

---

# Advances in Computer Aided Translation Beyond Post-Editing

Philipp Koehn

31 October 2015



- Post-editing
- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- Interactive translation prediction
- Model adaptation
- Logging, eye tracking, and user studies
- CASMACAT Home Edition

# 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, Musharraf sent the Pakistani army into the tribal areas bordering Afghanistan for the first time since Pakistan's independence.

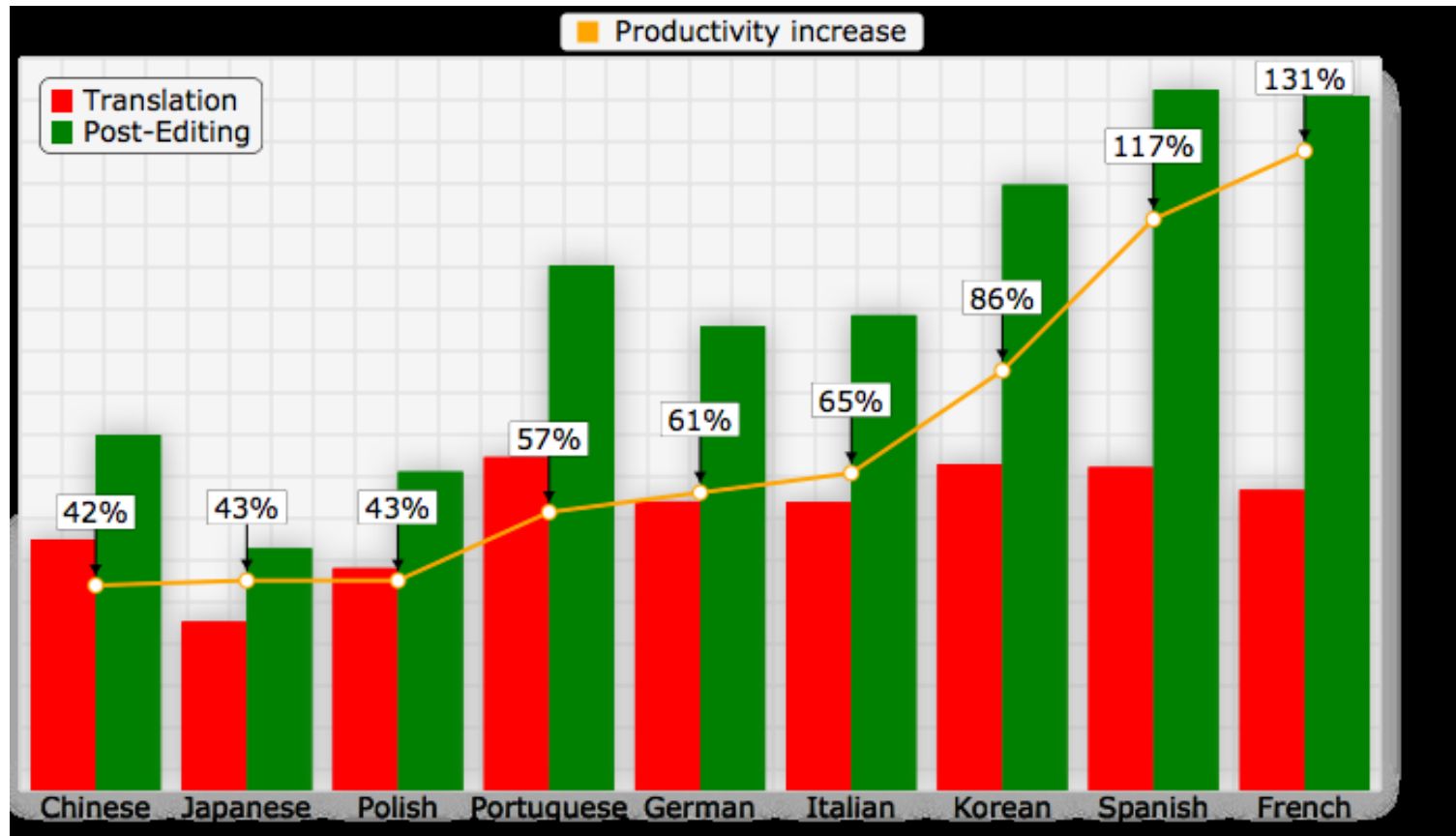
8 Les opérations contre les forces des Talibans et d'Al-Qaeda ont obtenu des résultats mitigés. >

visualization >>

ITP T→ DRAFT **TRANSLATED**

- Screenshot from CASMACAT post-editing mode (same as MATECAT)
- Source on left, translation on right / context above and below

# Productivity Improvements



(source: Autodesk)

# MT Quality and Productivity



- What is the relationship between MT Quality and Postediting Speed
- One study (English–German, news translation, non-professionals)

| System       | Speed      |           | Metric |        |
|--------------|------------|-----------|--------|--------|
|              | sec./word. | wrds./hr. | BLEU   | MANUAL |
| ONLINE-B     | 5.46       | 659       | 20.7   | 0.637  |
| UEDIN-SYNTAX | 5.38       | 669       | 19.4   | 0.614  |
| UEDIN-PHRASE | 5.45       | 661       | 20.1   | 0.571  |
| UU           | 6.35       | 567       | 16.1   | 0.361  |

# Translator Variability



- Translator differ in
  - ability to translate
  - motivation to fix minor translation
- High variance in translation time (again: non-professionals)

| Post-editor | Speed      |           |
|-------------|------------|-----------|
|             | sec./ wrd. | wrds./hr. |
| 1           | 3.03       | 1,188     |
| 2           | 4.78       | 753       |
| 3           | 9.79       | 368       |
| 4           | 5.05       | 713       |

- Post-editing
- **Richer information**
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- Interactive translation prediction
- Model adaptation
- Logging, eye tracking, and user studies
- CASMACAT Home Edition

# Word Alignment



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 **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, Musharraf 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. >

ITP T→ DRAFT TRANSLATED

- Caret alignment (green)
- Mouse alignment (yellow)



# 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

# 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    | , | <b>erupted</b>   | 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

# 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 | في | موعد | اقصاه | الاول | من | نيسان | @/ @ | ابريل |
|---------------------------------|---------------------------------|--------------------|----------------|-----------------------------|------------------------------|------------------------------|--------------|------------------|----------------|-----------------|-----------|------|--------|----|------|-------|-------|----|-------|------|-------|
| the                             | the us house of representatives | adopted            | thursday       | legally                     | calls for the withdrawal of  | combat troops                | us           | iraq             | in             | no later than   | the first | from | april  |    |      |       |       |    |       |      |       |
| the us house of representatives | the                             | thursday ,         | law            |                             | the fighting forces          | the us                       | from iraq    |                  | the latest     | the first of    | april     |      |        |    |      |       |       |    |       |      |       |
| the us house                    | adopted the                     | thu                | the legally    |                             | fighting forces              | us                           | from iraq in |                  |                | i               | april     |      |        |    |      |       |       |    |       |      |       |
| it was                          | us house of representatives     | was adopted        | thursday , the | the law                     | demands withdrawal of troops | fighter                      | the us       |                  | no later than  | first           | on april  |      |        |    |      |       |       |    |       |      |       |
| he was                          | the us house                    | adopted by         | thursday 's    | a law                       | calls for withdrawal of      | combat forces                | of           | in the           | not later than | first of        |           |      |        |    |      |       |       |    |       |      |       |
| he                              | us house                        | adopted by the     | on thursday    | a legally                   | calls for the withdrawal     | forces                       | the fighter  | from             |                |                 |           |      |        |    |      |       |       |    |       |      |       |
| earlier ,                       |                                 | us                 | adopted a      | on thursday ,               | by law                       | demands the withdrawal of    | troops       | iraq             |                |                 |           |      |        |    |      |       |       |    |       |      |       |
| was                             |                                 |                    | , was adopted  | thursday the                | legally ,                    | demands withdrawal of        |              | of the           |                |                 |           |      |        |    |      |       |       |    |       |      |       |
| it was the                      |                                 |                    | adopted ,      | thu ,                       | the legal                    | calls for withdrawal         |              | from iraq in the |                |                 |           |      |        |    |      |       |       |    |       |      |       |
| earlier , the                   |                                 |                    | adopted , the  | thursday , a                | legally @-@                  | demands the withdrawal       |              | the american     |                | by the first of |           |      |        |    |      |       |       |    |       |      |       |
| 2008                            | متحديا                          | مرة                | جديدة          | الرئيس                      | جورج                         | بوش                          | 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 |              |        |

# Bilingual Concordancer



TIP [Menu] T→ DRAFT TRANSLATED

abandonner

**abandon**

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

**give up**

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

**to**

|                       |                   |                        |  |                       |           |                        |
|-----------------------|-------------------|------------------------|--|-----------------------|-----------|------------------------|
| t ne veulent donc pas | <b>abandonner</b> | leurs prérogatives dar |  | olicy and do not want | <b>to</b> | delegate this prerogat |
|-----------------------|-------------------|------------------------|--|-----------------------|-----------|------------------------|

**to abandon**

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



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**.



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* · [...]

© Linguee Dictionary, 2015

Wikipedia

External sources (not reviewed)

|   |   |
|---|---|
| The implementing provisions applicable to the machine translation system would have to be established by the Select Committee [...] | Die Durchführungsbestimmungen für das System der maschinellen Übersetzung müssten vom engeren Ausschuss des EPO-Verwaltungsrats [...] |
|---|---|

|   |  |
|---|--|
| By user licence agreements relating to the SYSTRAN machine translation software program concluded between the applicants' [...] | Durch Lizenzverträge über die Benutzung der Software für maschinelle Übersetzungen SYSTRAN zwischen den Rechtsvorgängern der Klägerinnen [...] |
|---|--|

|   |  |
|---|--|
| [...] curriculum vitae, in forms suitable for multilingual machine translation, without restricting a user's option of adding other [...] | [...] standardisierten Lebenslauf zu prüfen, die für eine automatische Übersetzung in mehrere Sprachen geeignet sind, wobei der Nutzer [...] |
|---|--|

