

Bad to the Bone: AI-Enabled SmartLQA





Alex Yanishevsky

-
-
-
-
-
-
-
-

**Director,
AI Deployments
Welocalize**



SmartLQA Agenda



WHAT IS IT?



WHEN IS IT USED?



HOW IS IT USED?



WHAT'S NEXT?



What is it?

**Methodology to
inform strategic
global content
business
decisions
through AI**



SOURCE SUITABILITY



PREDICT AT-RISK CONTENT



“SPENDING SMART” VIA TARGETED LQA



MTQE CORRELATION



PE DISTANCE CORRELATION



What is it?

AI-Driven Quality Management

Inform data-driven content decisions through AI

1.



SOURCE SUITABILITY

AI can **identify** errors in poor source content and **predict** 'at-risk' content:

- Content written by non-native authors
- Content created by technical specialists for a non-technical audience
- Dated content not adhering to brand tone and voice

Does the source content need to be re-written before translation?



What is it?

AI-Driven Quality Management

Inform data-driven content
decisions through AI

2.



TARGET SUITABILITY

- Does the translation deviate from previous style?
- Does the translation introduce unnecessary complexity?

Does the target need go through LQA for data-driven checks and corrections?



What is it?

AI-Driven Quality Management

Inform data-driven content
decisions through AI

3.



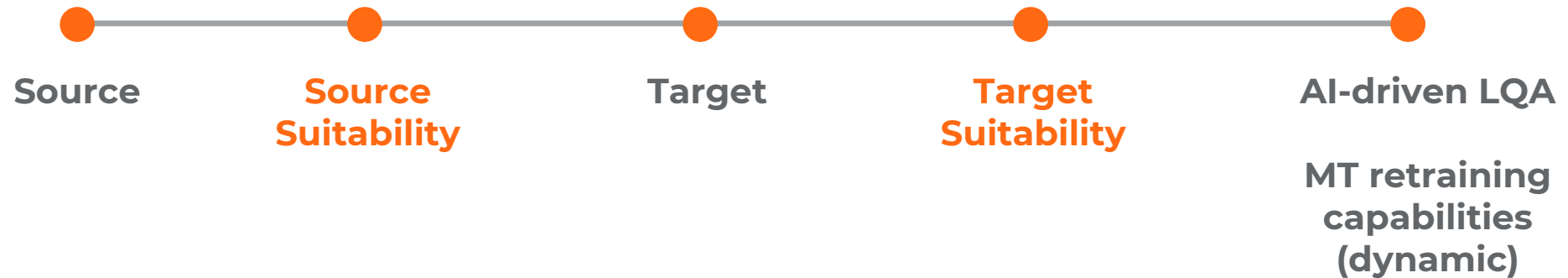
AI-DRIVEN LQA + MT RETRAINING

- Targeted “SmartLQA” focuses on problematic files and segments within them
- Data can be used to **retrain engines (dynamically)**



When Is It Used?

Where this fits into the Content Lifecycle



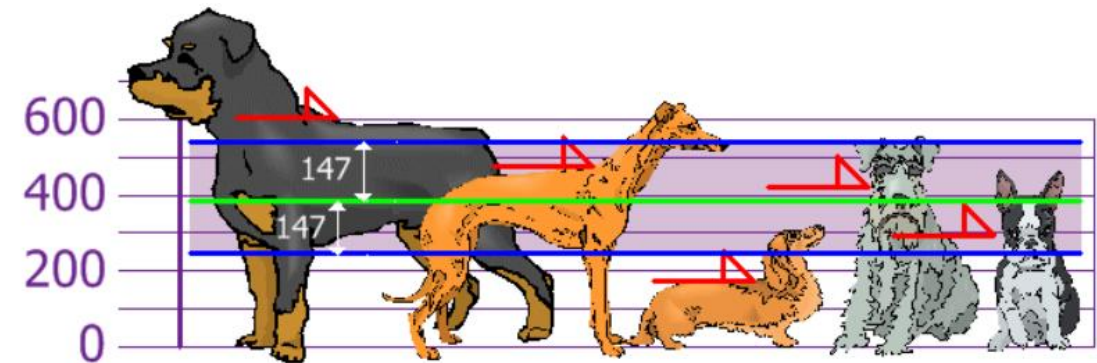
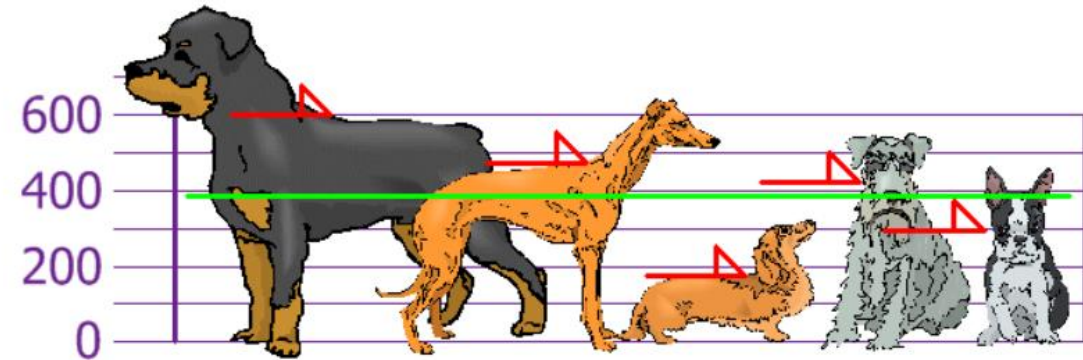
How is it Used? Configuring Thresholds

