

Evaluation Report for Customization of Amazon Translate Active Custom Translation with TAUS Financial Domain Data

Language Pair

English → Dutch

Domain

Financial

Introduction

Online machine translation engines provide easy access to high quality machine translations. These machine translation engines are optimized for content like news articles and social media posts that end users of these online platforms frequently translate.

Businesses often want to translate text with a different style and a specific topic. For enterprise use, online machine translation engines offer customization with existing translations that reflect the desired style and topic.

TAUS makes such customization data available via the TAUS Data Marketplace and TAUS Matching Data platforms, and now AWS Marketplace.

So that business can clearly assess the value of TAUS data, Polyglot Technology LLC was tasked to independently evaluate the quality of machine translation of Amazon Translate customized with TAUS Data (using [Amazon Translate Active Custom Translation](#)) compared to non-customized Amazon Translate.

BLEU Scores for Amazon Translate and Amazon Translate Active Custom Translation

Machine Translation Evaluation

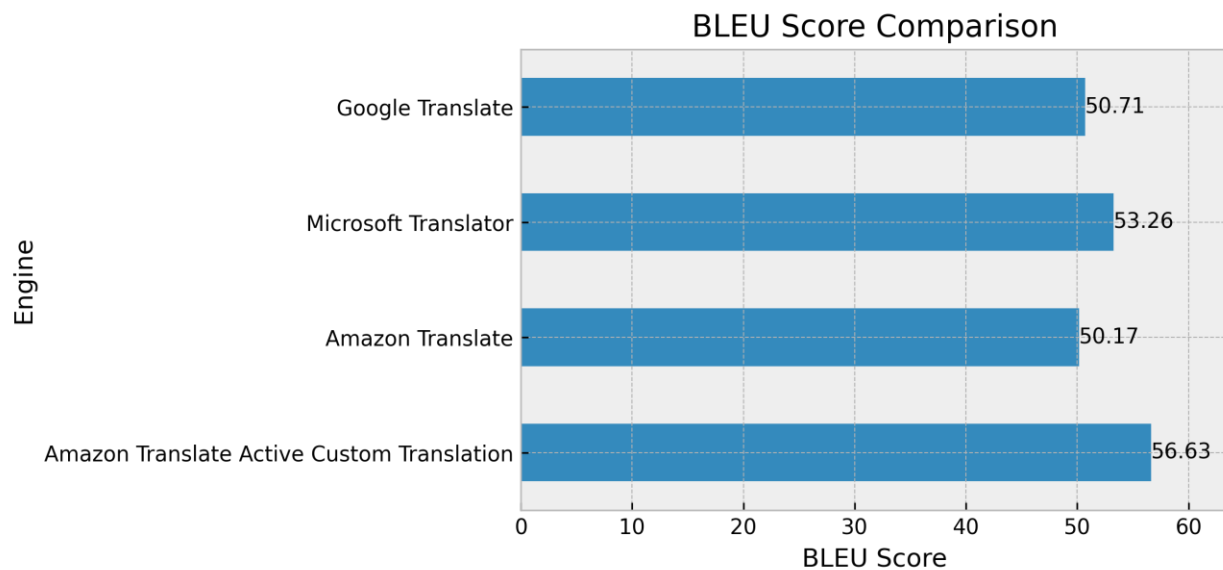
To judge whether machine translation is good or not, human evaluation is the best method. We can ask speakers of the source language and the target language, or better professional translators, to judge whether a machine translation is an adequate and fluent translation of the original text. Or we can ask how close the machine translation is to a human reference translation. Human evaluation however, is slow and hard to scale along language pairs and domains.

Automatic metrics that also use human reference translations have been developed to calculate a numeric score for machine translation quality. For close to 20 years the predominant automatic metric is BLEU, measuring the similarity of machine translations to human reference translations on a scale from 0 to 1 (or 0 to 100 when expressed as percentages). More details on BLEU and how to interpret it can be found in the section “Interpreting BLEU Scores” below.

TAUS Test Set

TAUS selects the machine translation customization data by querying its large repository of high-quality translation data with a domain-specific text. The resulting customization dataset is then split at random into a larger training set for Amazon Translate Active Custom Translation and a smaller 2,000 sentence test set that was provided to Polyglot Technology for evaluation.