# Verification of Terminology

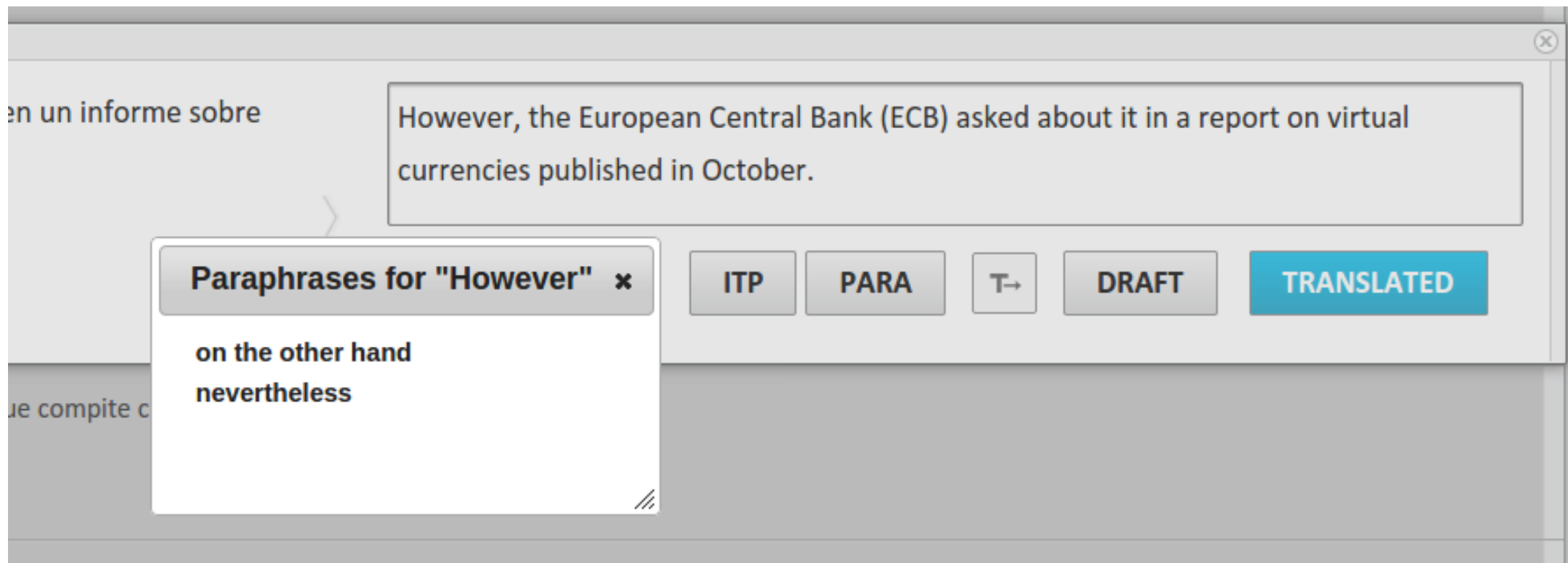
- Translation of German *Windkraft*

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

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

# Paraphrasing

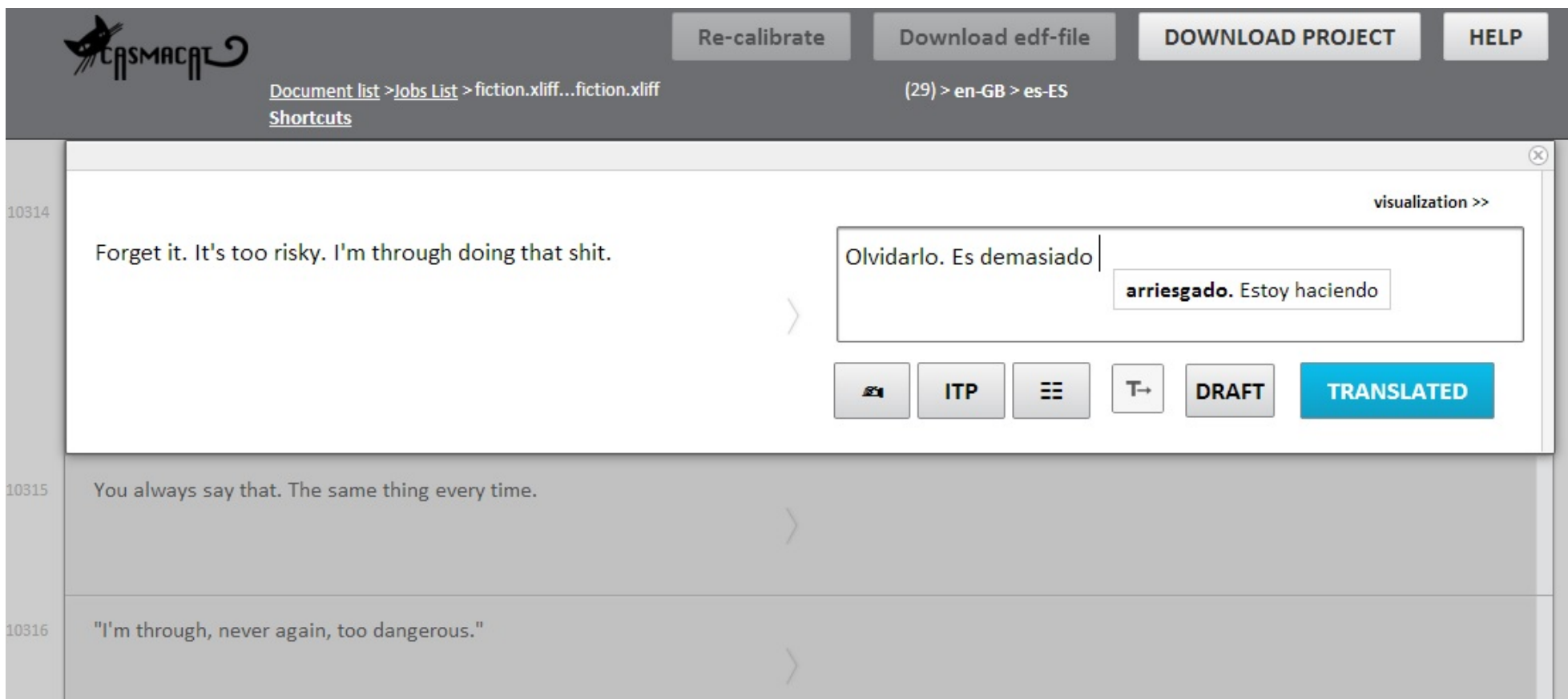
18



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

- Post-editing
- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- **Interactive translation prediction**
- Model adaptation
- Logging, eye tracking, and user studies
- CASMACAT Home Edition

# Interactive Translation Prediction



The screenshot displays the CASMACAT web interface. At the top left is the CASMACAT logo. The top navigation bar includes buttons for "Re-calibrate", "Download edf-file", "DOWNLOAD PROJECT", and "HELP". Below the navigation bar, the breadcrumb path is "Document list > Jobs List > fiction.xliff...fiction.xliff" and the language pair is "(29) > en-GB > es-ES". There are also links for "Shortcuts" and "visualization >>".

The main content area shows a list of text segments. The first segment, labeled "10314", contains the text "Forget it. It's too risky. I'm through doing that shit." To its right, a translation prediction box shows "Olvidarlo. Es demasiado" with a dropdown menu open displaying "arriesgado. Estoy haciendo". Below this box are several buttons: a flag icon, "ITP", a menu icon, "T→", "DRAFT", and a prominent blue "TRANSLATED" button. The subsequent segments, labeled "10315" and "10316", are dimmed and show the text "You always say that. The same thing every time." and "I'm through, never again, too dangerous." respectively.

# Shade Off Translated



L'intervention israélienne dans la bande de Gaza et les bombardements américains en Irak pour lutter contre les djihadistes de l'État islamique en Irak et au Levant ont également ajouté de la nervosité sur les marchés.

Israeli intervention in the Gaza Strip and the

American bombing in

- Word alignment visualization for interactive translation prediction
- Shade off words that are already translated
- Highlight words aligned to first predicted translation word

# 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 themselves cannot **E**

**D** Dans le cadre de l'Institut Jedlička, nous transférerons ce projet dans un nouveau bâtiment. Les étudiants eux-mêmes ne peuvent pas se rendre à des cours, nous essayons de les aider de cette manière.

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 à apprendre.

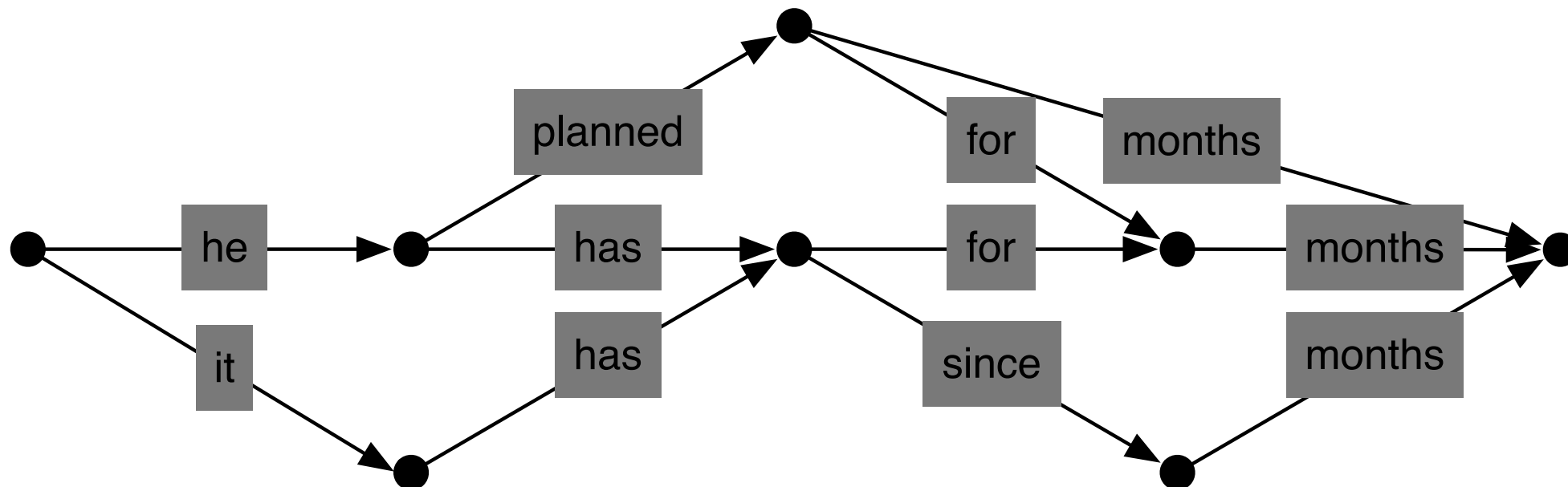
Teachers regularly visit Jedlička's activities and help them and their students.

Les étudiants eux-mêmes ne peuvent pas se rendre à des cours, nous essayons de les aider de cette manière.