1.



THRESHOLDS

- Based on average plus standard deviation(s)
- Relative measure
- Captures outliers for that specific domain/product



How is it Used? Configuring Thresholds

2.



THRESHOLDS

- Based on average plus standard deviation(s)
- Relative measure
- Captures outliers for that specific domain/product

Content Type	Avg. ADJ Count	Avg. NOUN Count	Avg. PROPON Count	Avg. Word Count	Avg. Long Word Count	Avg. Complex Word Count	Avg. FleschReadingEase
Legal	3.89	18.55	0.49	57.13	17.60	11.84	66.38
Legal	4.66	18.77	0.46	54.43	17.91	12.44	51.97
Legal	3.60	14.61	0.27	48.10	14.88	9.71	68.19
Legal	3.25	18.42	0.11	46.48	15.15	8.89	63.59
Legal	2.76	14.23	0.25	45.24	12.51	7.35	82.17
Legal	5.05	20.30	0.40	67.33	19.90	13.33	60.53
Repair instructions	0.36	2.71	0.68	9.05	1.80	0.81	49.87
Repair instructions	0.36	2.71	0.68	9.05	1.80	0.81	49.87
Life Sciences	0.00	4.00	0.00	6.00	3.00	1.00	31.55
Life Sciences	1.00	4.00	0.00	16.00	7.00	6.00	31.97
Life Sciences	1.00	4.00	4.00	22.00	5.00	4.00	87.86
Life Sciences	1.08	2.67	0.42	12.08	4.50	2.75	64.97
Transactional 1	1.05	5.27	0.17	15.39	4.36	2.89	48.77
Transactional 2	1.14	6.12	0.06	19.45	5.22	3.25	37.26
Transactional 3	1.94	6.54	0.18	19.90	5.76	3.60	41.68
Transactional 4	1.24	6.52	0.02	20.85	5.65	3.72	35.98
Transactional 5	1.36	5.98	0.60	20.23	5.43	3.38	35.69
Transactional 6	1.23	5.65	0.10	16.12	5.00	3.05	30.40
Transactional 7	1.61	5.80	0.43	18.52	5.56	4.09	31.82
Marketing	0.75	3.36	0.25	13.89	1.93	1.18	87.45
Marketing	0.67	3.00	0.27	12.17	1.77	1.17	86.95
Marketing	0.77	3.50	0.77	17.09	3.73	2.23	80.60
Marketing	0.80	3.00	0.65	16.20	3.15	1.45	78.34
Marketing	0.68	3.96	1.42	16.99	3.79	1.99	85.50
Marketing	0.88	3.42	0.54	13.71	3.38	1.54	97.83
Marketing	0.92	4.58	0.21	16.96	3.04	0.88	89.07
Average score	1.62	7.37	0.52	24.24	6.88	4.36	60.63



How is it Used?

