
Recent Trends in Computer Aided Translation

Philipp Koehn

16 September 2016



Overview



- A practical introduction: the CASMACAT workbench
- Postediting
- Types of assistance
- Logging, eye tracking and user studies
- :

- Cognitive studies of translators leading to insights into interface design
 - better understanding of translator needs
- Workbench with novel types of assistance to human translators
 - interactive translation prediction
 - interactive editing and reviewing
 - adaptive translation models
 - better tools for translators
- Demonstration of effectiveness in field tests with professional translators
 - increased translator productivity

Postediting Interface



6 Le Pakistan a donc été récompensé par l'assistance et les armes des États-Unis. > As a result, Pakistan was rewarded with American financial assistance and arms.

7 Pour mieux redistribuer ses cartes, Moucharraaf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan. > In furtherance of his re-alignment, Musharraaf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

8 Les opérations contre les forces des Talibans et d'Al-Qaeda ont obtenu des résultats mitigés. >

visualization >>

ITP T→ DRAFT **TRANSLATED**

- Source on left, translation on right
- Context above and below

Confidence Measures



And on that the signs are mixed. > Y en que los indicios son desiguales.

Translation matches

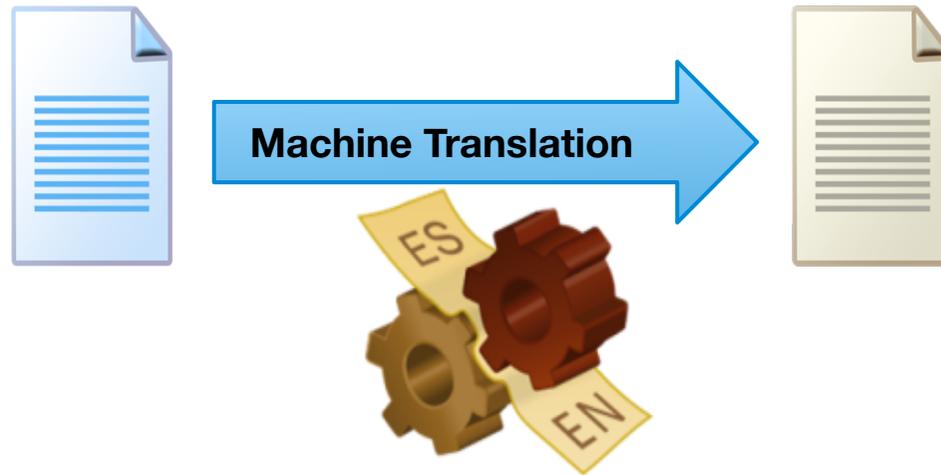
And on that the signs are mixed. Y en que los indicios son desiguales.

Source: ITP Fri Apr 12 2013 18:03:17 GMT+0200 (CEST) 42

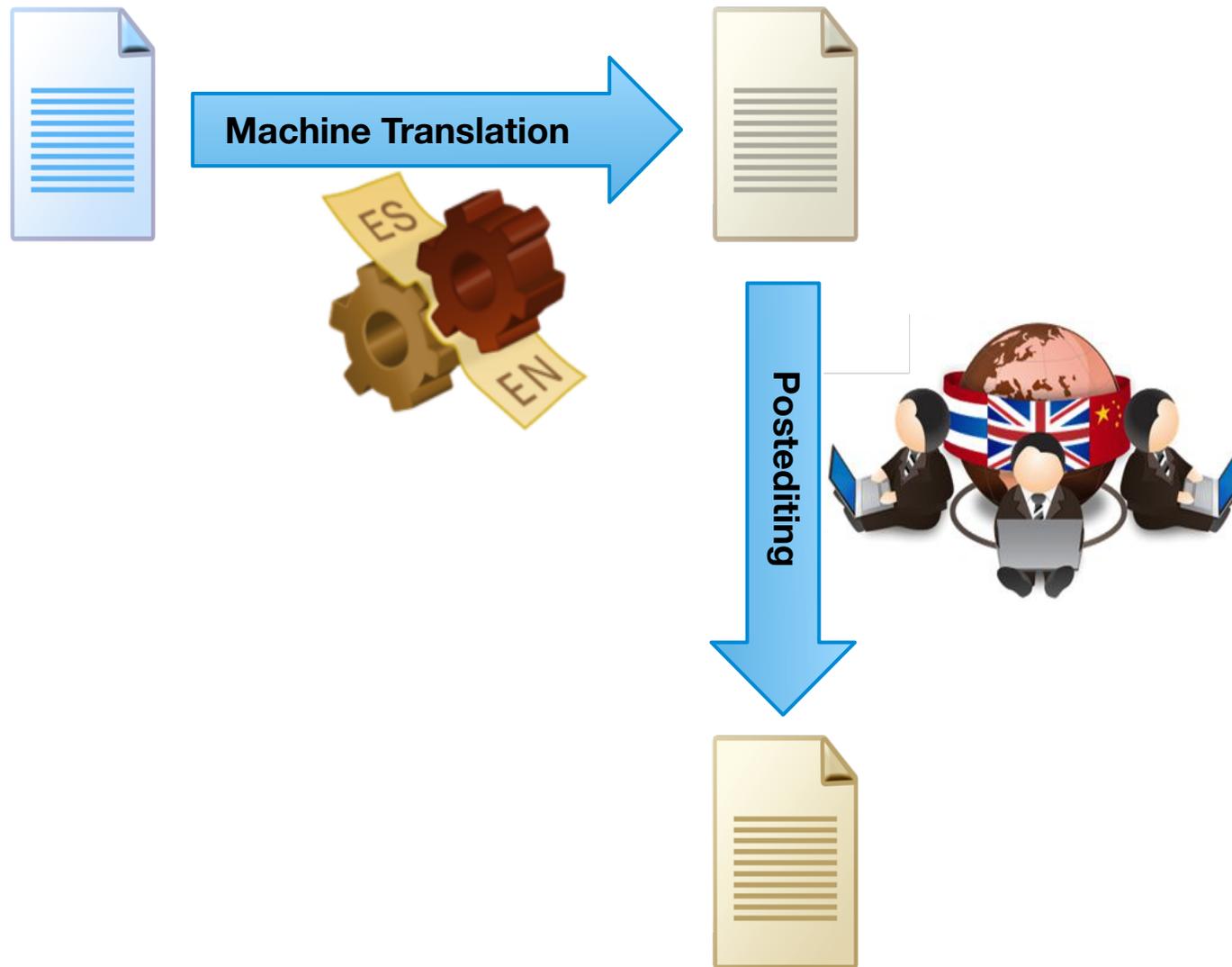
The screenshot shows a machine translation interface. At the top, the source sentence "And on that the signs are mixed." is on the left, and the target sentence "Y en que los indicios son desiguales." is on the right. The target sentence has some words highlighted in orange and red. Below the target sentence are several buttons: a pencil icon, "ITP", "T→", "DRAFT", and "TRANSLATED". A "Translation matches" box is visible on the left. At the bottom, the source and target sentences are repeated, and a status bar shows "Source: ITP Fri Apr 12 2013 18:03:17 GMT+0200 (CEST) 42".

- Sentence-level confidence measures
→ estimate usefulness of machine translation output
- Word-level confidence measures
→ point posteditor to words that need to be changed

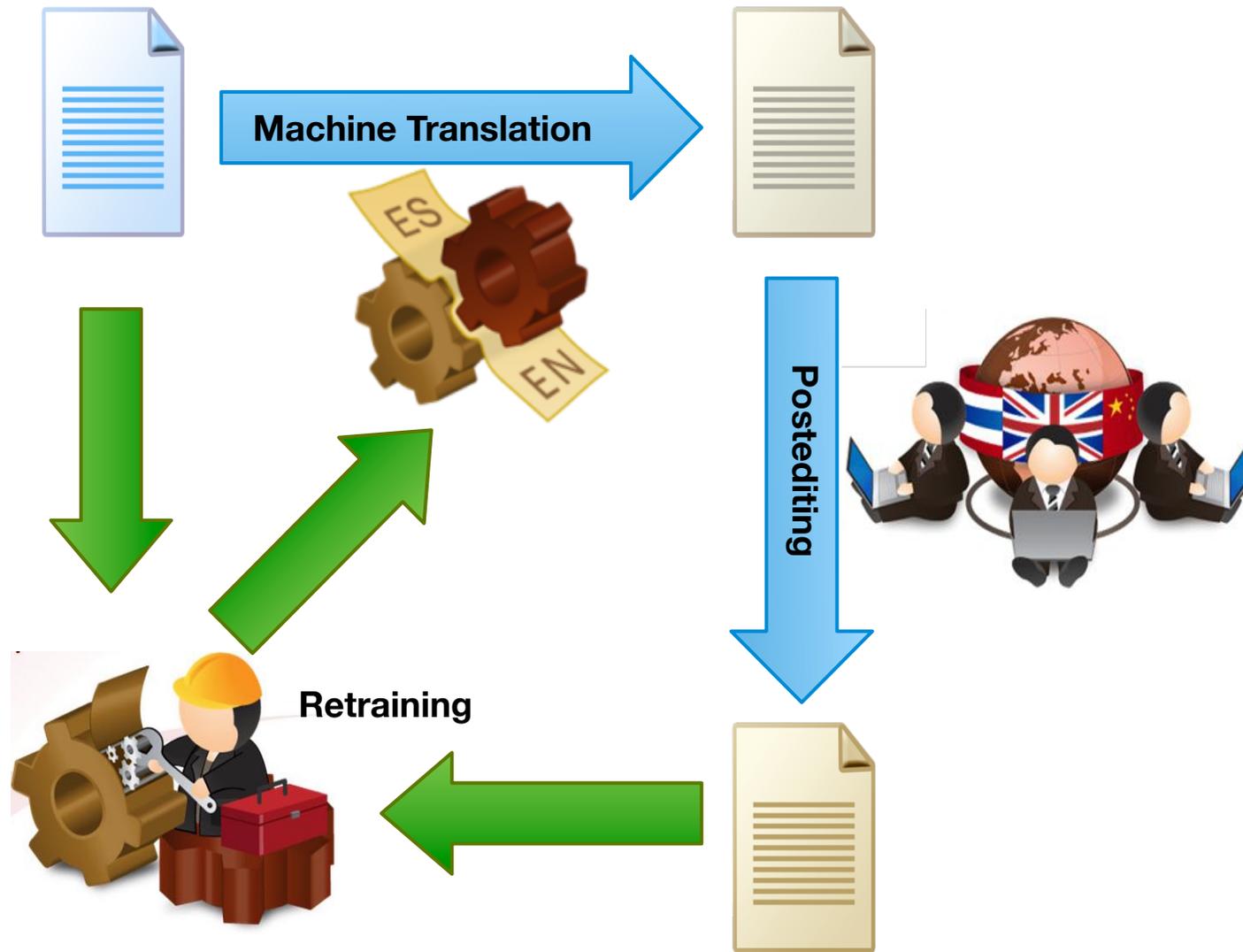
Incremental Updating



Incremental Updating



Incremental Updating



Interactive Translation Prediction



CASMACAT Re-calibrate Download edf-file **DOWNLOAD PROJECT** **HELP**

Document list > Jobs List > cyberpresse-2012-12-01-...2012-12-01-1566244.xliff (10) > fr > en

Shortcuts

visualization >>

Pour la science, cela sert à vérifier la validité du Modèle standard (MS), et cela permet aussi aux physiciens de scruter tout écart entre les observations et les prédictions du MS.

For science, this serves to

ITP ☰ SRC→ DRAFT **TRANSLATED**

Ils sont d'ailleurs plusieurs à souhaiter ardemment qu'on en trouve, car la moindre différence pourrait ouvrir une porte sur une "nouvelle physique" et boucher certains trous du Modèle.

Word Alignment



9

visualization >> displayMouseAlign displayCaretAlign displayShadeOffTranslatedSource displayConfidences highlightValidated highlightPrefix highlightLastValidated limitSuffixLength

Pour mieux redistribuer ses cartes, Moucharraaf a envoyé l'armée pakistanaise dans les zones ethniques qui longent l'Afghanistan, pour la première fois depuis l'indépendance du Pakistan.

In furtherance of his re-alignment, Musharraaf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

ITP T→ DRAFT **TRANSLATED**

Word Alignment

10



The screenshot shows a translation interface with two text boxes. The left box contains the French text: "Ils sont d'ailleurs plusieurs à souhaiter ardemment qu'on en trouve, car la moindre différence pourrait ouvrir une porte sur une "nouvelle physique" et boucher certains trous du Modèle." The word "ardemment" is highlighted in yellow. The right box contains the English text: "There are elsewhere several who wish". A tooltip above the English text shows the alignment: "fervently that,, because". Below the text boxes are several buttons: "ITP", a menu icon, "SRC→", "DRAFT", and a blue "TRANSLATED" button. A "visualization >>" link is in the top right corner of the interface.

- With interactive translation prediction
- Shade off translated words, highlight next word to translate

Translation Option Array

... climbers are severely injured, and ten people are missing.
 after Mount Ontake (御嶽山, Ontake-san), a popular climbing spot in central Japan, **erupted** for the first time in five years.

Kletterer sind schwer verletzt, und zehn Menschen werden vermisst, nachdem Mount Ontake (御嶽山, Ontake-san), ein beliebter Kletterplatz im zentralen Japan,

ITP ≡ T→ DRAFT **TRANSLATED**

Translation Options

ke	-	san)	,	a	popular	climbing	spot	in central	Japan	,	erupted	for the first time in five years	.
ke	-	san)	,	ein	beliebtes	Klettern	vor Ort	in Mittel-	Japan,		ausbrach	zum ersten Mal in fünf Jahren	.
	und	San)	,	ein	populär	Bergsteigen	vor	zentrale	Japan	,	ausbrach,	zum ersten Mal in	fünf Jahre.
	/)	,	die	beliebt	Aufstieg	Fleck	zentralen	Japans,		platzte	zum ersten Mal	fünf Jahre
	der)		eine	beliebte	abhalten,	ein, in	zentraler	Japan		Ausbruch		in fünf Jahren
	bis)	,	in	populär	Erklimmen	Vor - Ort @-@	zentral	Japans	.	ausgebrochen	zum ersten Mal in der	von fünf Jahren.
	von)	,	.	populär ist,	beim Besteigen	in	mittel-	in Japan	-	ausgebrochen ist	zum ersten Mal seit	fünf Jahren sind.

