Human Translation and Machine Translation

Philipp Koehn

1 December 2009

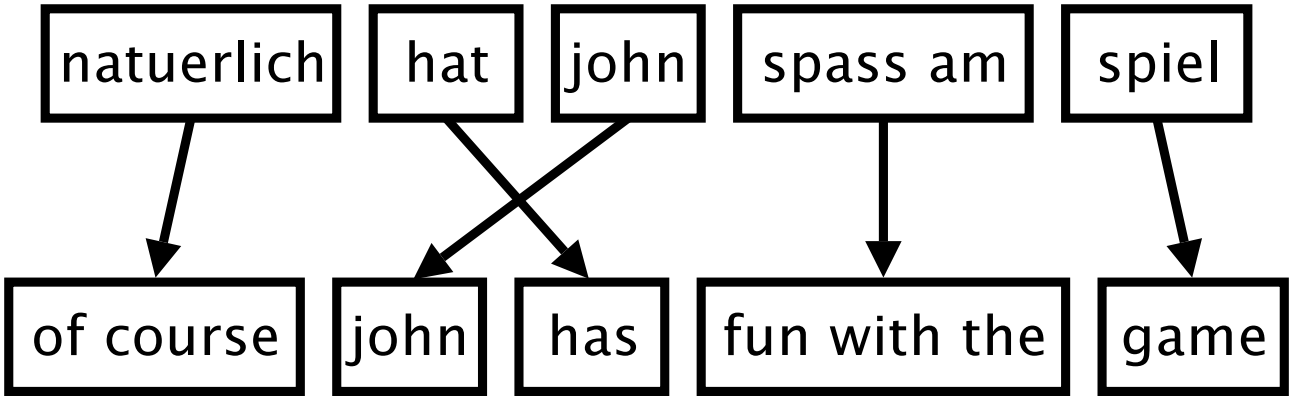
Overview

- Machine Translation
- Human Translation
- Assistance to Human Translators
- User Study 1
- User Study 2

Overview

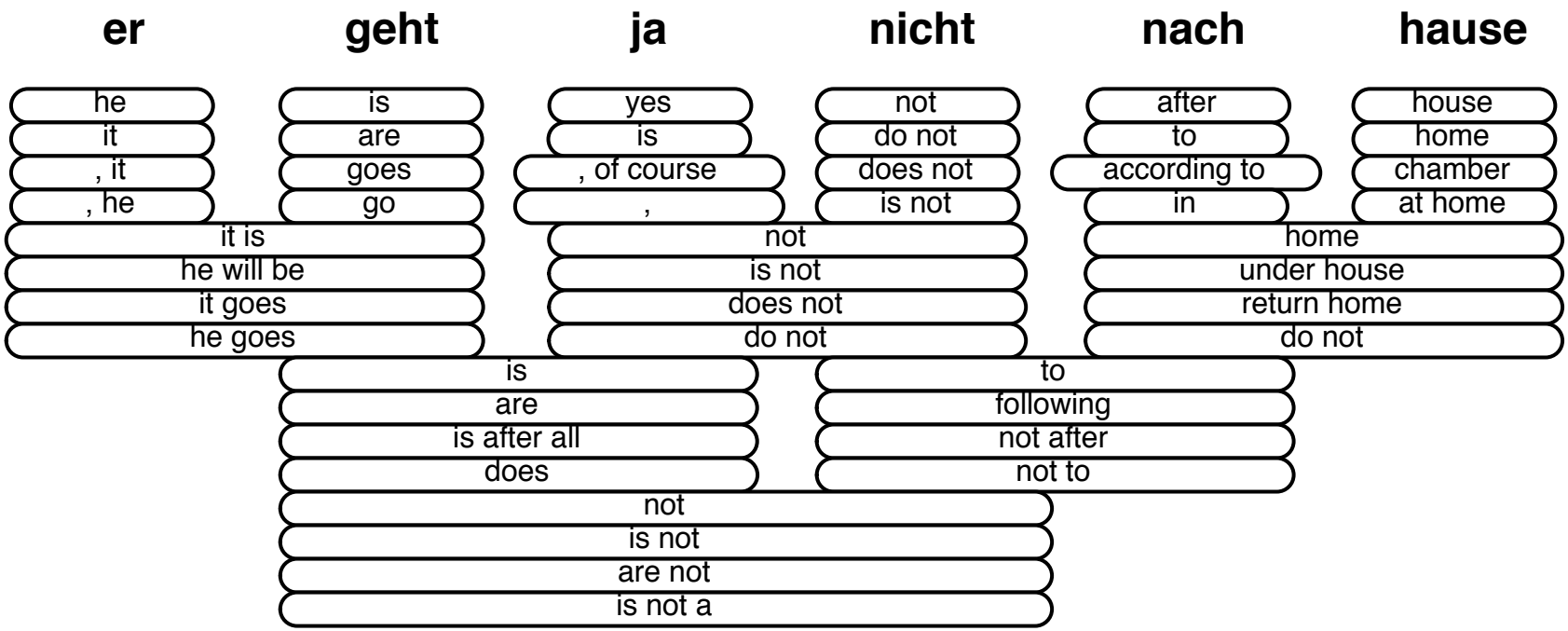
- **Machine Translation**
- Human Translation
- Assistance to Human Translators
- User Study 1
- User Study 2

Phrase-Based Translation



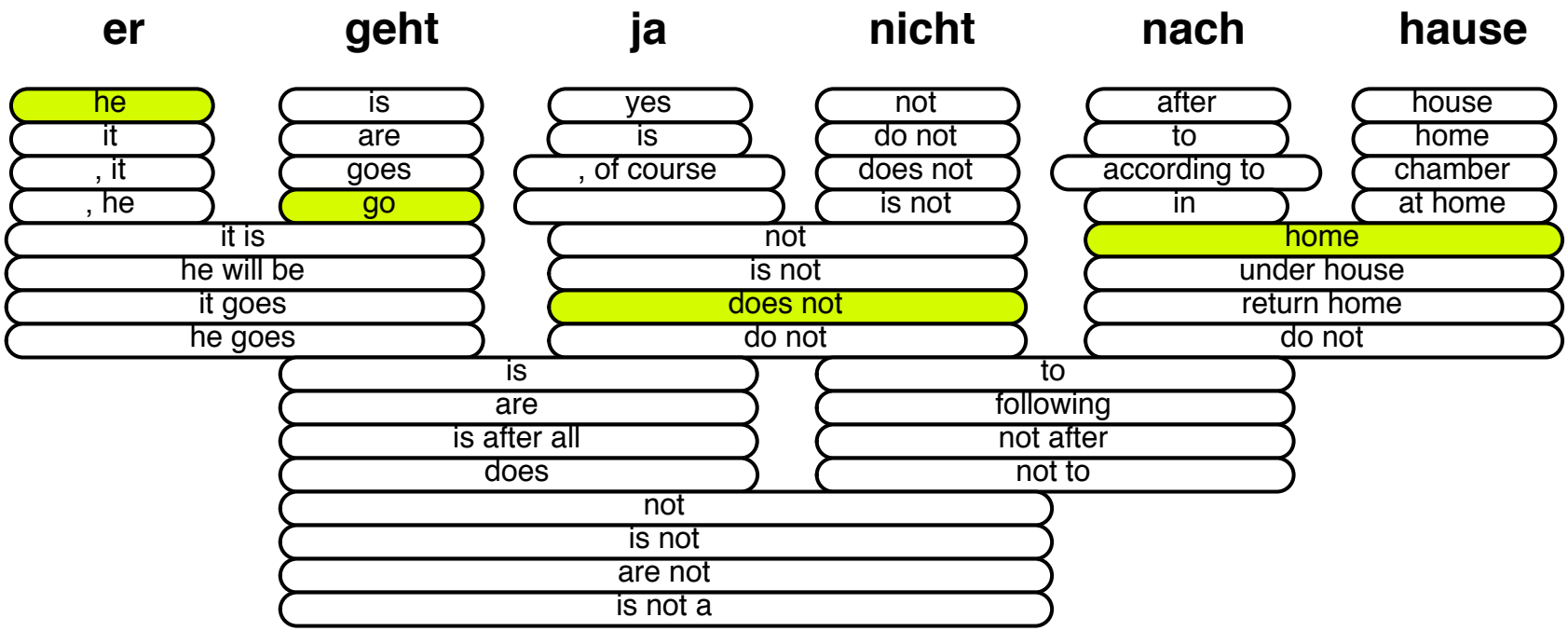
- Foreign input is segmented in phrases
  - any sequence of words, not necessarily linguistically motivated
- Each phrase is translated into English
- Phrases are reordered

