Computer Aided Translation
Advances and Challenges

Philipp Koehn

7 August 2016

Overview

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

:

part I

CASMACAT workbench

CASMACAT workbench

CASMACAT Project 2011-2014

- Cognitive studies of translators leading to insights into interface design

  → better understanding of translator needs

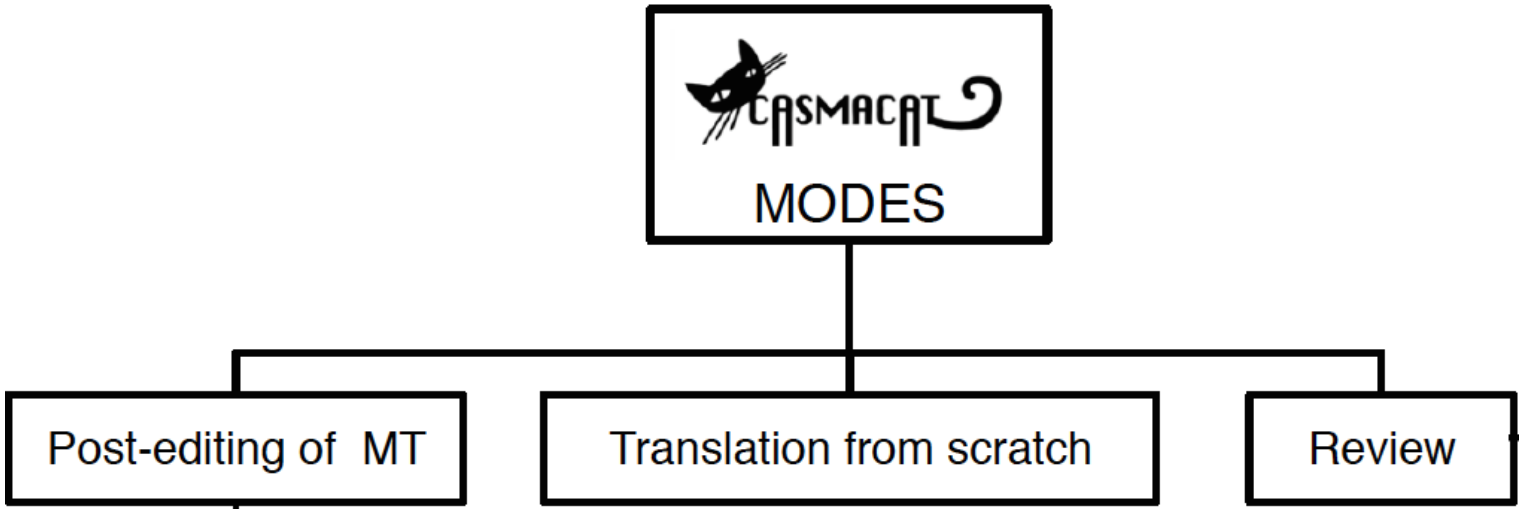
- Workbench with novel types of assistance to human translators
  - interactive translation prediction
  - interactive editing and reviewing
  - adaptive translation models

  → better tools for translators

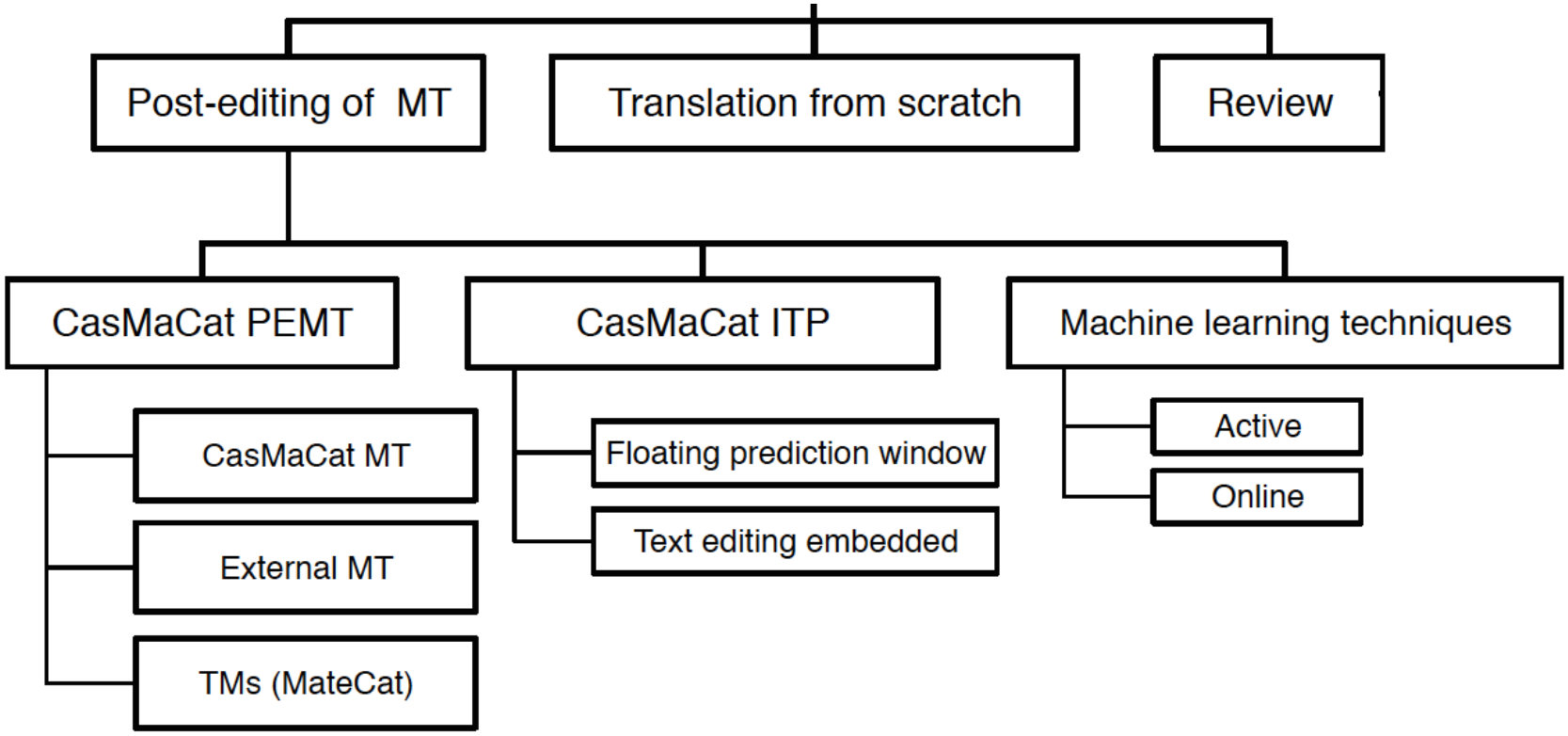
- Demonstration of effectiveness in field tests with professional translators

  → increased translator productivity

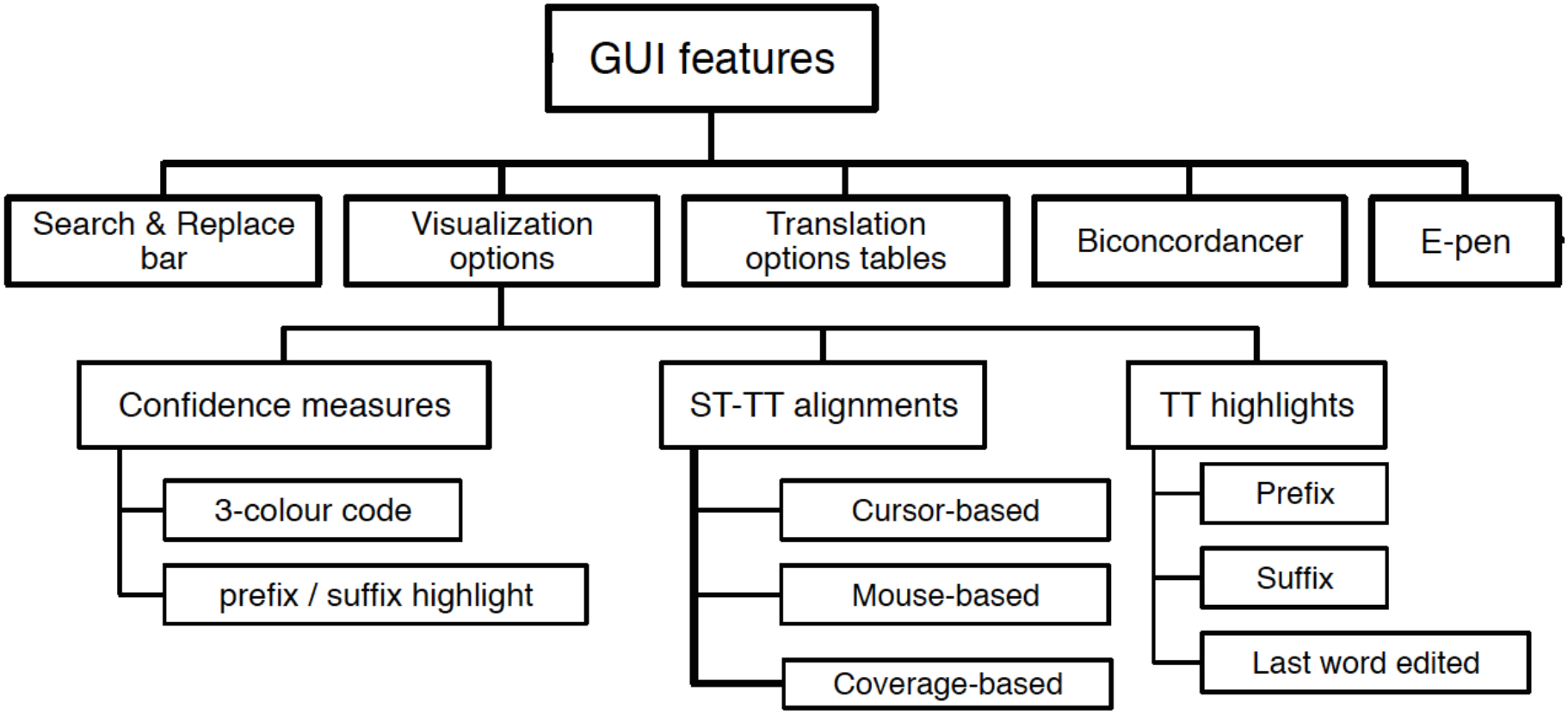
Core Modes


CASMACAT
MODES
Post-editing of MT
Translation from scratch
Review

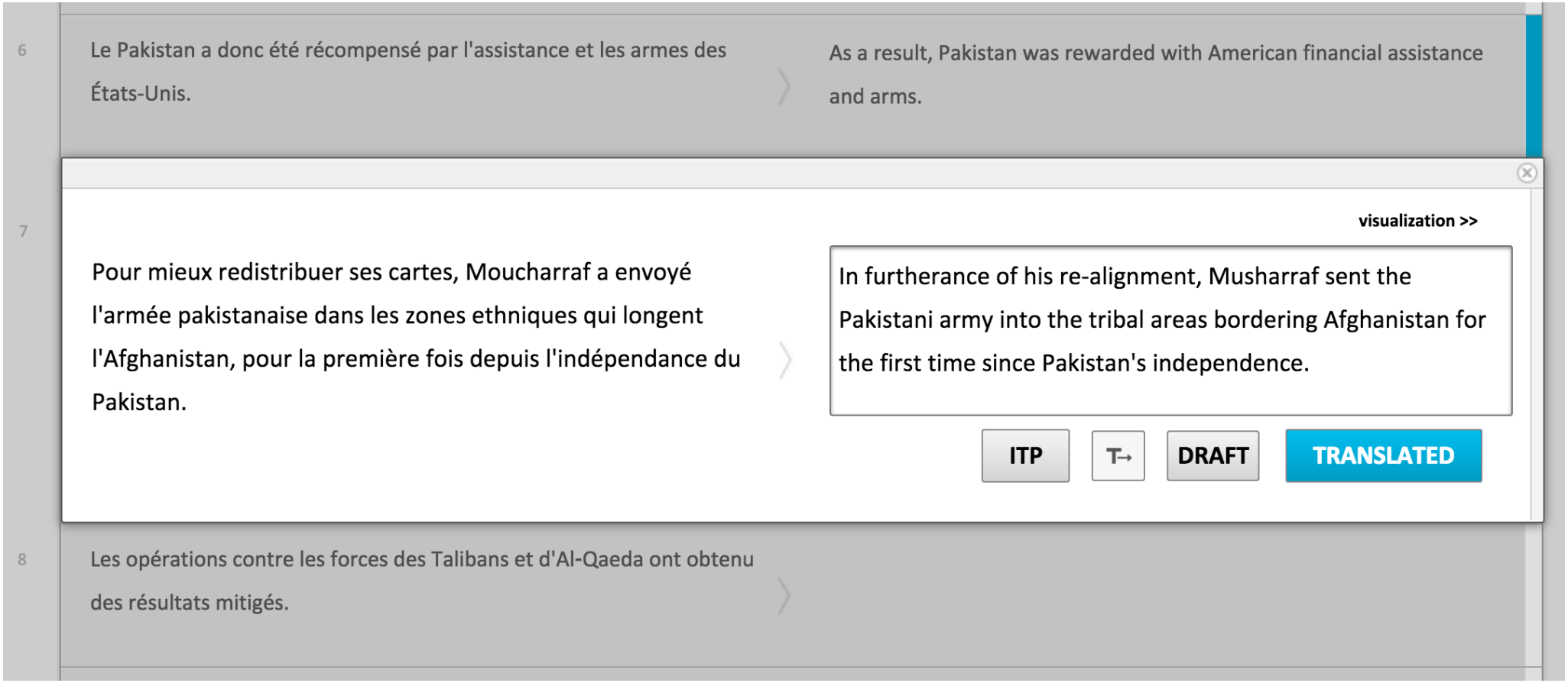
Postediting Modes



GUI Features

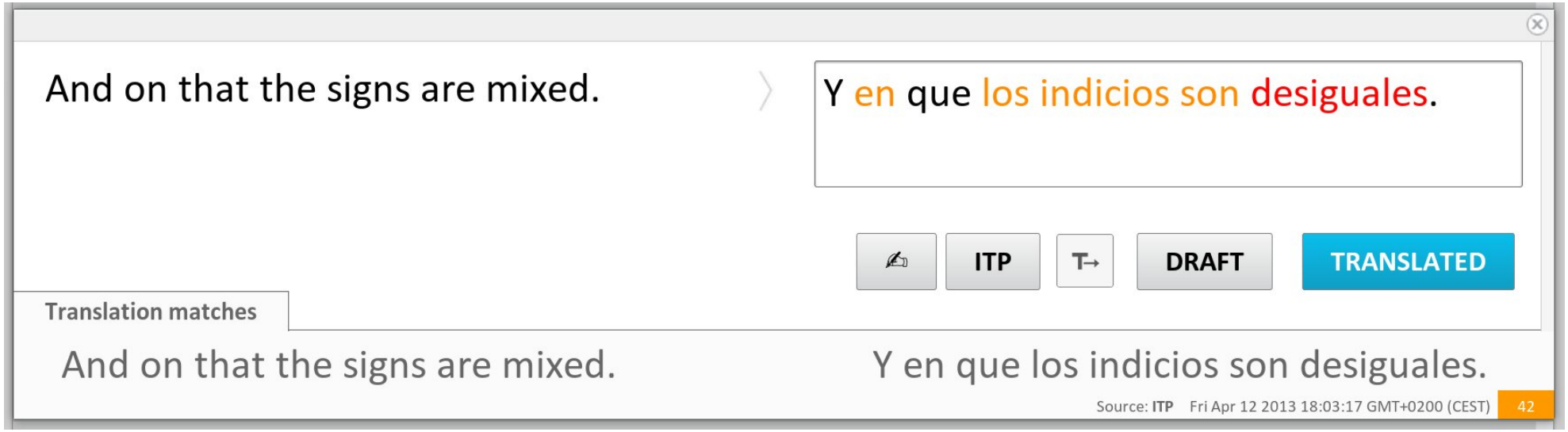


Postediting Interface



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

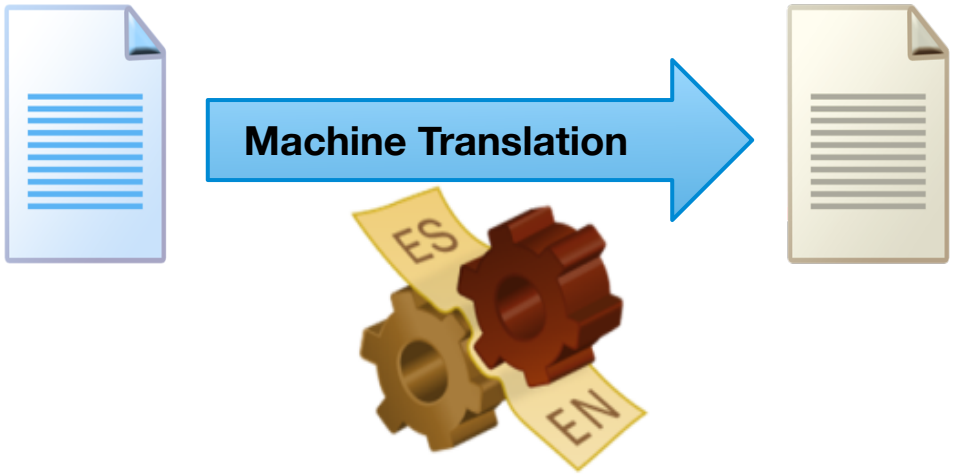
Confidence Measures



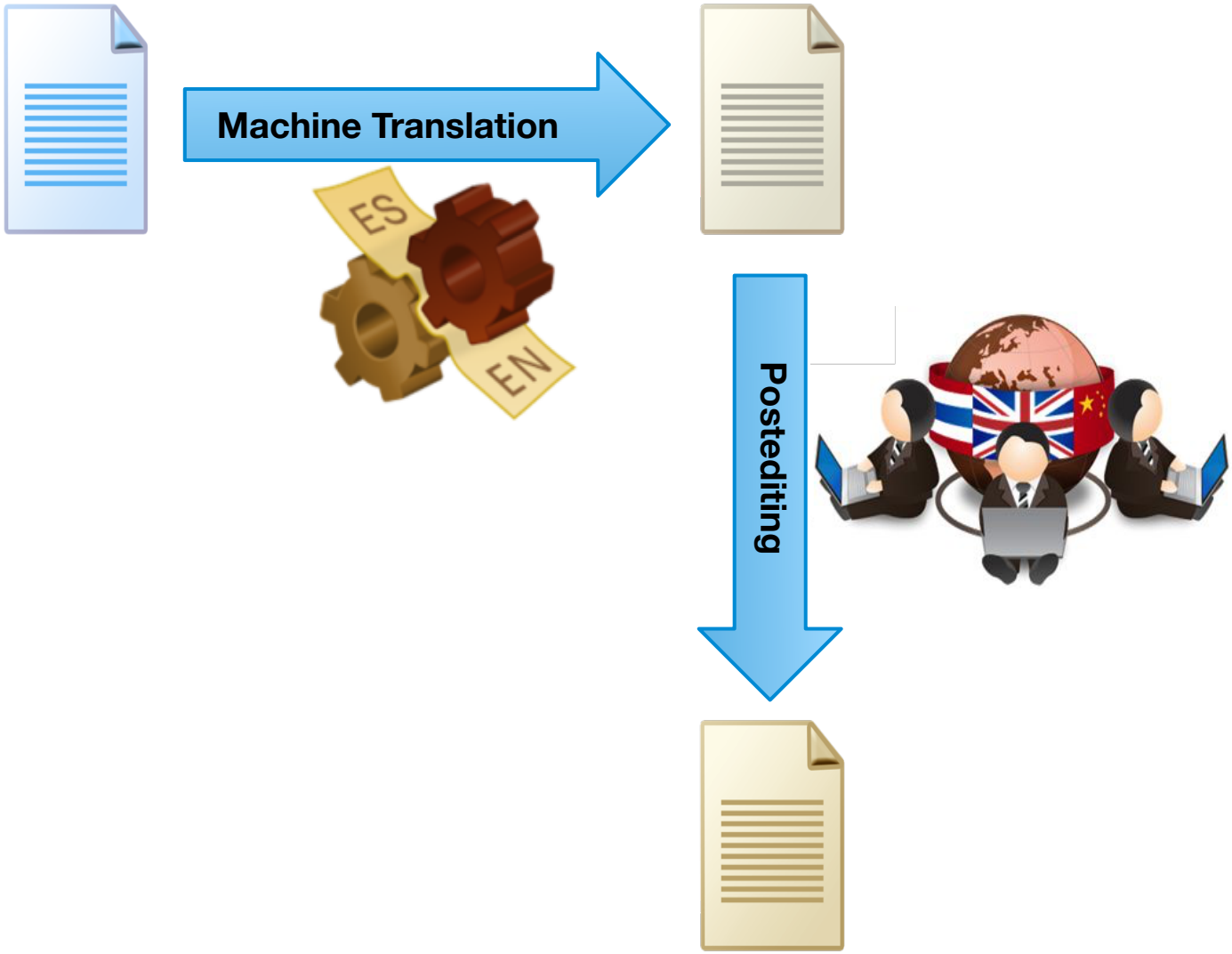
- Sentence-level confidence measures
  → estimate usefulness of machine translation output

- Word-level confidence measures
  → point posteditor to words that need to be changed