are regularly

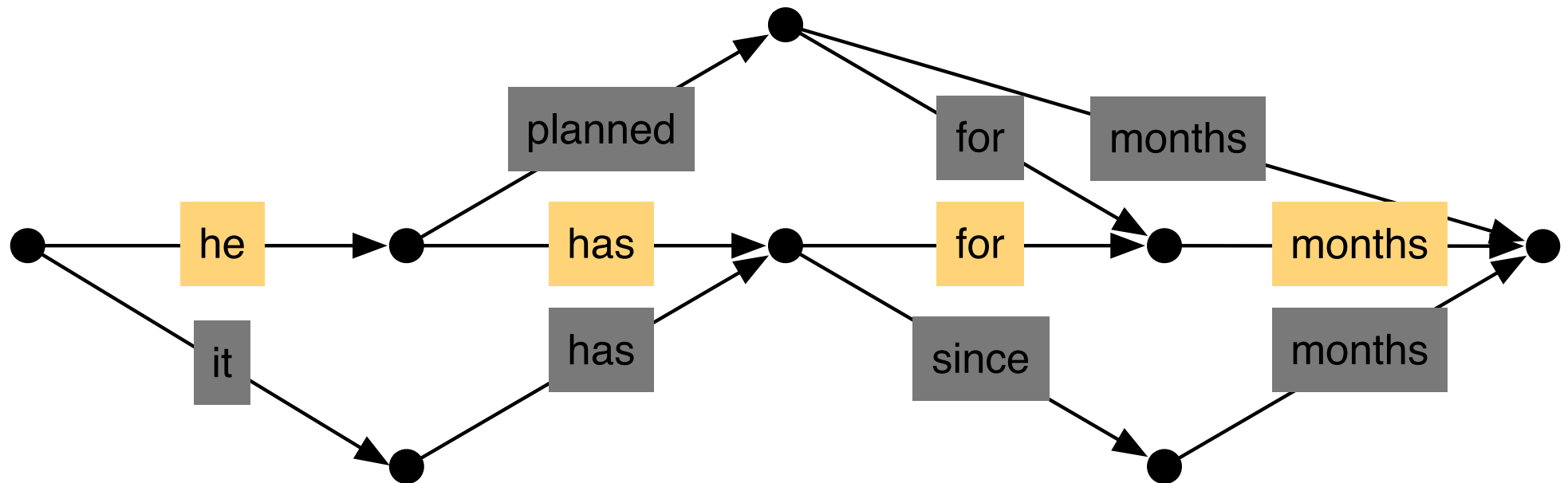


# Prediction from Search Graph



Search for best translation creates a graph of possible translations

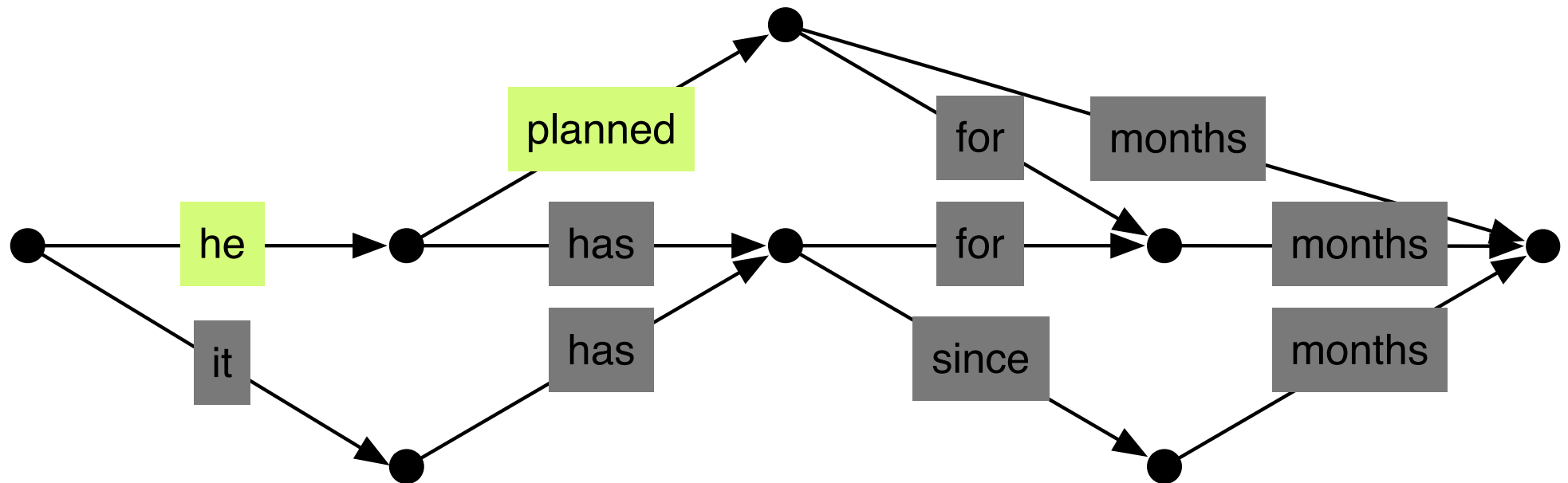
# Prediction from Search Graph



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

This path is suggested to the user

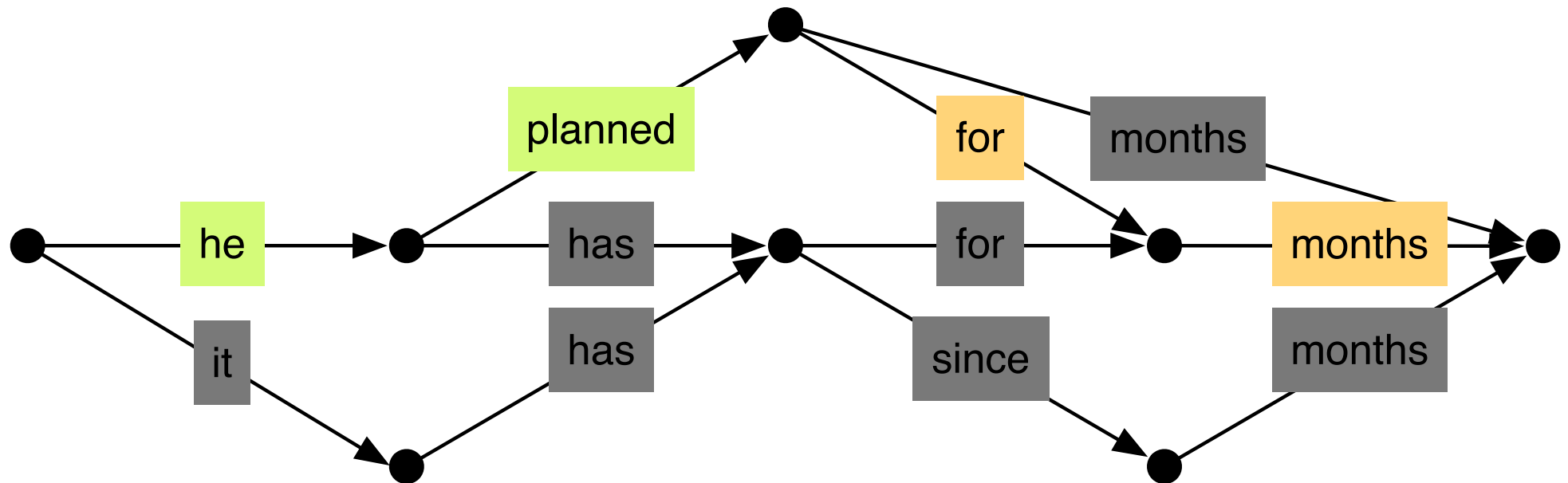
# Prediction from Search Graph



The user may enter a different translation for the first words

We have to find it in the graph

# Prediction from Search Graph



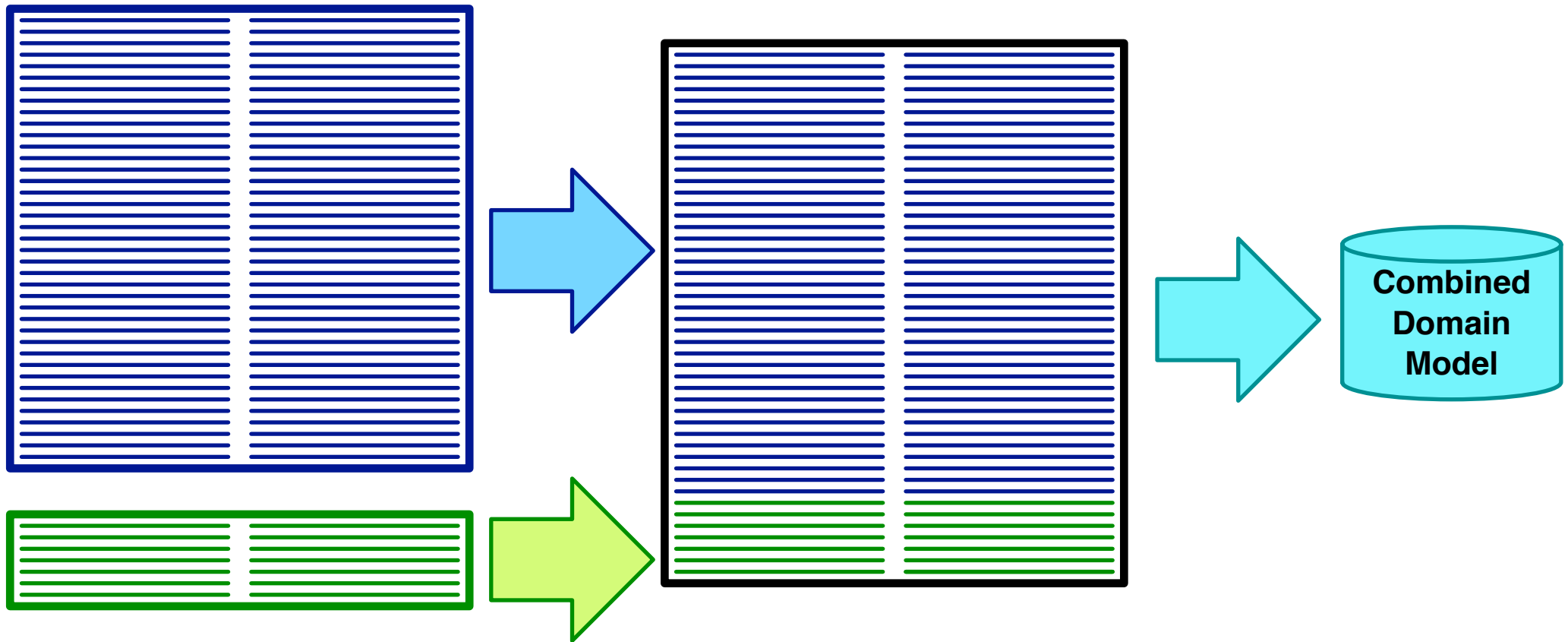
We can predict the optimal completion (according to the model)

- Post-editing
- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- Interactive translation prediction
- **Model adaptation**Interactive translation prediction
- Logging, eye tracking, and user studies
- CASMACAT Home Edition

# Adaptation

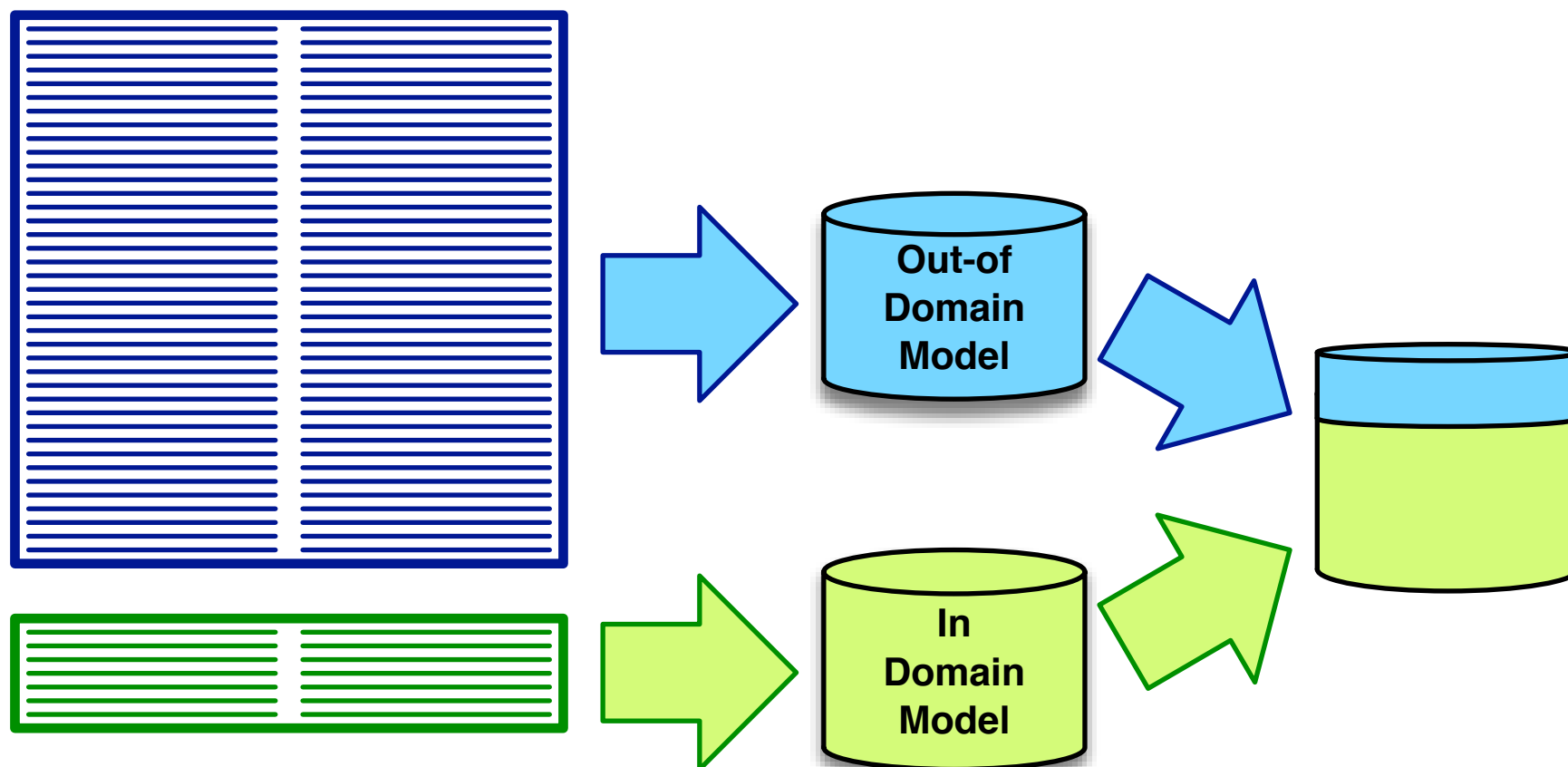
- Machine translation works best if optimized for domain
- Typically, large amounts of out-of-domain data available
  - European Parliament, United Nations
  - unspecified data crawled from the web
- Little in-domain data (maybe 1% of total)
  - information technology data
  - more specific: IBM's user 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