BLEU Score Results for the TAUS Test Set



BLEU Score Results for Publicly Available Test Sets for the Language Pair

No widely shared public test sets are available for the language pair English-Dutch.

Description of Publicly Available Test Sets

wmt: News text test sets published by the Conference on Machine Translation¹

iwslt: Transcribed TED talk test sets published by the Conference on Spoken Language Translation²

Use Case Specific Evaluation

When employing machine translation for a specific use case, it is advisable to evaluate the systems with usage-scenario specific source text and its human reference translation. Maybe you have already data from a previous, similar project, or your translation vendor can help you create the test data. Polyglot Technology can assist in implementing a robust evaluation program.

¹ More details on the data sets can be found out e.g. for the 2020 edition on <http://statmt.org/wmt20/translation-task.html>

² Conference website: <https://iwslt.org/>

Interpreting BLEU Scores

The paragraphs in this section are adapted from Google AutoML Translate's [documentation page on evaluation](#) which is licensed under the [Creative Commons 4.0 Attribution License](#).

[BLEU \(BiLingual Evaluation Understudy\)](#) is a metric for automatically evaluating machine-translated text. The BLEU score is a number between zero and one that measures the similarity of the machine-translated text to a set of high quality reference translations. A value of 0 means that the machine-translated output has no overlap with the reference translation (low quality) while a value of 1 means there is perfect overlap with the reference translations (high quality).

It has been shown that BLEU scores correlate well with human judgment of translation quality. Note that even human translators do not achieve a perfect score of 1.0 (for the reason that a source sentence can have several valid, equally appropriate translations).

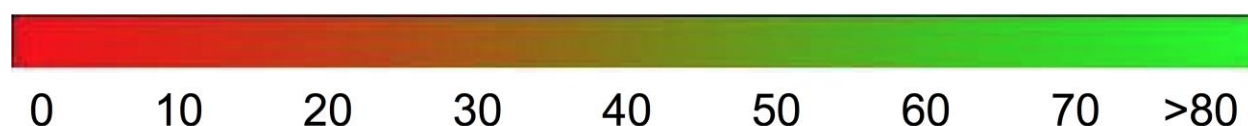
Interpretation

Trying to compare BLEU scores across different corpora and languages is strongly discouraged. Even comparing BLEU scores for the same corpus but with different numbers of reference translations can be highly misleading.

However, as a rough guideline, the following interpretation of BLEU scores (expressed as percentages rather than decimals) might be helpful.

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

The following color gradient can be used as a general scale [interpretation of the BLEU score](#):



Segment Analysis

To determine which translations improved the most between the non-customized and the customized Amazon Translate we ranked the translations by the most improved [COMET](#) metric³.

The table below shows the 30 most improved translations with a visualization of edits that have to be applied to transform the Amazon Translate translation into the Amazon Translate Active Custom Translation.

source	reference	difference
supply of wholesale gas for Denmark (and potentially also for Sweden),	levering van groothandelsgas voor Denemarken (en mogelijkerwijs ook voor Zweden);	levering van wholesale <u>groothandel</u> gas voor Denemarken (en mogelijk ook voor Zweden);
9.3.1. Short-term impact	9.3.1. Effecten op korte termijn	9.3.1. Impact op korte <u>termijn</u>
When the first marketing takes place in a Member State other than the flag Member State, the competent authorities of the Member State where the first marketing takes place shall ensure that a	Wanneer de producten voor de eerste keer op de markt worden gebracht in een andere lidstaat dan de vlaggenlidstaat, zorgen de bevoegde autoriteiten van de lidstaat waar de producten voor de eerste keer op de markt worden gebracht, ervoor dat een kopie van de gegevens uit het verkoopdocument na	Wanneer de <u>verkoop voor de</u> eerste hande <u>maal</u> plaatsvindt in een andere lidstaat dan de vlaggenlidstaat, zorgen de bevoegde autoriteiten van de lidstaat waar de eerste handel <u>plaatsvindt, maal op de markt wordt gebracht</u> ervoor dat een kopie van de gegevens van de verkoop bijetten <u>notities</u> langs