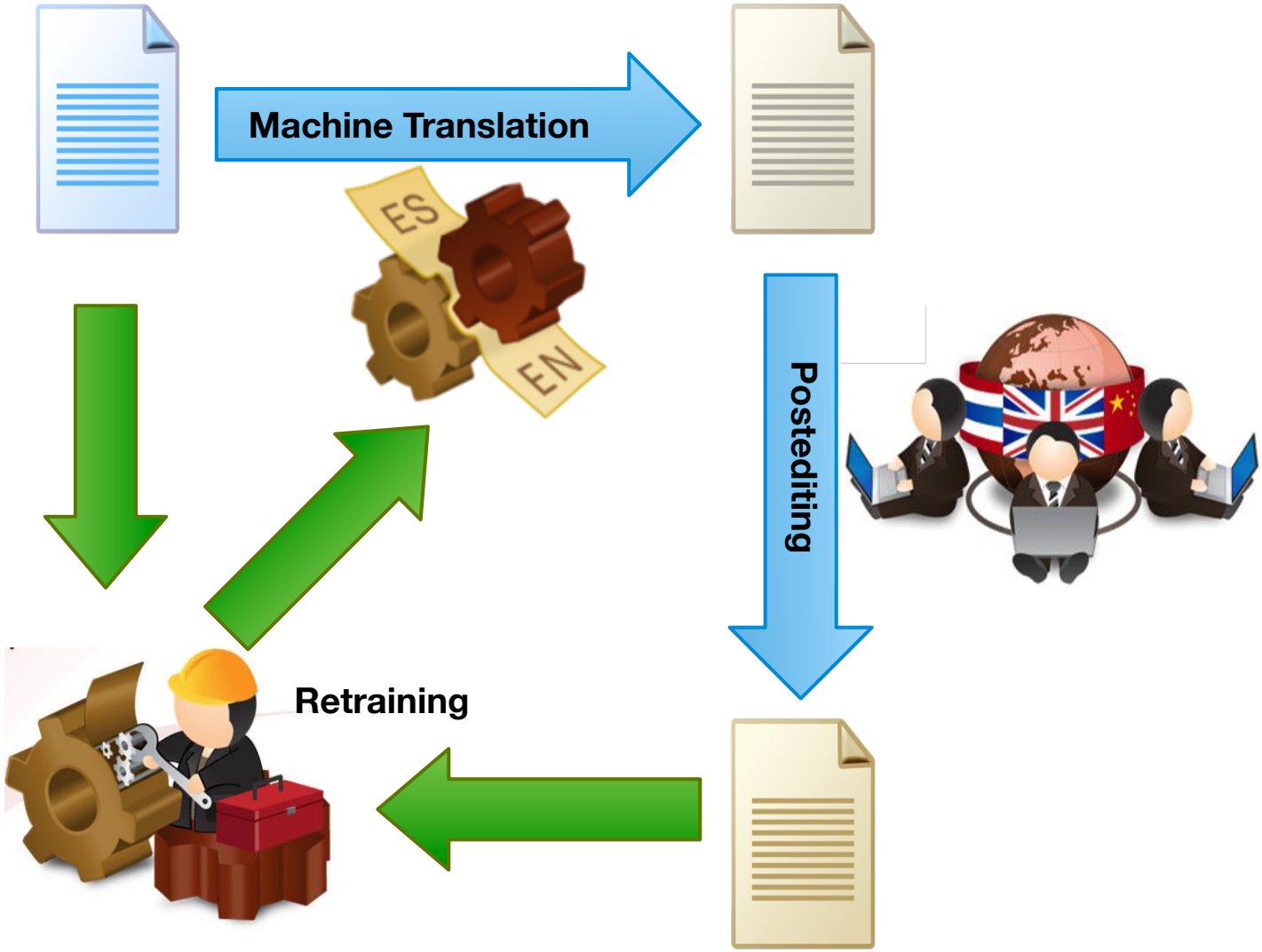
Incremental Updating



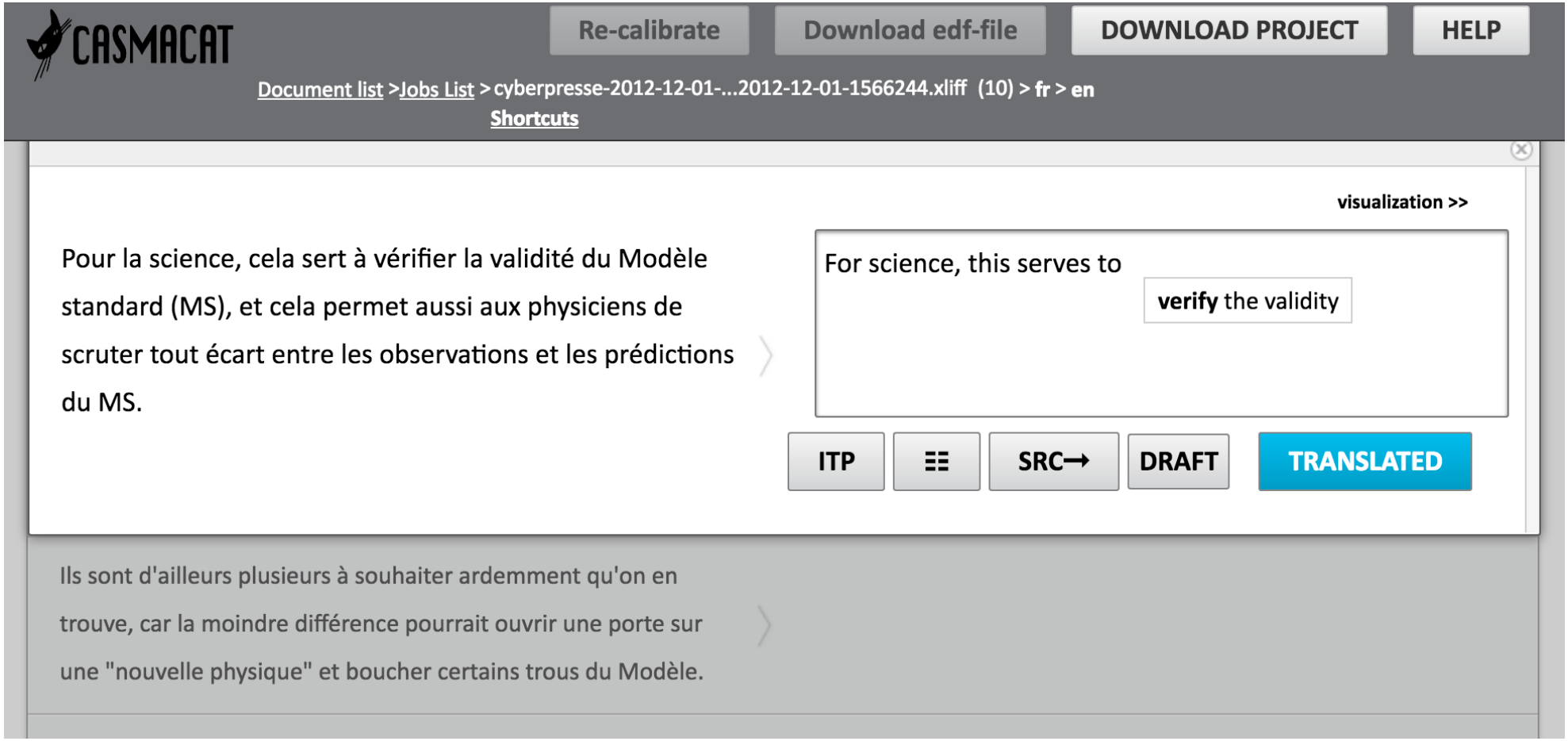
Incremental Updating



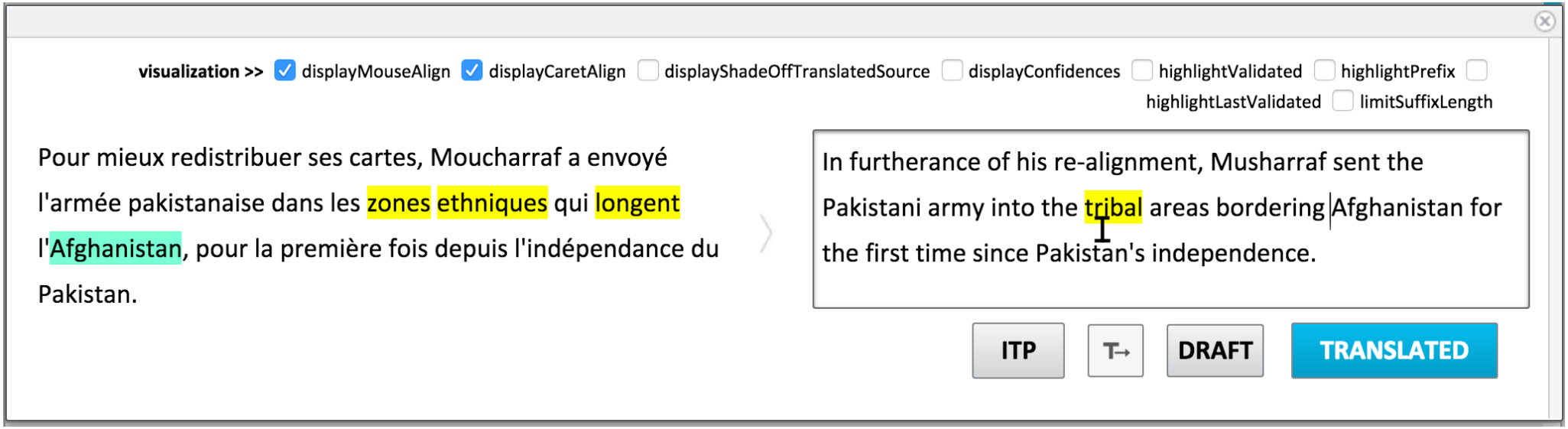
Incremental Updating



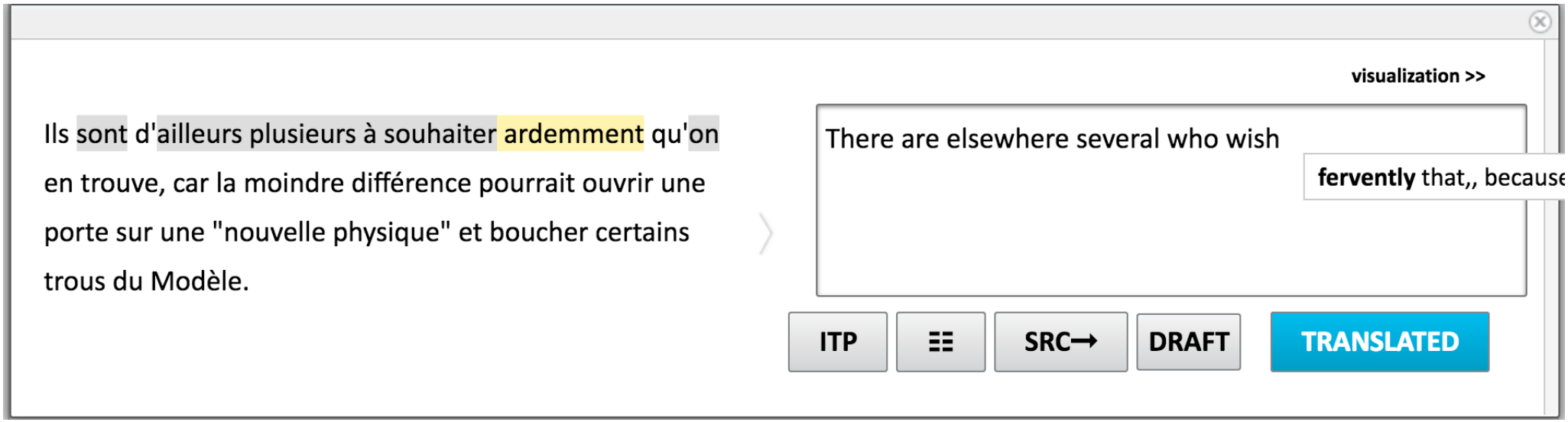
Interactive Translation Prediction



Word Alignment

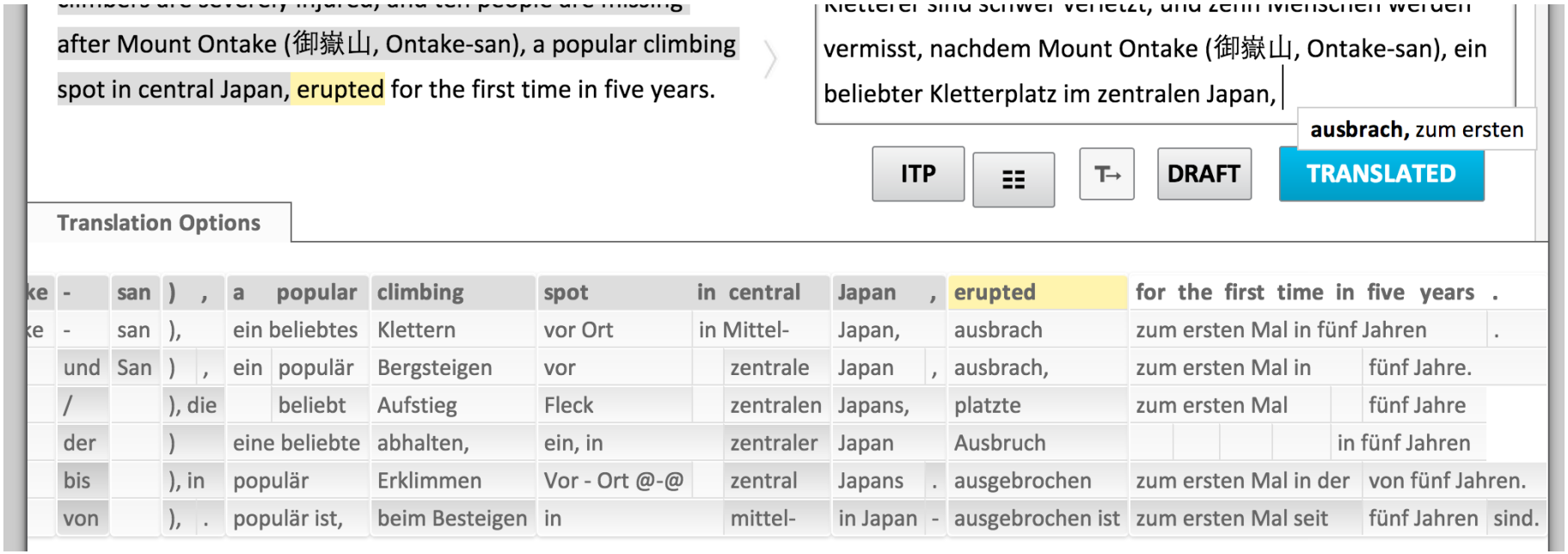


Word Alignment



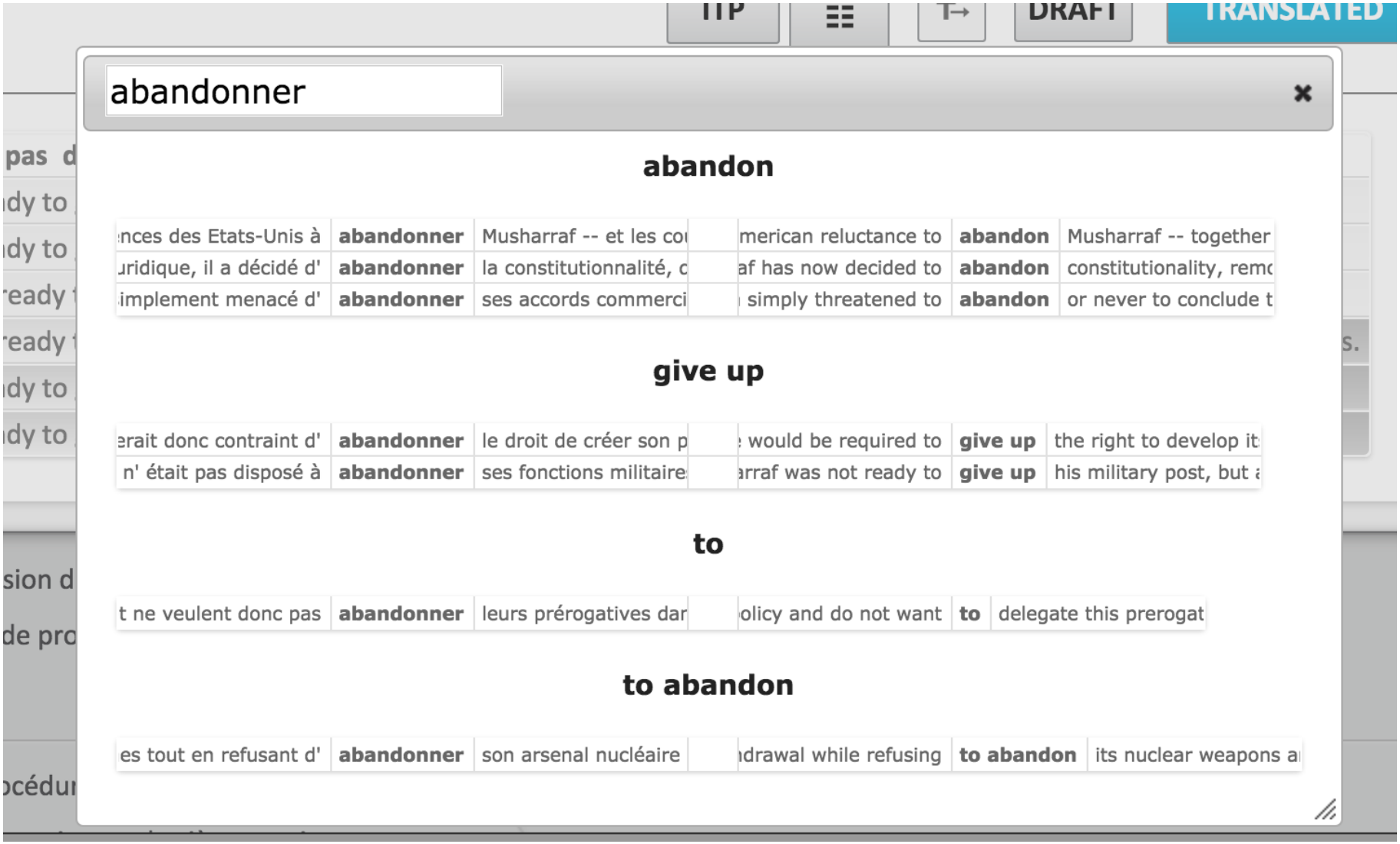
- With interactive translation prediction
- Shade off translated words, highlight next word to translate

Translation Option Array

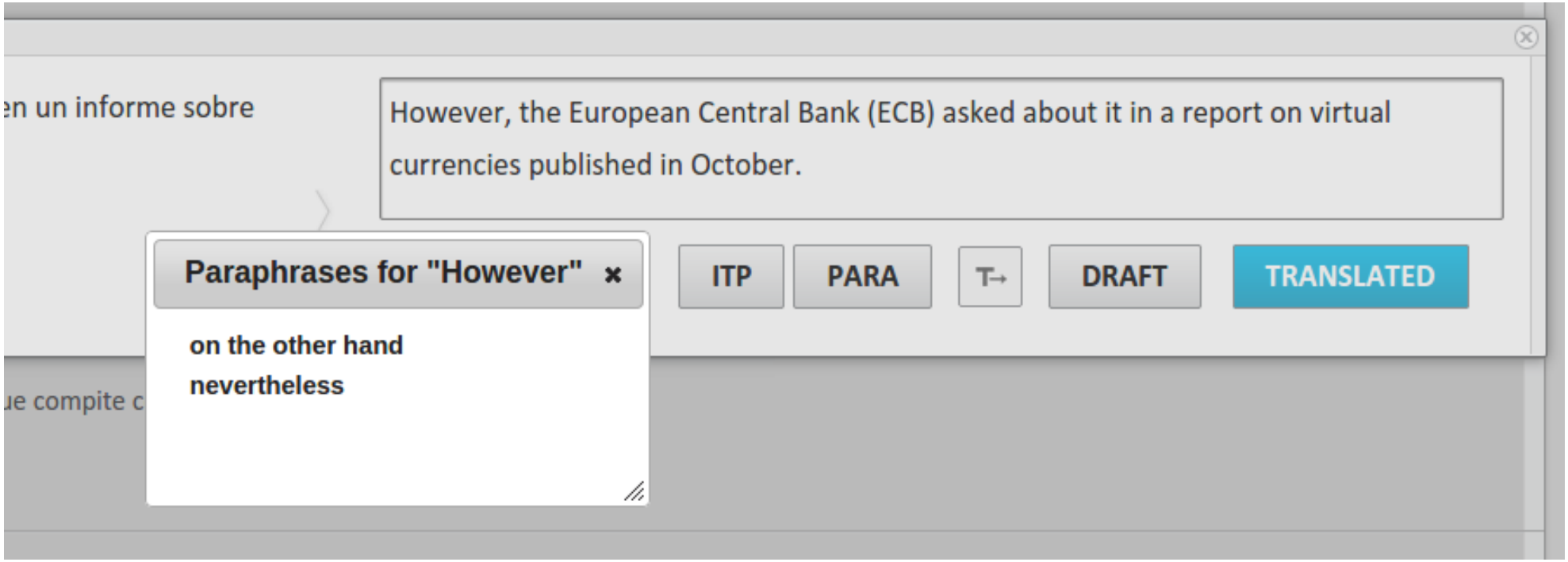


- Visual aid: non-intrusive provision of cues to the translator
- Clickable: click on target phrase → added to edit area
- Automatic orientation
  - most relevant is next word to be translated
  - automatic centering on next word

Bilingual Concordancer



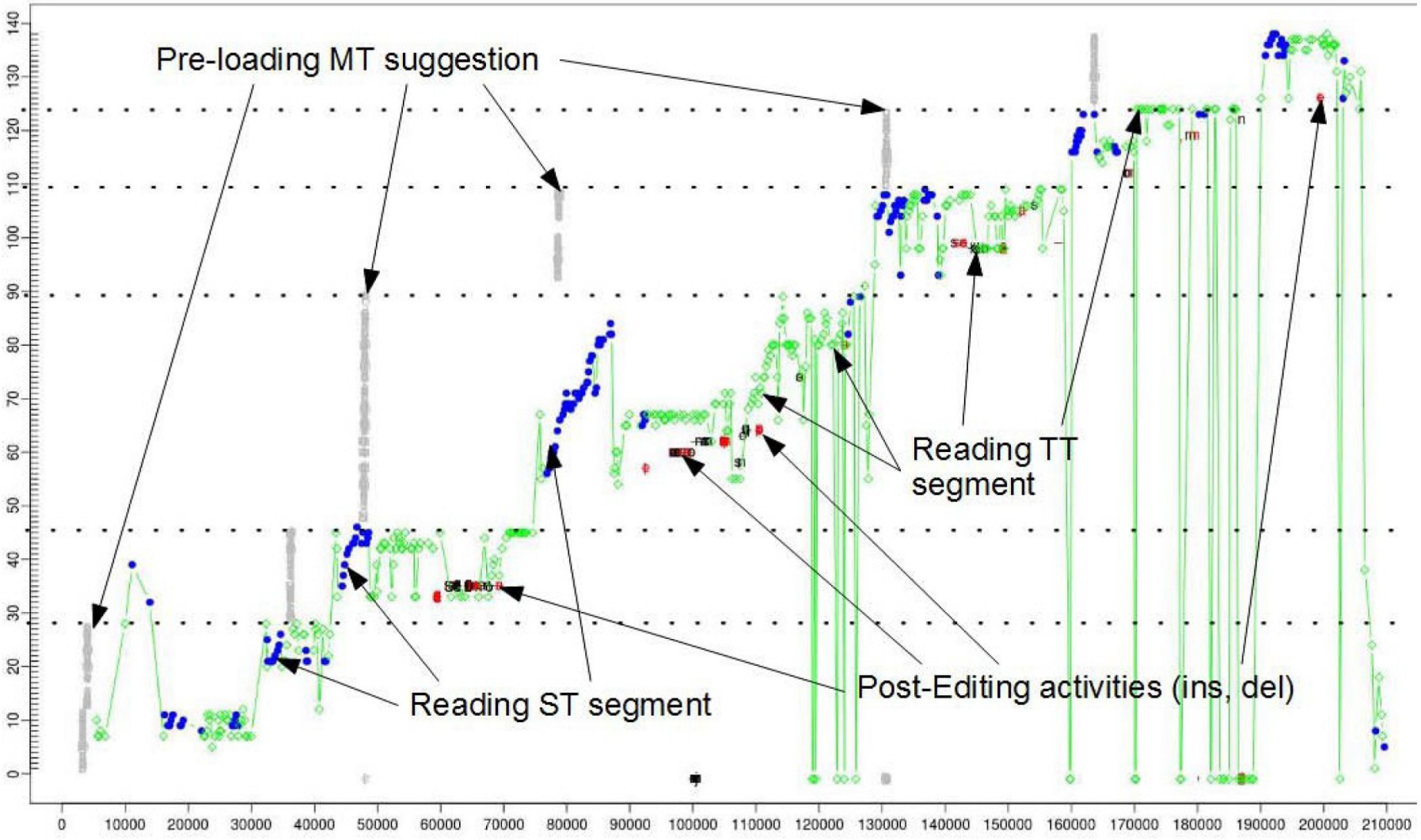
Paraphrasing



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?

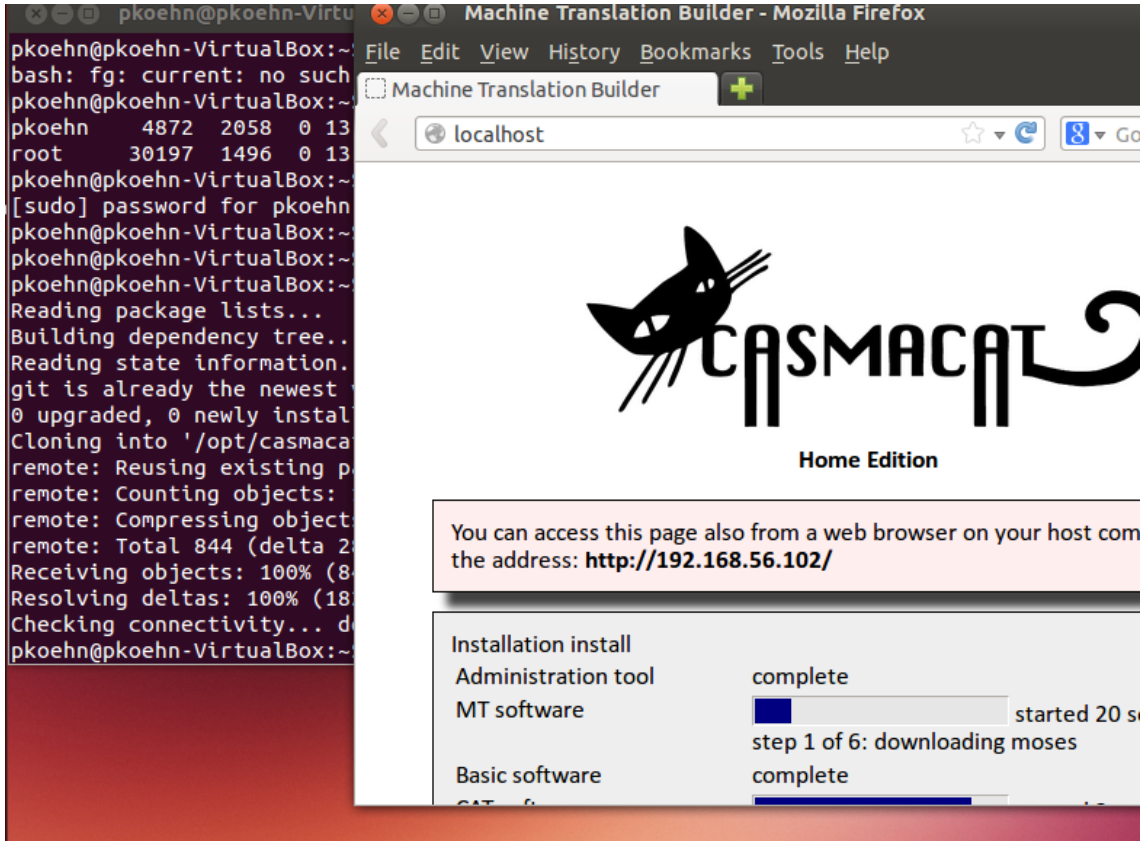
Logging and Eye Tracking



Home Edition

- 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



Administration through Web Browser

Administration

Translate

- Translate new document
- List documents

Engines

- Manage engines
- Upload engine
- Build new prototype

Settings

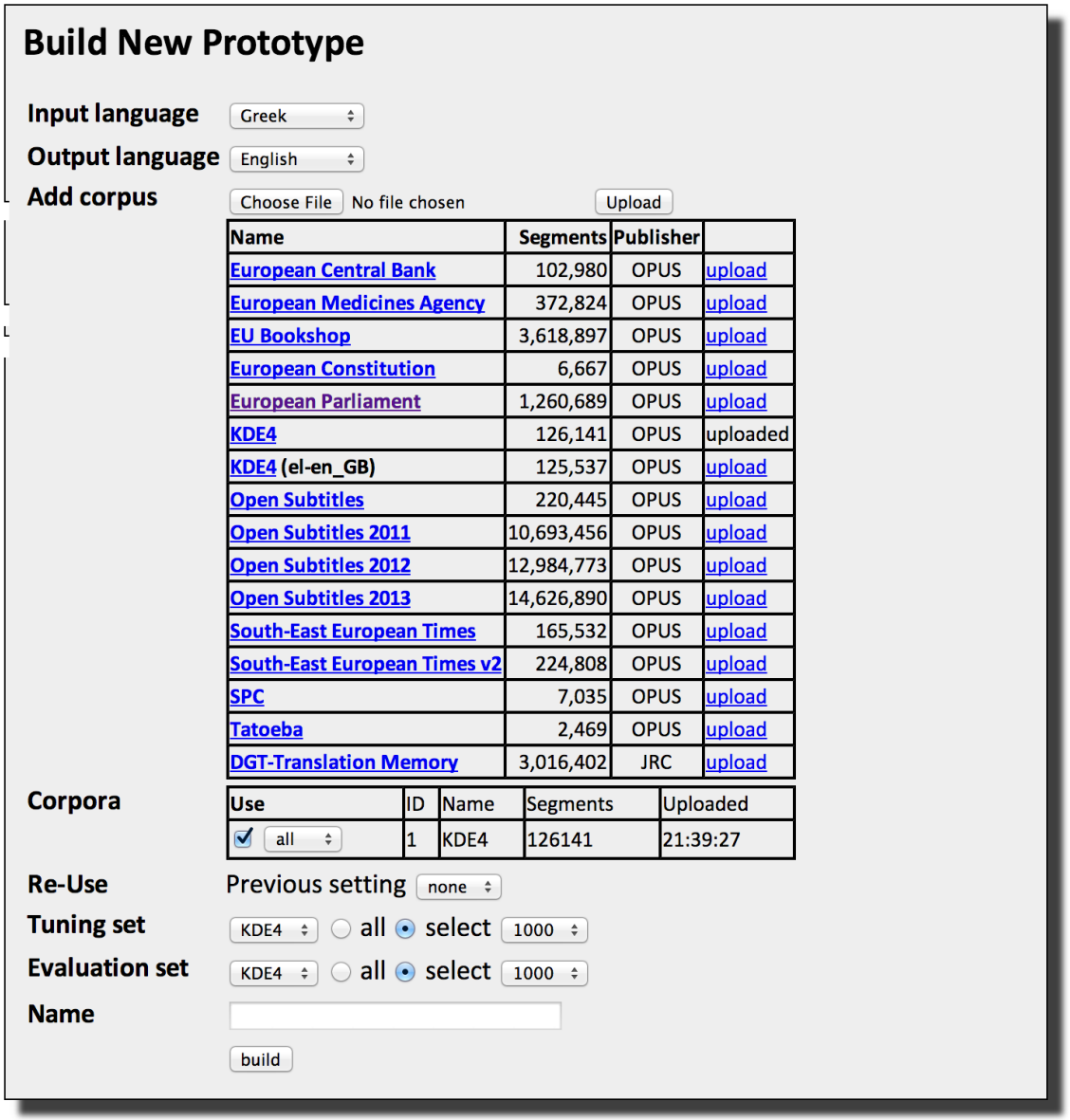
- Reset CAT and MT server
- CAT Settings
- Update Software

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 Greek

Output language English

Add corpus Choose File No file chosen Upload

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
☑ all	1	KDE4	126141	21:39:27

Re-Use Previous setting none

Tuning set KDE4 ○ all ◉ select 1000

Evaluation set KDE4 ○ all ◉ select 1000

Name

build

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

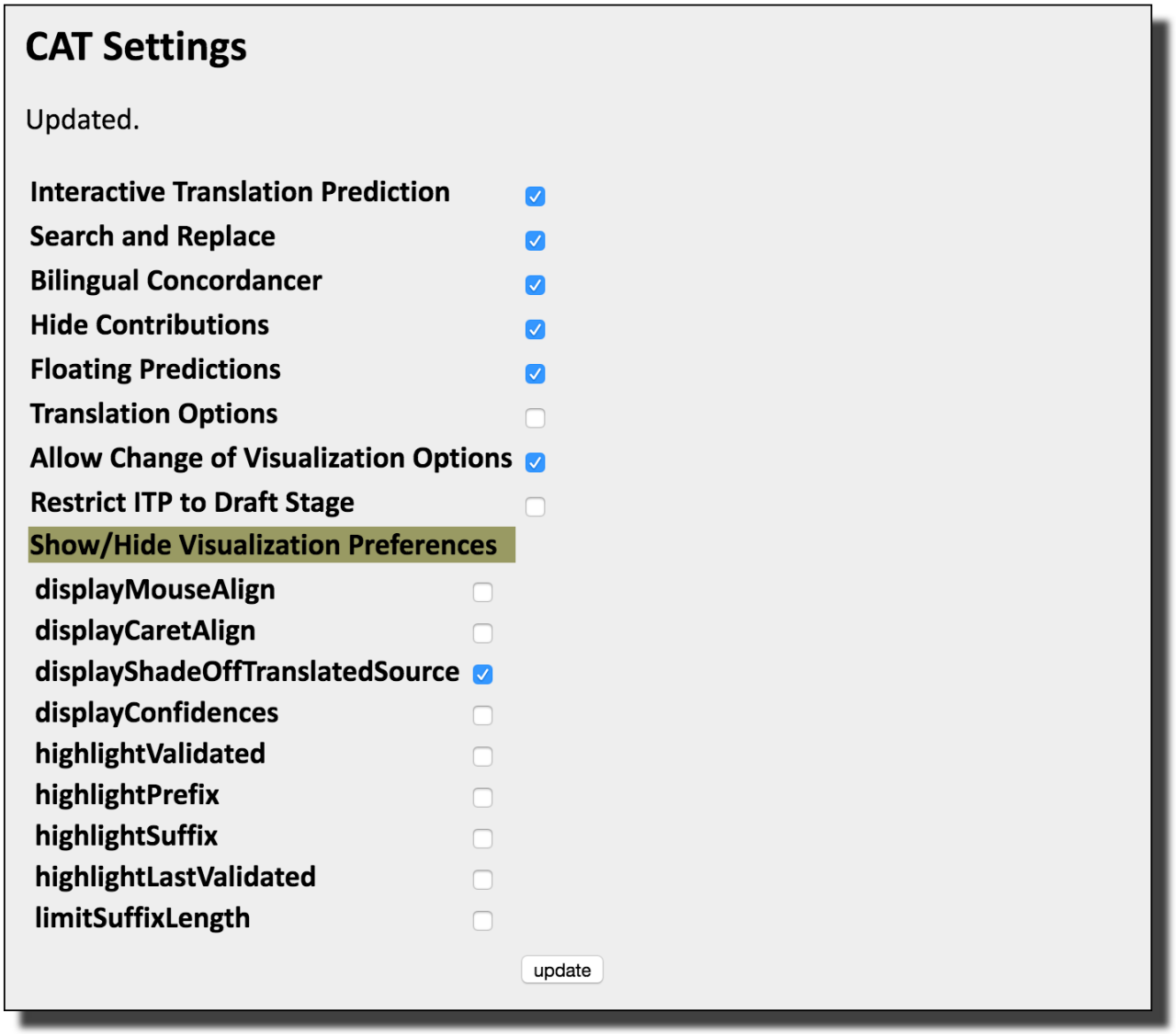
Available Engines

#	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

CAT Settings

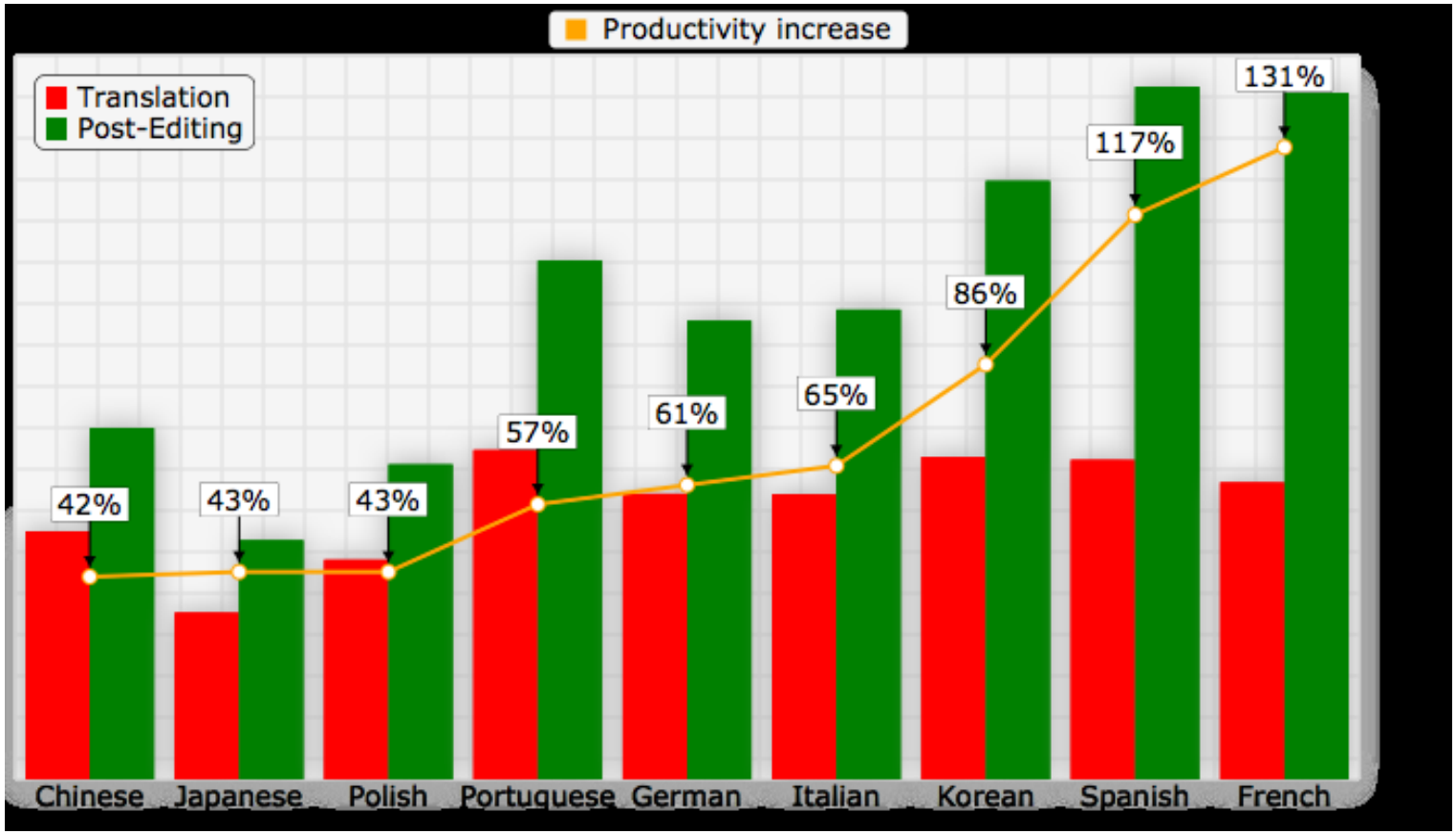


part II

cat methods

post-editing

Productivity Improvements



(source: Autodesk)

