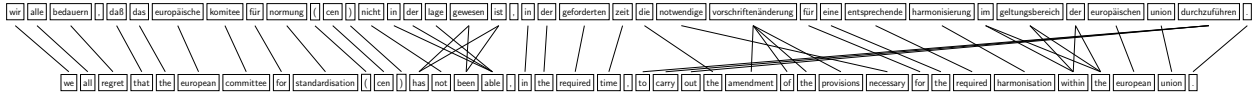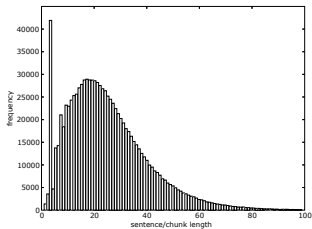# Yet Another Word Alignment Tool

Ulrich Germann

germann@cs.toronto.edu

Alignment visualization by drawing lines. Especially when word order differences are large and sentences are long, the display tends to become very cluttered.

The most common visualization techniques for word alignment relations are the drawing of lines (↑) and alignment matrices (→). These techniques have several drawbacks.

- **Static display of information:** *All* information is shown *all at once*. The resulting visual clutter can easily overwhelm the human annotator and makes individual alignment relations hard to track.
- **Inflexible layout:** each sentence must be displayed in a single row or column.
  - The text is harder to read in this form than in regular text flow with line breaks.
  - The sentences often do not fit on the screen entirely and the human viewer/annotator has to scroll back and forth to get the full picture.

Distribution of segment lengths in the English track of the Europarl corpus. 1,733 out of 974,507 segments are 100 words or longer.
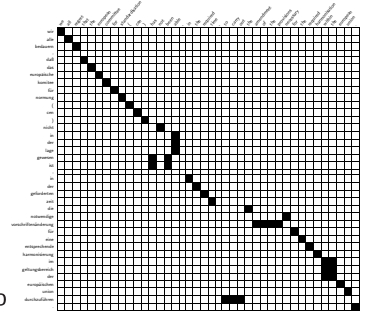
Alignment visualization with an alignment matrix.

Yawat is different.

Dynamic HTML reacts to annotator activity. Only the alignment information for the word currently under the mouse pointer is shown by highlighting the word and the words associated with it. By using highlighting for alignment visualization, we can show the parallel corpus as naturally flowing text.
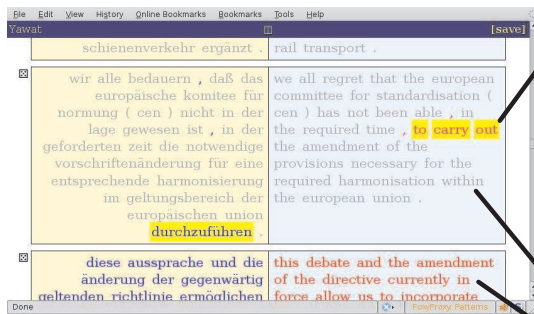
Yawat is efficient.

The GUI was designed to minimize mouse travel required for alignment. Most functions are accomplished by clicking on words. A left-click on a word opens an 'alignment group', subsequent left-clicks on other words add them to the alignment group, a final right click on one of the words closes the group.

Different font colors are used to indicate which words have been aligned already, and which ones still need to be aligned.

Yawat can do more than just align words.
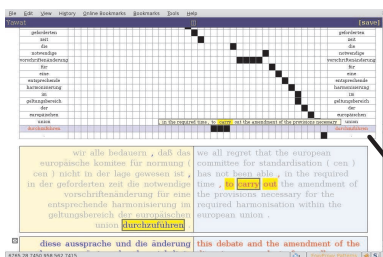
Alignments can be labeled with a customizable tag set.

Alignment matrix display is available upon request.

Computational Linguistics Group