



The World Speaks One Language - Yours™

Company & Technology Overview

Product and Technology Platform Overview

Kirti Vashee – kirti.vashee@asiaonline.net



Asia Online Company Objectives



Revolutionize the Internet experience for non-English speakers in Asia



Revolutionize the enterprise translation process with a comprehensive, continuous learning SMT platform



Provide 1 billion+ local-language pages online using mostly translated open license content, combined with compelling portal and social networking style services in Thailand, Indonesia, India, Malaysia, Philippines, Vietnam and China, Japan & Korea



SaaS environment that allows data cleaning and preparation, develop SMT engines on demand and enable ongoing comprehensive post editing and correction to continuously improve engines



The Consumer Market



The Enterprise Market

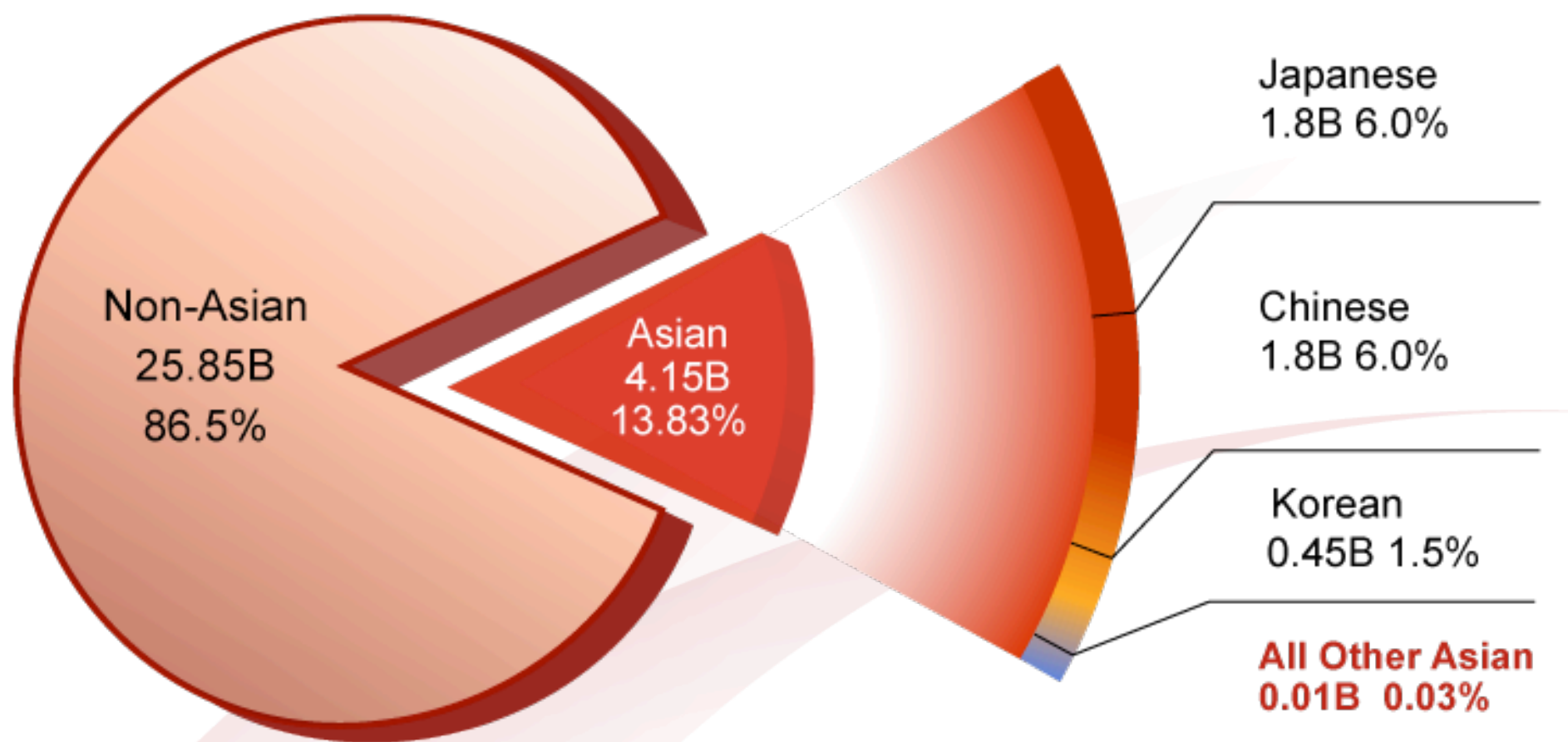


Large Buyer & Publisher Perspective



Translation Tools Vendor Perspective

Asian Language Content Gap



What makes Asia Online unique?

- The only SMT technology provider that is **also a major user** of ALT technology on one of the largest translation projects in the world - English Wikipedia (1B Words+) into 11 Asian languages using MT and crowdsourcing.
- The translation tools and technology platform used to accomplish this, is also made available as a SaaS product and onsite servers for the enterprise translation market.

