
Computer Aided Translation

Philipp Koehn

7 September 2012



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



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



- **Human Translation**
- Assistance to Human Translators
- User Study
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT

Setup

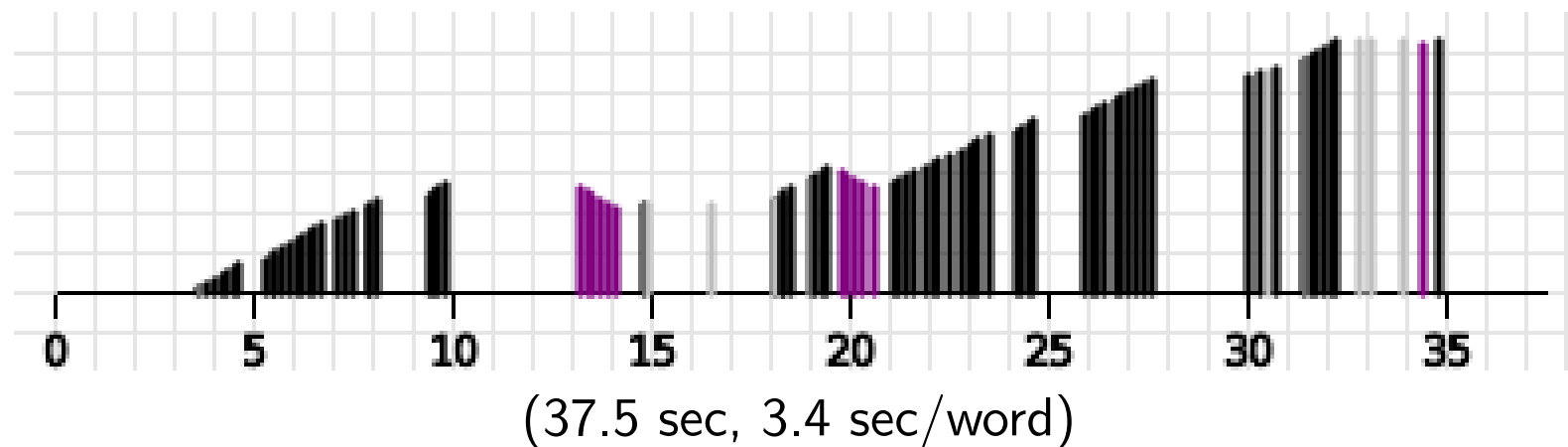


- 10 students at the University of Edinburgh
 - half native French speakers
 - half native English speakers with advanced French
- Each student translated
 - news stories
 - French-English
 - about 40 sentences
 - easy task: familiar content, no specialized terminology
- Keystroke log

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■
 - 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	
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

User	total	not much time		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

User	total	not much time			Pauses		similar
		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

User	total	not much time			Pauses	differences		similar
		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	

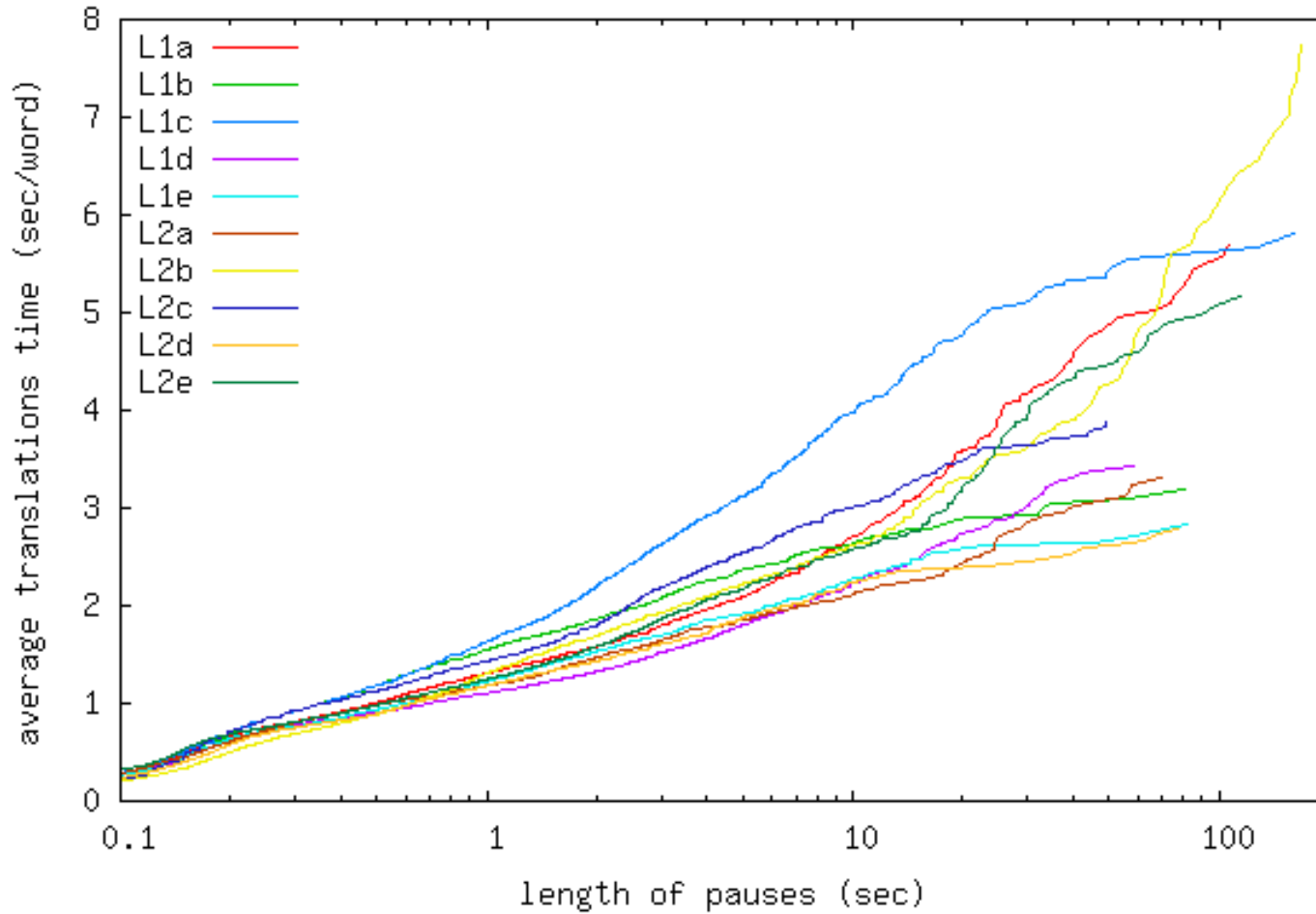
L1 = native French, L2 = native English
average time per input word

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



Overview



- Human Translation
- **Assistance to Human Translators**
- User Study
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT

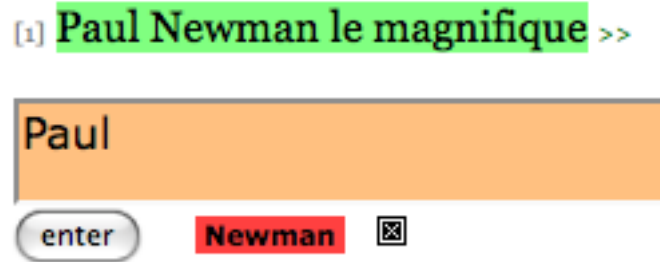
Our Types of Assistance

- 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 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?²⁰



- 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)

Overview



- Human Translation
- Assistance to Human Translators
- **User Study**
- Assistance to Monolingual Translators
- Integration of Translation Memory and MT

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.	
<hr/>		
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

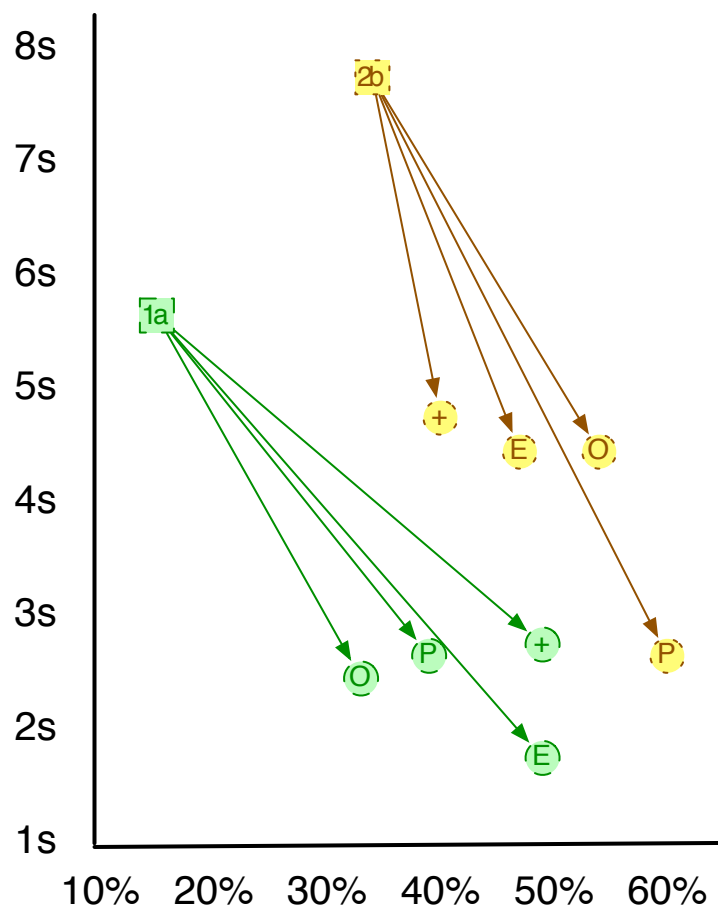


Assistance	Speed	Quality
Unassisted	4.4s/word	47% correct
Postedit	2.7s (-1.7s)	55% (+8%)
Options	3.7s (-0.7s)	51% (+4%)
Prediction	3.2s (-1.2s)	54% (+7%)
Prediction+Options	3.3s (-1.1s)	53% (+6%)

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%

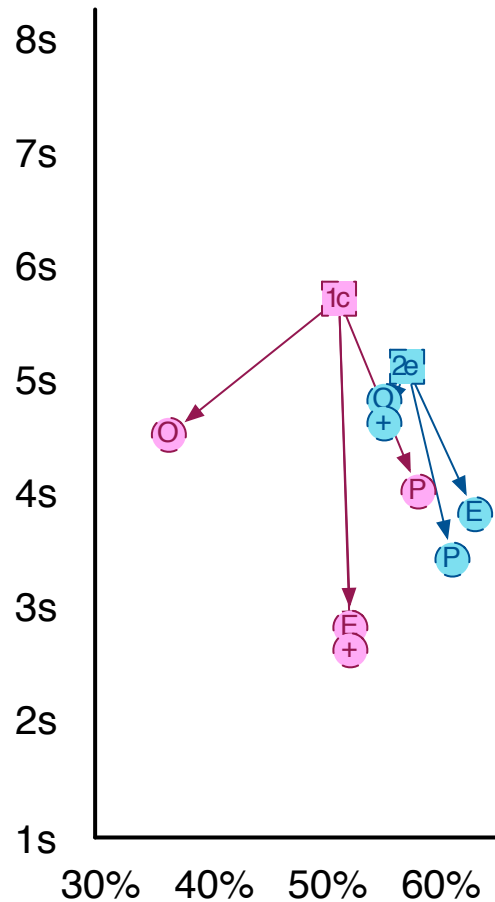
Slow Users 1: Faster and Better



- Unassisted
 - more than 5 seconds per input word
 - very bad (35%, 16%)

