



# When the Past, Present and Future Collide

Agenor Hofmann-Delbor, Localize.pl

# Translation industry loves typewriters

---



Exactly 60 characters in 30 lines = 1800 chars per page.

Other variants (incl. whitespaces): 1800, 1125, 1500.

Traditionally, billing is based on the target text.

What about the 250 words per page standard?



# Similarity of text

---



# What is hidden in the document?

word[whitespace]word[whitespace]word[whitespace]word.

Analysis: 4 words, 17 chars  
(whitespaces not included).

Estimating the no. of characters with whitespaces  
included:

No. of chars (whitespaces not incl.) + no. of words.



# Databases - the heart of all technologies

---



# Fuzzy matching

Cake is not the issue.



Cake	is	not	the	issue.
------	----	-----	-----	--------

100%



Coffee	is	not	the	issue.
--------	----	-----	-----	--------

81%



Cake	is	not	the	problem.
------	----	-----	-----	----------

81%



Cake	is	not	the	issue.
------	----	-----	-----	--------

100%



# Estimating translation time and workload



Total	Type	Segments	Words	Characters	Percent
Files:1 Chars/Word:5.30	Repetitions	2332	8704	47391	40.87%
	100%	0	0	0	0.00%
	95% - 99%	80	651	3378	3.06%
	85% - 94%	303	2519	13269	11.83%
	75% - 84%	264	1962	10257	9.21%
	50% - 74%	76	592	3123	2.78%
	New	907	6868	35520	32.25%
	<b>Total</b>	<b>3962</b>	<b>21296</b>	<b>112938</b>	<b>100%</b>





# Artificial standard of fuzzy bands

## Fuzzy Bands

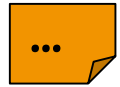


Specify the bands to use for translation analysis. Fuzzy translation match statistics will be collected for every specified band.

Min		Max
50	▼	74
75	▼	84
85	▼	94
95	▼	99



# Can we still find new ways of editing?



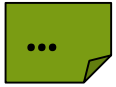
Source files



	Source text	Translation
1	This is a sentence.	To jest zdanie.
2	This is a new sentence.	To jest nowe zdanie.
3	This is a sentence.	To jest zdanie.
4	This is another new sentence.	To jest kolejne nowe zdanie.
5	This sentence contains a term.	
6	Terms are added to the termbase.	



Final Translation (target files)



CAT tools use their own bilingual formats for saving draft translations

# Using multiple resources

---



Alicia has a cat.

1  
2

Alicia has a cat.

Alicia has a cat.

CM

Ala ma kota.

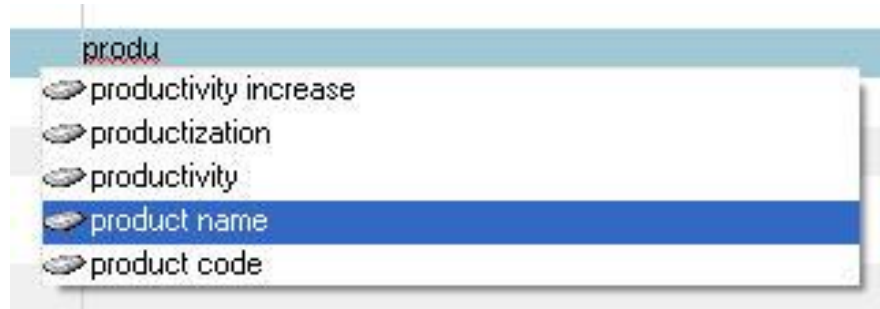
80%

Alicja ma kota.



# Predictive typing

---



Source: D. Brockmann, SDL

# Repairing matches



Sample 1 M-en-uk\_tr-tr 9/27/2016 11:42:27 AM GLOBAL\dbrockmann

Translation Results... Fragment Matches... Concordance Search... Comments TQAs (0) Messages Term Recognition Termbase Search

Connect the blue cable to the probe.docx.sdlxliff [Translation]\*

Source Text	Score	Target Text	Status
1 Connect the blue cable to the probe.		Connectez le câble bleu à la sonde.	P
2 Connect the green cable to the probe.	91%	Connectez le câble vert à la sonde.	P
3 To run the spelling check:			P
4 To run the verification:			P

