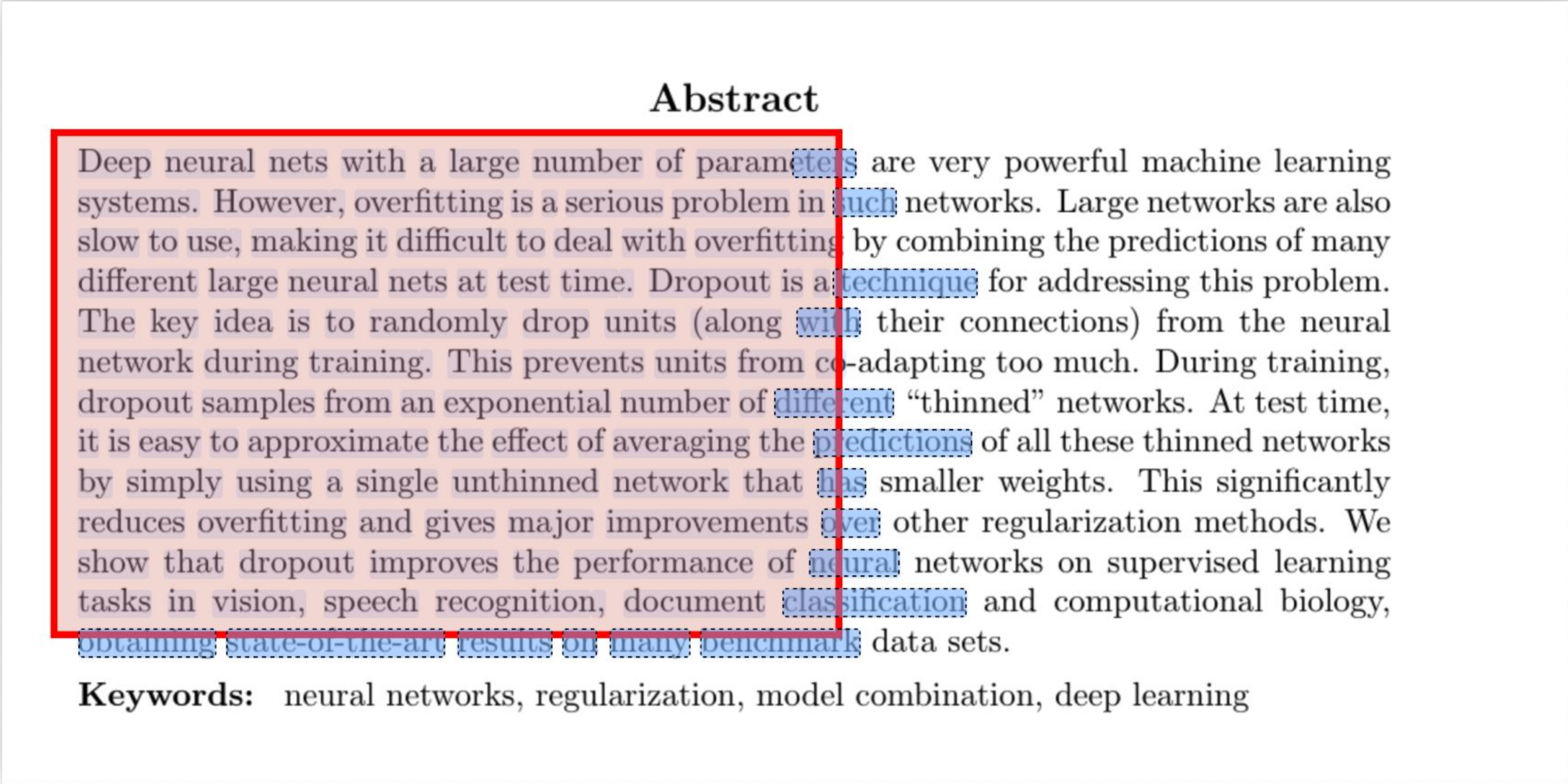
