

---

# Computer Aided Translation a User Study

Philipp Koehn

29 September 2009



# Why Machine Translation?



**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)



# Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don't speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

# Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don't speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

**Dissemination** — publisher wants to make content available in other languages

- high demands for quality
- currently almost exclusively done by human translators

# Why Machine Translation?

**Assimilation** — reader initiates translation, wants to know content

- user is tolerant of inferior quality
- focus of majority of research (GALE program, etc.)

**Communication** — participants don't speak same language, rely on translation

- users can ask questions, when something is unclear
- chat room translations, hand-held devices
- often combined with speech recognition, IWSLT campaign

**Dissemination** — publisher wants to make content available in other languages

- high demands for quality
- currently almost exclusively done by human translators

**OUR  
FOCUS**

# Goal: Helping Human Translators



If you can't beat them, join them.

# Goal: Helping Human Translators



If you can't beat them, join them.

- How can machine translation help human translators?

# Goal: Helping Human Translators



If you can't beat them, join them.

- How can machine translation help human translators?
- First question: What do translators do?



# Setup



- 10 students at the University of Edinburgh
  - half native French speakers
  - half native English speakers with advanced French

# Setup

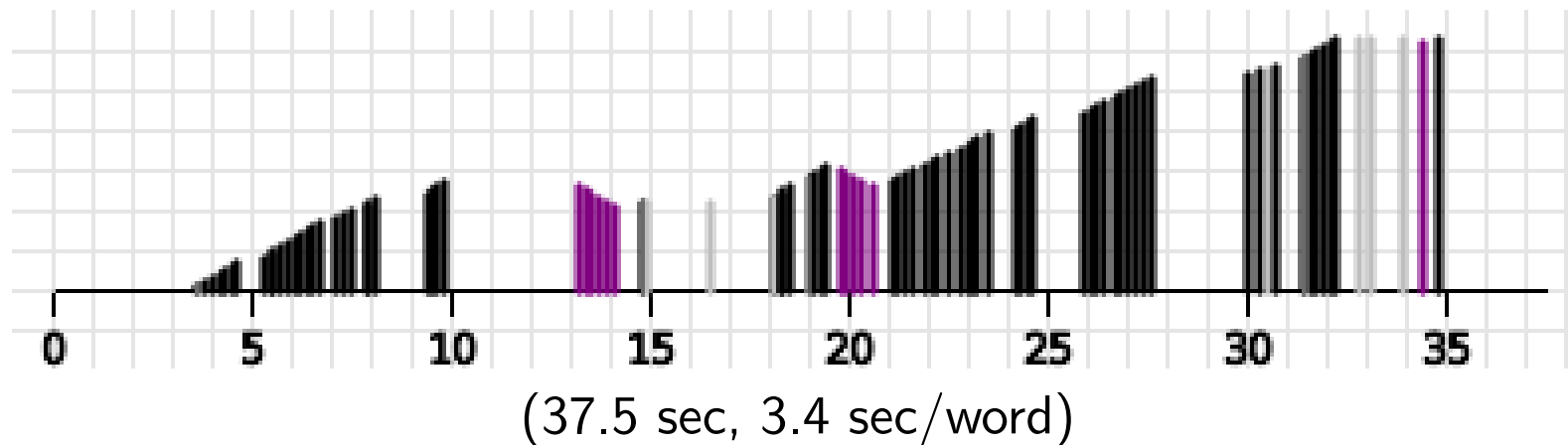


- 10 students at the University of Edinburgh
  - half native French speakers
  - half native English speakers with advanced French
- Each student translated
  - news stories
  - French-English
  - about 40 sentences
  - easy task: familiar content, no specialized terminology
- Keystroke log

# Keystroke Log

Input: *Au premier semestre, l'avionneur a livr 97 avions.*

Output: *The manufacturer has delivered 97 planes during the first half.*



black: keystroke, purple: deletion, grey: cursor move  
 height: length of sentence

# Analysis



- We can observe
  - slow typing

# Analysis



- We can observe
  - slow typing
  - fast typing

# Analysis



- We can observe
  - slow typing
  - fast typing
  - pauses

# Analysis



- We can observe
  - slow typing
  - fast typing
  - pauses
- Pauses
  - beginning pause: reading the input sentence
  - final pause: reviewing the translation

# Analysis



- We can observe
  - slow typing
  - fast typing
  - pauses
- Pauses
  - beginning pause: reading the input sentence
  - final pause: reviewing the translation
  - short pauses (2-6 seconds): hesitation
  - medium pauses (6-60 seconds): problem solving
  - big pauses (>60 seconds): serious problem



## Time Spent on Activities

User	total	Pauses					keystroke
		initial	final	short	medium	big	
L2a	3.3s	0.1s	0.1s	0.2s	1.0s	0.1s	1.8s
L2b	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s
L2c	3.9s	0.2s	0.2s	0.3s	0.7s	-	2.5s
L2d	2.8s	0.2s	0.0s	0.2s	0.4s	0.1s	1.8s
L2e	5.2s	0.3s	0.0s	0.3s	1.9s	0.5s	2.2s
L1a	5.7s	0.5s	0.1s	0.3s	1.8s	0.7s	2.2s
L1b	3.2s	0.1s	0.1s	0.2s	0.4s	0.1s	2.2s
L1c	5.8s	0.3s	0.2s	0.5s	1.5s	0.3s	3.1s
L1d	3.4s	0.7s	0.1s	0.3s	0.6s	-	1.8s
L1e	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s

L2 = native French, L1 = native English  
average time per input word

# Time Spent on Activities

User	total	not much time			Pauses		keystroke
		initial	final	short	medium	big	
L2a	3.3s	0.1s	0.1s	0.2s	1.0s	0.1s	1.8s
L2b	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s
L2c	3.9s	0.2s	0.2s	0.3s	0.7s	-	2.5s
L2d	2.8s	0.2s	0.0s	0.2s	0.4s	0.1s	1.8s
L2e	5.2s	0.3s	0.0s	0.3s	1.9s	0.5s	2.2s
L1a	5.7s	0.5s	0.1s	0.3s	1.8s	0.7s	2.2s
L1b	3.2s	0.1s	0.1s	0.2s	0.4s	0.1s	2.2s
L1c	5.8s	0.3s	0.2s	0.5s	1.5s	0.3s	3.1s
L1d	3.4s	0.7s	0.1s	0.3s	0.6s	-	1.8s
L1e	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s

L2 = native French, L1 = native English  
 average time per input word

# Time Spent on Activities

User	total	not much time			Pauses		similar keystroke
		initial	final	short	medium	big	
L2a	3.3s	0.1s	0.1s	0.2s	1.0s	0.1s	1.8s
L2b	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s
L2c	3.9s	0.2s	0.2s	0.3s	0.7s	-	2.5s
L2d	2.8s	0.2s	0.0s	0.2s	0.4s	0.1s	1.8s
L2e	5.2s	0.3s	0.0s	0.3s	1.9s	0.5s	2.2s
L1a	5.7s	0.5s	0.1s	0.3s	1.8s	0.7s	2.2s
L1b	3.2s	0.1s	0.1s	0.2s	0.4s	0.1s	2.2s
L1c	5.8s	0.3s	0.2s	0.5s	1.5s	0.3s	3.1s
L1d	3.4s	0.7s	0.1s	0.3s	0.6s	-	1.8s
L1e	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s

L2 = native French, L1 = native English  
average time per input word

# Time Spent on Activities

User	total	not much time			differences		similar keystroke
		initial	final	short	medium	big	
L2a	3.3s	0.1s	0.1s	0.2s	1.0s	0.1s	1.8s
L2b	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s
L2c	3.9s	0.2s	0.2s	0.3s	0.7s	-	2.5s
L2d	2.8s	0.2s	0.0s	0.2s	0.4s	0.1s	1.8s
L2e	5.2s	0.3s	0.0s	0.3s	1.9s	0.5s	2.2s
L1a	5.7s	0.5s	0.1s	0.3s	1.8s	0.7s	2.2s
L1b	3.2s	0.1s	0.1s	0.2s	0.4s	0.1s	2.2s
L1c	5.8s	0.3s	0.2s	0.5s	1.5s	0.3s	3.1s
L1d	3.4s	0.7s	0.1s	0.3s	0.6s	-	1.8s
L1e	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s