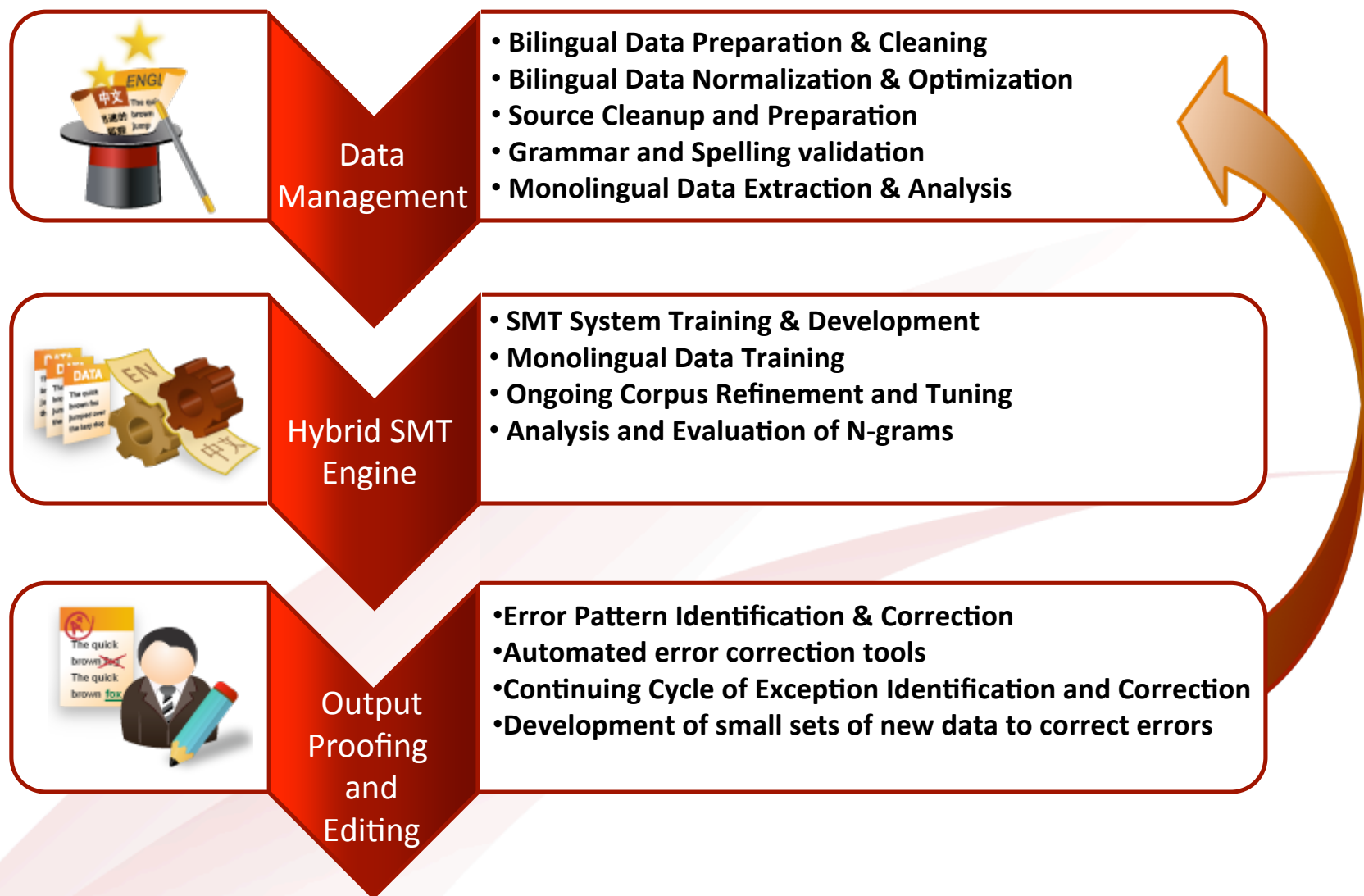


Constant dialogue between the Users and Developers to ensure that rapid translation quality evolution is a reality