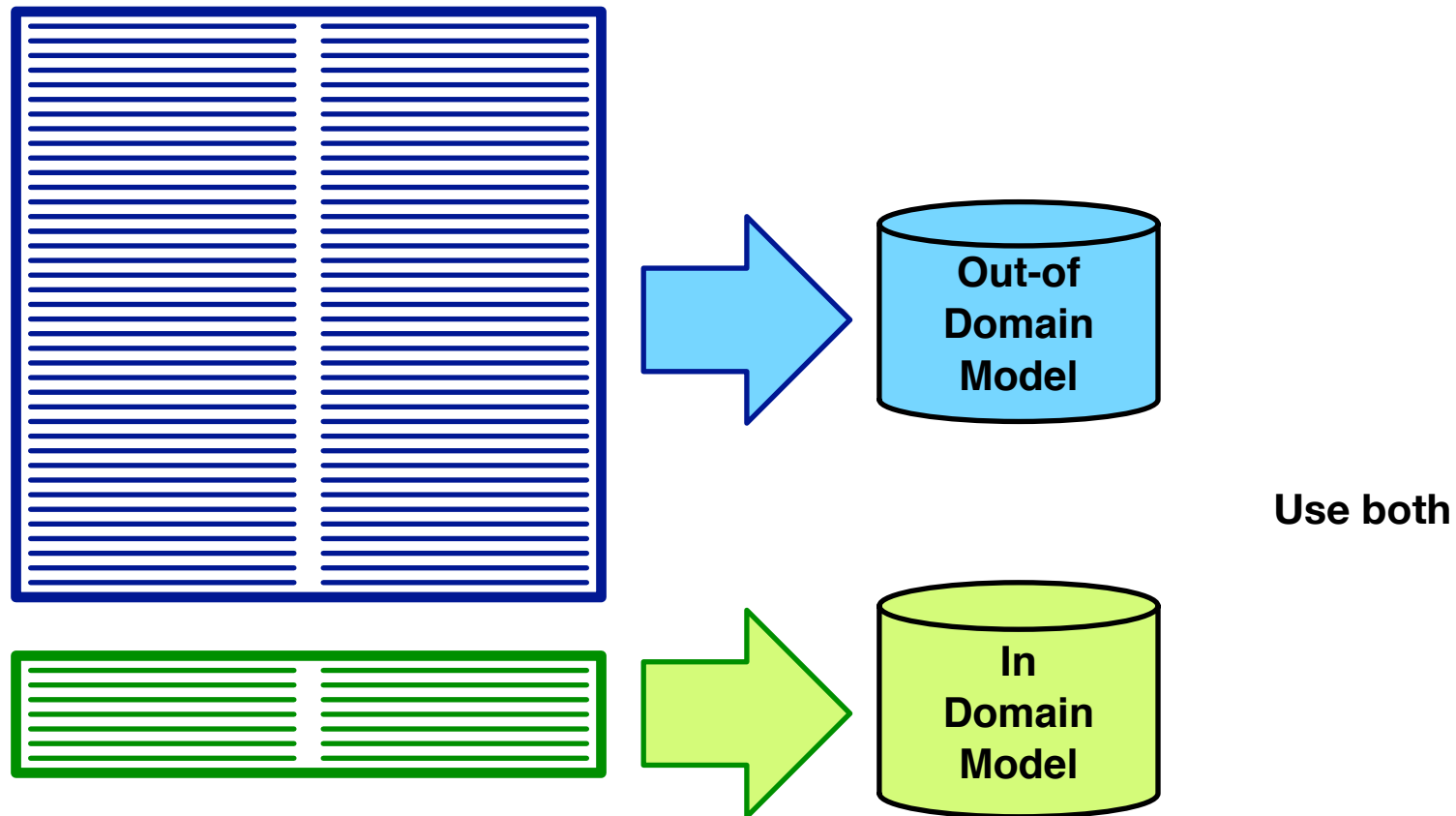
# Interpolate Models



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

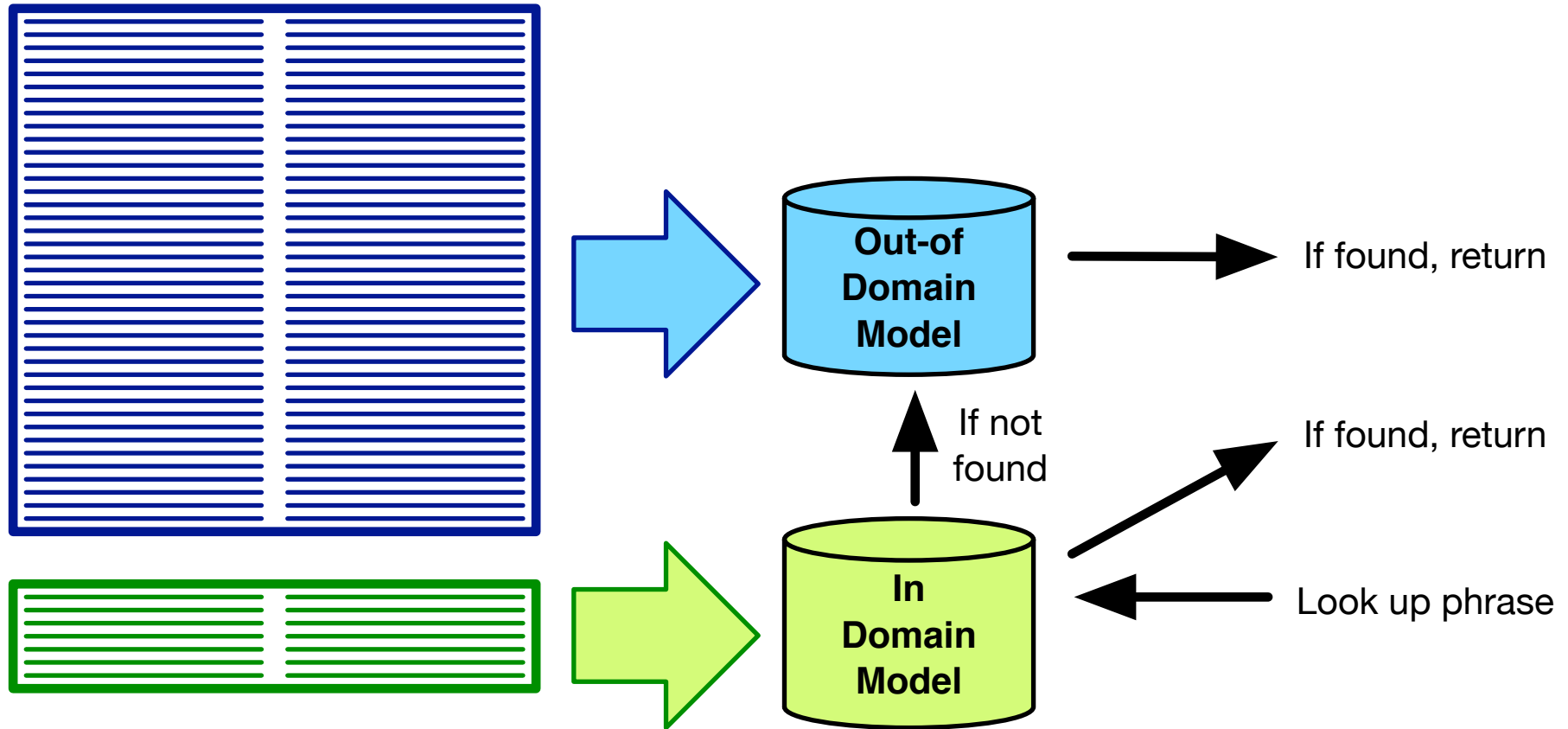


# Multiple Models

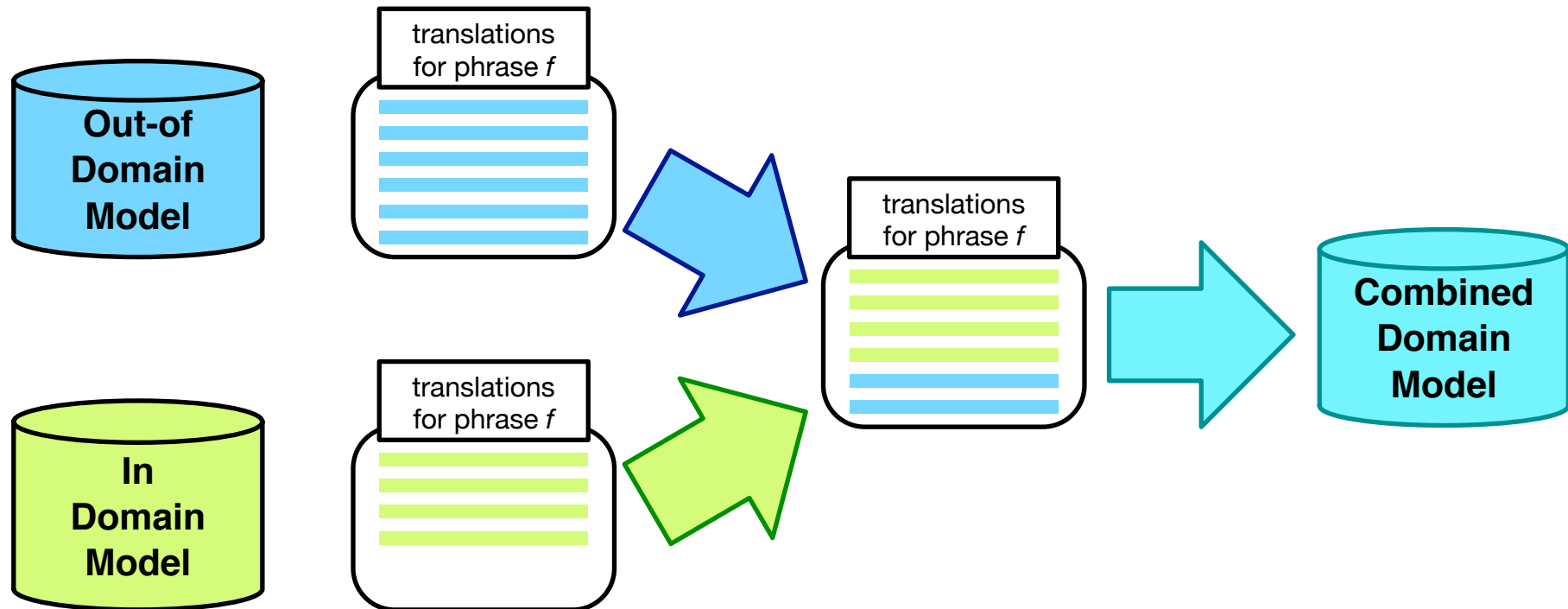


- Multiple models  $\rightarrow$  multiple feature functions

# Backoff

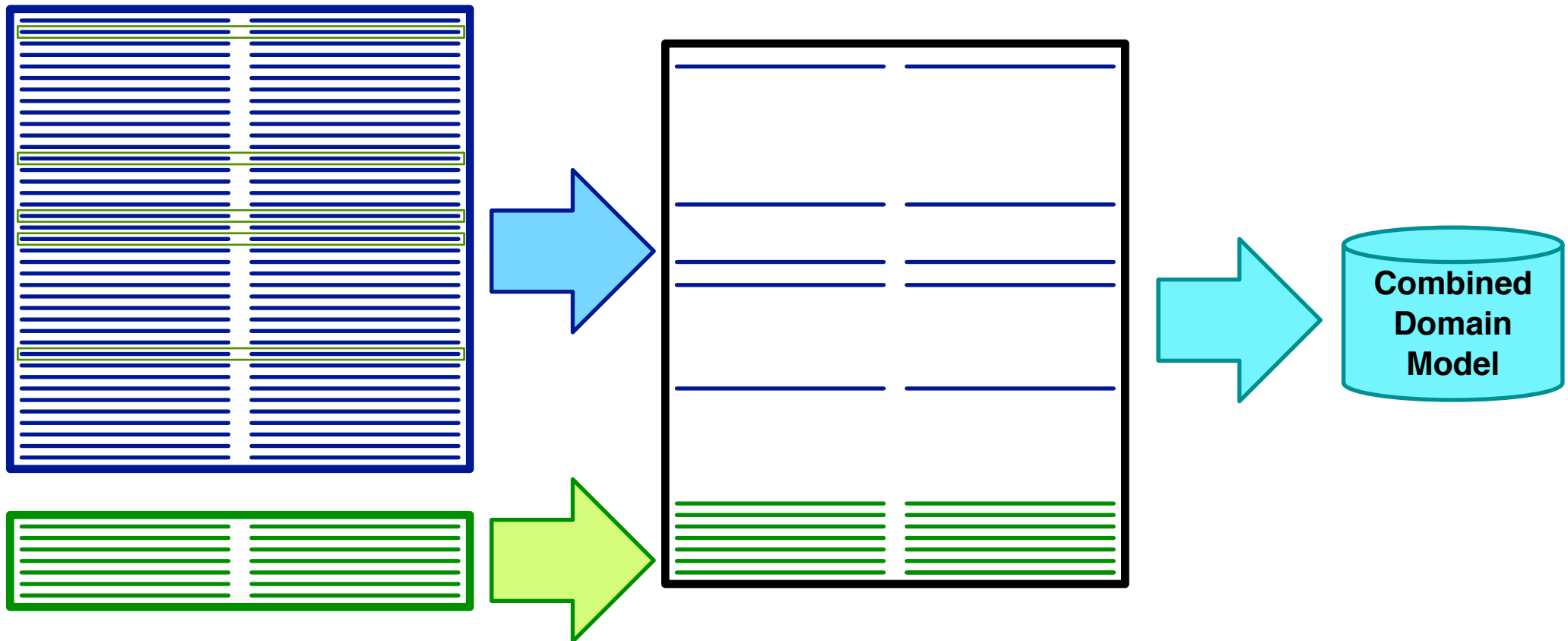


# Fill-Up



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

# Sentence Selection

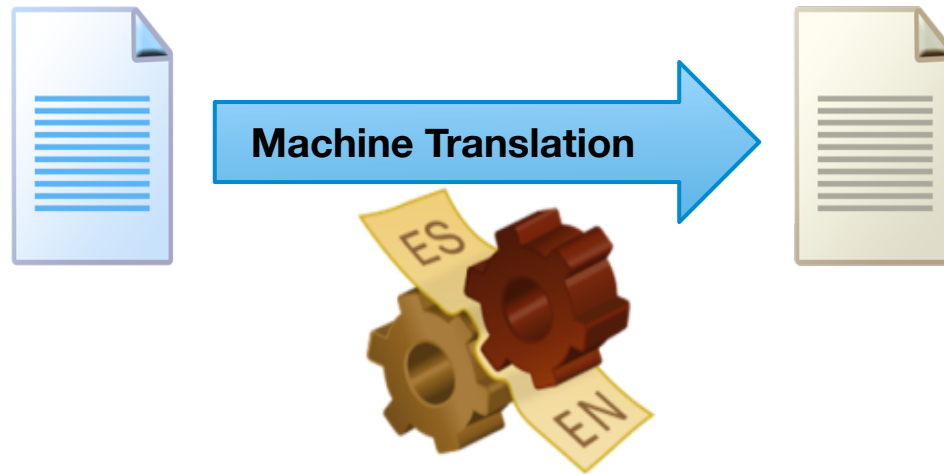


- Select out-of-domain sentence pairs that are similar to in-domain data
- Score similarity with language model, other means