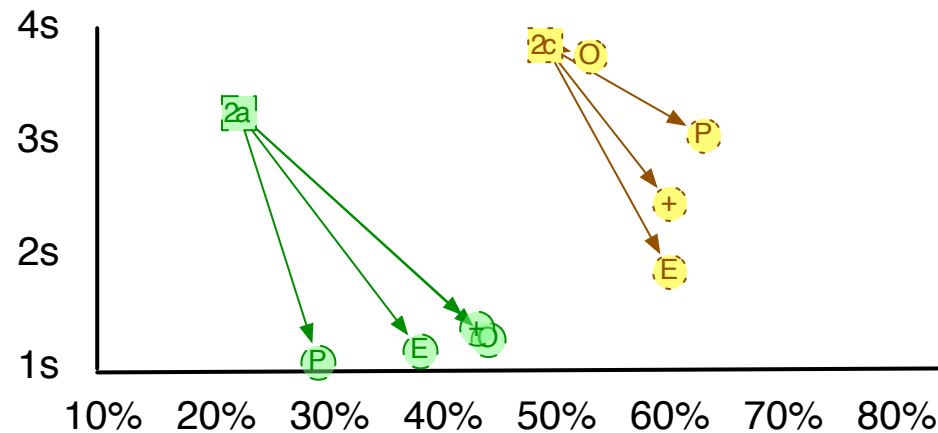
- With assistance
 - much faster and better
 - reaching roughly average performance

Slow Users 2: Only Faster



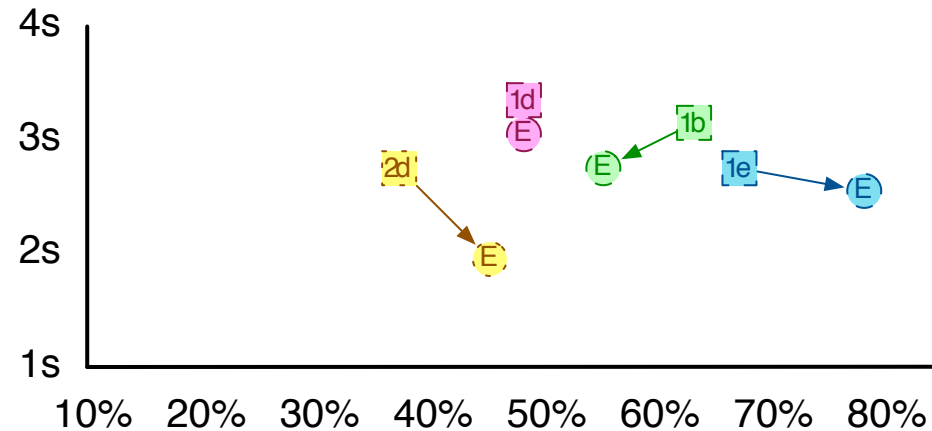
- Unassisted
 - more than 5 seconds per input word
 - average quality
- With assistance
 - faster and but not better

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

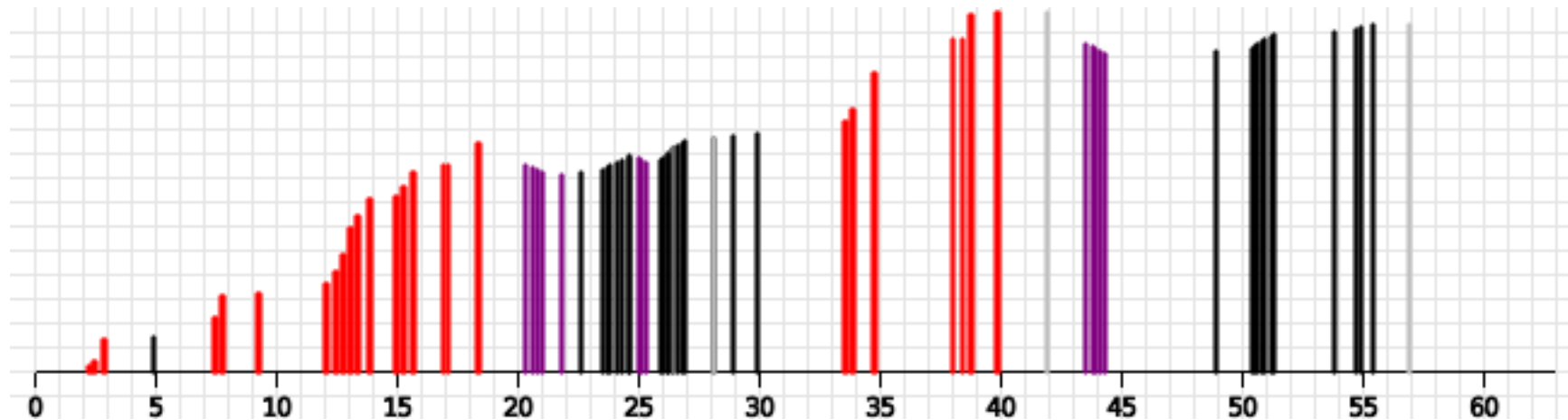


- Use the assistance sparingly or not at all, and see generally no gains
- The two best translators are in this group
- Postediting
 - mixed on quality (2 better, 1 worse, 1 same), but all faster
 - best translator (L2e, 68%) becomes much better (record 79%)

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

Slightly 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

Slightly 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

Slightly 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 L1b⁴³



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 L1b⁴⁴



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 L2e⁴⁵



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

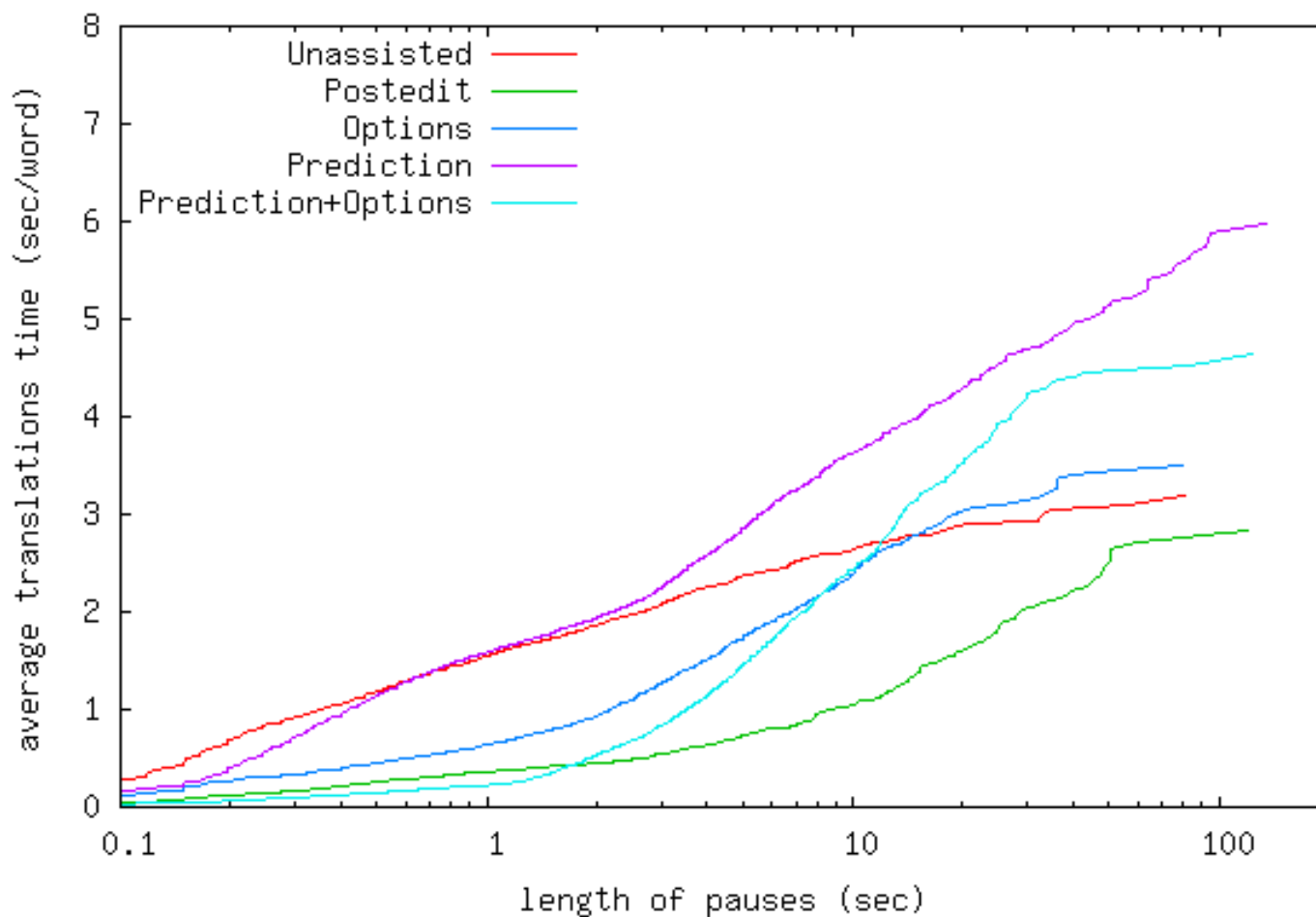
Origin of Characters: Native English L2e⁴⁶