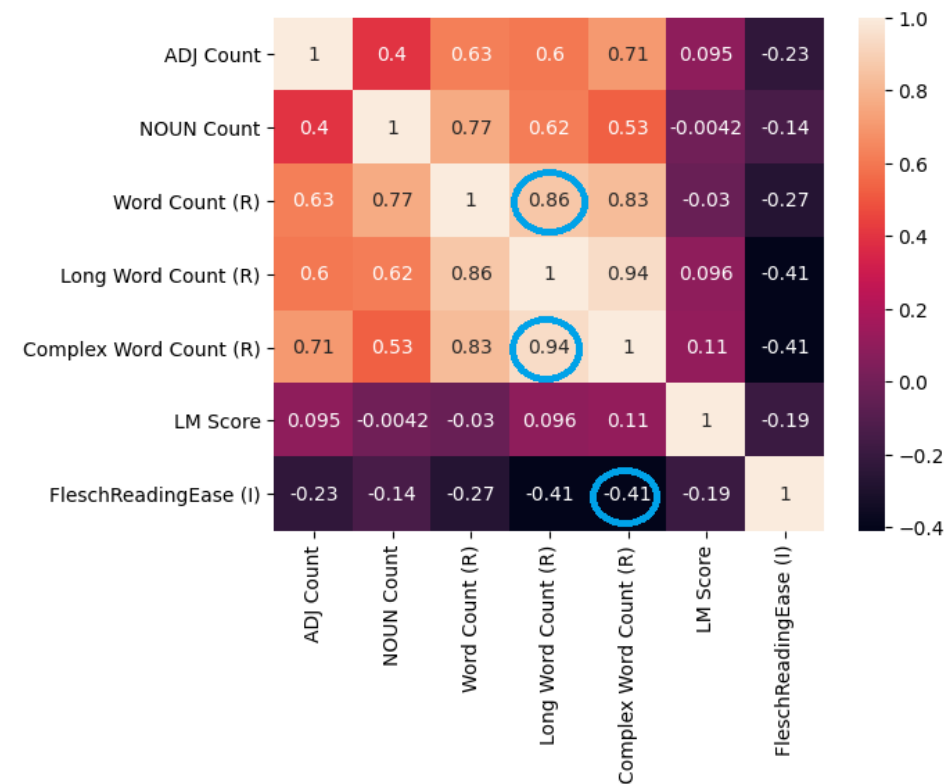
Identifying Salient Features

1.



FEATURES

- Parts of speech such as adjectives, nouns, proper nouns, numbers
- Adjective/noun density
- Long words, complex words, short and long sentences
- Stylistic similarity/dissimilarity
- Readability and complexity metrics
- Correlations to PE Distance and MT



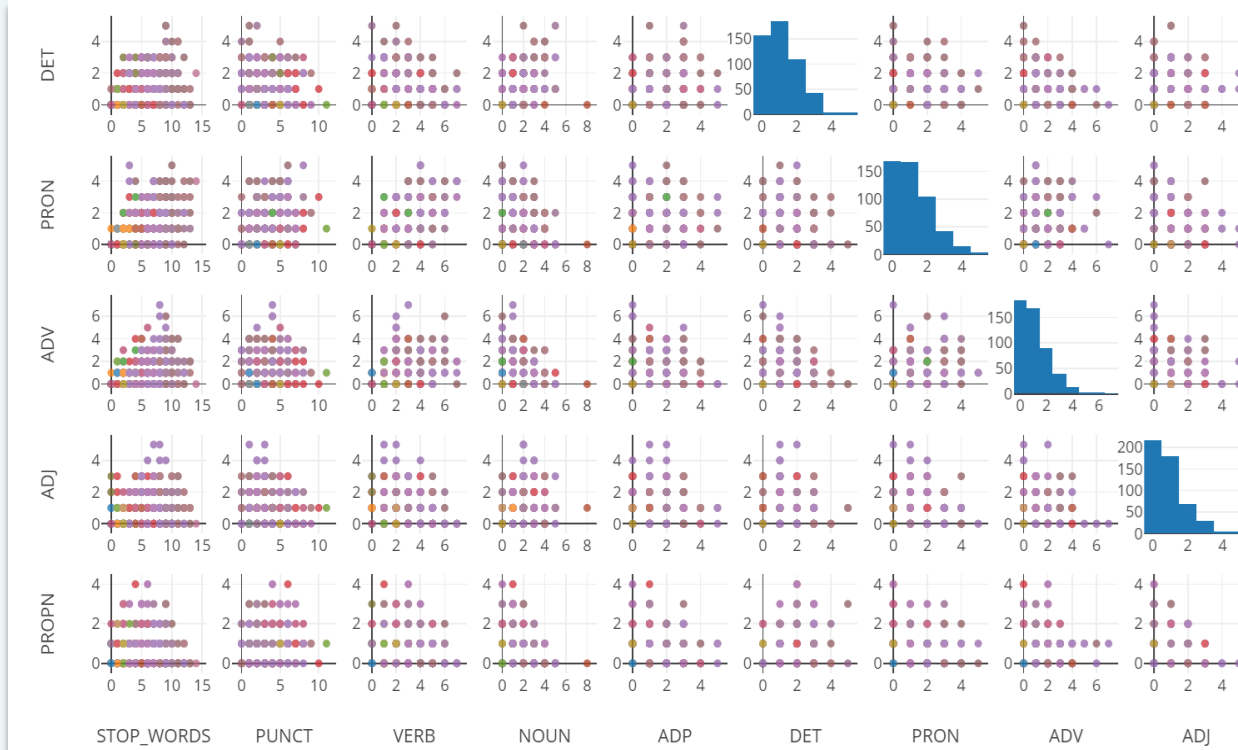
How is it Used? Identifying Salient Features

2.