Translation options



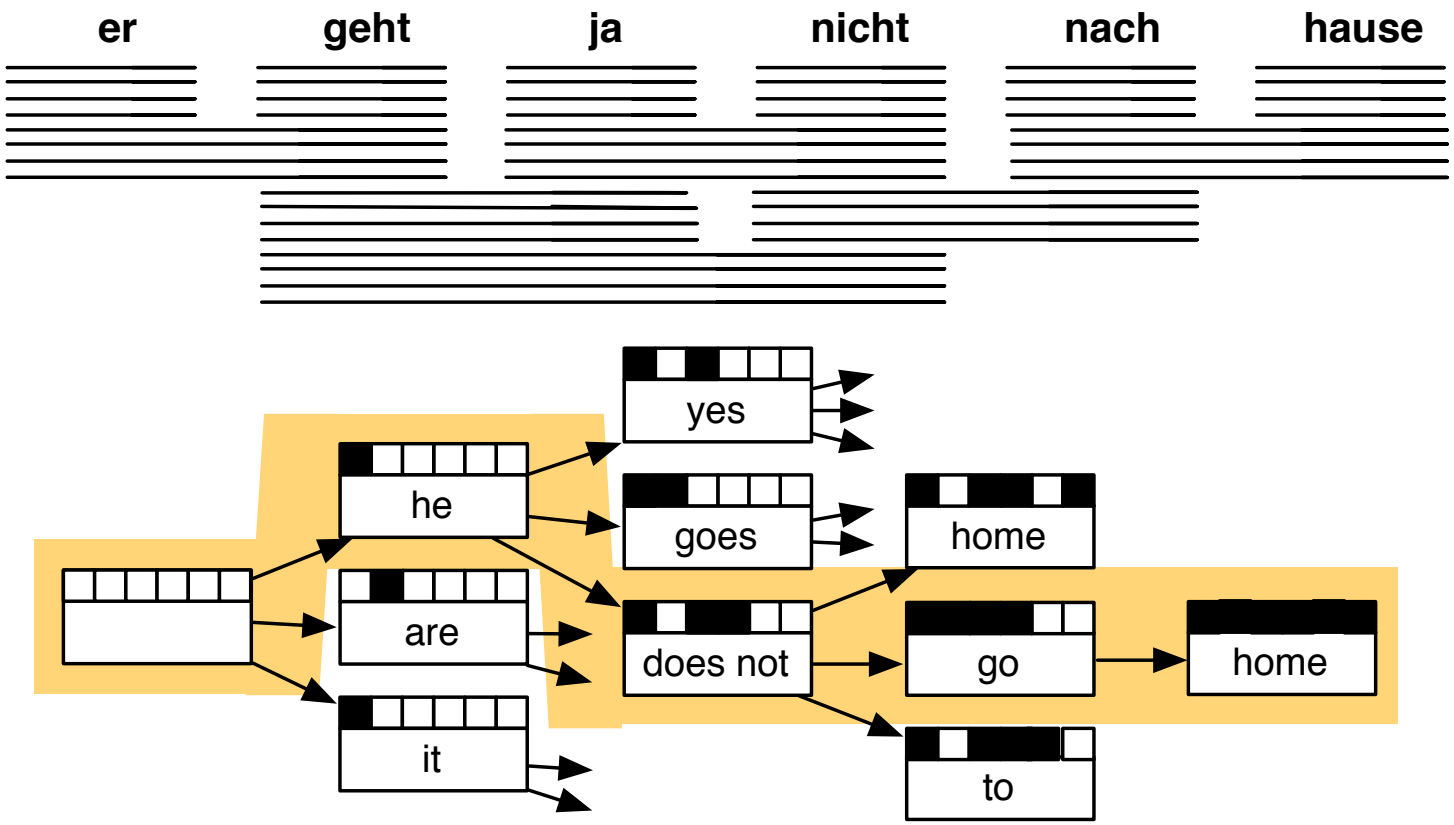
- Many translation options to choose from

Translation options



- Many translation options to choose from

Decoding process: find best path



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.

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

Overview

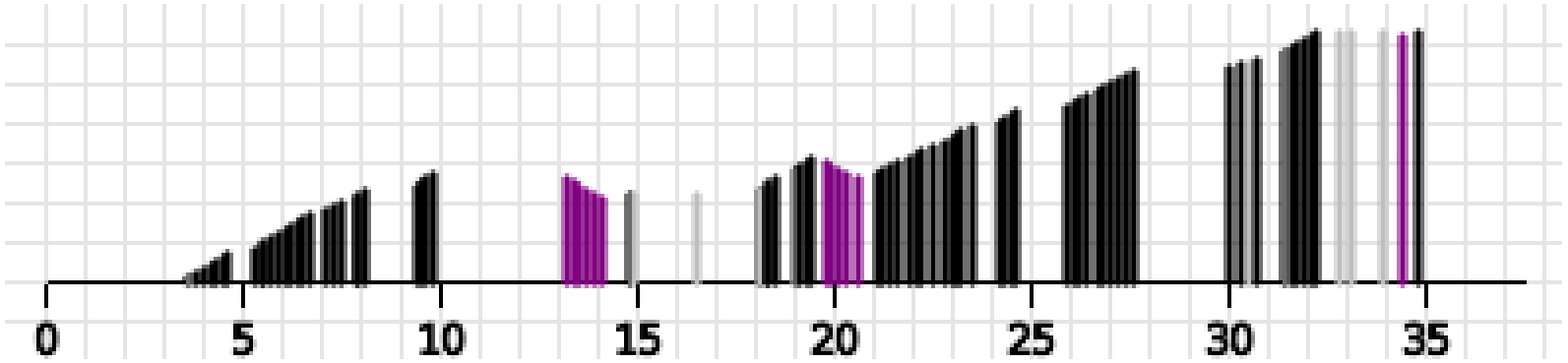
- Machine Translation
- **Human Translation**
- Assistance to Human Translators
- User Study 1
- User Study 2

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.



(37.5 sec, 3.4 sec/word)

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

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

L1 = native French, L2 = native English

average time per input word

Time Spent on Activities

not much time

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

L1 = native French, L2 = native English

average time per input word

Time Spent on Activities

not much time

similar

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

L1 = native French, L2 = native English

average time per input word

Time Spent on Activities

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

L1 = native French, L2 = native English

average time per input word

Pauses Reconsidered

- Our classification of pauses is arbitrary (2-6sec, 6-60sec, >60sec)