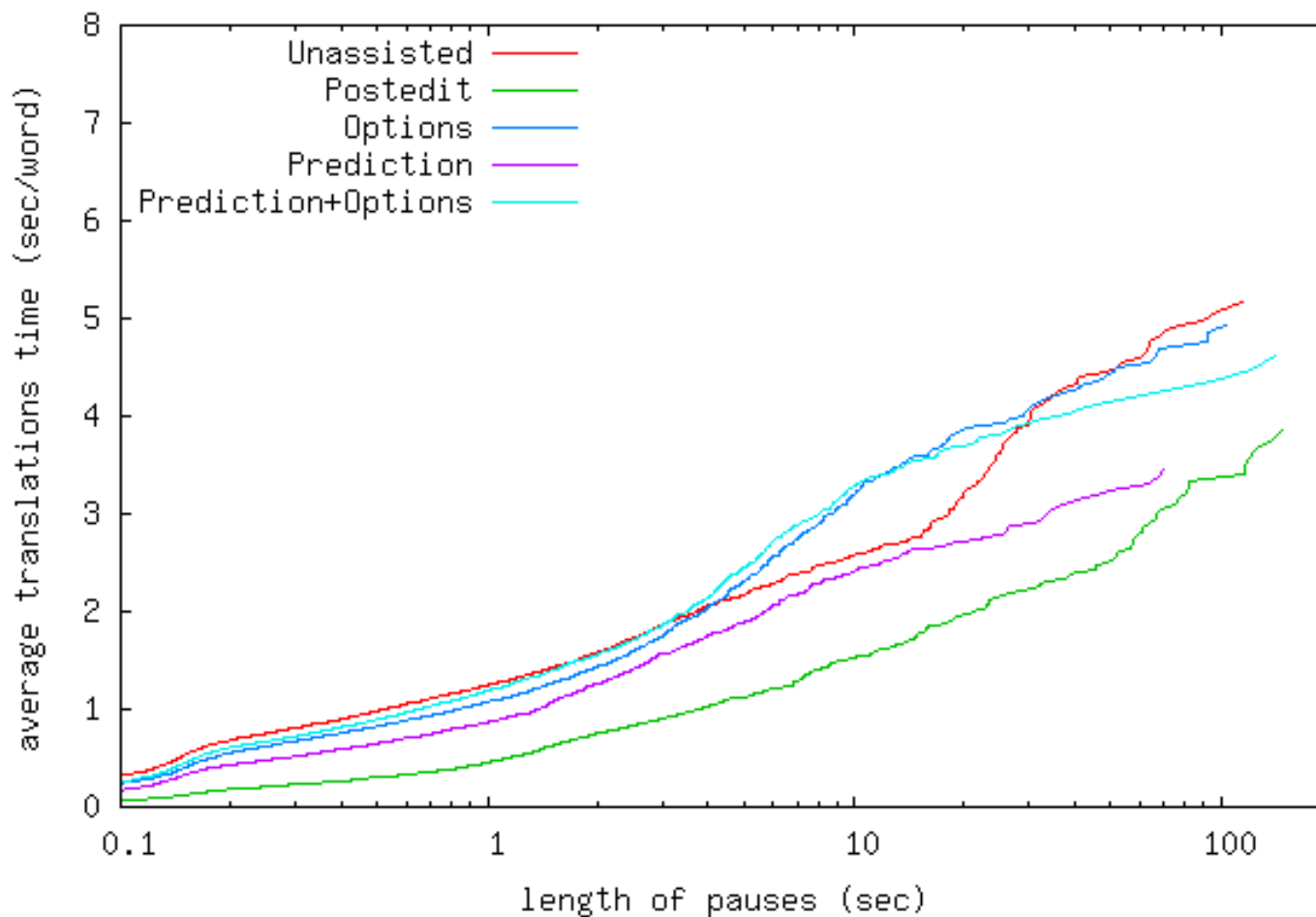
User: L2e	key	click	tab	mt
Postedit	20%	-	-	79%
Options	77%	22%	-	-
Prediction	61%	-	38%	-
Prediction+Options	100%	-	-	-

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

Pauses: French-Native User L1b

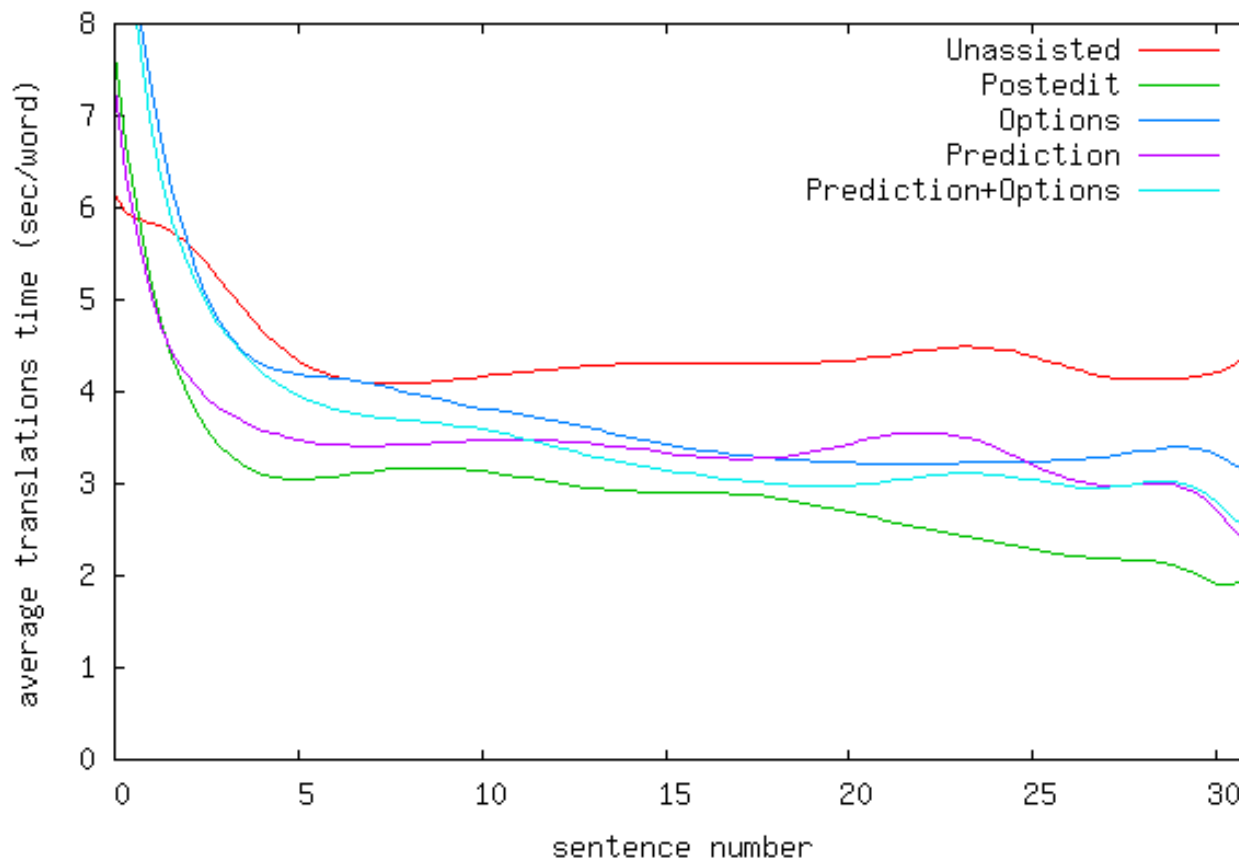


Pauses: English-Native User L2e



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, 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■
- 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?■
- What are the hard problems?
 - unknown words
 - words without direct translation
 - syntactic re-arrangement■
- What do translators change in post-editing?

⇒ We will investigate this
in a new EU project




Related Work: Tools used by Translators⁵⁴



- 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

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/abo...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)

Overview



- Human Translation
- Assistance to Human Translators
- User Study
- **Assistance to Monolingual Translators**
- Integration of Translation Memory and MT

Enabling Monolingual Translators



- Monolingual translator
 - wants to understand a foreign document
 - has no knowledge of foreign language
 - uses a machine translation system■
- Questions
 - Is current MT output sufficient for understanding?
 - What else could be provided by a MT system?

Good Enough



- An MT system might produce this:

The girl entered **into** room.■

- We know what is meant:

The girl entered **the** room.■

- We understood.

Process



- MT system translates foreign story■
- Person A edits it
 - goal: fluent translation that represents the meaning (as it was understood)
 - without access to reference translation■
- Person B checks if edited sentences are correct
 - with access to reference translation

Example

- MT system translates foreign sentence

The girl goes the room.■

- Person A edits it

The girl goes into the room.■

- Reference

The girl enters the room.■

- Person B checks edited sentence: **CORRECT**

Real Example

- MT system output:

The study also found that one of the genes **in the improvement in people with** prostate cancer risk, it also reduces the risk of suffering from diabetes.

- What does this mean?■

- Monolingual translator:

The research also found that one of the genes **increased people's risk of** prostate cancer, but at the same time lowered people's risk of diabetes.■

- Document context helps

Experiment



- Language pairs
 - Arabic–English
 - Chinese–English■
- Machine translation systems
 - Edinburgh’s 2009 GALE systems
 - Moses system with all available parallel data■
- Stories taken from NIST 2008 test sets

Stories



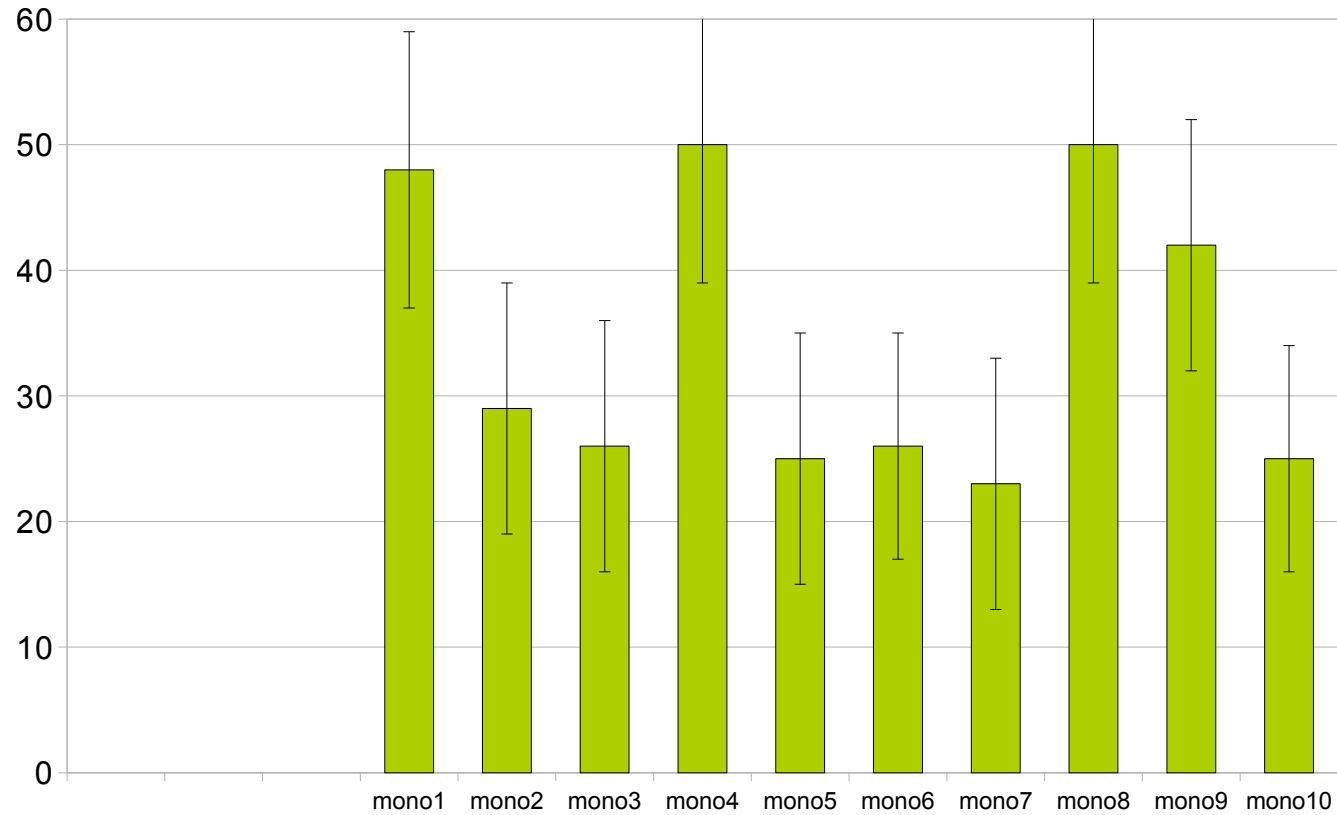
Story	Headline	Sent.	Words
1: chi	White House Pushes for Nuclear Inspectors to Be Sent as Soon as Possible to Monitor North Korea's Closure of Its Nuclear Reactors	6	207
2: chi	Torrential Rains Hit Western India, 43 People Dead	10	204
3: chi	Research Shows a Link between Arrhythmia and Two Forms of Genetic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