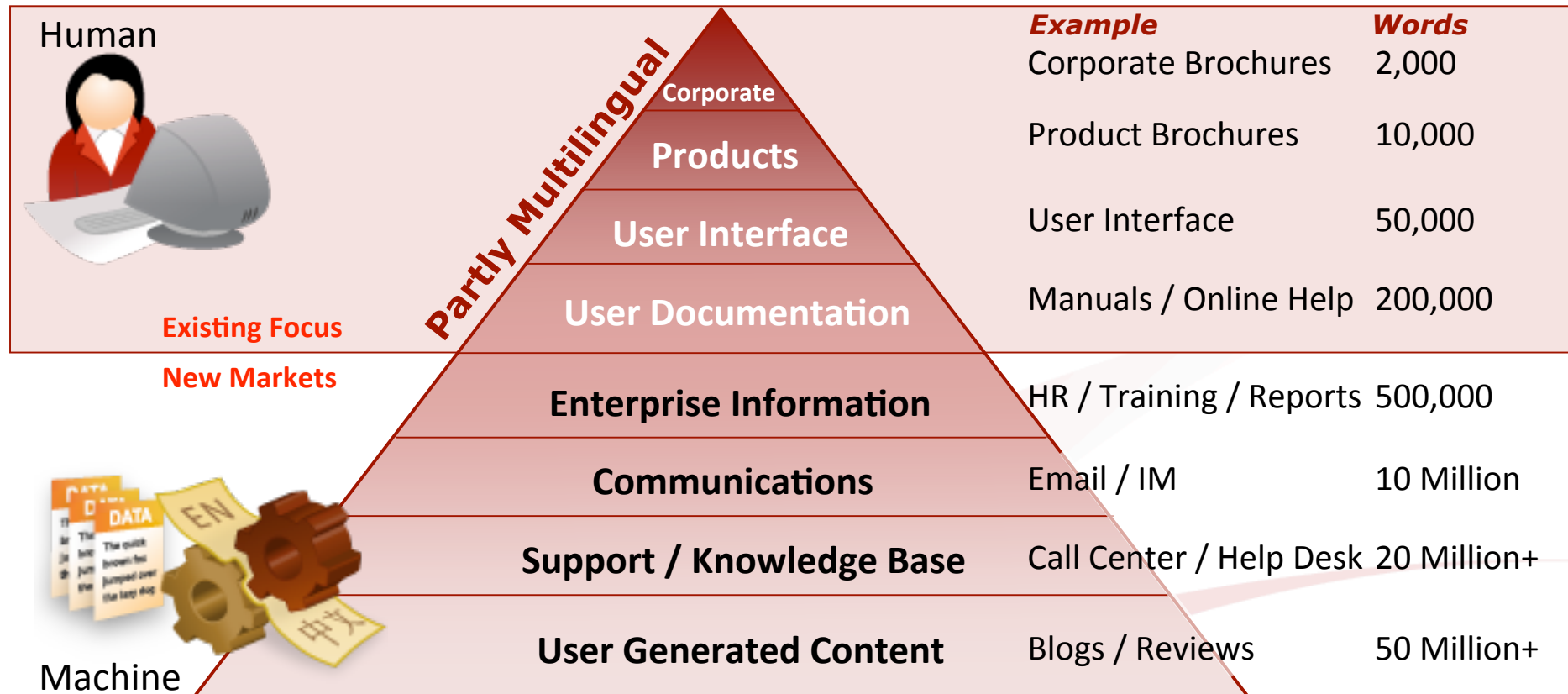
Technology Overview



The Translation Platform Overview



Expanding the Reach of Translation



Problem: • Only 0.5% of what needs to be translated today is being translated due to cost and time constraints.

Solution: • Machine translation offers , “good enough” quality alternative to human that is faster and lower cost and can deliver huge volumes of content quickly – millions of words per day.
• Where “gold-standard” translations are required, MT + human post editing can deliver the same content at high quality, significantly faster.