L2 = native French, L1 = native English  
average time per input word

# Pauses Reconsidered



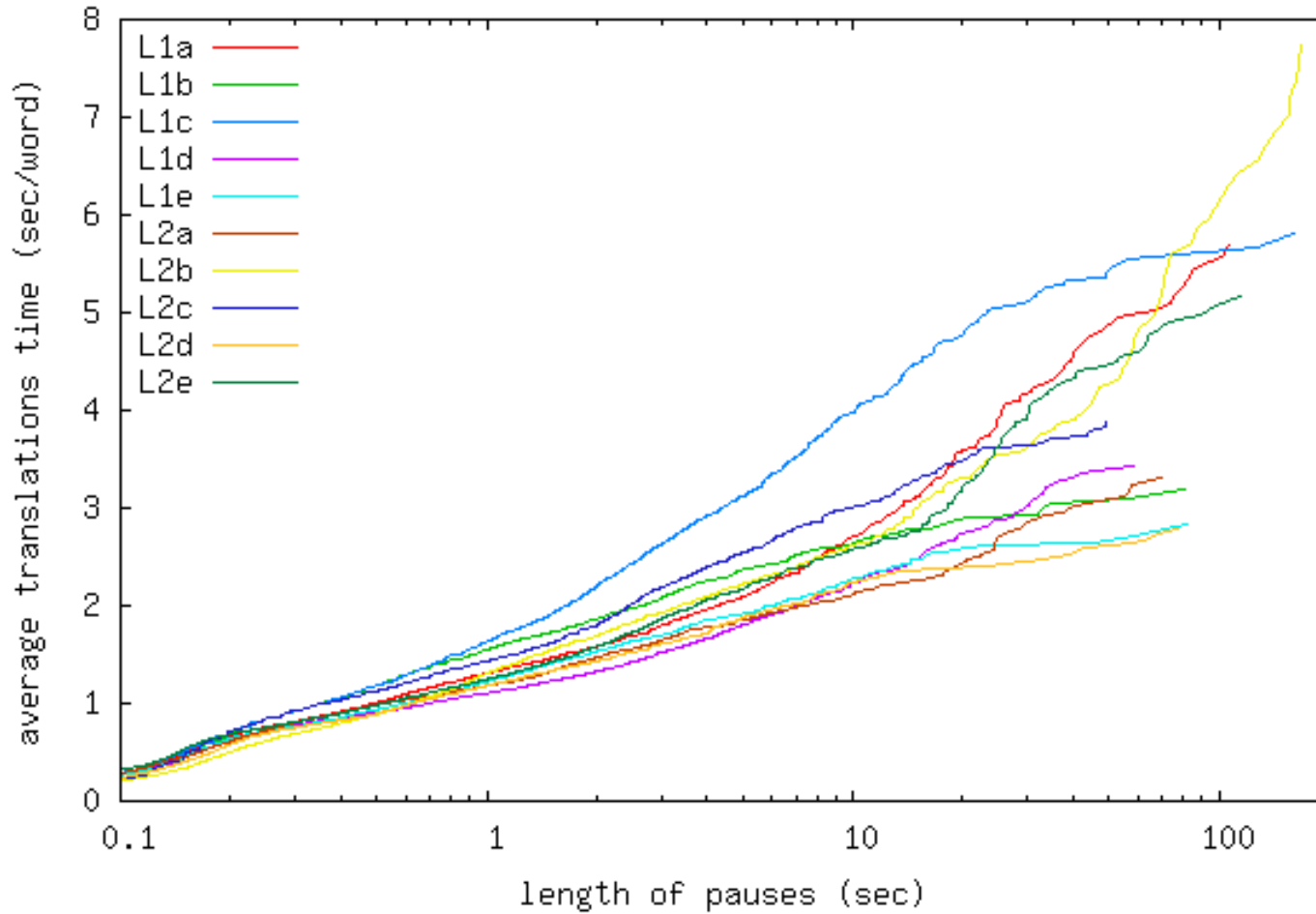
- Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)
- Extreme view: all you see is pauses
  - keystrokes take no observable time
  - all you see is pauses between action points

# Pauses Reconsidered

- Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)
- Extreme view: all you see is pauses
  - keystrokes take no observable time
  - all you see is pauses between action points
- Visualizing range of pauses:  
time  $t$  spent in pauses  $p \in P$  up to a certain length  $l$

$$sum(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)$$

# Results



# Related Work: Tools used by Translators<sup>13</sup>



- Translators often use standard text editors and additional tools
- Bilingual dictionary
- Spell checker, grammar checker
- Monolingual concordancer
- Terminology database
- Web search to establish and verify meaning of terms



# Translation Memory

- Source:

This feature is available for free in the QX 3400.



- Fuzzy match in translation memory:

This feature is available for free in the QX 3200.

Diese Funktion ist kostenlos im Modell QX 3200 verfügbar.

- Translator inspects the fuzzy match and uses it in her translation.

# Bilingual Concordancer

Examples + -	<b>Windkraft</b> (noun, feminine) (also: Windenergie)	 <b>wind power</b> (noun)	✓
	Zum Vergleich: <b>Windkraft</b> schafft fast sieben Mal mehr. ↳ German: <a href="http://www.goethe.de/wis/umw/thm/ntr/de92305.htm">www.goethe.de/wis/umw/thm/ntr/de92305.htm</a>	By way of comparison, <b>wind power</b> generates almost seven times as much. ↳ English: <a href="http://www.goethe.de/wis/umw/thm/ntr/en92305.htm">www.goethe.de/wis/umw/thm/ntr/en92305.htm</a>	
	Einführung von Windcube, einer neuen Generation von Wind Lidar für <b>Windkraft</b> . ↳ German: <a href="http://www.husumwindenergy.com/index.php?L...howUid]=1177">www.husumwindenergy.com/index.php?L...howUid]=1177</a>	Introducing Windcube, a new generation of <b>wind</b> Lidar for <b>wind power</b> . ↳ English: <a href="http://www.husumwindenergy.com/index.php?L...howUid]=1177">www.husumwindenergy.com/index.php?L...howUid]=1177</a>	
	<b>Windkraft</b> ist eine etablierte, wettbewerbsfähige Technologie mit hoher Zuverlässigkeit ↳ German: <a href="http://www.powergeneration.siemens.de/abo...ns-services/">www.powergeneration.siemens.de/abo...ns-services/</a>	<b>Wind power</b> is an established, competitive technology with high reliability ↳ English: <a href="http://www.powergeneration.siemens.com/abo...ns-services/">www.powergeneration.siemens.com/abo...ns-services/</a>	
Examples + -	<b>Windkraft</b> (noun, feminine) (also: Windenergie)	 <b>wind energy</b> (noun)	✓
	Je mehr aber klimapolitische Sonntagsreden von der Politik auch in Taten umgesetzt werden, desto höher steigt dieser Preis und desto wettbewerbsfähiger werden saubere Energien wie die <b>Windkraft</b> . ↳ German: <a href="http://emagazine.credit-suisse.com/app/art...4382">emagazine.credit-suisse.com/app/art...4382</a> (=DE)	But as the focus of the climate change issue shifts increasingly from policy to action, this price will increase and cleaner <b>energy</b> sources like <b>wind</b> will become more competitive. ↳ English: <a href="http://emagazine.credit-suisse.com/app/art...4382">emagazine.credit-suisse.com/app/art...4382</a> (=en)	
	Nur wenige befürchten hingegen, dass dies auch bei erneuerbaren Energieträgern wie Biomasse oder <b>Windkraft</b> der Fall sein wird. ↳ German: <a href="http://www.eu2006.gv.at/de/News/Press_Rele...1proell.html">www.eu2006.gv.at/de/News/Press_Rele...1proell.html</a>	However, only a few fear that this will also be the case with renewable <b>energy</b> sources such as biomass or <b>wind energy</b> . ↳ English: <a href="http://www.eu2006.gv.at/en/News/Press_Rele...1proell.html">www.eu2006.gv.at/en/News/Press_Rele...1proell.html</a>	

show translations in context (www.linguee.com)

# Our Types of Assistance



- Sentence completion
  - tool suggests how to complete the translation
  - one phrase at a time

# Our Types of Assistance



- Sentence completion
  - tool suggests how to complete the translation
  - one phrase at a time
- Translation options
  - most likely translations for each word and phrase
  - ordered and color-highlighted by probability

# Our Types of Assistance

- Sentence completion
  - tool suggests how to complete the translation
  - one phrase at a time
- Translation options
  - most likely translations for each word and phrase
  - ordered and color-highlighted by probability
- Postediting machine translation
  - start with machine translation output
  - user edits, tool shows changes

# Technical Notes



- Online at <http://www.caitra.org/>
- User uploads source text, translates one sentence at a time
- Implementation
  - AJAX Web 2.0 using Ruby on Rails, mySQL
  - Back end: Moses machine translation system

# Predicting Sentence Completion

[1] Paul Newman le magnifique >>

Paul

enter Newman

- Tool makes a suggestion how to continue (in red)

# Predicting Sentence Completion

[1] Paul Newman le magnifique >>

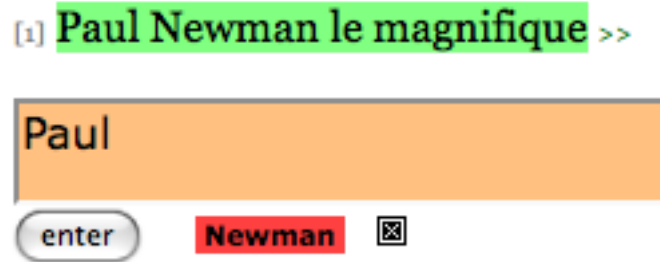
Paul

enter Newman

- Tool makes a suggestion how to continue (in red)
- User can accept it (by pressing TAB), or type in her own translation