Experiment



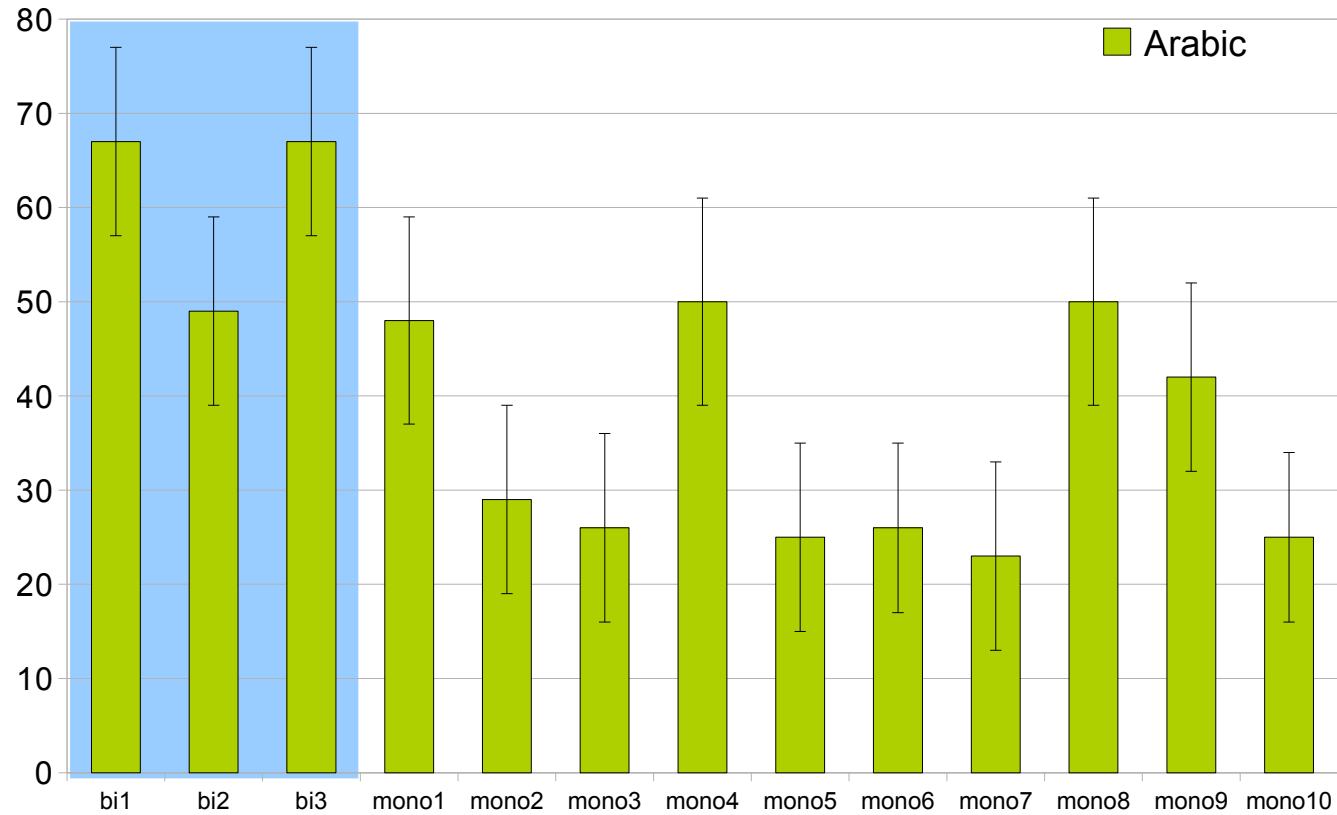
- Monolingual translators
 - 10 students/staff at the University of Edinburgh
 - none knew Arabic or Chinese
 - have access to full stories at a time, may correct prior sentences
- Bilingual translators
 - 3 of the 4 reference translations in NIST test set
- Remaining reference translation as truth

Results: Arabic



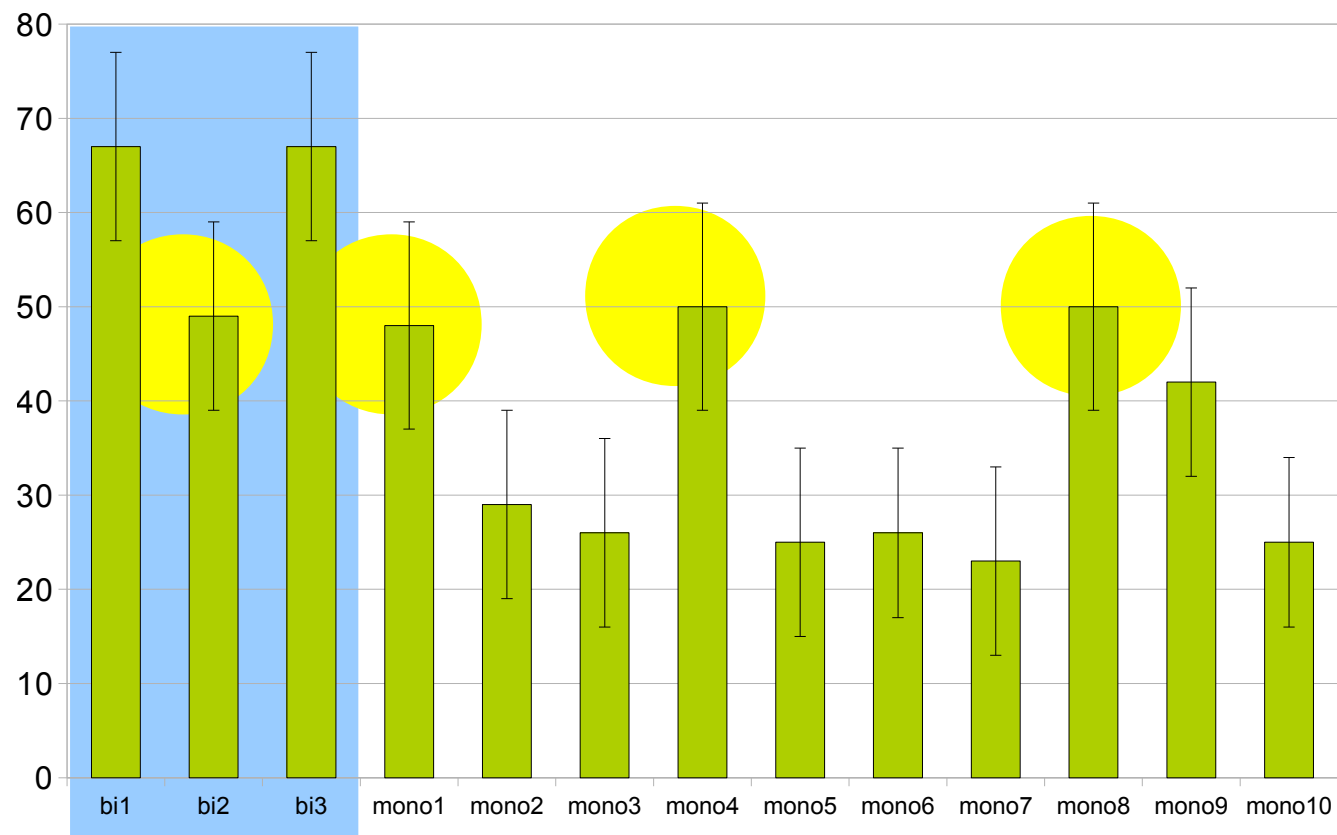
percentage of sentences judged as correct