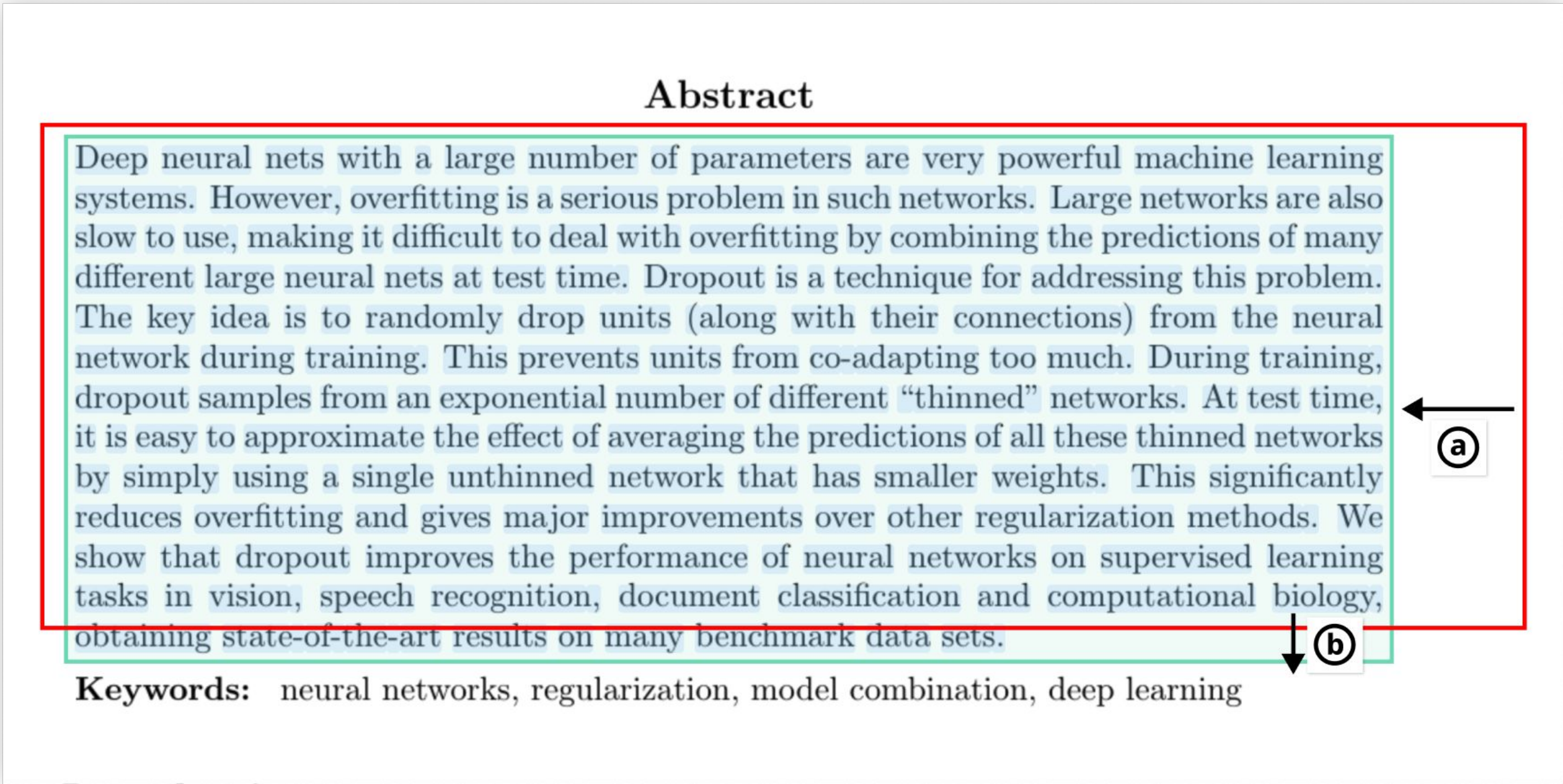