Connecting the Language Studio™ Suite

Customer Site

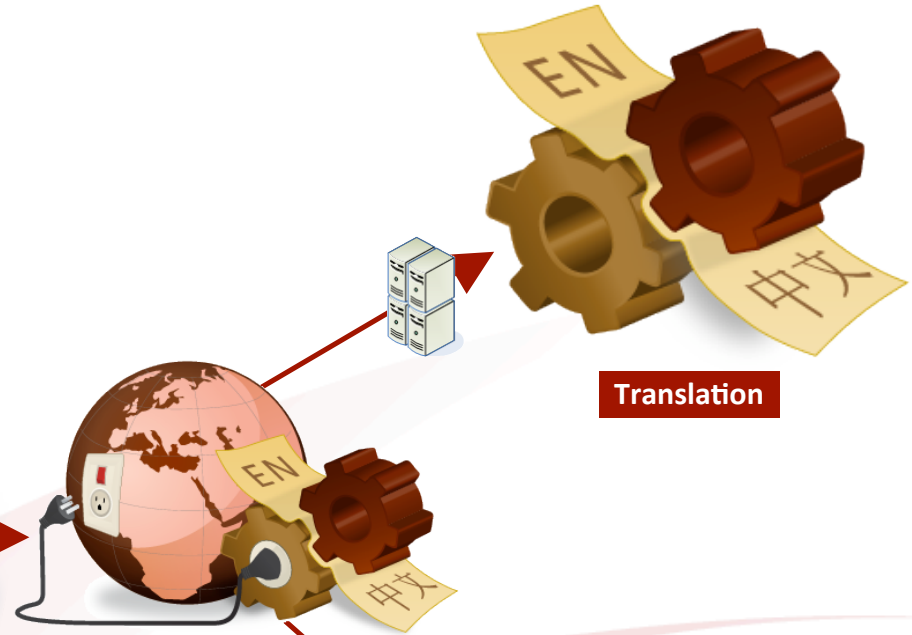
Private Data Storage

Project and Resource Management

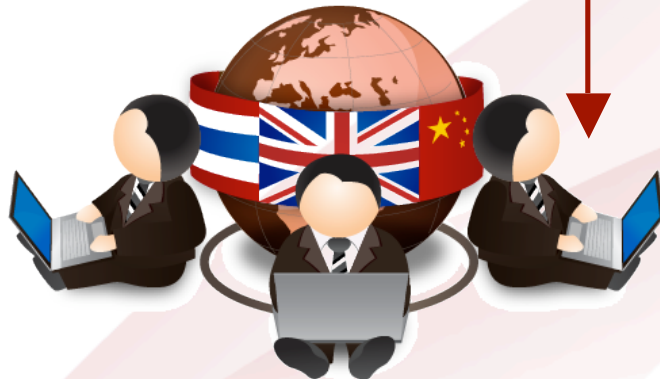


Asia Online Software as a Service

Translation



External Editors and Translators



Distributed Post Editing

Training

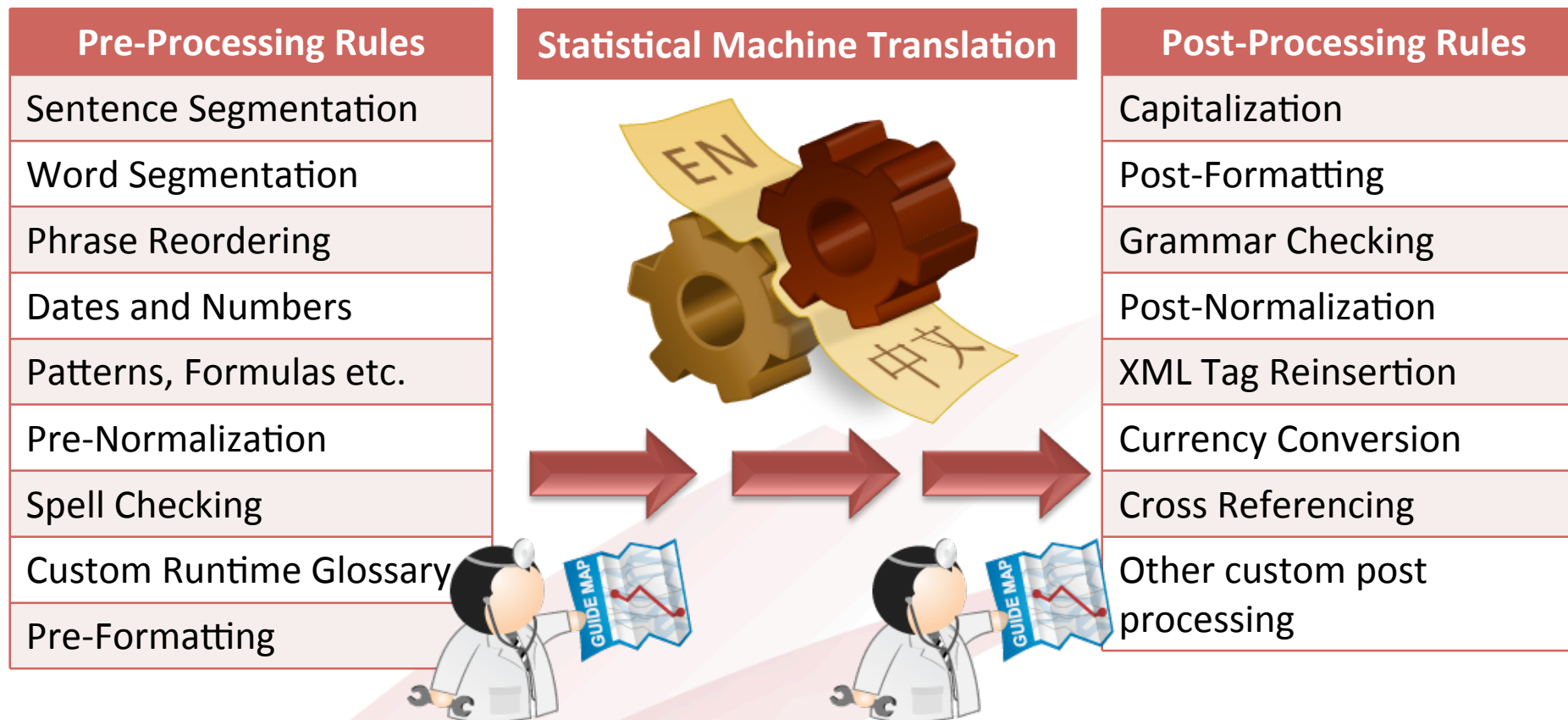


Key

- Other language pairs will be created as demand requires or based on customer request.

		AR	ID	MS	BN	BG	ZH	CS	DA	NL	EN	ET	FI	FR	DE	EL	GU	HI	HU	GA	IT	JA	KO	LV	LT	MT	PL	PT	PA	RO	RU	SK	SL	ES	SV	TL	TM	TH	VI		
Arabic	AR																																								
Bahasa Indonesia	ID																																								
Bahasa Malay	MS																																								
Bengali	BN																																								
Bulgarian	BG																																								
Chinese	ZH																																								
Czech	CS																																								
Danish	DA																																								
Dutch	NL																																								
English	EN																																								
Estonian	ET																																								
Finnish	FI																																								
French	FR																																								
German	DE																																								
Greek	EL																																								
Gujarati	GU																																								
Hindi	HI																																								
Hungarian	HU																																								
Irish	GA																																								
Italian	IT																																								
Japanese	JA																																								
Korean	KO																																								
Latvian	LV																																								
Lithuanian	LT																																								
Maltese	MT																																								
Polish	PL																																								
Portuguese	PT																																								
Punjabi	PA																																								
Romanian	RO																																								
Russian	RU																																								
Slovak	SK																																								
Slovene	SL																																								
Spanish	ES																														</										

Understanding the Language Studio™ Hybrid Rules and SMT Engine Model



Pre-Processing and Post-Processing Rules are implemented in a C# script, which can be customized to fit almost any requirements. This is a true hybrid approach, where rules are used to guide the statistical engine to greater accuracy.

Training Data	Preparation	Train	Diagnostic Engine	Train	Version Release
1 Foundation Language Pair 2 Foundation Domain 3 Custom Data → 4 Bilingual Data 5 Glossaries → 6 Phrase Pairs → (TMX, XLIFF, CSV) 7 Bilingual Documents → (MS Word, Plain Text, etc.) 8 Monolingual Data 9 URLs → 10 Plain Text 11 Tuning Set 12 Test Set	13 Cleaning 14 Data Generation 15 Validate Alignment and Quality 16 Alignment 17 Data mining for grammatical patterns <u>Optional</u> (Recommended) 18 High Frequency Term & Phrase Analysis 19 Normalization	 	20 Error Pattern Analysis & Correction <u>Optional</u> (Recommended) 21 Unknown Word Analysis 22 Normalization 23 Error Analysis	24 Post Edit and Train as required	

The System Interface & Screenshots





Library in London, says a "naive"

desire to help Moscow beat

fascism motivated him.

Blunt wrote the 30,000-word document after former prime minis

Margaret Thatcher exposed his treachery in 1979.

The revelations had led to a man who had worked as Surveyor

Queen's Pictures being stripped of his knighthood.

His version of events was given to the library in 1984, the year

his death, on condition that it was not displayed for 25 years.

In it, he describes his recruitment by Moscow: "I found that

Cambridge had been hit by Marxism and that most of my friend

among my junior contemporaries - including Guy Burgess - had

joined the Communist Party or were at least very close to it

politically."

Blunt's memoirs revealed little

about his spying activities

Aussi la

nouvelle

Blunt a écrit le document en 30 000 mot après l'ancien premier

ministre Margaret Thatcher avoir exposés son trahison en

1979.

Les révélations avait un homme qui avaient fait comme

"inspecteur des images de la Reine s'arrogant de ses

chevauchée.

Son version des événement a été donnée à la bibliothèque en

1984, l'année après son mort, à condition qu'il n'était pas pour

25 ans.

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

La BBC

politique

TOP R

panc

1 £m

"con

histoir

parta

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

2 L

1 B

Source

Target

Confidence

?

?

?

?

?

?