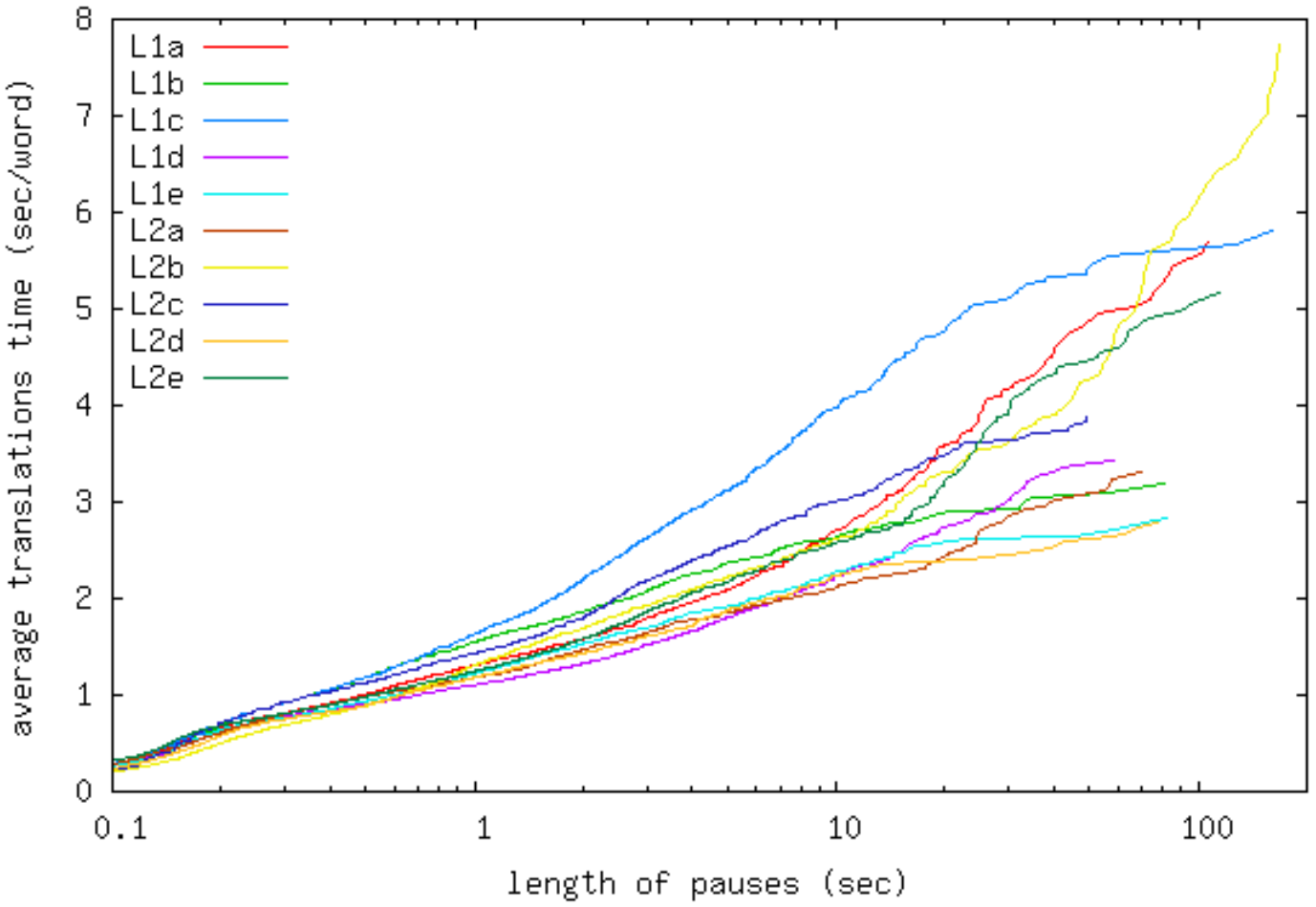
- Extreme view: all you see is pauses
  - keystrokes take no observable time
  - all you see is pauses between action points

- 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


L1a
L1b
L1c
L1d
L1e
L2a
L2b
L2c
L2d
L2e
average translations time (sec/word)
length of pauses (sec)
0
1
2
3
4
5
6
7
8
0.1
1
10
100

Overview

- Machine Translation
- Human Translation
- **Assistance to Human Translators**
- User Study 1
- User Study 2

Related Work: Tools used by Translators

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

show translations in context (www.linguee.com)

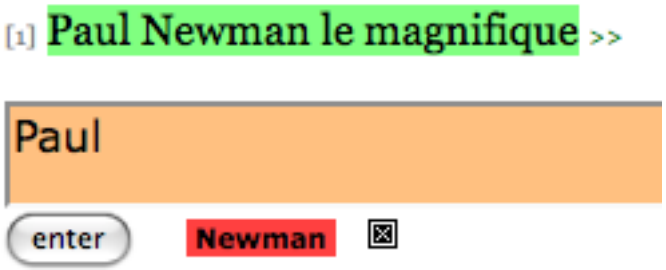
Our Types of Assistance

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



- 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

- 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 brower)

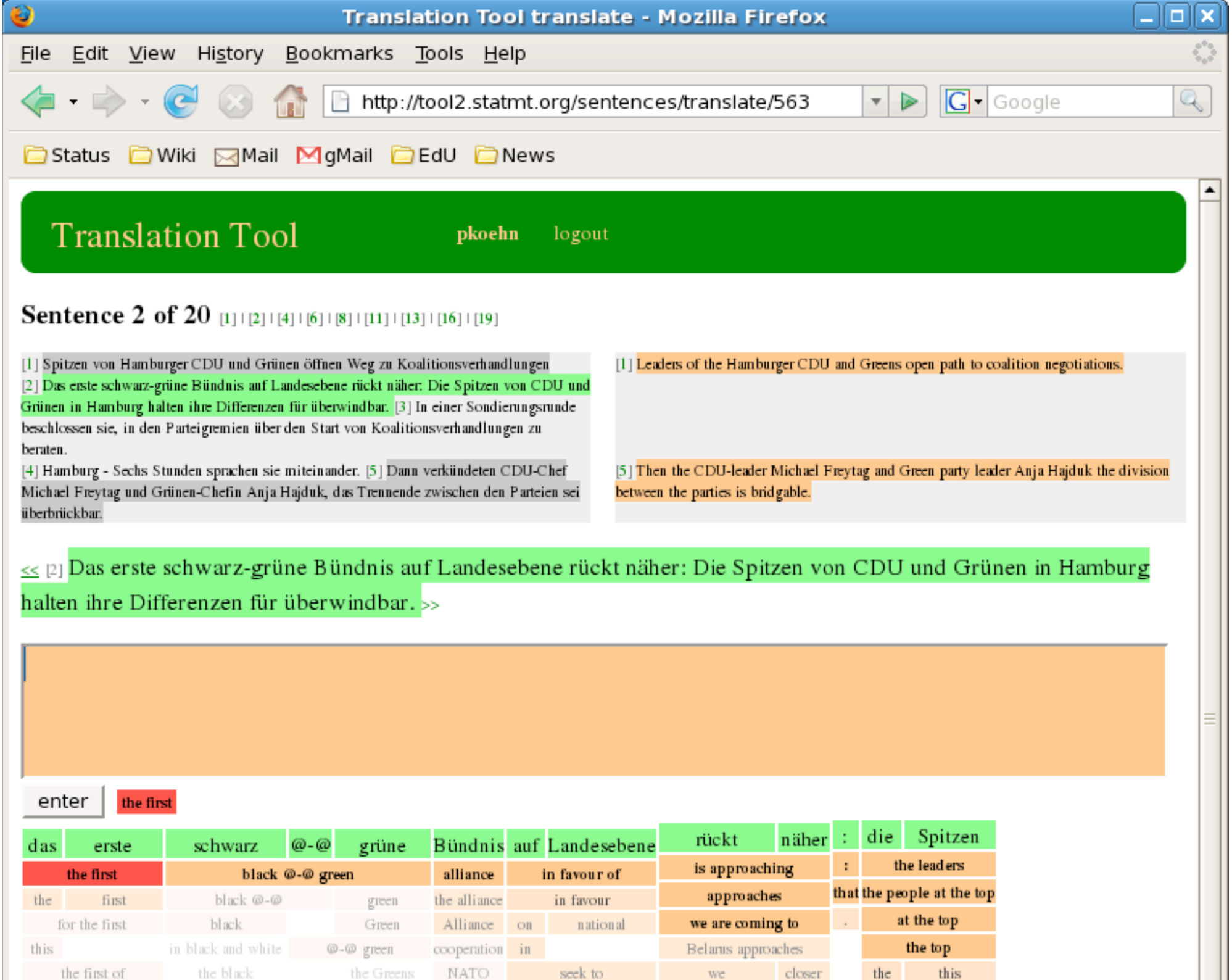
Translation Options

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?