MT Quality and Postediting Effort

- Postediting effort = number of words changed
- Evaluation metric at IWSLT 2014
  - TER = automatic metric, comparison against a reference translation
  - HTER = postediting metric, actual words changed

English–German

Ranking	HTER	TER
EU-BRIDGE	19.2	54.6
UEDIN	19.9	56.3
KIT	20.9	54.9
NTT-NAIST	21.3	54.7
KLE	28.8	59.7

English–French

Ranking	HTER	TER
EU-BRIDGE	16.5	42.6
RWTH	16.6	41.8
KIT	17.6	42.3
UEDIN	17.2	43.3
MITLL-AFRL	18.7	43.5
FBK	22.3	44.3
MIRACL	32.9	52.2

Translator Variability

- Professional translators

English–German

Posteditor	HTER	TER
PE 1	32.2	56.1
PE 2	19.7	56.3
PE 3	40.9	56.2
PE 4	27.6	55.9
PE 5	25.0	55.6

English–French

Posteditor	HTER	TER
PE 1	35.0	42.6
PE 2	17.5	42.8
PE 3	23.7	43.0
PE 4	39.7	42.3
PE 5	19.7	42.9

- Also very high variability

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

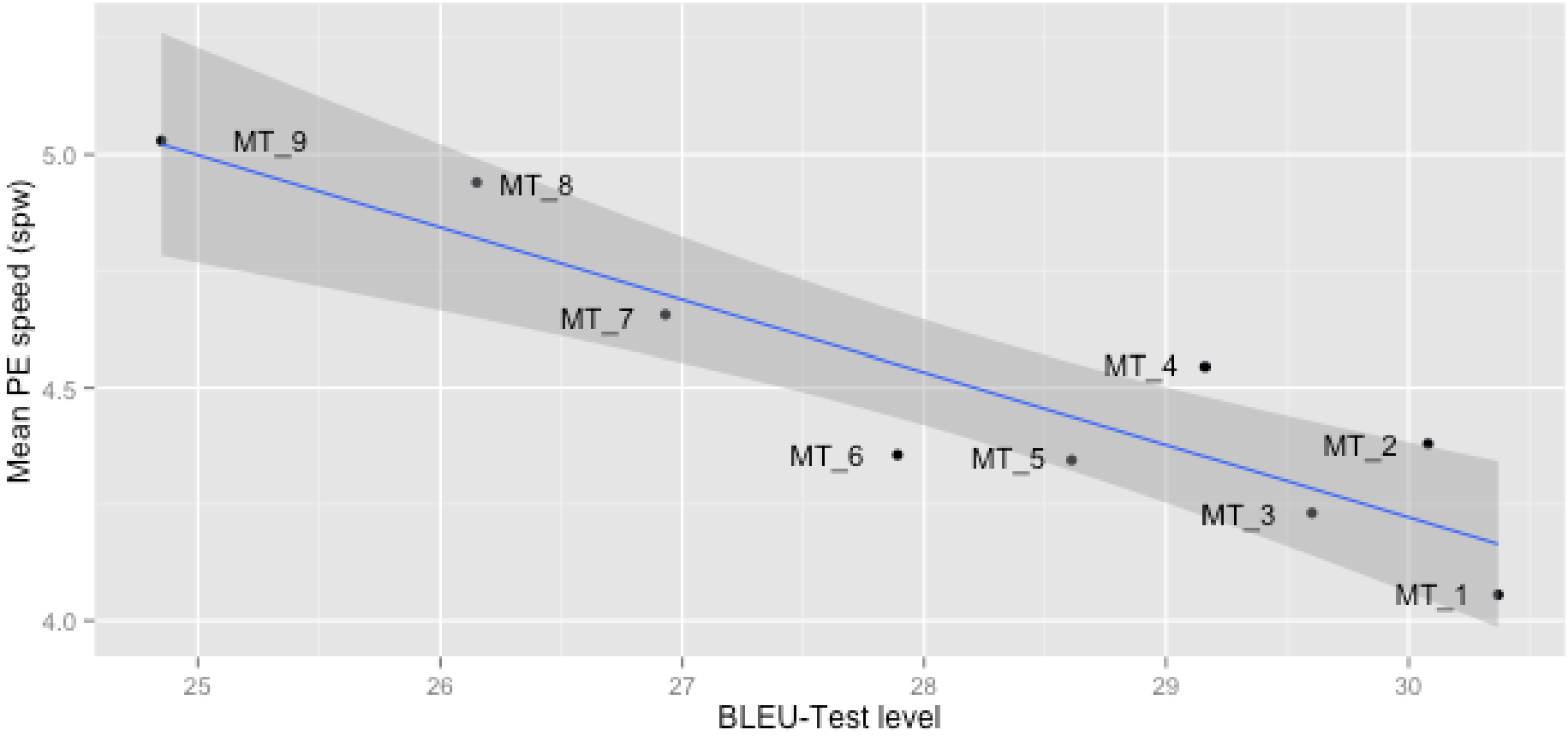
- Same type of system (Spanish–English, phrase-based, Moses)
- Trained on varying amounts of data

MT Quality and Productivity

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

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

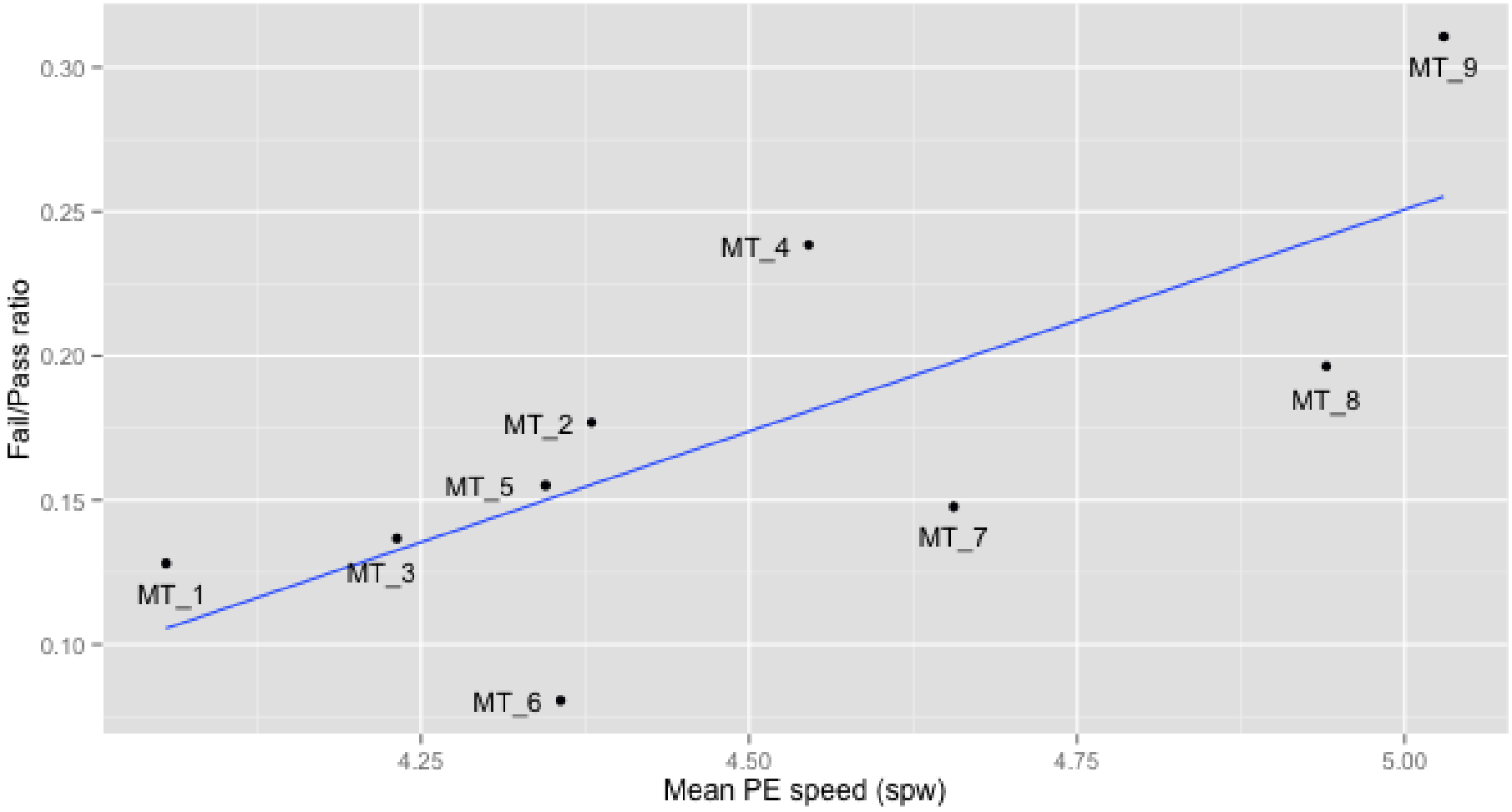
MT Quality and Productivity



BLEU against PE speed and regression line with 95% confidence bounds

+1 BLEU $\leftrightarrow$ decrease in PE time of $\sim$0.16 sec/word

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

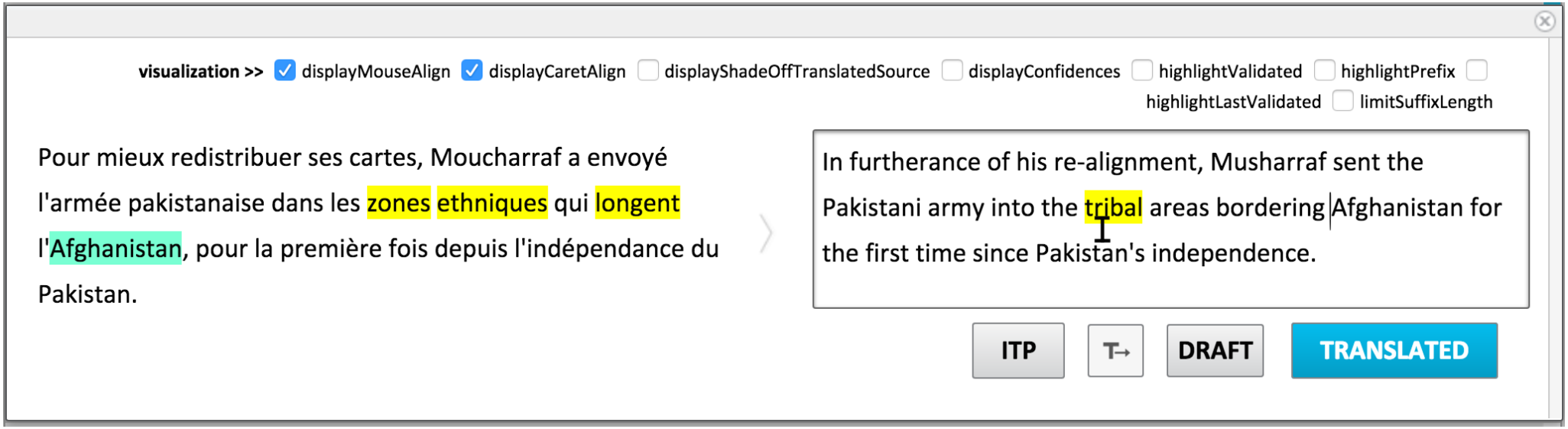
- Higher variability between translators than between MT systems

Postediting and MT Metrics

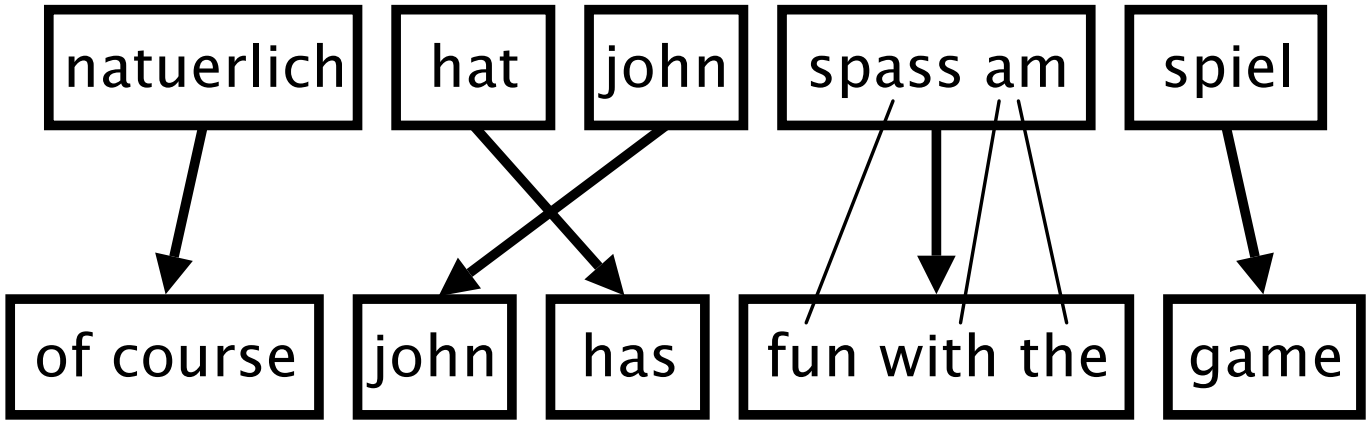
- Goal of MT quality metrics not clear
  - understandability: do you get the meaning?
  - post-editing effort: how much effort to change?
- Example: dropping of the word "not"
  - understandability: big mistake
  - post-editing effort: quick add of just one word
- Not clear, what tradition manual metrics prefer (adequacy, fluency)
- Not clear, what BLEU score etc. prefer

word alignment

Word Alignment



Word Alignment from MT



- Machine translation output is constructed by phrase mappings
- Each phrase mapping has internal word alignment

$\Rightarrow$ This can be used to visualize word alignments

- But: word alignment points become invalid after user edits

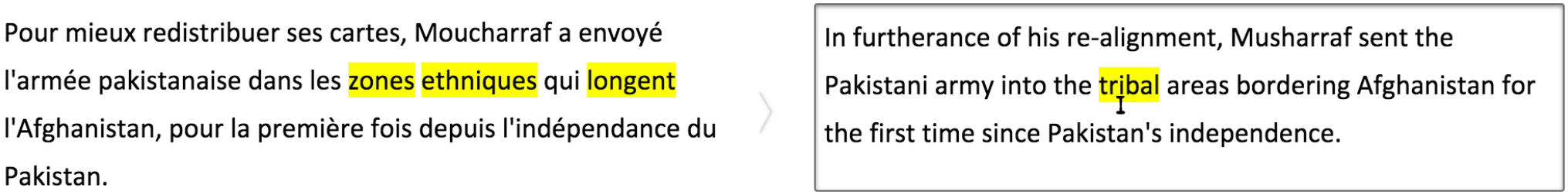
Word Alignment from Alignment Tools

- During machine translation training, standard component is word alignment

- Standard tools
  - old workhorse: GIZA++
  - currently popular tool: fast-align

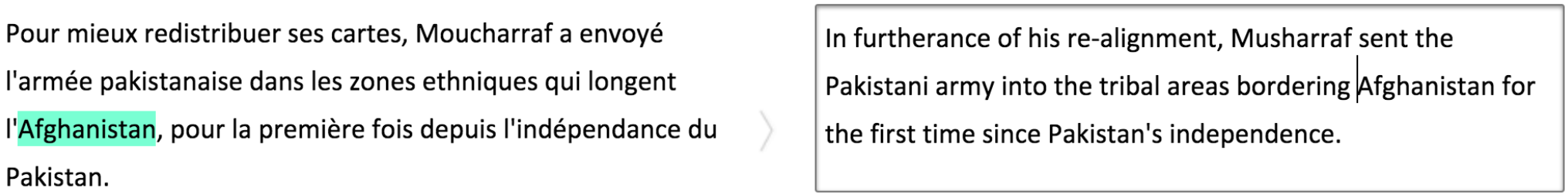
- These tools have been adapted to align new sentence pairs

Mouse Over Alignment



- Highlight the source word aligned to the word at the current **mouse** position

Caret Alignment



- Highlight the source word aligned to the word at the current **caret** position

Shade Off Translated

- Use in interactive prediction mode
- Shade off words that are already translated
- Highlight words aligned to first predicted translation word

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

  $\rightarrow$ 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)

QuEst

- 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

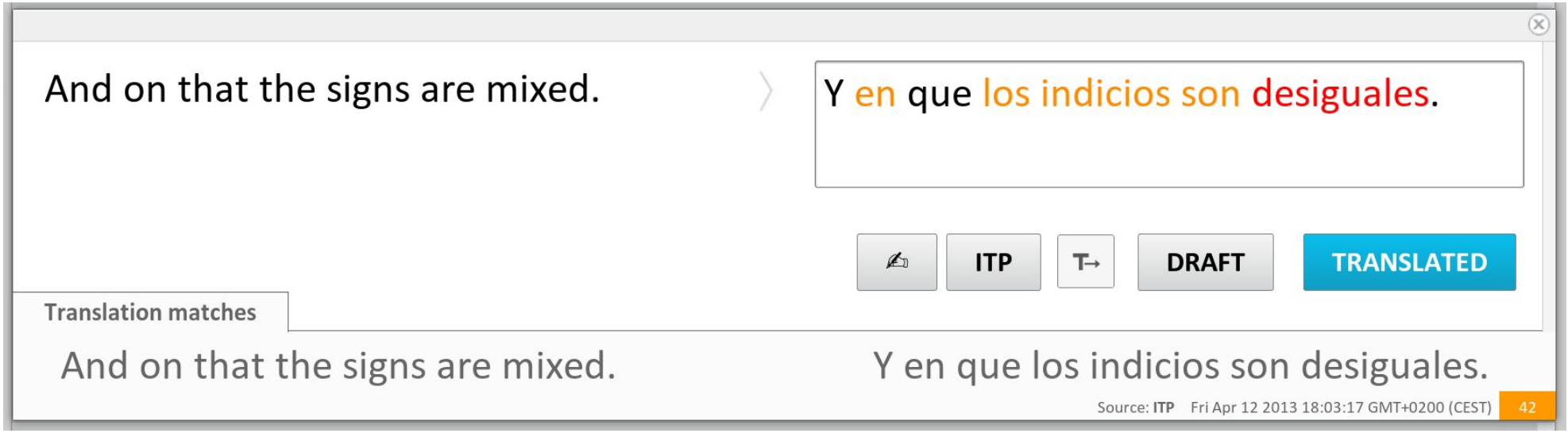
- Yandex School of Data Analysis (Kozlova et al., 2016)

- 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

Visualization



- Highlight words less likely to be correct

Methods

- 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

Fast	*brown*	*fox*	*jumps*	*on*	*the*	*dog*	*lazy*	*.*
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

automatic reviewing

Automatic Reviewing

- Can we identify errors in human translations?
  - missing / added information
  - inconsistent use of terminology

Input Sentence

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

Human Translation

Moreover, he planned for months to give a lecture in Miami.

Reviewing with E-Pen

- Intuition
  - reviewing more efficient with pen and paper
  - e-pen enables this work process in digital environment
- Work carried out
  - fronted modified for larger drawing area
  - backend support for hand-written text recognition (HTR)
  - development of methods for HTR
- Field trial carried out $\rightarrow$ corpus of reviewing edits

Analysis of Reviewer Edits

- 171 insertions — vast majority function words
- 152 deletions — about half substantial content
- 621 replacements — of which:
  - 75 changes to punctuation only
  - 28 change to lowercase / uppercase
  - 29 cases that are mostly deletions
  - 8 cases that are mostly insertions
  - 289 morphological/spelling changes (Levenshtein distance of less than 50%)
  - 190 other changes, about equal amounts function words and content words

Automatic Reviewing

- Focus on translation errors
  - not: basic spell checking
  - not: basic grammar checking
- Do not try the impossible
  - semantic errors
  - errors in function words
- What is left?
  - added content (insertions)
  - non-translated content (deletions)
  - inconsistency in terminology

Method

- Word alignment of human translation and source
- Detect unaligned words
  - insertion of content words:
    unaligned sequence of words in the draft translation
  - deletion of content words:
    unaligned sequence of words in the source sentence
  - inconsistent terminology:
    source word occurs multiple times, aligned to different word
- Only content words (minimum 4 characters)

Evaluation on Field Trial Data

- Two evaluation metrics
  - strict: predicted word X deleted / inserted
  - generous: predicted any deletion / insertion

	Strict Scoring		Generous Scoring		
Edit type	Precision	Recall	Precision	Recall	Baseline Precision
Deletion	7%	27%	11%	48%	7%
Insertion	-	-	5%	35%	4%
Any edit	-	-	20%	60%	14%

- Good enough to be useful?

Subjective Evaluation

- Evaluation on community translation platform data
- English–German
- Predict insertions and deletions
- Manually check if these are valid suggestions (i.e., precision only) by native German speaker

Results

- 4 cases of detection of valid errors (3 deletions, 1 inset ion)
- 31 false alarms

Count	Type
16 cases	unaligned verb
6 cases	one-to-many alignment
2 cases	non-literal
6 cases	misalignment, often due to unknown word
1 case	valid verb ellipsis, repeated in sub clause

- Good enough to be useful?

interactive translation prediction

Interactive Translation Prediction

Input Sentence

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

Professional Translator

|

Interactive Translation Prediction

Input Sentence

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

Professional Translator

| He

Interactive Translation Prediction

Input Sentence

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

Professional Translator

He | has

Interactive Translation Prediction

Input Sentence

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

Professional Translator

He has | for months

Interactive Translation Prediction

Input Sentence

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

Professional Translator

He planned |

Interactive Translation Prediction

Input Sentence

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

Professional Translator

He planned | for months

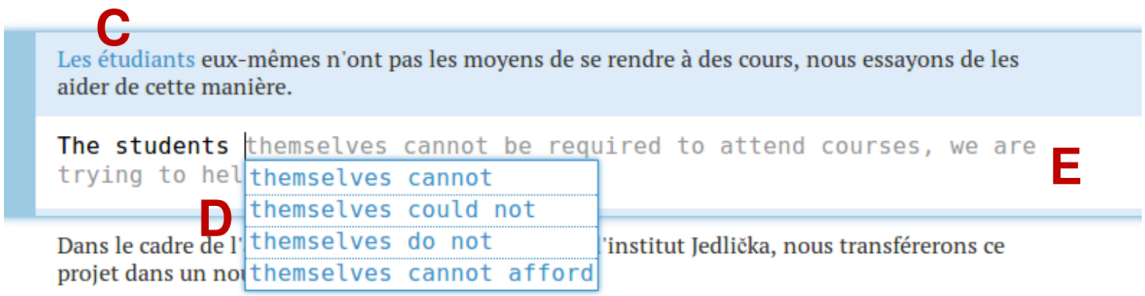
Visualization

- Show n next words

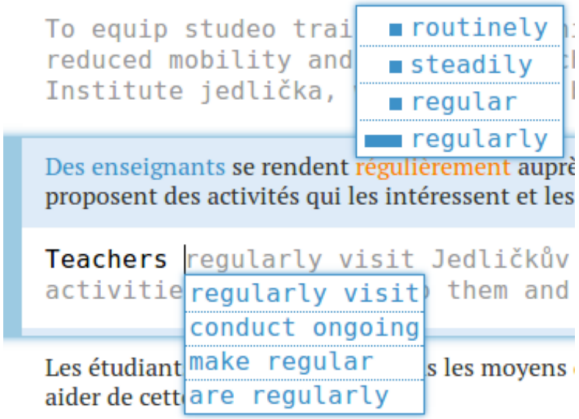
- Show rest of sentence

Spence Green's Lilt System

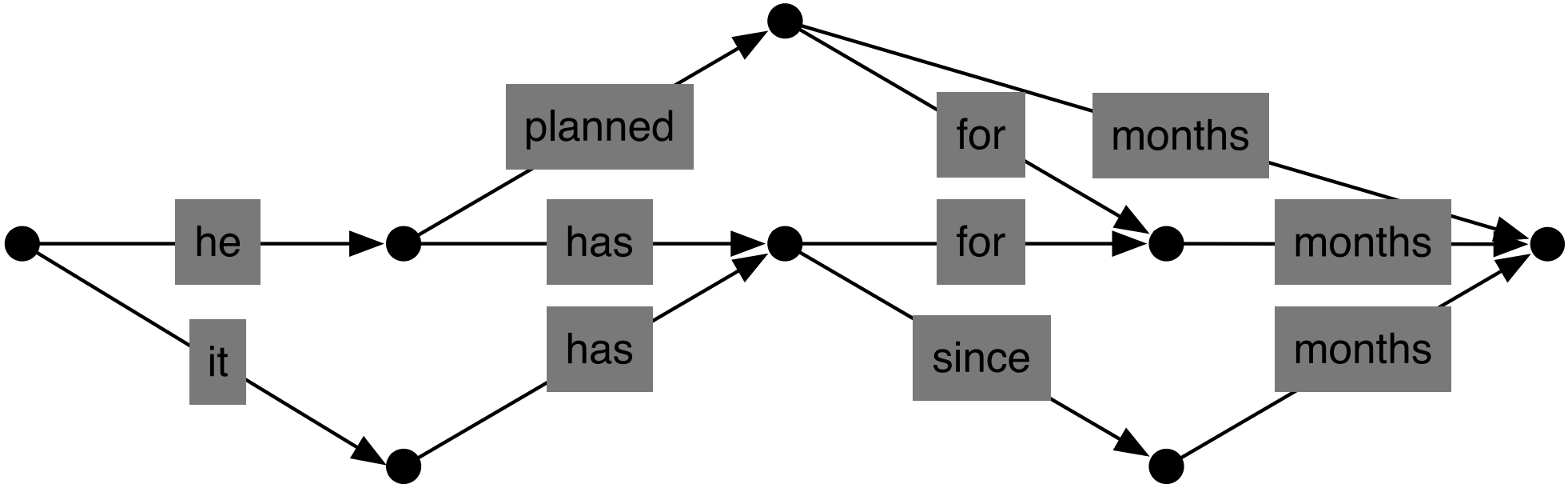
- Show alternate translation predictions