He supplied hundreds of secret documents to the Soviets while a wartime agent for MI5.

Il a fourni des centaines de des documents secrets à les Soviétiques alors qu'un agent de guerre pour

MI5.

Blunt was part of the infamous Cambridge spy ring, with Kim Philby, Guy Burgess and Donald

Maclean.

Blunt faisait partie de l'infameuse Cambridge espion ring, avec Kim philby, Guy burgess et Donald

MacLean.

His manuscript, at the British Library in London, says a "naive" desire to help Moscow beat fascism

motivated him.

Son manuscrit, à la bibliothèque à Londres, dit un "naïf" volonté d'aider Moscou battre fascisme le

motivée.

Blunt wrote the 30,000-word document after former prime minister Margaret Thatcher exposed his

treachery in 1979.

Blunt a écrit le document en 30 000 mot après l'ancien premier ministre Margaret Thatcher avoir

exposés son trahison en 1979.

The revelations had led to a man who had worked as Surveyor of the Queen's Pictures being

stripped of his knighthood.

Les révélations avait un homme qui avaient fait comme l'inspecteur des images de la Reine

s'arrogant de ses chevauchée.

His version of events was given to the library in 1984, the year after his death, on condition that it

was not displayed for 25 years.

Son version des événement a été donnée à la bibliothèque en 1984, l'année après son mort, à

condition qu'il n'était pas pour 25 ans.

In it, he describes his recruitment by Moscow

Dans cette fenêtre, il décrit son recrutement de Moscou

"I found that Cambridge had been hit by Marxism and that most of my friends among my junior

contemporaries - including Guy Burgess - had either joined the Communist Party or were at least

very close to it politically."

"J'ai trouvé que Cambridge a été touché par le marxisme et que la plupart de mes amis au milieu de

mes contemporaries de junior - y compris Guy Burgess- soit il a rejoint le parti communiste ou soit

il était au moins très proche de ce parti de façon politique."

Edit

Edit

Edit

Edit

Edit

Edit

Edit

Edit

Edit

Edit

Edit

Target :

Confidence : 50

Source :

EN

In it, he describes his recruitment by Moscow

FR

Dans cette fenêtre, il décrit son recrutement de Moscou

Version

Choices

Confidence level of the MT output

Changed

Completed

Proof Read

Not Reviewed

	Confidence	?			
agent de guerre pour	42	☒	☐	☐	☐
uy burgess et Donald	70	☒	☐	☐	☐
atre fascisme le	68	☒	☐	☐	☐
voir	72	☒	☐	☐	☐



Asia
Online™
The World Speaks One Language - Yours™

Language Studio™ Pro

1.0.0.0

offline

Account : Asia Online

User : Demo

**Edits are reflected real time
in the target document.**

Blunt's Soviet spying a mistake

Start Page

Views :



Programmes

Have Your Say

In Pictures

Country Profiles

Special Reports

Related BBC sites

Search

His version of events was given to the library in 1984, the year of his death, on condition that it was not displayed for 25 years.

In it, he describes his recruitment by Moscow: "I found that Cambridge had been hit by Marxism and that most of my friends among my junior contemporaries - including Guy Burgess - had joined the Communist Party or were at least very close to it politically."

Les profils de pays

Rapports spéciaux

liées BBC sites

Sport

temps

à ce jour

éditeurs de blog

Son version des événement a été donnée à la bibliothèque en 1984, l'année après son mort, à condition qu'il n'était pas pour 25 ans.

Dans cela, il décrit son recrutement de Moscou: "J'ai trouvé que Cambridge a été touché par le marxisme et que la plupart de mes amis au milieu de mes contemporains de junior - y compris Guy Burgess- soit il a rejoint le parti communiste ou soit il était au moins très proche de ce parti de façon

historique

parta

1 B

2 L

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Source

Target

Confidence

?

?

?

?

?

Blunt was part of the infamous Cambridge spy ring, with Kim Philby, Guy Burgess and Donald Maclean.
His manuscript, at the British Library in London, says a "naïve" desire to help Moscow beat fascism motivated him.
Blunt wrote the 30,000-word document after former prime minister Margaret Thatcher exposed his treachery in 1979.
The revelations had led to a man who had worked as Surveyor of the Queen's Pictures being stripped of his knighthood.
His version of events was given to the library in 1984, the year after his death, on condition that it was not displayed for 25 years.
In it, he describes his recruitment by Moscow
"I found that Cambridge had been hit by Marxism and that most of my friends among my junior contemporaries - including Guy Burgess - had either joined the Communist Party or were at least very close to it politically."

Blunt faisait partie de l'infameuse Cambridge espion ring, avec Kim philby, Guy burgess et Donald MacLean.
Son manuscrit, à la bibliothèque à Londres, dit un "naïf" volonté d'aider Moscou battre fascisme le motivée.
Blunt a écrit le document en 30 000 mot après l'ancien premier ministre Margaret Thatcher avoir exposés son trahison en 1979.
Les révélations avait un homme qui avaient fait comme l'inspecteur des images de la Reine s'arrogant de ses chevauchée.
Son version des événement a été donnée à la bibliothèque en 1984, l'année après son mort, à condition qu'il n'était pas pour 25 ans.
Dans cela, il décrit son recrutement de Moscou
"J'ai trouvé que Cambridge a été touché par le marxisme et que la plupart de mes amis au milieu de mes contemporains de junior - y compris Guy Burgess- soit il a rejoint le parti communiste ou soit il était au moins très proche de ce parti de façon politique."

70

71

68

72

71

80

50

84

?

?

?

?

?

?

?

?

?

?

?

?

?

?

?

?

?

?

?