- 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 translate - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://tool2.statmt.org/sentences/translate/563
Google
Status Wiki Mail gMail EdU News
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 ~~,~~ at the age of 83 ~~years~~ ~~,~~ 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)

Overview

- Machine Translation
- Human Translation
- Assistance to Human Translators
- **User Study 1**
- User Study 2

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.

- 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.	
1/3	Without fail, he has been concise and accurate.	*(Prediction+Options, L2a)*
4/0	Without getting flustered, he showed himself to be concise and precise.	*(Unassisted, L2b)*
4/0	Without falling apart, he has shown himself to be concise and accurate.	*(Postedit, L2c)*
1/3	Unswayable, he has shown himself to be concise and to the point.	*(Options, L2d)*
0/4	Without showing off, he showed himself to be concise and precise.	*(Prediction, L2e)*
1/3	Without dismantling himself, he presented himself consistent and precise.	*(Prediction+Options, L1a)*
2/2	He showed himself concise and precise.	*(Unassisted, L1b)*
3/1	Nothing daunted, he has been concise and accurate.	*(Postedit, L1c)*
3/1	Without losing face, he remained focused and specific.	*(Options, L1d)*
3/1	Without becoming flustered, he showed himself concise and precise.	*(Prediction, L1e)*

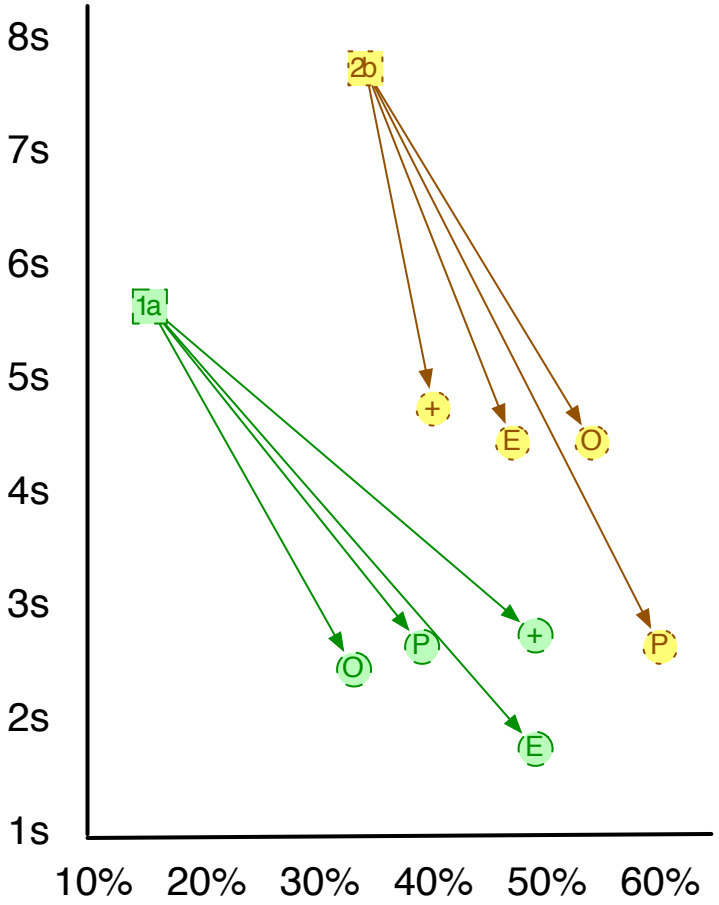
Faster and Better

Assistance	**Speed**	**Quality**
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

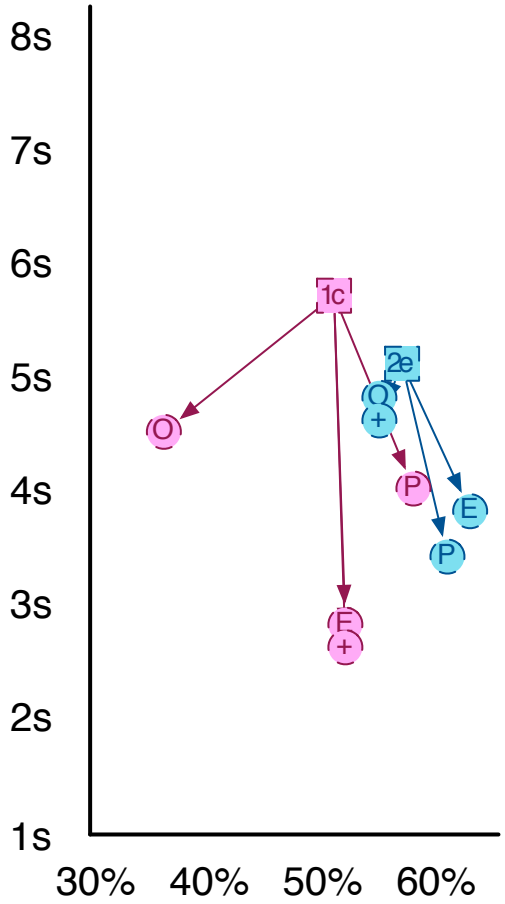
User	Unassisted	Postedit		Options		Prediction		Prediction+Options	
L1a	3.3sec/word 23% correct	**1.2s** **39%**	**-2.2s** **+16%)**	**2.3s** **45%**	**-1.0s** **+22%**	**1.1s** **30%**	**-2.2s** **+7%)**	**2.4s** **44%**	**-0.9s** **+21%**
L1b	7.7sec/word 35% correct	**4.5s** **48%**	**-3.2s)** **+13%**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

Slow Users 1: Faster and Better



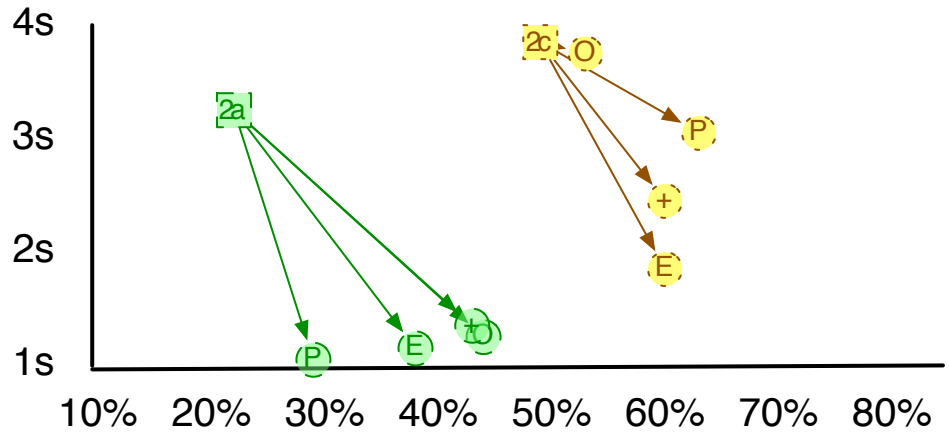
- Unassisted
  - more than 5 seconds per input word
  - very bad (35%, 16%)
- With assistance
  - much faster and better
  - reaching roughly average performance

Slow Users 2: Only Faster



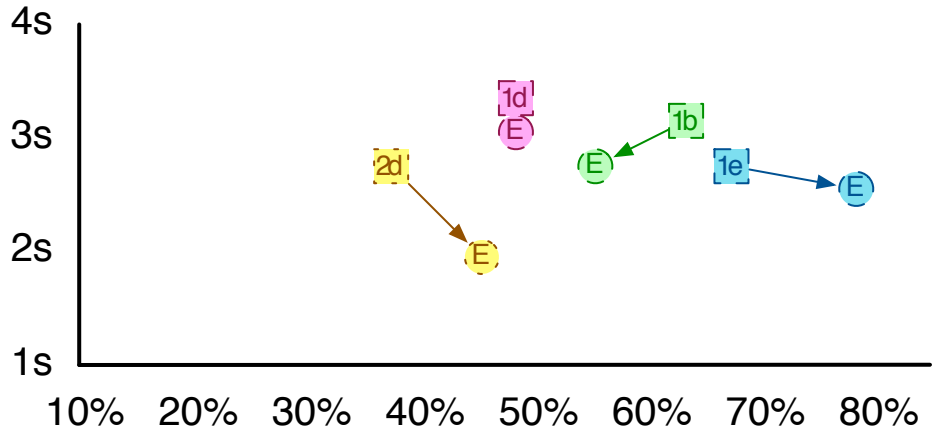
- Unassisted
  - more than 5 seconds per input word
  - average quality
- With assistance
  - faster and but not better

