ROCZNIKI HUMANISTYCZNE
Tom LXXIII, zeszyt 10 – 2025
ZESZYT SPECJALNY / SPECIALE UITGAVE
DOI: https://doi.org/10.18290/rh257310.4s

GYS-WALT VAN EGDOM
CHRISTOPHE DECLERCQ

EEN PSYCHOLOGISCH PERSPECTIEF OP MACHINEVERTALING: DE PERCEPTIE VAN VERTAALKWALITEIT IN HET VERTAALONDERWIJS

Abstract. Binnen de vertaalwetenschap wordt er al jaren onderzoek naar de kwaliteit van machinevertaling verricht, maar de perceptie van die kwaliteit door vertaalstudenten is tot op heden onderbelicht. Toch heeft de perceptie van kwaliteit een grote invloed op de beslissing om machinevertaling te gebruiken en op de verwerking van de output als onderdeel van het vertaalproces. In dit artikel wordt verslag uitgebracht over een kleinschalig onderzoek naar de perceptie van de kwaliteit van machinevertaling door tweede- en derdejaarsstudenten, meer bepaald de perceptie van output in hun moedertaal en in hun tweede taal. De waargenomen kwaliteit werd gemeten met 'confidence scores'. Evaluaties werden uitgevoerd onder twee condities: een directe beoordeling en een beoordeling na een begeleide discussie over de kenmerken van de brontekst en het doel van de doeltaalpublicatie. De resultaten bieden inzichten in hoe toekomstige vertalers zich verhouden tot synthetische taalproducten. In het besluitvormingsproces omtrent gebruik van en omgang met machinevertaling blijken vertrouwdheid met machineoutput en de kritische houding ten overstaan van die vertalingen een doorslaggevende rol te spelen.

Trefwoorden: machinale vertaling; kwaliteitsperceptie; T1-vertaling; T2-vertaling; synthetische tekst

PSYCHOLOGICZNA PERSPEKTYWA NA TŁUMACZENIE MASZYNOWE: PERCEPCJA JAKOŚCI TŁUMACZENIA W NAUCZANIU TRANSLACJI

Abstrakt. W ramach badań nad przekładem opublikowano już wiele prac dotyczących jakości tłumaczenia maszynowego, jednak sposób, w jaki studenci translatoryki postrzegają jakość tłumaczeń

GYS-WALT VAN EGDOM, PhD is docent aan de Universiteit Utrecht, Sectie Vertalen, Interculturele Communicatie en Educatie; correspondentieadres: Trans 10, 3512 JK Utrecht, Nederland; e-mail: g.m.w.vanegdom@uu.nl; ORCID: https://orcid.org/0000-0002-7521-6792.

CHRISTOPHE DECLERCQ, PhD is docent aan de Universiteit Utrecht, Sectie Vertalen, Interculturele Communicatie en Educatie; correspondentieadres: Trans 10, 3512 JK Utrecht, Nederland; e-mail: c.j.m.declercq@uu.nl; ORCID: https://orcid.org/0000-0002-6687-120X. Christophe is ook gastdocent aan de Universiteit Antwerpen en Senior Honorary Research Fellow aan University College London.

maszynowych, otrzymał dotychczas niewielką uwagę. Jest to dość zaskakujące, ponieważ percepcja jakości tłumaczenia maszynowego wpływa zarówno na decyzję o jego użyciu, jak i na sposób pracy z takim tłumaczeniem. Niniejszy artykuł przedstawia wyniki małoskalowego badania dotyczącego percepcji jakości tłumaczeń maszynowych wśród studentów. Analizowano postrzeganą jakość tłumaczeń na język ojczysty i na język obcy w grupie studentów drugiego i trzeciego roku. Wskaźniki postrzeganej jakości mierzono za pomocą tzw. confidence scores (wskaźników pewności). Badanie przeprowadzono w dwóch różnych warunkach: jedna grupa oceniała jakość tłumaczenia natychmiast, podczas gdy druga grupa dokonywała oceny po przeprowadzonej dyskusji na temat cech tekstu źródłowego oraz skuteczności przekładu. Wyniki dostarczają pewnych wskazówek dotyczących stopnia przyzwyczajenia do syntetycznego języka współczesnej automatyzacji tłumaczeń oraz poszerzają perspektywę na rolę technologii językowych w edukacji tłumaczeniowej.

Słowa kluczowe: tłumaczenie maszynowe; postrzegana jakość; tłumaczenie na język ojczysty; tłumaczenie na język obcy; tekst syntetyczny

A PSYCHOLOGICAL PERSPECTIVE ON MACHINE TRANSLATION:
THE PERCEPTION OF TRANSLATION QUALITY IN TRANSLATOR TRAINING

Abstract. Despite decade-long attention to machine translation quality in translation studies, the psychological aspect of the perception of this quality by translation students has been largely overlooked. Yet, the perception of machine translation quality significantly influences the decision to utilize machine translation as well as managing its output as part of the translation process. This article reports on a small-scale study examining second- and third-year students' perceptions of machine translation quality in both their native language and a second language. Perceived quality was measured using "confidence scores." Assessments were conducted under two conditions: immediate evaluation and post-discussion evaluation. The latter concerned the source text's characteristics and the target publication's purpose. The results provide insights into how aspiring translators relate to the synthetic language of machine translation and enrich our perspective on the role of language technology within translator training.

