

Word Transition Entropy as an Indicator for Expected Machine Translation Quality

Michael Carl¹ and Moritz Schaeffer²

CRITT, Copenhagen Business School¹
University of Edinburgh²

Overview

- Ambiguity and Translation Entropy in Post-editing
- Correlation Translation Entropy and Gaze Duration
- MT Search Graph Transition Entropy
- Correlation MT Search Graph Transition Entropy and PE Duration
- Data taken from CFT13 in TPR-DB v 1.4
Poster on Wednesday

Translation Ambiguity and Entropy

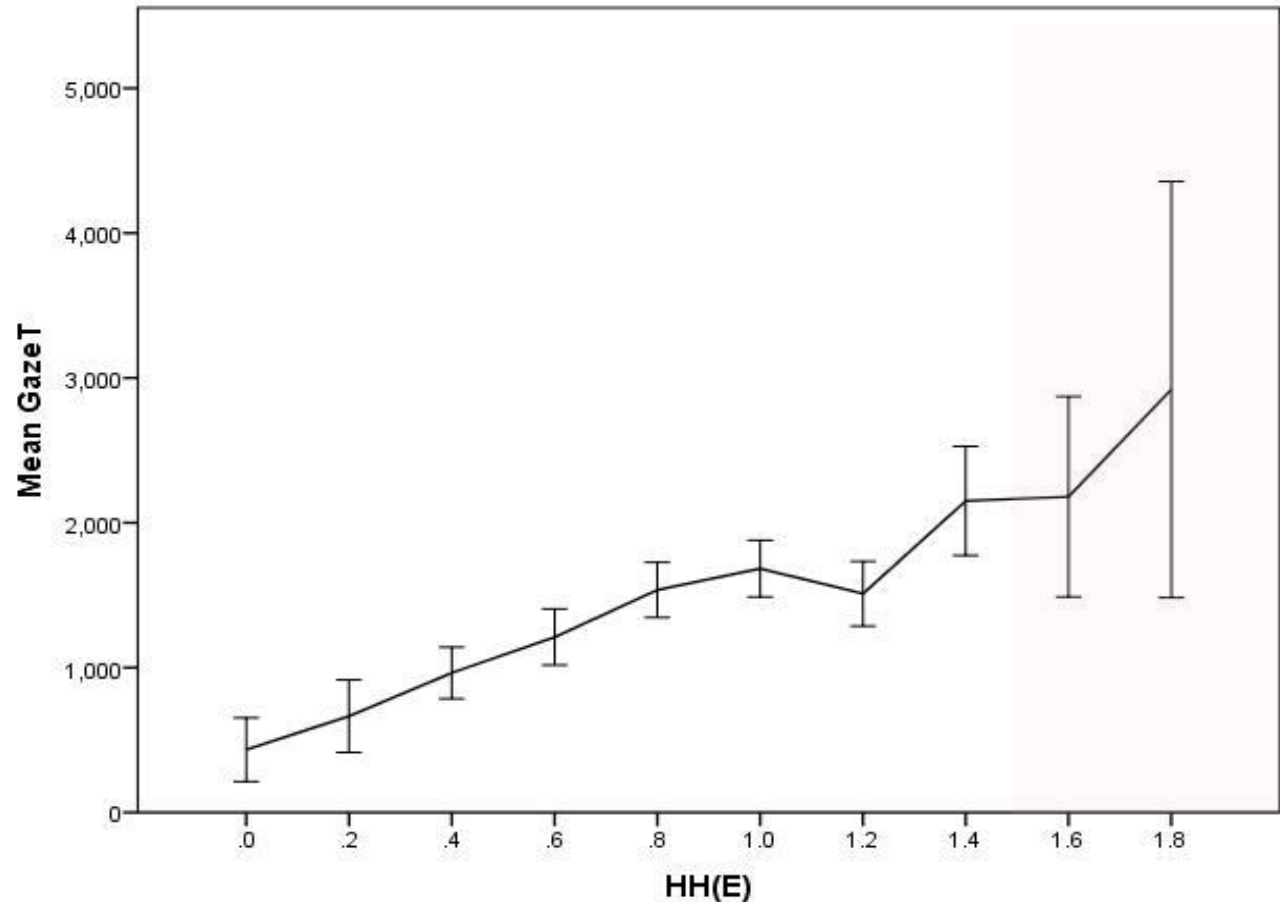
	Killer	nurse	receives	four	life	sentences
11	asesino	7 el_ enfermero	15 recibe	28 cuatro	12 perpetuas	13 cadenas
6	el_ asesino	5 enfermero_ asesino	3 es_ condenado		12 cadenas	11 perpetuas
3	el_ enfermero	4 enfermero	3 condenado _a			2 asesino
2	Enfermero _asesino	4 asesino	2 recibe_a			
		3 un_ enfermero				
		2 enfermera				

