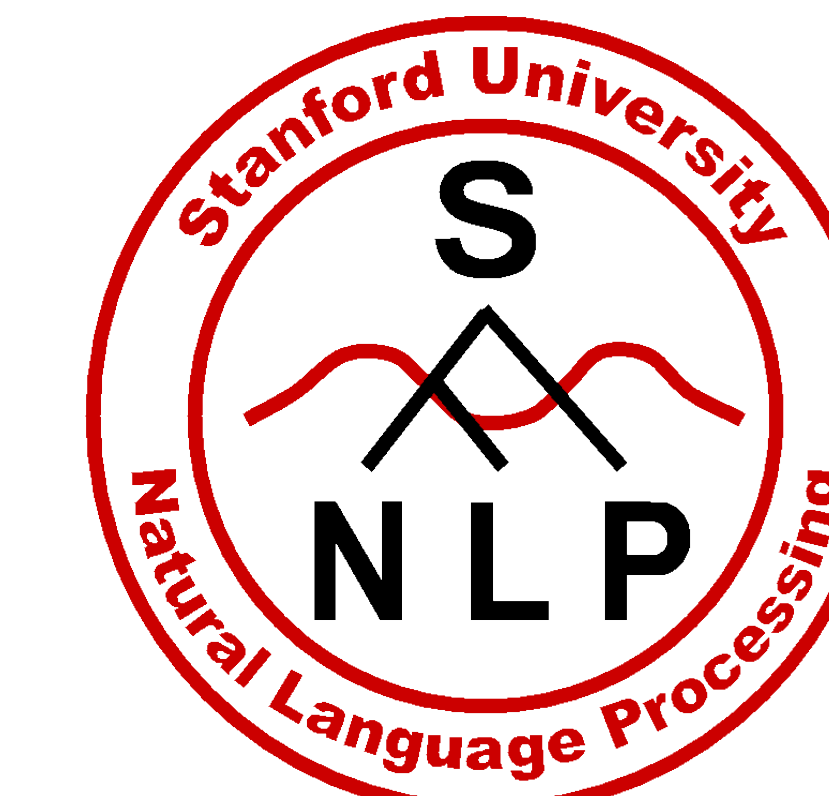
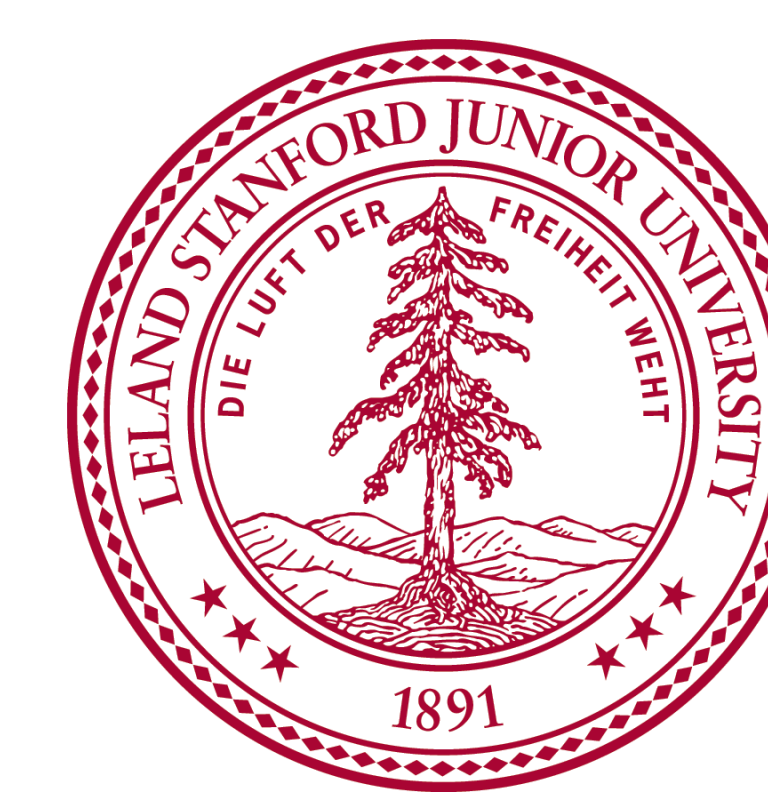


# The Stanford Post-Editing Corpus: Collection and Analysis

Spence Green, Jeffrey Heer, and Christopher D. Manning

<http://nlp.stanford.edu/data/postedit.shtml>



## Introduction: What is it?

A collection of English Wikipedia excerpts translated to **Arabic, French, and German**

16 professional human translations per language pair

**Human ratings of translation quality** from Mechanical Turk using the Workshop on Machine Translation (WMT) pairwise ranking framework.