- Visual aid: non-intrusive provision of cues to the translator
- Clickable: click on target phrase → added to edit area
- Automatic orientation
 - most relevant is next word to be translated
 - automatic centering on next word

Bilingual Concordancer



TIP [Menu] T→ DRAFT TRANSLATED

abandonner

abandon

ances des Etats-Unis à	abandonner	Musharraf -- et les col		merican reluctance to	abandon	Musharraf -- together
uridique, il a décidé d'	abandonner	la constitutionnalité, c		af has now decided to	abandon	constitutionality, remc
implement menacé d'	abandonner	ses accords commerci		simply threatened to	abandon	or never to conclude t

give up

erait donc contraint d'	abandonner	le droit de créer son p		would be required to	give up	the right to develop it
n' était pas disposé à	abandonner	ses fonctions militaire		arraf was not ready to	give up	his military post, but a

to

t ne veulent donc pas	abandonner	leurs prérogatives dar		olicy and do not want	to	delegate this prerogat
-----------------------	-------------------	------------------------	--	-----------------------	-----------	------------------------

to abandon

es tout en refusant d'	abandonner	son arsenal nucléaire		drawal while refusing	to abandon	its nuclear weapons a
------------------------	-------------------	-----------------------	--	-----------------------	-------------------	-----------------------

Paraphrasing

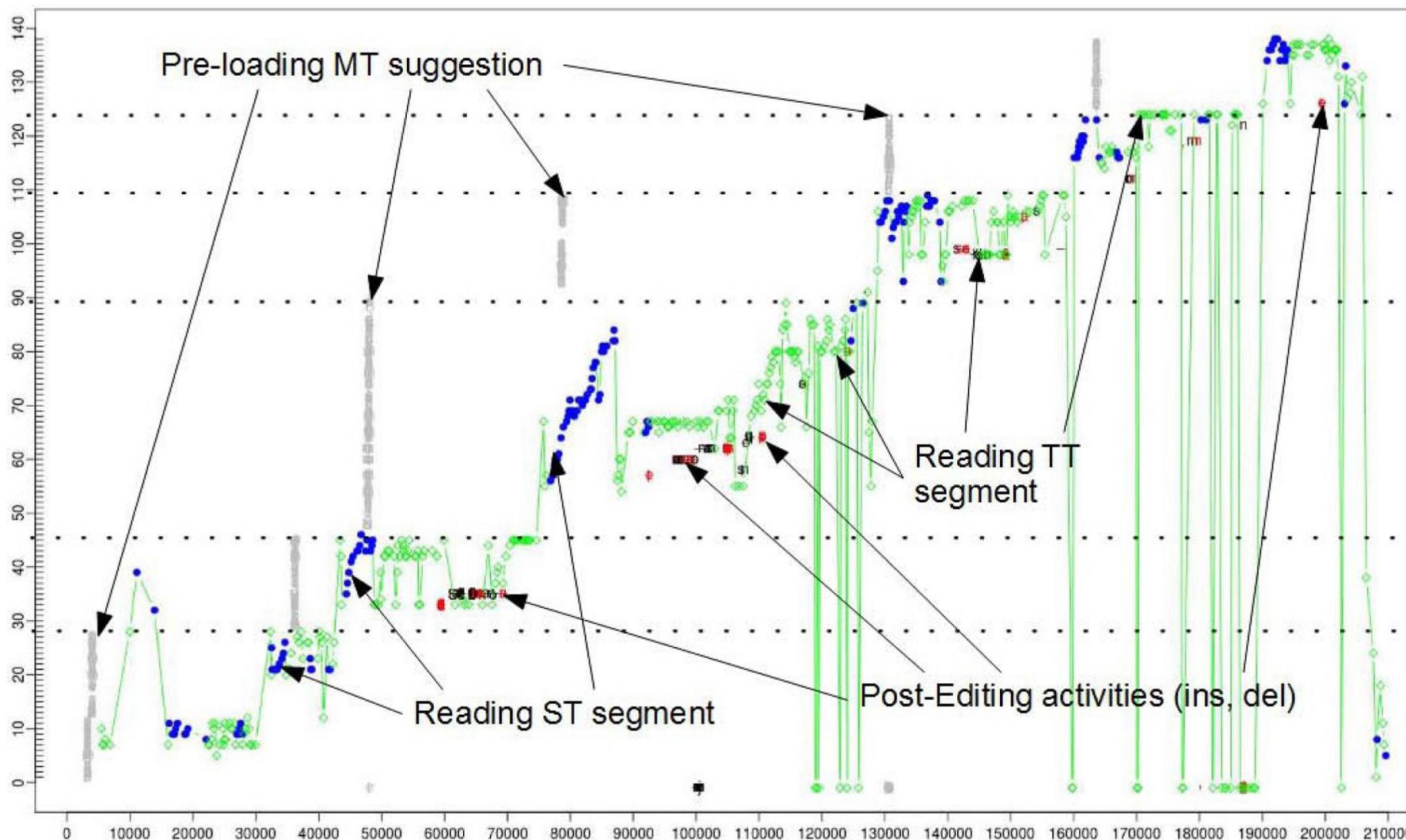


The screenshot shows a software interface for translation. At the top, there is a text input field containing the sentence: "However, the European Central Bank (ECB) asked about it in a report on virtual currencies published in October." Below this field is a row of buttons: "ITP", "PARA", "T→", "DRAFT", and "TRANSLATED". The "TRANSLATED" button is highlighted in blue. A dropdown menu is open under the "PARA" button, titled "Paraphrases for 'However' ✕". The menu lists two options: "on the other hand" and "nevertheless".

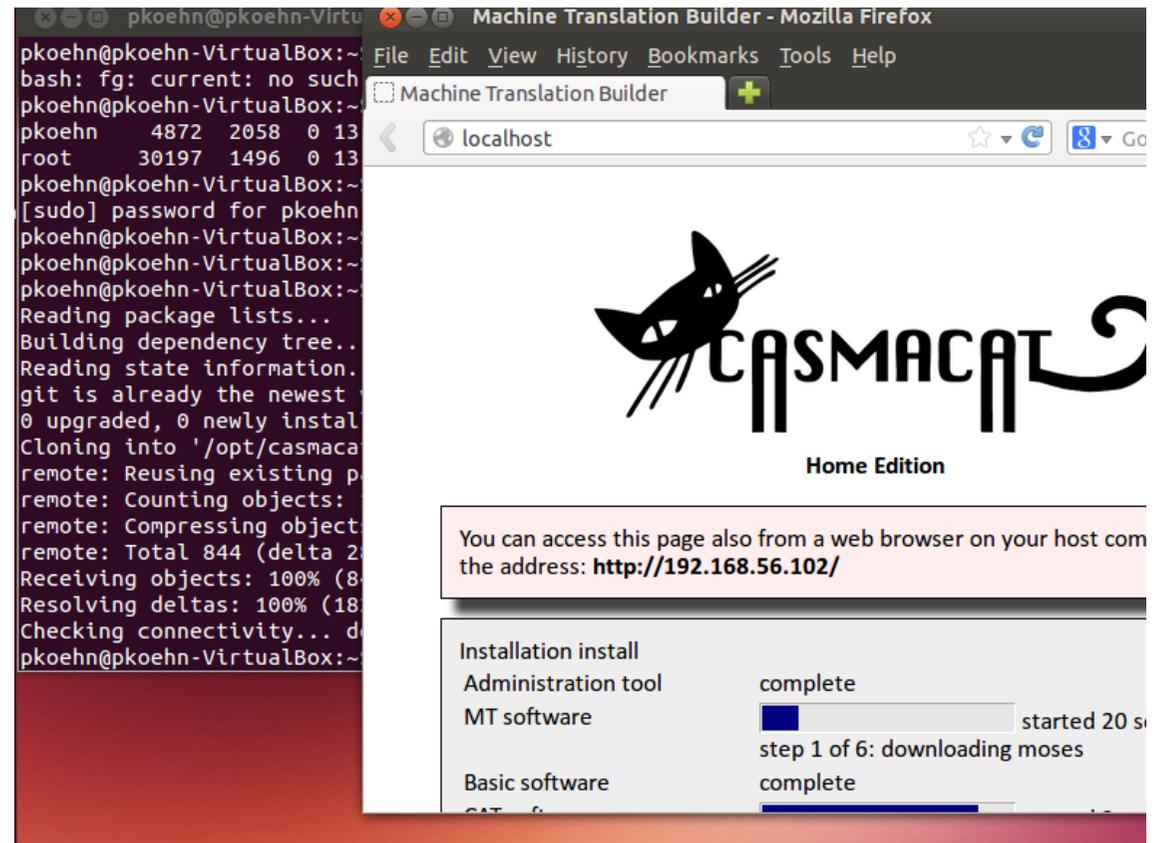
How do we Know it Works?

- Intrinsic Measures
 - word level confidence: user does not change words generated with certainty
 - interactive prediction: user accepts suggestions
- User Studies
 - professional translators faster with post-editing
 - ... but like interactive translation prediction better
- Cognitive studies with eye tracking
 - where is the translator looking at?
 - what causes the translator to be slow?

Logging and Eye Tracking



- Running CASMACAT on your desktop or laptop
- Installation
 - Installation software to run virtual machines (e.g., Virtualbox)
 - installation of Linux distribution (e.g., Ubuntu)
 - installation script sets up all the required software and dependencies



The screenshot shows a terminal window on the left and a web browser on the right. The terminal window displays the output of a script, including the installation of git and the cloning of the CASMACAT repository. The web browser shows the CASMACAT Home Edition website, which includes a progress bar for the installation process.

```
pkoehn@pkoehn-Virtu
bash: fg: current: no such
pkoehn@pkoehn-Virtu
pkoehn 4872 2058 0 13
root 30197 1496 0 13
pkoehn@pkoehn-Virtu
[sudo] password for pkoehn
pkoehn@pkoehn-Virtu
pkoehn@pkoehn-Virtu
pkoehn@pkoehn-Virtu
Reading package lists...
Building dependency tree..
Reading state information.
git is already the newest
0 upgraded, 0 newly instal
Cloning into '/opt/casmaca
remote: Reusing existing p
remote: Counting objects:
remote: Compressing object
remote: Total 844 (delta 2
Receiving objects: 100% (8
Resolving deltas: 100% (18
Checking connectivity... d
pkoehn@pkoehn-Virtu
```

Machine Translation Builder - Mozilla Firefox

Machine Translation Builder

localhost

 CASMACAT

Home Edition

You can access this page also from a web browser on your host com
the address: <http://192.168.56.102/>

Installation install	
Administration tool	complete
MT software	<div style="width: 50%;"></div> started 20 s
Basic software	complete

Administration through Web Browser

17



Administration

Translate

- [Translate new document](#)
- [List documents](#)

Engines

- [Manage engines](#)
- [Upload engine](#)
- [Build new prototype](#)

Settings

- [Reset CAT and MT server](#)
- [CAT Settings](#)
- [Update Software](#)

Deployed: fr-en-upload-1
Memory: 1.2 GB used, 6.6 GB free
Disk: 12.9 GB used, 10.2 GB free
Uptime: 22:24
Load: 0.01, 0.05, 0.08
Monday, 06 October 2014, 21:22:41



Training MT Engines

- Train MT engine on own or public data

Build New Prototype

Input language

Output language

Add corpus No file chosen

Name	Segments	Publisher	
European Central Bank	102,980	OPUS	upload
European Medicines Agency	372,824	OPUS	upload
EU Bookshop	3,618,897	OPUS	upload
European Constitution	6,667	OPUS	upload
European Parliament	1,260,689	OPUS	upload
KDE4	126,141	OPUS	uploaded
KDE4 (el-en_GB)	125,537	OPUS	upload
Open Subtitles	220,445	OPUS	upload
Open Subtitles 2011	10,693,456	OPUS	upload
Open Subtitles 2012	12,984,773	OPUS	upload
Open Subtitles 2013	14,626,890	OPUS	upload
South-East European Times	165,532	OPUS	upload
South-East European Times v2	224,808	OPUS	upload
SPC	7,035	OPUS	upload
Tatoeba	2,469	OPUS	upload
DGT-Translation Memory	3,016,402	JRC	upload

Corpora

Use	ID	Name	Segments	Uploaded
<input checked="" type="checkbox"/> all	1	KDE4	126141	21:39:27

Re-Use Previous setting

Tuning set all select

Evaluation set all select

Name

Managing MT Engines

Manage Engines

English-French

Available Engines

#	Name	Size	Build date	Action
2	NC+TED	2.3G	27 Mar 14	deploy delete download

Prototypes ([Inspect Details in Prototype Factory](#))

#	Name	Status	Build date	Action
2	NC+TED	done	Fri 20:34	delete
1	NC	done	Fri 20:34	create engine delete

English-Spanish

Available Engines

#	Name	Size	Build date	Action
2	NC+TED	2.3G	27 Mar 14	deploy delete download

Prototypes ([Inspect Details in Prototype Factory](#))

#	Name	Status	Build date	Action
3	NC+TED+EP	stopped	Fri 20:34	resume delete
2	NC+TED	done	Fri 20:34	delete
1	NC	done	Fri 20:34	create engine delete



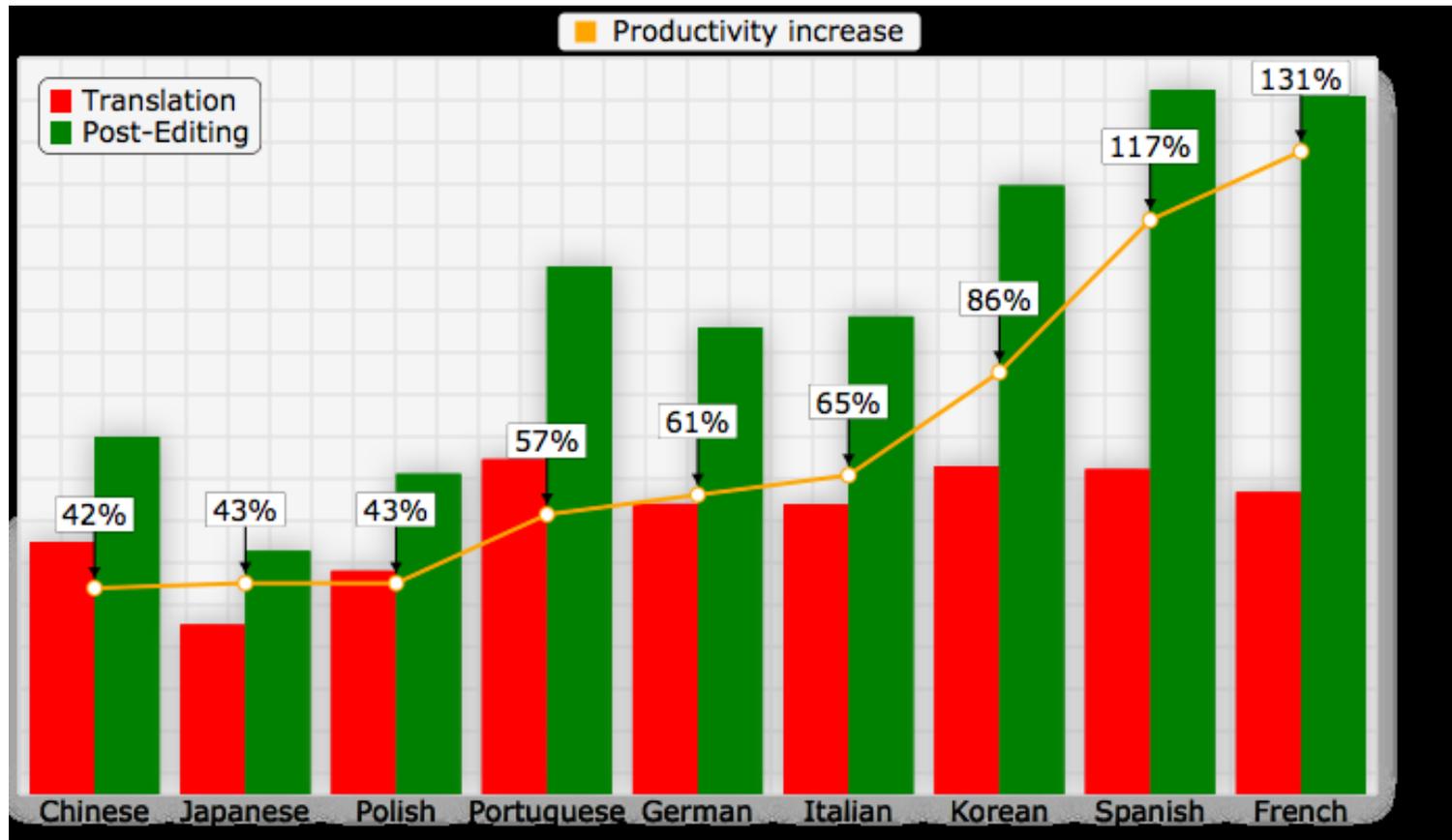
part II

cat methods



post-editing

Productivity Improvements



(source: Autodesk)

MT Quality and Productivity

System	BLEU	Training Sentences	Training Words (English)
MT1	30.37	14,700k	385m
MT2	30.08	7,350k	192m
MT3	29.60	3,675k	96m
MT4	29.16	1,837k	48m
MT5	28.61	918k	24m
MT6	27.89	459k	12m
MT7	26.93	230k	6.0m
MT8	26.14	115k	3.0m
MT9	24.85	57k	1.5m