Fast Users



- Unassisted
  - fast: 3-4 seconds per input word
  - L1a is very bad (23%), L1c is average (50%)
- With assistance
  - faster and better
  - L1a closer to average (30-45%), L1c becomes very good (54-61%)

Refuseniks

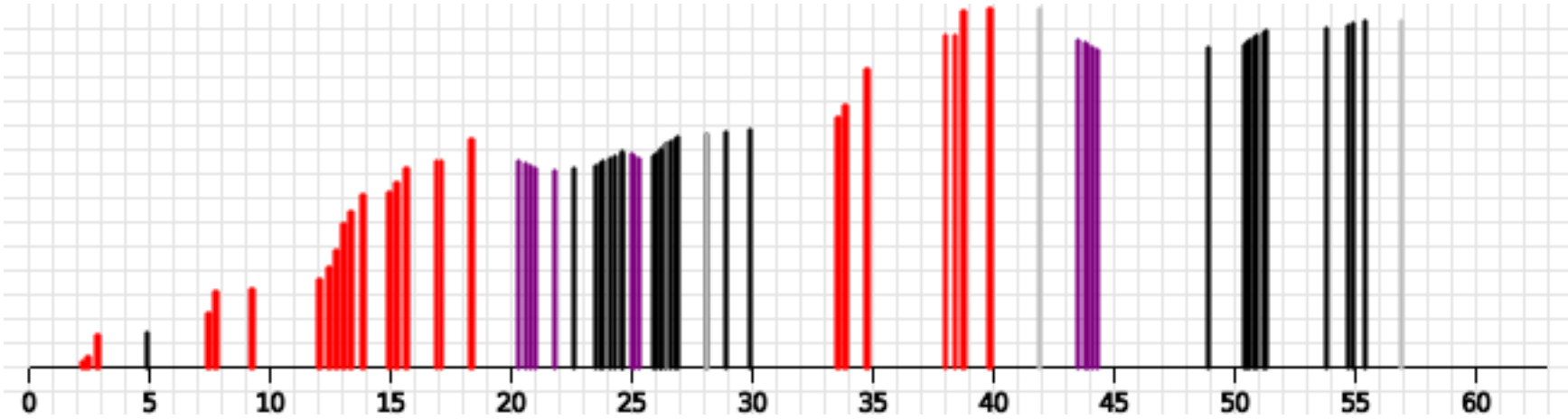


- Use the assistance sparingly or not at all, and see generally no gains
- The two best translators are in this group
- Postediting
  - mixed on quality (2 better, 1 worse, 1 same), but all faster
  - best translator (L2e, 68%) becomes much better (record 79%)

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 L1b

User: L1b	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 L1b

User: L1b	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

Slighly less time spent on typing

Activities: Native French User L1b

User: L1b	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

Slighly less time spent on typing

Activities: Native French User L1b

User: L1b	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

Slighly less time spent on typing

Activities: Native English User L2e

User: L2e	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 L2e

User: L2e	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 L2e

User: L2e	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 L2e

User: L2e	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 L1b

User: L1b	key	click	tab	mt
Postedit	18%	-	-	81%
Options	59%	40%	-	-
Prediction	14%	-	85%	-
Prediction+Options	21%	44%	33%	-

Origin of Characters: Native French L1b

User: L1b	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 L2e

User: L2e	key	click	tab	mt
Postedit	20%	-	-	79%
Options	77%	22%	-	-
Prediction	61%	-	38%	-
Prediction+Options	100%	-	-	-

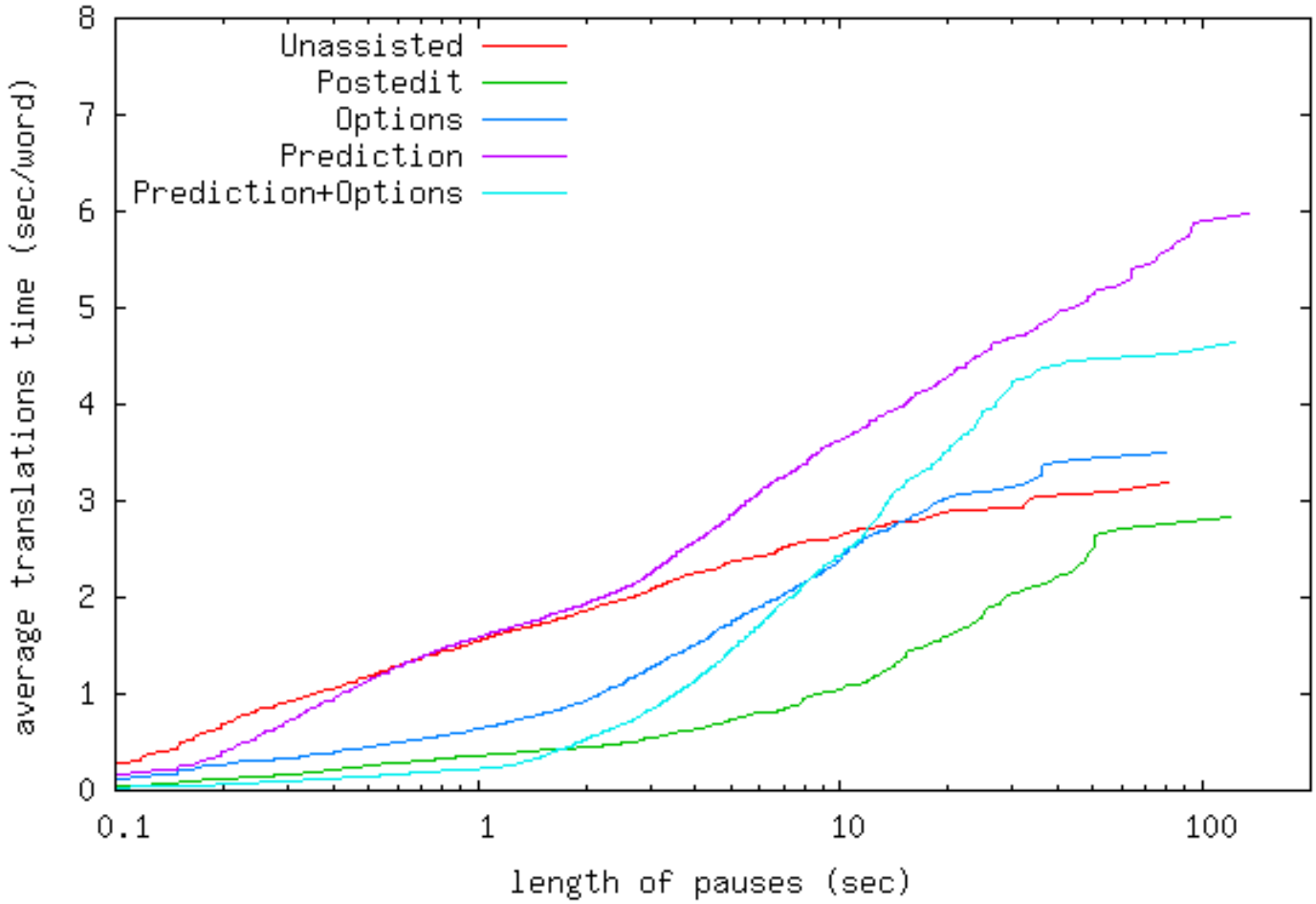
Origin of Characters: Native English L2e