Edit



Source :

EN

EN

**Version tracking with visual edits that
show the changes made by the user.
These changes are used later to build
corrective data.**

Version



Compare

Compare Version Result :

Version No.	Modified Date	Version	Modified By
1	9/14/2009 4:06:48 PM	Dans ce ~~tte fenêtre~~ la, il décrit son recrutement de Moscou	User1
0	9/14/2009 3:52:41 PM	Dans cette fenêtre, il décrit son recrutement de Moscou	

Choices



Asia
Online™

The World Speaks One Language - Yours™

Language Studio™ Pro

0.0.20091203.0

offline

Account :

User : ppuumm

Menu

- Home
- Start Page
- Custom Engines
- Overview
- Monolingual Data
- Bilingual Data
- Tuning Test Set Data
- Custom Engine Package
- Quality Assessment
- Human Quality Assessment
- Automated Quality Metrics
- Sentence Evaluation

Human Quality Assessment

Automated Quality Metrics Report

Automated Quality Metrics

Sentence Evaluation Report

Sentence Evaluation

Start Page

Human Quality Assessment

Enter details about the human quality assessment that you are creating

Overview

Review

Report

Assessment Title **Batch 1**

Reviewer Name: **lcr**

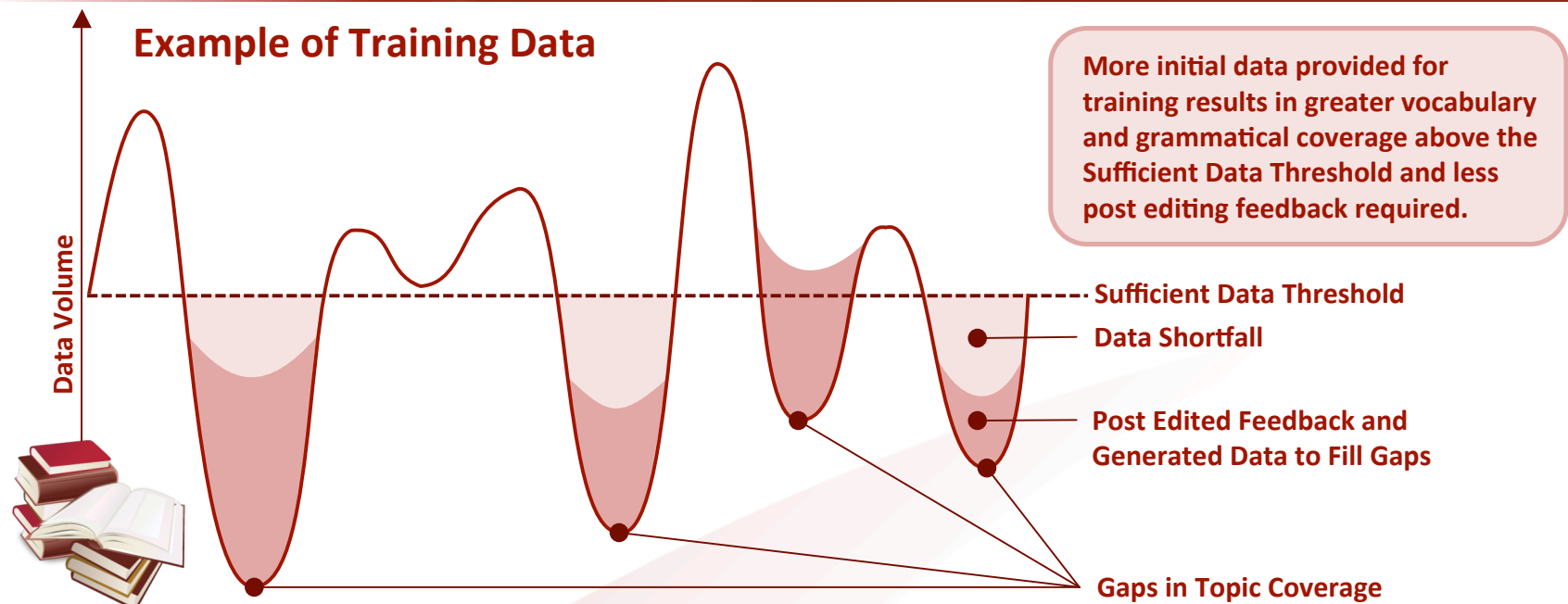
A+

A+

Save

No.	Source	Reference	Candidate	Words	x	x	x	x	x	x
1	空気送風用チャンバーは送風機にダクトを介して設けた。	The chamber for ventilating air provided in a blower through a duct.	An air chamber for the <u>air blower</u> through a duct provided.	11	0	0	2	0	0	0
2	NPC2は第二キャリアP C 2の回転速度を示す。	NPC2 represents the rotational speed of the secondary carrier PC2.	NPC2 <u>second</u> carrier PC2 represents the rotational speed of the device.	11	0	0	1	0	0	0
3	λ1は歯数比を示す。	λ1 is the gear ratio.	λ1 tooth gear ratios is shown.	6	0	0	0	0	0	0
4	送風用のチャンバーは断面視して卵形状としたチャンバー本体と、チャンバー本体の収れん部に設けたノズルとで構成した。	The chamber for ventilation includes a chamber body having an oval-shaped cross-section, and a nozzle provided in a collective part of the chamber body.	Chamber for ventilation of the oval-shaped cross section and a bottom chamber body <u>10, screws of a</u> chamber body provided with a nozzle.	23	0	3	0	0	0	0
5	レンガ解体装置はその他の種々の構造のレンガ解体装置であっても良い。	The brick dismantling device may also be a brick dismantling device having various other structures.	The brick dismantling device brick dismantling device good even in a variety of other structures.	15	0	0	0	0	0	0
6	レンガ解体装置は先端に耐火レンガの解体部材が取り付けられた少なくとも一対の伸縮棒状部材からなるものである必要は必ずしもない。	The brick dismantling device <u>does not necessarily need to be composed of at least a pair of telescopic rod-shaped members</u> attached to a refractory brick dismantling members at the tip.	A <u>distal end</u> of the brick dismantling device <u>fire</u> brick dismantling member is attached to at least one pair of telescopic rod-shaped members is that which comprises all necessary.	29	0	0	2	1	0	1
7	NPC1は第一キャリアP C 1の回転速度を示す。	NPC1 represents the rotational speed of the primary carrier PC1.	NPC1 first carrier PC1 represents the rotational speed of the <u>vehicle</u> .	11	0	1	0	0	0	0
8	バイパス通路20の両側には切換弁28, 30を設ける。	A bypass passage 20 is provided on both sides of the change-over valves 28, 30.	A bypass passage 20 of the change-over valves 28, 30 on both sides is provided.	15	0	0	0	0	0	0
9	前記切り欠き部に前記電磁膨張弁の外方を覆うカバー部材を設けることもできる。	The above notched section can have a cover member to cover the outer side of the above solenoid expansion valve	The notch can be provided in the cover member for covering the outer side of the solenoid expansion valve	19	0	0	0	0	0	0