- Same type of system (Spanish–English, phrase-based, Moses)
- Trained on varying amounts of data [Sanchez-Torron and Koehn, AMTA 2016]

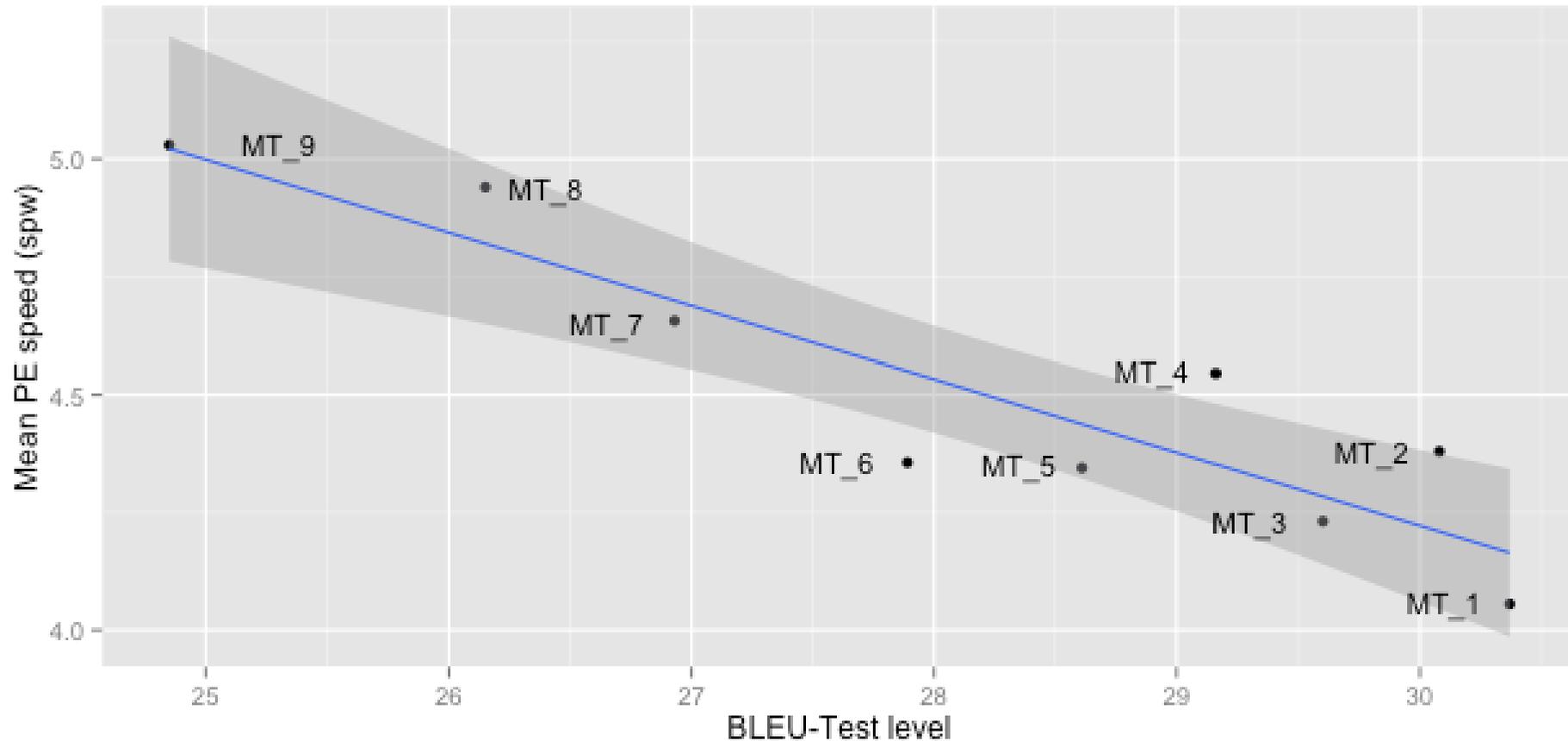
MT Quality and Productivity



System	BLEU	Training Sentences	Training Words (English)	Post-Editing Speed
MT1	30.37	14,700k	385m	4.06 sec/word
MT2	30.08	7,350k	192m	4.38 sec/word
MT3	29.60	3,675k	96m	4.23 sec/word
MT4	29.16	1,837k	48m	4.54 sec/word
MT5	28.61	918k	24m	4.35 sec/word
MT6	27.89	459k	12m	4.36 sec/word
MT7	26.93	230k	6.0m	4.66 sec/word
MT8	26.14	115k	3.0m	4.94 sec/word
MT9	24.85	57k	1.5m	5.03 sec/word

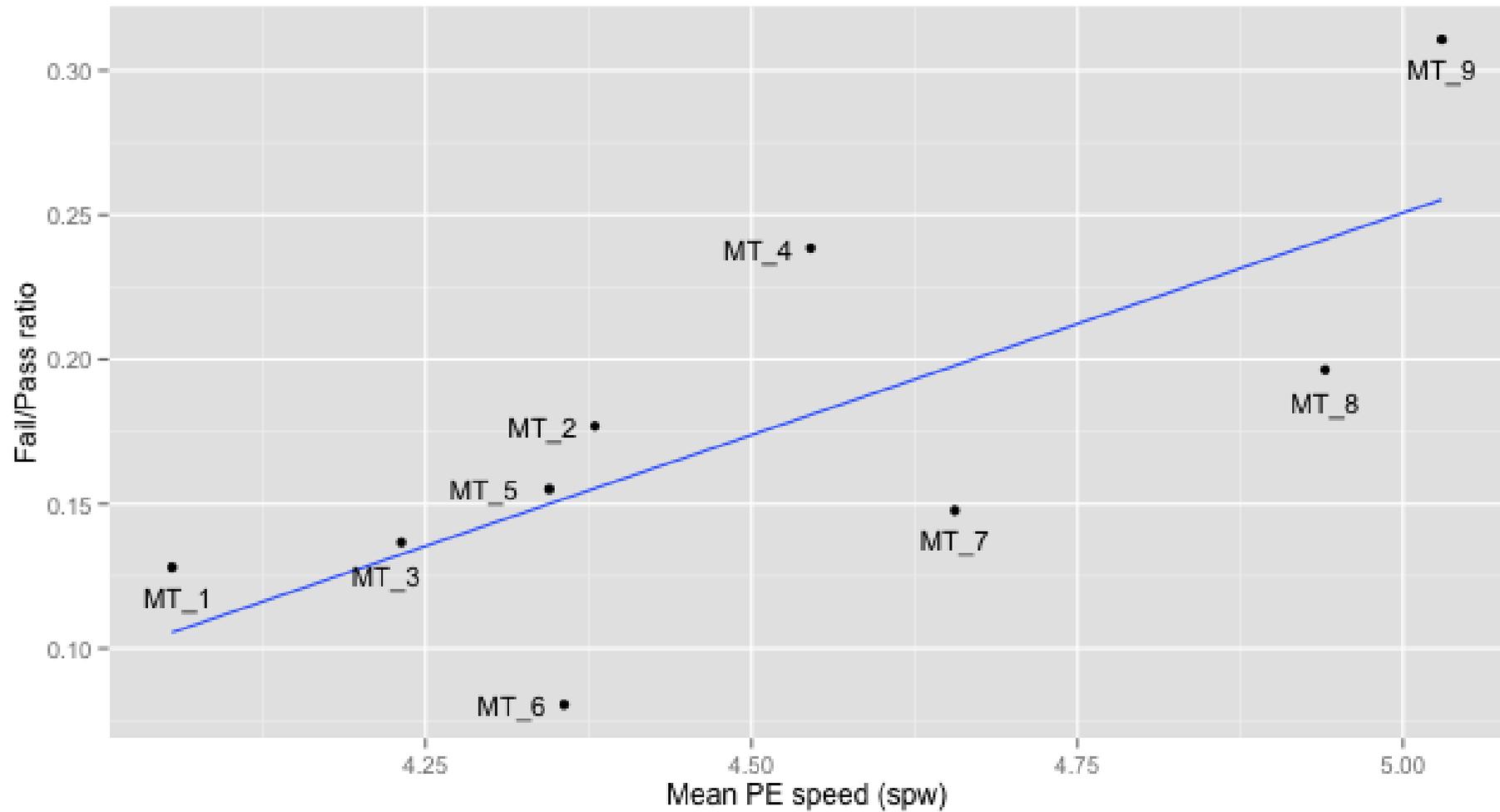
- User study with professional translators
- Correlation between BLEU and post-editing speed?

MT Quality and Productivity



BLEU against PE speed and regression line with 95% confidence bounds
+1 BLEU \leftrightarrow decrease in PE time of ~ 0.16 sec/word, or 3-4% speed-up

MT Quality and PE Quality



better MT \leftrightarrow fewer post-editing errors

Translator Variability

	HTER	Edit Rate	PE speed (spw)	MQM Score	Fail	Pass
TR1	44.79	2.29	4.57	98.65	10	124
TR2	42.76	3.33	4.14	97.13	23	102
TR3	34.18	2.05	3.25	96.50	26	106
TR4	49.90	3.52	2.98	98.10	17	120
TR5	54.28	4.72	4.68	97.45	17	119
TR6	37.14	2.78	2.86	97.43	24	113
TR7	39.18	2.23	6.36	97.92	18	112
TR8	50.77	7.63	6.29	97.20	19	117
TR9	39.21	2.81	5.45	96.48	22	113

- Higher variability between translators than between MT systems



confidence measures ("quality estimation")

Levels



- Machine translation engine indicates where it is likely wrong
- Different Levels of granularity
 - document-level (SDL's "TrustScore")
 - sentence-level
 - word-level

Sentence-Level Confidence

- Translators are used to “Fuzzy Match Score”
 - used in translation memory systems
 - roughly: ratio of words that are the same between input and TM source
 - if less than 70%, then not useful for post-editing
- We would like to have a similar score for machine translation■
- Even better
 - estimation of post-editing time
 - estimation of from-scratch translation time
 - can also be used for pricing
- Very active research area

Quality Estimation Shared Task

- Shared task organized at WMT since 2012
- Given
 - source sentence
 - machine translation
- Predict
 - human judgement of usefulness for post-editing (2012, 2014)
 - HTER score on post-edited sentences (2013–2016)
 - post-editing time (2013, 2014)
- Also task for word-level quality estimation (2014–2016) and document-level quality estimation (2015)

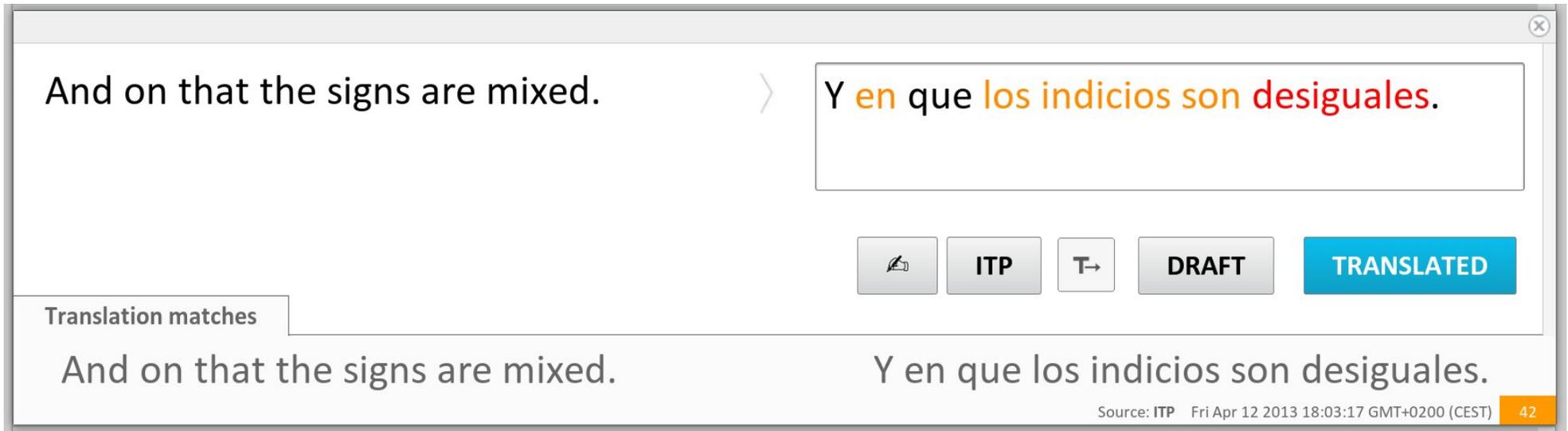
- Open source tool for quality estimation
- Source sentence features
 - number of tokens
 - language model (LM) probability
 - 1–3-grams observed in training corpus
 - average number of translations per word
- Similar target sentence features
- Alignment features
 - difference in number of tokens and characters
 - ratio of numbers, punctuation, nouns, verbs, named entities
 - syntactic similarity (POS tags, constituents, dependency relationships)
- Scores and properties of the machine translation derivation
- Uses Python's `SCIKIT-LEARN` implementation of SVM regression

WMT 2016: Best System

- Yandex School of Data Analysis (Kozlova et al., 2016)
- QuEst approach with additional features
 - syntactically motivated features
 - language model and statistics on web-scale corpus
 - pseudo-references and back-translations
 - other miscellaneous features
- Performance
 - mean average HTER difference 13.53
 - ranking correlation 0.525



word level confidence



The screenshot shows a translation interface. On the left, the source text is "And on that the signs are mixed." On the right, the target text is "Y en que los indicios son desiguales." The words "en", "que", "los", "indicios", "son", and "desiguales" in the target text are highlighted in orange and red. Below the source text, there is a tab labeled "Translation matches" which is currently empty. Below the target text, there is a row of buttons: a pencil icon, "ITP", "T→", "DRAFT", and "TRANSLATED" (which is highlighted in blue). At the bottom right, there is a status bar with the text "Source: ITP Fri Apr 12 2013 18:03:17 GMT+0200 (CEST)" and a small orange box containing the number "42".

- Highlight words less likely to be correct



- Simple methods quite effective
 - IBM Model 1 scores
 - posterior probability of the MT model

- Machine learning approach
 - similar features as for sentence-level quality estimation

Annotation

- Machine translation output

Quick brown fox jumps on the dog lazy.

- Post-editing

The quick brown fox jumps over the lazy dog.

- Annotation

<i>Fast</i>	<i>brown</i>	<i>fox</i>	<i>jumps</i>	<i>on</i>	<i>the</i>	<i>dog</i>	<i>lazy</i>	.
bad	good	good	good	bad	good	good	good	good

- Problems: dropped words? reordering?

Quality Requirements

- Evaluated in user study
- Feedback
 - could be useful feature
 - but accuracy not high enough
- To be truly useful, accuracy has to be very high
- Current methods cannot deliver this

WMT 2016: Best System

- Unbabel (Martins et al., 2016)
- Viewed as tagging task
- Features: black box and language model features
- Method: Combination of
 - feature-rich linear HMM model
 - deep neural networks
(feed-forward, bi-directionally recurrent, convolutional)
- Performance
 - F-score for detecting **good** words: 88.45
 - F-score for detecting **bad** words: 55.99



interactive translation prediction

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

|

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

| He

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

He | has



Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

He has | for months

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

He planned |

Input Sentence

Er hat seit Monaten geplant, im Oktober einen Vortrag in Miami zu halten.

Professional Translator

He planned | for months

Visualization

- Show n next words

Olvidarlo. Es demasiado | **arriesgado.** Estoy haciendo

- Show rest of sentence

Spence Green's Lilt System

- Show alternate translation predictions

C Les étudiants eux-mêmes n'ont pas les moyens de se rendre à des cours, nous essayons de les aider de cette manière.