- Show alternate translations predictions with probabilities

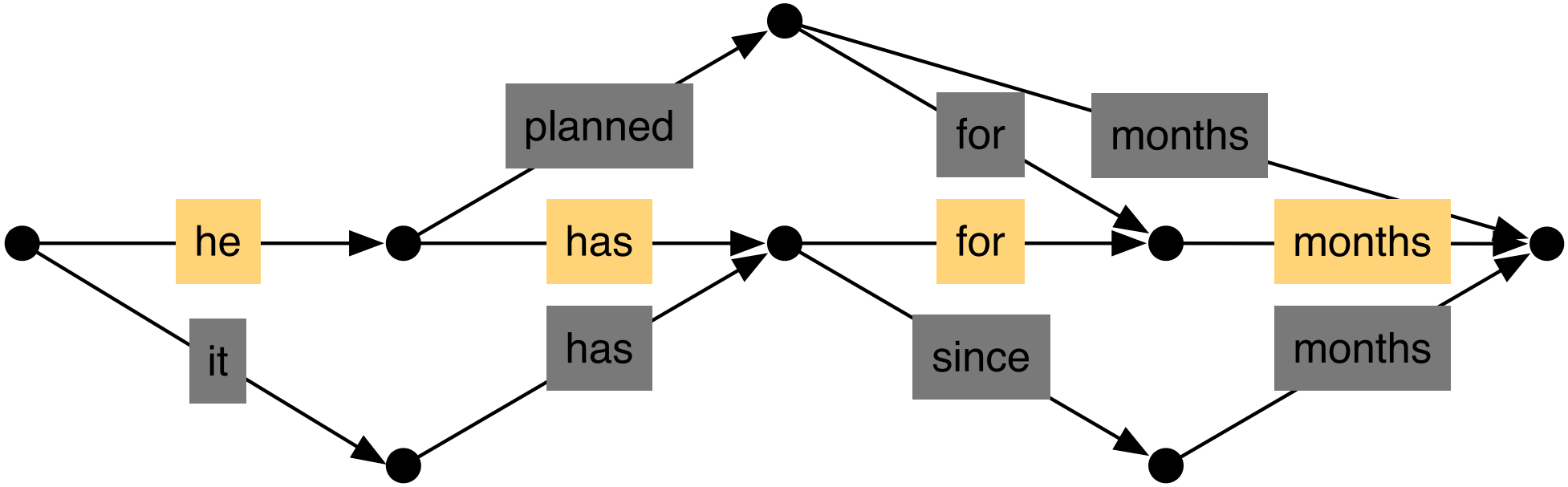


Prediction from Search Graph



Search for best translation creates a graph of possible translations

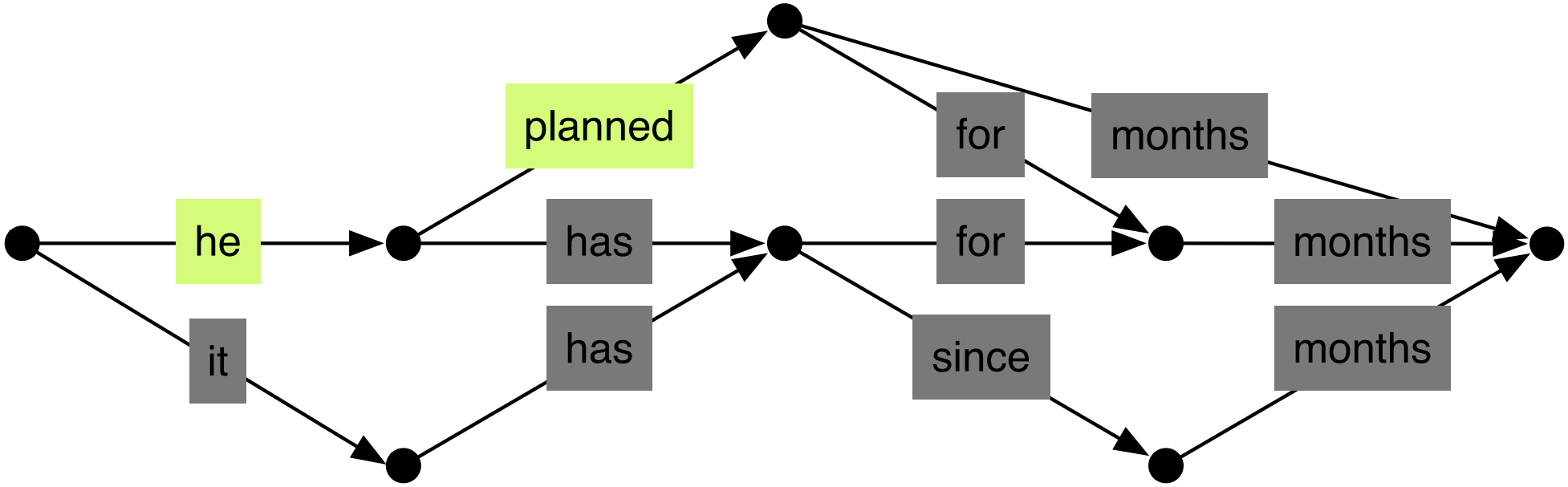
Prediction from Search Graph



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

This path is suggested to the user

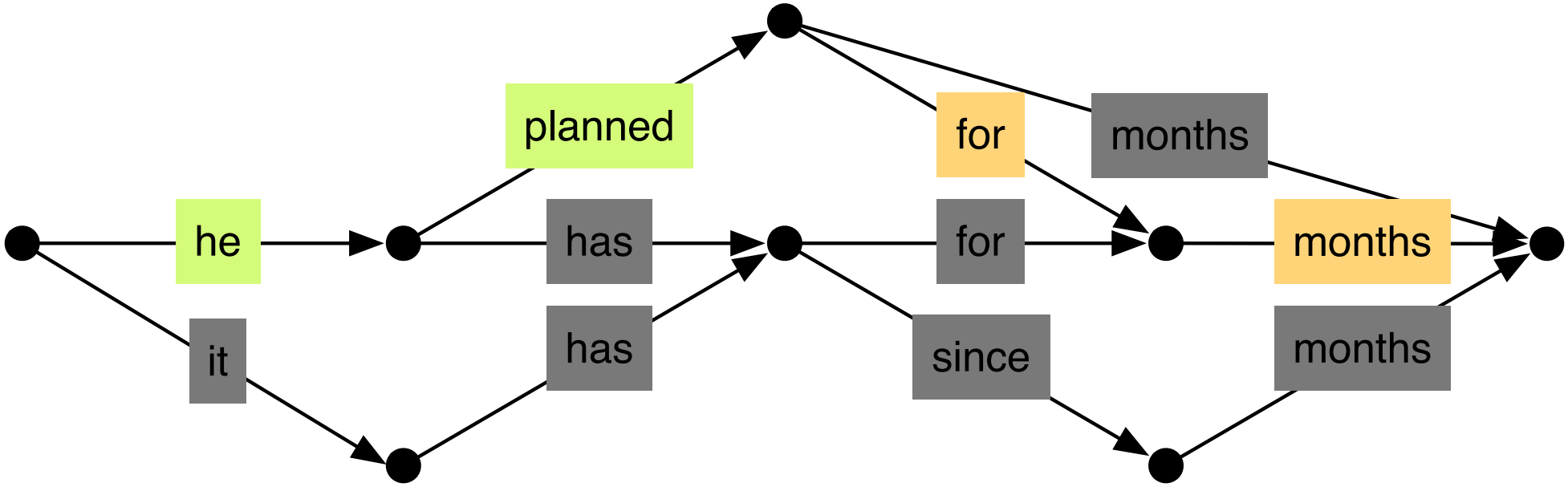
Prediction from Search Graph



The user may enter a different translation for the first words

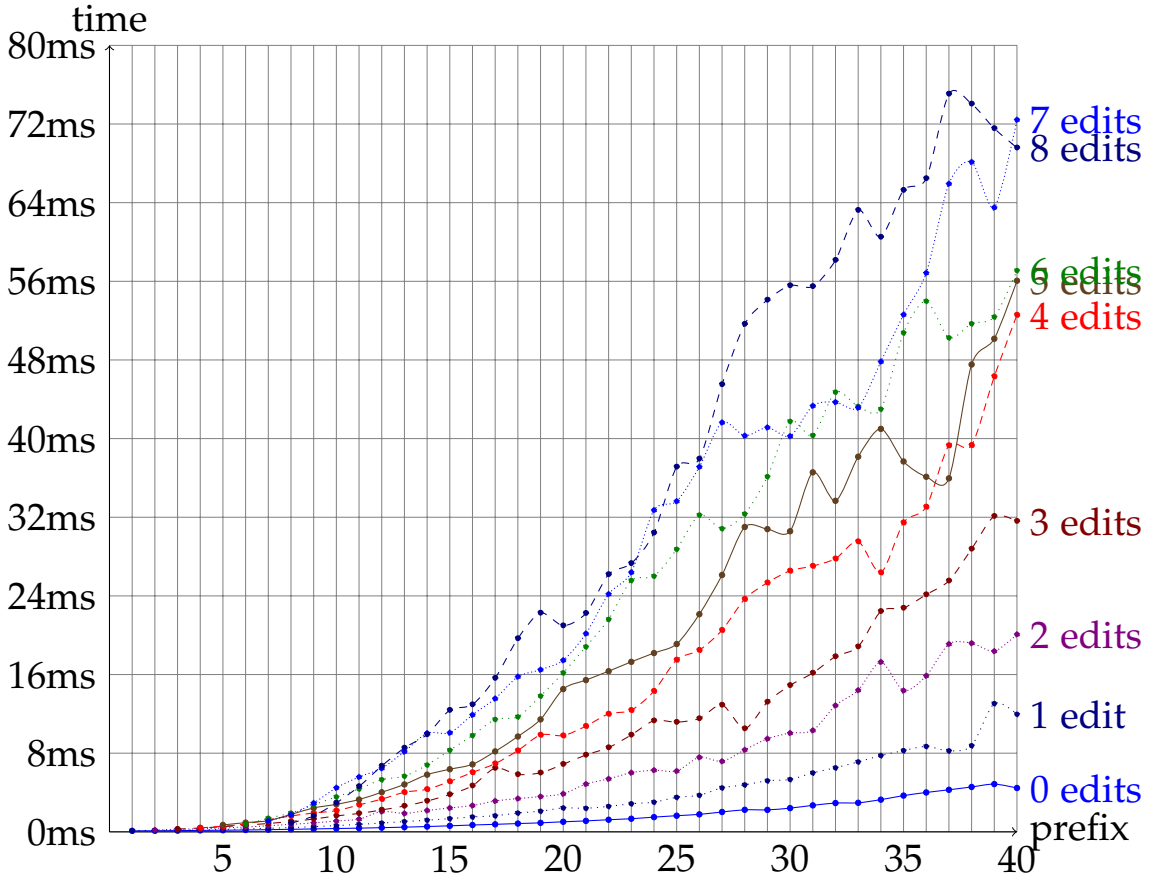
We have to find it in the graph

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 $u \rightarrow$ change prediction
- "Desperate" word completion: find any word that matches

Redecoding

- 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?

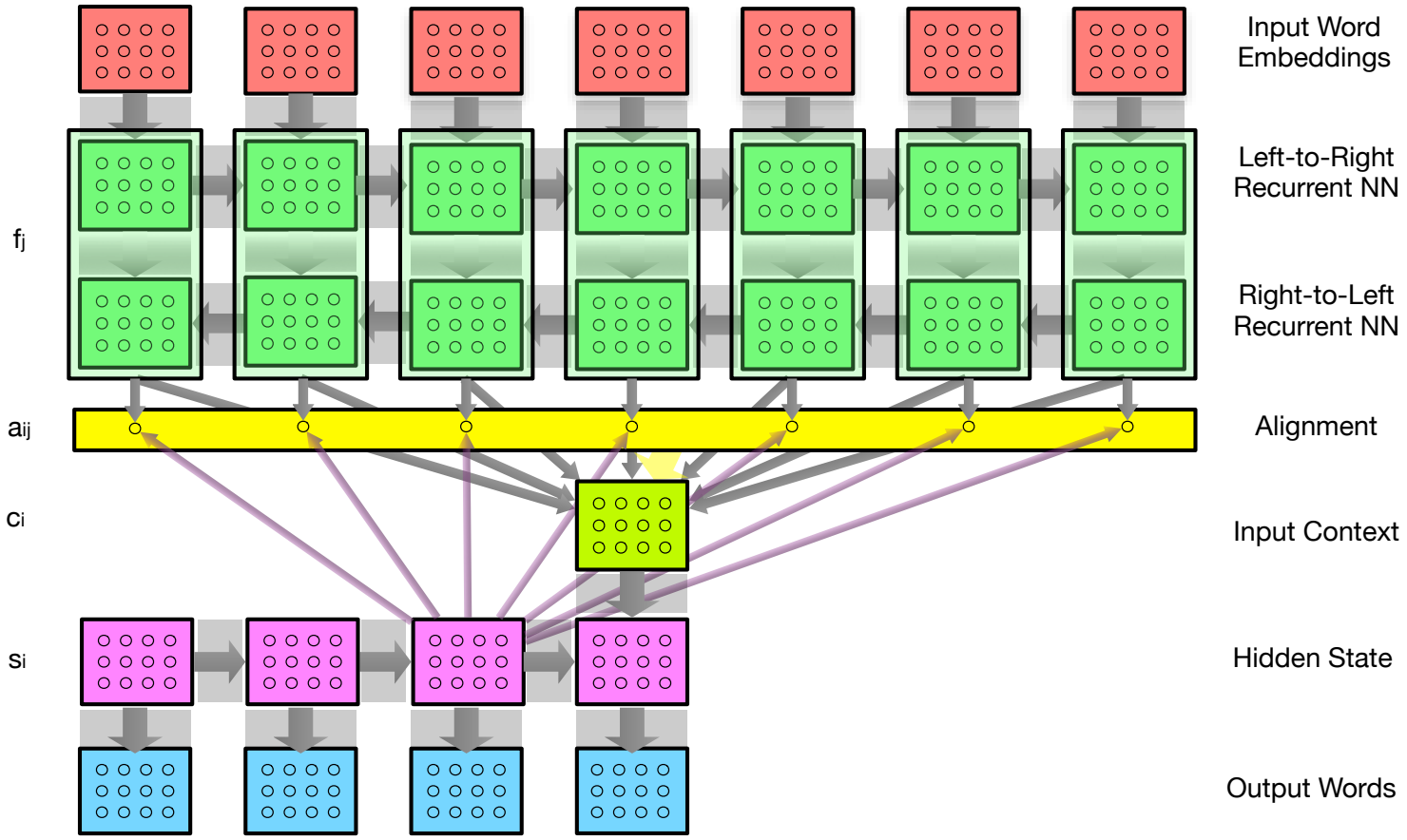
  $\Rightarrow$ Wuebker et al. [ACL 2016] report 51-89ms per sentence

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

- Recent success of neural machine translation (see WMT 2016)
- For instance, attention model



Neural MT: Sequential Prediction

- The model produces words in sequence

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

- 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%)**

Knowles and Koehn [AMTA 2016]

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

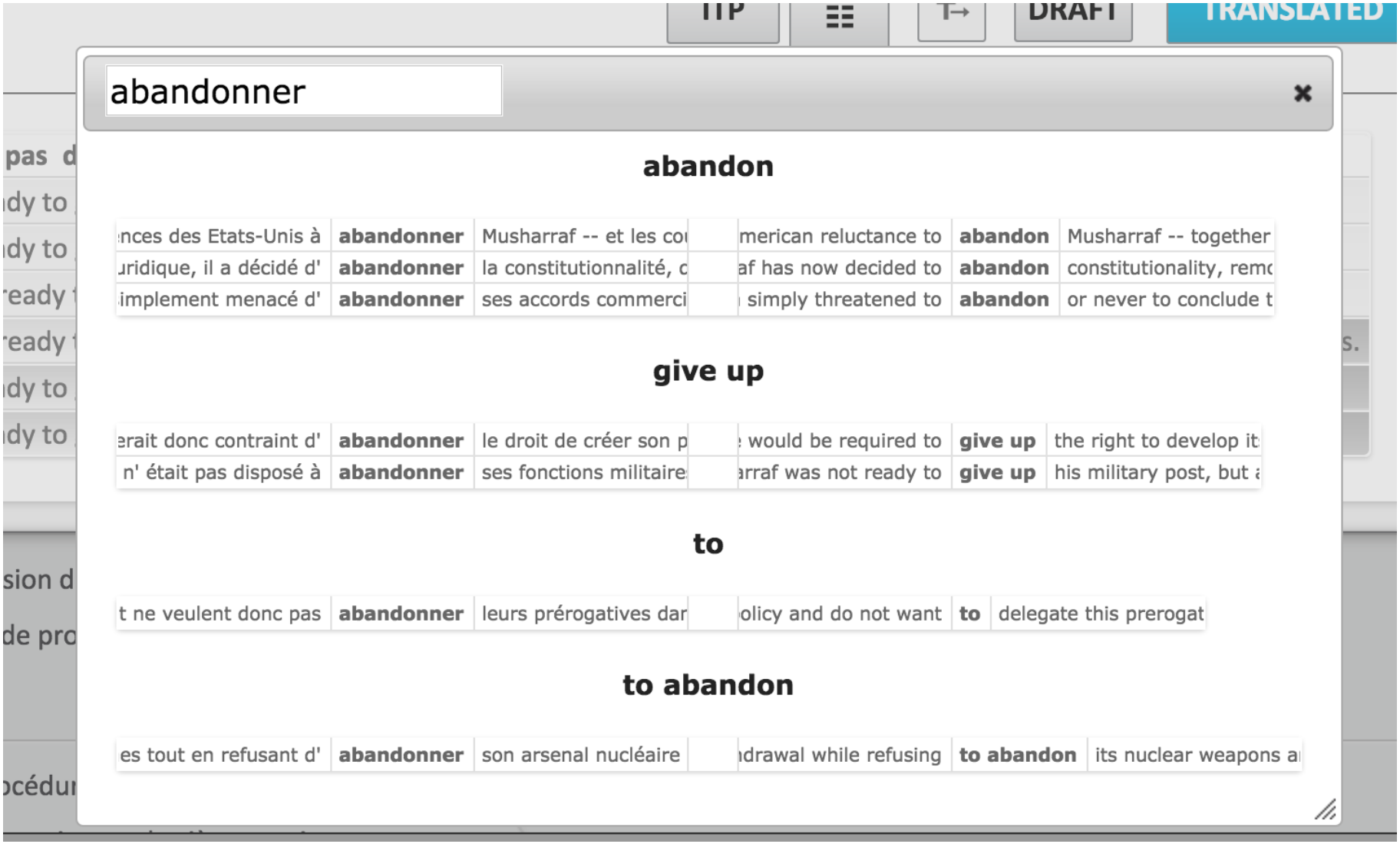
System	Configuration	BLEU	Word Prediction Accuracy	Letter Prediction Accuracy
Neural	no beam search	34.5	61.6%	86.8%
	beam size 12	36.2	63.6%	87.4%
Phrase-based	-	34.5	43.3%	72.8%

- Better recovery from failure

- Fast enough with GPU
  - translation speed with CPU: 100 ms/word
  - translation speed with GPU: 7ms/word

bilingual concordancer

Bilingual Concordancer



How does it Work?

- Have word-aligned parallel corpus

- Efficient data structure to quickly look up queried phrases
  (suffix arrays, we'll come back to them later)

- Translation spotting
  - look up queried phrase
  - use word alignment to identify target phrase
  - some edge cases (unaligned words at beginning/end)

Linguee

English ↔ German ä ö ü ß

machine translation

Dictionary German-English

machine translation *noun*

maschinelle Übersetzung *f*

Maschinenübersetzung *f*

translation machine *noun*

Übersetzungsmaschine *f*

See also:

machine *n* – Maschine *f* · Gerät *nt* · Automat *m* · Anlage *f* · Apparat *m* · […]

machine *v* – bearbeiten *vt* · maschinell herstellen *v* · spanen *v* · zerspanen *v* · maschinell bearbeiten *v* · […]

translation *n* – Übersetzung *f* · Translation *f* · Übersetzen *nt* · Verschiebung *f* · Sprachübersetzung *f* · […] […]

Wikipedia

External sources (not reviewed)

The implementing provisions applicable to the machine translation system would have to be established by the Select Committee [...] cep.eu

Die Durchführungsbestimmungen für das System der maschinellen Übersetzung müssten vom engeren Ausschuss des EPO-Verwaltungsrats [...] cep.eu

By user licence agreements relating to the SYSTRAN machine translation software program concluded between the applicants' [...] eur-lex.europa.eu

Durch Lizenzverträge über die Benutzung der Software für maschinelle Übersetzungen SYSTRAN zwischen den Rechtsvorgängern der Klägerinnen [...] eur-lex.europa.eu

[...] curriculum vitae, in forms suitable for multilingual machine translation, without restricting a user's option of adding other [...] europarl.europa.eu

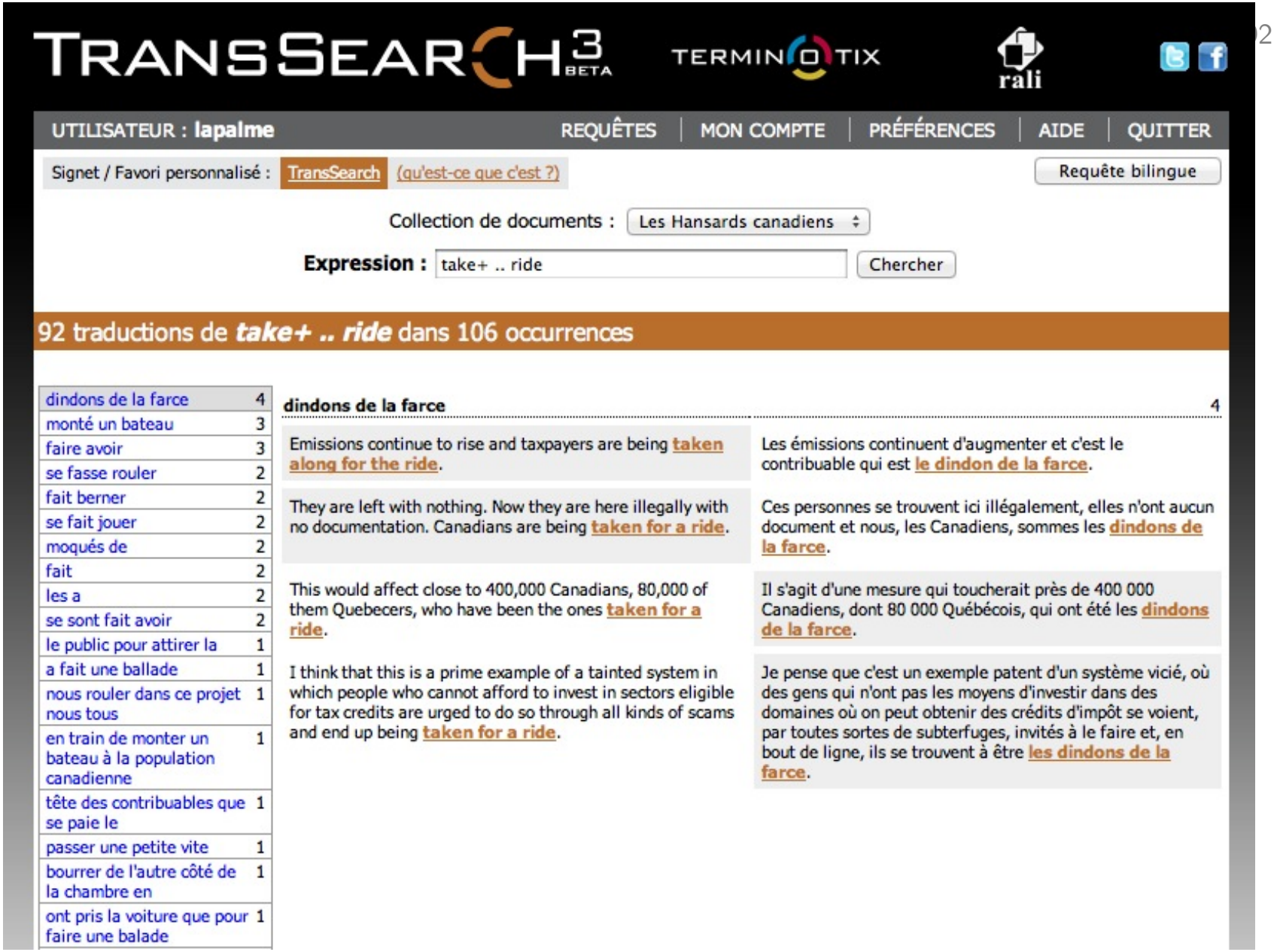
[...] standardisierten Lebenslauf zu prüfen, die für eine automatische Übersetzung in mehrere Sprachen geeignet sind, wobei der Nutzer [...] europarl.europa.eu

Verification of Terminology

- Translation of German *Windkraft*

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

- Context shows when each translation is used
- Indication of source supports trust in translations

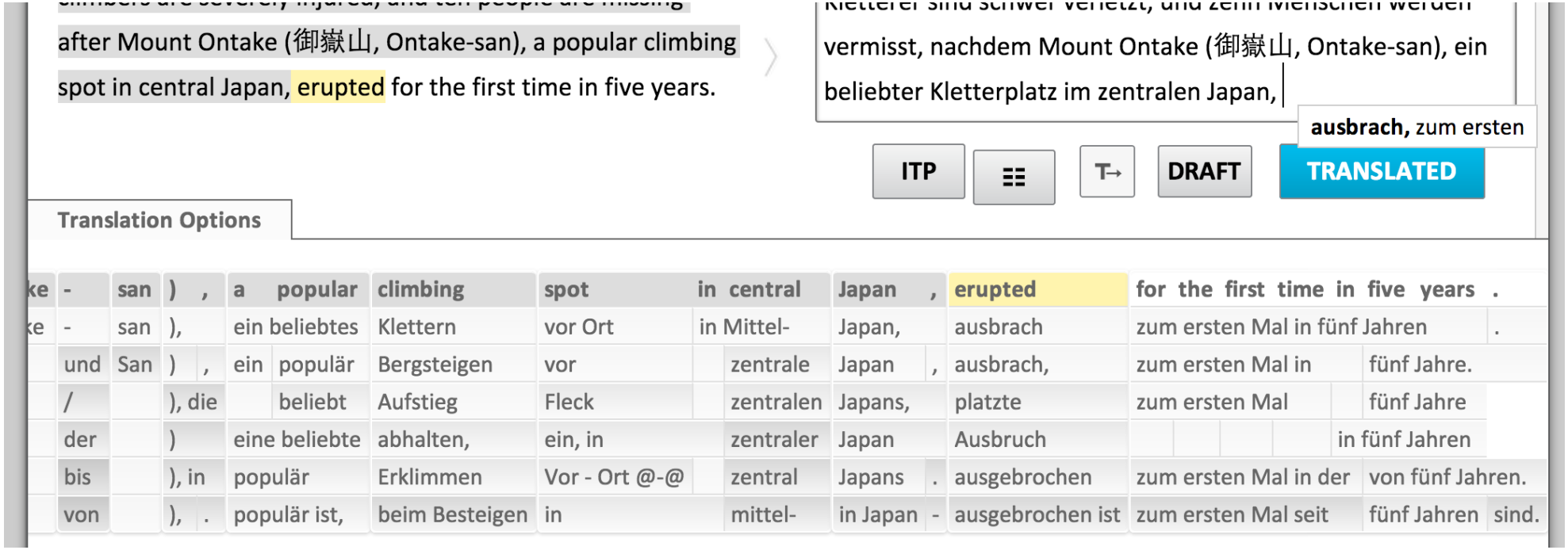

TRANSSEARCH3 BETA
TERMINOTIX
rali
UTILISATEUR : lapalme
REQUÊTES | MON COMPTE | PRÉFÉRENCES | AIDE | QUITTER
Signet / Favori personnalisé : TransSearch (qu'est-ce que c'est ?)
Requête bilingue
Collection de documents : Les Hansards canadiens
Expression : take+ .. ride
Chercher
92 traductions de take+ .. ride dans 106 occurrences
dindons de la farce 4
monté un bateau 3
faire avoir 3
se fasse rouler 2
fait berner 2
se fait jouer 2
moqués de 2
fait 2
les a 2
se sont fait avoir 2
le public pour attirer la 1
a fait une ballade 1
nous rouler dans ce projet nous tous 1
en train de monter un bateau à la population canadienne 1
tête des contribuables que se paie le 1
passer une petite vite 1
bourrer de l'autre côté de la chambre en 1
ont pris la voiture que pour faire une balade 1
dindons de la farce 4
Emissions continue to rise and taxpayers are being taken along for the ride.
Les émissions continuent d'augmenter et c'est le contribuable qui est le dindon de la farce.
They are left with nothing. Now they are here illegally with no documentation. Canadians are being taken for a ride.
Ces personnes se trouvent ici illégalement, elles n'ont aucun document et nous, les Canadiens, sommes les dindons de la farce.
This would affect close to 400,000 Canadians, 80,000 of them Quebecers, who have been the ones taken for a ride.
Il s'agit d'une mesure qui toucherait près de 400 000 Canadiens, dont 80 000 Québécois, qui ont été les dindons de la farce.
I think that this is a prime example of a tainted system in which people who cannot afford to invest in sectors eligible for tax credits are urged to do so through all kinds of scams and end up being taken for a ride.
Je pense que c'est un exemple patent d'un système vicié, où des gens qui n'ont pas les moyens d'investir dans des domaines où on peut obtenir des crédits d'impôt se voient, par toutes sortes de subterfuges, invités à le faire et, en bout de ligne, ils se trouvent à être les dindons de la farce.

TransSearch: Improved Transpotting

- Used to solve **difficult** translation problems
  - 7.2 million queries submitted to the system over a 6-year period
  - 87% contain at least two words
  - mainly search for idiomatic expressions such as *in keeping with*
- Improved translation spotting [Bourdaillet et al., MT Journal 2011]
- Filtering with classifier (45 features, trained on annotated data)
  - relative word count
  - word alignment scores
  - ratio of function words
- Merging of translations that only differ in function words, morphology
- Pseudo-relevance feedback

translation options

Translation Option Array



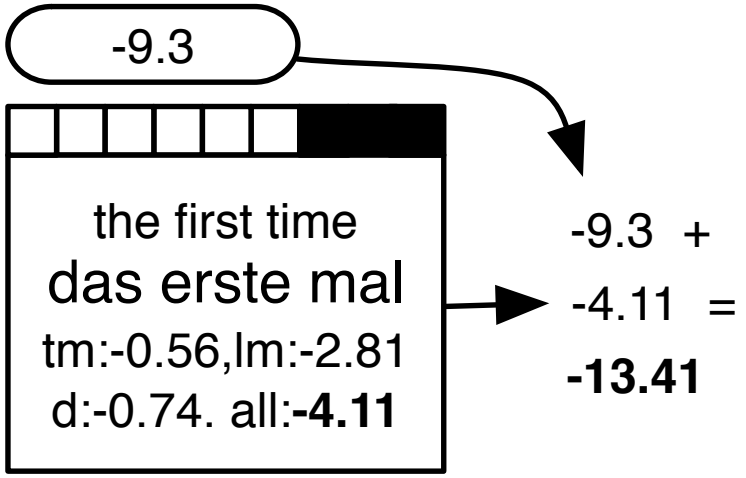
- Visual aid: non-intrusive provision of cues to the translator
- Trigger passive vocabulary

Visualization

- Show up to 6 options per word or phrase
- Rank best option on top
- Use color highlighting to show likelihood
  (grey = less likely to be useful)
- Clickable: click on target phrase → added to edit area
- Automatic orientation
  - most relevant is next word to be translated
  - automatic centering on next word

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.

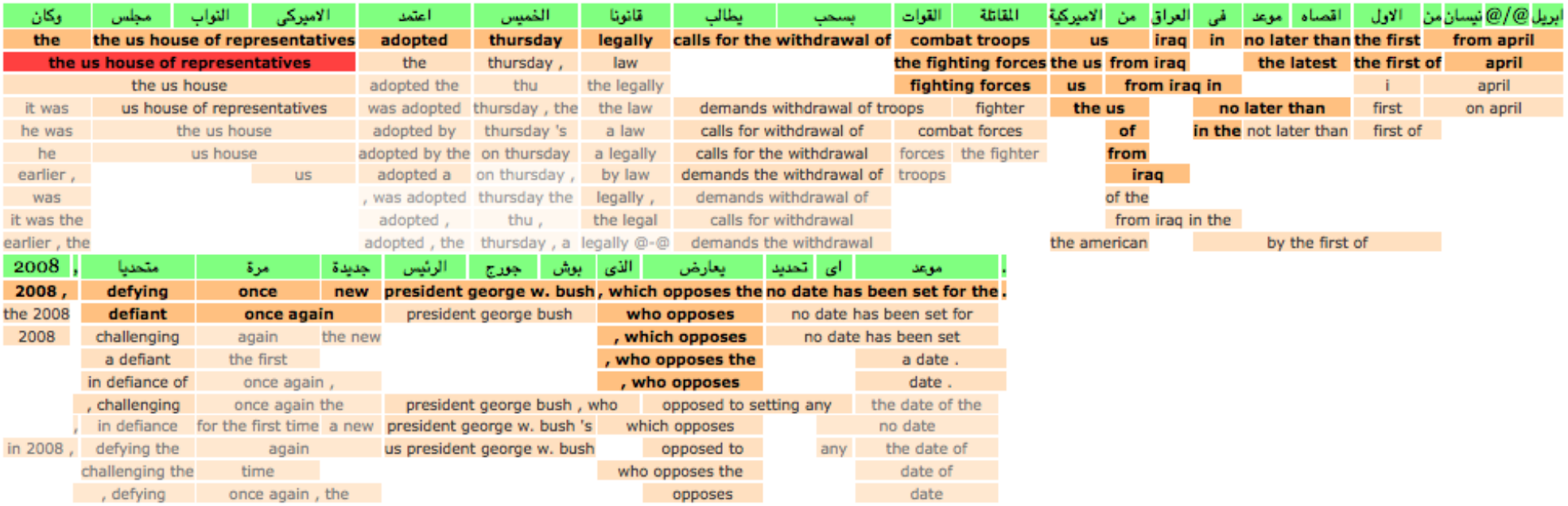
- What does this mean?