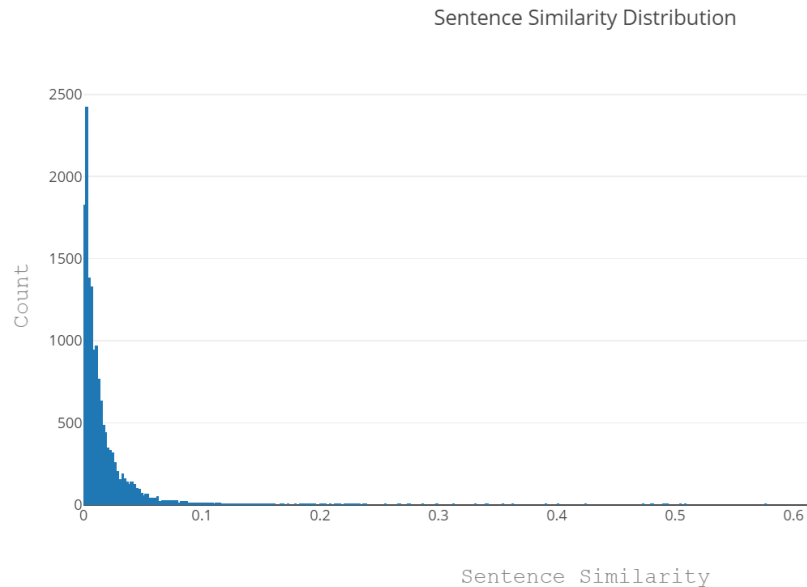


FEATURES

- Parts of speech such as adjectives, nouns, proper nouns, numbers
- Adjective/noun density
- Long words, complex words, short and long sentences
- Stylistic similarity/dissimilarity
- Readability and complexity metrics
- Correlations to PE Distance and MT Quality Estimation metrics



How is it Used? **Source Suitability**



POSSIBLE REMEDIES

- Don't run the project till source is improved
- Route to transcreation, human translation, different MT engines
- Alert of higher LQA risk to all production people (PM, linguists, LQA)



How is it Used? Source Query Analysis

PROCESS

- Analyzed over historical 600 segments for potential DNT
- Analyzed almost historical 400 segments for source ambiguity and meaning (almost 200 for each category)
- Identified thresholds for each category
- Ran thresholds for all categories and identified over 400 potential queries
- Savings of 6K

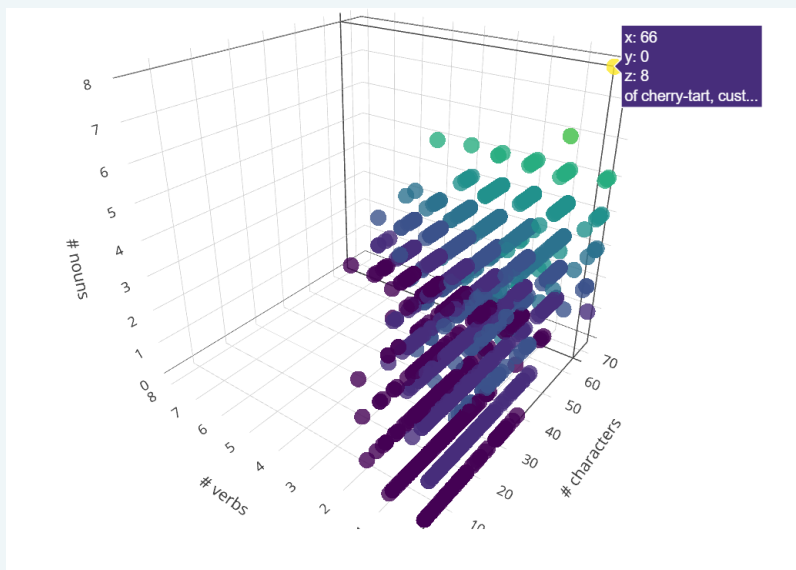
Text		NOUN Count	PROPN	PROPN Count	ADJ/NOUN Density	Long Word Count (R)	Word Count (R)	FleschReadin
Boomi Molecule	0	0	Boomi Molecule	2		1	2	-6.695
Delete incomplete target configuration failed, suspect permission or driver issue.	0	5		0	'NOUN 2', 'NOUN 2'	4	10	-6.355
Drive error recovery FW improvements and enhancements	0	5		0	'NOUN 4'	3	7	-5.727142857
In addition, on November 11th, Sheltered Harbor announced that	0	8	November Sheltered Harbor	11	'NOUN 4'	15	31	-2.017096774
Identity query failed user=1000 to name status=STATUS_ACCESS_DENIED.	0	5	PowerProtect Cyber Recovery Sheltered	0	'NOUN 2', 'NOUN 2'	2	9	0.3
IR Camera (User-Facing fixed focus) with low light +TNR +capability +IPU6 +Proximi	0	27	ExpressSign	1	'NOUN 2', 'NOUN 2', 'NOUN 2'	8	41	4.273658537
Standardized earned MDF expiration timelines aligned to fiscal quarter end dates	0	7		0	'NOUN 3', 'NOUN 3'	7	16	5.5325
Disable Lock Terminal	0	2		0	'NOUN 2'	2	3	6.39