# Predicting Sentence Completion



- Tool makes a suggestion how to continue (in red)
- User can accept it (by pressing TAB), or type in her own translation
- Same idea as TransType, with minor modifications
  - show only short text chunks, not full sentence completion
  - show only one suggestion, not alternatives

# How does it work?



- Uses search graph of SMT decoding

## How does it work?

- Uses search graph of SMT decoding
- Matches partial user translation against search graph, by optimizing
  1. minimal string edit distance between path in graph and user translation
  2. best full path probability, including best completion to end

# How does it work?

- Uses search graph of SMT decoding
- Matches partial user translation against search graph, by optimizing
  1. minimal string edit distance between path in graph and user translation
  2. best full path probability, including best completion to end
- Technical notes
  - search graph is pre-computed and stored in database
  - matching is done server-side, typically takes less than 1 second
  - completion path is returned to client (web browser)

# Translation Options

Paul	Newman	le magnifique
Paul	Newman	the wonderful
Mr	Newman ,	the magnificent
Mr Paul	Newman here	the wonderful
as Paul	Committee	beautiful
another	Newman , who speaks	magnificent
with Paul		the splendid
, Paul		the excellent
of Paul		the beautiful
work of Paul		It
the words of Paul		great

- For each word and phrases: suggested translations
- Ranked (and color-highlighted) by probability
- User may click on suggestion → appended to text box

# Translation Options - How does it work?<sup>21</sup>



- Uses phrase translation table of SMT system

# Translation Options - How does it work?<sup>21</sup>



- Uses phrase translation table of SMT system
- Translation score: future cost estimate
  - conditional probabilities  $\phi(\bar{e}|\bar{f}), \phi(\bar{f}|\bar{e})$
  - lexical probabilities  $\text{lex}(\bar{e}|\bar{f}), \text{lex}(\bar{f}|\bar{e})$
  - word count feature
  - language model estimate

# Translation Options - How does it work?<sup>21</sup>



- Uses phrase translation table of SMT system
- Translation score: future cost estimate
  - conditional probabilities  $\phi(\bar{e}|\bar{f})$ ,  $\phi(\bar{f}|\bar{e})$
  - lexical probabilities  $\text{lex}(\bar{e}|\bar{f})$ ,  $\text{lex}(\bar{f}|\bar{e})$
  - word count feature
  - language model estimate
- Ranking of shorter vs. longer phrases by including outside future cost estimate



Translation Tool
pkoehn
logout

**Sentence 2 of 20** [\[1\]](#) [\[2\]](#) [\[4\]](#) [\[6\]](#) [\[8\]](#) [\[11\]](#) [\[13\]](#) [\[16\]](#) [\[19\]](#)

[1] Spitzen von Hamburger CDU und Grünen öffnen Weg zu Koalitionsverhandlungen  
 [2] Das erste schwarz-grüne Bündnis auf Landesebene rückt näher: Die Spitzen von CDU und Grünen in Hamburg halten ihre Differenzen für überwindbar. [3] In einer Sondierungsrunde beschlossen sie, in den Parteigremien über den Start von Koalitionsverhandlungen zu beraten.  
 [4] Hamburg - Sechs Stunden sprachen sie miteinander. [5] Dann verkündeten CDU-Chef Michael Freytag und Grünen-Chefin Anja Hajduk, das Trennende zwischen den Parteien sei überbrückbar.

[1] Leaders of the Hamburger CDU and Greens open path to coalition negotiations.  
 [5] Then the CDU-leader Michael Freytag and Green party leader Anja Hajduk the division between the parties is bridgable.

<< [2] Das erste schwarz-grüne Bündnis auf Landesebene rückt näher: Die Spitzen von CDU und Grünen in Hamburg halten ihre Differenzen für überwindbar. >>



enter the first

das	erste	schwarz	@-@	grüne	Bündnis	auf	Landesebene	rückt	näher	:	die	Spitzen
the first	black	@-@	green	alliance	in favour of	is approaching	:	the leaders				
the	first	black	@-@	green	the alliance	in favour	approaches	that	the people at the top			
for the first	black		Green	Alliance	on	national	we are coming to	.	at the top			
this		in black and white	@-@	green	cooperation	in	Belarus approaches		the top			
the first of	the black		the Greens	NATO	seek to	we	closer	the	this			

# Postediting Machine Translation

<< [2] L'inoubliable interprète de "Butch Cassidy et le Kid" est mort des suites d'un cancer, à l'âge de 83 ans, dans sa maison du Connecticut. >>  
The unforgettable ~~interpreter~~ actor of " Butch Cassidy and the Sundance Kid " died as a result of cancer ~~7~~ at the age of 83 ~~years~~ ~~7~~ in his house in Connecticut . (9 edits)

The unforgettable actor of "Butch Cassidy and the Sundance Kid" died as a result of cancer at the age of 83 in his house in Connecticut.

- Textbox is initially filled with machine translation
- User edits translation
- String edit distance to machine translation is shown (blue background)

# Evaluation



- Recall setup
  - 10 students, half native French, half native English
  - each student translated French-English news stories
  - about 40 sentences for each condition of assistance

# Evaluation

- Recall setup
  - 10 students, half native French, half native English
  - each student translated French-English news stories
  - about 40 sentences for each condition of assistance
- Five different conditions
  - unassisted
  - prediction (sentence completion)
  - options
  - predictions and options
  - post-editing

# Quality

- We want faster translators, but not worse
- Assessment of translation quality
  - show translations to bilingual judges, with source
  - judgment: fully correct? yes/no

Indicate whether each user's input represents a fully fluent and meaning-equivalent translation of the source. The source is shown with context, the actual sentence is bold.

# Quality

- We want faster translators, but not worse
- Assessment of translation quality
  - show translations to bilingual judges, with source
  - judgment: fully correct? yes/no

Indicate whether each user's input represents a fully fluent and meaning-equivalent translation of the source. The source is shown with context, the actual sentence is bold.

- Average score: 50% correct — lower than expected
  - judges seemed to be too harsh
  - when given several translations, tendency to judge half as bad

# Example of Quality Judgments

---

Src.	Sans se démonter, il s'est montré concis et précis.	
MT	Without dismantle, it has been concise and accurate.	
<hr/>		
1/3	Without fail, he has been concise and accurate.	(Prediction+Options, L1a)
4/0	Without getting flustered, he showed himself to be concise and precise.	(Unassisted, L1b)
4/0	Without falling apart, he has shown himself to be concise and accurate.	(Postedit, L1c)
1/3	Unswayable, he has shown himself to be concise and to the point.	(Options, L1d)
0/4	Without showing off, he showed himself to be concise and precise.	(Prediction, L1e)
1/3	Without dismantling himself, he presented himself consistent and precise.	(Prediction+Options, L2a)
2/2	He showed himself concise and precise.	(Unassisted, L2b)
3/1	Nothing daunted, he has been concise and accurate.	(Postedit, L2c)
3/1	Without losing face, he remained focused and specific.	(Options, L2d)
3/1	Without becoming flustered, he showed himself concise and precise.	(Prediction, L2e)

# Faster and Better



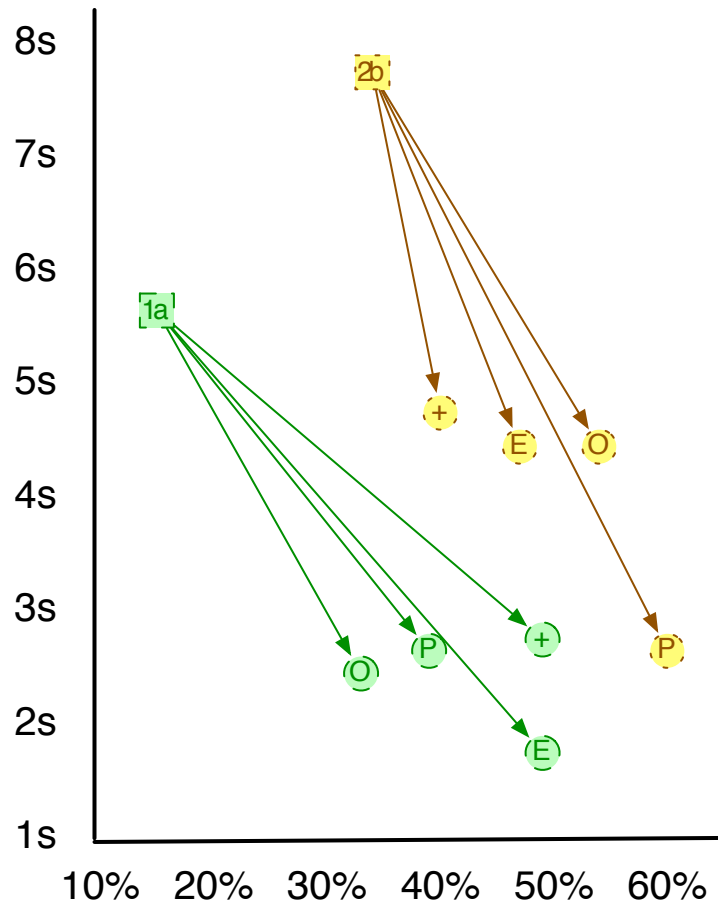
<b>Assistance</b>	<b>Speed</b>	<b>Quality</b>
Unassisted	4.4s/word	47% correct
Postedit	2.7s (-1.7s)	55% (+8%)
Options	3.7s (-0.7s)	51% (+4%)
Prediction	3.2s (-1.2s)	54% (+7%)
Prediction+Options	3.3s (-1.1s)	53% (+6%)