- Semi-automatically word-aligned
- Human Word Translation Entropy:

$$HH(e) = -\sum_i p(e \rightarrow s_i) * \log_2 (p(e \rightarrow s_i))$$

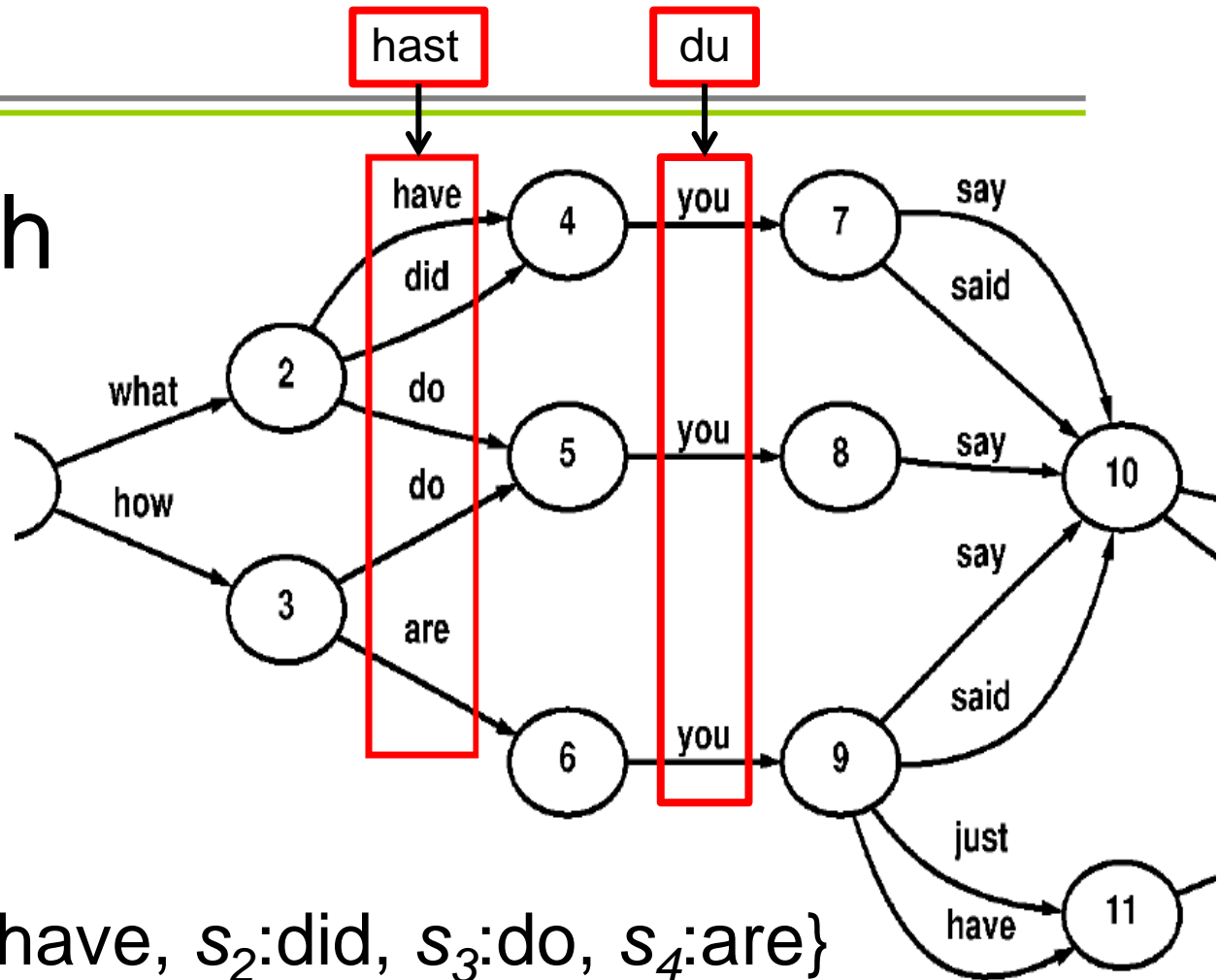
Correlation $HH(e)$ and Gaze Target

Human word translation entropy $HH(e)$ correlates with longer gaze activity on the target word.