Quick calculation: 405 queries save 15 mins per query = 6075 minutes = 101 hours at \$60/hr (if not more) = **\$6075** saved



How is it Used?

Target Suitability - “Spending Smart”



POSSIBLE REMEDIES

- Go back to linguist for more editing
- Alert of higher LQA risk
- Use information to retrain MT engine (dynamic?)
- Map to client LQA methodology
- Spend LQA \$\$ where it counts
- Confirm MTQE
- Confirm PE Distance and/or TER
- Confirm productivity metrics



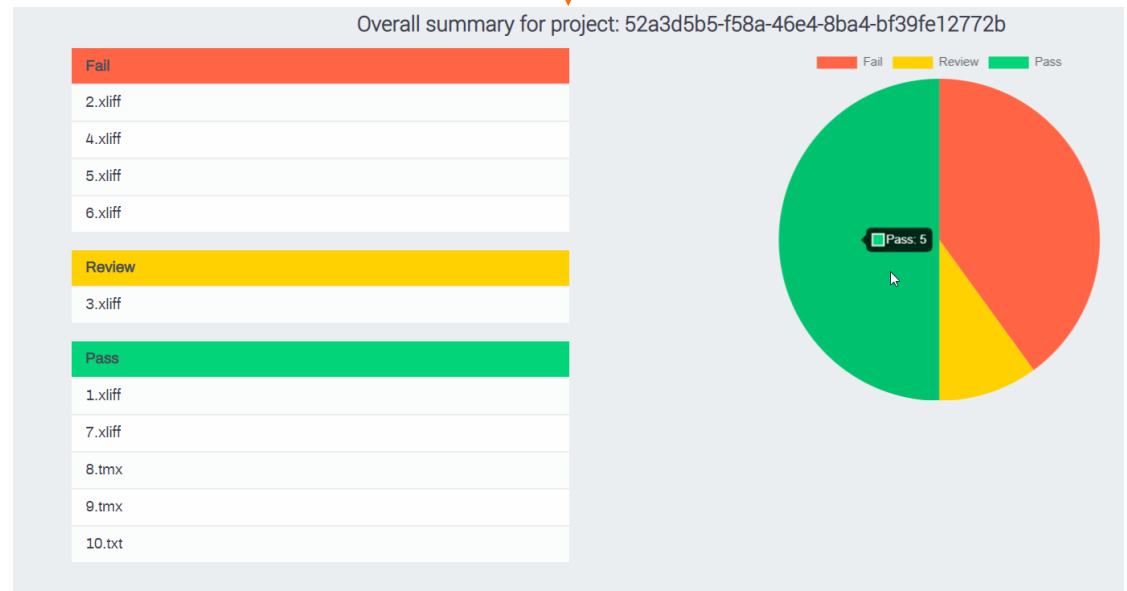
How is it Used?

Summary View

- How many features failed?
- Pass/Fail/Review per segment
- Aggregated to pass/fail per file

1.

Text	ADJ Count Pass	Noun Count Pass	PROPN Count Pass	Long Word Count Pass	Complex Word Count Pass	Nominalization Count Pass	Word Count Pass	LM Pass	FleschReadingEase (l) Pass	Segment Pass/Fail/Review	Segment comment
In addition to the game's deep	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	Fail	
With twelve maps, five modes, and	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	Review	
As easy it is to drop into MP and pick it up, Nathan	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	Pass	