- 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.

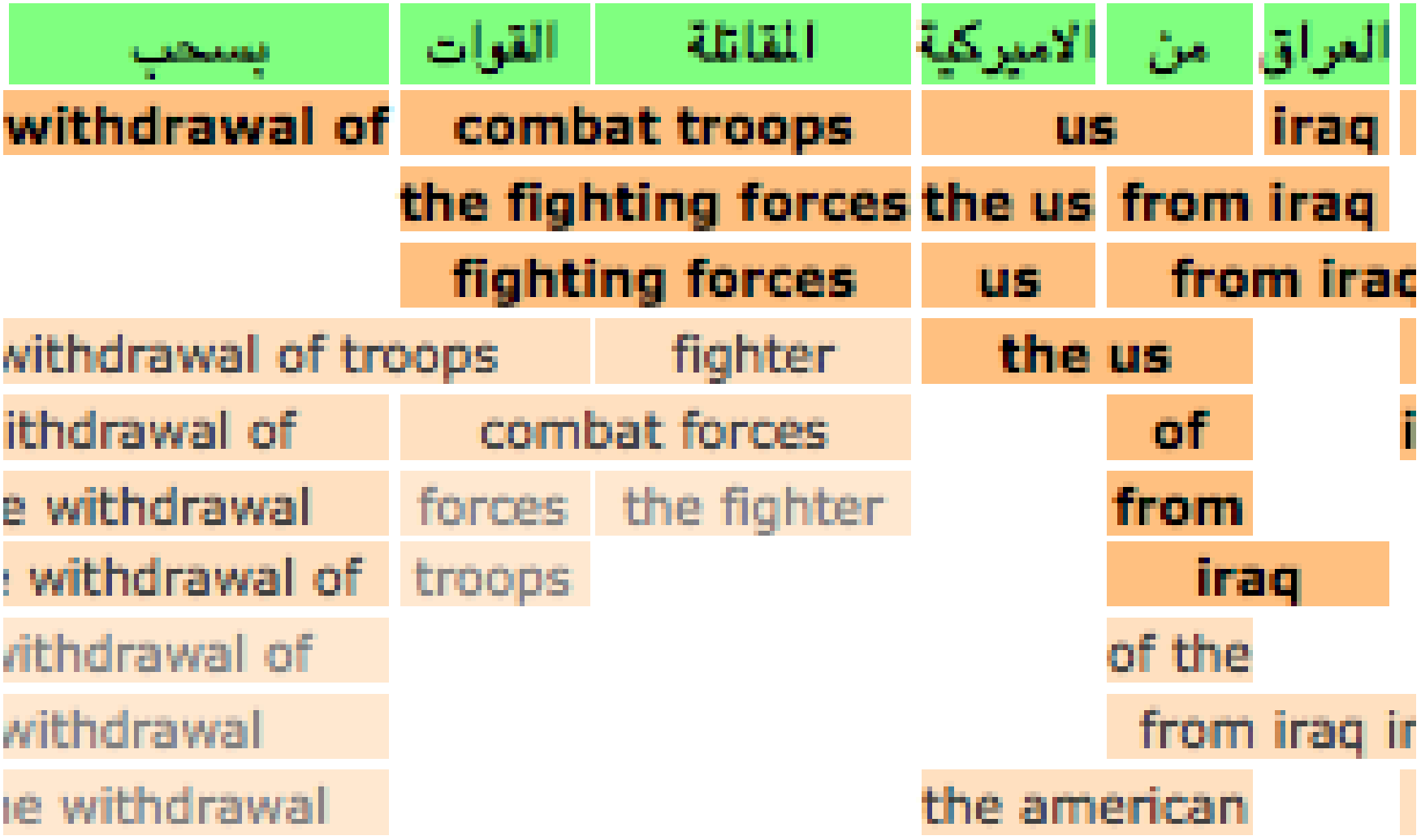
- Document context helps

Example: Arabic

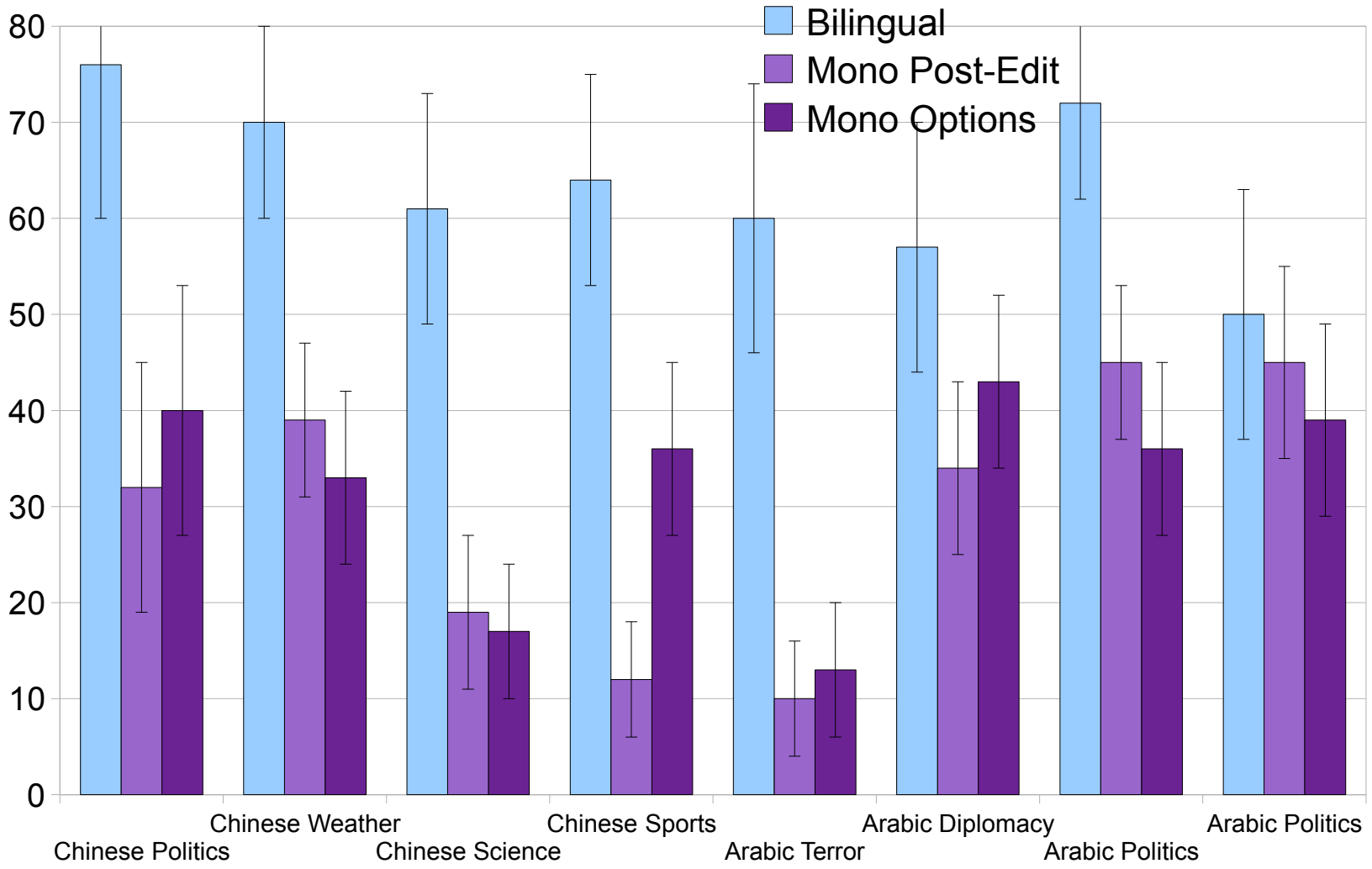


up to 10 translations for each word / phrase

Example: Arabic



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

paraphrasing

Paraphrasing

Input Sentence

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

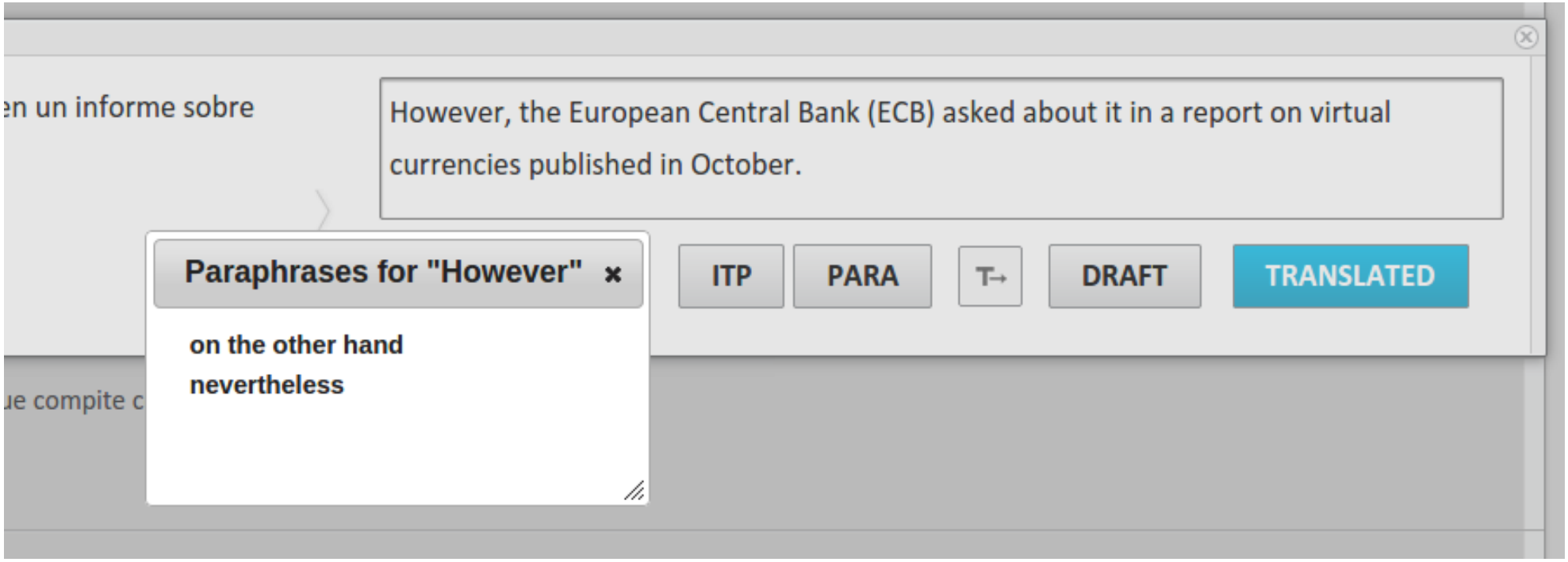
Professional Translator

He planned for months to give a lecture in Miami in October.

- give a presentation
- present his work
- give a speech
- speak

User requests alternative translations for parts of sentence.

Visualization in CASMACAT



- User marks part of translation
- Clicks on paraphrasing button
- Alternative translations appear

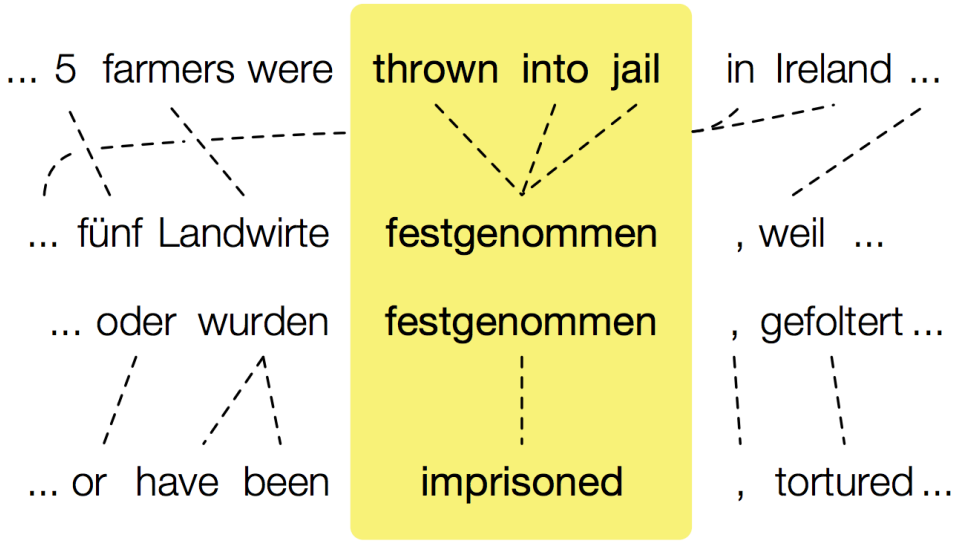
Paraphrasing Research

- Somewhat popular research area

- Popular method: extract from parallel data

  - goal: find paraphrases for phrase e
  - look up likely translations $f_1, f_2, ...$ for e
  - for each f_i, look up likely translations $e'_{i1}, e'_{i2}, ...$

  $\Rightarrow$ these are the paraphrases



- Refinement: collect over several foreign languages, intersect

- Paraphrase database for several languages:

  `http://paraphrase.org/`

Paraphrasing in Context

- Our problem: paraphrasing in context
  - driven by source
  - considers sentence context
  - ranking and diversity important
  - real time performance

- Approach
  - target span is mapped to source span
  - search graph is consulted for alternative translations for source span
  - additional translations generated by combining translation options

  $\Rightarrow$ initial list of translations

  - various components to distill n-best paraphrases

Components

- Filtering: remove some translations
  - with extraneous punctuation
  - too similar to others
  - additional function words
- Scoring: score translations
  - translation model scores
  - language model score in context
  - compare alternate translations against best path
- Sorting: rank list
  - cluster translations by similarity
  - picks best translation from each cluster

Automatic Evaluation

- Motivation
  - alternative translations should fix translation errors

  → create bad translations by back-translation

- Process
  - Train machine translation system for both directions
  - Translate test set `target` → `source` → `target*`
  - Spot differences between `target` and `target*`
  - Use span in `target*` as "marked by user", span in `target` as correct

Example

- Translate

Unlike in Canada , **the American states** *are responsible for the organisation of federal elections.*

- Into

в отличие от канады, американские штаты ответственны за организацию федеральных выборов в соединенных штатах .

- Back into English

Unlike in Canada , **US states** *are responsible for the organization of federal elections.*

Manual Evaluation

- Web based interactive evaluation tool
- Same setup as automatic evaluation
  - shows target span
  - 5 selectable paraphrases
  - user accepts one → correct
- Four users (U1–U4)
- Number of instances where one translation is correct

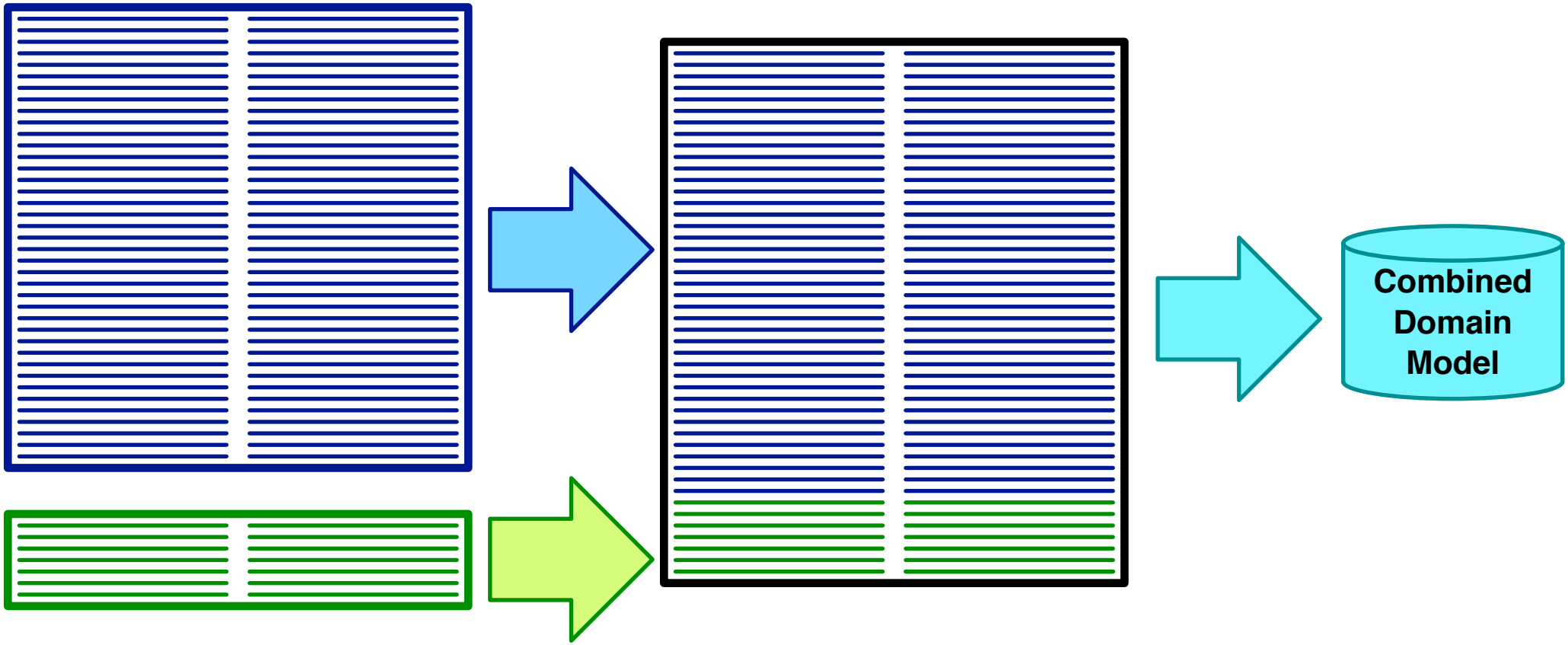
Method	U1	U2	U3	U4	average score
1	8	6	9	6	6/50
7	15	17	12	10	13/50
10	24	20	26	29	26/50

adaptation

Adaptation

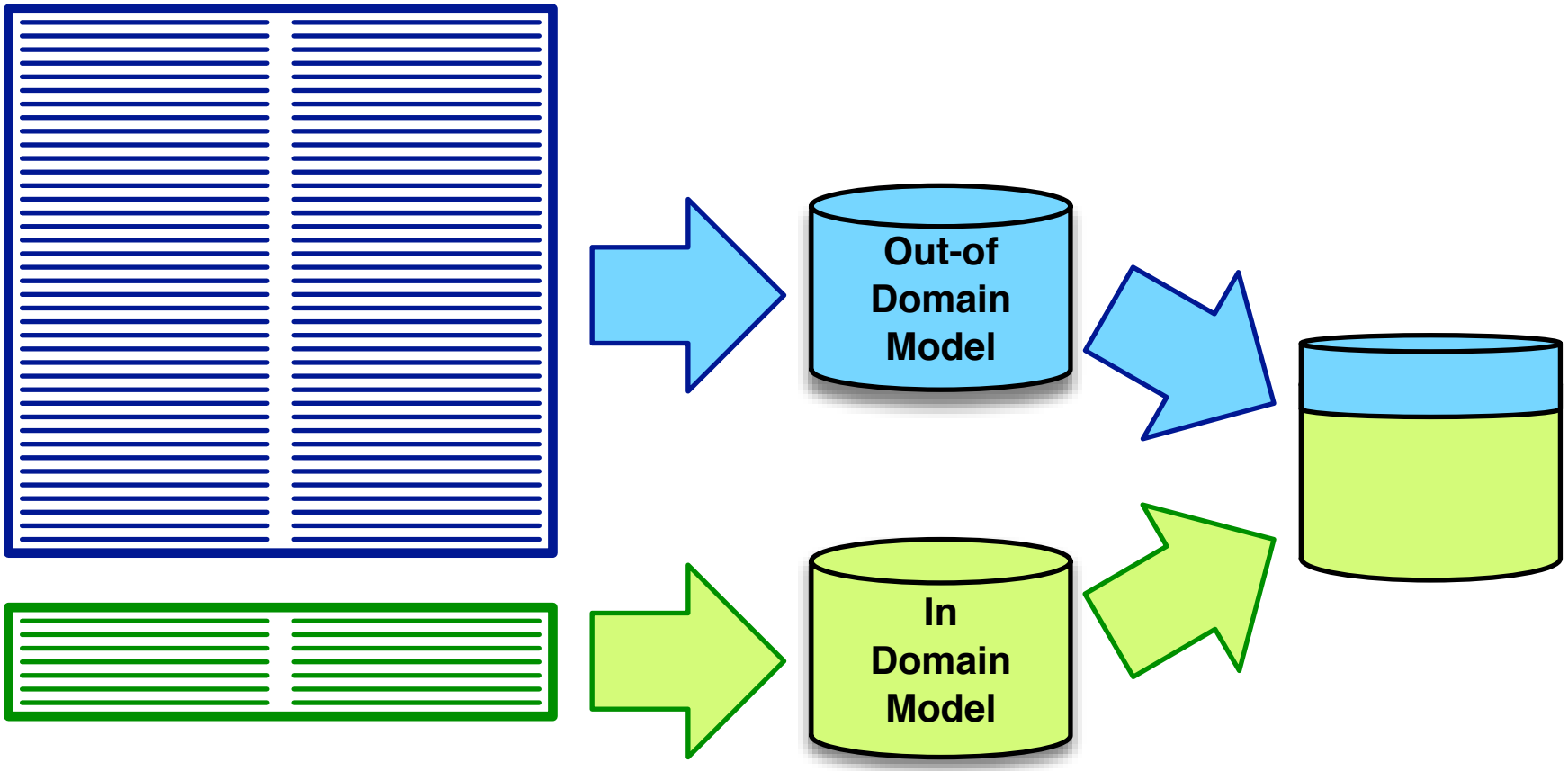
- Machine translation works best if optimized for domain
- Typically, large amounts of out-of-domain data available
  - European Parliament, United Nations
  - unspecified data crawled from the web
- Little in-domain data (maybe 1% of total)
  - information technology data
  - more specific: IBM's user manuals
  - even more specific: IBM's user manual for same product line from last year
  - and even more specific: sentence pairs from current project
- Various domain adaptation techniques researched and used

Combining Data



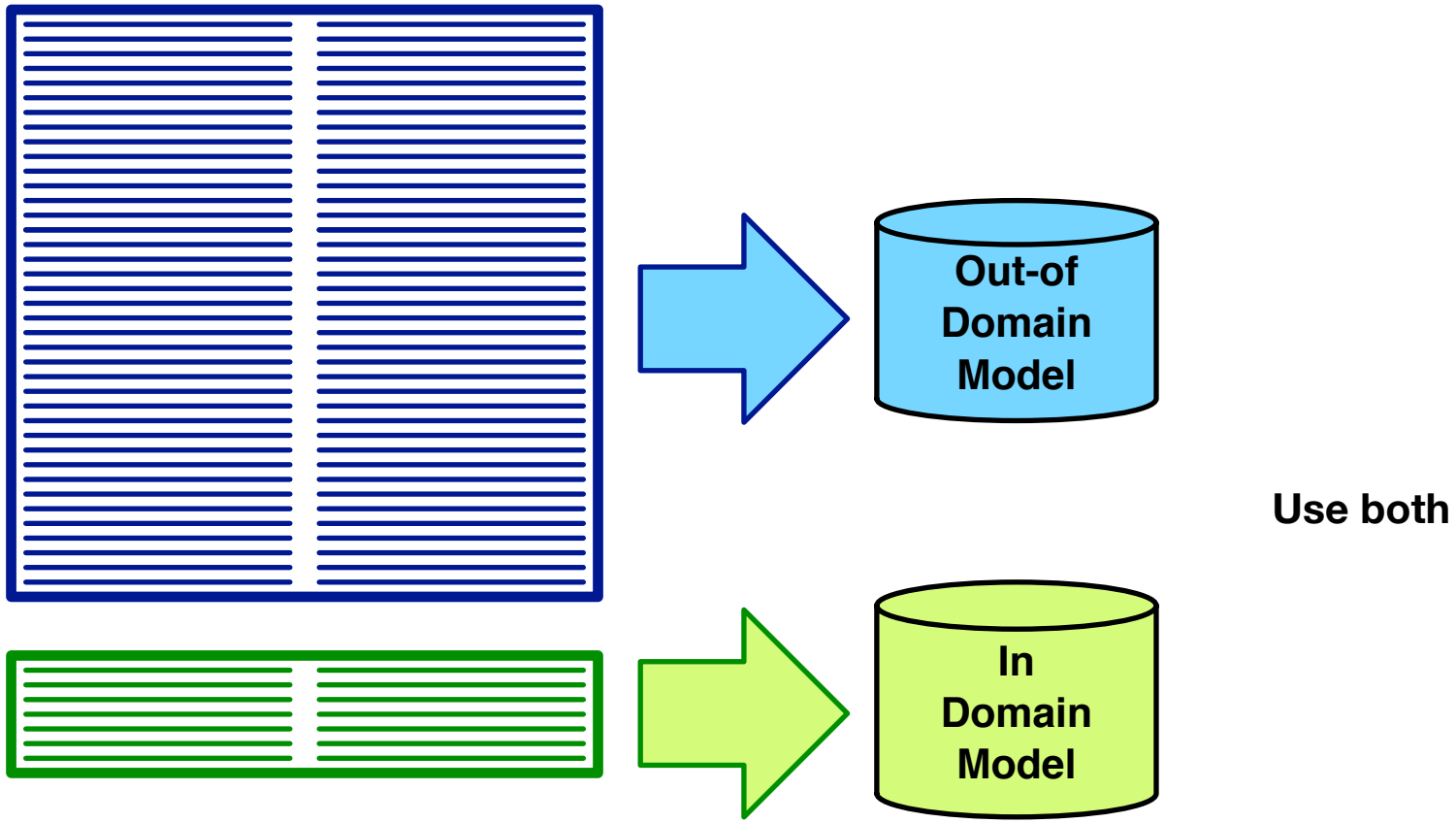
- Too biased towards out of domain data
- May flag translation options with indicator feature functions

Interpolate Models



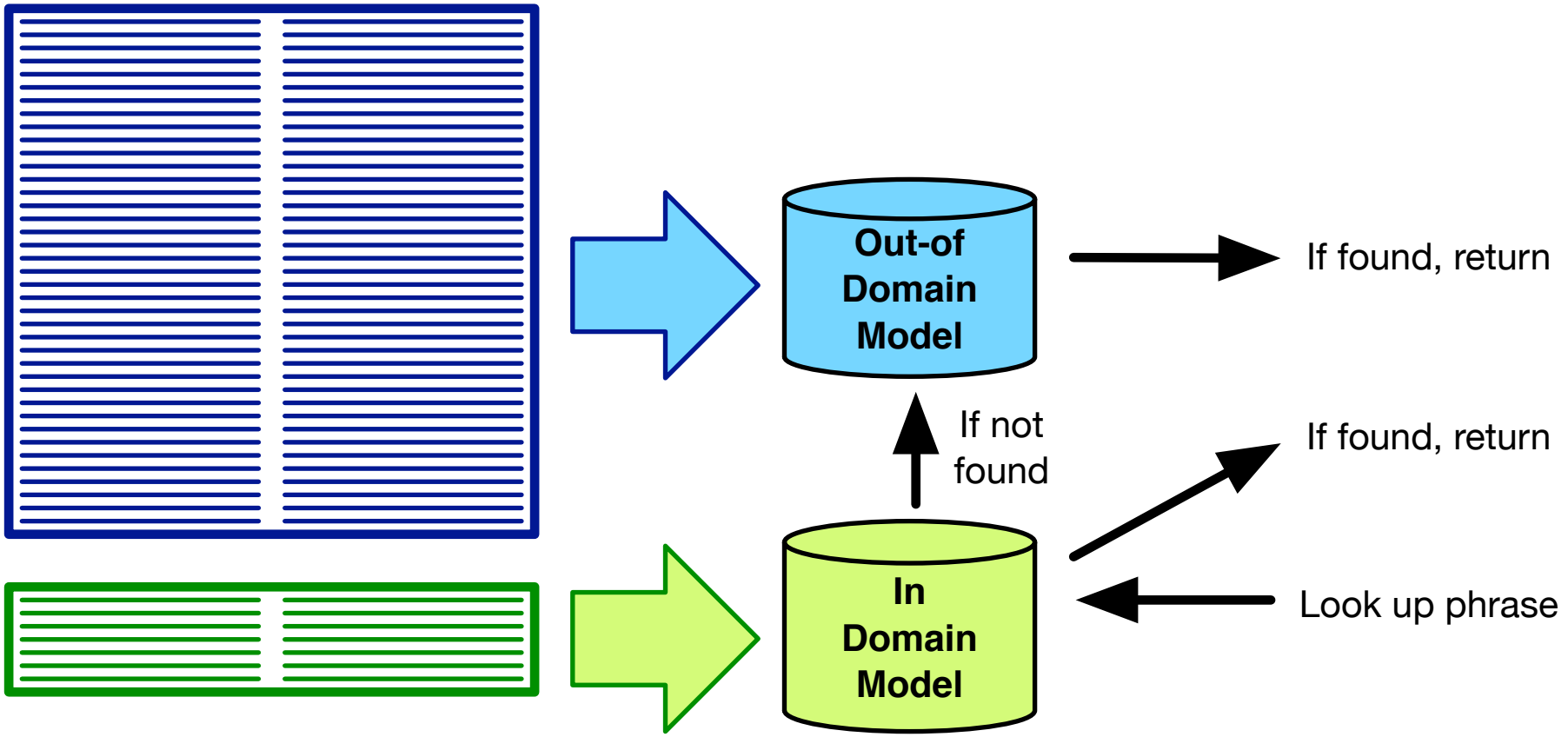
- $p_c(e|f) = \lambda_{\text{in}} p_{\text{in}}(e|f) + \lambda_{\text{out}} p_{\text{out}}(e|f)$
- Quite successful for language modelling