Results: Arabic



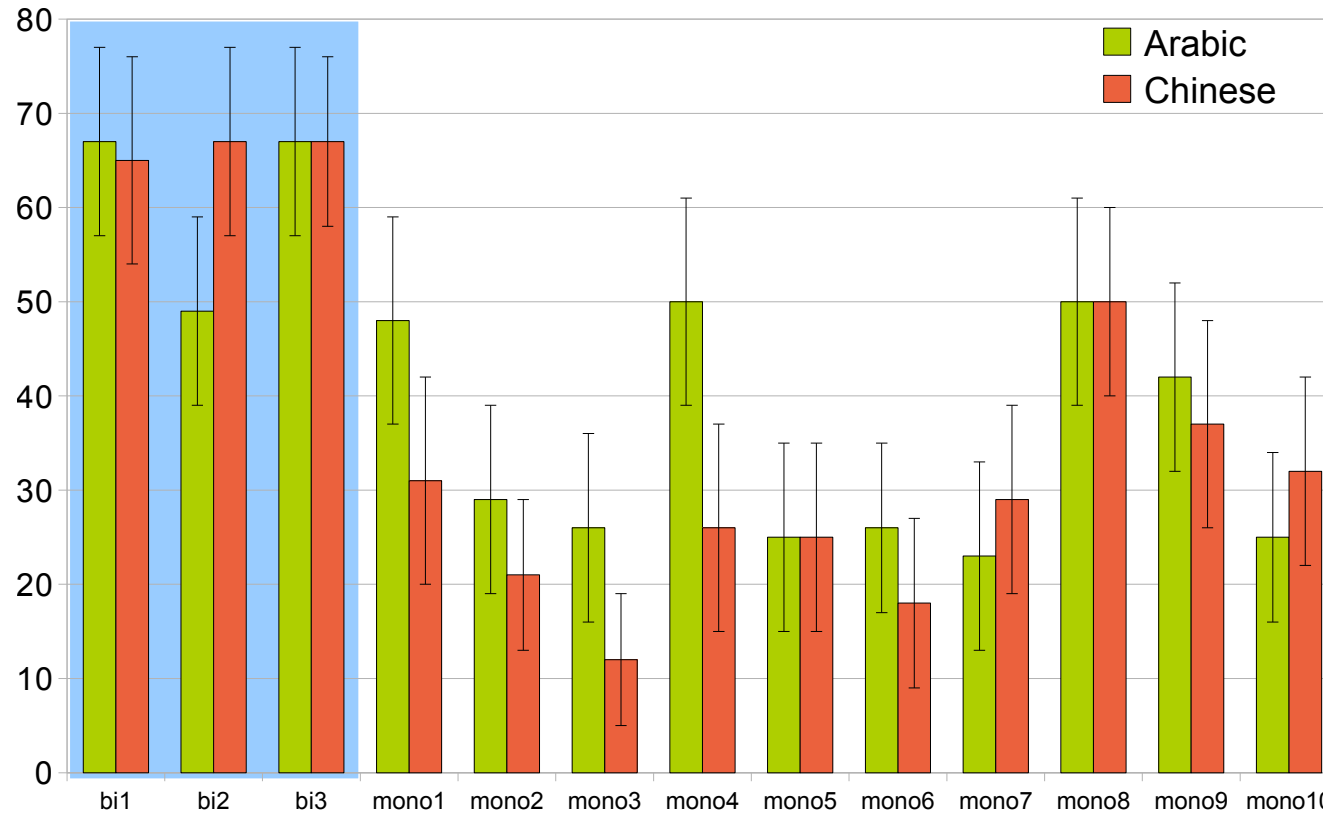
compared to bilingual translators

Results: Arabic



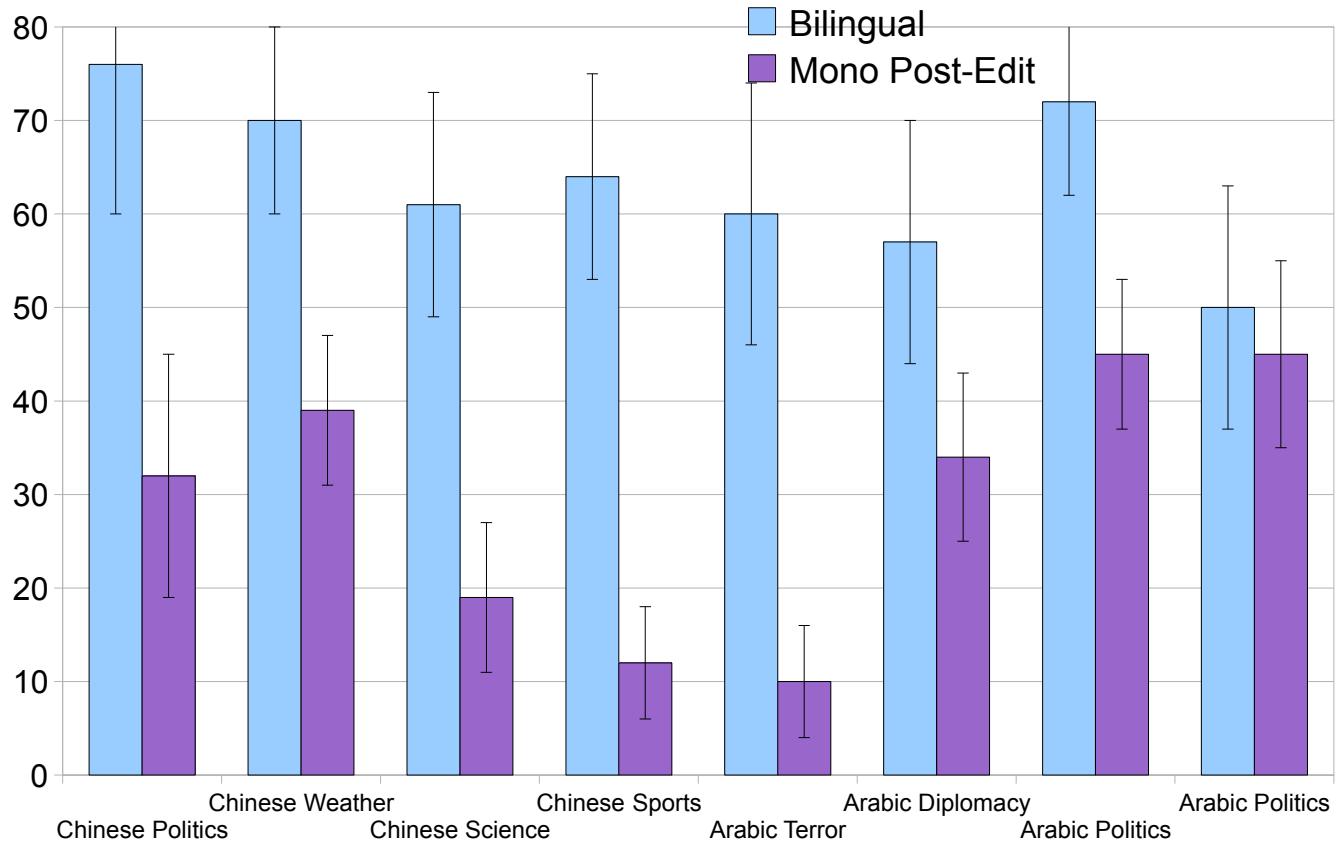
best monolinguals as good as worst bilingual

Results: Arabic and Chinese



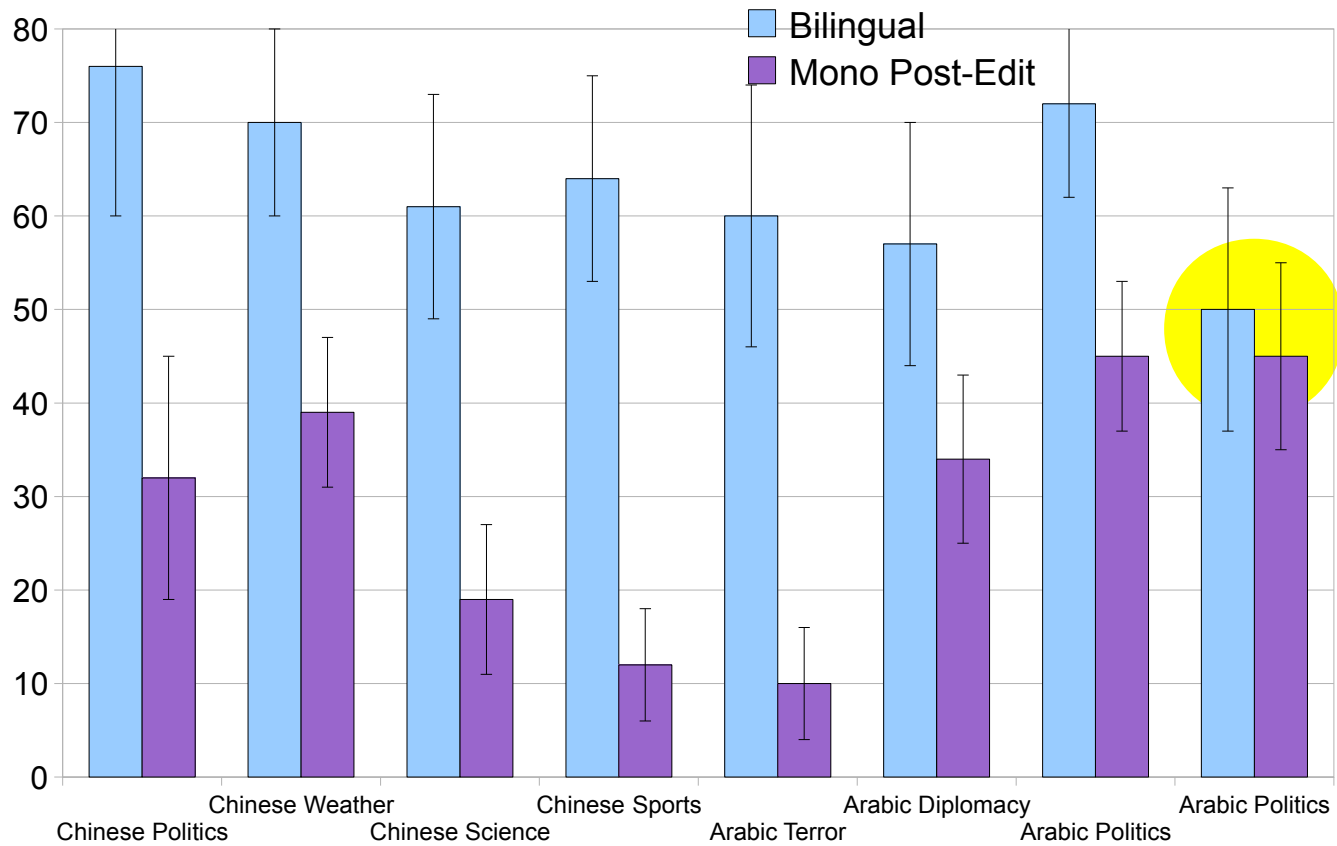
mostly worse performance for Chinese

Results per Story



performance differs widely per story

Results per Story



one story: monolinguals as good as bilinguals

Offering more assistance

- Progress in computer aided translation■
- Interactive machine translation (TransType)
 - show prediction of sentence completion
 - recompute when user types own translation■
- Alternative translations [Koehn and Haddow, 2009]
 - display translation options from translation model
 - ranked by translation score