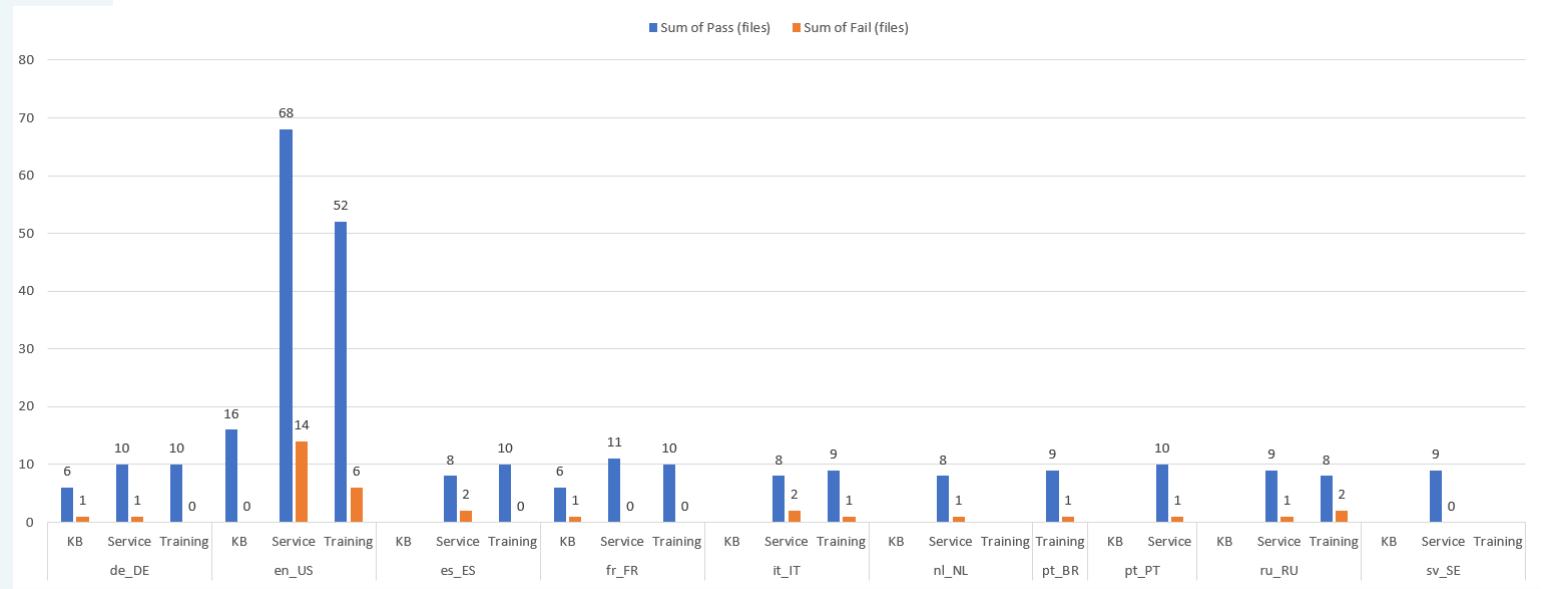


How is it Used?

Summary View

- Passes/fails per domain
- Passes/fails per locale pair

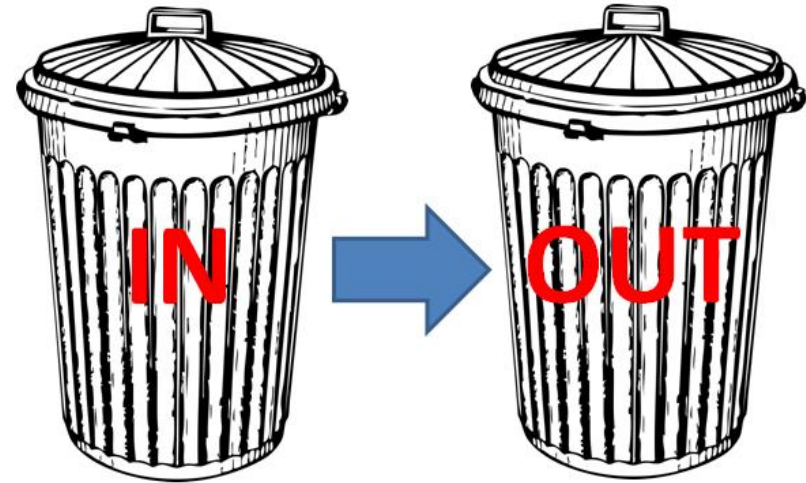
2.



How is it Used?