PAWLS is a tool that enables easy annotation of NLP datasets on PDF files. An online demo is available at pawls.apps.allenai.org.



1 Auto detection of the contained tokens

For a user-drawn region bounding box (red), PAWLS automatically detects the contained text, including tokens at the boundaries (highlighted in blue).



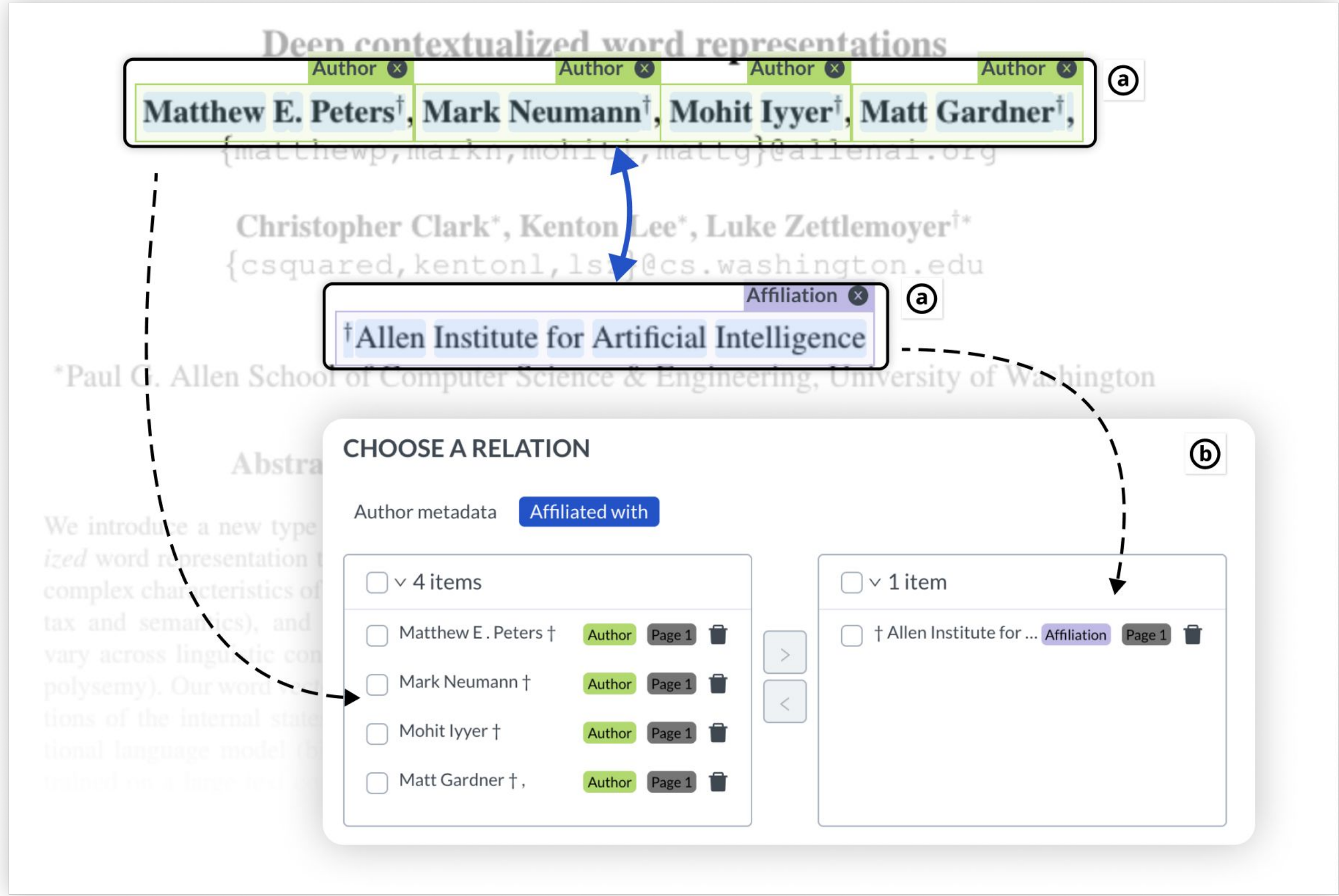
2 Box Snapping

For a user-drawn region box (red), PAWLS automatically trims white space (arrow a) and recovers partially labeled word regions (arrow b), generating accurate and normalized region boundaries (green).



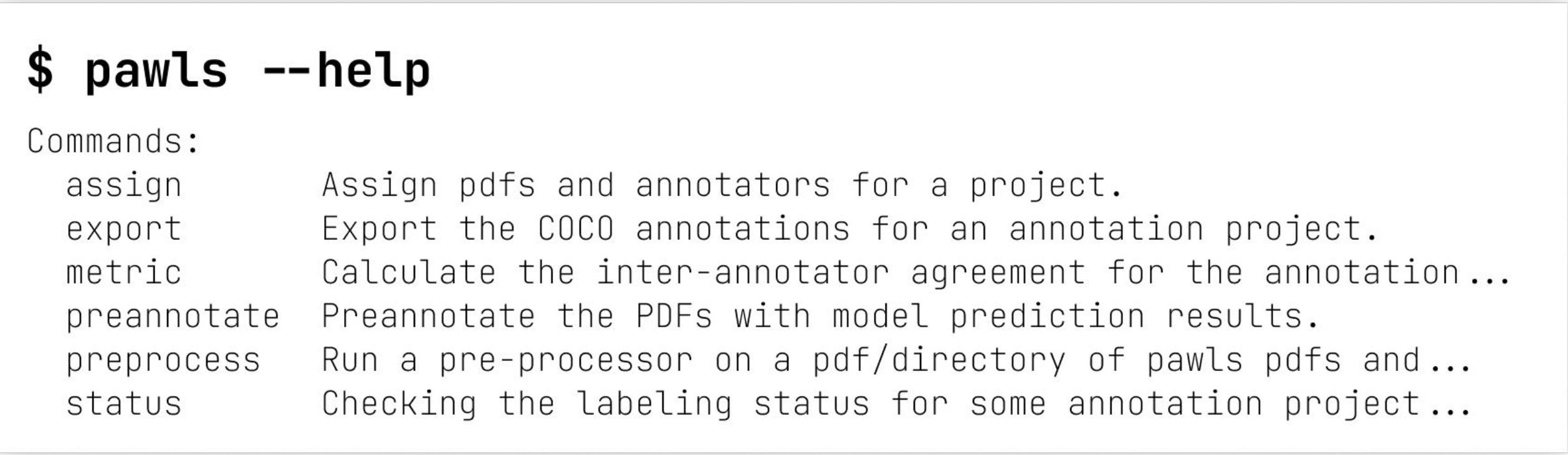
PDF Annotation With Labels and Structure

Mark Neumann, Zejiang Shen, Sam Skjonsberg
{markn, shannons, sams}@allenai.org



3 Relation Annotation

After selecting the bounding boxes with Shift + Click (step a) , users select a relation label and organize the annotations using a Select View (step b), allowing groups, directed annotations, and multi-entity events.



4 Command Line Interface

The PAWLS Command Line Interface (CLI) supports a wide range of functions, including PDF pre-processing, PDF pre-annotation with existing models, labeling status monitoring, annotation export, and many others.