Multiple Models

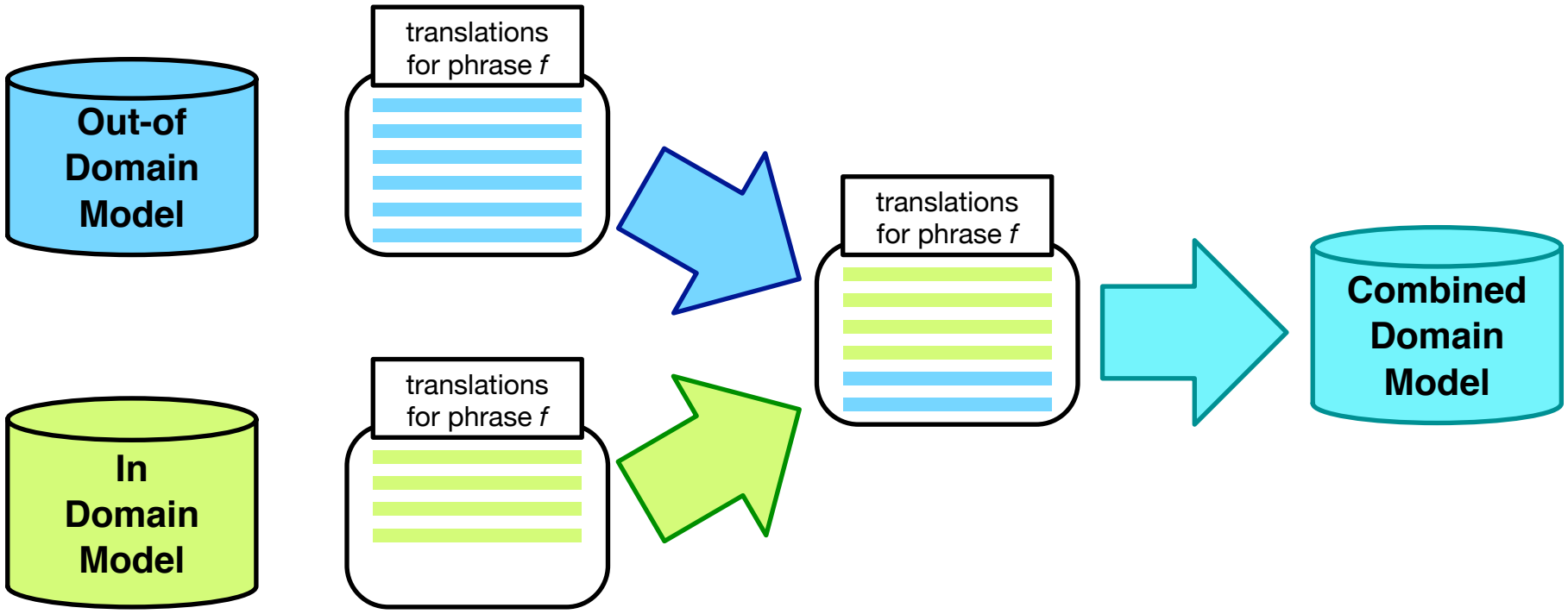


- Multiple models → multiple feature functions

Backoff

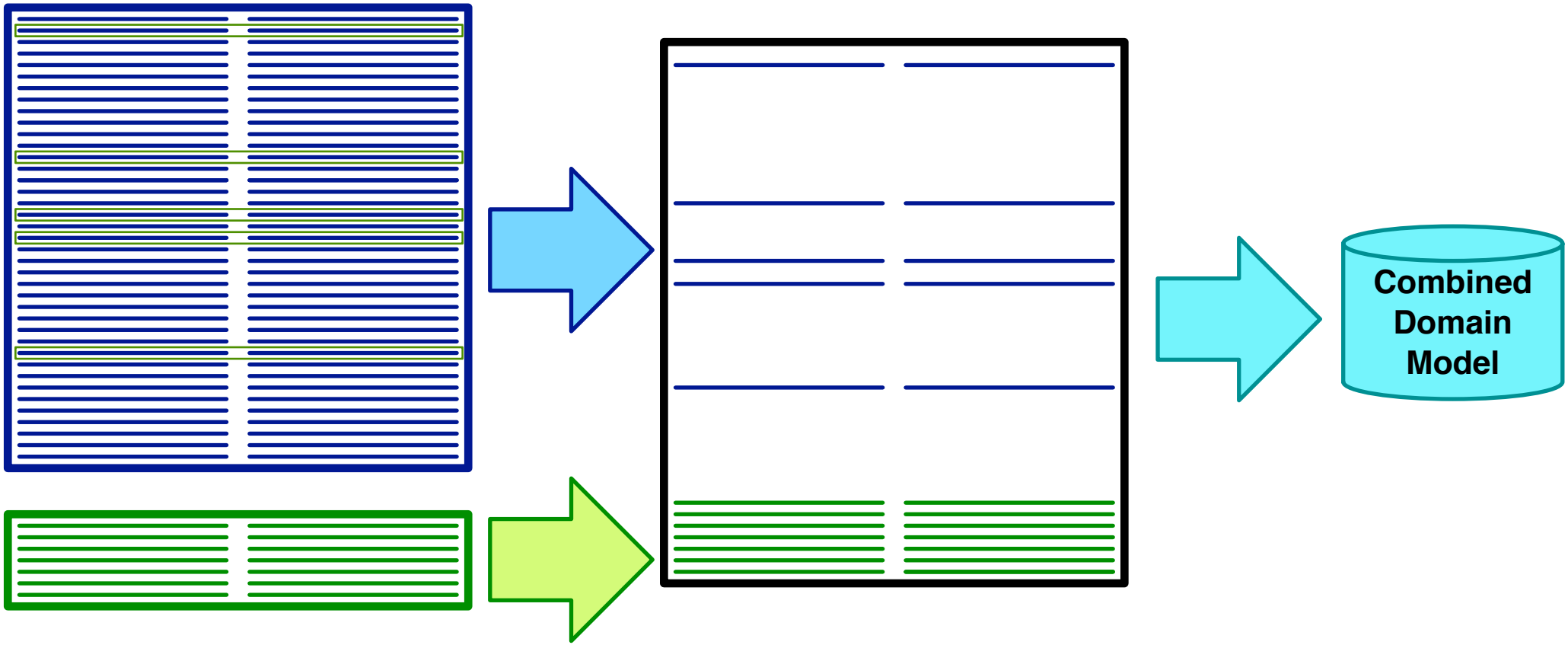

Out-of Domain Model
If found, return
If not found
If found, return
In Domain Model
Look up phrase

Fill-Up



- Use translation options from in-domain table
- Fill up with additional options from out-of-domain table

Sentence Selection

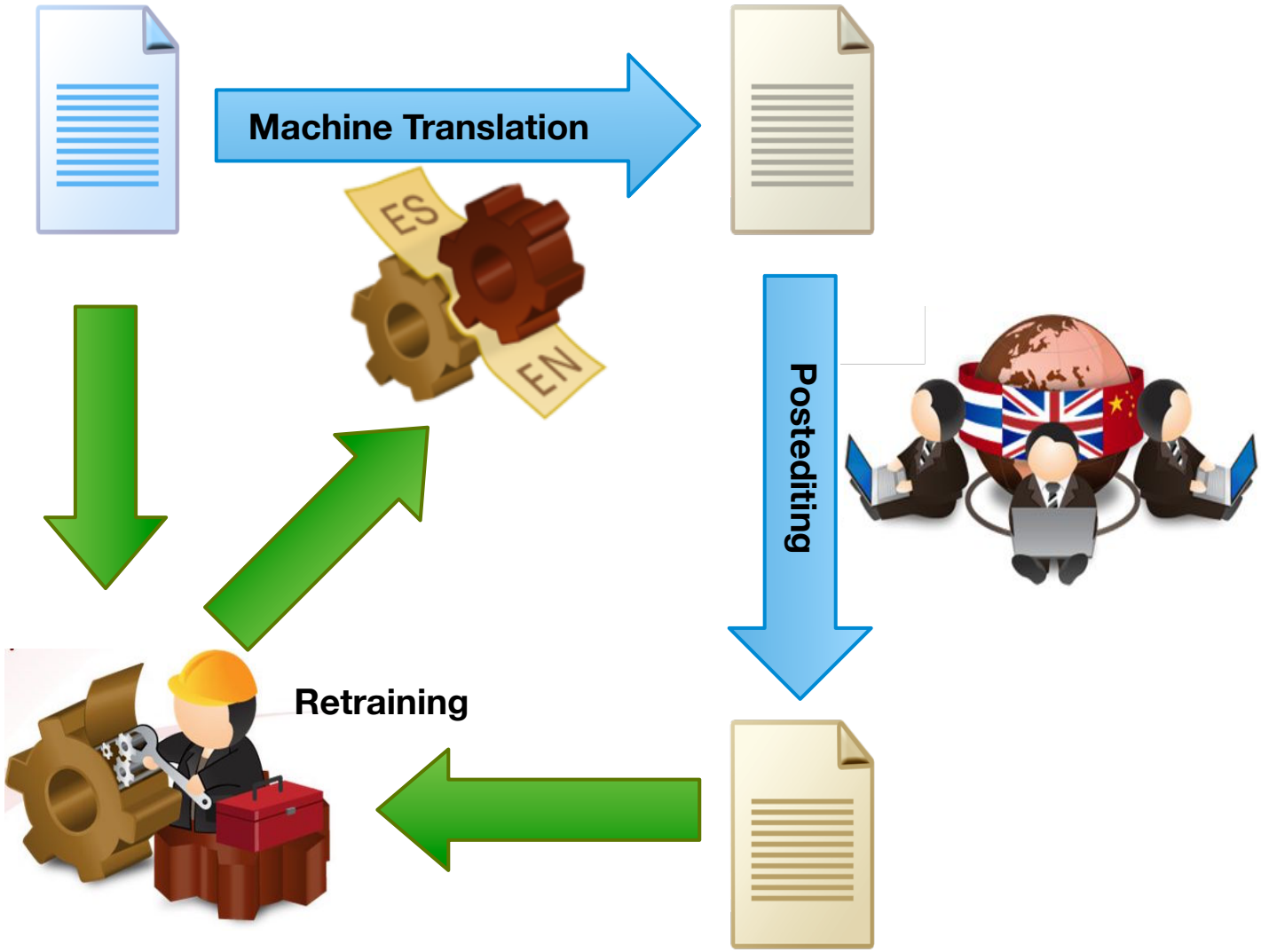


- Select out-of-domain sentence pairs that are similar to in-domain data
- Score similarity with language model, other means

Project Adaptation

- Method developed by the Matecat project
- Update model during translation project
- After each day
  - collected translated sentences
  - add to model
  - optimize
- Main benefit after the first day

Instant Adaptation



Adaptable Translation Model

- Store in memory
  - parallel corpus
  - word alignment
- Adding new sentence pair
  - word alignment of sentence pair
  - add sentence pair
  - update index (suffix array)
- Retrieve phrase translations on demand

On-Demand Word Alignment

- Needed: word alignment method that scores a sentence pairs
- Online EM algorithm
  - keep sufficient statistics of corpus in memory
  - run EM iteration on single sentence pair
  - update statistics
  - return word alignment
- For efficiency reason, a static model may be sufficient
- Implementations in bith mGIZA and fast-align

Suffixes

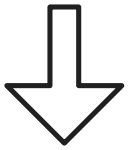
1 government of the people , by the people , for the people
2 of the people , by the people , for the people
3 the people , by the people , for the people
4 people , by the people , for the people
5 , by the people , for the people
6 by the people , for the people
7 the people , for the people
8 people , for the people
9 , for the people
10 for the people
11 the people
12 people

Sorted Suffixes

5	, by the people , for the people
9	, for the people
6	by the people , for the people
10	for the people
1	government of the people , by the people , for the people
2	of the people , by the people , for the people
12	people
4	people , by the people , for the people
8	people , for the people
11	the people
3	the people , by the people , for the people
7	the people , for the people

Suffix Array

5	, by the people , for the people
9	, for the people
6	by the people , for the people
10	for the people
1	government of the people , by the people , for the people
2	of the people , by the people , for the people
12	people
4	people , by the people , for the people
8	people , for the people
11	the people
3	the people , by the people , for the people
7	the people , for the people



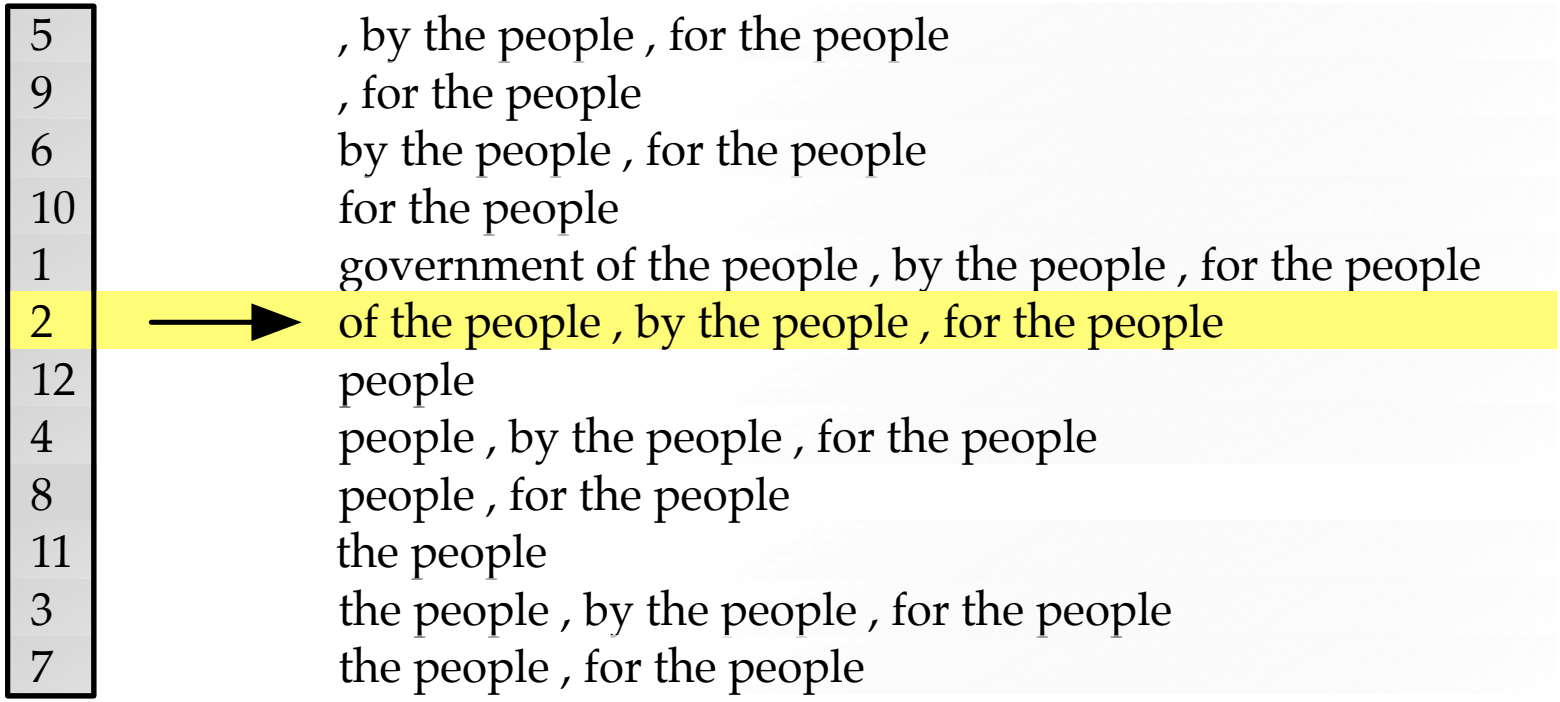
suffix array: sorted index of corpus positions

Querying the Suffix Array

5	, by the people , for the people
9	, for the people
6	by the people , for the people
10	for the people
1	government of the people , by the people , for the people
2	of the people , by the people , for the people
12	people
4	people , by the people , for the people
8	people , for the people
11	the people
3	the people , by the people , for the people
7	the people , for the people

Query: people

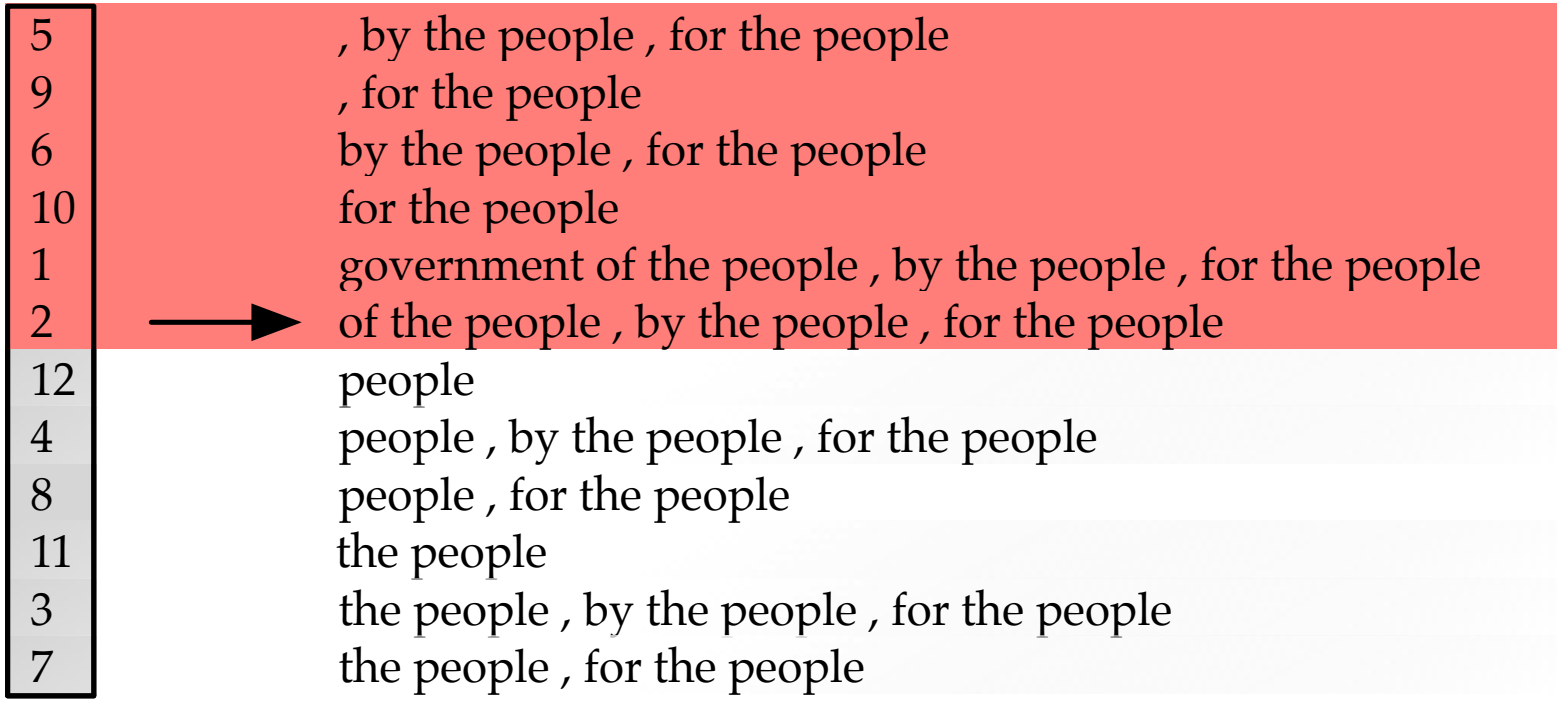
Querying the Suffix Array



Query: people

Binary search: start in the middle

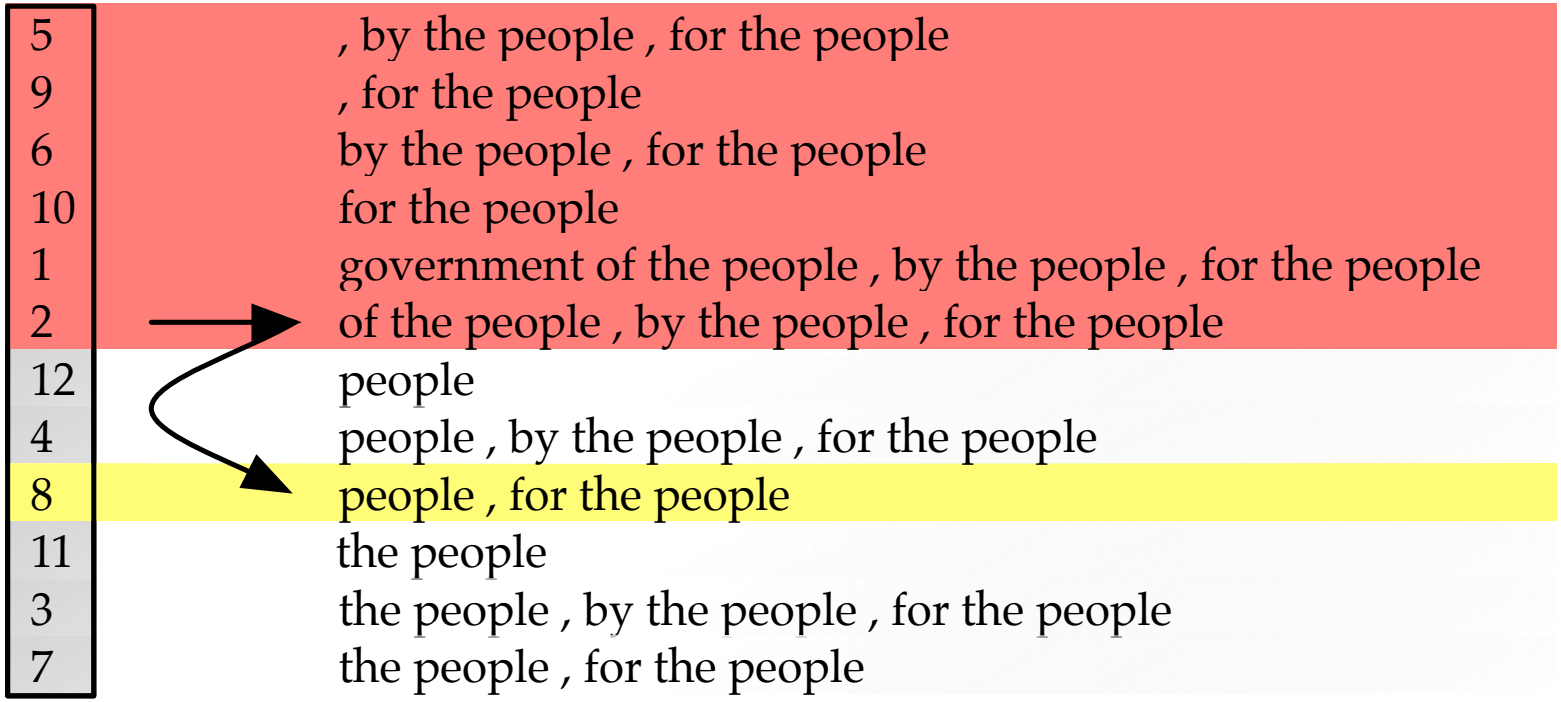
Querying the Suffix Array



Query: people

Binary search: discard upper half

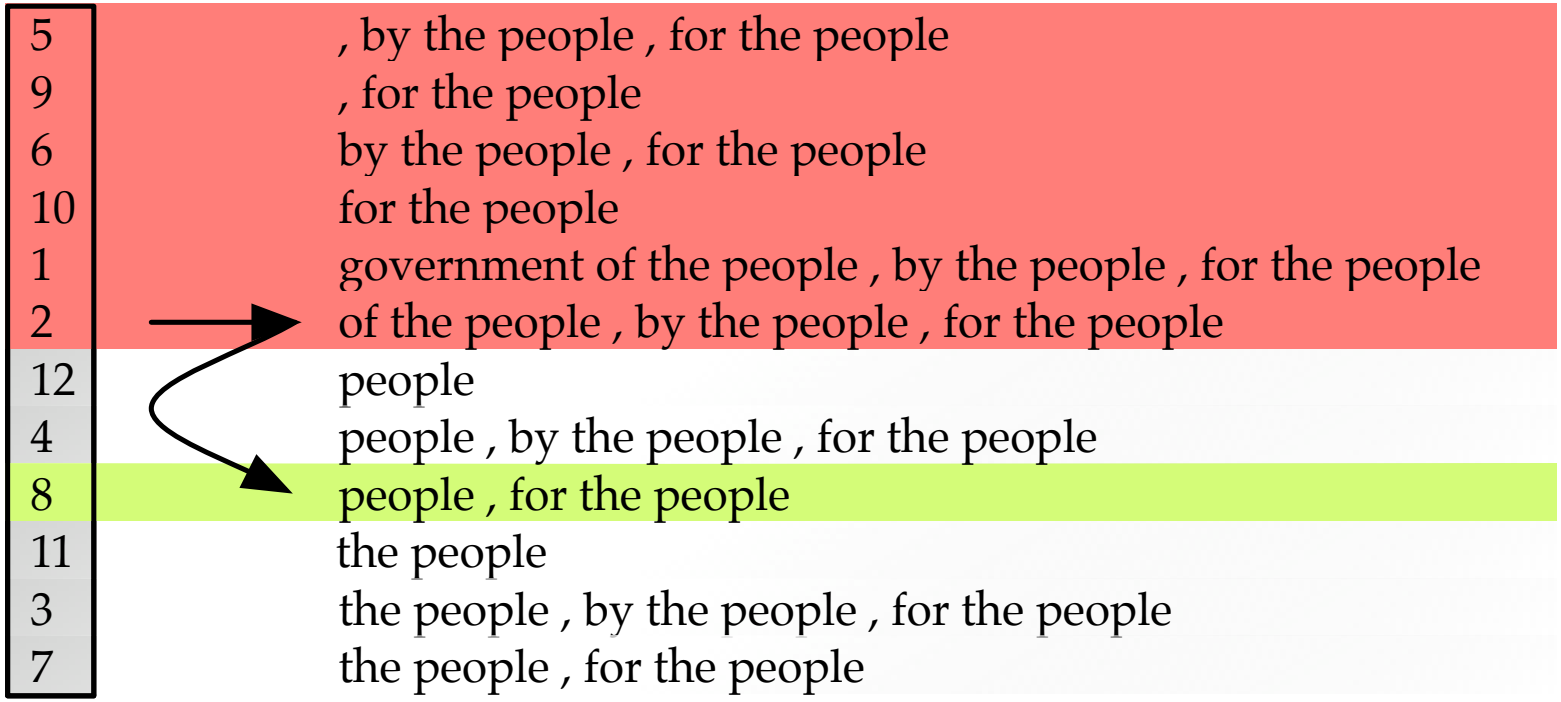
Querying the Suffix Array



Query: people

Binary search: middle of remaining space

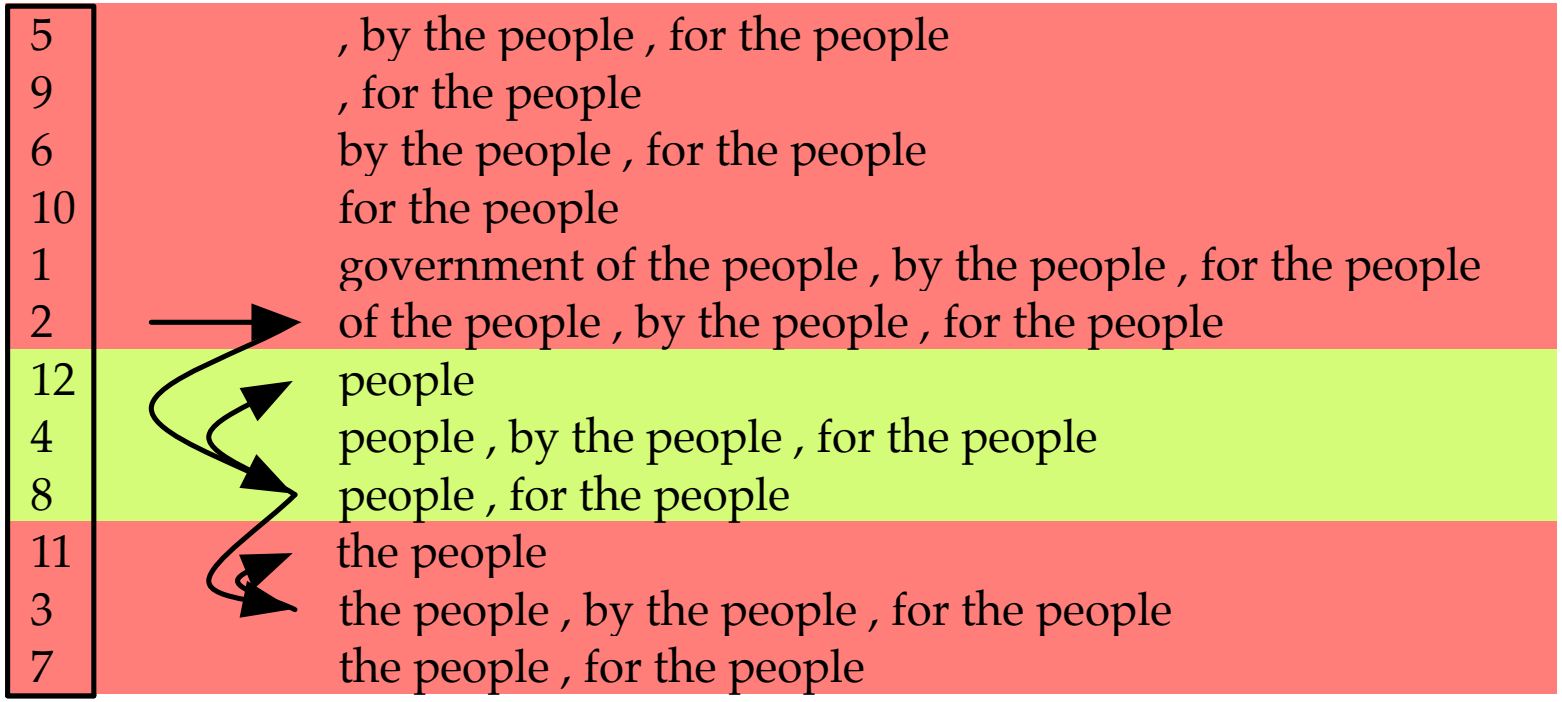
Querying the Suffix Array



Query: people

Binary search: match

Querying the Suffix Array



Query: people

Finding matching range with additional binary searches for start and end

Bias Towards User Translation

- Cache-based models

- Language model

  → give bonus to n-grams in previous user translation

- Translation model

  → give bonus to translation options in previous user translation

- Decaying score for bonus (less recent, less relevant)

- More details: Bertoldi et al. [MT Summit 2013]

integration of translation memories

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

TM	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)

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 .

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 deuxiéme 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 deuxiéme 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 deuxiéme 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 deuxiéme 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 deuxiéme 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 deuxiéme alinéa est supprimé .

- XML frame (input to Moses)

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

<xml translation=" , le texte du deuxiéme 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 deuxiéme alinéa est supprimé .