The students themselves cannot be required to attend courses, we are trying to help them. **E**

D Dans le cadre de l'Institut Jedlička, nous transférerons ce projet dans un nouveau bâtiment.

themselves cannot
themselves could not
themselves do not
themselves cannot afford

- Show alternate translations predictions with probabilities

To equip students with training and reduce mobility and Institute jedlička,

- routinely
- steadily
- regular
- regularly

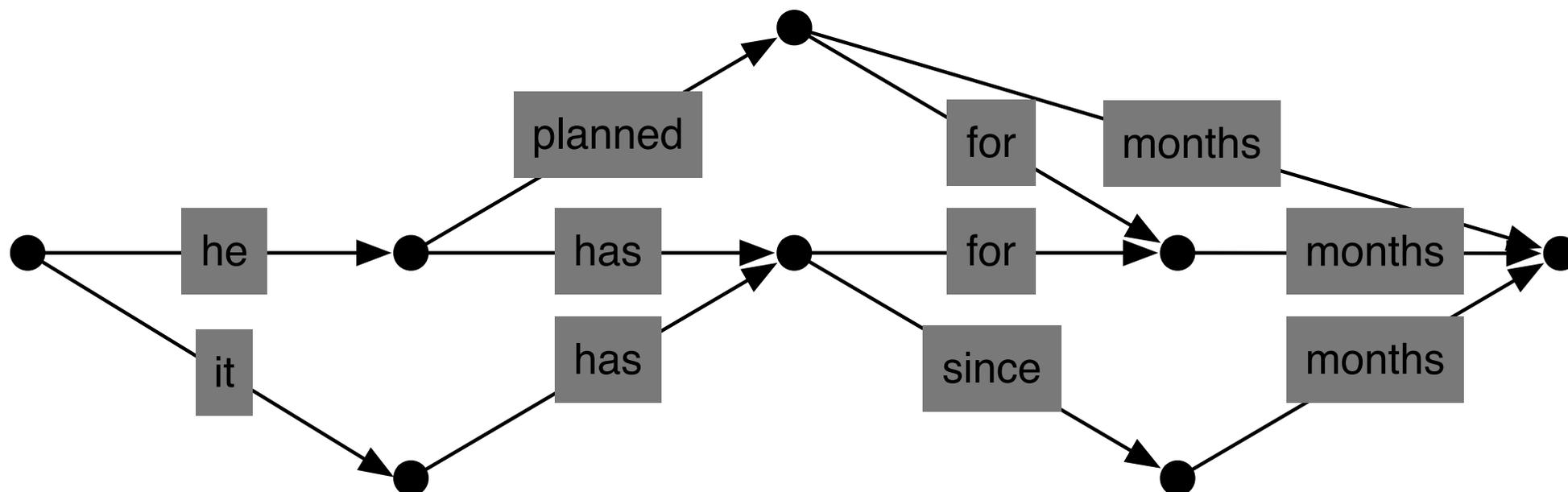
Des enseignants se rendent régulièrement auprès d'eux et proposent des activités qui les intéressent et les aident.

Teachers regularly visit Jedlička's activities and help them and their students.

Les étudiants eux-mêmes n'ont pas les moyens de se rendre à des cours, nous essayons de les aider de cette manière.

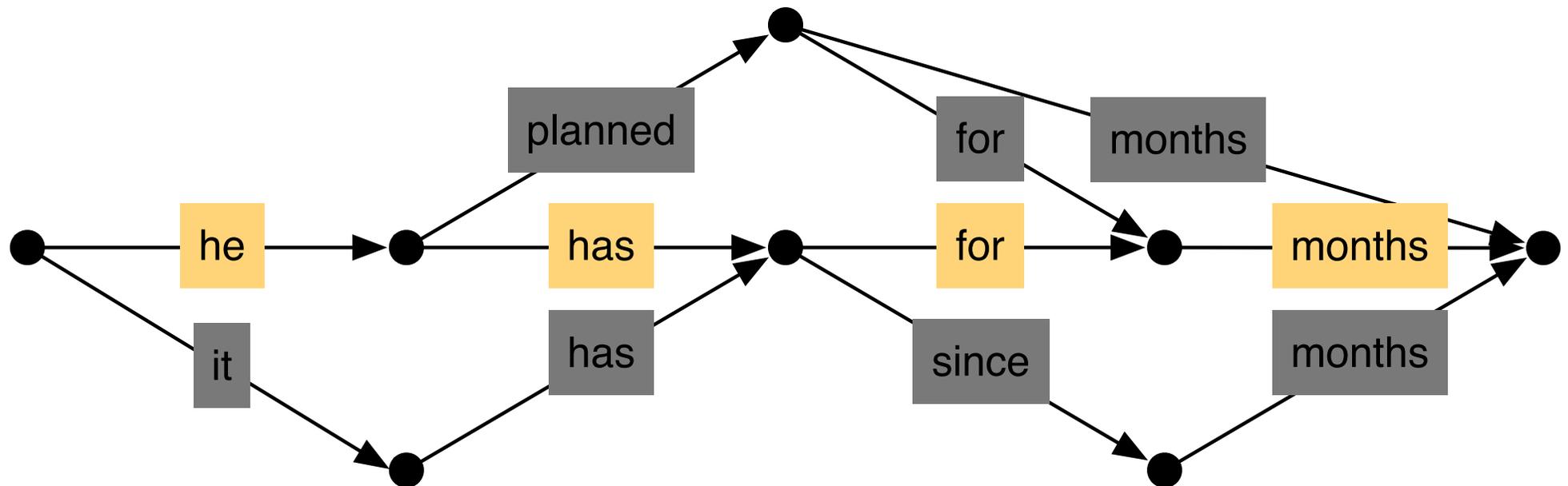
regularly visit
conduct ongoing
make regular
are regularly

Prediction from Search Graph



Search for best translation creates a graph of possible translations

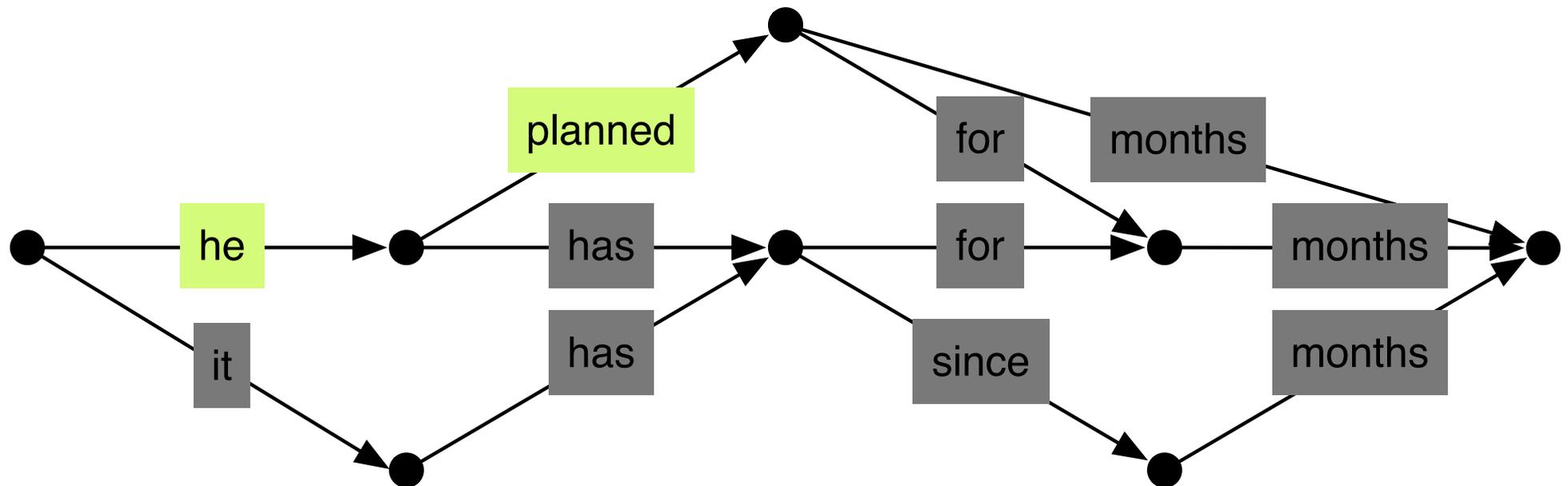
Prediction from Search Graph



One path in the graph is the best (according to the model)

This path is suggested to the user

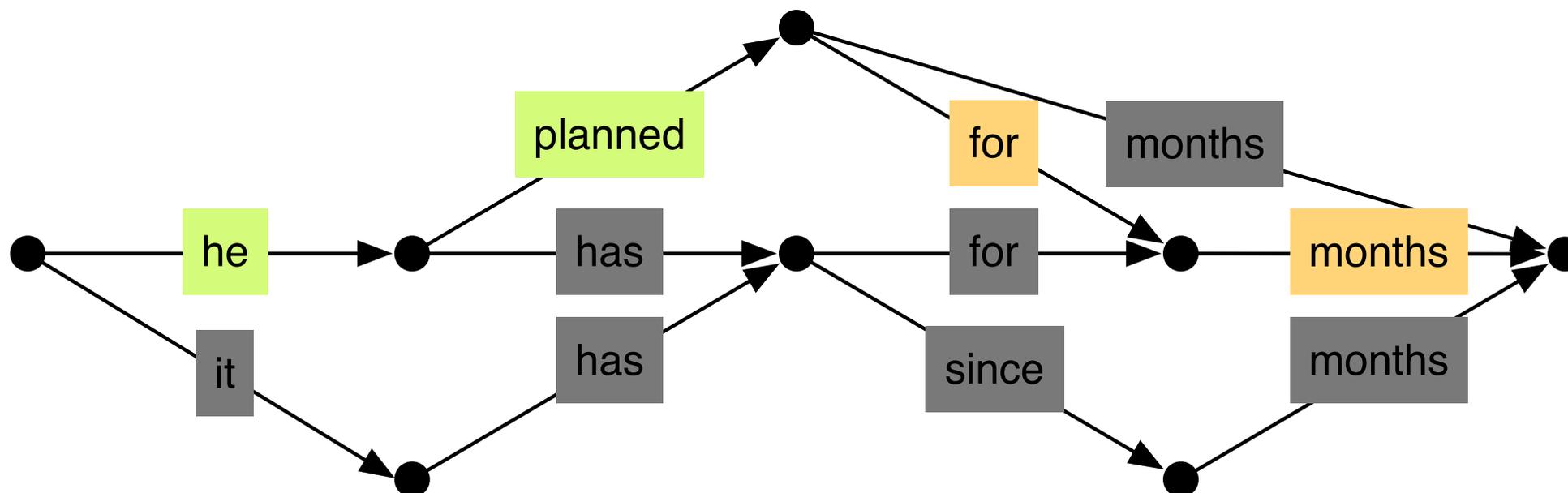
Prediction from Search Graph



The user may enter a different translation for the first words

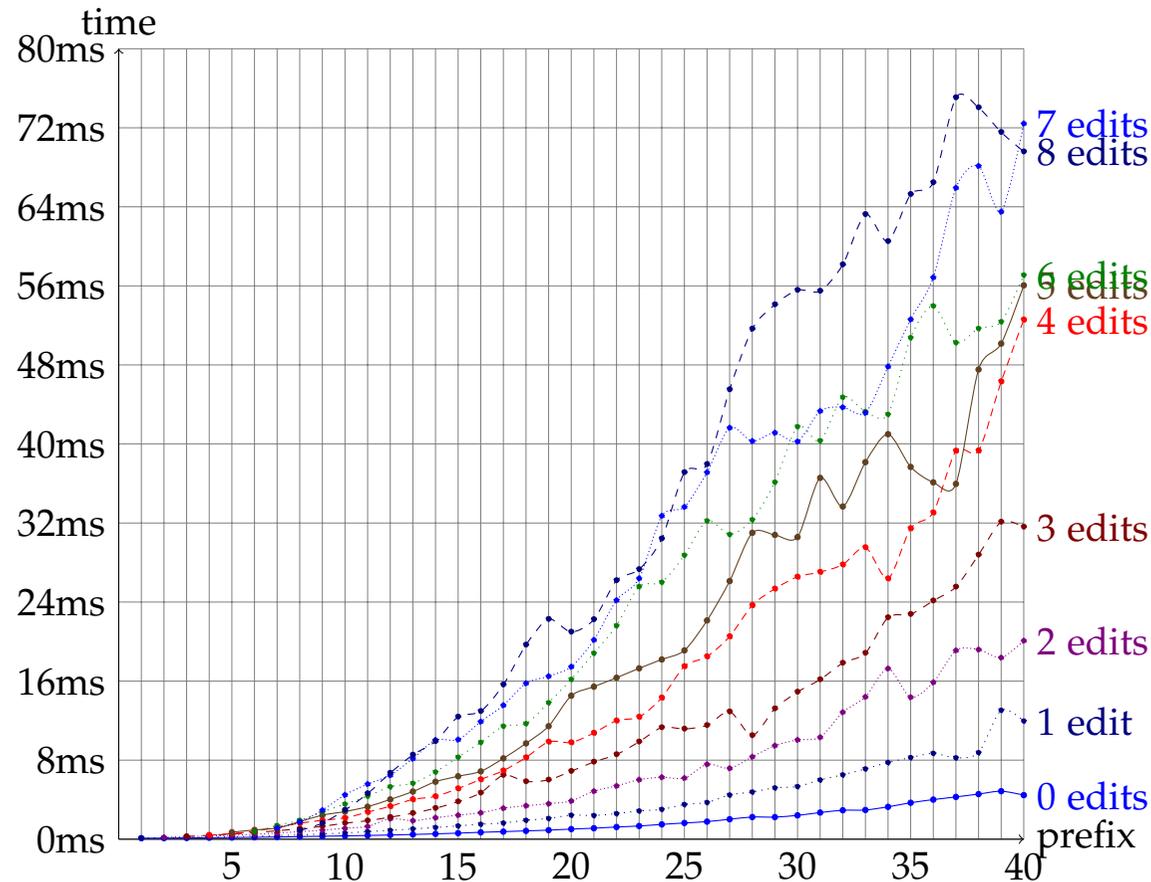
We have to find it in the graph

Prediction from Search Graph



We can predict the optimal completion (according to the model)

Speed of Algorithm



- Average response time based on length of the prefix and number of edits
- Main bottleneck is the string edit distance between prefix and path.