Translation conditions: unaided and Google Translate

Source metadata: POS tags, dependency parses, named entity tags

**Human metadata:** source and target language proficiency tests, hourly rate, demographic information

**Activity metadata:** timestamped Javascript event logs

## Source Text

Four **universal topics**: 1896 Olympics (easy), Flag of Japan (easy), Schizophrenia (hard), Infinite monkey theorem (hard)

Topic had “Featured” status in at least two of the four languages.

27 sentences, 606 tokens, 22.4 tokens/sentence

It was the first international Olympic Games held in the Modern era.

Any physical process that is even less likely than such monkeys’ success is effectively impossible, and it may safely be said that such a process will never happen.

## Data Collection

### Subjects

Freelance translators on oDesk, most held at least a bachelor’s degree. **Three had Ph.Ds**

Solicited bids at \$0.085 per source word (\$52 in total). Bids centered near the target price: Arabic (\$50.50) French (\$52.32) German (\$49.93)

Completed 40-minute, 40-question tests of both source and target language proficiency

### Experimental Design

Two conditions: unaided and Google Translate

Two timed translation sessions separated by an untimed break

**Maximum pause length:** 3 minutes

### Translation Quality Assessment

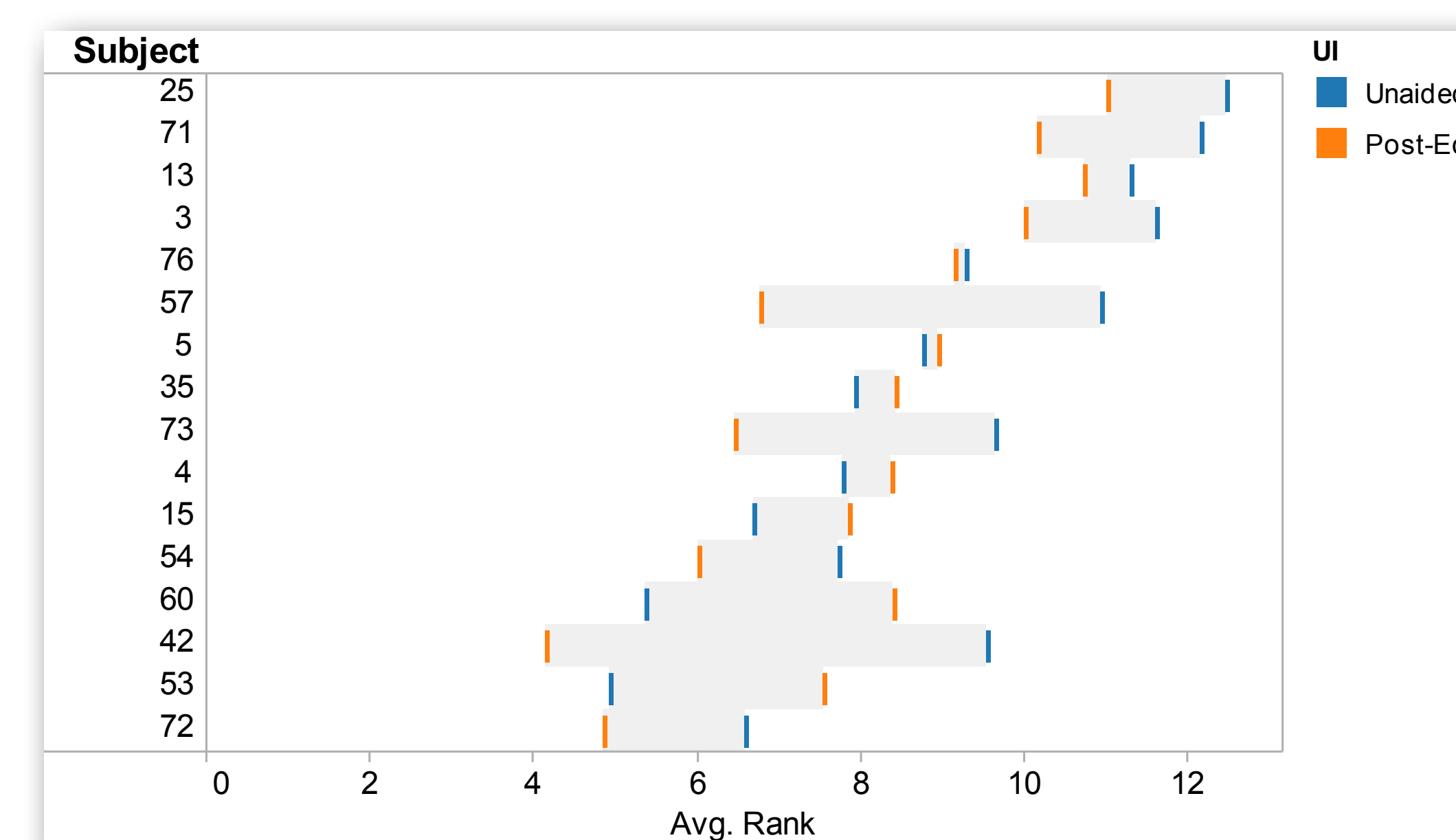
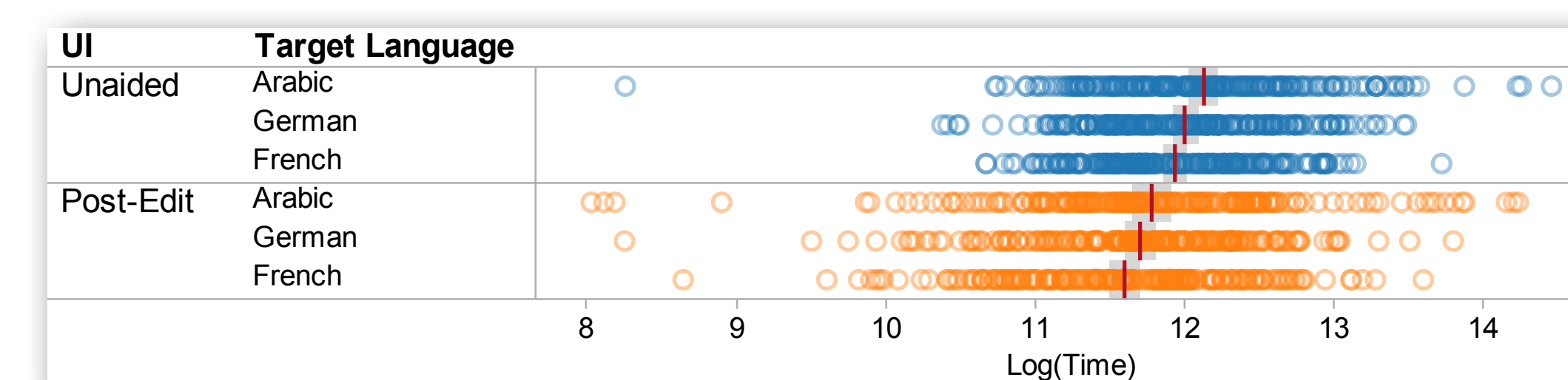
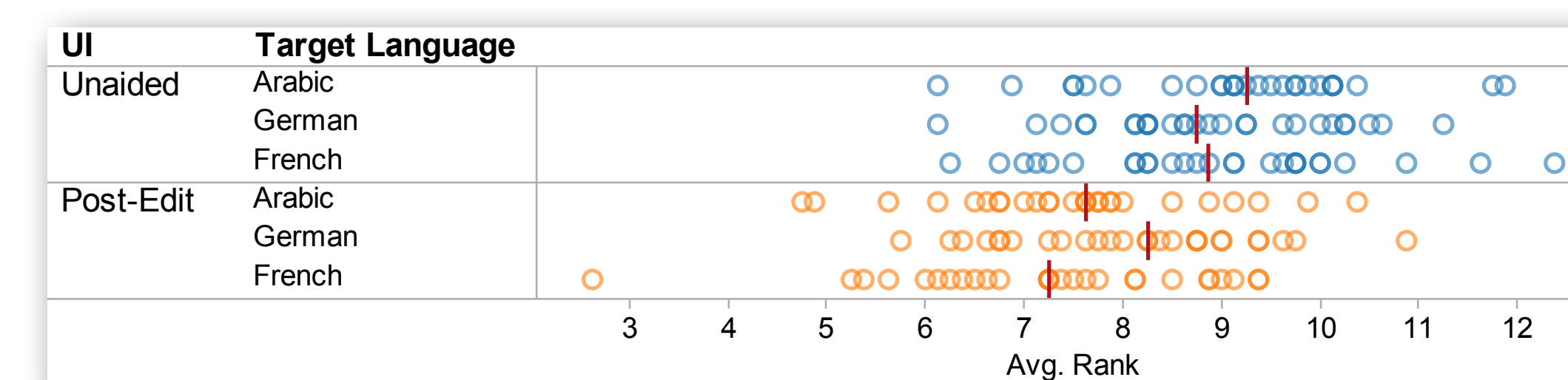
**Pairwise assessment** identical to the WMT 2012 evaluation campaign

\$0.04 per HIT, 3 sentence pairs per HIT

**Sentence-level ranking** based on the minimum-feedback arc set solver of Lopez (2012)

## Visual Summary of the Data

In short: **high post-editor variance**



**Example of Interface B (Translate to French)**

Gandhi spent 21 years in South Africa, where he developed his political views, his ethics, and his political leadership skills.

Gandhi a passé 21 ans en Afrique du Sud, où il a développé ses opinions politiques, son éthique et ses compétences en leadership politiques.

Submit

Feel free to delete the suggested translation if you think that it would be quicker to type your own translation.

Idle: 00:26 (maximum: 03:00)

## Initial Analysis

- Translators are faster and more accurate** in the post-edit condition for all three languages
- MT primes the translator:** post-edit translations are closer to MT by edit distance
- Simple source lexical features** predict time
- Interaction patterns measurably different in post-edit condition