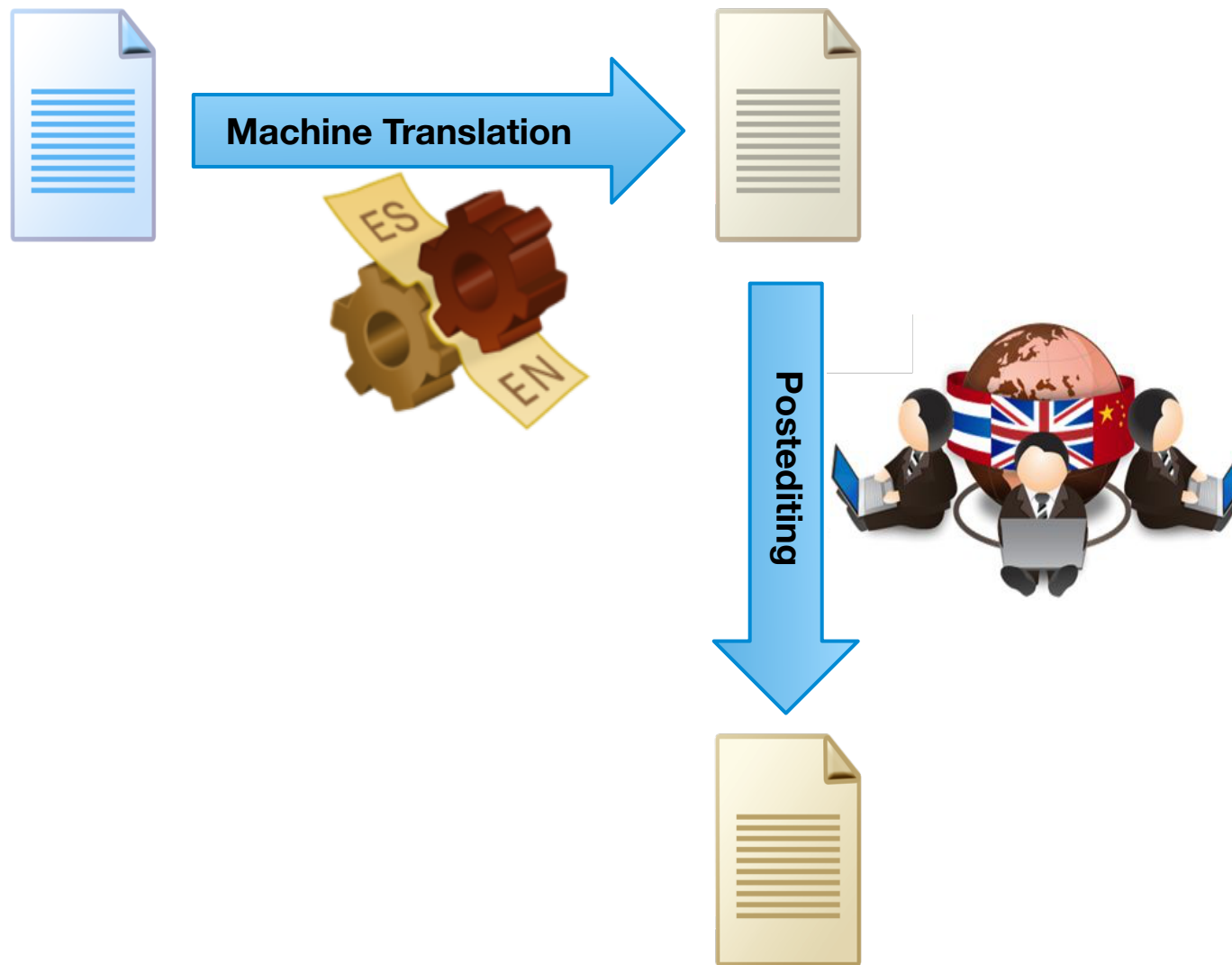
# Project Adaptation

- 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

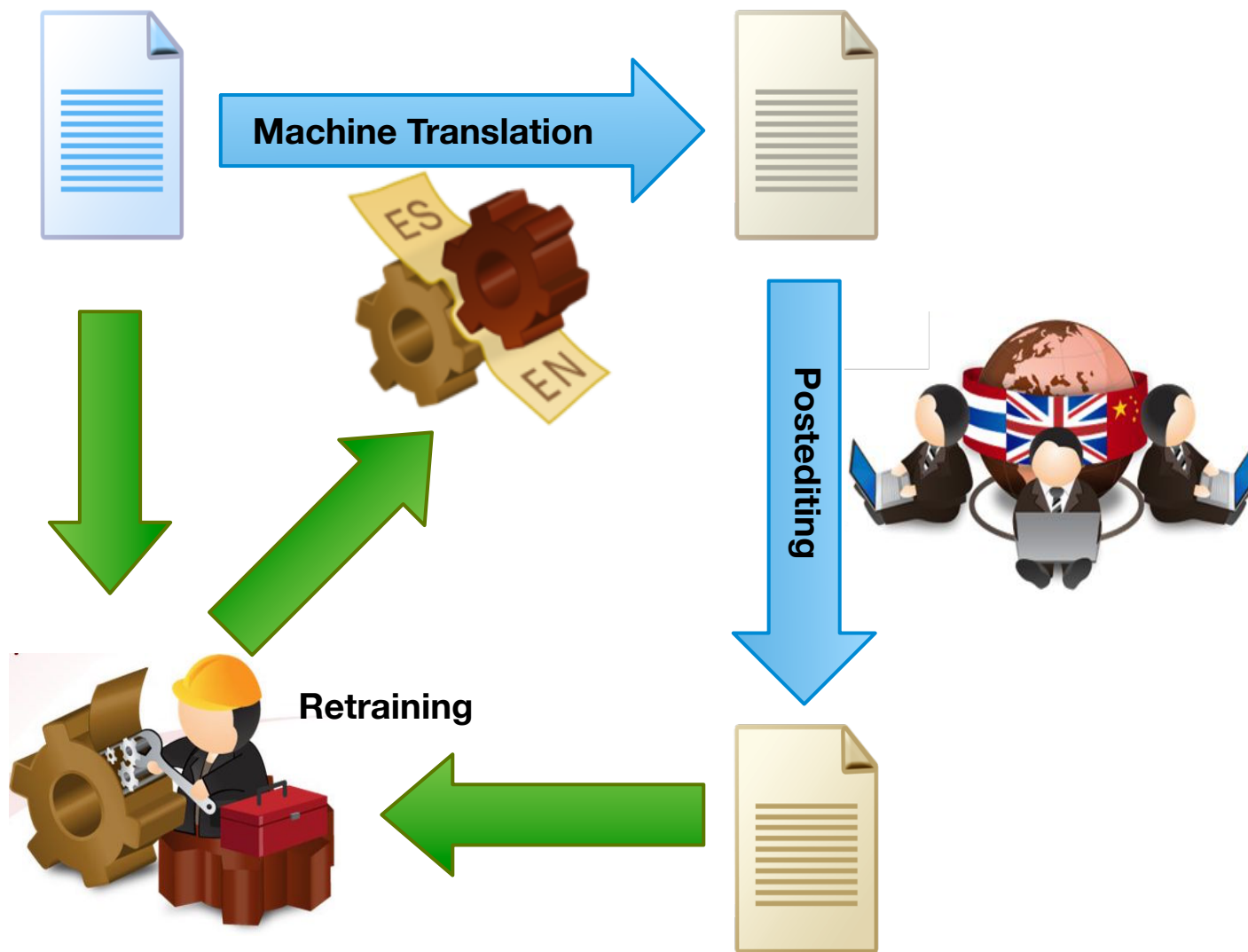
# Incremental Updating



# Incremental Updating



# Incremental Updating





# 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

# 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)

- Post-editing
- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- Interactive translation prediction
- Model adaptation
- **Logging, eye tracking, and user studies**
- CASMACAT Home Edition

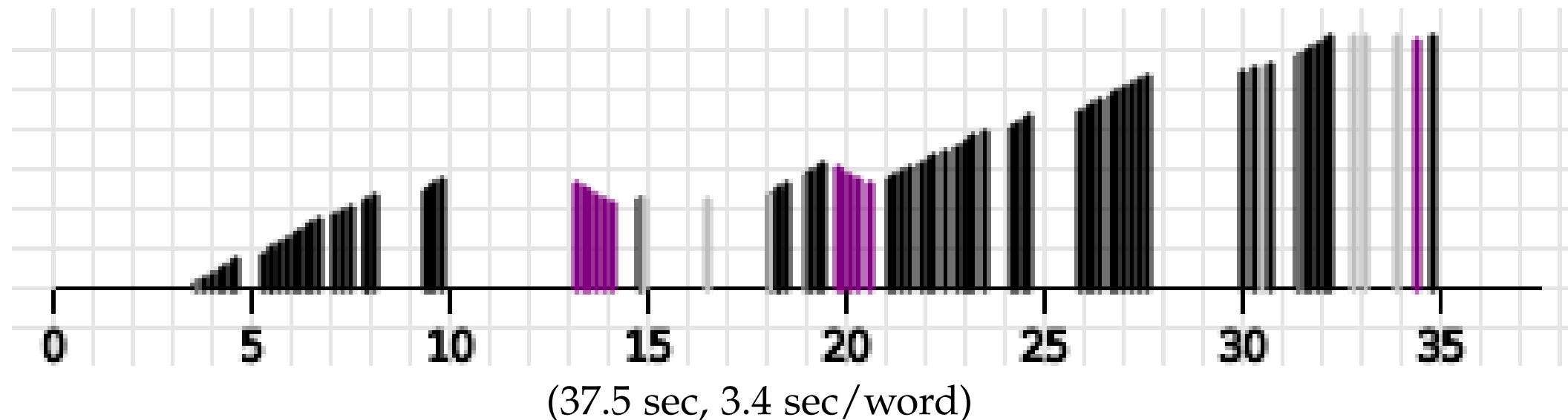
# How do we Know it Works?