- More compact formalism for the purposes of this presentation:

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

Two Solutions

- XML frames

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

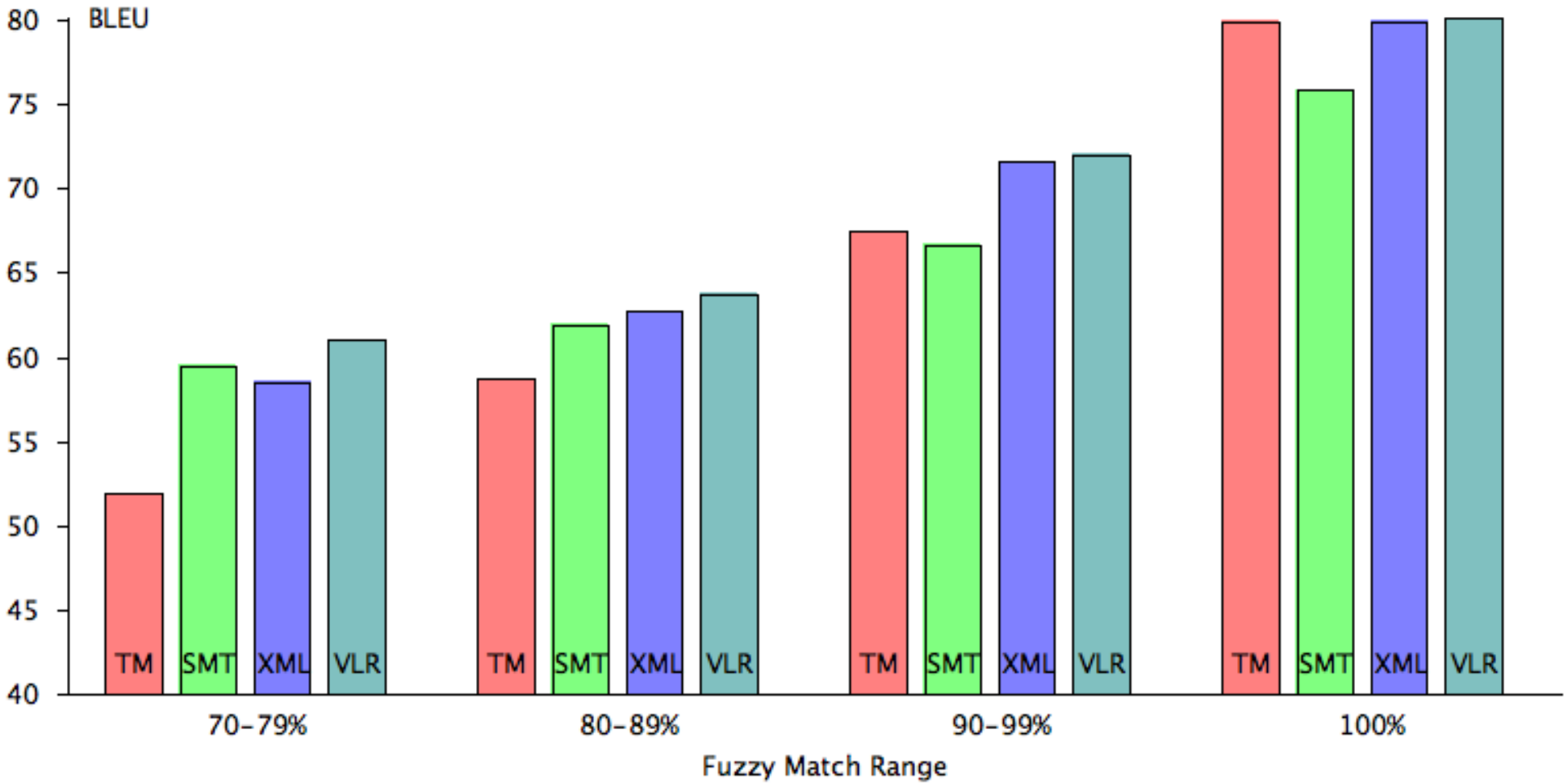
  for input

  The second paragraph of Article 21 is deleted .

- Very large hierarchical rule

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

Result: Acquis



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

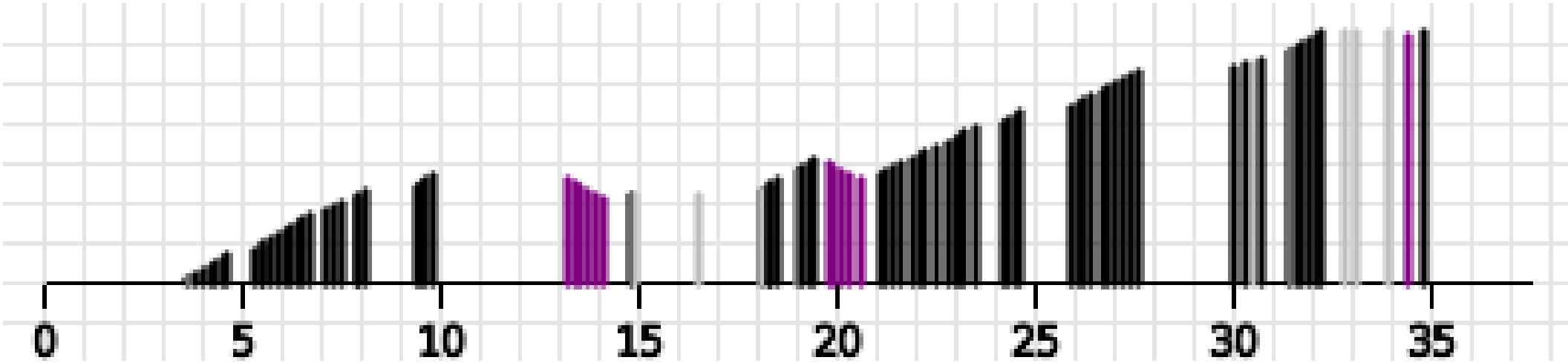
- In every event we save:
  - Type
  - In which element was produced
  - Time

- 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

$\Rightarrow$ 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.



(37.5 sec, 3.4 sec/word)

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)*

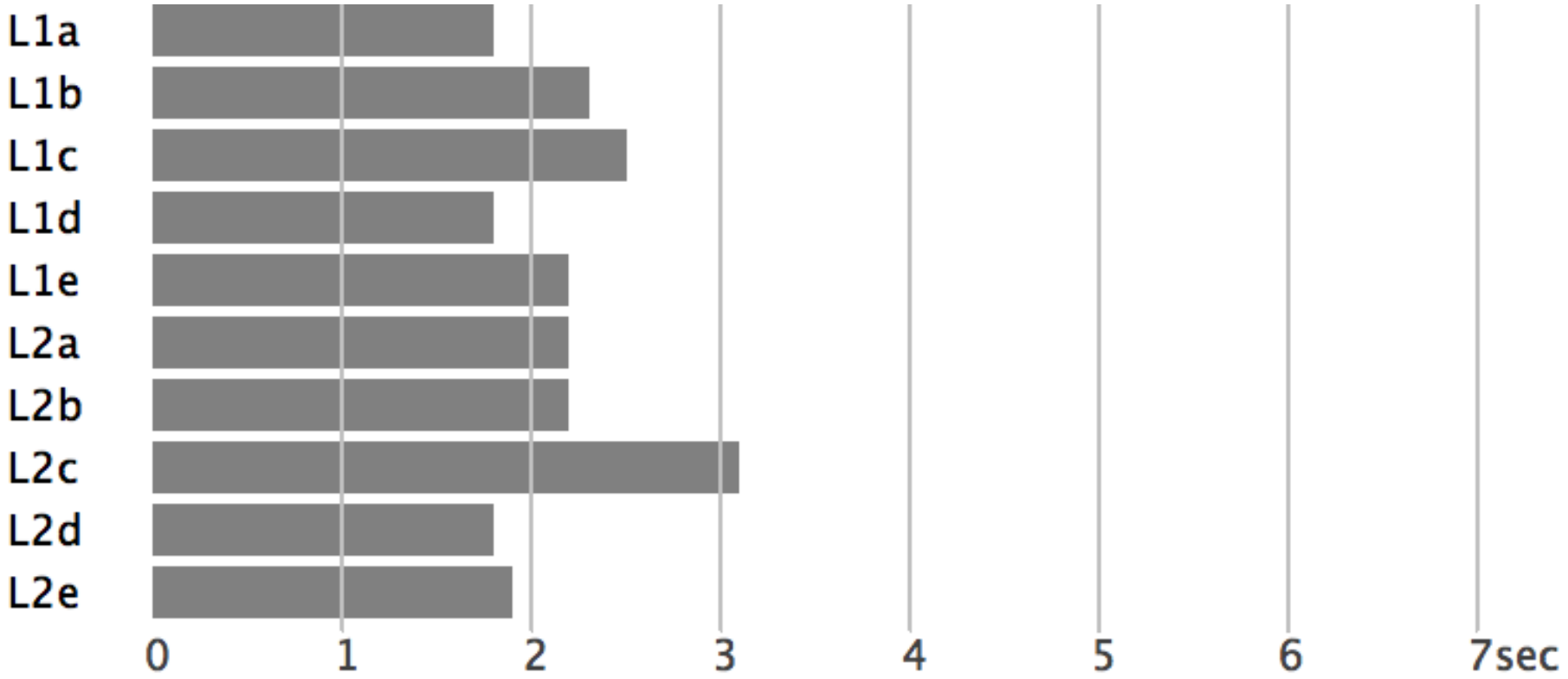
Main Measure: Productivity

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** **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%**

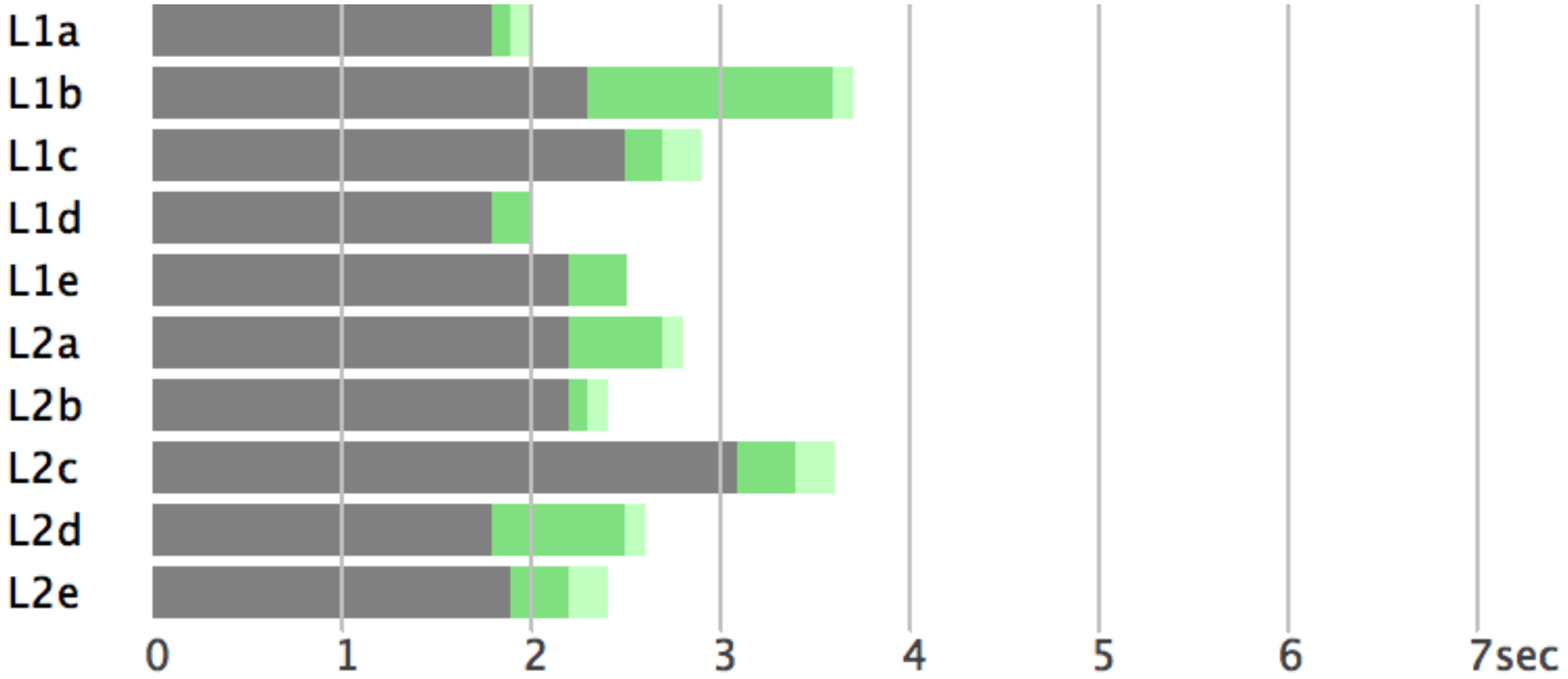
Unassisted Novice Translators



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

only typing

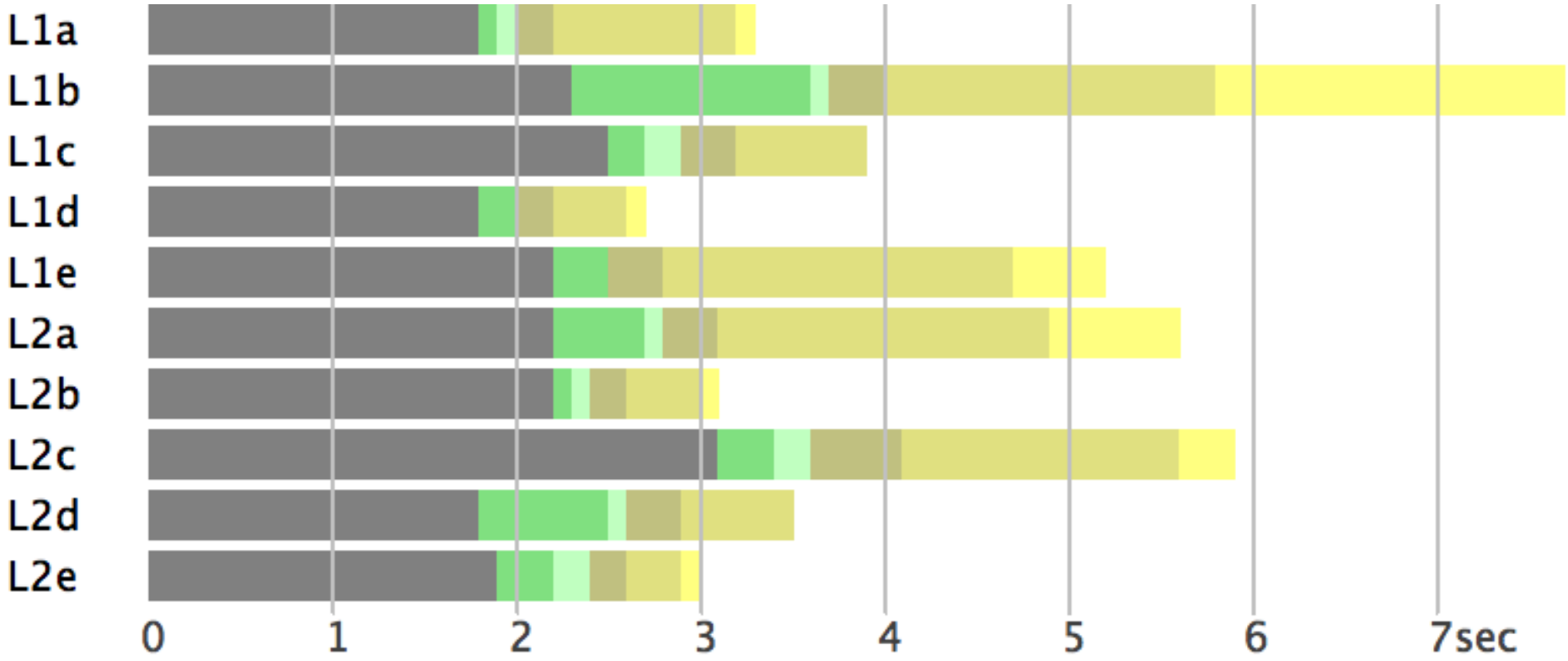
Unassisted Novice Translators



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%	-

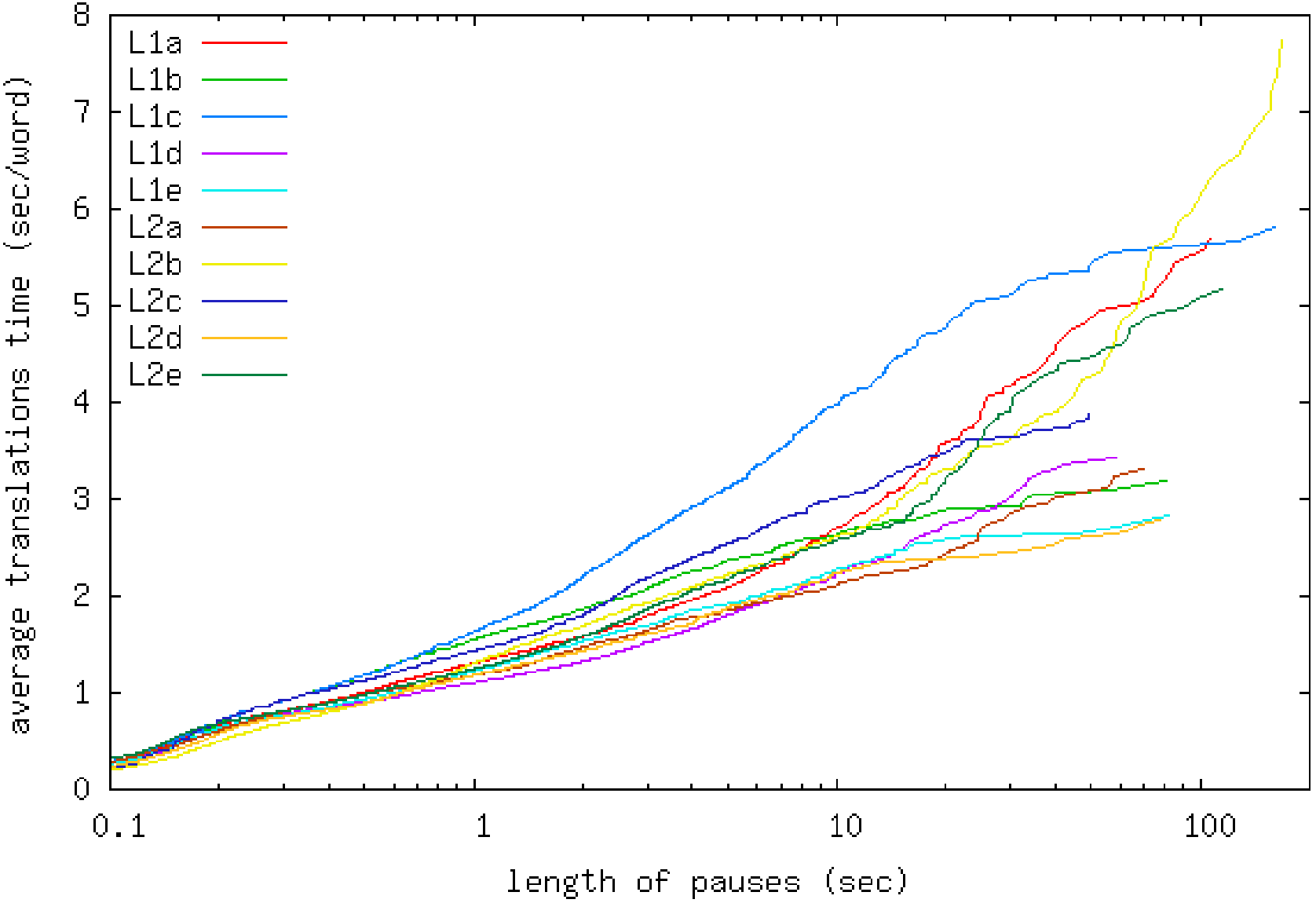
Translation comes to large degree from assistance

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

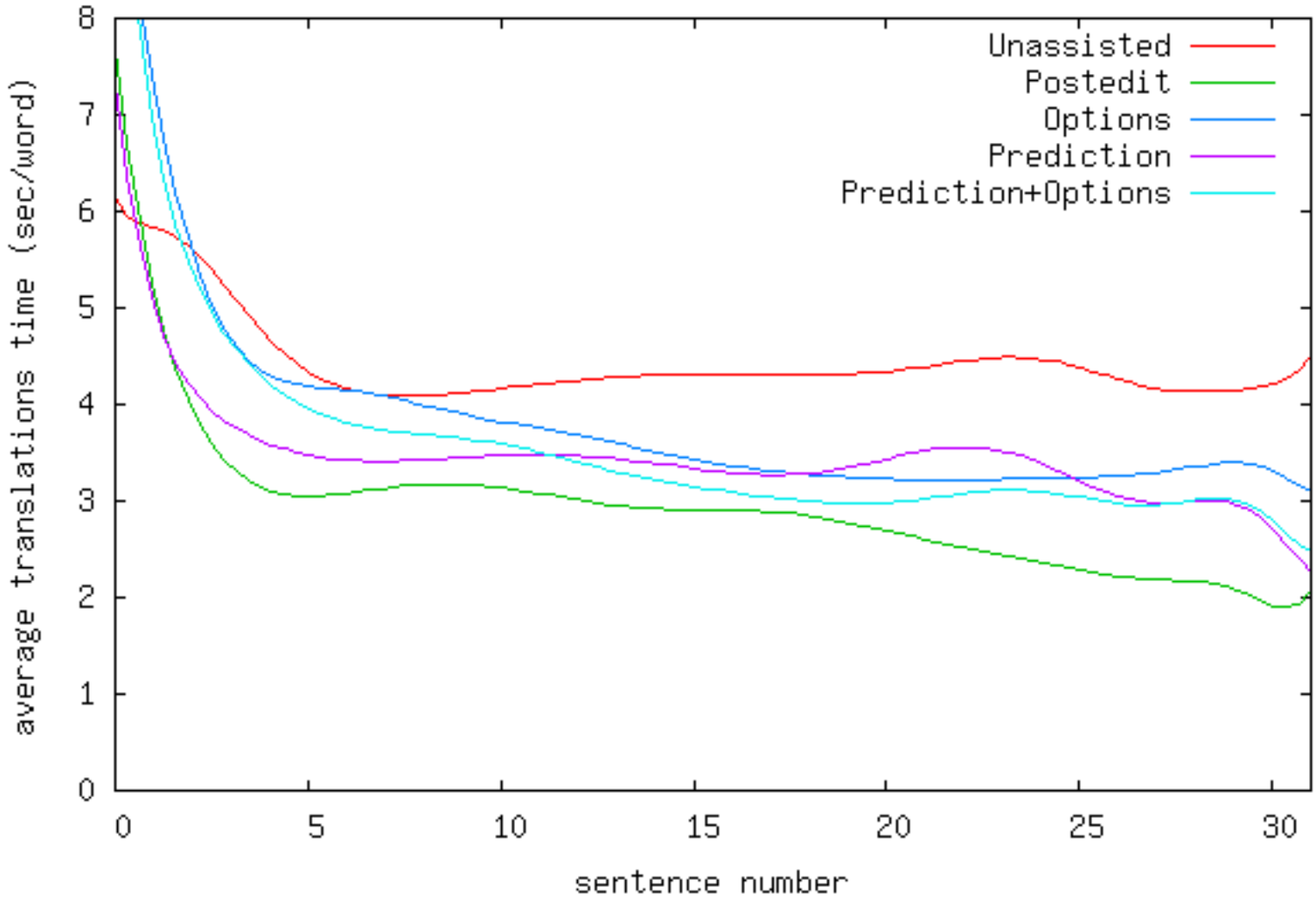
$$sum(t) = \frac{1}{Z} \sum_{p \in P, l(p) \leq t} l(p)$$

Results

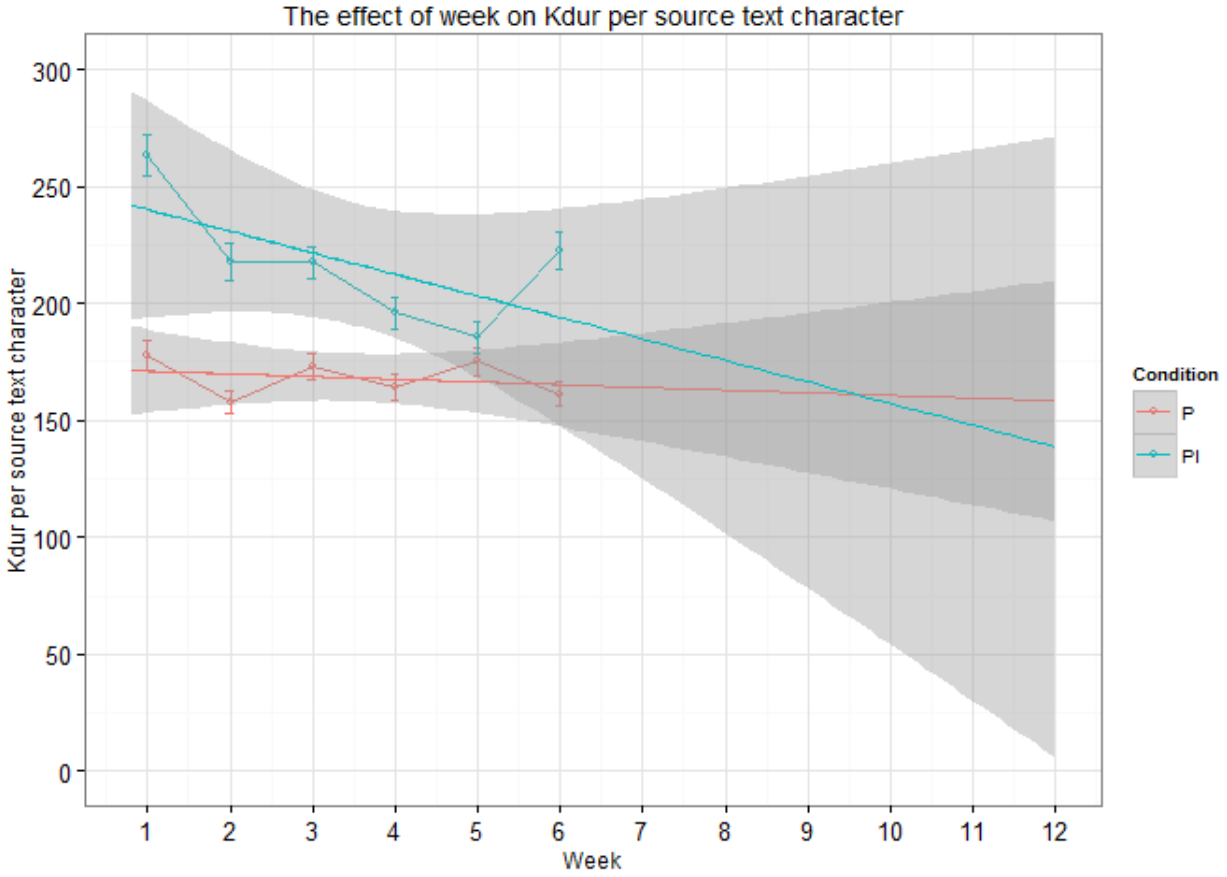

L1a
L1b
L1c
L1d
L1e
L2a
L2b
L2c
L2d
L2e
average translations time (sec/word)
length of pauses (sec)
0
1
2
3
4
5
6
7
8
0.1
1
10
100

Learning Effects

Users become better over time with assistance



Learning Effects: Professional Translators

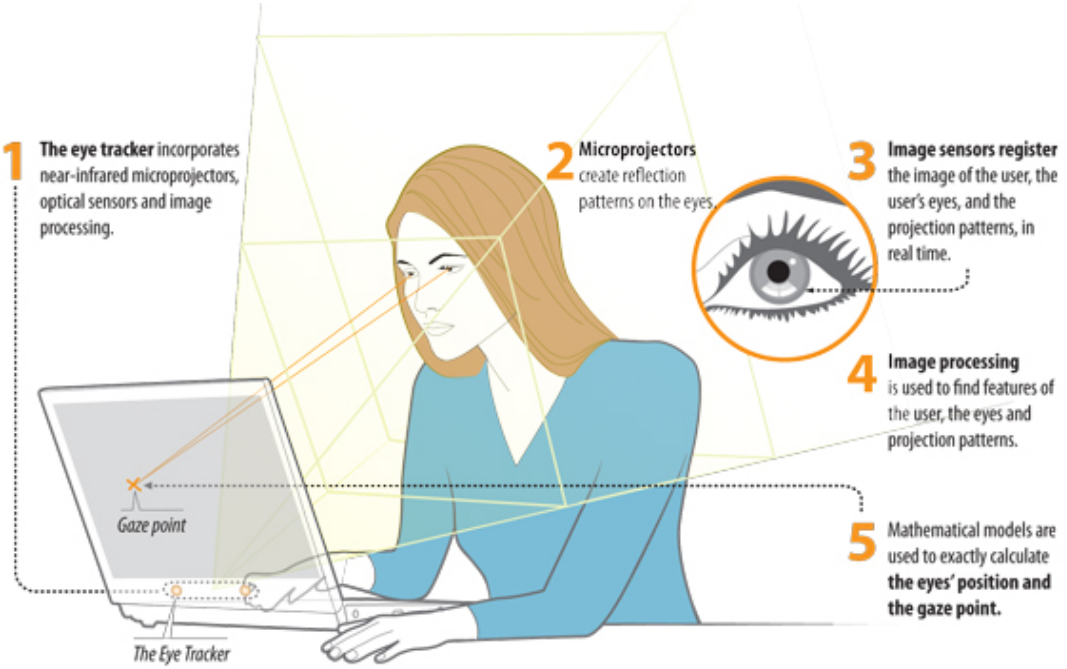


CASMACAT longitudinal study

Productivity projection as reflected in Kdur taking into account six weeks

(Kdur = user activity excluding pauses $>$ 5 secods)