Word Completion

- Complete word once few letters are typed
- Example: predict *college* over *university*?
- User types the letter *u* → change prediction
- “Desperate” word completion: find any word that matches

- Translate the sentence again, enforce matching the prefix
- Recent work on this: Wuebker et al. [ACL 2016]

Models and Inference for Prefix-Constrained Machine Translation

**Joern Wuebker, Spence Green,
John DeNero, Saša Hasan**
Lilt, Inc.
`first_name@lilt.com`

Minh-Thang Luong
Stanford University
`lmthang@stanford.edu`

Prefix-Matching Decoding

- Prefix-matching phase
 - only allow translation options that match prefix
 - prune based on target words matched
- Ensure that prefix can be created by system
 - add synthetic translation options from word aligned prefix (but with low probability)
 - no reordering limit
- After prefix is match, regular beam search
- Fast enough?
 - ⇒ Wuebker et al. [ACL 2016] report 51-89ms per sentence

Tuning

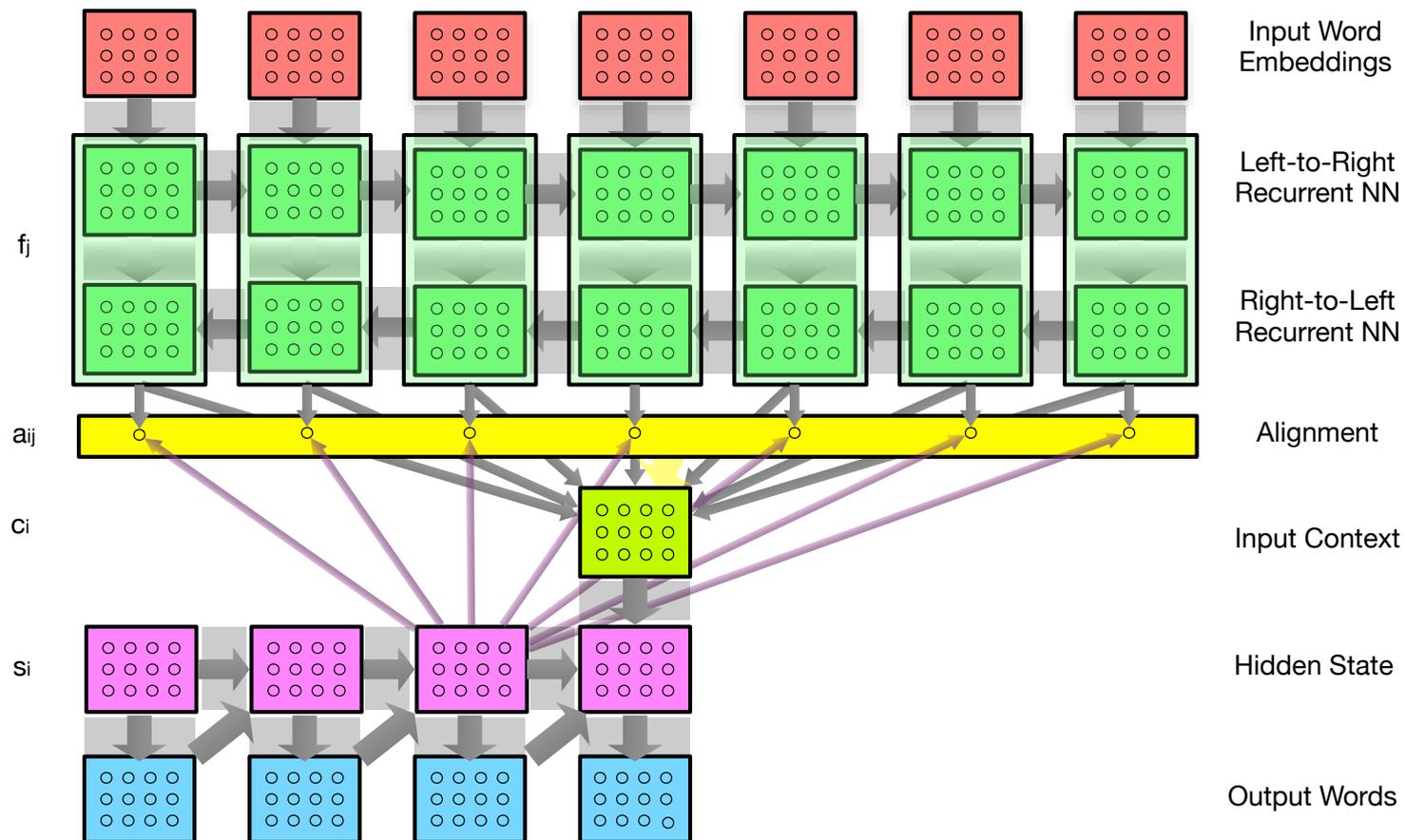
- Optimize to produce better predictions
- Focus on next few words, not full sentence
- Tuning metric
 - prefix BLEU (ignoring prefix to measure score)
 - word prediction accuracy
 - length of correctly predicted suffix sequence
- Generate diverse n-best list to ensure learnability
- Wuebker et al. [ACL 2016] report significant gains

Neural Interactive Translation Prediction

58



- Recent success of neural machine translation
- For instance, attention model



Neural MT: Sequential Prediction



- The model produces words in sequence

$$p(\text{output}_t | \{\text{output}_1, \dots, \text{output}_{t-1}\}, \vec{\text{input}}) = g(\hat{\text{output}}_{t-1}, \text{context}_t, \text{hidden}_t)$$

- Translation prediction: feed in user prefix

Example

Input: *Das Unternehmen sagte, dass es in diesem Monat mit Bewerbungsgesprächen beginnen wird und die Mitarbeiterzahl von Oktober bis Dezember steigt.*

	Correct	Prediction	Prediction probability distribution
✓	the	the	the (99.2%)
✓	company	company	company (90.9%) , firm (7.6%)
✓	said	said	said (98.9%)
✓	it	it	it (42.6%) , this (14.0%), that (13.1%), job (2.0%), the (1.7%), ...
✓	will	will	will (77.5%) , is (4.5%), started (2.5%), 's (2.0%), starts (1.8%), ...
✓	start	start	start (49.6%) , begin (46.7%)
	inter@@	job	job (16.1%), application (6.1%), en@@ (5.2%), out (4.8%), ...
✗	viewing	state	state (32.4%), related (5.8%), viewing (3.4%) , min@@ (2.0%), ...
✗	applicants	talks	talks (61.6%), interviews (6.4%), discussions (6.2%), ...
✓	this	this	this (88.1%) , so (1.9%), later (1.8%), that (1.1%)
✓	month	month	month (99.4%)
✗	,	and	and (90.8%), , (7.7%)
✗	with	and	and (42.6%), increasing (24.5%), rising (6.3%), with (5.1%) , ...
✓	staff	staff	staff (22.8%) , the (19.5%), employees (6.3%), employee (5.0%), ...
✗	levels	numbers	numbers (69.0%), levels (3.3%) , increasing (3.2%), ...
✗	rising	increasing	increasing (40.1%), rising (35.3%) , climbing (4.4%), rise (3.4%), ...
✓	from	from	from (97.4%)
✓	October	October	October (81.3%) , Oc@@ (12.8%), oc@@ (2.9%), Oct (1.2%)
✗	through	to	to (73.2%), through (15.6%) , until (8.7%)
✓	December	December	December (85.6%) , Dec (8.0%), to (5.1%)
✓	.	.	. (97.5%)

- Better prediction accuracy, even when systems have same BLEU score (state-of-the-art German-English systems, compared to search graph matching)

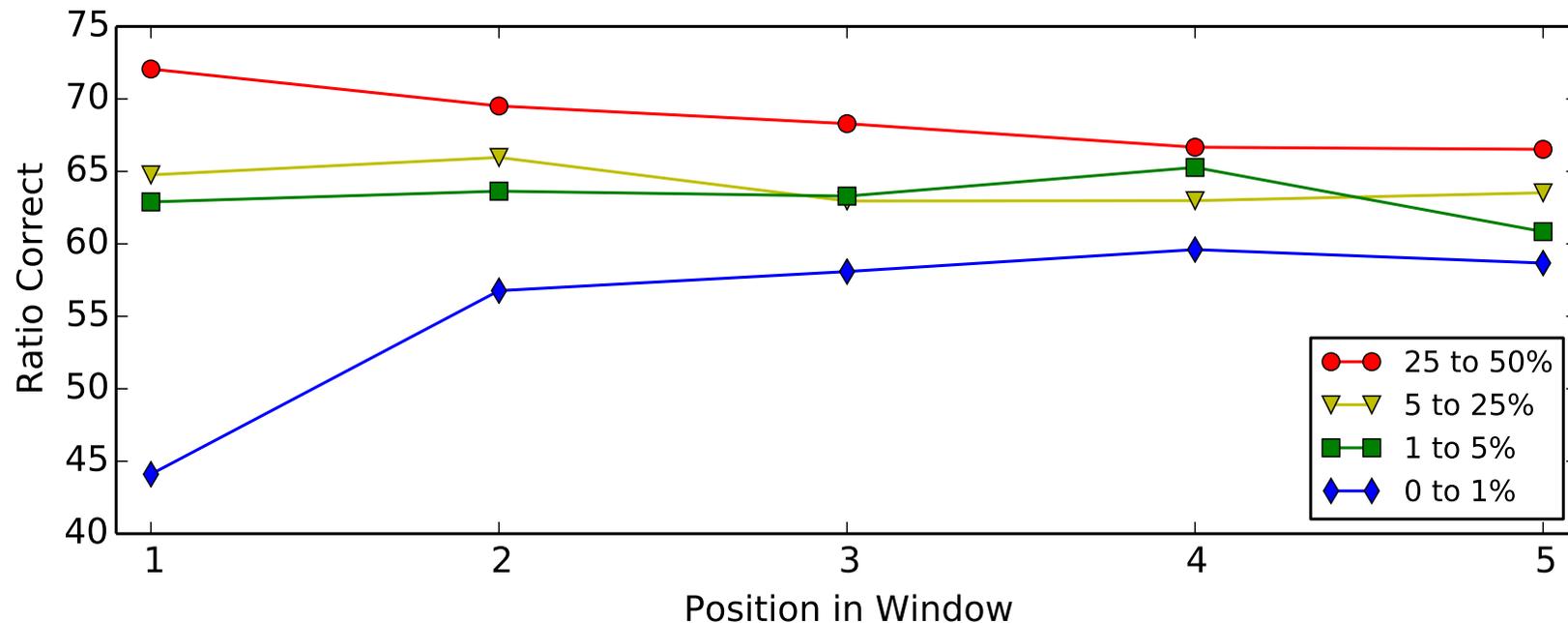
System	Configuration	BLEU	Word Prediction Accuracy	Letter Prediction Accuracy
Neural	no beam search	34.5	61.6%	86.8%
	beam size 12	36.2	63.6%	87.4%
Phrase-based	-	34.5	43.3%	72.8%

Recovery from Failure

- Ratio of words correct after first failure

System	Configuration	1	2	3	4	5
Neural	no beam search	55.9%	61.8%	61.3%	62.2%	61.1%
	beam size 12	58.0%	62.9%	62.8%	64.0%	61.5%
Phrase-based	-	28.6%	45.5%	46.9%	47.4%	48.4%

- Depending on probability of user word (neural, no beam)



Patching Translations

- Decoding speeds
 - translation speed with CPU: 100 ms/word
 - translation speed with GPU: 7ms/word
- To stay within 100ms speed limit
 - predict only a few words ahead (say, 5, in $5 \times 7\text{ms} = 35\text{ms}$)
 - patch new partial prediction with old full sentence prediction
 - uses KL divergence to find best patch point in ± 2 word window
- May compute new full sentence prediction in background, return as update
- Only doing quick response reduces word prediction accuracy 61.6% \rightarrow 56.4%



translation options

Translation Option Array

climbers are severely injured, and ten people are missing
 after Mount Ontake (御嶽山, Ontake-san), a popular climbing
 spot in central Japan, **erupted** for the first time in five years.

Kletterer sind schwer verletzt, und zehn Menschen werden
 vermisst, nachdem Mount Ontake (御嶽山, Ontake-san), ein
 beliebter Kletterplatz im zentralen Japan,

ausbruch, zum ersten

ITP ☰ T→ DRAFT **TRANSLATED**

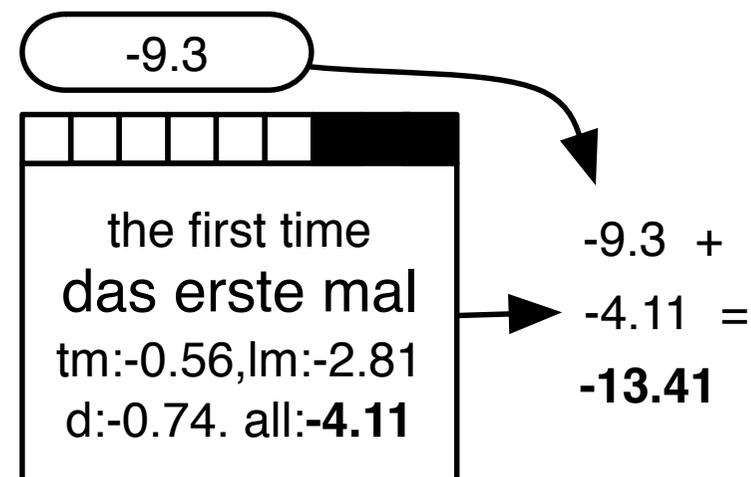
Translation Options

ke	-	san)	,	a	popular	climbing	spot	in central	Japan	,	erupted	for the first time in five years	.
ke	-	san)	,	ein	beliebtes	Klettern	vor Ort	in Mittel-	Japan,		ausbruch	zum ersten Mal in fünf Jahren	.
	und	San)	,	ein	populär	Bergsteigen	vor	zentrale	Japan	,	ausbruch,	zum ersten Mal in	fünf Jahre.
	/)	, die		beliebt	Aufstieg	Fleck	zentralen	Japans,		platzte	zum ersten Mal	fünf Jahre
	der)		eine	beliebte	abhalten,	ein, in	zentraler	Japan		Ausbruch		in fünf Jahren
	bis)	, in	populär		Erklimmen	Vor - Ort @-@	zentral	Japans	.	ausgebrochen	zum ersten Mal in der	von fünf Jahren.
	von)	.	populär ist,	beim Besteigen		in	mittel-	in Japan	-	ausgebrochen ist	zum ersten Mal seit	fünf Jahren sind.

- Visual aid: non-intrusive provision of cues to the translator
- Trigger passive vocabulary

How to Rank

- Basic idea: best options on top
- Problem: how to rank word translation vs. phrase translations?
- Method: utilize future cost estimates
- Translation score
 - sum of translation model costs
 - language model estimate
 - outside future cost estimate



Improving Rankings

- Removal of duplicates and near duplicates

bad	good
erupted	climbing
ausbrach	Klettern
ausbrach,	Bergsteigen
platzte	Aufstieg
Ausbruch	abhalten,
ausgebrochen	Erklimmen
ausgebrochen ist	beim Besteigen

- Ranking by likelihood to be used in the translation
→ can this be learned from user feedback?

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?