Error Bars: 95% CI

MT Search
Graph
Word
Transition
Entropy



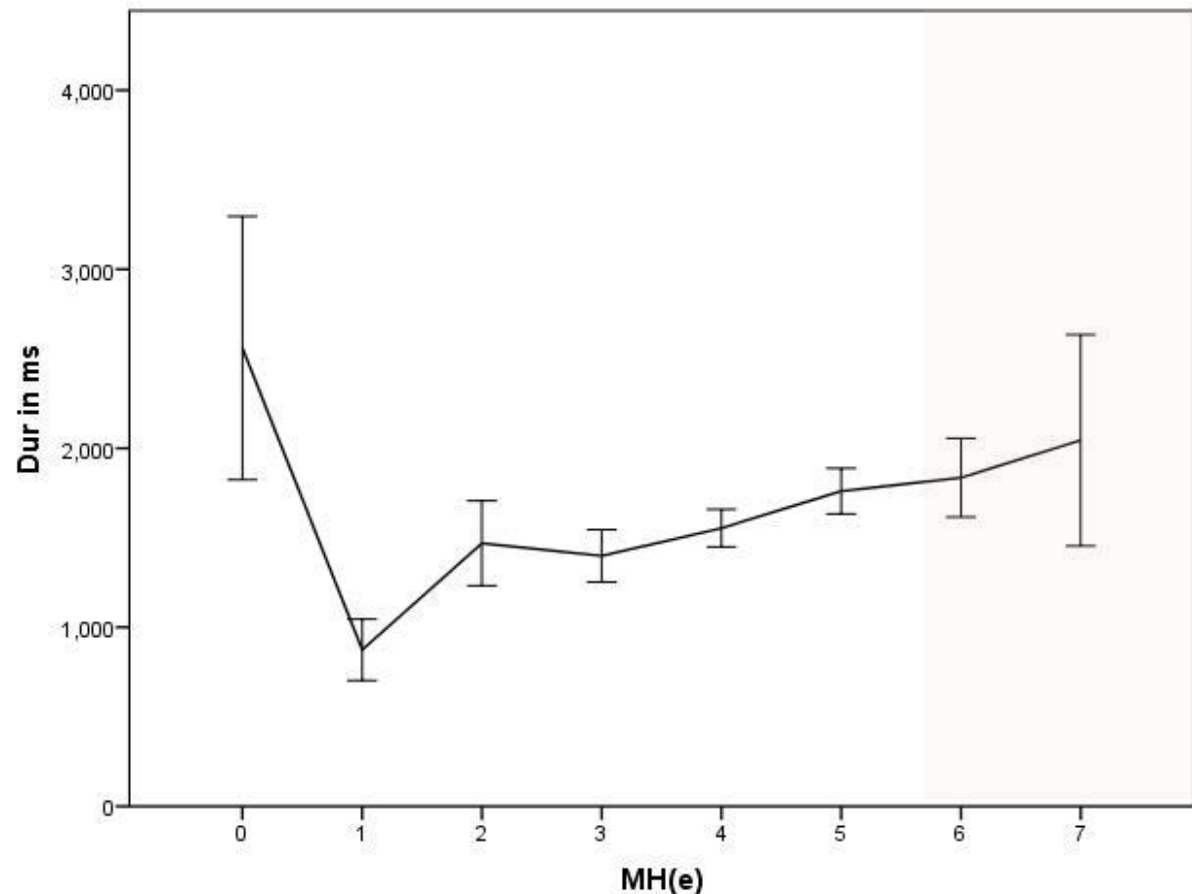
$e_1: \text{hast} \rightarrow \{s_1: \text{have}, s_2: \text{did}, s_3: \text{do}, s_4: \text{are}\}$
 $e_2: \text{du} \rightarrow \{s_1: \text{you}\}$

$$MH(e) = -\sum_i p(e \rightarrow s_i) * \log_2 (p(e \rightarrow s_i))$$

Correlation $MH(e)$ and PE duration

Entropy of MT search graph transitions $MH(e)$ correlates with post-editing duration.