Garbage In, Garbage Out

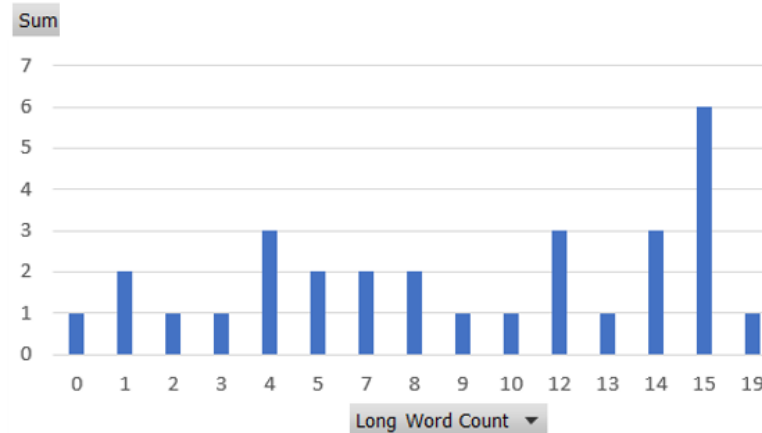
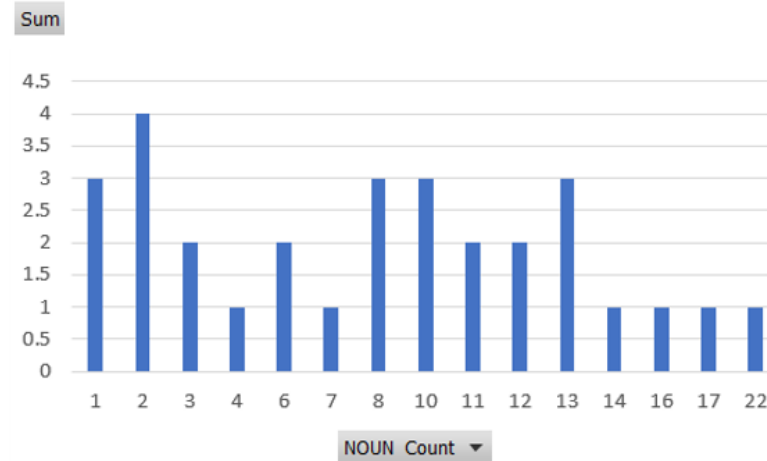
- TRACING SOURCE TO TARGET CORRELATIONS
- POOR SOURCE LEADS TO POOR TARGET



File name	EN									DE								
	ADJ Count	NOUN Count	PROP N Count	Word Count	Long Word Count	Complex Word Count	Nominalization Count	LM Score	FleschReadingEase	ADJ Count	NOUN Count	PROP N Count	Word Count	Long Word Count	Complex Word Count	Nominalization Count	LM Score	LIX
TASK10196529	0.666667	3.333333	1.311111	9.266667	3.555556	0.577778	0.355555556	648.734	53.9952	0.888889	2.333333	1.977778	9.088889	3.844444	0.911111	0.266666667	569.1765	55.52887
TASK10196533	0.954545	4.318182	0.681818	12.45455	5.363636	3.181818	0.454545455	257.5985	36.08856	1.727273	3.363636	1.363636	12.04545	6.136364	2.272727	0.272727273	372.648	59.00989
TASK10196537	0.766667	3.266667	1.366667	9.866667	3.633333	1.633333	0.3	411.9258	55.52377	1.121212	2.272727	2.30303	9.242424	3.878788	0.969697	0.212121212	2095.755	50.82482
TASK10276202	1.338983	3.966102	0.711864	14	4.525424	2.694915	0.355932203	445.9728	52.33312	1.542373	3.932203	0.881356	15.45763	5.745763	1.474576	0.4406677966	607.4004	55.07588
TASK10294494	1.142857	3.97619	0.619048	12.42857	4.380952	2.238095	0.428571429	1075.118	50.6495	1.452381	3.214286	1.309524	12.28571	4.785714	1.047619	0.357142857	1761.157	58.67438
TASK10294496	2.433333	8.266667	1	23.83333	9.266667	6.166667	0.833333333	227.824	29.01318	2.266667	7.333333	1.366667	22.93333	10.56667	2.866667	0.366666667	456.5975	66.27447
TASK10354283	0.608696	2.717391	0.73913	6.902174	2.706522	1.141304	0.293478261	2668.863	42.92559	0.684783	2.26087	1.336957	7.141304	3.108696	0.652174	0.217391304	1856.129	58.27621

How is it Used? How Bad is the File?

More than half of the file
has 6 or more nouns
Half of the file has 8 long
words or more



How is it Used?

A Telling Example



Today's machines enable industrial workers to carry out complex Computer Aided Design, Manufacturing and Engineering (CAD, CAM, CAE) operations, model Computational Fluid Dynamics (CFD), accomplish thermal, stress and fatigue analysis, or visualise and test designs and models using immersive Virtual Reality (VR).