Post Editing Reduces Data Shortfall



- Training data can often have gaps in coverage and an excess of data in other areas.
- Gaps in coverage reduce translation quality.
- Gaps can quickly be filled via post editing the machine translated output and submitting the data back to the system for further learning.
- Many gaps can be filled with monolingual data only.
- Further gaps can be identified and resolved by analyzing the text that is to be translated for high frequency terms and unknown words
- In some cases incorrect data may be statistically more relevant. Post editing will raise the relevance of the correct grammar.



Asia
Online™
The World Speaks One Language - Yours™

Language Studio™ Pro

1.0.0.0

offline

Account : Asia Online

User : Demo

Menu

Corrective Data - Blunt's Soviet spying a mistake

Blunt's Soviet spying a mistake

Start Page

Corrective Data - Blunt's Soviet spying a mistake

Provide corrective data to improve future translations.

Overview

Monolingual Data

Bilingual Data

Phrase :

Add

Process

Stop

☒	Phrase	Count	
☒	Dans cela	6	X
☒	cela veut	5	X

Extracted from
the corrections in
the post editor

Advanced concordance
search facilities

Scope ☐ All ☐ Select... ☐ Regular Expression
Find
Replace ☐ Whole Words Only
☐ Match Case

☒	Concordance	?	X	✓
☐	Audio Archive dans cela neant	☒	☐	☐
☐	whiteoutwhiteout - dans cela neant	☒	☐	☐
☐	whiteout" dans cela neant" epKuK 004december 2007free web-releaseThis audio is part of the collection: Open Source AudioArtist/Composer whiteoutSource: http://www.koksundkohlede	☒	☐	☐
☐	Je n'ai aucun intérêt dans cela si ce n'est de rendre ou plutôt de donner l'occasion à des personnes de se rendre compte qu'ils peuvent gérer eux même leur santé ... et éventuellement aller plus loin en apprenant en autre les fondements de la Bio-Electronique Vincent qui est un système qui affranchi l'homme de ses peurs en lui apprenant que le milieu conditionne l'apparition de tel ou tel syndrome (ou maladie) et que de fait cet homme devient responsable de ce qui lui échoie, et donc ne peut se plaindre à personne des conséquences de son ignorance et des ses mauvaises habitudes de vie.	☒	☐	☐
☐	Il a plusieurs avantages uniques par rapport à d'autres systèmes dans cela agit en tant que passeport pour MSN, Office de phase, et emplacements de passeport de Microsoft, y compris les goûts de Hotmail (maintenant 5 gigaoctets de l'espace mémoire), espaces, et OneCare.	☒	☐	☐
☐	aide dans cela je considère Vos suggestions et idées Seront fortement	☒	☐	☐
☐	Experts cela veut dire: "Mistinguette"?	☒	☐	☐
☐	French Language - Qu'est-ce que cela veut dire: "Mistinguette"?	☒	☐	☐
☐	qu'est ce que cela veut dire?	☒	☐	☐
☐	Contributions: ce que cela veut dire	☒	☐	☐
☐	5 commentaires pour "Contributions: ce que cela veut dire"	☒	☐	☐

Advanced Concordance Generator Settings:

Save Settings

Words: Include
Exclude
Sites: Include
Exclude
Language/Country
Case-sensitive ☐
Performance 1
Maximum: URLs 5
URLs/Site 5
Sentences/URL 5

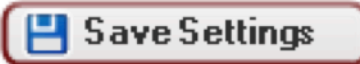
Search Provider

☒ Google
☒ Bing
☐ Yahoo
☐ AltaVista

The best translations come from clean data.
Tools determine clean data, but give user the
chance to review and remove out of domain
data and other unwanted data.

Save

Advanced Concordance Generation

Advanced Concordance Generator Settings: 

Words: Include
Exclude

Sites: Include
Exclude

Language/Country

Case-sensitive ☐

Performance 1

Maximum: URLs 5
URLs/Site 5
Sentences/URL 5

☒ Search Provider
☒ Google
☒ Bing
☐ Yahoo
☐ All Sites

Searches can be constrained and refined by many factors.

Example:

- For Mexican Spanish, set language to Spanish and Country to Mexico.
- For Table (as in Database), set Sites to include “Oracle, Microsoft, ZDNet etc.”
- To avoid overload of one particular style, limit the amount of return data by site.
- Add custom search providers such as existing TMS or public search engines.