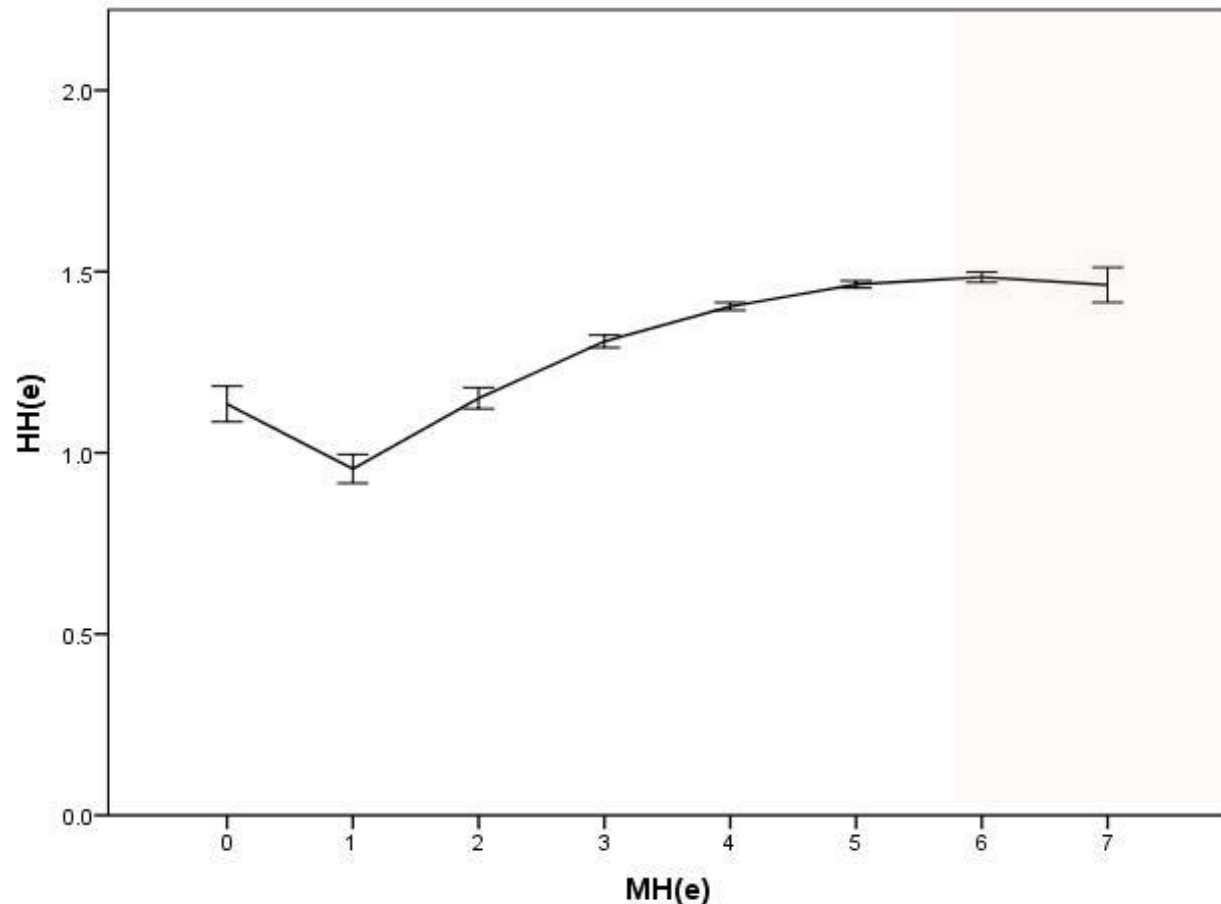
$MH(e) = 0$
due to OOV more PE effort



Error Bars: 95% CI

Correlation $HH(e)$ and $MH(e)$.

MT search
graph transition
entropy $MH(e)$
correlates with
human
translation
entropy
 $HH(e)$,
up to a certain
point.



Error Bars: 95% CI

Conclusion

- The higher $HH(e)$ entropy the longer gaze activity
- The higher $MH(e)$ entropy the longer PE duration
- MT search graph transition entropy $MH(e)$ correlates with human translation entropy $HH(e)$ up to a certain point

→ Visit the CasMaCat booth in the HLT Village