Translation Options

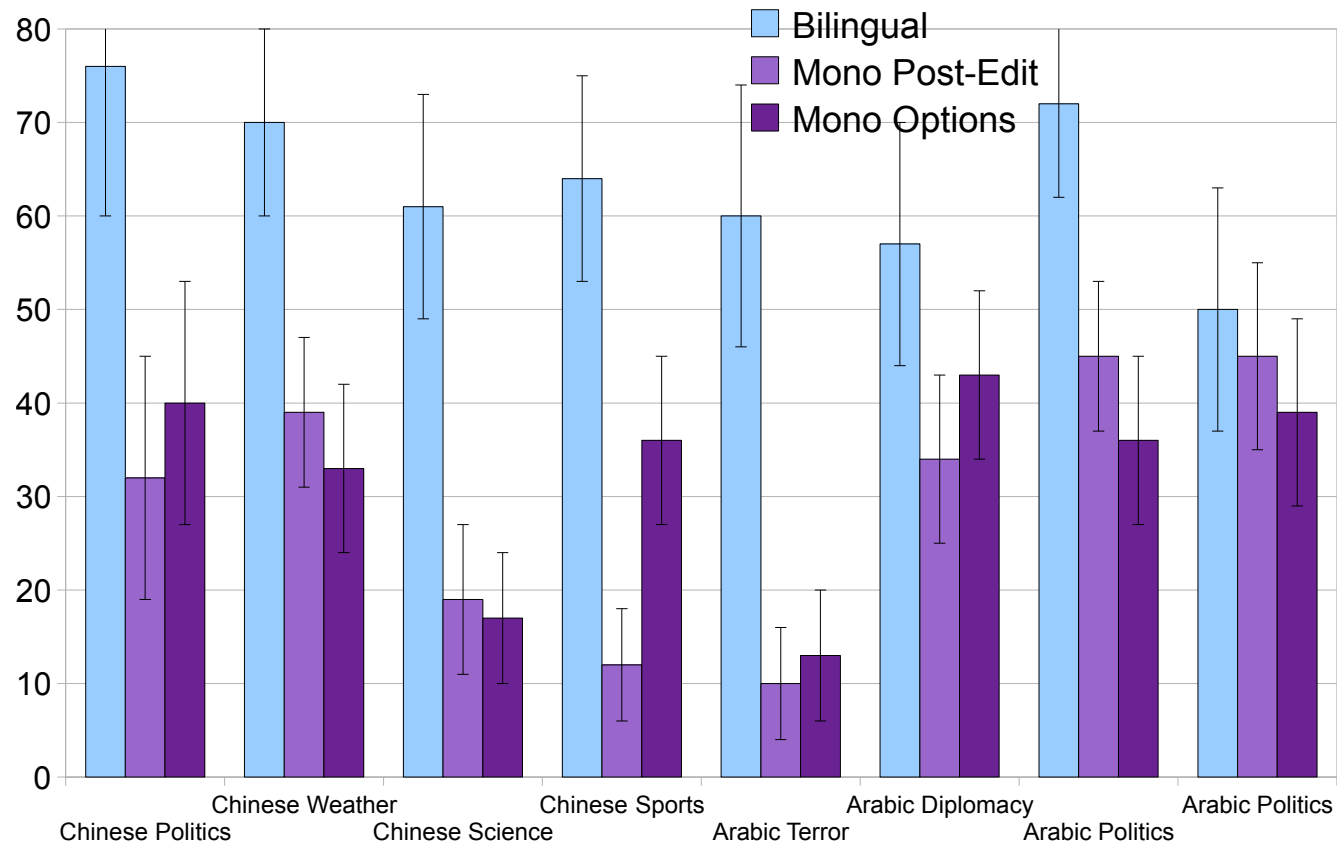
وكان	مجلس	التواب	الاميركى	اعتمد	الخميس	قانونا	يطالب	يسحب	القوات	المقاتلة	الاميركية	من	العراق	في	موعد	اقصاه	الاول	من نيسان	ابريل @/ @
the	the us house of representatives	adopted	thursday	legally	calls for the withdrawal of	combat troops	us	iraq	in	no later than	the first	from	april						
the us house of representatives	the	thursday ,	law		the fighting forces	the us	from iraq	the latest	the first of	april									
the us house	adopted the	thu	the legally		fighting forces	us	from iraq in	i	april										
it was	us house of representatives	was adopted	thursday , the	the law	demands withdrawal of troops	fighter	the us	no later than	first	on april									
he was	the us house	adopted by	thursday 's	a law	calls for withdrawal of	combat forces	of	in the	not later than	first of									
he	us house	adopted by the	on thursday	a legally	calls for the withdrawal	forces	the fighter	from											
earlier ,	us	adopted a	on thursday ,	by law	demands the withdrawal of	troops	iraq												
was		, was adopted	thursday the	legally ,	demands withdrawal of		of the												
it was the		adopted ,	thu ,	the legal	calls for withdrawal		from iraq in the												
earlier , the		adopted , the	thursday , a	legally @-@	demands the withdrawal		the american	by the first of											
2008	متحديا	مرة	جديدة	الرئيس	جورج	بوش	الذي	يعارض	اي	تحديد	موعد								
2008 ,	defying	once	new	president george w. bush	which opposes the	no date has been set for the													
the 2008	defiant	once again		president george bush	who opposes	no date has been set for													
2008	challenging	again	the new		, which opposes	no date has been set													
	a defiant	the first			, who opposes the	a date .													
	in defiance of	once again ,			, who opposes	date .													
	, challenging	once again the		president george bush , who	opposed to setting any	the date of the													
	, in defiance	for the first time	a new	president george w. bush 's	which opposes	no date													
in 2008 ,	defying the	again		us president george w. bush	opposed to	any	the date of												
	challenging the	time			who opposes the	date of													
	, defying	once again , the			opposes	date													

up to 10 translations for each word / phrase

Translation Options

يسحب	القوات	المقاتلة	الاميركية	من	العراق
withdrawal of	combat troops		us		iraq
	the fighting forces		the us	from	iraq
	fighting forces		us	from	iraq
withdrawal of troops		fighter	the us		
withdrawal of	combat forces			of	
the withdrawal	forces	the fighter		from	
the withdrawal of	troops			iraq	
withdrawal of				of the	
withdrawal				from iraq in	
the withdrawal			the american		

Results with Options



no big difference — once significantly better

Error Analysis

(a) Critical Judges

- Reference

Torrential Rains Hit Western India, 43 People Dead

- Bilingual translator

Heavy Rains Plague Western India Leaving 43 Dead

Error Analysis

(b) Mistakes by the professional translators

- Reference

Over just two days on the 29th and 30th, rainfall in Mumbai reached **243 mm**.

- Bilingual translator

The rainfall in Mumbai had reached **243 cm** over the two days of the 29th and 30th alone.

Error Analysis

(b2) Domain knowledge vs. language skills

- Bilingual translator

With **Munchen-Gladbach** falling to the German Bundesliga 2, ...

- Monolingual translator

The **Mönchengladbach** team fell into the second German league, ...

Error Analysis

(c) Bad English by monolingual translators

- Monolingual translator

The western region of **india heavy rain** killed 43 people.

Error Analysis

(d) Mistranslated / untranslated name

- Reference

Johndroe said that the two leaders ...

- Machine translation

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

- Monolingual translator

Qiang Zhuo pointed out that the two presidents ...

Error Analysis

(e) Wrong relationship between entities

- Machine translation

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

- Monolingual translator 6

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

- Monolingual translator 8

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

Error Analysis

(f) Badly muddled machine translation

- Reference

In the current America's cup, he has, just as before, been given an important job to do by head coach Bradley, but he clearly cannot win the match singlehanded. The US team, made up of "young guards," ...

- Machine translation

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...

Conclusions

- Main findings
 - monolingual translators may be as good as bilinguals■
 - widely different performance by translator / story■
 - named entity translation critically important■
- Various human factors important
 - domain knowledge■
 - language skills■
 - effort

Overview



- Human Translation
- Assistance to Human Translators
- User Study
- Assistance to Monolingual Translators
- **Integration of Translation Memory and MT**

Progress in Translation Automation



- **Translation Memory (TM)**

- translators store past translation in database
- when translating new text, consult database for similar segments
- fuzzy match score defines similarity

widely used by translation agencies■

- **Statistical Machine Translation (SMT)**

- collect large quantities of translated text
- extract automatically probabilistic translation rules
- when translating new text, find most probable translation given rules

wide use of free web-based services
not yet used by many translation agencies

TM

vs.

SMT



used by
human translator

used by
target language information seeker

restricted domain
(e.g. product manual)

open domain translation
(e.g. news)

very repetitive content

huge diversity (esp. web)

corpus size:
1 million words

corpus size:
100-1000 million words

commercial developers
(e.g., SDL Trados)

academic/commercial research
(e.g., Google)

Our Goal



Better TM
using SMT methods

Main Idea

- Input

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory

The second paragraph of Article 5 is deleted .

⇒ **Part of the translation from TM fuzzy match**

Part of the translation with SMT

The second paragraph of Article **21** is deleted .

Related Work

- Work inspired by EBMT

[Smith and Clark, 2009]

[Zhechev and van Genabith, 2010]

- uses syntactic information in alignment
- lower performance than reported here

- Encode fuzzy match as rule with gaps

[Biçici and Dymetman, 2008]

- similar to our second method
- impressive improvements, but weak baseline SMT

Two Solutions



- XML frames
- Very large hierarchical rules

Example



- Input sentence:

The second paragraph of Article 21 is deleted .

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

- Detect mismatch (string edit distance)

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

- Detect mismatch (string edit distance)
- Align mismatch (using word alignment from GIZA++)

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

Output word(s) taken from the target TM

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

Output word(s) taken from the target TM

Input word(s) that still need to be translated by SMT

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

- XML frame (input to Moses)

<xml translation=" À l' article " /> 21

<xml translation=" , le texte du deuxième alinéa est supprimé . " />

Example

- Input sentence:

The second paragraph of Article 21 is deleted .

- Fuzzy match in translation memory:

The second paragraph of Article 5 is deleted .

=

À l' article 5 , le texte du deuxième alinéa est supprimé .

- More compact formalism for the purposes of this presentation:

< À l' article > 21 < , le texte du deuxième alinéa est supprimé . >

Steps

- Fuzzy matching

- based on string edit distance on words

$$FMS = 1 - \frac{\text{edit-distance}(\text{source}, \text{tm-source})}{\max(|\text{source}|, |\text{tm-source}|)}$$

- string edit distance on letters as tie breaker
- details see [Koehn and Senellart, AMTA 2010]