User: L2e	key	click	tab	mt
Postedit	20%	-	-	79%
Options	77%	22%	-	-
Prediction	61%	-	38%	-
Prediction+Options	100%	-	-	-

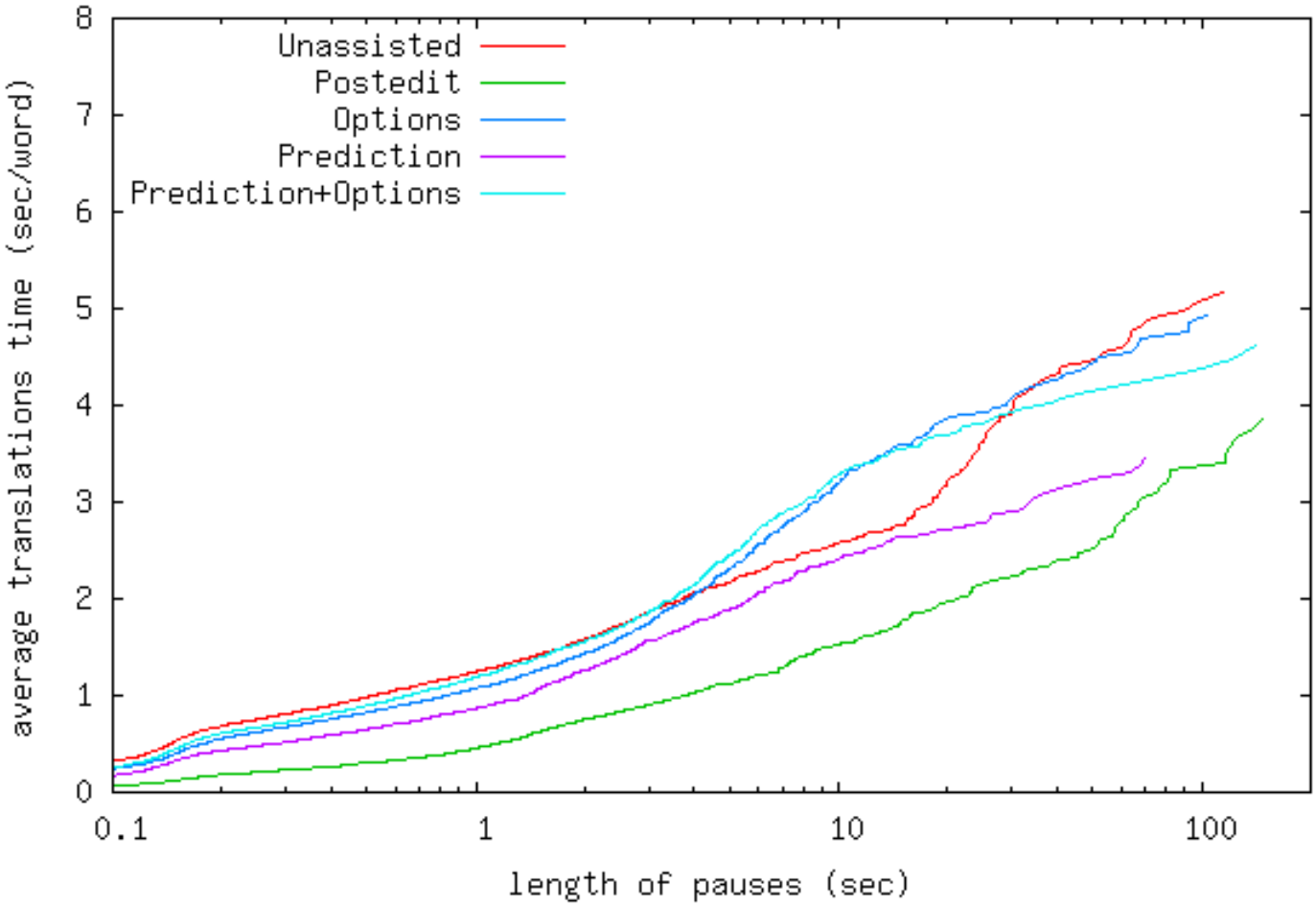
Although hardly any time spent on assistance,

fair amount of characters produced by it

Pauses: French-Native User L1b

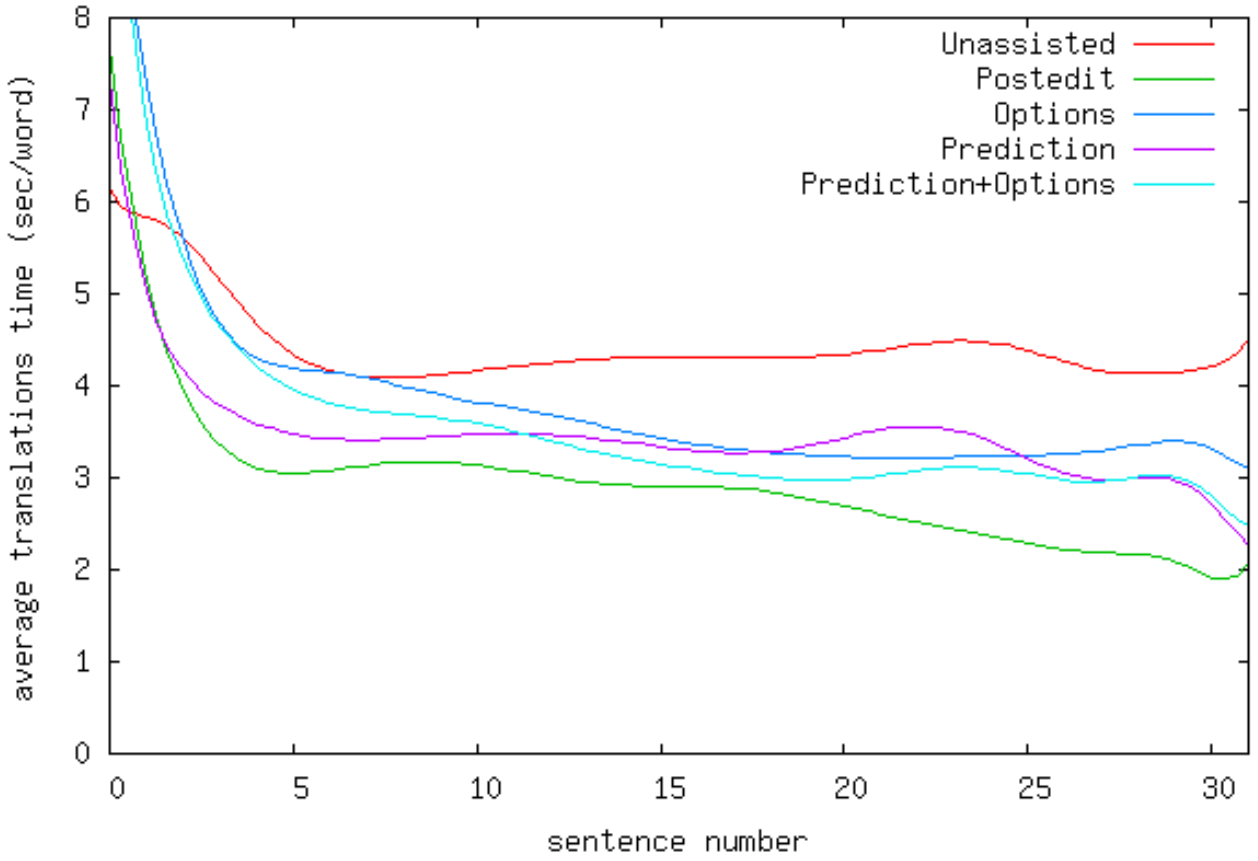


Pauses: English-Native User L2e



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
- Q: Rank the different types of assistance on a scale from 1 to 5, where1 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, where1 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
- 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

Overview

- Machine Translation
- Human Translation
- Assistance to Human Translators
- User Study 1
- **User Study 2**

Monolingual Translators

- Translating when only knowing the target language?
- Why?
  - Low-cost first draft
  - Evaluating machine translation quality: meaning preservation