And now the statistics

- 42 words
- 22 nouns
- 19 long words
- 9 complex words

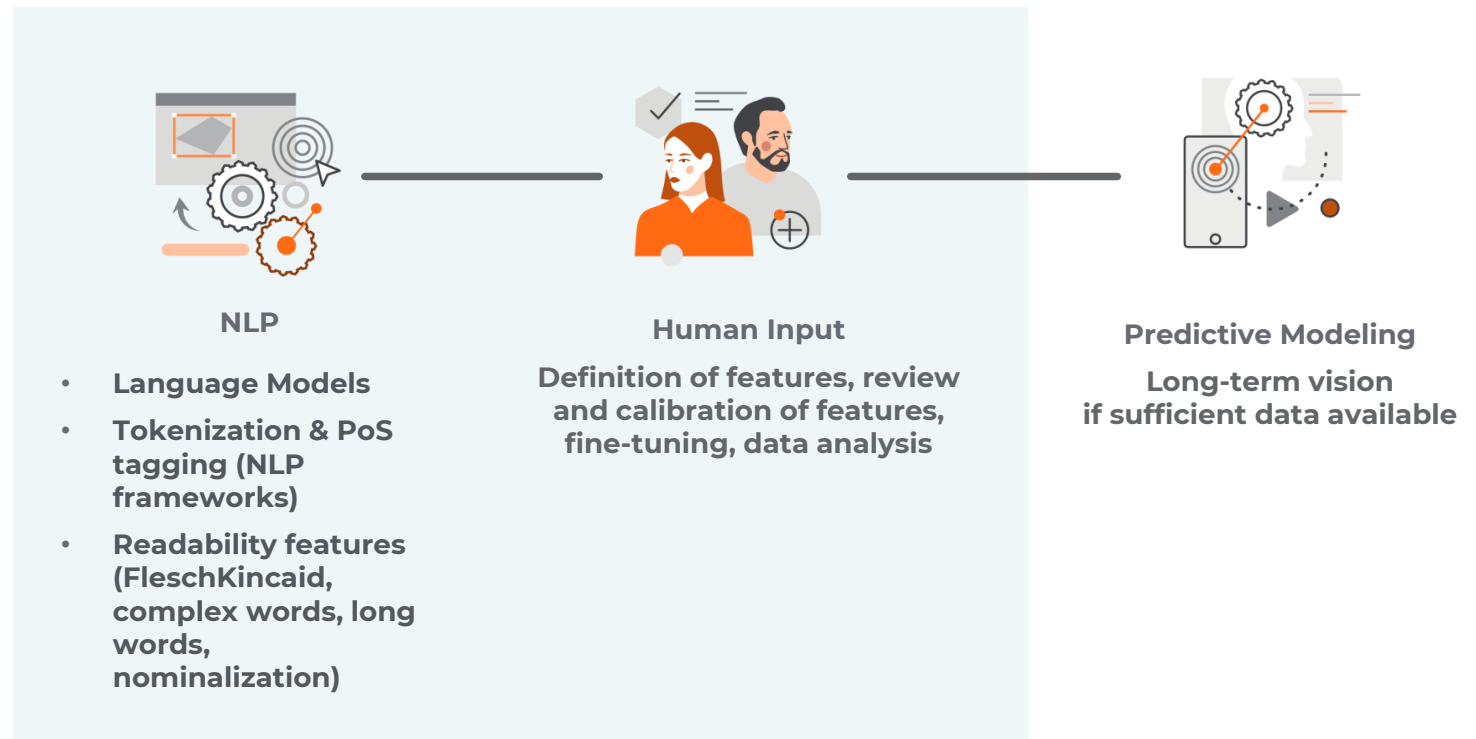
List of nouns

Today | machines | workers | Computer | Design | Manufacturing | Engineering | CAD | CAM | CAE | operations | model | Computational | Fluid | Dynamics | CFD | thermal | stress | fatigue | analysis | designs | models



How is it Used? Under the Hood

NLP frameworks
Human validation
Predictive modeling



How is it Used?

Process Optimization

Reducing time to market and costs while improving linguist acquisition and retention



15-20%

LQA Time Saved



20%

LQA Pass Rate Improvement



10%

LQA Spend Reduction



What's Next?

- Continued human validation
- Build predictive models using machine learning (ML) algorithms
- Human validation comment

“I think this is a very interesting tool that has a **lot of potential**. The output statistics provide some **interesting insights about the nature and style of the source**, and more importantly, also **the target text**. With the help of these figures, a source text can be analyzed for its complexity, while **a translation can be characterized and possibly rated** with regard to certain stylistic guidelines.”





Questions?





Thank you