# Faster and Better, Mostly

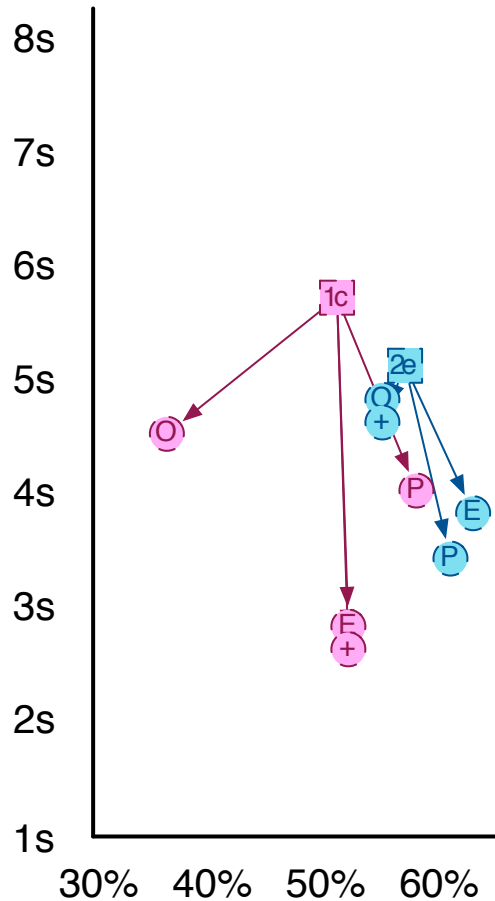
User	Unassisted	Postedit		Options		Prediction		Prediction+Options	
L2a	3.3sec/word 23% correct	<b>1.2s</b> <b>39%</b>	<b>-2.2s</b> <b>+16%)</b>	<b>2.3s</b> <b>45%</b>	<b>-1.0s</b> <b>+22%</b>	<b>1.1s</b> <b>30%</b>	<b>-2.2s</b> <b>+7%)</b>	<b>2.4s</b> <b>44%</b>	<b>-0.9s</b> <b>+21%</b>
L2b	7.7sec/word 35% correct	<b>4.5s</b> <b>48%</b>	<b>-3.2s)</b> <b>+13%</b>	<b>4.5s</b> <b>55%</b>	<b>-3.3s</b> <b>+20%</b>	<b>2.7s</b> <b>61%</b>	<b>-5.1s</b> <b>+26%</b>	<b>4.8s</b> <b>41%</b>	<b>-3.0s</b> <b>+6%</b>
L2c	3.9sec/word 50% correct	<b>1.9s</b> <b>61%</b>	<b>-2.0s</b> <b>+11%</b>	<b>3.8s</b> <b>54%</b>	<b>-0.1s</b> <b>+4%</b>	<b>3.1s</b> <b>64%</b>	<b>-0.8s</b> <b>+14%</b>	<b>2.5s</b> <b>61%</b>	<b>-1.4s</b> <b>+11%</b>
L2d	2.8sec/word 38% correct	<b>2.0s</b> <b>46%</b>	<b>-0.7s</b> <b>+8%</b>	2.9s 59%	(+0.1s) (+21%)	2.4s 37%	(-0.4s) (-1%)	<b>1.8s</b> <b>45%</b>	<b>-1.0s</b> <b>+7%</b>
L2e	5.2sec/word 58% correct	<b>3.9s</b> <b>64%</b>	<b>-1.3s</b> <b>+6%</b>	4.9s 56%	(-0.2s) (-2%)	<b>3.5s</b> <b>62%</b>	<b>-1.7s</b> <b>+4%</b>	4.6s 56%	(-0.5s) (-2%)
L1a	5.7sec/word 16% correct	<b>1.8s</b> <b>50%</b>	<b>-3.9s</b> <b>+34%</b>	<b>2.5s</b> <b>34%</b>	<b>-3.2s</b> <b>+18%</b>	<b>2.7s</b> <b>40%</b>	<b>-3.0s</b> <b>+24%</b>	<b>2.8s</b> <b>50%</b>	<b>-2.9s</b> <b>+34%</b>
L1b	3.2sec/word 64% correct	2.8s 56%	(-0.4s) (-8%)	3.5s 60%	+0.3s -4%	6.0s 61%	+2.8s -3%	4.6s 57%	+1.4s -7%
L1c	5.8sec/word 52% correct	<b>2.9s</b> <b>53%</b>	<b>-3.0s</b> <b>+1%</b>	4.6s 37%	(-1.2s) (-15%)	<b>4.1s</b> <b>59%</b>	<b>-1.7s</b> <b>+7%</b>	<b>2.7s</b> <b>53%</b>	<b>-3.1s</b> <b>+1%</b>
L1d	3.4sec/word 49% correct	3.1s 49%	(-0.3s) (+0%)	4.3s 51%	(+0.9s) (+2%)	3.8s 53%	(+0.4s) (+4%)	3.7s 58%	(+0.3s) (+9%)
L1e	2.8sec/word 68% correct	<b>2.6s</b> <b>79%</b>	<b>-0.2s</b> <b>+11%</b>	3.5s 59%	+0.7s -9%	2.8s 64%	(-0.0s) (-4%)	3.0s 66%	+0.2s -2%
avg.	4.4sec/word 47% correct	<b>2.7s</b> <b>55%</b>	<b>-1.7s</b> <b>+8%</b>	<b>3.7s</b> <b>51%</b>	<b>-0.7s</b> <b>+4%</b>	<b>3.2s</b> <b>54%</b>	<b>-1.2s</b> <b>+7%</b>	<b>3.3s</b> <b>53%</b>	<b>-1.1s</b> <b>+6%</b>

# Slow Users 1: Faster and Better



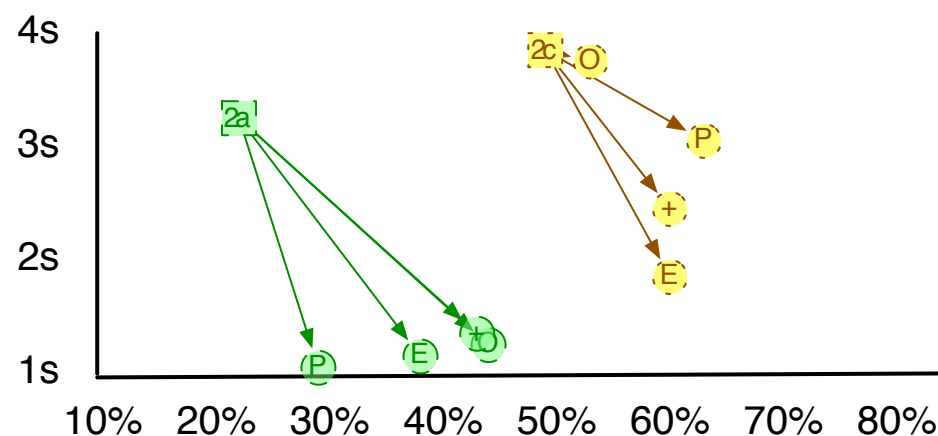
- Unassisted
  - more than 5 seconds per input word
  - very bad (35%, 16%)
  
- With assistance
  - much faster and better
  - reaching roughly average performance

# Slow Users 2: Only Faster



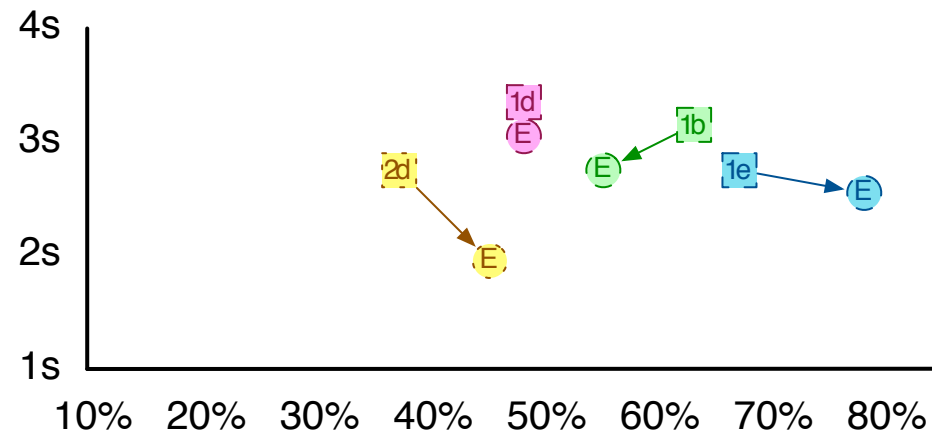
- Unassisted
  - more than 5 seconds per input word
  - average quality
- With assistance
  - faster and but not better