Setup

- 10 monolingual translators
- 2 types of assistance: Postediting and Options+Prediction
- 2 language pairs: Arabic–English and Chinese–English
  - systems developed under the GALE program
  - close to state of the art (at least for Arabic)
- 8 news stories (4 Arabic, 4 Chinese) from NIST Eval 2008 set
- also in competition: 3 human reference translations

Stories

Story	Headline	Sent.	Words
1: Chinese	White House Pushes for Nuclear Inspectors to Be Sent to Monitor North Korea's Closure of Its Nuclear Reactors	6	207
2: Chinese	Torrential Rains Hit Western India, 43 People Dead	10	204
3: Chinese	Research Shows a Link between Arrhythmia and Two Forms of Genetic Variation	7	247
4: Chinese	Veteran US Goalkeeper Keller May Retire after America's Cup	10	367
5: Arabic	Britain: Arrests in Several Cities and Explosion of Suspicious Car	7	224
6: Arabic	Ban Ki-Moon Withdraws His Report on the Sahara after Controversy Surrounding Its Content	8	310
7: Arabic	Pakistani Opposition Leaders Call on Musharraf to Resign.	11	312
8: Arabic	Al-Maliki: Iraqi Forces Are Capable of Taking Over the Security Dossier Any Time They Want	8	255

Results

Assistance	Arabic	Chinese
Bilingual	63±8%	68±7%
Postediting	39±6%	23±5%
Options+Prediction	40±6%	34±6%

- Judges are very critical... again...
- Monolingual translators not much worse than bilingual translators
- Options help (especially for Chinese)

Individual Translators

Translator	Arabic	Chinese
bi1	73±14%	**62±14%**
bi2	**48±15%**	65±14%
bi3	70±14%	76±11%
mono1	48±15%	30±14%
mono2	37±15%	**12±9%**
mono3	36±16%	15±11%
mono4	60±14%	28±13%
mono5	37±16%	31±15%
mono6	26±12%	18±12%
mono7	20±12%	21±12%
mono8	53±15%	59±14%
mono9	41±15%	31±15%
mono10	35±14%	37±14%

Individual Stories

Story	Language	Bilingual	Postediting	Options+Prediction	
1	Chinese	80±20%	34±18%	56±19%	
2	Chinese	76±13%	36±11%	34±12%	
3	Chinese	61±16%	10±9%	16±10%	bad
4	Chinese	64±13%	13±8%	41±11%	
5	Arabic	69±25%	12±12%	14±13%	
6	Arabic	50±20%	39±13%	54±15%	good
7	Arabic	74±12%	45±11%	36±11%	
8	Arabic	55±16%	42±12%	45±13%	

Story 3: political news, story 4: sports report about American soccer player

What was Hard?

- Mistranslated / untranslated name

  MT: Strong zhuo, pointing out that the two presidents ...

  Mono: Qiang Zhuo pointed out that the two presidents ...

  Bi: Johndroe said the two presidents ...

  No chance to recover...

What was Hard?

- Relationship between entities

  MT: The colombian team for the match, and it is very likely that the united states and kai in the americas cup final performance.

  Mono1: The Colombian team and the United States are very likely to end up in the Americas Cup as the final performance.

  Mono2: The next match against Colombia is likely to be the United States' and Keller's final performance in the current Copa America.

  Bi: The next game against the Colombian team will very probably be the last performance of the U.S. team and Keller in this year's Copa America.

What was Hard?

- Badly muddled machine translation

  MT: He is still being head coach bradley appointed to important, it's even a fist ", four young guards at the beginning of the ", the united states is...

  Mono: He is still being considered important by head coach Bradley who appointed him. It is a fight with "four young guards at the beginning of their careers", but the United States..

  Bi: He was still entrusted with the important task by head coach Bradley, but what can one man do against overwhelming odds; a US team of "Elite Young Guards" setting out to do battle...

Summary

- Very good results
  - assisted monolingual translators competitive with bilingual translators
  - most of the meaning comes across
- Some outstanding issues can be addressed: name transliteration
- Domain knowledge important

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
  - familiar content for translator?
  - very similar to previously translated text?

Try it at home!

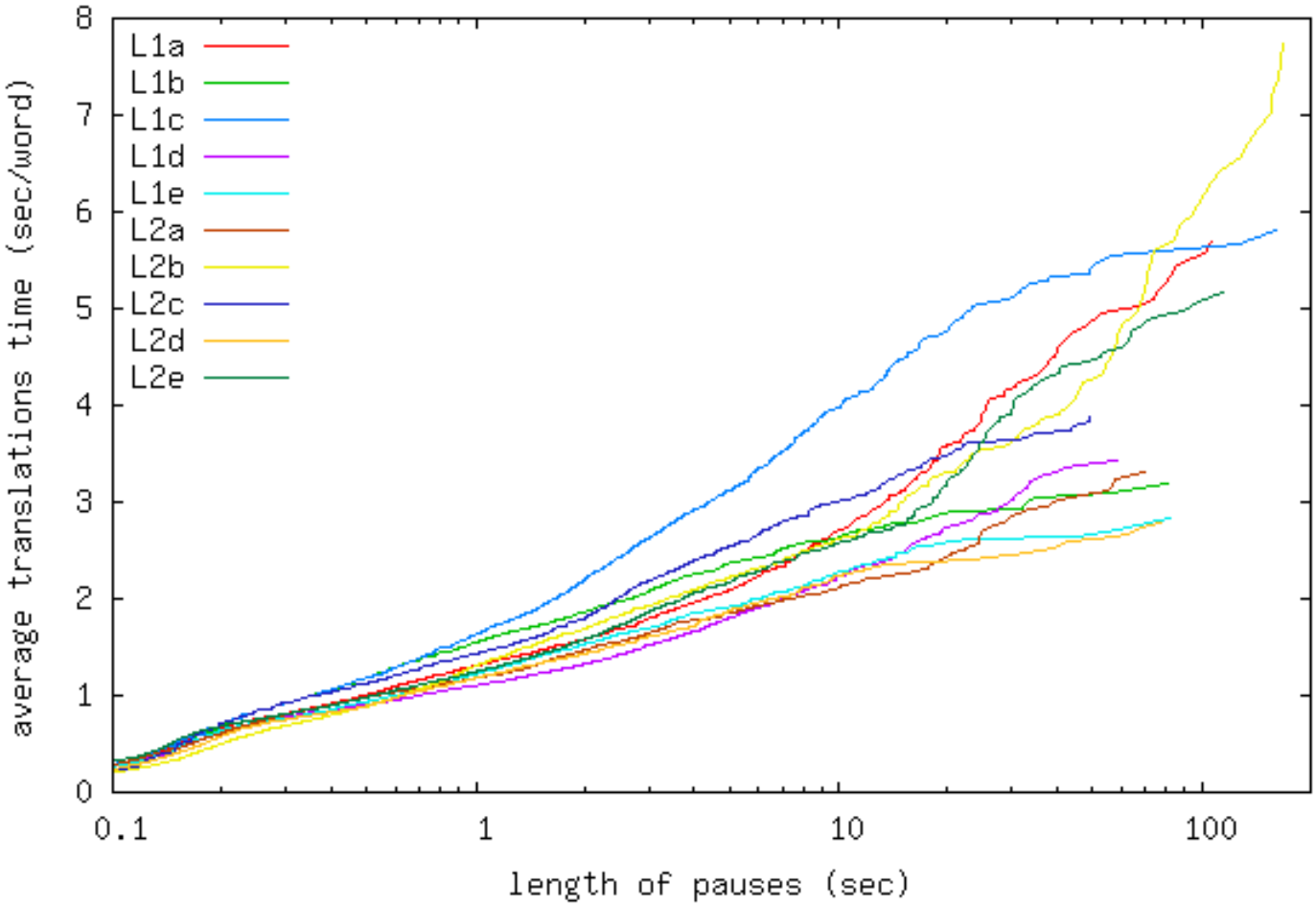
http://www.caitra.org/

questions?