straight-forward

- Word alignment of TM source and target

standard method

- Construction of XML frame

- = linking mismatch(input, TM source) to TM target

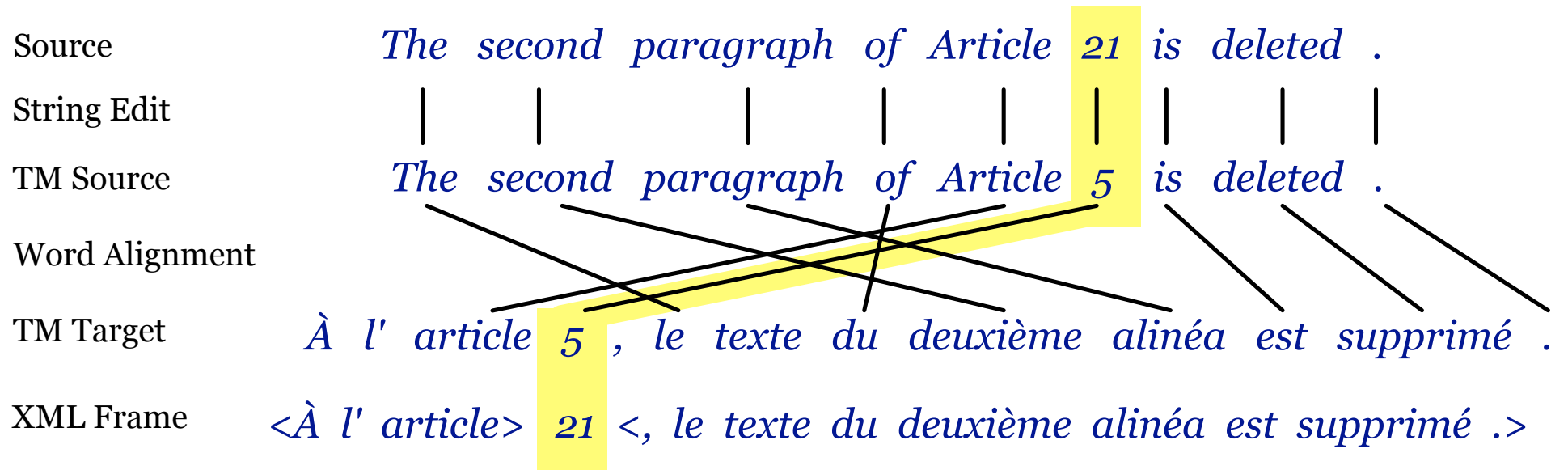
can be tricky

Construction of XML Frame

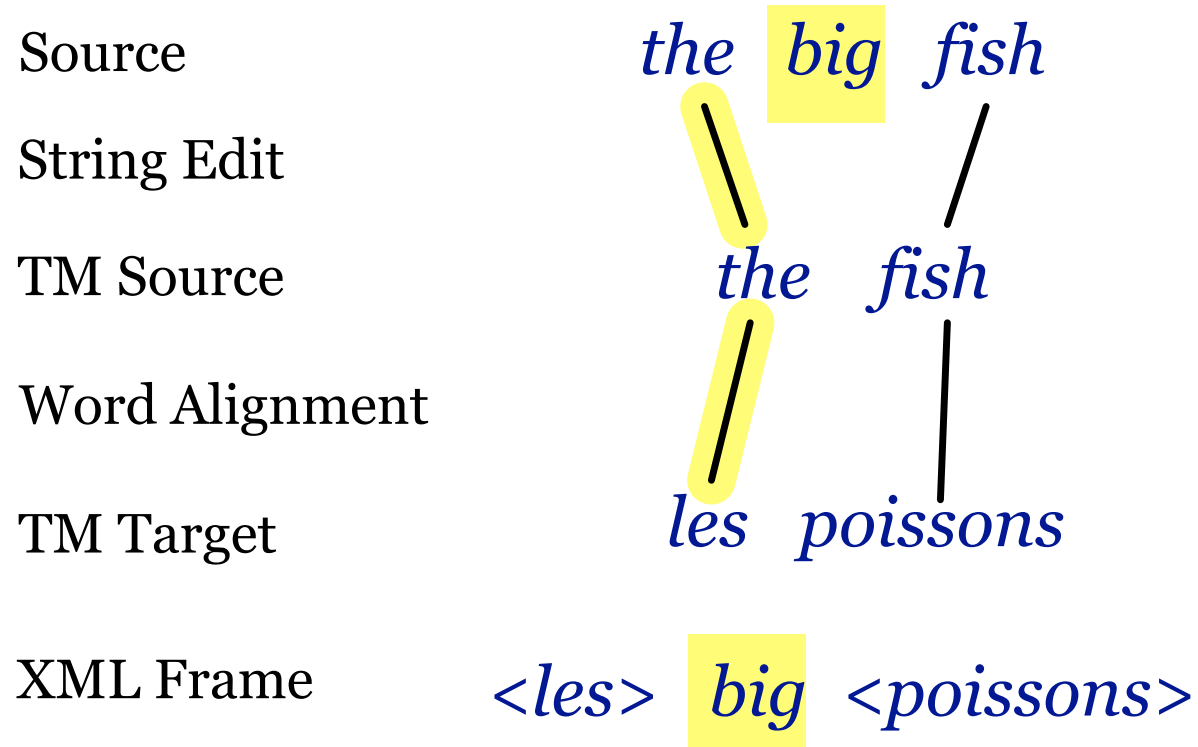


- Basic principles
 - start with fully specified XML frame
 - all mismatched source words have to be translated by SMT
 - all TM target words aligned to mismatched TM source words are removed
 - if the alignment to the TM target words fails
 - go to the previous TM source word and follow its alignment
- See paper for algorithm

Example

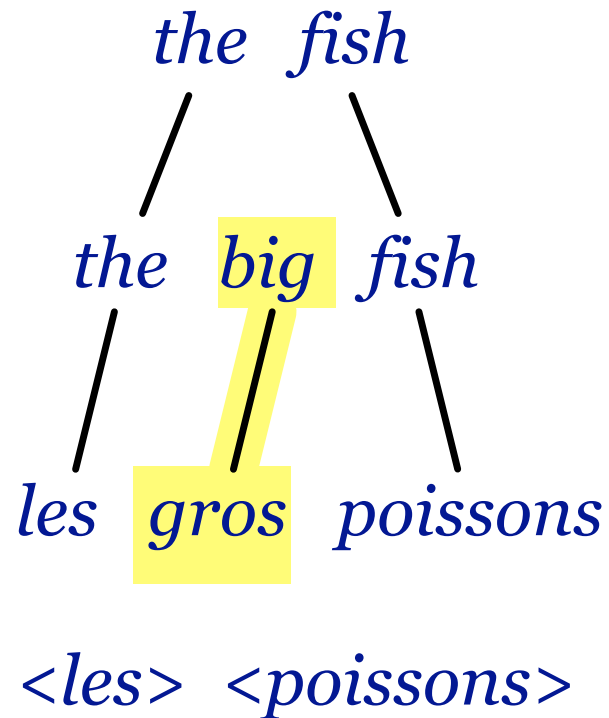


Special Case: Insertion



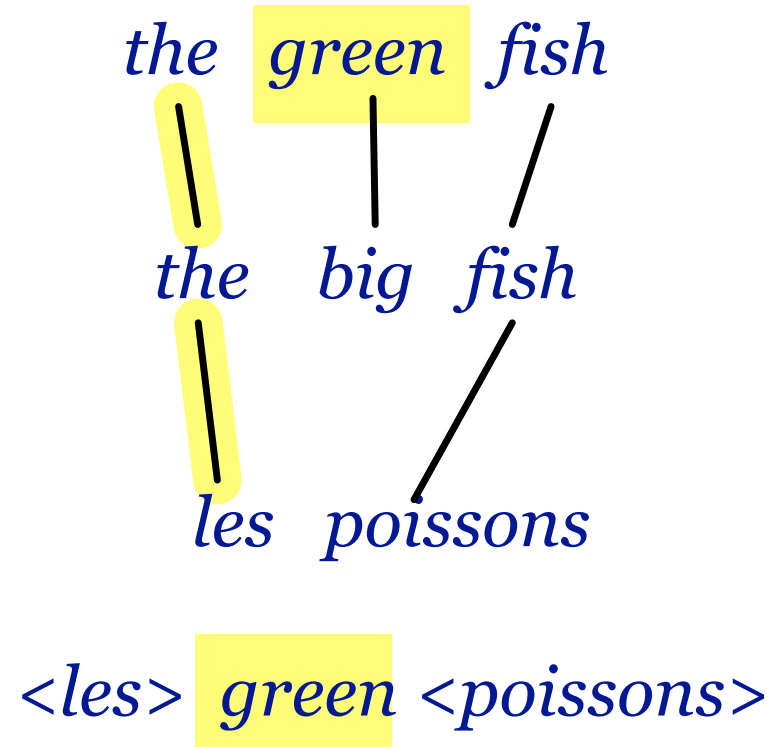
Special Case: Deletion

102

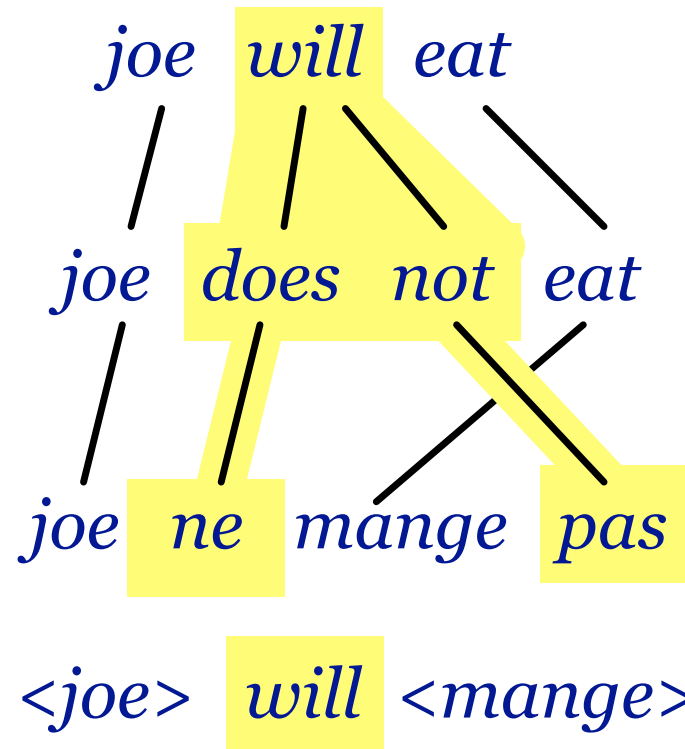


Special Case: Unaligned Mismatch

103



Special Case: Disconnected Alignments¹⁰⁴



Experiments



- Baseline 1: Unmodified TM matches
- Baseline 2: SMT system trained on TM data
- Our XML frame method

Corpora: Size

Acquis

	Corpus	Test
segments	1,165,867	4,107
English words	24,069,452	129,261
French words	25,533,259	135,224

Product

	Corpus	Test
segments	83,461	2,000
English words	1,038,762	24,643
French words	1,110,284	26,248

Corpora: Fuzzy Matches

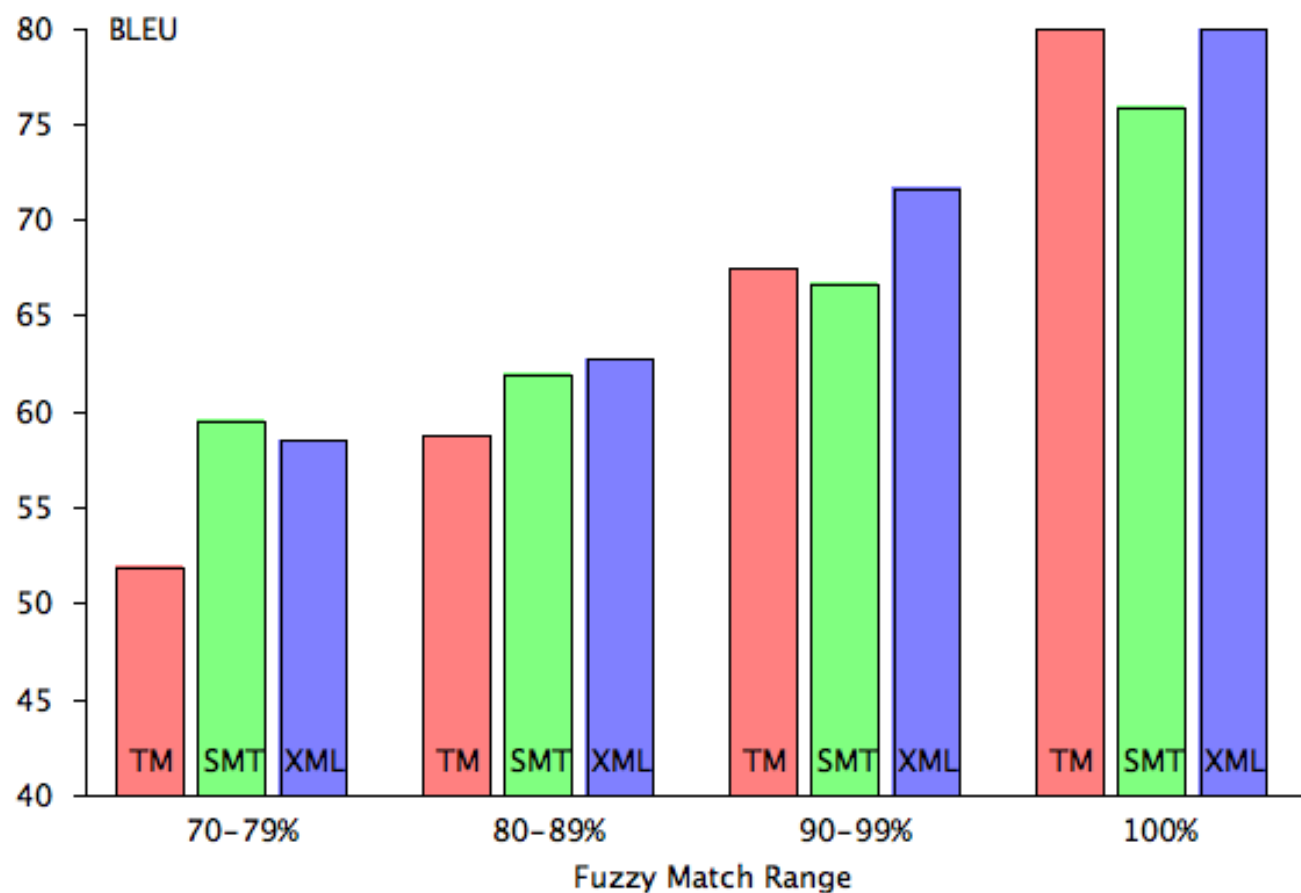
Acquis

	Sentences	Words	W/S
100%	1395	14,559	10.4
90-99%	433	12,775	29.5
80-89%	154	5,347	34.7
70-79%	250	6,767	27.1