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



Example: Arabic

وكان	مجلس	النواب	الاميركى	اعتمد	الخميس	قانونا	يطالب	يسحب	القوات	المقاتلة	الاميركية	من	العراق	فى	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	الذى	يعارض	اي	تحدد	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															

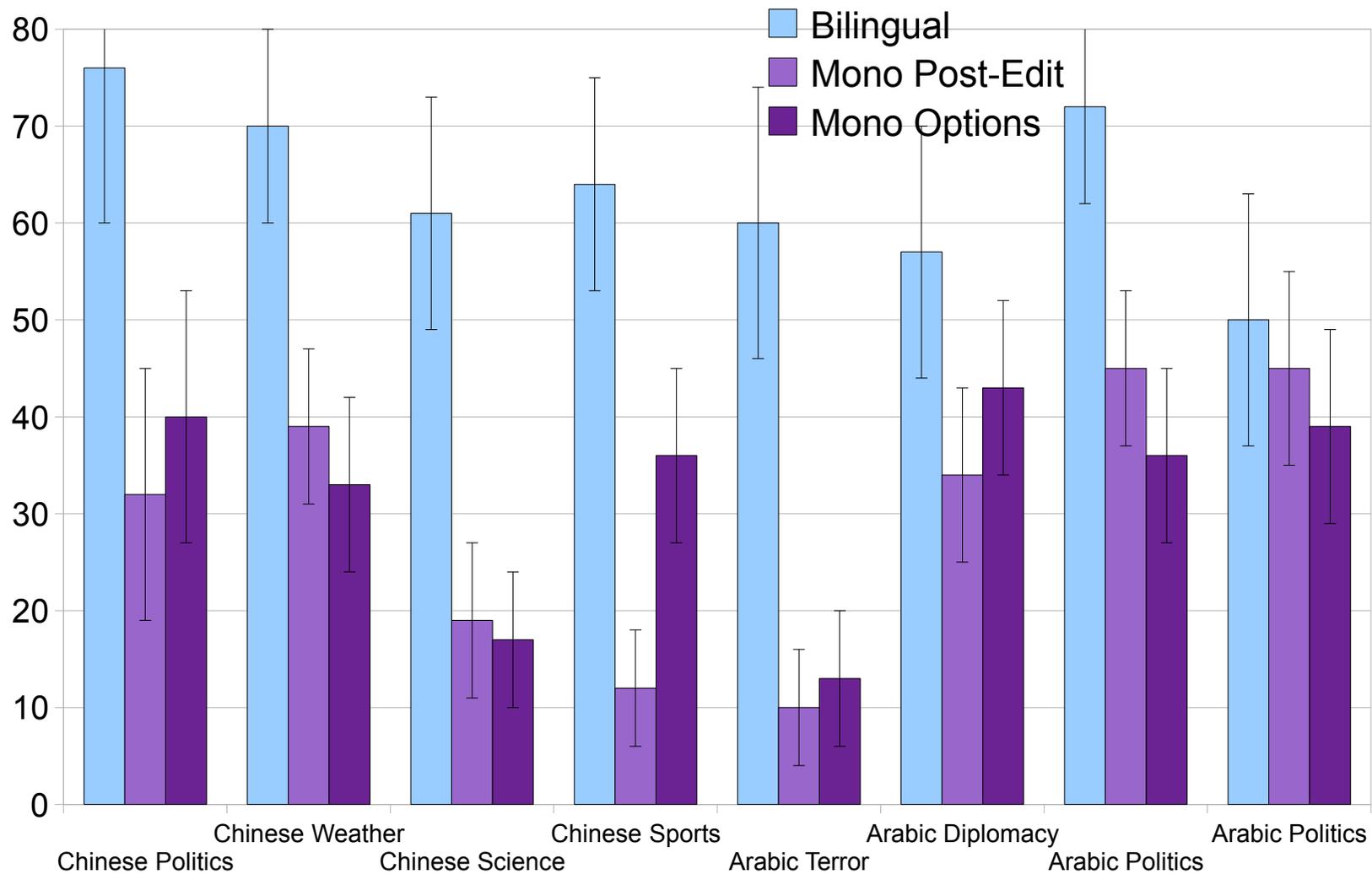
up to 10 translations for each word / phrase

Example: Arabic



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	iraq
the withdrawal	forces	the fighter		from	
the withdrawal of	troops			iraq	
withdrawal of				of the	
withdrawal				from iraq in	
the withdrawal			the american		

Monolingual Translation with Options



No big difference — once significantly better

Monolingual Translation Triage

- Study on Russian–English (Schwartz, 2014)
- Allow monolingual translators to assess their translation
 - confident → accept the translation
 - verify → proofread by bilingual
 - partially unsure → part of translation handled by bilingual
 - completely unsure → handled by bilingual
- Monolingual translator highly effective in triage

Monolingual Translation: 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

logging and eye tracking

Logging functions

- Different types of events are saved in the logging.
 - configuration and statistics
 - start and stop session
 - segment opened and closed
 - text, key strokes, and mouse events
 - scroll and resize
 - search and replace
 - suggestions loaded and suggestion chosen
 - interactive translation prediction
 - gaze and fixation from eye tracker

Logging functions

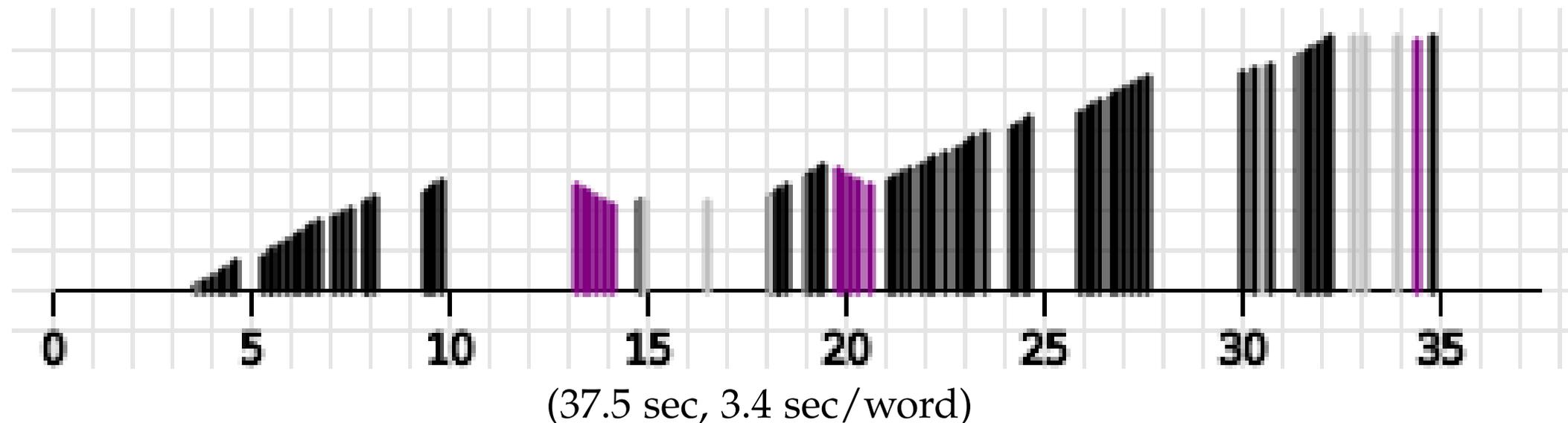
- In every event we save:
 - Type
 - In which element was produced
 - Time
- Special attributes are kept for some types of events
 - Diff of a text change
 - Current cursor position
 - Character looked at
 - Clicked UI element
 - Selected text

⇒ Full replay of user session is possible

Keystroke Log

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

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



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

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*)

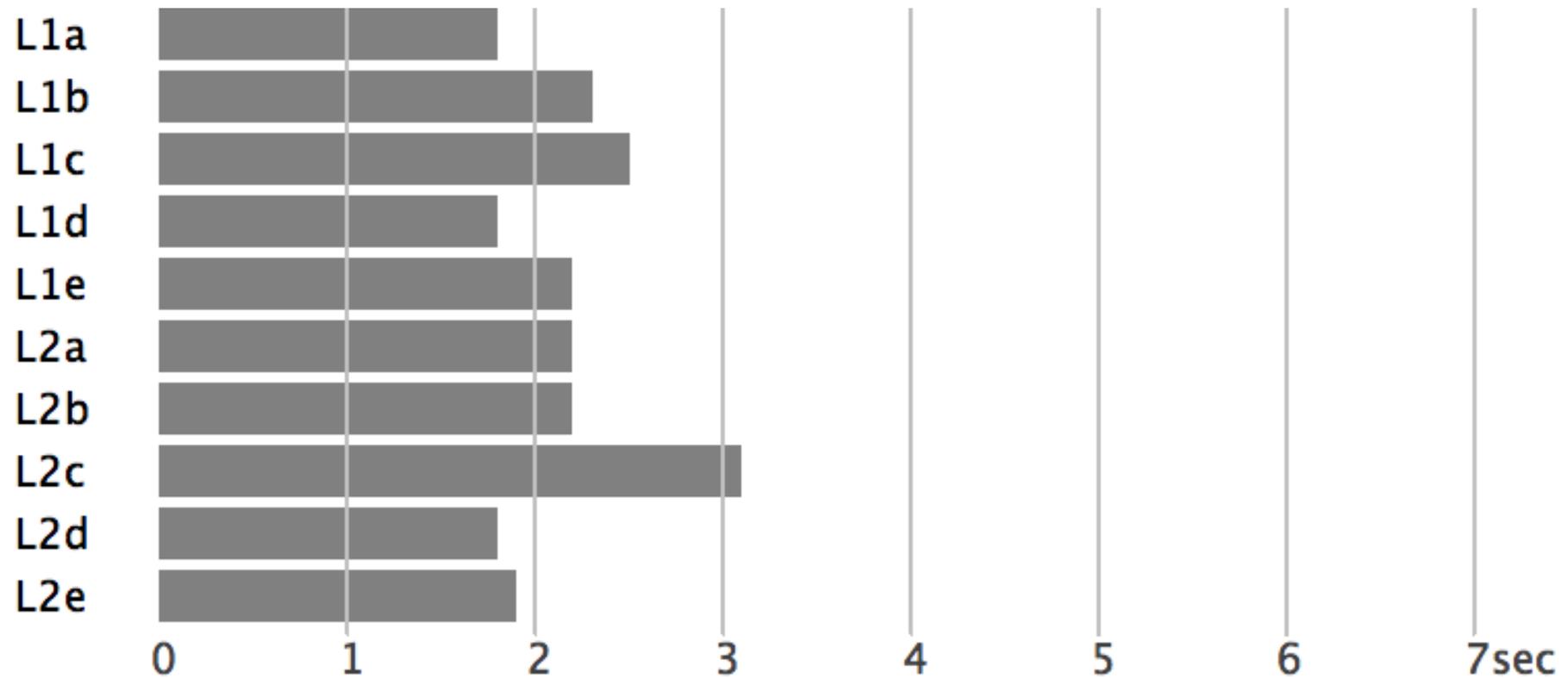
Main Measure: Productivity

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

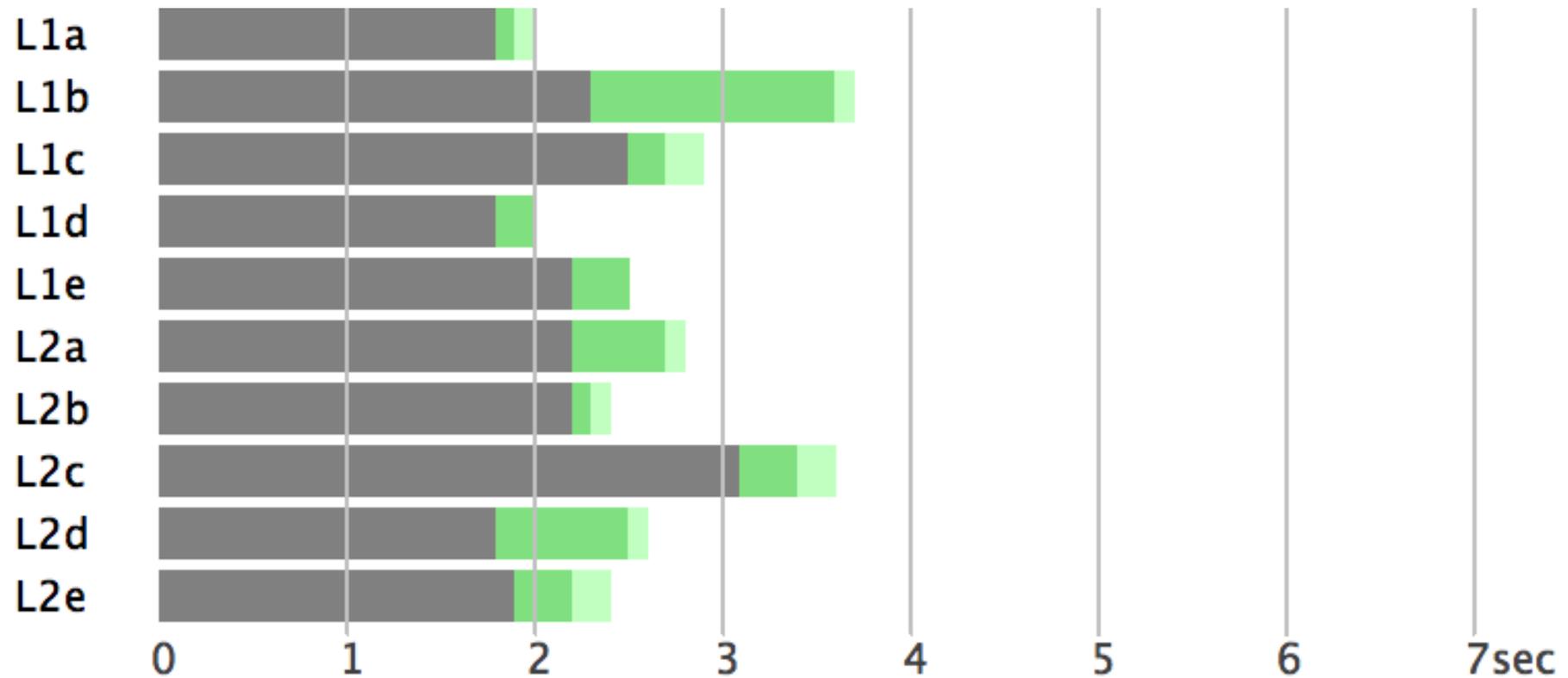
User	Unassisted	Postedit	Options	Prediction	Prediction+Options
L1a	3.3sec/word 23% correct	1.2s -2.2s 39% +16%)	2.3s -1.0s 45% +22%	1.1s -2.2s 30% +7%)	2.4s -0.9s 44% +21%
L1b	7.7sec/word 35% correct	4.5s -3.2s) 48% +13%	4.5s -3.3s 55% +20%	2.7s -5.1s 61% +26%	4.8s -3.0s 41% +6%
L1c	3.9sec/word 50% correct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

Unassisted Novice Translators



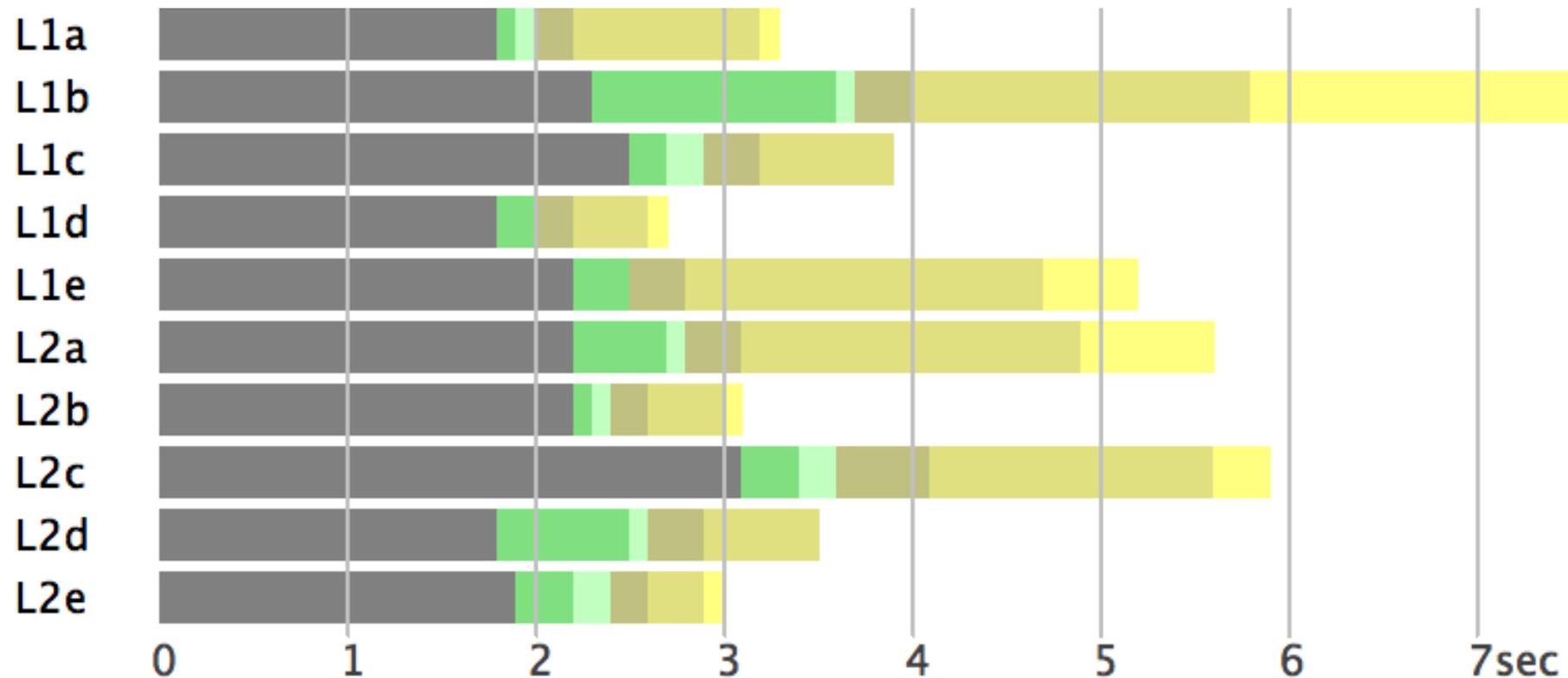
L1 = native French, L2 = native English, average time per input word
only typing