³ BLEU works only really well on a corpus (entire text) level than on an individual sentence level

source	reference	difference
copy of the sales notes data is submitted by electronic means to the competent authorities of the flag Member State on receipt of relevant information.	ontvangst van de desbetreffende informatie elektronisch bij de bevoegde autoriteiten van de vlaggenlidstaat wordt ingediend.	elektronische weg <u>wordt ingediend</u> bij de bevoegde autoriteiten van de vlaggenlidstaat <u>wordt ingediend op</u> ontvangst van relevante informatie.
Road tolls for means of transport with a mass greater than 3500 kg other than means of transport for paying passengers	Wegentol voor vervoermiddelen met een massa van meer dan 3500 kg, andere dan voor het vervoer van betalende passagiers	<u>Tolheffing</u> <u>Wegentol</u> voor vervoermiddelen met een massa van meer dan 3500 kg, <u>andere</u> dan <u>vervoermiddelen voor</u> <u>oor het vervoer van</u> betalende passagiers
Without prejudice to the applicable rules of penal law, The Contracting Parties shall refrain from instituting proceedings in respect of unpremeditated or inadvertent infringements of the law which come to their attention only because they have been reported under the national mandatory occurrence-reporting scheme, except in cases of gross negligence.	Onverminderd de toepasselijke strafbepalingen, onthouden de partijen zich van het instellen van procedures naar aanleiding van onopzettelijk of uit onachtzaamheid begane overtredingen, indien hiervan alleen maar kennis is verkregen door een melding in het kader van de nationale regeling voor de verplichte melding van voorvallen, behalve indien er sprake is van grove nalatigheid.	<u>Onverminderd de toepasselijke strafrechtelijke regels, regels van het strafrecht</u> onthouden de overeenkomstsluitende partijen <u>zich van het instellen van een procedure ten aanzien van ongevoorboden of onbedoeld</u> <u>geen procedures in te stellen met betrekking tot onberaamde of onopzettelijke</u> inbreuken op het recht, die alleen onder hun aandacht komen omdat zij zijn gemeld <u>in het kader van onder</u> de nationale verplichte <u>voorvalmelding, gebeurtenisrapportage</u> -regeling, behalve in geval van grove nalatigheid.
However, the containers may be used only once during each stay in a Member State, for transporting goods which are loaded within the territory of that Member State and are intended to be unloaded within the territory of the same Member State, where the containers would otherwise have to make a journey unladen within that territory.	De containers mogen echter tijdens elk verblijf in een Lid-Staat één enkele keer worden gebruikt voor het vervoer van goederen die binnen het grondgebied van die Lid-Staat worden geladen en gelost, wanneer de containers in die Lid-Staat anders leeg vervoerd zouden moeten worden.	<u>De containers mogen echter slechts eenmaal worden gebruikt</u> tijdens elk verblijf in een lidstaat, <u>één enkele keer worden gebruikt</u> voor het vervoer van goederen die <u>op binnen</u> het grondgebied van die lidstaat worden geladen en <u>bestemd zijn om op het grondgebied van dezelfde lidstaat te worden</u> gelost, wanneer de containers anders <u>leen reis onbeladen vervoerd</u> zouden moeten <u>mak</u> worden. binnen dat <u>gebied</u> <u>territorium</u> .
(1) The amendments introduced with regard to the place of supply of services by Council Directive 2008/8/EC of 12 February 2008 amending	(1) De bij Richtlijn 2008/8/EG van de Raad van 12 februari 2008 tot wijziging van Richtlijn 2006/112/EG wat betreft de plaats van een dienst aangebrachte wijzigingen met betrekking tot de plaats van een dienst [3] brengen met zich dat	(1) De <u>door bij</u> Richtlijn 2008/8/EG van de Raad van 12 februari 2008 tot wijziging van Richtlijn 2006/112/EG <u>van de Raad van 12 februari 2008 tot wijziging van Richtlijn 2006/112/EG wat betreft de plaats van een dienst aangebrachte</u>