Post Editing Errors Amplifies Corrections

The quick brown fox **over jumps** the lazy dog

The quick brown fox **jumps over** the lazy dog

Additional corrective data
generated by Language Studio™ Pro

Buddha	jumps over	the wall
Siemens Wind Power CEO	jumps over	to Repower
Judge	jumps over	bench in courtroom melee
Military surveillance bot	jumps over	25 foot walls
Robbie Maddison	jumps over	Tower Bridge on motorbike
Man	jumps over	Grand Canyon
Cow	jumps over	Moon
With IE9 in sight, Firefox	jumps over	50% market share mark
Long jumper Brian Thomas	jumps over	a car to raise money
Rally car	jumps over	a crazy fan!
Kobe	jumps over	a speeding Aston Martin
A deer	jumps over	a motorcycle
A woman jogging in a California state park	jumps over	a 100-foot cliff to get away from attacker
An Afghan Army soldier	jumps over	a irrigation canal while conducting a foot patrol

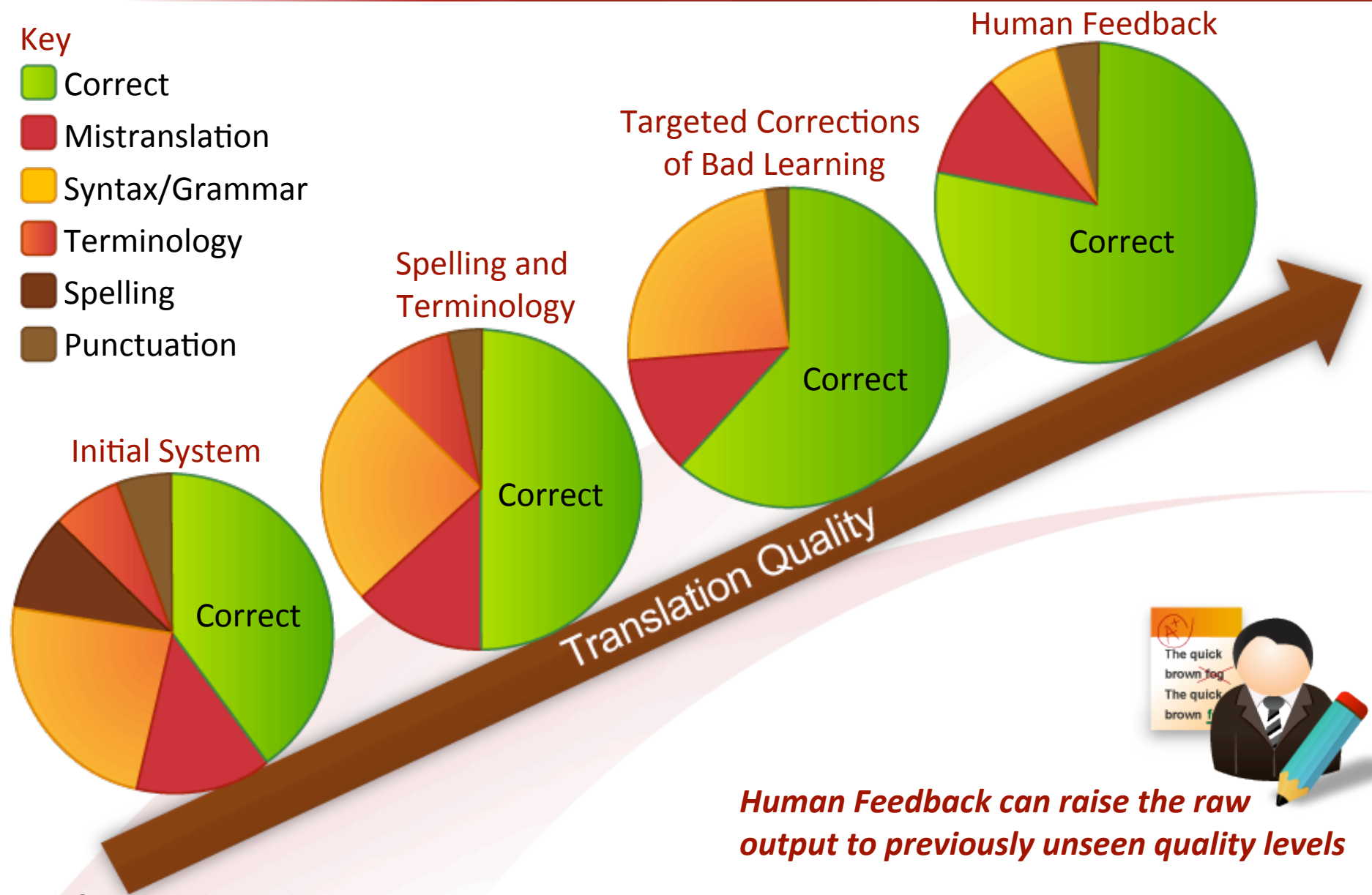


Language Studio™ Pro analyzes corrections and generates other examples that include the corrected phrase to fill gaps in grammatical patterns. Each post edited correction is amplified producing many other corrective patterns, improving future translations.

Quality Evolution via Corrective Feedback

Key

- Correct
- Mistranslation
- Syntax/Grammar
- Terminology
- Spelling
- Punctuation



15 Domains – Focused Quality Translation

- The highest quality translation engines are specialized on domains.
- 15 pre-created domains provide a solid foundation upon which high quality customized engines for clients can quickly be created.
- A major element and accelerator when implementing stylized translations specific to clients.
- This is unique in the industry and gives Asia Online a significant competitive advantage.
- Domains can be combined.

	Domain
1	Law and Justice
2	Sports and Leisure
3	Technology and High-Tech
4	Military and Defense
5	Medical and Healthcare
6	Religion and Philosophy
7	Politics and Government
8	Economics and Finance
9	Science and Nature
10	Arts and Entertainment
11	Geography and Places
12	History and Events
13	Social and Culture
14	Blogs and Freeform Text
15	News and Commentary

Thank You

kirti.vashee@asiaonline.net

Follow me on Twitter:
[@kvashee](https://twitter.com/kvashee)



Join the Automated Language Translation
Translation Group in Linked In to
monitor MT discussions