- 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?

# Keystroke Log

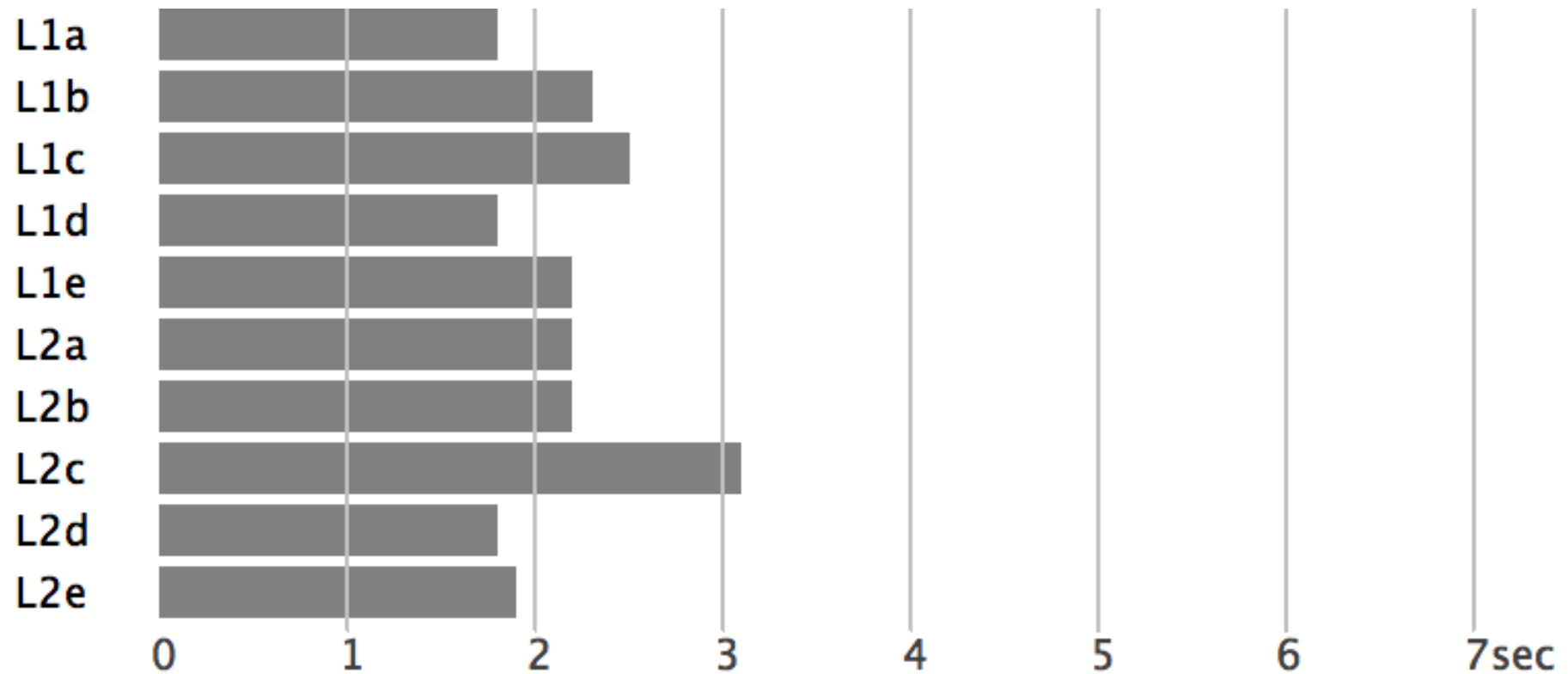
Input: *Au premier semestre, l'avionneur a livré 97 avions.*

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



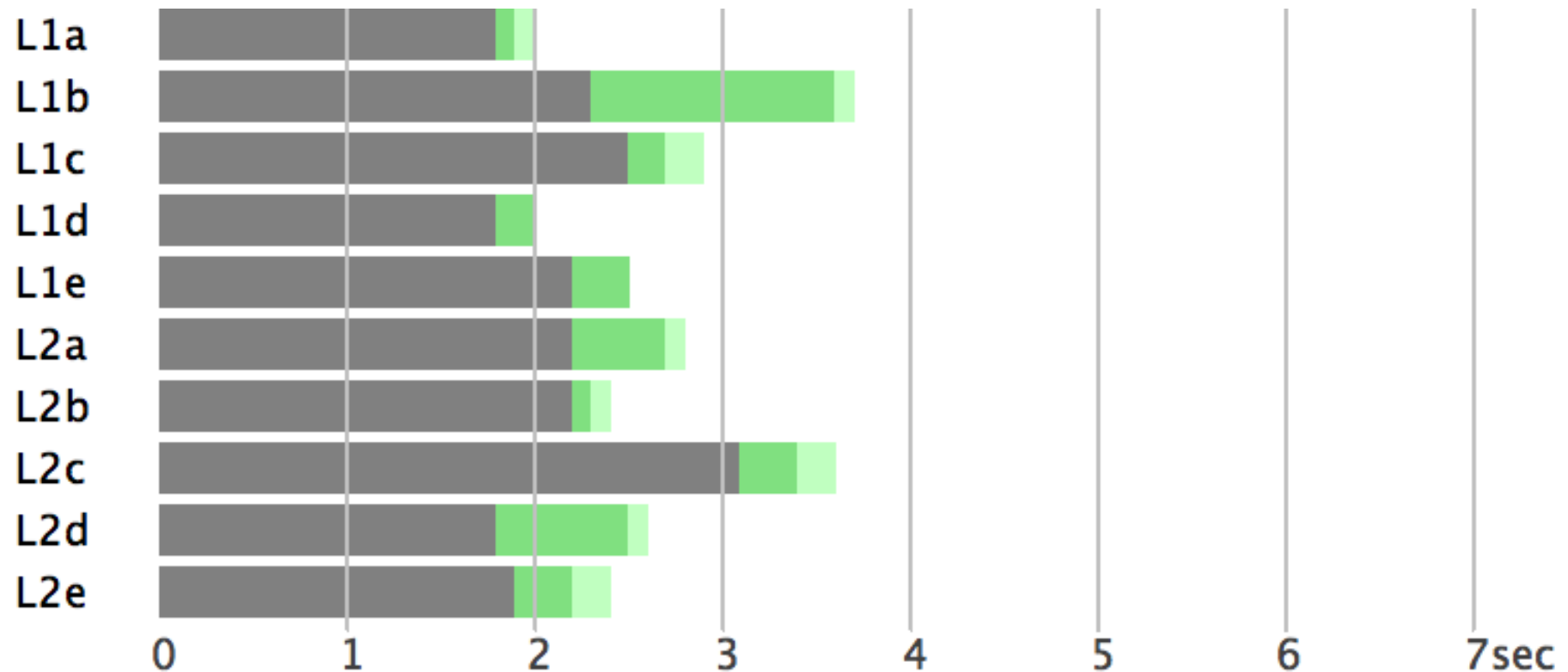
black: keystroke, purple: deletion, grey: cursor move  
height: length of sentence