## Fast Users



- Unassisted
  - fast: 3-4 seconds per input word
  - L2a is very bad (23%), L2c is average (50%)
- With assistance
  - faster and better
  - L2a closer to average (30-45%), L2c becomes very good (54-61%)

# Refuseniks

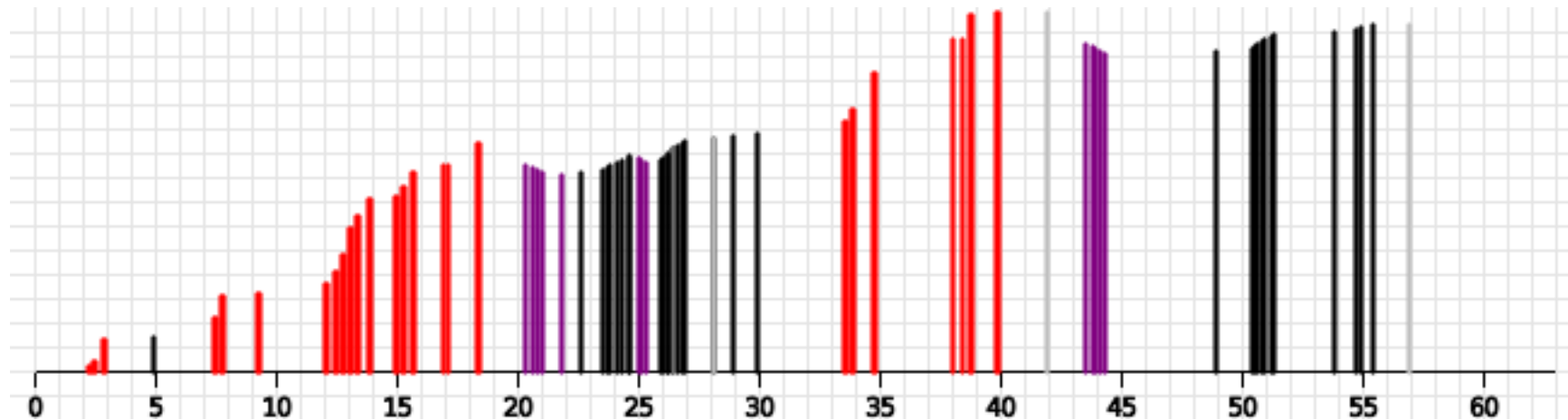


- Use the assistance sparingly or not at all, and see generally no gains
- The two best translators are in this group
- Postediting
  - mixed on quality (2 better, 1 worse, 1 same), but all faster
  - best translator (L1e, 68%) becomes much better (record 79%)

## Further Analysis

- How does the assistance change translator behaviour?
- How do translators utilize assistance?
- How is the translation produced?

# Keystroke Log



black: keystroke, purple: deletion, grey: cursor move  
red: sentence completion accept  
orange: click on translation option

**Analysis:** Segment into periods of activity: typing, **tabbing**, **clicking**, pauses

one second before and after a keystroke is part of typing interval

# Activities: Native French User L2b



User: L2b	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s	-	-
Postedit	4.5s	1.5s	0.4s	0.1s	1.0s	0.4s	1.1s	-	-
Options	4.5s	0.6s	0.1s	0.4s	0.9s	0.7s	1.5s	0.4s	-
Prediction	2.7s	0.3s	0.3s	0.2s	0.7s	0.1s	0.6s	-	0.4s
Prediction+Options	4.8s	0.6s	0.4s	0.4s	1.3s	0.5s	0.9s	0.5s	0.2s



# Activities: Native French User L2b



User: L2b	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s	-	-
Postedit	4.5s	1.5s	0.4s	0.1s	1.0s	0.4s	1.1s	-	-
Options	4.5s	0.6s	0.1s	0.4s	0.9s	0.7s	1.5s	0.4s	-
Prediction	2.7s	0.3s	0.3s	0.2s	0.7s	0.1s	0.6s	-	0.4s
Prediction+Options	4.8s	0.6s	0.4s	0.4s	1.3s	0.5s	0.9s	0.5s	0.2s

Slightly less  
time spent  
on typing

# Activities: Native French User L2b

User: L2b	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s	-	-
Postedit	4.5s	1.5s	0.4s	0.1s	1.0s	0.4s	1.1s	-	-
Options	4.5s	0.6s	0.1s	0.4s	0.9s	0.7s	1.5s	0.4s	-
Prediction	2.7s	0.3s	0.3s	0.2s	0.7s	0.1s	0.6s	-	0.4s
Prediction+Options	4.8s	0.6s	0.4s	0.4s	1.3s	0.5s	0.9s	0.5s	0.2s

Less  
pausing

Slightly less  
time spent  
on typing

# Activities: Native French User L2b

User: L2b	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	7.7s	1.3s	0.1s	0.3s	1.8s	1.9s	2.3s	-	-
Postedit	4.5s	1.5s	0.4s	0.1s	1.0s	0.4s	1.1s	-	-
Options	4.5s	0.6s	0.1s	0.4s	0.9s	0.7s	1.5s	0.4s	-
Prediction	2.7s	0.3s	0.3s	0.2s	0.7s	0.1s	0.6s	-	0.4s
Prediction+Options	4.8s	0.6s	0.4s	0.4s	1.3s	0.5s	0.9s	0.5s	0.2s

Less  
pausing

Especially  
less time  
in big  
pauses

Slightly less  
time spent  
on typing

# Activities: Native English User L1e



<b>User: L1e</b>	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s	-	-
Postedit	2.6s	0.4s	0.3s	0.2s	1.0s	0.1s	0.7s	-	-
Options	3.5s	0.1s	0.3s	0.4s	0.6s	0.2s	1.7s	0.1s	-
Prediction	2.8s	0.1s	0.3s	0.3s	0.3s	-	1.4s	-	0.3s
Prediction+Options	3.0s	0.1s	0.3s	0.2s	0.5s	-	1.9s	-	-

# Activities: Native English User L1e



User: L1e	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s	-	-
Postedit	2.6s	0.4s	0.3s	0.2s	1.0s	0.1s	0.7s	-	-
Options	3.5s	0.1s	0.3s	0.4s	0.6s	0.2s	1.7s	0.1s	-
Prediction	2.8s	0.1s	0.3s	0.3s	0.3s	-	1.4s	-	0.3s
Prediction+Options	3.0s	0.1s	0.3s	0.2s	0.5s	-	1.9s	-	-

Little time  
spent on  
assistance

# Activities: Native English User L1e

User: L1e	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s	-	-
Postedit	2.6s	0.4s	0.3s	0.2s	1.0s	0.1s	0.7s	-	-
Options	3.5s	0.1s	0.3s	0.4s	0.6s	0.2s	1.7s	0.1s	-
Prediction	2.8s	0.1s	0.3s	0.3s	0.3s	-	1.4s	-	0.3s
Prediction+Options	3.0s	0.1s	0.3s	0.2s	0.5s	-	1.9s	-	-

Does not use both assistances, little overall change

Little time spent on assistance

# Activities: Native English User L1e

User: L1e	total	init-p	end-p	short-p	mid-p	big-p	key	click	tab
Unassisted	2.8s	0.3s	0.2s	0.2s	0.3s	0.1s	1.9s	-	-
Postedit	2.6s	0.4s	0.3s	0.2s	1.0s	0.1s	0.7s	-	-
Options	3.5s	0.1s	0.3s	0.4s	0.6s	0.2s	1.7s	0.1s	-
Prediction	2.8s	0.1s	0.3s	0.3s	0.3s	-	1.4s	-	0.3s
Prediction+Options	3.0s	0.1s	0.3s	0.2s	0.5s	-	1.9s	-	-

Does not use both assistances, little overall change

Postediting:  
less typing (-1.2s)  
more medium pauses (+0.7s)

Little time spent on assistance

# Origin of Characters: Native French L2b<sup>43</sup>



User: L2b	key	click	tab	mt
Postedit	18%	-	-	81%
Options	59%	40%	-	-
Prediction	14%	-	85%	-
Prediction+Options	21%	44%	33%	-



# Origin of Characters: Native French L2b<sup>44</sup>



User: L2b	key	click	tab	mt
Postedit	18%	-	-	81%
Options	59%	40%	-	-
Prediction	14%	-	85%	-
Prediction+Options	21%	44%	33%	-

Translation comes to large degree from assistance

# Origin of Characters: Native English L1e<sup>45</sup>



<b>User: L1e</b>	key	click	tab	mt
Postedit	20%	-	-	79%
Options	77%	22%	-	-
Prediction	61%	-	38%	-
Prediction+Options	100%	-	-	-