Unassisted Novice Translators



L1 = native French, L2 = native English, average time per input word
typing, **initial and final pauses**

Unassisted Novice Translators



L1 = native French, L2 = native English, average time per input word
typing, **initial and final pauses**, **short, medium, and long pauses**
most time difference on intermediate pauses

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

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%	-

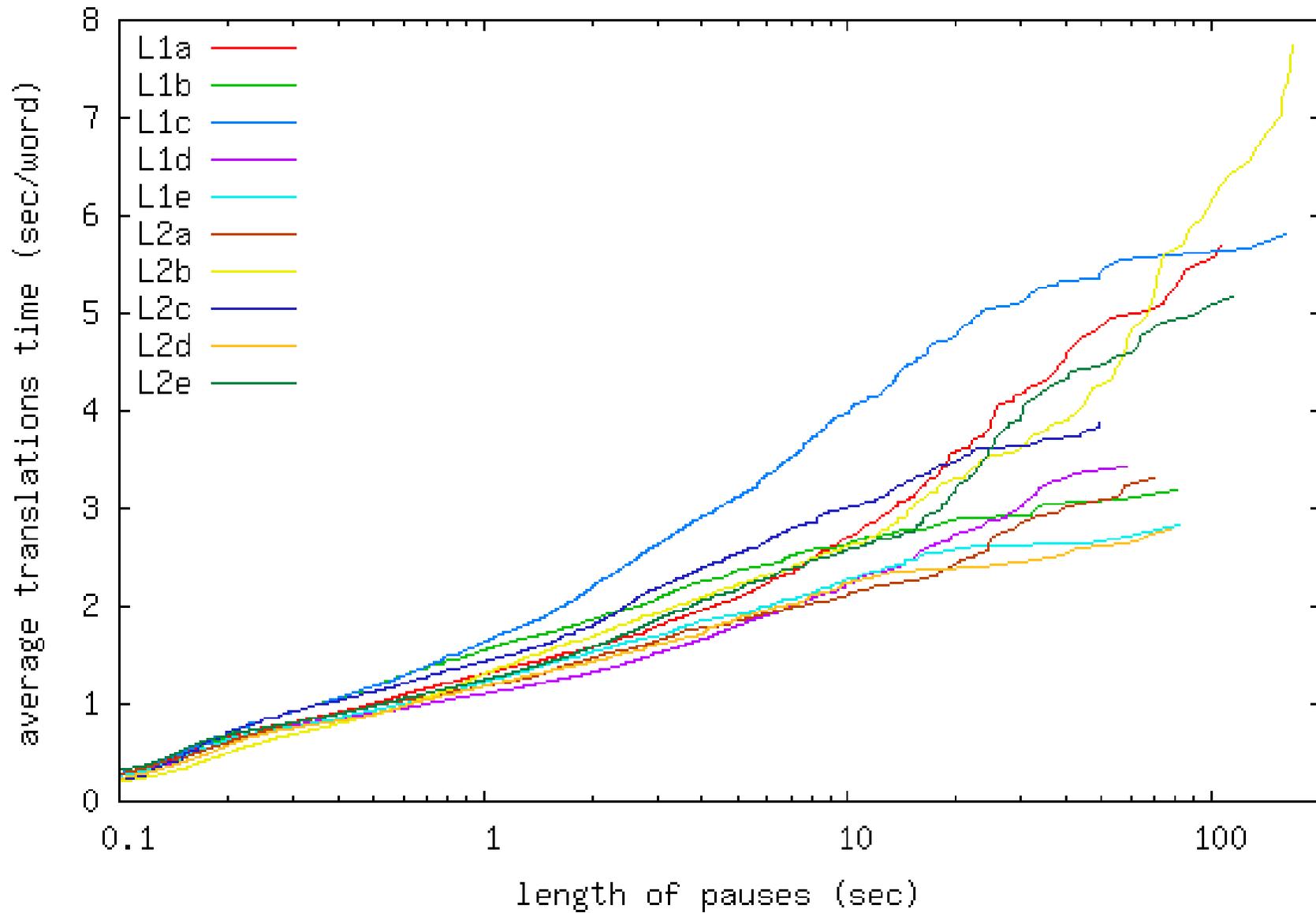
Translation comes to large degree from assistance

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

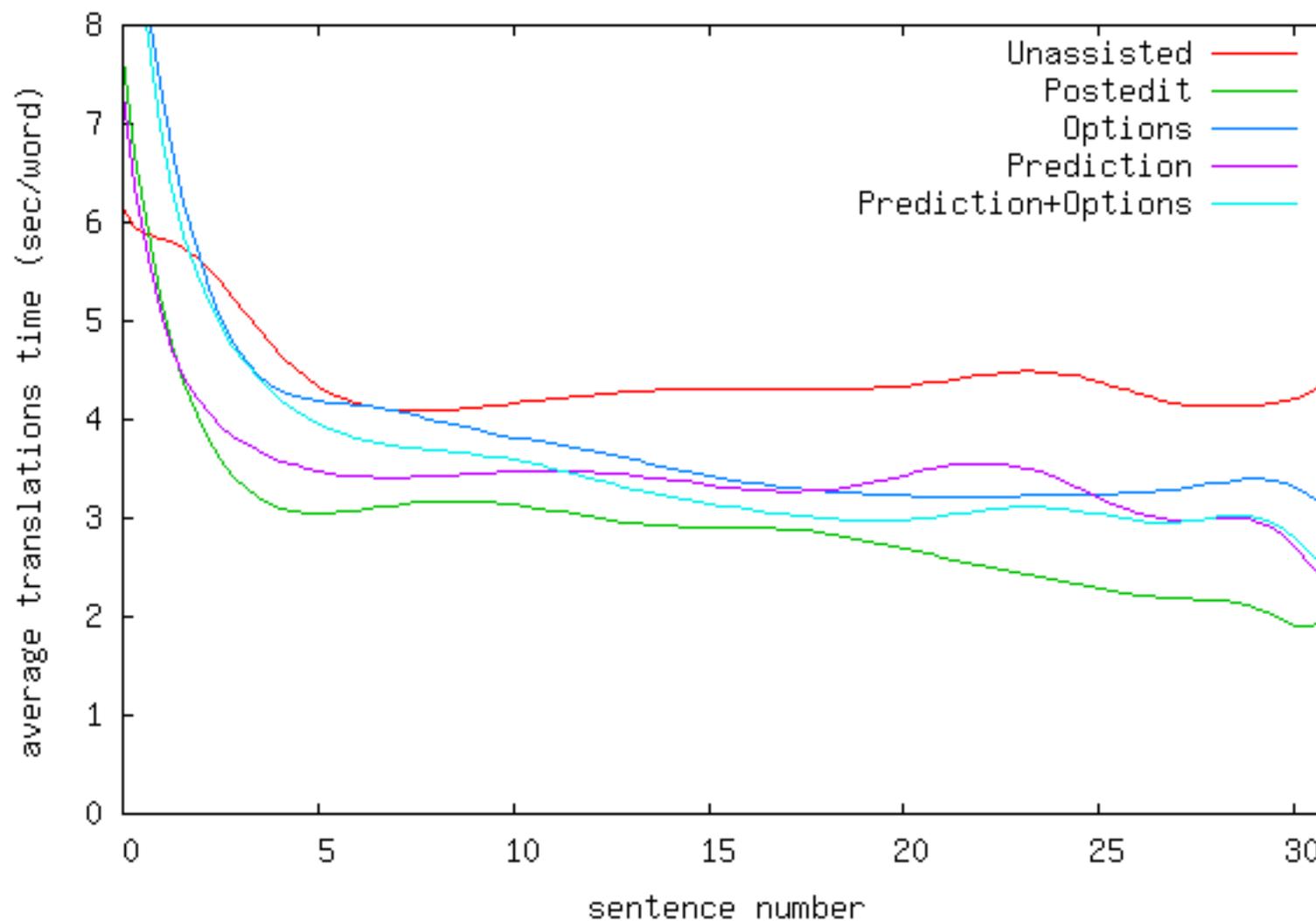
$$sum(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)$$

Results



Learning Effects

Users become better over time with assistance



Learning Effects: Professional Translators

94

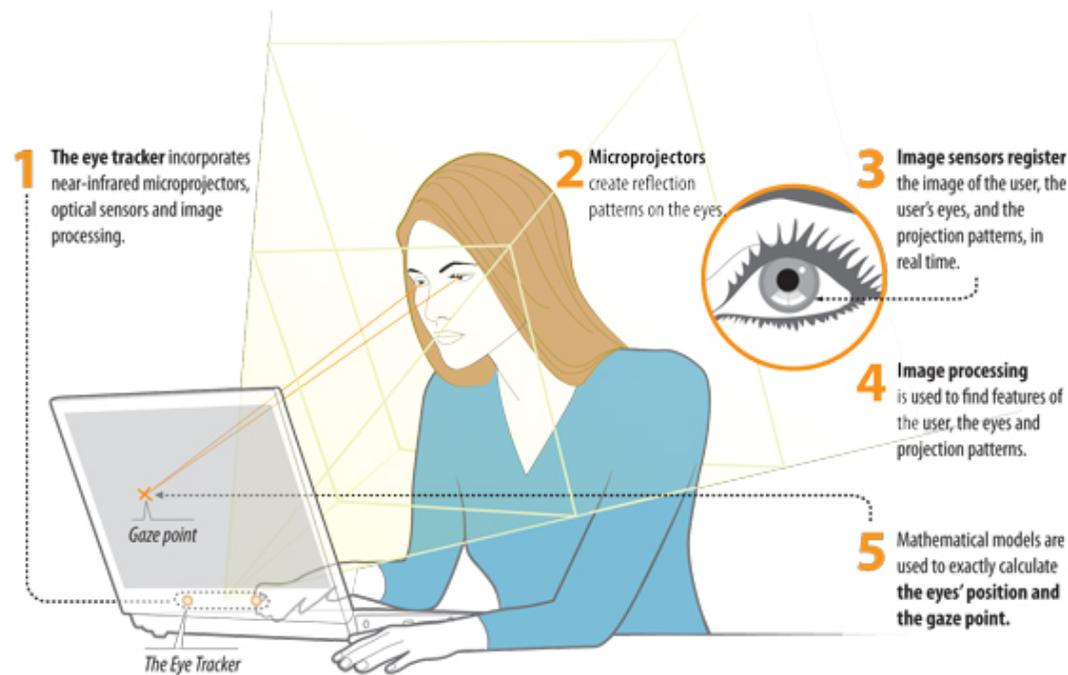


CASMACAT longitudinal study

Productivity projection as reflected in Kdur taking into account six weeks

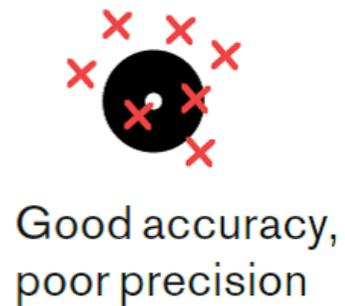
(Kdur = user activity excluding pauses > 5 seconds)

Eye Tracking



- Eye trackers extensively used in cognitive studies of, e.g., reading behavior
- Overcomes weakness of key logger: what happens during pauses
- Fixation: where is the focus of the gaze
- Pupil dilation: indicates degree of concentration

- Problem: Accuracy and precision of gaze samples



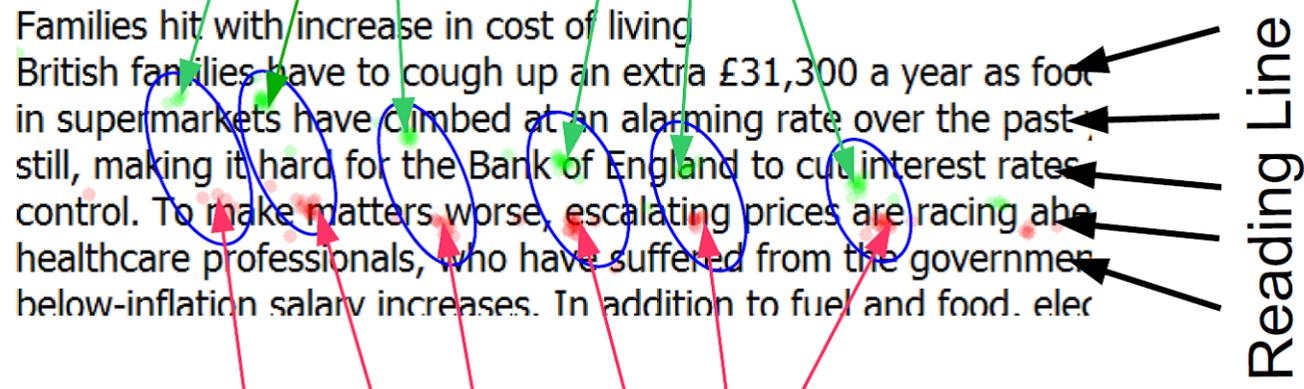
× = *eye tracker result*
● = *target looked at*

Gaze-to-Word Mapping

- Recorded gaze locations and fixations

Right eye gaze samples

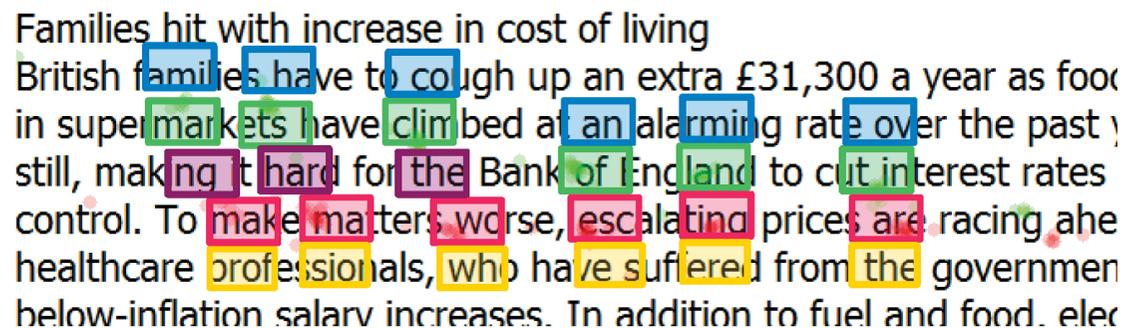
Families hit with increase in cost of living
 British families have to cough up an extra £31,300 a year as food
 in supermarkets have climbed at an alarming rate over the past
 still, making it hard for the Bank of England to cut interest rates
 control. To make matters worse, escalating prices are racing ahead
 healthcare professionals, who have suffered from the government
 below-inflation salary increases. In addition to fuel and food. elec