# Unassisted Novice Translators



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

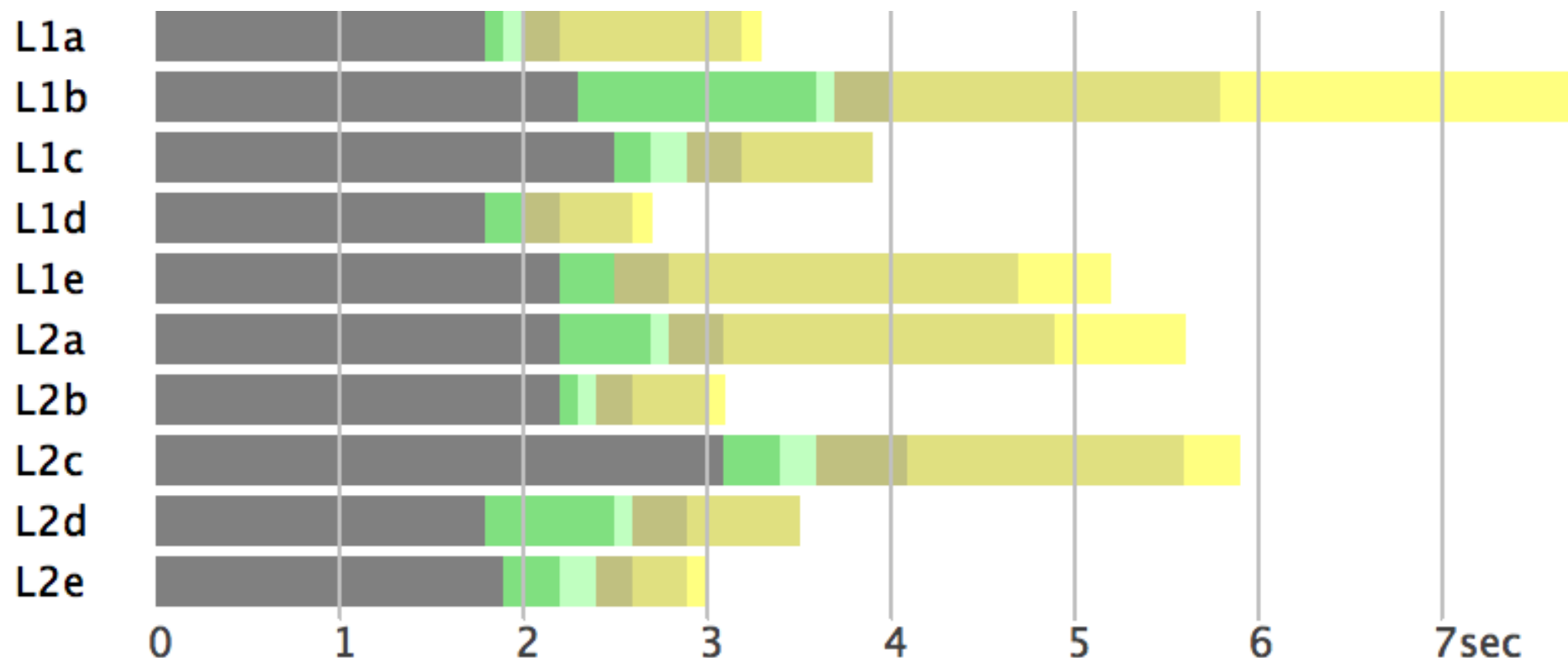
# Unassisted Novice Translators



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

| <b>User: L1b</b>   | 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



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

# Origin of Characters: Native French L1b

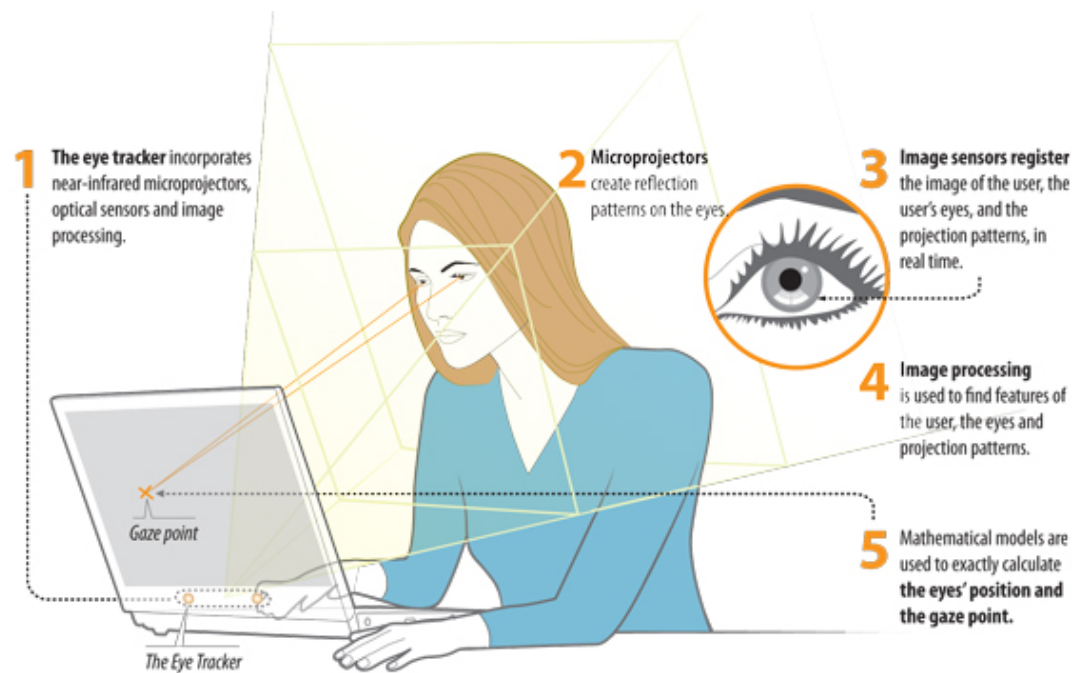
53



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

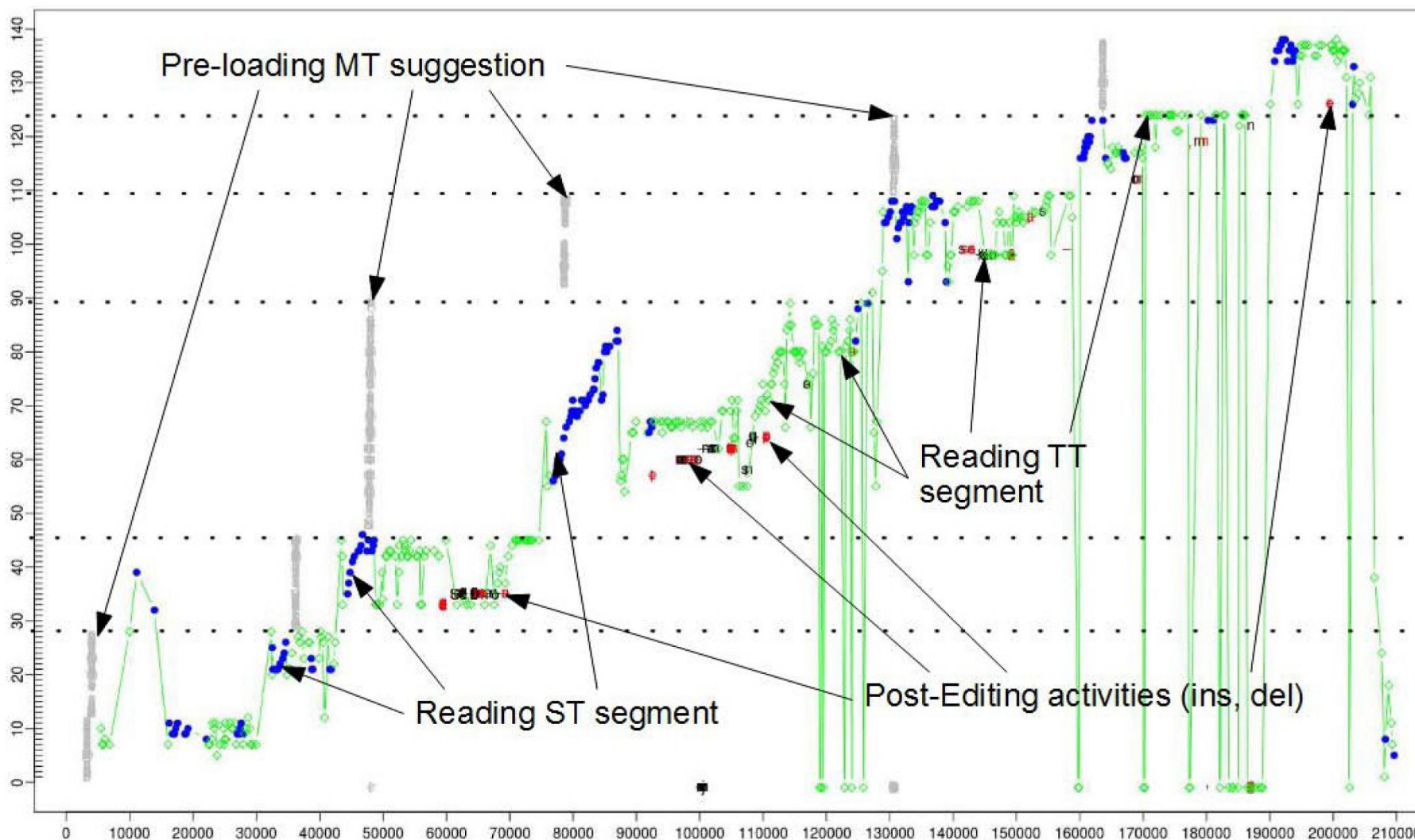
Translation comes to large degree from assistance

# 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

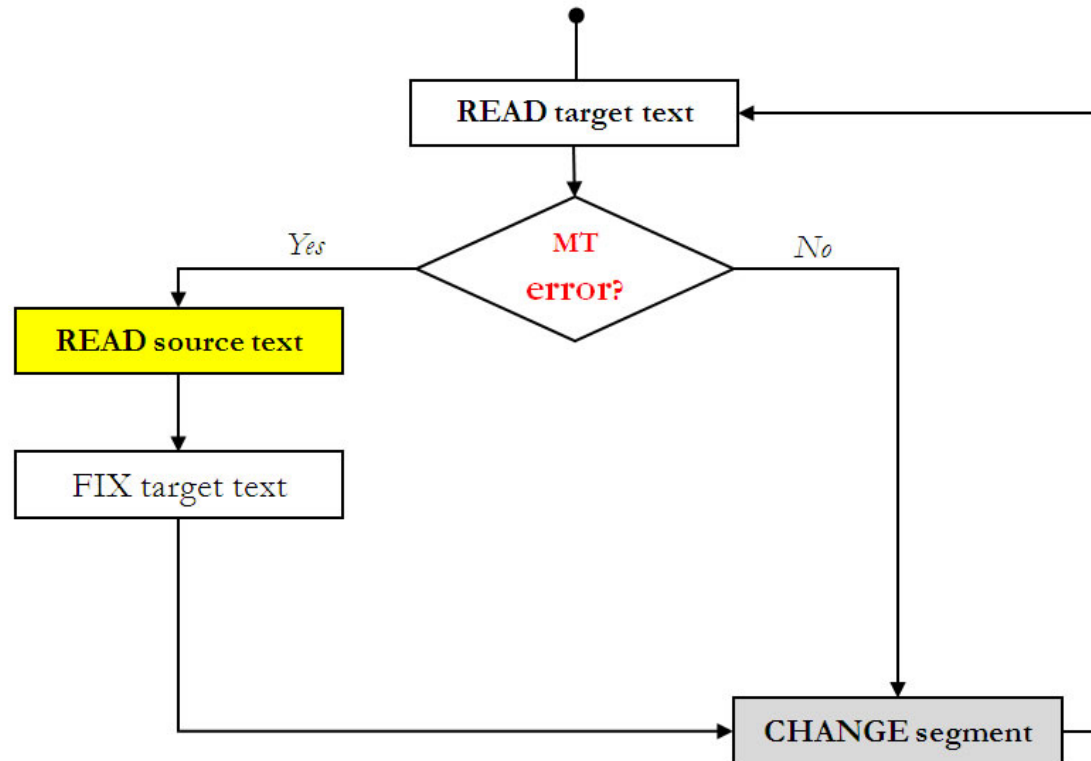
# Eye Tracking Chart



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

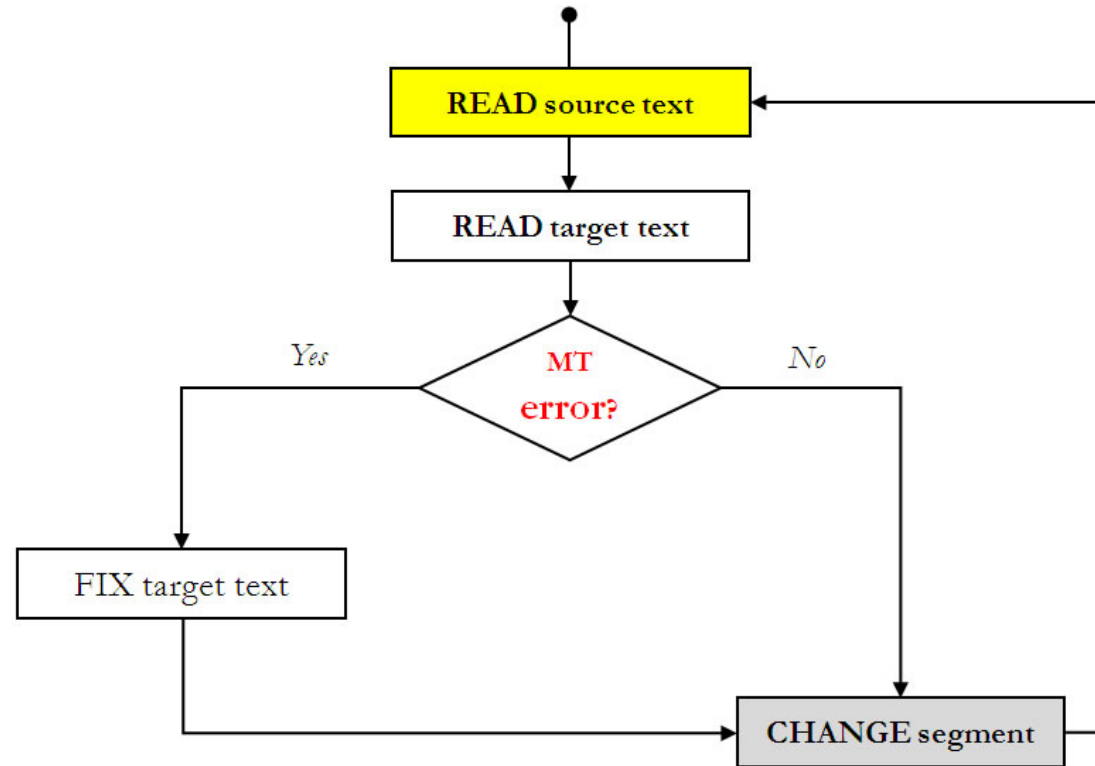


# Cognitive Studies: User Styles



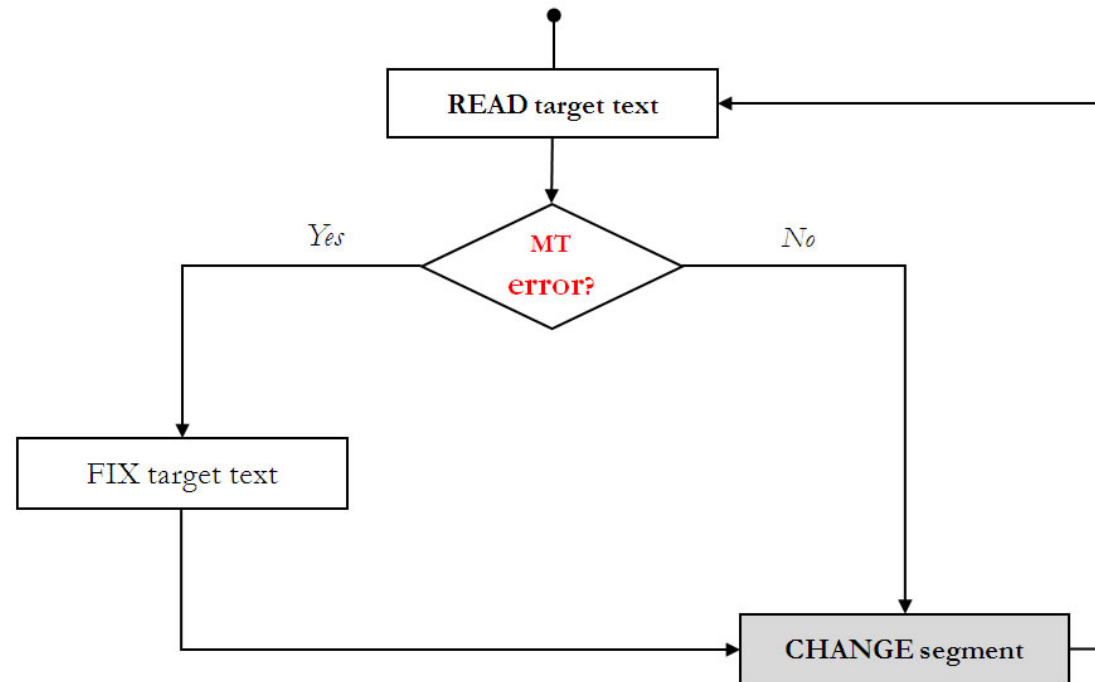
- User style 1: Verifies translation just based on the target text, reads source text to fix it