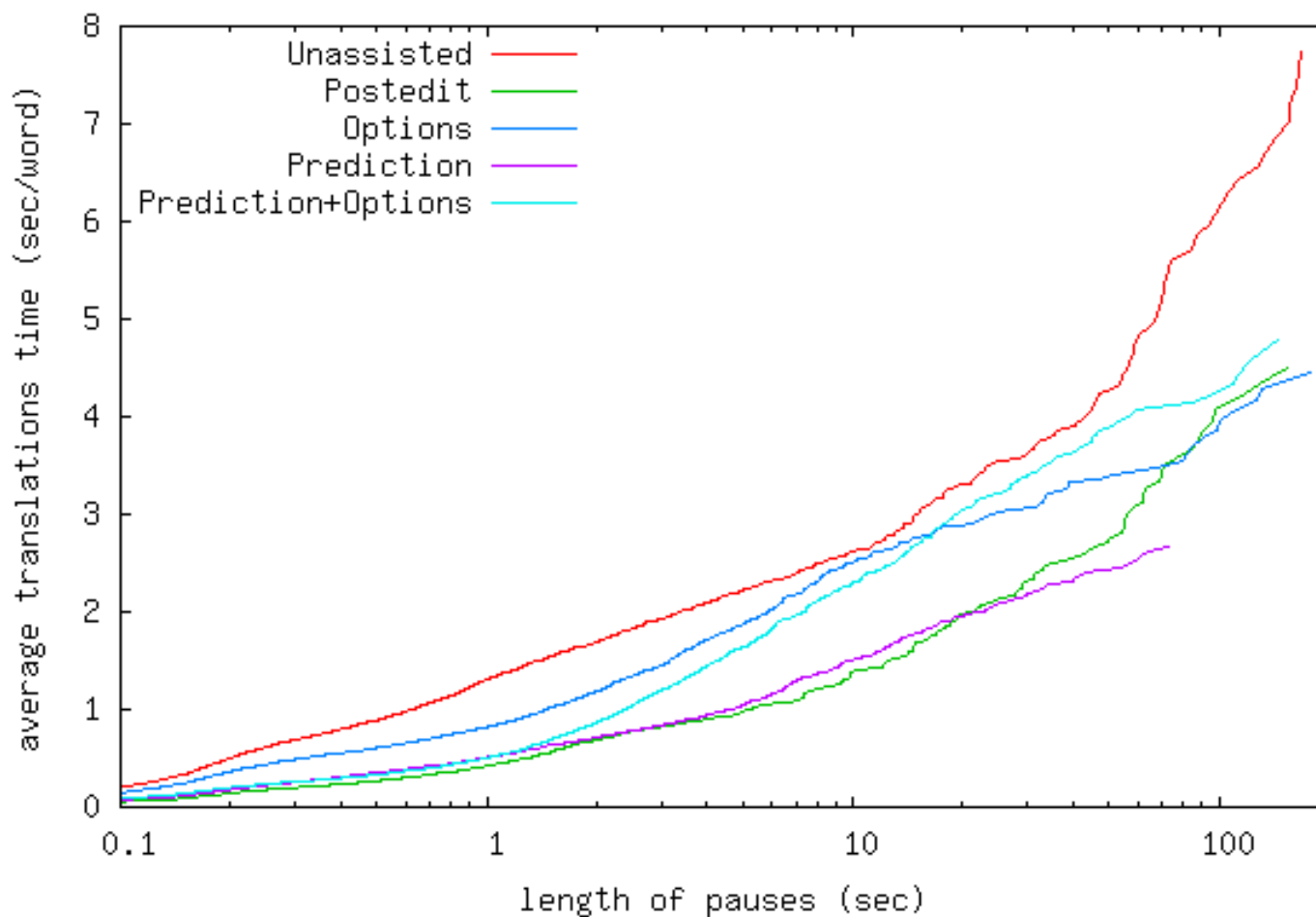
# Origin of Characters: Native English L1e<sup>46</sup>



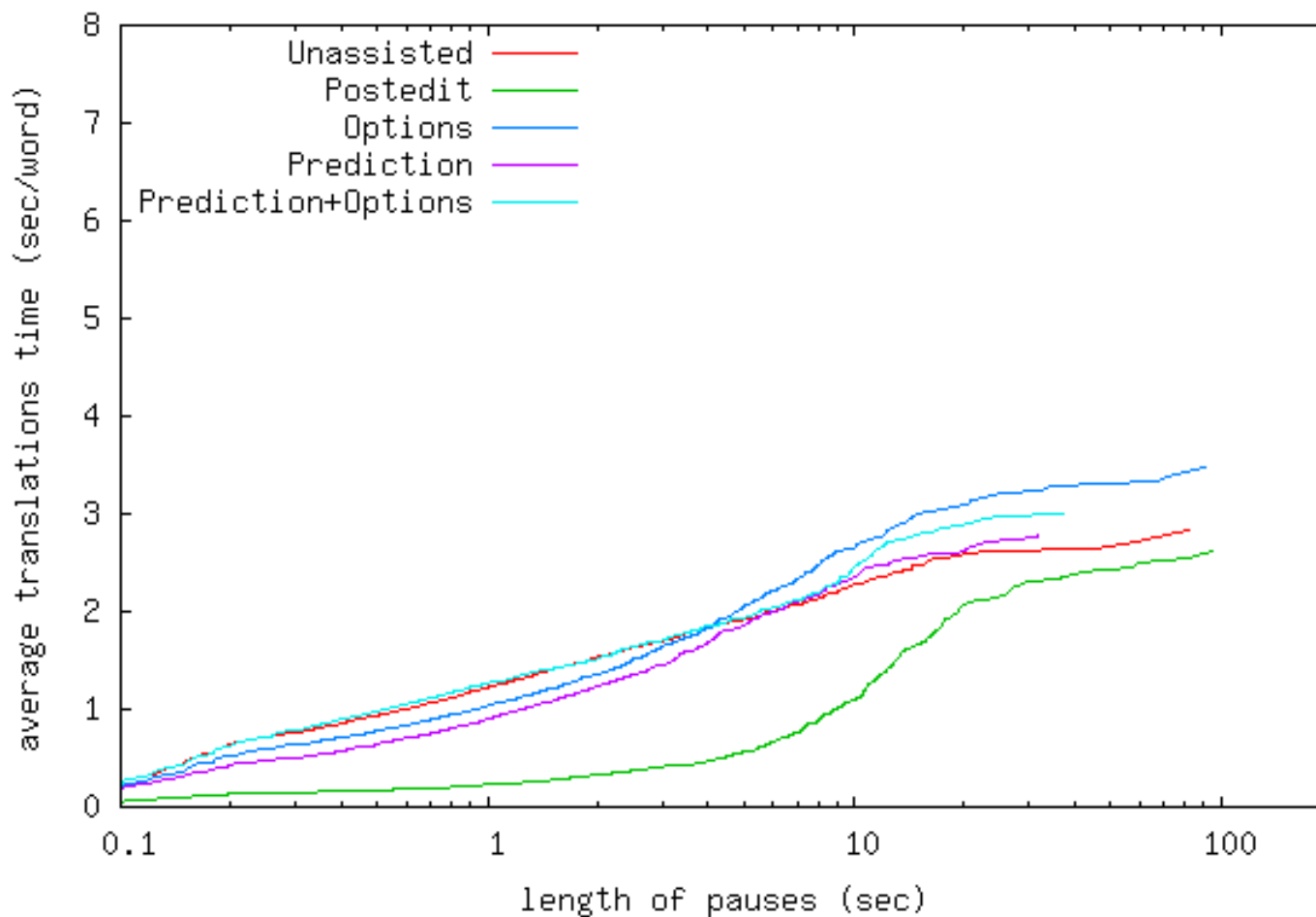
User: L1e	key	click	tab	mt
Postedit	20%	-	-	79%
Options	77%	22%	-	-
Prediction	61%	-	38%	-
Prediction+Options	100%	-	-	-

Although hardly any time spent on assistance, fair amount of characters produced by it