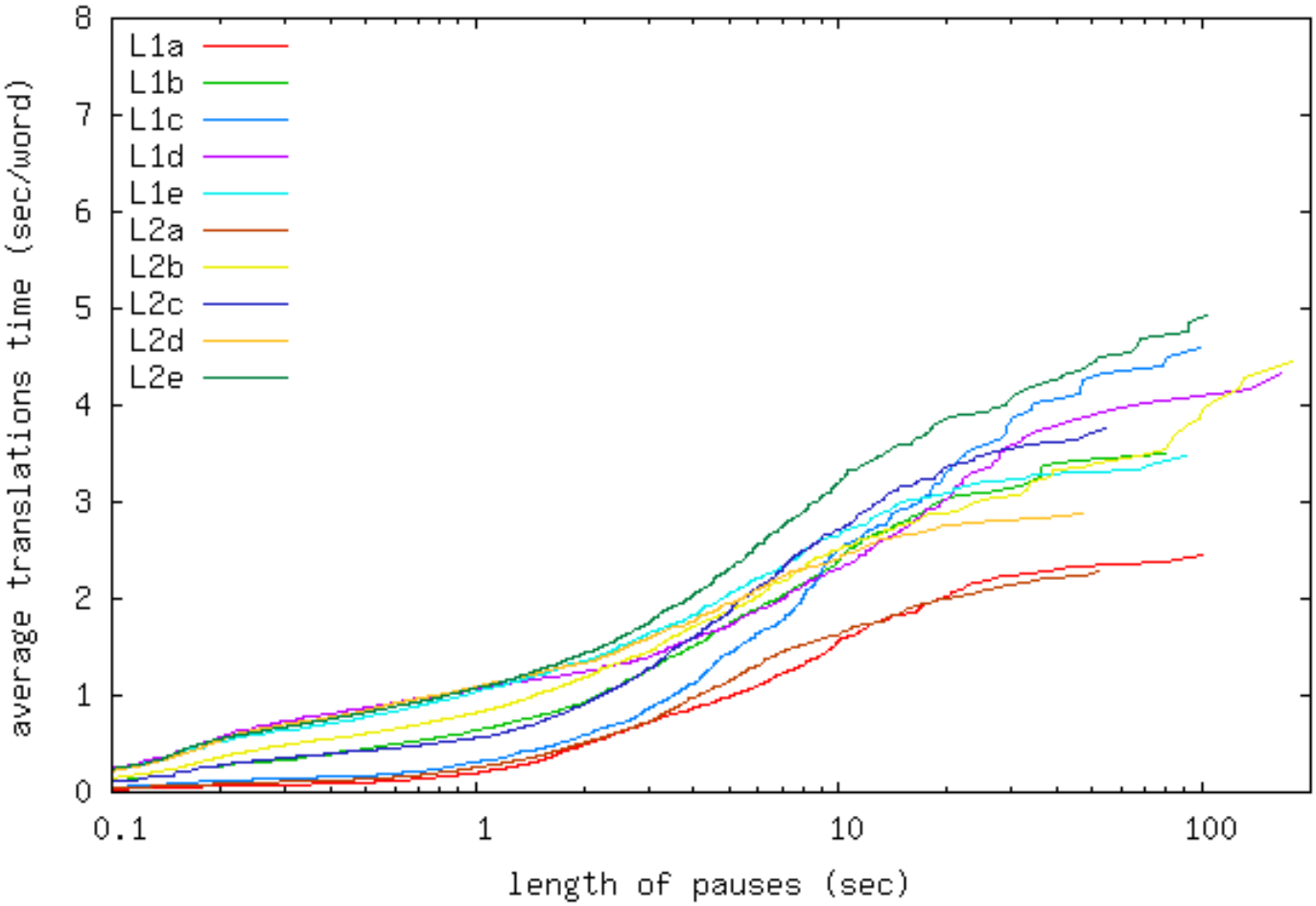
Interactive Post-Editing?

- word alignment to source
- confidence estimation of likely faulty parts
- integration with translation memory

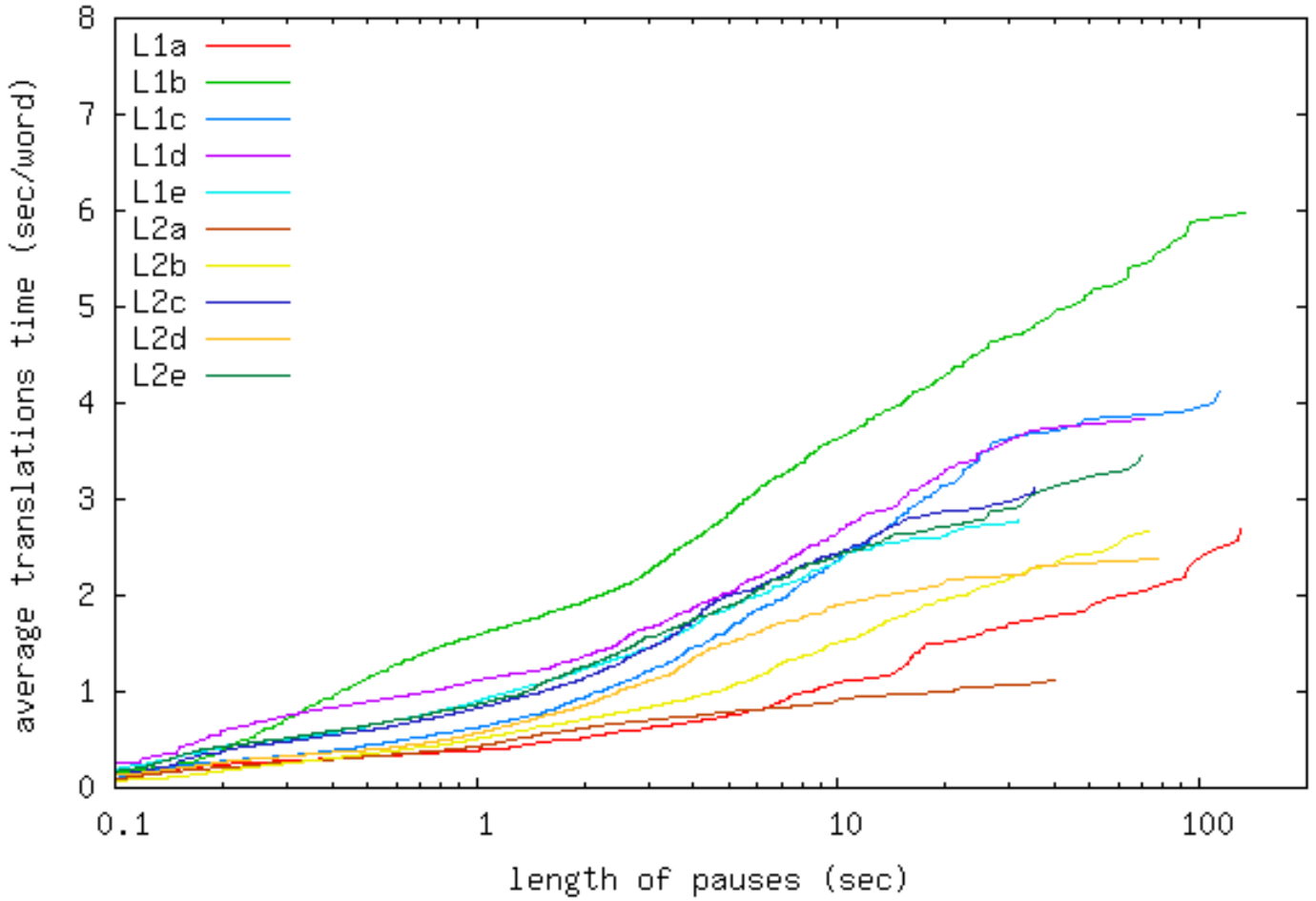
Pauses: Unassisted



Pauses: Options



Pauses: Prediction of sentence completion



Pauses: Postediting