WordprocessingML v. 2

All segments INS 39.13% 30.43% 30.43% Chars: 35 0/104

Source: D. Brockmann, SDL

# Big Brother, aka post-editing



The screenshot displays the SDL Trados Studio interface. The top navigation bar includes 'Document Overview', 'Document Records', and 'Document Reports'. The main window shows a document overview for 'PL\_Watch your step.docx.sdxiff' with a total elapsed time of 00:15:13. Below this, there are two tables: 'Translation Modifications' and 'Post-Edit Modifications Analysis'. The 'Translation Modifications' table shows various match types like Perfect Match, Context Match, etc. The 'Post-Edit Modifications Analysis' table shows segments, words, characters, and price. A 'Confirmation Statistics' table is also visible. Below these tables is a list of document activities with columns for ID, Date/Time, Status, Match, Words, and Source. The bottom part of the screenshot shows a browser window displaying an 'Average Throughput by Sentence Length, per Target Language' bar chart. The chart shows 'Words per hour' on the y-axis (0 to 800,000) and 'Sentence Length (Number of Words)' on the x-axis (2 to 36). The chart shows a significant peak in throughput for sentence lengths between 17 and 27 words, with a red bar indicating 'Post-edit' activity.

Match Type	Total	Modified	%	Words	Characters	Tags
Perfect Match	0	0	0%	0	0	0
Context Match	32	0	0%	0	0	0
Exact Match	13	0	0%	0	0	0
Automated Translation	40	16	40%	218	1439	0
Fuzzy Match	13	8	62%	60	399	0
New	1	2	200%	19	135	0
Sub-Total				297	1973	0
Total	99	26	26%	297	1973	0

Type	Segments	Words	Characters	Percent	Tags	Price	Total
100%	6	92	627	30.98%	0	0.06	5.52 (EUR)
95%							
85%							
75%							
50%							
New							
Total							

Document Name	Source	Target	Activity Type	Translation Mod
PL_Watch your step.docx.sdxiff	pl-PL	en-GB	Translation	2
PL_Watch your step.docx.sdxiff	pl-PL	en-GB	Translation	23

ID	Date/Time	Status	Match	Words	Source
2	15/06/2015 02:22:15	Draft	78%	3	na placu budowy
6	15/06/2015 02:23:08	Draft	95%	13	W przypadku około tysiąca tych wypadków dochodzi do zmianna kości lub zwichnięcia stawów.
7	15/06/2015 03:09:31	Draft	AT	18	Każdego roku, zwiyczajne wypadki tego typu oznaczają, że ludzie łacy jak Ty muszą zrezygnować z pracy w budownictwie.
8	15/06/2015 03:09:52	Draft	AT	15	Nie możesz sobie na to pozwolić i nie może sobie na to pozwolić branża budowlana.
9	15/06/2015 03:10:07	Draft		19	Wielkośći tych wypadków można było zapobiec, gdyby wszystkie osoby na placu budowy obaly o większe bezpieczeństwo w miejscu pracy.
10	15/06/2015 03:10:15	Draft Translated	100%	13	W interesie osób na wszystkich szczeblach organizacji leży utrzymanie porządku na placach budowy.
11	15/06/2015 03:10:43	Draft		14	Praca mogą być wykonywane szybciej i łatwiej bez narażania pracowników na poważne obrażenia ciała.
12	15/06/2015 03:12:57	Draft	88%	16	Niniejsza broszura prezentuje kilka przykładów rozsądnych działań, które można podjąć, aby zapewnić porządek na placu budowy.
13	15/06/2015 03:13:17	Draft	AT	7	Jednak samo przeczytanie tej broszury nie wystarczy.
	15/06/2015 03:13:23	Draft	AT		O ile wszystkie osoby na placu budowy nie

**Average Throughput by Sentence Length, per Target Language**

Polish > English (United Kingdom)

Words per hour

Sentence Length (Number of Words)

Post-edit

Edit Distance Graph (Info)

# Neural Machine Translation is here to stay



The screenshot displays the SDL Trados Studio interface with the following components:

- Toolbar:** Includes icons for Project Settings, Cut, Copy, Paste, Batch Tasks, Formatting, QuickInsert, Translation Memory, Terminology, Segment Actions, and Navigation.
- Editor:** Shows the source text: "England.docx [Translation en-U...]"
- Translation Results:** A list of translation pairs with a yellow highlight for the DeepL Translator provider.
- Table of Results:**