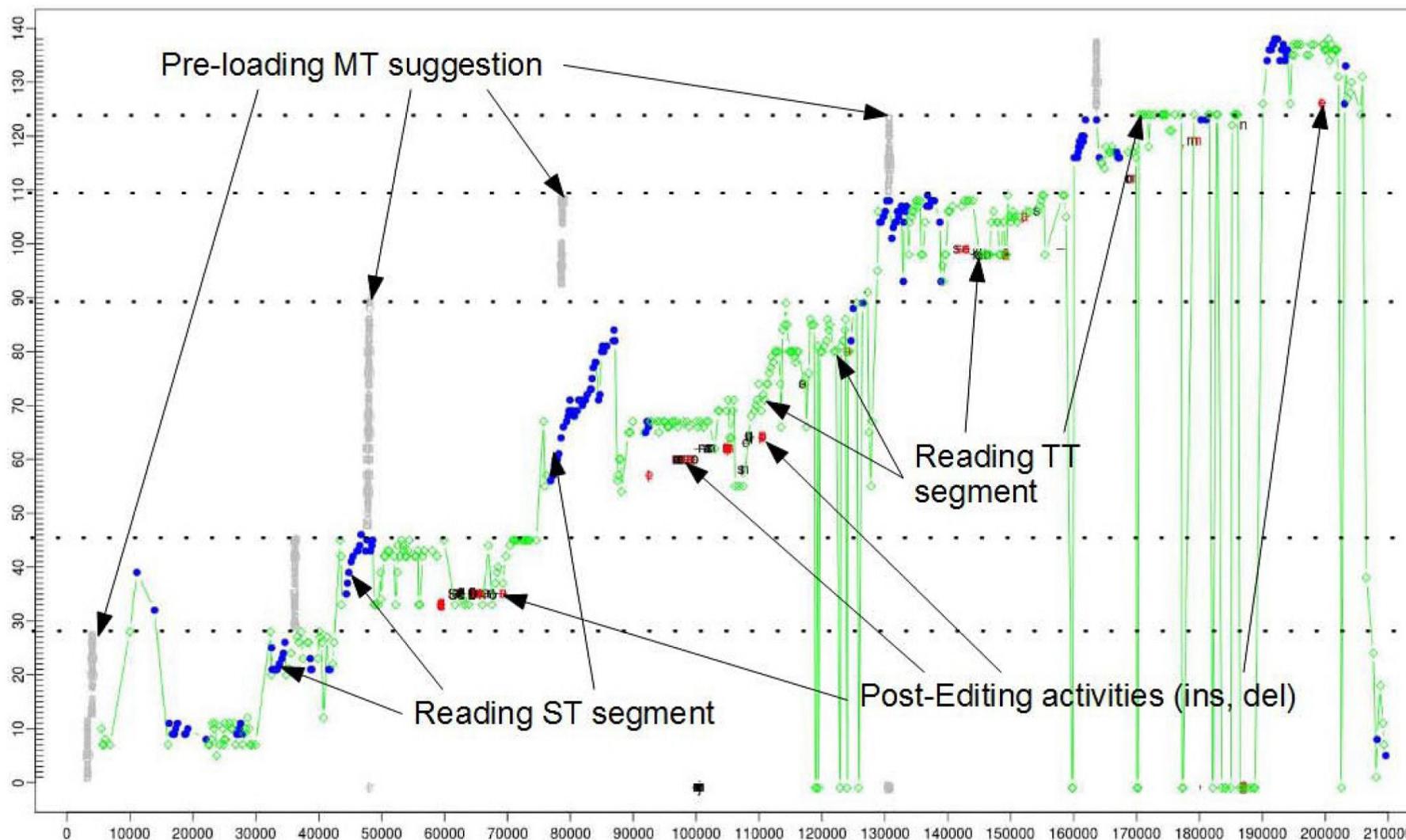
Left eye gaze samples

- Gaze-to-word mapping

Families hit with increase in cost of living
 British families have to cough up an extra £31,300 a year as food
 in supermarkets have climbed at an alarming rate over the past
 still, making it hard for the Bank of England to cut interest rates
 control. To make matters worse, escalating prices are racing ahead
 healthcare professionals, who have suffered from the government
 below-inflation salary increases. In addition to fuel and food. elec

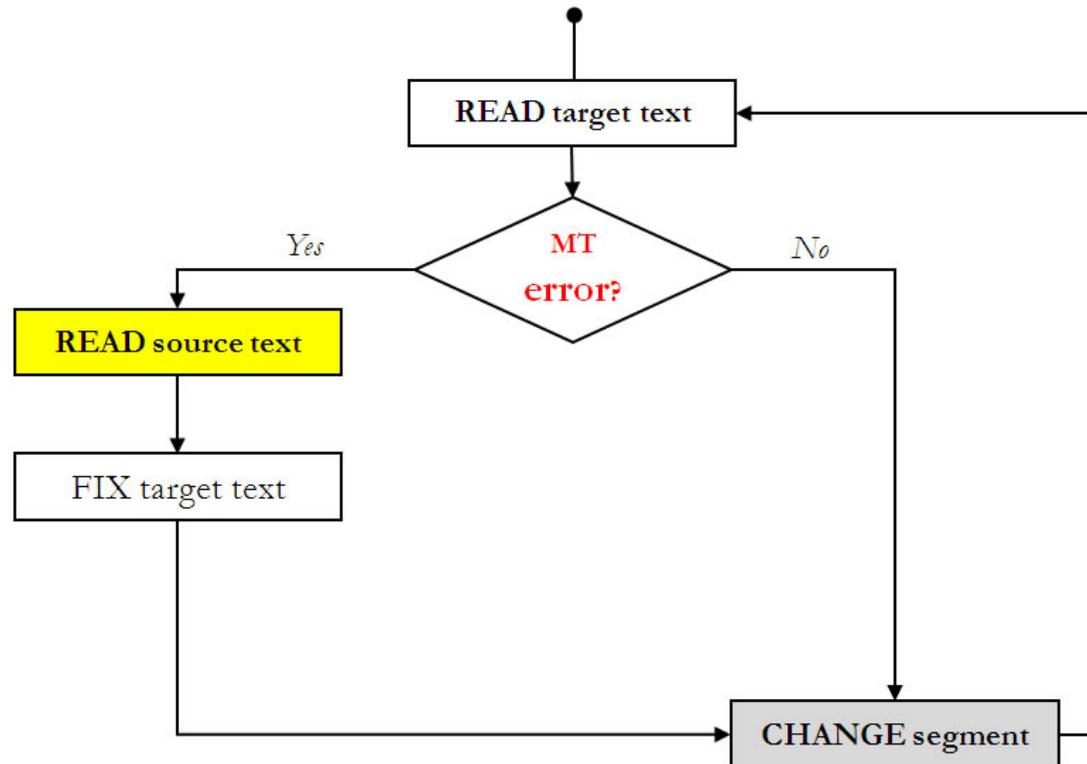


Logging and Eye Tracking



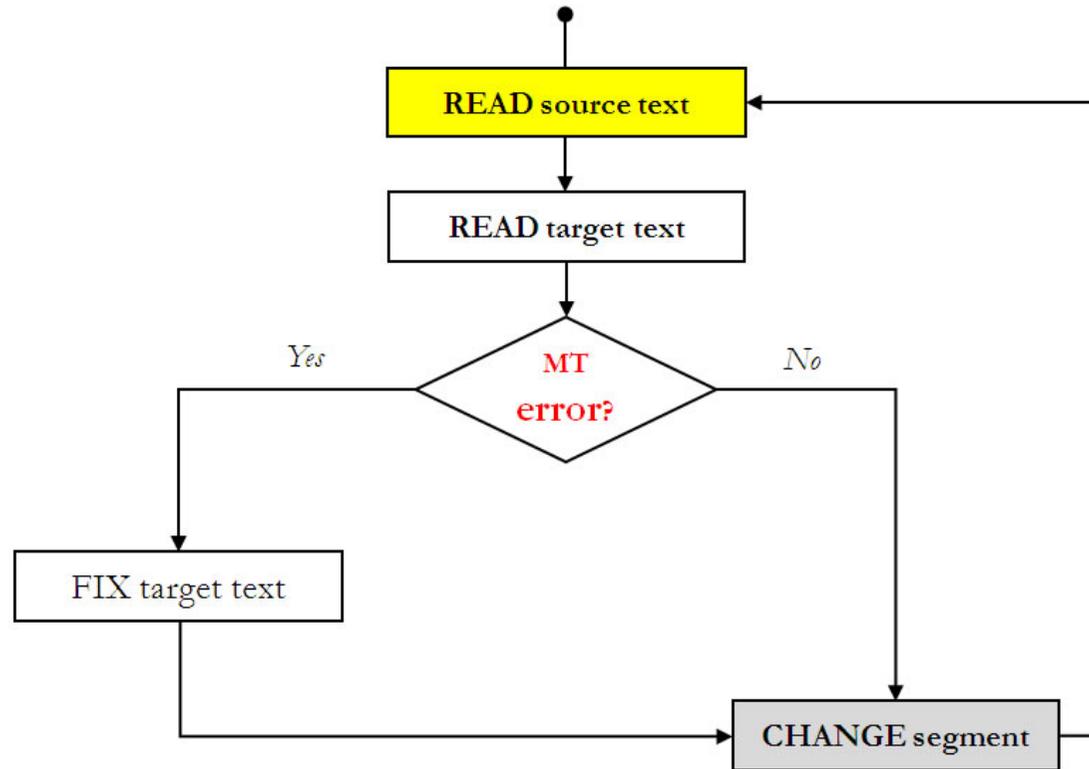
focus on **target word (green)** or **source word (blue)** at position x

Cognitive Studies: User Styles



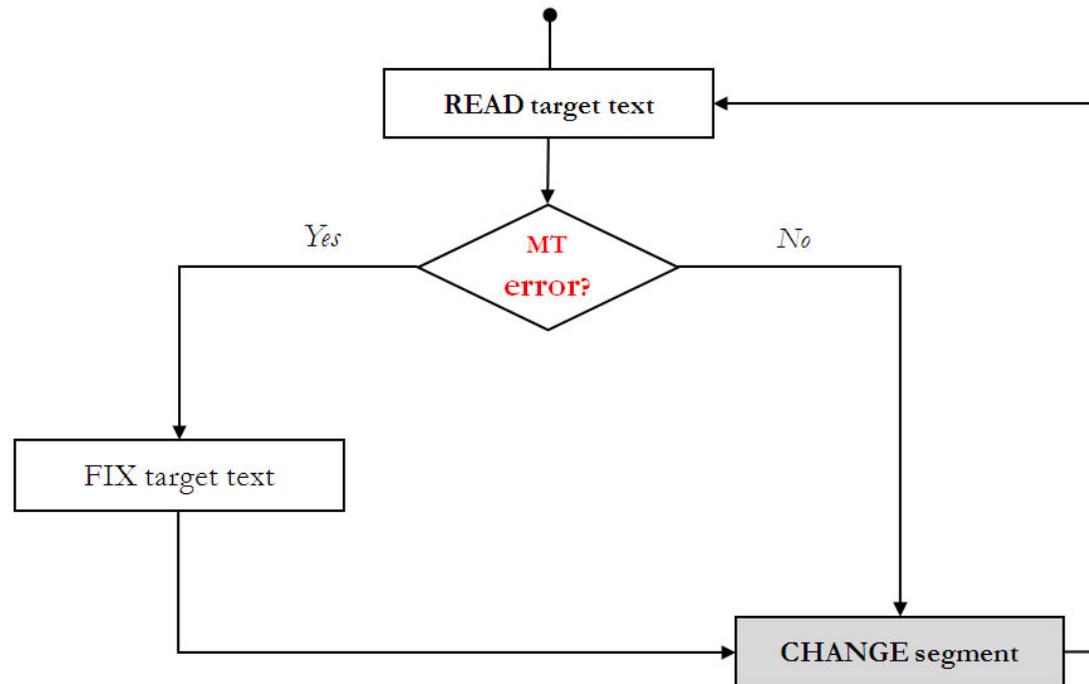
- User style 1: Verifies translation just based on the target text, reads source text to fix it

Cognitive Studies: User Styles



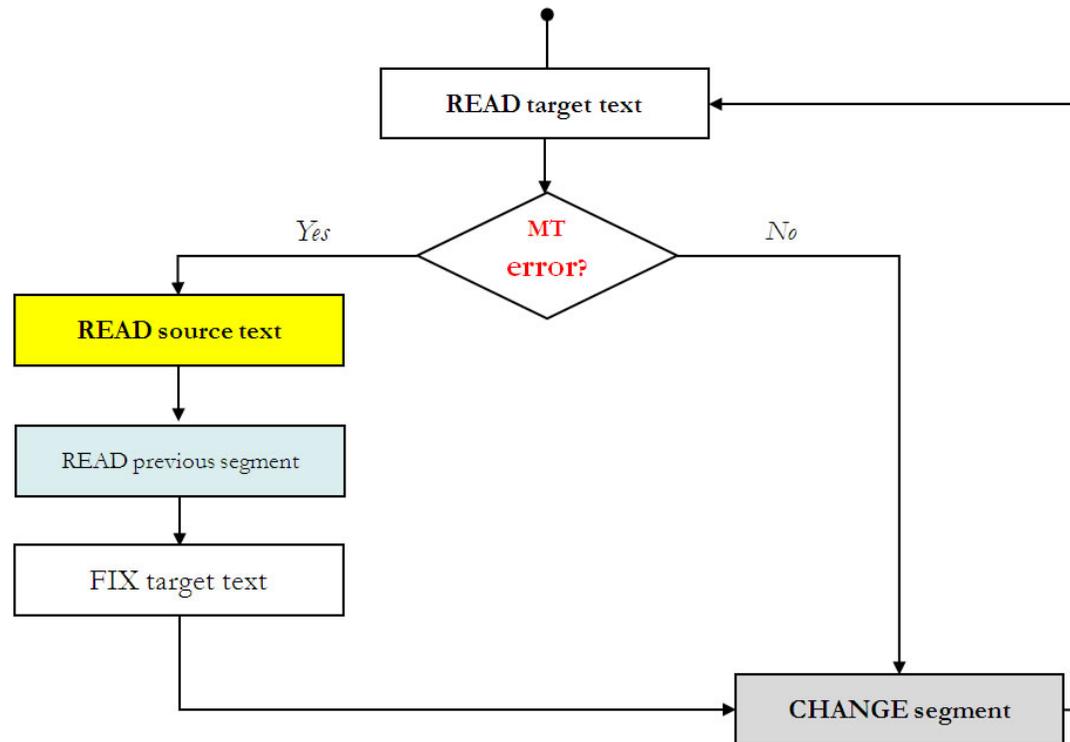
- User style 2: Reads source text first, then target text

Cognitive Studies: User Styles



- User style 3: Makes corrections based on target text only

Cognitive Studies: User Styles



- User style 4: As style 1, but also considers previous segment for corrections

Users and User Styles

	Style 1			Style 2			Style 3			Style 4		
	target / source-fix			source-target			target only			wider context		
	P	PI	PIA	P	PI	PIA	P	PI	PIA	P	PI	PIA
P02	*	*	*	●	●	●	●			●	●	●
P03												
P04	●	*	*				*	●	●	●	●	●
P05	●	●	●				*	*	*	●	●	●
P07	*	*	*				●	●	●	●	●	●
P08	*	*	*	●	●	●				●	●	●
P09	●	●	●				*	*	*	●	●	●

- Individual users employ different user styles
- But: consistently across different types of assistance
(P = post-editing, PI = interactive post-editing, PIA = interactive post-editing with additional annotations)



- Local backtracking
 - **Immediate repetition:** the user immediately returns to the same segment (e.g. AAAA)
 - **Local alternation:** user switches between adjacent segments, often singly (e.g. ABAB) but also for longer stretches (e.g. ABC-ABC).
 - **Local orientation:** very brief reading of a number of segments, then returning to each one and editing them (e.g. ABCDE-ABCDE).
- Long-distance backtracking
 - **Long-distance alternation:** user switches between the current segment and different previous segments (e.g. JCJDJFJG)
 - **Text final backtracking:** user backtracks to specific segments after having edited all the segments at least once
 - **In-text long distance backtracking:** instances of long distance backtracking as the user proceeds in order through the text.

Thank You

105



questions?