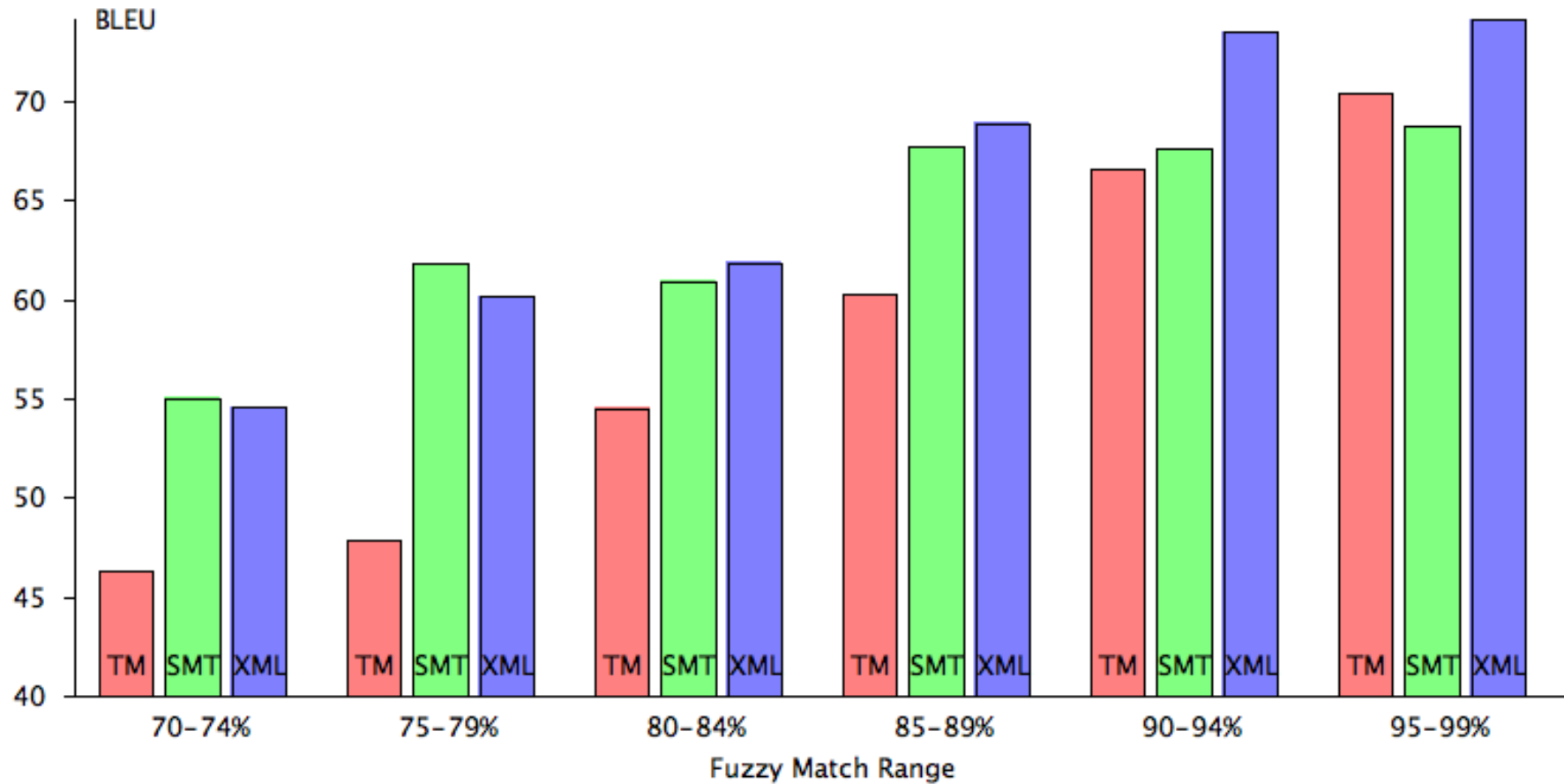
Product

	Sentences	Words	W/S
95-99%	230	3,006	13.1
90-94%	225	2,968	13.2
85-89%	177	2,000	11.3
80-84%	185	1,950	10.5
75-79%	152	1,350	8.9
70-74%	98	987	10.1

Results: Acquis



Results: Product



Recap



- TM provides fuzzy matches
- SMT translates mismatched words
- TM match encoded in XML frame
... but is that not just a very large translation rule?

Background: Hierarchical Phrase Rules ¹¹¹



- Given: sentence pair with monotone 1-to-1 alignment

the big fish = les gros poissons

- Phrase translation rules
 - (the ; les)
 - (the big ; les gros)
 - (the big fish ; les gros poissons)
 - (big ; gros)
 - (big fish ; gros poissons)
 - (fish ; poissons)

- Hierarchical phrase-based rule are constructed by subtraction
 - large rule: (the big fish ; les gros poissons)
 - small rule: (big ; gros) (contained in large rule)
 - hierarchical rule: (the x fish ; les x poissons)

XML Frame as Very Large Rule



- XML frame

<À l' article> 21 <, le texte du deuxième alinéa est supprimé .>

for input

The second paragraph of Article 21 is deleted .

- Very large rule

(The second paragraph of Article x is deleted .
; À l' article x , le texte du deuxième alinéa est supprimé .)

Very Large Rules in SMT

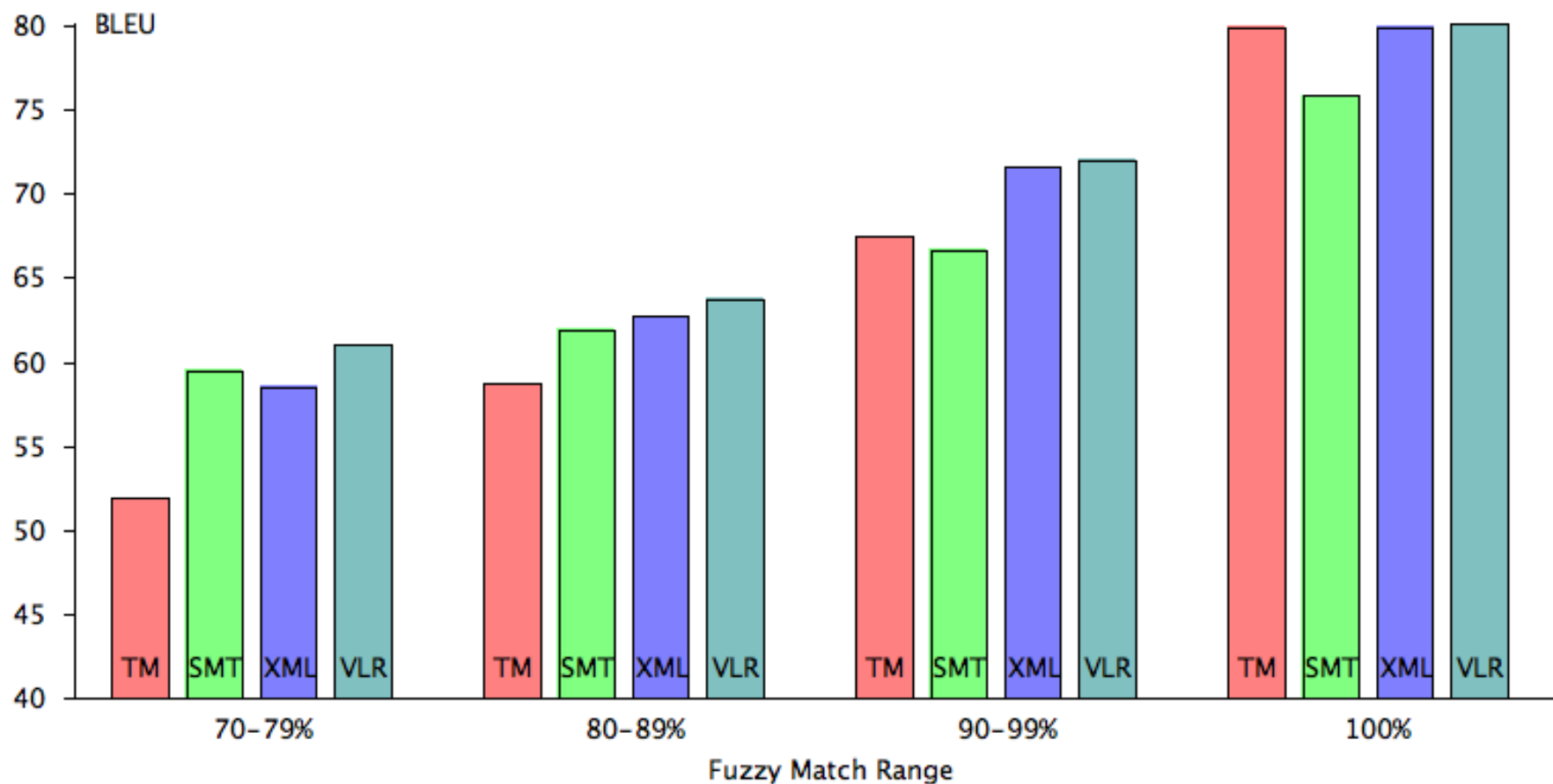


- Rule size limited in SMT
 - maximum number of words, e.g. 5
 - maximum number of non-terminals (x), e.g. 2
- ... but only due to storage limitations for large rule rule tables
- Rules may be generated on the fly [Lopez, 2007]

Advantage over XML Method

- Choices
 1. multiple fuzzy matches in TM with same score
 2. same TM source with multiple translations
 3. multiple SMT translations■
- Decisions in XML frame method
 1. randomly chosen
 2. most frequent
 3. highest model score (tried others, see paper)■
- Decisions for very large rules
 1. all
 2. all
 3. integrated scoring of VLR rules and basic translation rules (tunable)

Result: Acquis



Future Work: User Studies

- Significant increases in BLEU
- To do: validation in user studies
- Additional benefit: possible to highlight mismatch in translation

– input

The second paragraph of Article 21 is deleted .

– suggested translation

À l' article 21 , le texte du deuxième alinéa est supprimé .

Thank You



questions?