# Cognitive Studies: User Styles



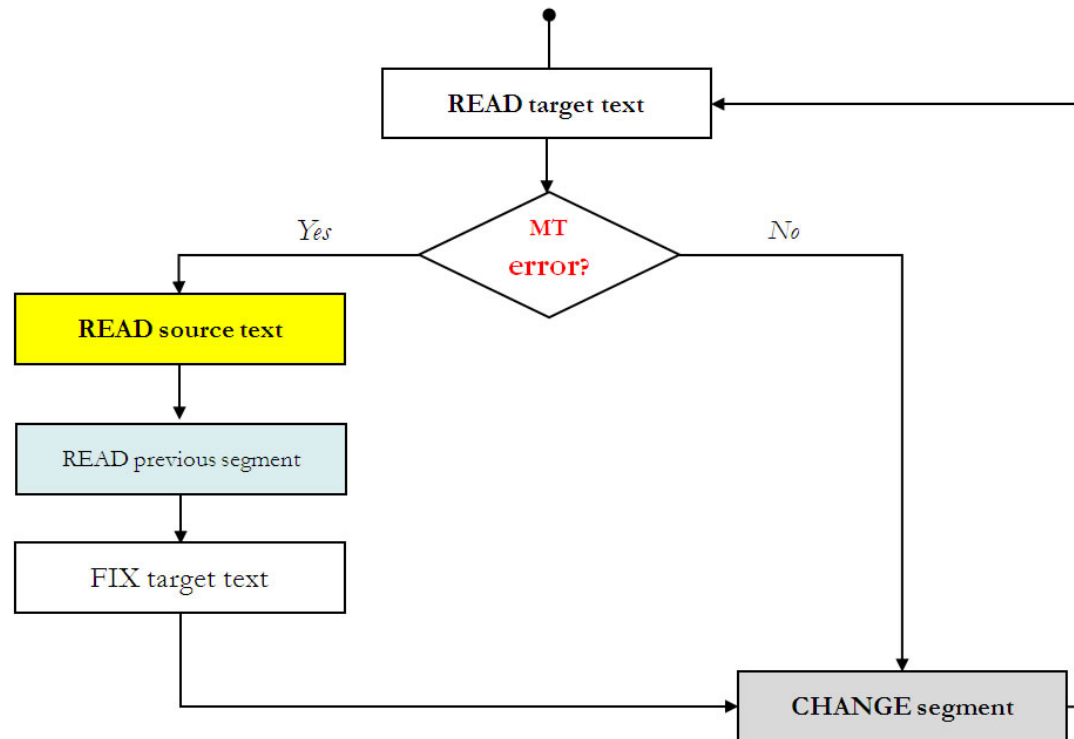
- User style 2: Reads source text first, then target text

# Cognitive Studies: User Styles



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

# Cognitive Studies: User Styles



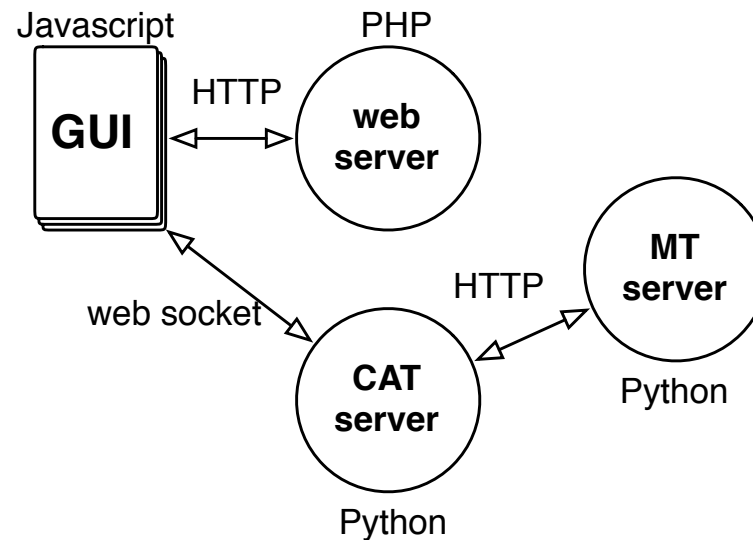
- User style 4: As style 1, but also considers previous segment for corrections

# Backtracking

- Local backtracking
  - immediate repetition
  - local alternation
  - local orientation
- Long-distance backtracking
  - long-distance alternation
  - text final backtracking
  - in-text long distance backtracking

- Post-editing
- Richer information
  - word alignment
  - confidence scores
  - translation option array
  - bilingual concordancer
  - paraphraser
- Interactive translation prediction
- Model adaptation
- Logging, eye tracking, and user studies
- **CASMACAT Home Edition**

# CASMACAT

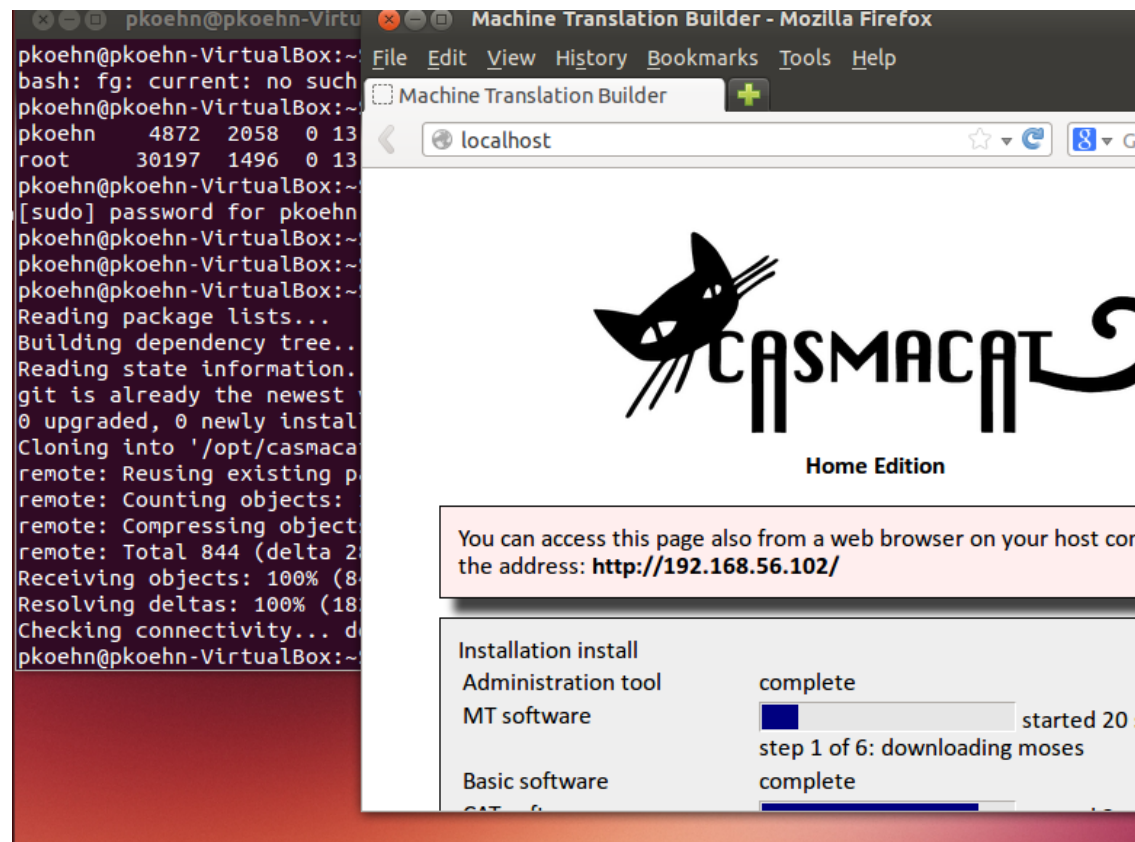


- European research project 2011-2014
- All describe methods implemented in CASMACAT workbench
  - builds on MATECAT open source implementation
  - typical web application: LAMP (Linux, Apache, MySQL, PHP)
  - uses model, view, controller breakdown
- Workbench freely available at <http://www.casmacat.eu/>

- 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](#)

|                                   |                            |   |
|-----------------------------------|----------------------------|---|
| <b>Deployed:</b>                  | fr-en-upload-1             |   |
| <b>Memory:</b>                    | 1.2 GB used, 6.6 GB free   |  |
| <b>Disk:</b>                      | 12.9 GB used, 10.2 GB free |  |
| <b>Uptime:</b>                    | 22:24                      |   |
| <b>Load:</b>                      | 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 |                        |
|--|------------|-----------|------------------------|
| <a href="#">European Central Bank</a>        | 102,980    | OPUS      | <a href="#">upload</a> |
| <a href="#">European Medicines Agency</a>    | 372,824    | OPUS      | <a href="#">upload</a> |
| <a href="#">EU Bookshop</a>                  | 3,618,897  | OPUS      | <a href="#">upload</a> |
| <a href="#">European Constitution</a>        | 6,667      | OPUS      | <a href="#">upload</a> |
| <a href="#">European Parliament</a>          | 1,260,689  | OPUS      | <a href="#">upload</a> |
| <a href="#">KDE4</a>                         | 126,141    | OPUS      | uploaded               |
| <a href="#">KDE4 (el-en_GB)</a>              | 125,537    | OPUS      | <a href="#">upload</a> |
| <a href="#">Open Subtitles</a>               | 220,445    | OPUS      | <a href="#">upload</a> |
| <a href="#">Open Subtitles 2011</a>          | 10,693,456 | OPUS      | <a href="#">upload</a> |
| <a href="#">Open Subtitles 2012</a>          | 12,984,773 | OPUS      | <a href="#">upload</a> |
| <a href="#">Open Subtitles 2013</a>          | 14,626,890 | OPUS      | <a href="#">upload</a> |
| <a href="#">South-East European Times</a>    | 165,532    | OPUS      | <a href="#">upload</a> |
| <a href="#">South-East European Times v2</a> | 224,808    | OPUS      | <a href="#">upload</a> |
| <a href="#">SPC</a>                          | 7,035      | OPUS      | <a href="#">upload</a> |
| <a href="#">Tatoeba</a>                      | 2,469      | OPUS      | <a href="#">upload</a> |
| <a href="#">DGT-Translation Memory</a>       | 3,016,402  | JRC       | <a href="#">upload</a> |

**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**

# Thank You



# questions?