source	reference	difference
Directive 2006/112/EC as regards the place of supply of services [3] mean that services to taxable persons are supplied principally where the recipient is established.	diensten aan een belastingplichtige hoofdzakelijk geacht worden te zijn verricht op de plaats waar de afnemer gevestigd is.	wijzigingen met betrekking tot de plaats van dienst verricht een diensten [3] betekenen rengen met zich dat diensten aan belastingplichtigen hoofdzakelijk worden verricht wanneer op de plaats waar de ontvanger is gevestigd is.
as listed in Annex IV, shall be frozen;	als genoemd in bijlage IV, worden bevroren.	zoals die vermeld staan in bijlage IV, moeten worden ingevroren;
The vote will take place in a moment.	De stemming vindt over enkele ogenblikken plaats.	De stemming vindt in een oogwenzo dadelijk plaats.
Member States may, however, grant exemption for more than Ö EUR Ö 200 provided that the value of each exempt present does not exceed Ö EUR Ö 1 000.	De lidstaten kunnen evenwel een vrijstelling van meer dan 200 Ö EUR Ö verlenen, mits de waarde van elk geschenk dat met vrijstelling mag worden ingevoerd, niet meer bedraagt dan 1 000 Ö EUR Ö.	De lidstaten kunnen echter evenwel een vrijstelling verlenen voor meer dan Ö EUR Ö 200, an meer dan Ö 200 EUR verlenen mits de waarde van elk geschenk dat met vrijgestelde aanwezig niet hoger is lling mag worden binnengebracht, niet meer bedraagt dan Ö EUR Ö 1 000 EUR.
First fixing date determined by the earliest date of all the fixings.	De eerste fixingsdatum, bepaald door de vroegste datum van alle fixingen.	Eerste vaststelt fixings datum, bepaald door de vroegste datum van alle bevestig fixingen.
is executed in a decentralised manner by the NCBs, without prejudice to Article 45(3).	zij worden decentraal uitgevoerd door de NCB's, zonder afbreuk te doen aan artikel 45, lid 3.	wordt gezij worden decentraaliseerd uitgevoerd door de NCB's, zoverminderder afbreuk te doen aan artikel 45, lid 3.
(e) information forwarded to one Member State by the competent authority of the other Member State has enabled information to be obtained which may be relevant in assessing liability to tax in the latter Member State.	e) de aan de ene lidstaat door de bevoegde autoriteit van de andere lidstaat verstrekte inlichtingen hebben informatie opgeleverd die voor de vaststelling van de belastingschuld in de andere lidstaat van nut kan zijn.	e) informatie diede aan de ene lidstaat door de bevoegde autoriteit van de andere lidstaat aan een lidstaat is toegezonden, heeft het mogelijk gemaakt informatie te verkrijgen die relevant kan zijn voor de beoordeling van de belastingverplichting in laatstgenoemde lidstaat verstrekte inlichtingen hebben informatie opgeleverd die voor de vaststelling van de belastingschuld in de andere lidstaat van nut kan zijn.
Start installing MSI Package ...	Installatie van MSI-pakket starten...	Begin met het installeren van MSI-pakket...
recommendation concerning medical care and sickness benefits	Aanbeveling betreffende geneeskundige verzorging en uitkeringen bij ziekte	aAanbeveling betreffende medische verzorging en ziektevoordelen uitkeringen bij ziekte
Deferred Income Tax Liability	Uitgestelde inkomstenbelastingverplichting	Verplichting tot uUitgestelde inkomstenbelastingverplichting
(g) the importation of goods by international bodies recognised as such by the public authorities of	g) de invoer van goederen verricht door internationale instellingen die als zodanig door de overheid van de lidstaat waar zij zijn gevestigd, zijn erkend, alsmede door	g) de invoer van goederen door internationale instantiesellingen die als zodusdanig zijn erkend door de overheidsinstanties van de lidstaat van