# Pauses: French-Native User L2b

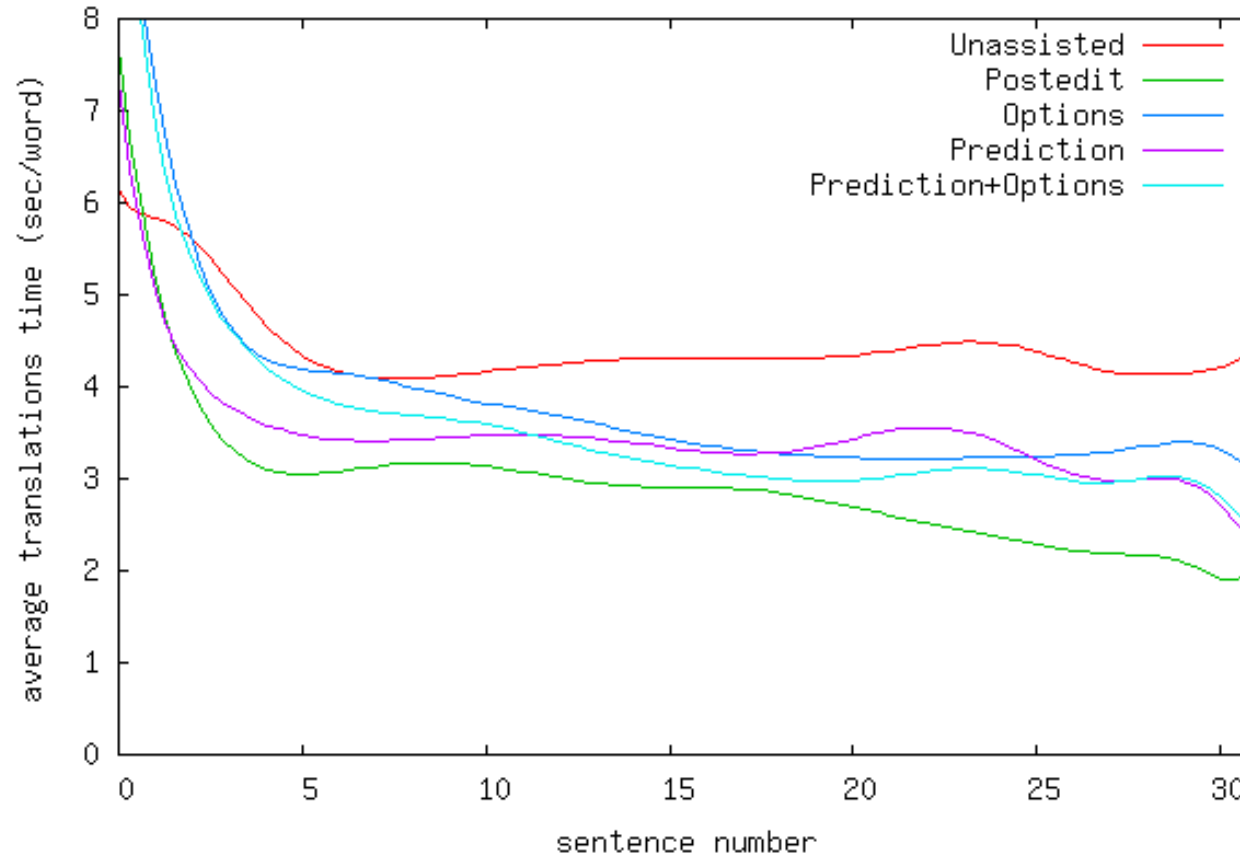


# Pauses: English-Native User L1e



# Learning Curve

users become better over time with assistance



# User Feedback



- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - postediting: 1 user

# User Feedback



- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - postediting: 1 user
- Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  - prediction+options: 4.6
  - prediction: 3.9
  - options: 3.7
  - postediting: 2.9



# User Feedback

- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - postediting: 1 user
  
- Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  - prediction+options: 4.6
  - prediction: 3.9
  - options: 3.7
  - postediting: 2.9

# User Feedback

- Q: In which of the five conditions did you think you were most accurate?
  - predictions+options: 5 users
  - options: 2 users
  - prediction: 1 user
  - **postediting: 1 user**
- Q: Rank the different types of assistance on a scale from 1 to 5, where 1 indicates not at all and 5 indicates very helpful.
  - prediction+options: 4.6
  - prediction: 3.9
  - options: 3.7
  - **postediting: 2.9**
- **Note: does not match empirical results**

# Summary



- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing

# Summary



- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing
- Assistance made translators better
  - average judgment increased from 47% to 51-55% with help
  - even good translators get better with postediting

# Summary



- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing
- Assistance made translators better
  - average judgment increased from 47% to 51-55% with help
  - even good translators get better with postediting
- Some good translators ignored the assistance

# Summary



- Assistance made translators faster
  - average speed improvement from 4.4s/word to 2.7-3.7s/word
  - reduction of big pauses
  - reduction of typing effort in post-editing
- Assistance made translators better
  - average judgment increased from 47% to 51-55% with help
  - even good translators get better with postediting
- Some good translators ignored the assistance
- Fastest and (barely) best with postediting, but did not like it

## Outlook: More analysis



- What do translators think about when they are pausing?

# Outlook: More analysis



- What do translators think about when they are pausing?
- What are the hard problems?
  - unknown words
  - words without direct translation
  - syntactic re-arrangement



## Outlook: More analysis

- What do translators think about when they are pausing?
- What are the hard problems?
  - unknown words
  - words without direct translation
  - syntactic re-arrangement
- What do translators change in post-editing?

# Outlook: More experiments



- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language

# Outlook: More experiments



- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language
- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem

# Outlook: More experiments



- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language
- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem
- Different types of translation tasks
  - content unfamiliar to translator
  - test material very similar to previously translated text

# Outlook: More experiments



- Different types of users
  - experienced professional translators
  - volunteer / amateur
  - no/little knowledge of source language
- Different types of language pairs
  - target-side morphology a problem
  - large-scale reordering maybe a problem
- Different types of translation tasks
  - content unfamiliar to translator
  - test material very similar to previously translated text

# Outlook: Better Assistance



- Confidence estimation for post-editing
  - we have training data: pairs of MT output, corrected sentence
  - supervised binary classification problem (correct/false)

# Outlook: Better Assistance



- Confidence estimation for post-editing
  - we have training data: pairs of MT output, corrected sentence
  - supervised binary classification problem (correct/false)
- Visualizing word alignment (not only for post-editing)

# Outlook: Better Assistance



- Confidence estimation for post-editing
  - we have training data: pairs of MT output, corrected sentence
  - supervised binary classification problem (correct/false)
- Visualizing word alignment (not only for post-editing)
- Better prediction model
  - different metrics to find best matching prefix
    - \* BLEU, TER, ...
    - \* word substitution scored by letter string edit distance
  - prediction as a machine learning multi-class classification problem
  - prediction based on parse forest from tree-based models



# Outlook: Additional Types of Assistance<sup>56</sup>



- Bilingual concordancer
  - show translations of words and phrases in context

# Outlook: Additional Types of Assistance<sup>56</sup>



- Bilingual concordancer
  - show translations of words and phrases in context
- Monolingual concordancer / language model
  - show which expressions are more or less common
  - language model to highlight unusual transitions in output

# Outlook: Additional Types of Assistance<sup>56</sup>



- Bilingual concordancer
  - show translations of words and phrases in context
- Monolingual concordancer / language model
  - show which expressions are more or less common
  - language model to highlight unusual transitions in output
- Translation memory
  - finding most similar sentence in parallel corpus
  - show translation, highlight difference, fill in difference with MT