Source Text	Provider	Target Text	Quality
1 England became a unified state in the 10th century and, since the Age of Discovery, has had a significant cultural and legal impact on the wider world.	AT	England wurde im 10. Jahrhundert zu einem vereinigten Staat und hat seit dem Zeitalter der Entdeckungen einen bedeutenden kulturellen und rechtlichen Einfluss auf die ganze Welt ausgeübt.	P
2 England is a country that is part of the United Kingdom that shares land borders with Scotland to the north and Wales to the west.	AT	England ist ein Land, das Teil des Vereinigten Königreichs ist, das Landgrenzen mit Schottland im Norden und Wales im Westen teilt.	P
3 The country covers five-eighths of the island of Great Britain.		Das Land bedeckt fünf Achtel der Insel Großbritannien.	P
4 England became a unified state in the 10th century and, since the Age of Discovery, has had a significant cultural and legal impact on the wider world.			P
5 Climate			P
6 England has a temperate maritime climate, i.e. it is mild with temperatures not much lower than zero degrees in winter and not much higher than thirty-two in summer.			P

Source: DeepL

Bottom status bar: All segments | INS | 61.54% | 38.46% | Chars: 0 | 0/430 | Language flags (EN, DE, FR, IT, PT, ES, NL, PL, RU, UK, US, JP, KR, AR, HI, VI, TH, ID, ZH, TW, PT, BR, PT, BR, PT, BR, PT, BR)

# Measuring the adaptive learning progress



Creating engine from the scratch. Running the initial analysis:

Type	Segments	Words	Characters	Percent	Recognized Tokens	Fragment words (whole TU)	Fragment words (TU fragment)	AdaptiveMT Impact	Tags
AdaptiveMT Baseline	261	1315	6632	100.00%	207	0	0		92
AdaptiveMT with Learnings	0	0	0	0.00%	0	0	0	0.00%	0
Total	261	1315	6632	100%	207	0	0	0.00%	92

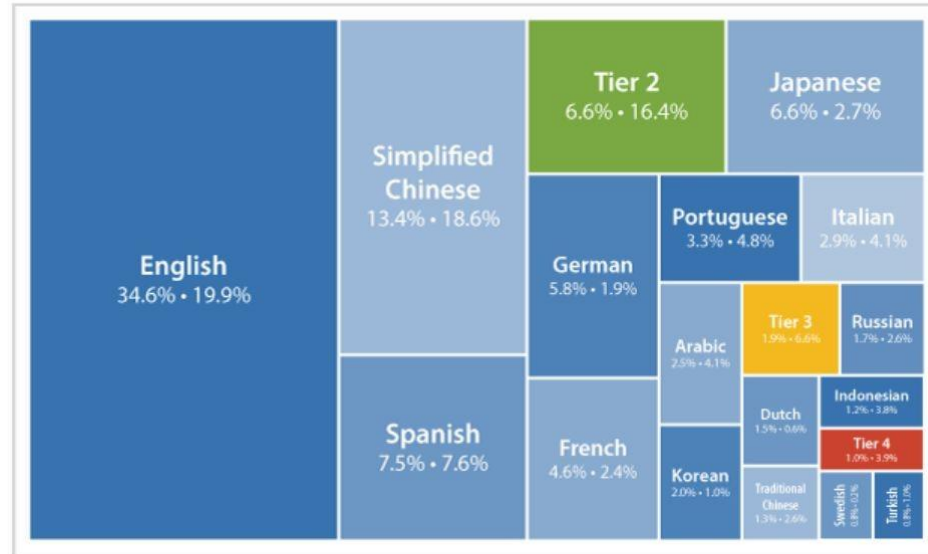
...and after some reviewing:

AdaptiveMT Baseline	255	1292	6511	98.25%	207	0	0		92
AdaptiveMT with Learnings	6	23	121	1.75%	0	0	0	44.00%	0
Total	261	1315	6632	100%	207	0	0	44.00%	92

# Languages of business, 2022



2022 Total Online Market and Audience Shares



**All Tiers**  
with Tier One broken  
down by language

Source: Common Sense Advisory

Key: **34.6% • 19.9%**  
↑ Share of online GDP    ↑ Share of Online Audience





# What can we expect?

---

- **Even more AI. Due to almost unlimited processing power, we can now apply AI to more domains.**
- **More AI in predictive text, editing and fixing matches.**
- **Even more automation in project management.**
- **Smart reviewing of NMT texts (forget about No Matches) - no more postediting as we know it from SMT.**
- **Focus on hourly rates and increased productivity.**



**THANK YOU!**

[agenorh@localize.pl](mailto:agenorh@localize.pl)  
Agenor Hofmann-Delbor, Localize.pl