source	reference	difference
the host Member State, or by members of such bodies, within the limits and under the conditions laid down by the international conventions establishing the bodies or by headquarters agreements;	de leden van deze instellingen, zulks binnen de beperkingen en onder de voorwaarden die zijn vastgesteld bij de internationale verdragen tot oprichting van deze instellingen of bij de vestigingsovereenkomsten;	ontvangst waar zij zijn gevestigd, zijn erkend, of door de leden van de gelijke organen , ze instellingen, zulks binnen de grenz beperkingen en onder de voorwaarden die zijn vastgesteld in bij de internationale verdragen tot oprichting van de organen ze instellingen of in bij de vestigings overeenkomsten inzake het hoofdkantoor;
Except Estonia, Latvia, Lithuania and Poland.;	Met uitzondering van Estland, Letland, Litouwen en Polen.”.	Behalve Met uitzondering van Estland, Letland, Litouwen en Polen. ;
Returns the base year for a given year and period.	Retourneert het basisjaar voor een gegeven jaar en periode.	Geeft als resultaat Retourneert het basisjaar voor een bepaald jaar en <u>een</u> <u>bepaalde</u> periode.
(b) a decision in accordance with point (a), but modifying the conditions under which the test is to be carried out;	b) een besluit als bedoeld onder a), waarbij echter de voorwaarden waaronder de test moet worden uitgevoerd worden gewijzigd;	b) een besluit overeenkomstig punt als bedoeld onder a), na waar tot wijziging van bij echter de voorwaarden waaronder de test moet worden uitgevoerd <u>worden</u> gewijzigd;
Please note that pursuant to Council Regulation (EC) No 1/2005 animals are to be checked by the competent authorities of the Member States to check if they are fit to continue the journey after entry into the Community.	overeenkomstig Verordening (EG) nr. 1/2005 van de Raad zullen de dieren door de bevoegde autoriteiten van de lidstaten worden gecontroleerd om na te gaan of zij na binnenkomst in de Gemeenschap geschikt zijn voor verder vervoer.	Houd er rekening mee dat overeenkomstig grond van Verordening (EG) nr. 1/2005 van de Raad dieren door de bevoegde autoriteiten van de lidstaten moeten worden gecontroleerd om na binnenkomst in de Gemeenschap na binnenkomst in de Gemeenschap te controleren of zij geschikt zijn om de reis voort te zetten.
Temporary difficulties in supply	Tijdelijke moeilijkheden bij de levering	Tijdelijke problemen bij het aanbod de levering
Foliage, branches and other parts of plants, without flowers or flower buds, and grasses, mosses and lichens, being goods of a kind suitable for bouquets or for ornamental purposes, fresh, dried, dyed, bleached, impregnated or otherwise prepared	Loof, bladeren, twijgen, takken en andere delen van planten, zonder bloemen, bloesems of bloemknoppen, alsmede grassen, mossen en korstmossen, voor bloemstukken of voor versiering, vers, gedroogd, gebleekt, geverfd, geïmpregneerd of op andere wijze geprepareerd	GeLoof, bladeren, takken en andere delen van planten, zonder bloemen of bloemknoppen, en grassen, mossen en korstmossen, zijnde goederen van de soort die geschikt zijn voor boeketten of voor versier doeleinden, vers, gedroogd, geverfd , gebleekt, geïmpregneerd of op anderszins bereide wijze geprepareerd
- if the participating Member State concerned acts in compliance with notices given in accordance with Article 104c (9).	- indien de betrokken deelnemende lidstaat gevolg geeft aan de aanmaningen op grond van artikel 104 C, lid 9.	- indien de betrokken deelnemende lidstaat handelt in overeenstemming met de overeenkomstig artikel 104 C, lid 9, gegeven kennisgevingen gevolg geeft aan de aanmaningen op grond van artikel 104 C, lid 9.