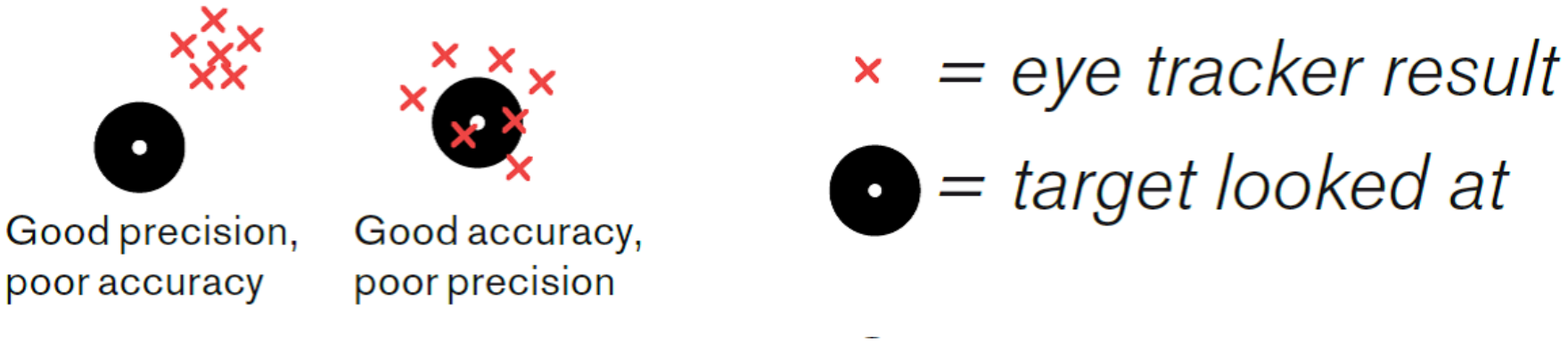
Eye Tracking



- 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

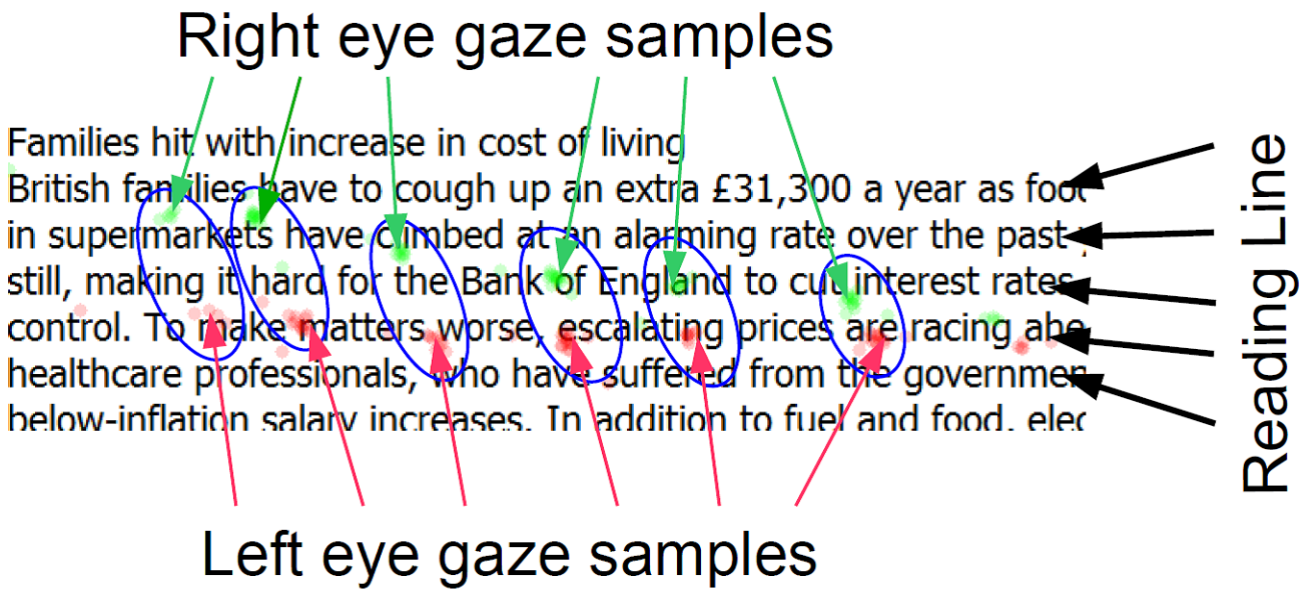
Eye Tracking

- Problem: Accuracy and precision of gaze samples

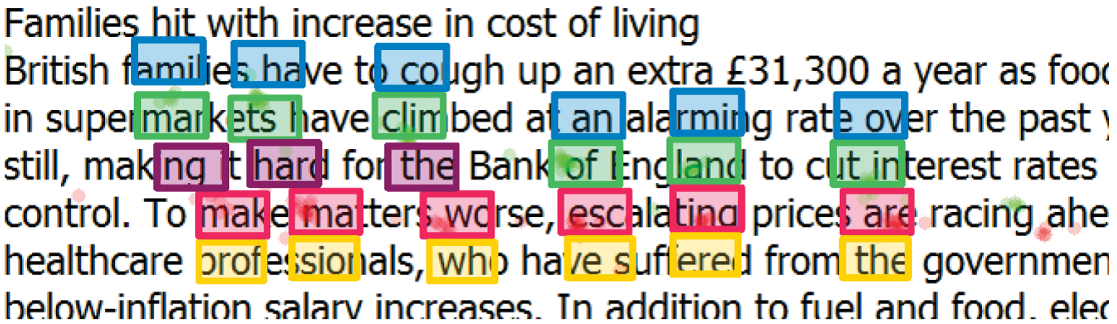


Gaze-to-Word Mapping

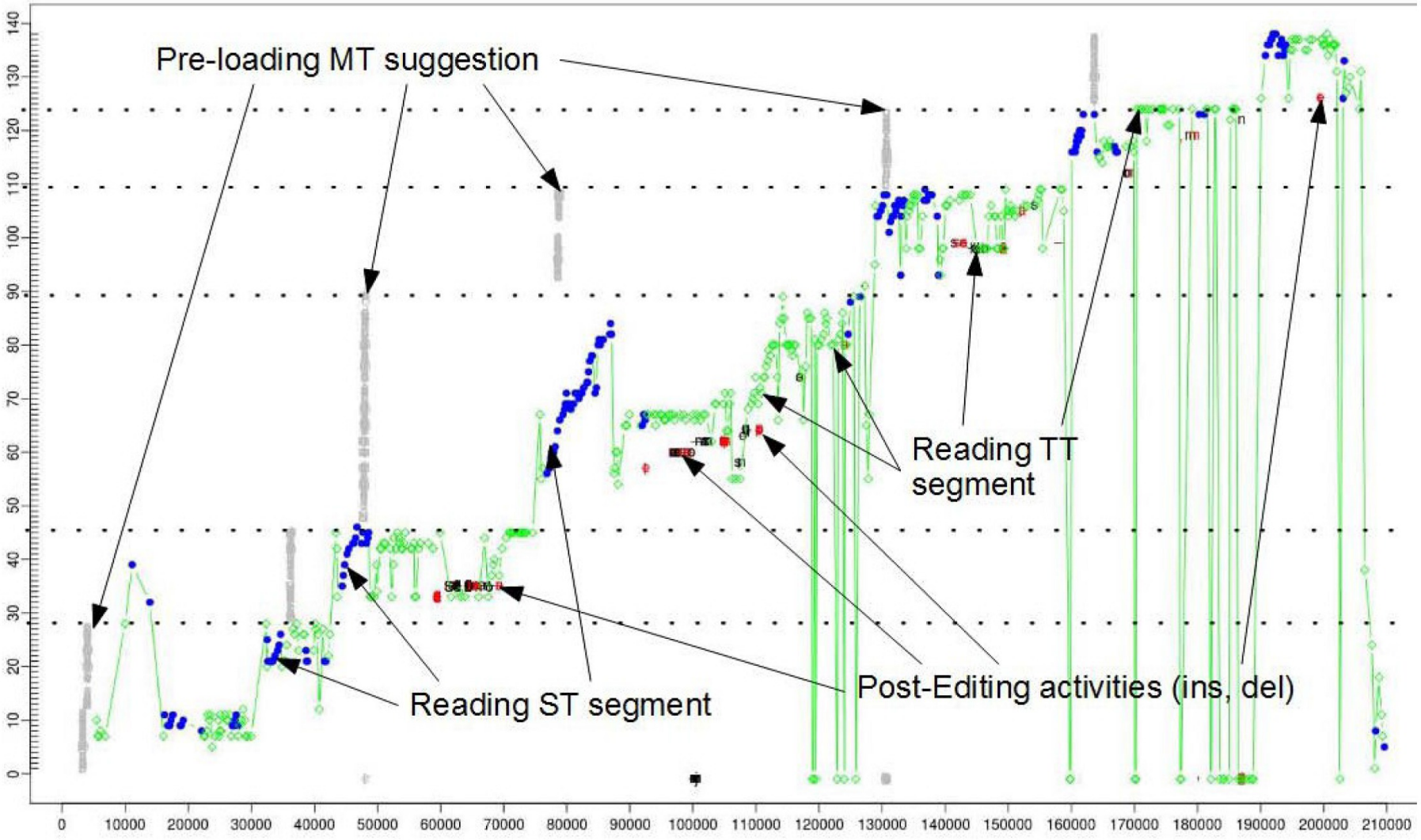
- Recorded gaze lacations and fixations



- Gaze-to-word mapping

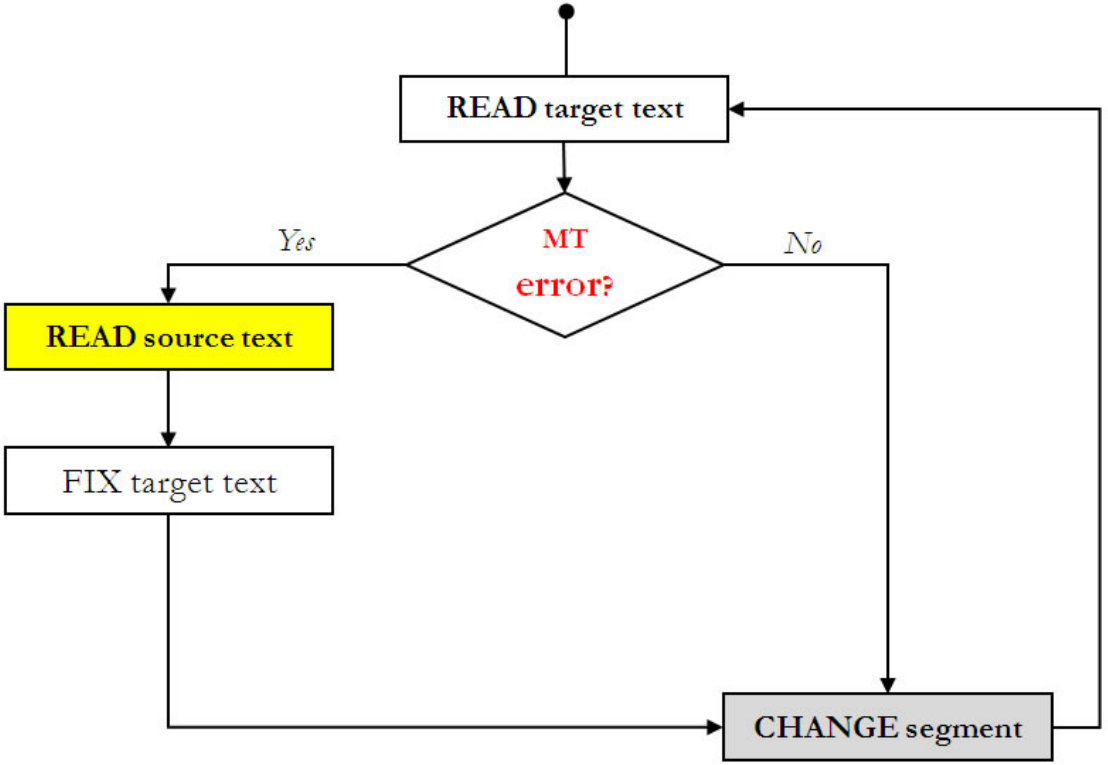


Logging and Eye Tracking



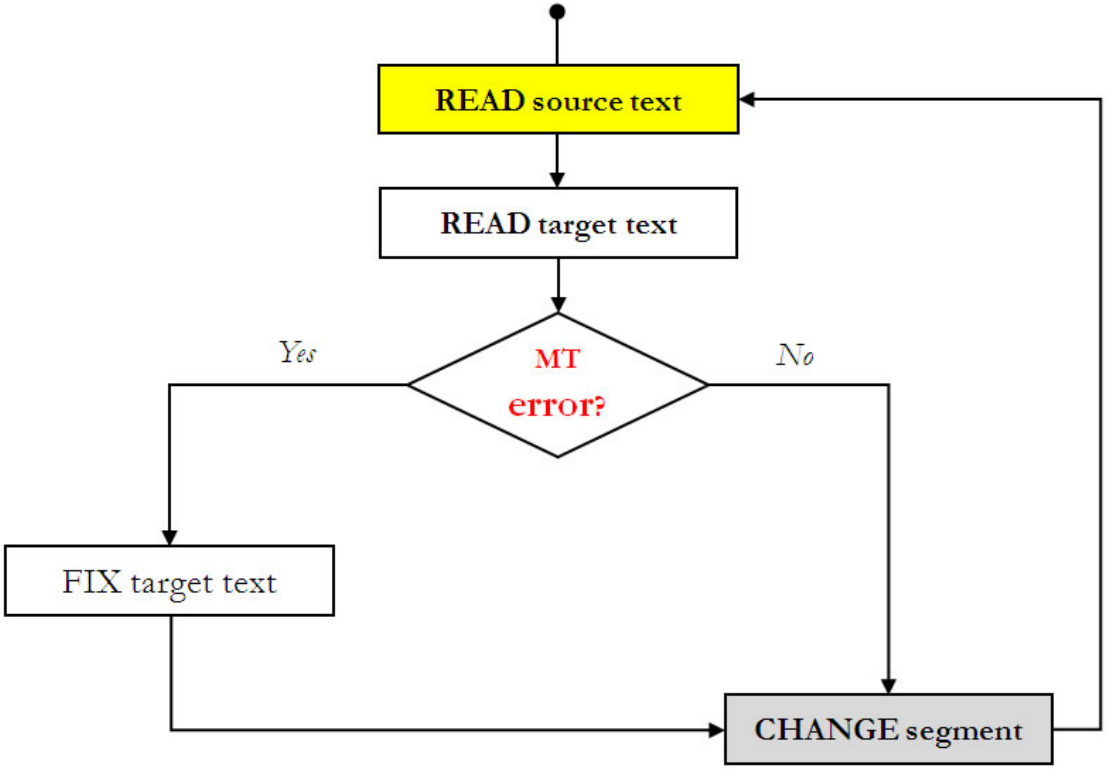
focus on target word (green) or source word (blue) at position x

Cognitive Studies: User Styles



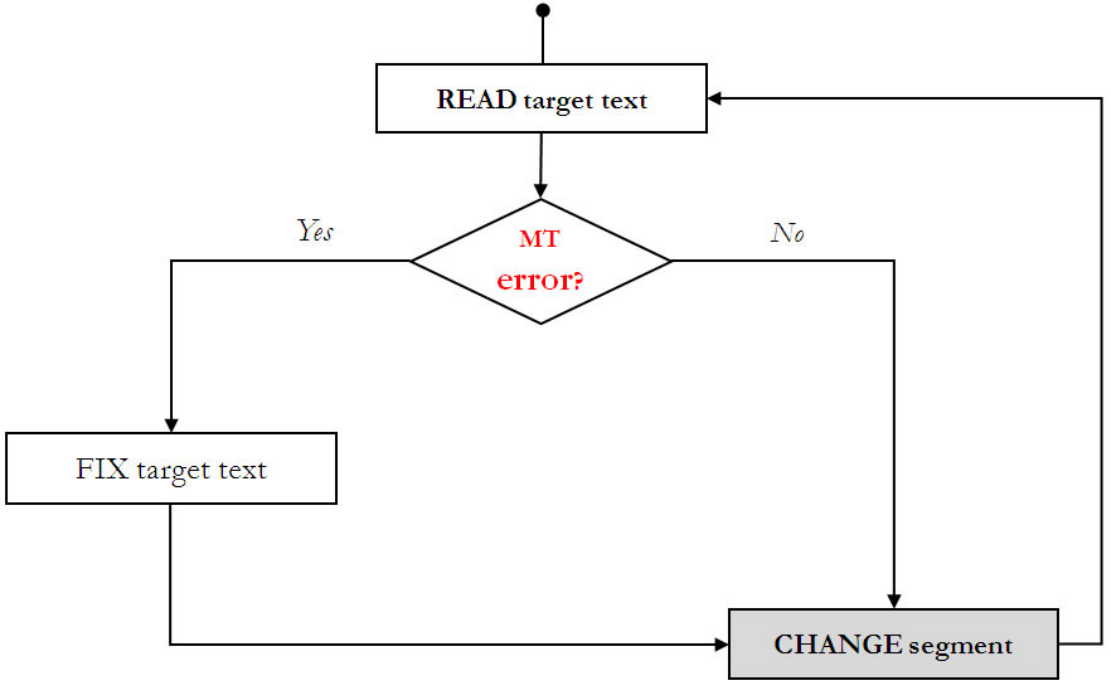
- User style 1: Verifies translation just based on the target text, reads source text to fix it

Cognitive Studies: User Styles



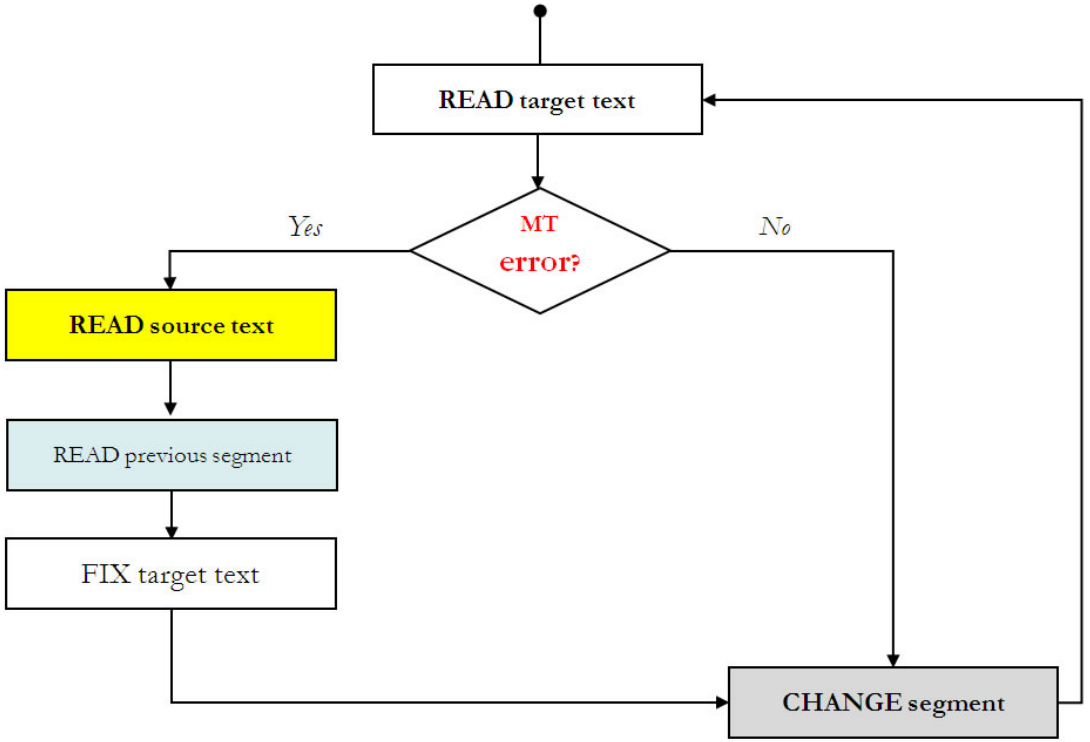
- User style 2: Reads source text first, then target text

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 assitance
  (P = post-editing, PI = interactive post-editing, PIA = interactive post-editing with additional annotations)

Backtracking

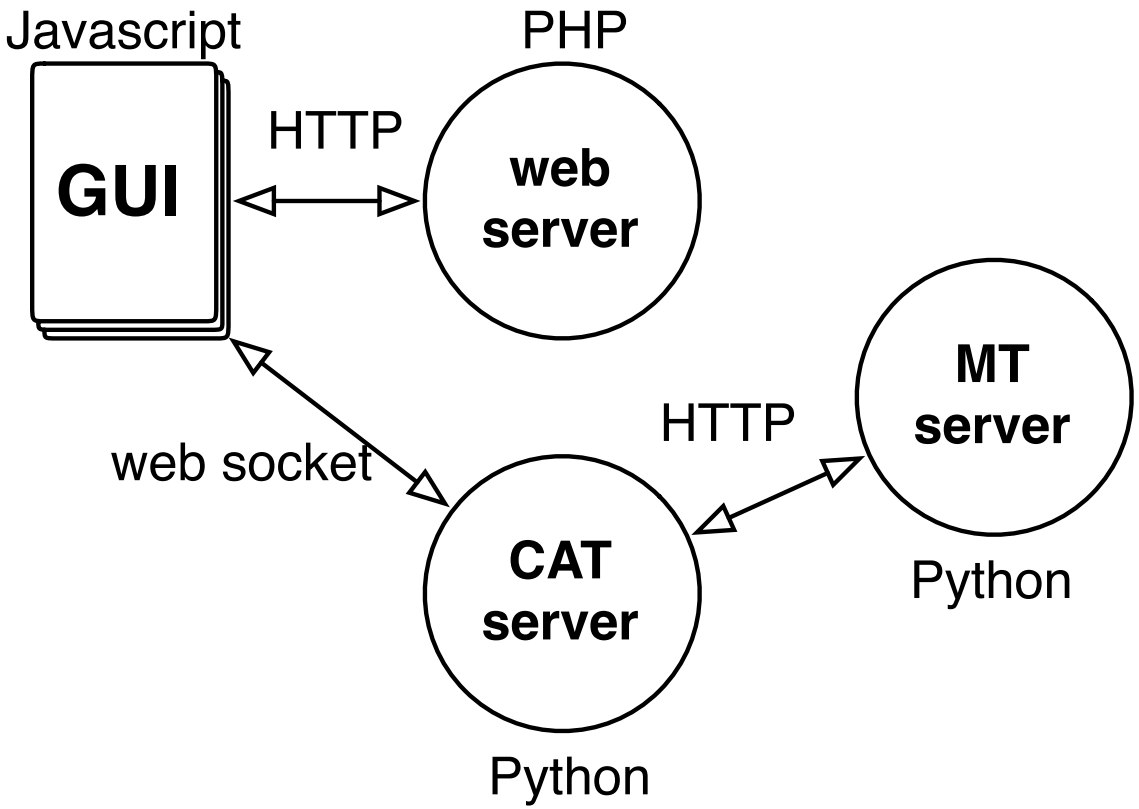
- Local backtracking
  - **Immediate repetition**: the user immediately returns to the same segment (e.g. AAAA)
  - **Local alternation**: user switches between adjacent segments, often singly (e.g. ABAB) but also for longer stretches (e.g. ABC-ABC).
  - **Local orientation**: very brief reading of a 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.

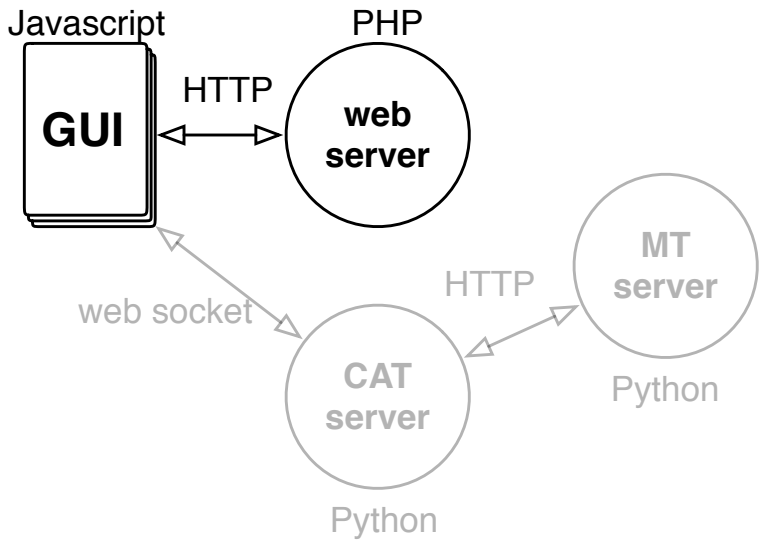
part III

CASMACAT workbench implementation

Components


Javascript
PHP
GUI
HTTP
web server
web socket
HTTP
MT server
CAT server
Python
Python

Web Server



- Builds on Matecat open source implementation
- Typical web application: LAMP (Linux, Apache, MySQL, PHP)
- Uses model, view, controller breakdown

Model

- Relevant data is stored in MySQL database `matecat_sandbox`
- Major database tables
  - Projects are stored in `projects`
  - They have a corresponding entry in `jobs`
  - Raw files (XLIFF) are stored in `files`
  - Segments are stored in `segments`
  - Translations of segments are stored in `segment_translations`
  - Log events are stored in `*_event`
  - etc.
- The major change from Matecat is the logging

Controller

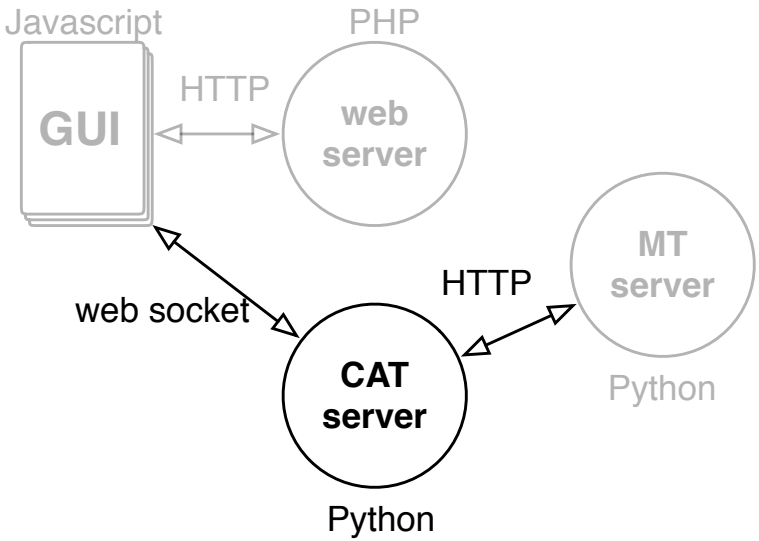
- Typical request: get information about a segment:
  `POST http://192.168.56.2:8000/?action=getSegments&time=1446185242727`

- Script `index.php` selects corresponding action in `lib/controller`
  e.g., `getSegmentsController.php`

- Response is HTML or JSON

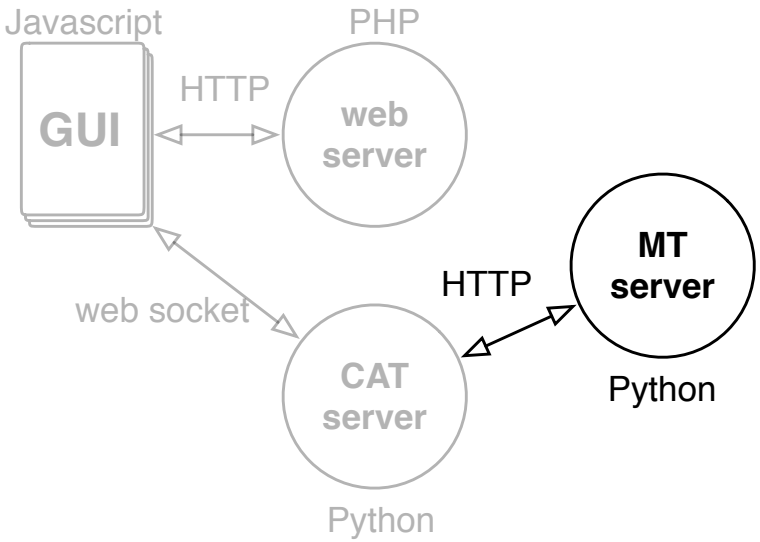
- The main action is really in the Javascript GUI `public/js`
  - core functionality from Matecat `public/js/cat.js`
  - CASMACAT extensions `public/js/casmacat`

CAT Server



- To a large degree middleware
- Calls external services such as
  - MT server
  - word aligner
  - interactive translation prediction
- Caches information about a sentence translation

MT Server



- Google-style API to MT Server
- Python wrapper for Moses
  - basic translation request
  - includes pre and post processing pipeline
  - other functions: word alignment, incremental updating, etc.
- Uses `mosesserver` XMLRPC server

server.py

- Requires `mosesserver` to run as a service

  ```
  mosesserver -config $MODELDIR/moses.ini --server-port 9010
  ```

- Script `server.py` requires a lot of parameters
  - preprocessing tools (tokenizer, truecaser, etc.)
  - IP address and port
  - URL of the `mosesserver` API
  - etc.

- Request to the script

  ```
  http://127.0.0.1:9000//translate?q=Un+test&key=0&source=xx&target=xx
  ```

- Response

  ```
  {"data": {"translations": [{"translatedText": "A test",
  "translatedTextRaw": "a test",
  "annotatedSource": "un test",
  "tokenization": {"src": [[0, 1], [3, 6]], "tgt": [[0, 0], [2, 5]]}}]}}
  ```

Home Edition

- Moses is installed in `/opt/moses`

- CASMACAT is installed in `/opt/casmacat`
  - web server / GUI in `/opt/casmacat/web-server`
  - MT server (server.py) in `/opt/casmacat/mt-server`
  - CAT server in `/opt/casmacat/cat-server`
  - installation scripts in `/opt/casmacat/install`
  - log files in `/opt/casmacat/logs`

- Home Edition
  - admin web server in `/opt/casmacat/admin`
  - corpus data in `/opt/casmacat/data`
  - prototype training in `/opt/casmacat/experiment`
  - engines stored in `/opt/casmacat/engines`

Home Edition MT Engine

- Demo engine in `/opt/casmacat/engines/fr-en-upload-1`
- Files

```
biconcor.1
biconcor.1.align
biconcor.1.src-vcb
biconcor.1.tgt
biconcor.1.tgt-vcb
corpus-1.binlm.1
fast-align.1
fast-align.1.log
fast-align.1.parameters
fast-align-inverse.1
fast-align-inverse.1.log
fast-align-inverse.1.parameters
info
moses.tuned.ini.1
phrase-table-mmsapt.1
reordering-table.1.wbe-msd-bidirectional-fe.minlexr
RUN
truecase-model.1.en
truecase-model.1.fr
```

- The script `RUN` starts the engine

Thank You

questions?