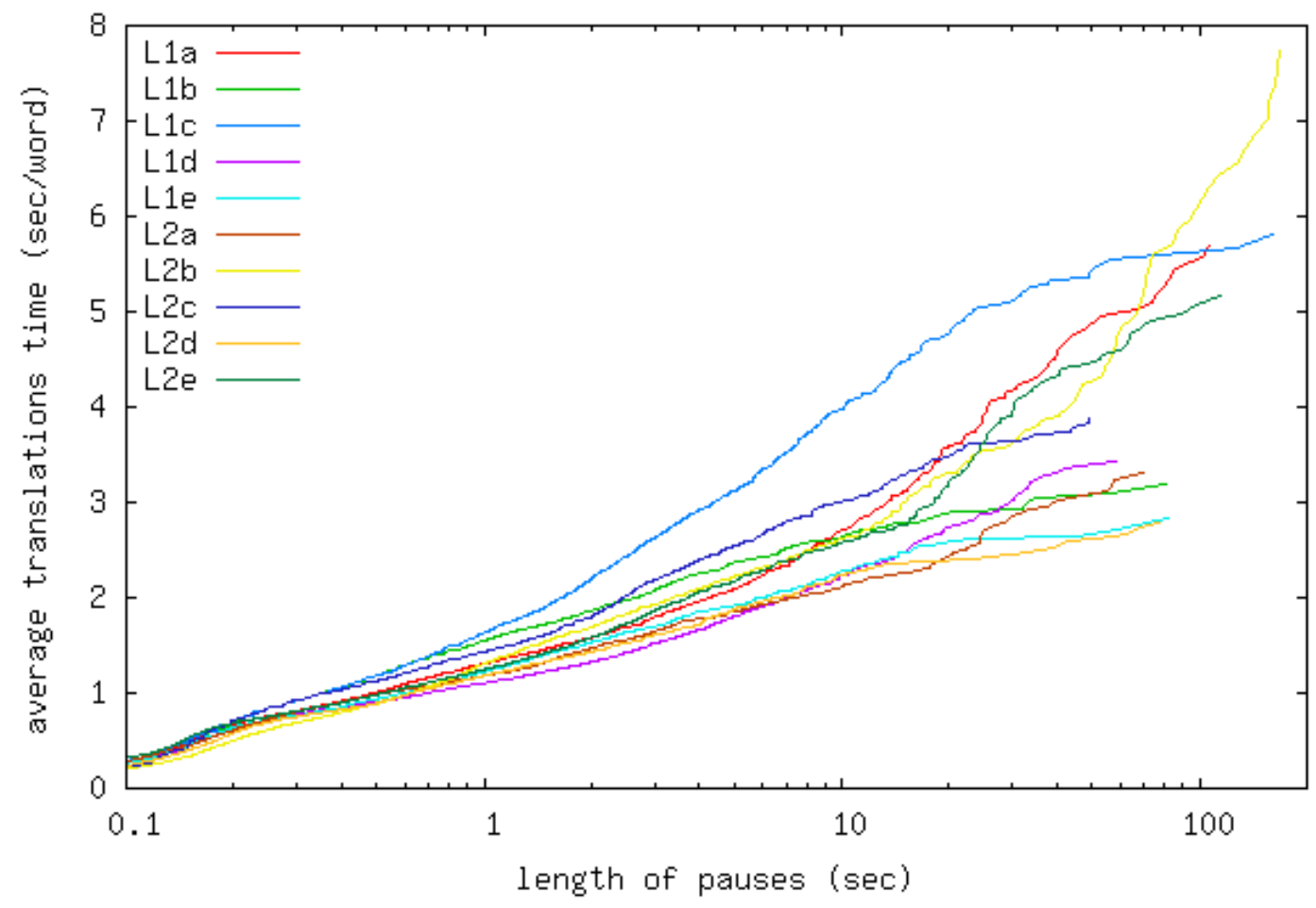
**Try it at home!**



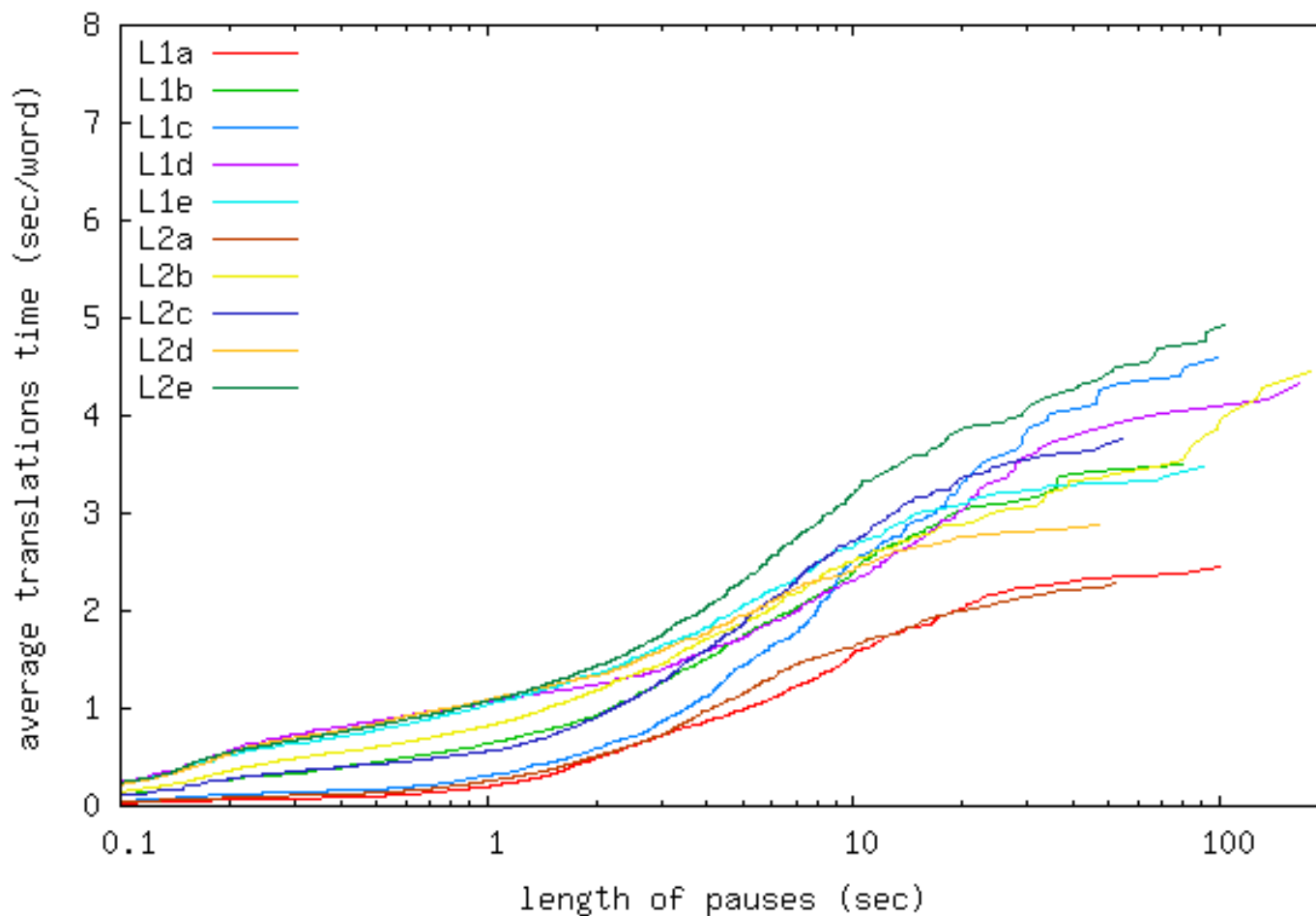
<http://www.caitra.org/>

questions?

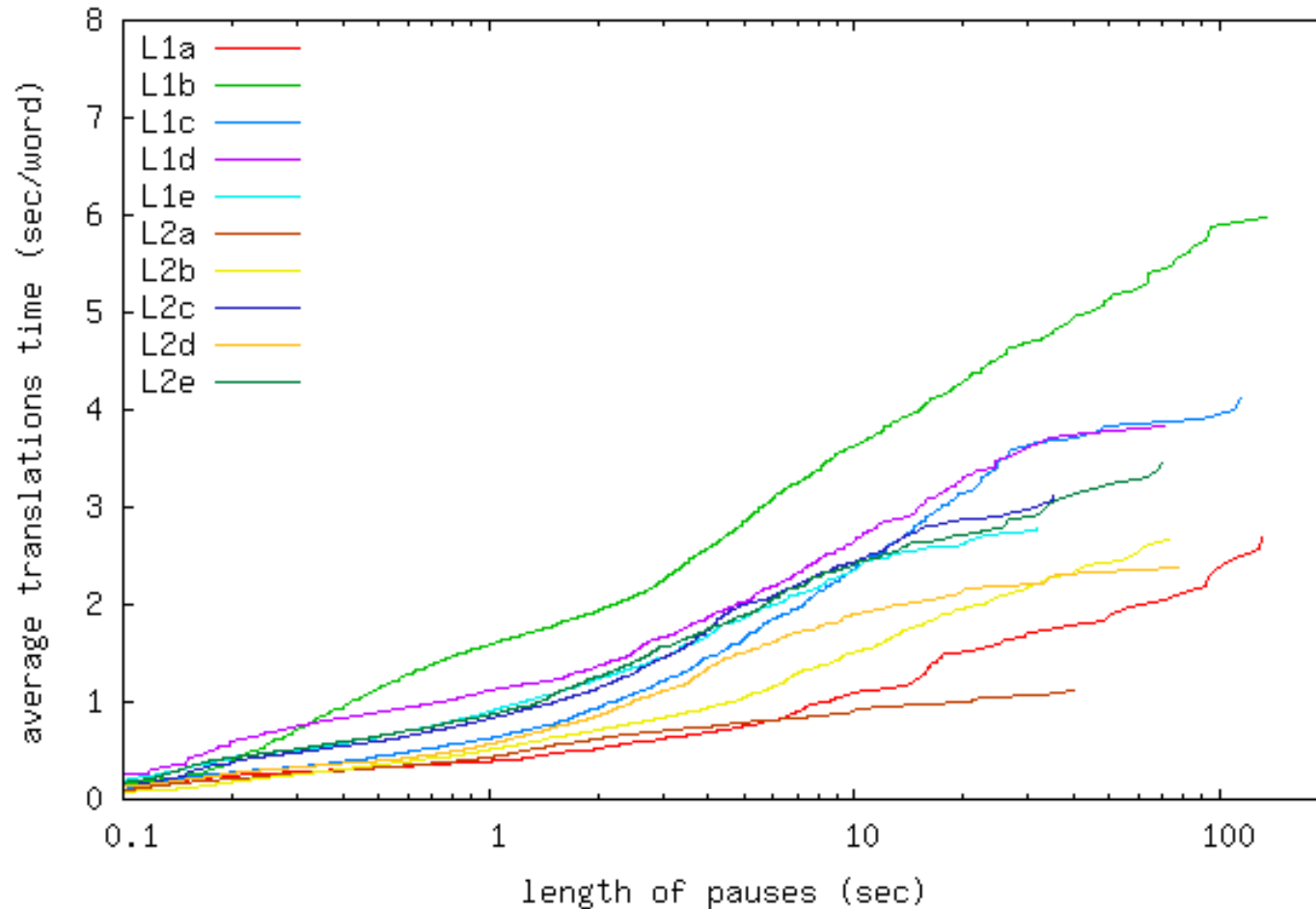
# Pauses: Unassisted



# Pauses: Options



# Pauses: Prediction of sentence completion



# Pauses: Postediting