source	reference	difference
trade and consumption	handel en verbruik	hHandel en consumptieverbruik
effected within a Member State which is a party to the North Atlantic Treaty and intended either for the use of the forces of other States which are parties to that Treaty or of the civilian staff accompanying them, or for supplying their messes or canteens when such forces take part in the common defence effort.	verricht in de lidstaten die partij bij het Noord-Atlantisch Verdrag zijn en bestemd voor de strijdkrachten van de andere Staten die partij bij dat Verdrag zijn, ten behoeve van deze strijdkrachten of het hen begeleidend burgerpersoneel of voor de bevoorrading van hun messes of kantines, voor zover de betreffende strijdkrachten deelnemen aan de gemeenschappelijke defensie-inspanning .	uitgevoerd in een lidstaatverricht in de Lid-Staten die partij is-bij het Noord-Atlantisch Verdrag zijn en bestemd is-voor het gebruik van de strijdkrachten van de andere sStaten die partij zijn-bij dat Verdrag of van het civiele personeel dat hen vergezelt, of voor het leveren zijn, ten behoeve van deze strijdkrachten of het hen begeleidend burgerpersoneel of voor de bevoorrading van hun rommelmesses of kantines wanneer dergelijk, voor zover de betreffende streepijdkrachten deelnemen aan de gemeenschappelijke defensie-inspanning.
Xylofagou (4107) (excluding the part within the limits of the sovereign bases area)	Xylofagou (4107) (met uitzondering van het deel binnen de grenzen van het gebied van de Soevereine Bases)	Xylofagou (4107) (exclusief met uitzondering van het deel binnen de grenzen van het gebied van sde Soevereine bBases)
CEN/TR First committee draft Munich (March 2005) — Mechanical vibration - Guideline for the assessment of exposure to wholebody vibration of ride on operated earthmoving machines.Using harmonised data measured by international institutes, organisations and manufacturers.	CEN/TR Eerste ontwerp comité München (maart 2005) — Mechanische trillingen - Richtlijn voor de beoordeling van de blootstelling van het menselijk lichaam aan trillingen van grondverzetmachines met meerrijdende bestuurder — Gebruik van geharmoniseerde gegevens gemeten door internationale instituten, organisaties en fabrikanten.	CEN/TR Ontwerp van de eerste commissieEerste ontwerp comité München (maart 2005) — Mechanische trillingen — Richtsnoerlijn voor de beoordeling van de blootstelling aan trillingen van het helevan het menselijk lichaam vaan trijgedrag op geopereerdellingen van grondverzetmachines.M met behulpmeerrijdende bestuurder — Gebruik van geharmoniseerde gegevens gemeten door internationale instituten, organisaties en fabrikanten.
Collection services of non-hazardous non-recyclable waste, other	Inzameling van ongevaarlijk niet-recycleerbaar afval, ander	Inzamelingsdiensten van niet- van ongevaarlijk niet-recycleerbaar afval, andere
‘Separately, the total fees for the financial year charged by the statutory auditor or audit firm for the statutory audit of the consolidated accounts, the total fees charged for other assurance services, the total fees charged for tax advisory services and the total fees charged for other non-audit services.’	afzonderlijk, de tijdens het boekjaar door de wettelijke auditor of het auditkantoor aangerekende totale honoraria voor de wettelijke controle van de jaarrekening, de totale honoraria voor andere assurance-opdrachten, de totale honoraria voor belastingadviesdiensten en de totale honoraria voor andere niet-controlediensten.”.	„Afzonderlijk, de totale vergoedingen voor het boekjaar die door de wettelijke auditor of het auditkantoor in rekening worden gebracht aangerekende totale honoraria voor de wettelijke controle van de geconsolideerde rekeningen, de totale vergoedingen die in rekening worden gebracht honoraria voor andere verzekeringsassurancediensten, de totale vergoedingen die in rekening worden gebracht voor belastingadviesdiensten en de totale vergoedingen die in rekening worden gebracht honoraria voor belastingadviesdiensten en de totale

source	reference	difference
		honoraria voor andere niet-auditcontrole diensten.”

Results Summary

By customizing Amazon Translate Active Custom Translation with TAUS Financial domain English-Dutch training data the test set BLEU score improves by more than 6 points. This is a significant improvement in machine translation quality for this language pair and domain.

This numerical score improvement is supported by analyzing segments that improved significantly according to the semantic quality measure COMET: translations became more fluent and adhered better to the terminology used in the Financial domain.

Polyglot Technology

Polyglot Technology LLC helps customers succeed with machine translation by ensuring that they make best use of data available to them, by assessing machine translation quality independent from MT vendors and by advising customers on how to best integrate the technology with people and processes.