Keywords: machine translation; perceived quality; L1 translation; L2 translation; synthetic text

INLEIDING

Het onderzoek naar de kwaliteit van vertaling kan bogen op een buitengewoon lange traditie (House, 1997; House, 2001; Colina, 2008; Moorkens e.a., 2018). De laatste jaren heeft het echter, zowel binnen de academische wereld als in de professionele wereld, nog een extra boost gekregen. Dit heeft niet alleen te maken met de opkomst van (ver)taaltechnologie en de ontwikkeling van nieuwe raamwerken en nieuwe begrippen om vertaalkwaliteit te duiden en te evalueren, maar ook zeker met de veronderstelde kwaliteitsverbetering van automatische vertaling. Technologisering heeft op vele wijzen een stempel op het discours rond kwaliteit gedrukt.

Vertaalkwaliteit is een belangrijke notie in de vergelijking tussen machinevertaling en wat tegenwoordig steeds vaker ‘menselijke’ vertaling wordt genoemd (zie Moorkens e.a., 2018). De afstand tussen mens en machine wordt in kaart gebracht om de bruikbaarheid van automatische vertaaloplossingen in te schatten. De technologisering heeft er ook toe te hebben geleid dat de kwaliteitsverwachtingen een grotere rol van betekenis zijn gaan spelen. Zo hebben de begrippen ‘good-enough quality’ en ‘fit-for-purpose’ hun intrede gedaan, doordat ook vertalingen van ‘suboptimale’ kwaliteit bruikbaar worden geacht (van Egdom & Segers, 2019). Met de machine vertaalde teksten kunnen immers ‘goed genoeg’ geacht worden. Indien gewenst worden machinevertalingen nog door een post-editor aangepast tot ze voldoende op hun doel zijn toegesneden. Ten slotte heeft de technologisering ook geleid tot een aantal toepassingen – computationele technieken en algoritmes – om vertaalkwaliteit te kwantificeren. De term ‘metrieken’ (*metrics*) is hierdoor in zwang geraakt Vanroy e.a., 2023).

In deze bijdrage wordt er geen poging gedaan om kwaliteit vast te leggen en wordt al evenmin een meetlat aangereikt om de ideale afstand tussen mens en machine te bepalen. Wel wordt er een poging gedaan om de invloed van technologisering op kwaliteitsperceptie op een originele manier in beeld te brengen, door de aandacht te vestigen op de perceptie van synthetische tekst binnen de context van het vertaalonderwijs. ‘Synthetische tekst’ verwijst naar tekst in zogenaamde ‘onnatuurlijke taal’ die door een computertoepassing is gegenereerd. Synthetische tekst lijkt vaak sterk op natuurlijke taal. Deze onnatuurlijke taal wordt geproduceerd met toepassingen zoals spraakherkennings- en tekst-naar-spraaktechnologie, maar ook met automatische tekstgeneratie zoals bij generatieve AI maar ook bij machinevertalingen. De productie van synthetische tekst erg laagdrempelig is geworden: in 2016 waren er elke dag 500 miljoen gebruikers van Google Translate, gebruikers die samen een totaal aan 100 miljard woorden lieten vertalen (Davis, 2024). In september 2024, nog geen twee jaar nadat het werd gelanceerd, had ChatGPT wekelijks 100 miljoen gebruikers (Duarte, 2024). Het is dan ook van groot belang dat de kwaliteit van die content kritisch wordt geëvalueerd (Cabrero-Daniel & Sanagustín Cabrero, 2023). Binnen het vertaalonderwijs is dat belang extra groot, omdat er binnen de vertaaldidactiek steeds steviger wordt gehamerd op de integratie van taaltechnologie binnen het vertaalonderwijs (Massey & Ehrensberger-Dow, 2017; Yuxiu, 2024), terwijl het onduidelijk is wat de effecten van versnelde introductie van taaltechnologie daadwerkelijk zijn.

De hypothese die ten grondslag ligt aan het onderzoek in deze bijdrage luidt: de vorm van blootstelling aan synthetische tekst zoals machinevertaling heeft een invloed op vertaalkwaliteitsperceptie binnen het vertaalonderwijs. De hypothese is ingegeven door de gedachte dat blootstelling aan synthetische tekst voor 'cognitieve priming' kan zorgen bij vertaalopdrachten en dat veelvuldige blootstelling zelfs zou kunnen leiden tot een gewenning aan en normalisering van synthetisch taalgebruik en afname van de menselijke taal(vaardigheid). Door een positievere perceptie van kwaliteit zullen vertalers sneller voor machinevertaling kiezen en zal de concrete omgang met machinevertaling worden beïnvloed. In het onderzoek waarin bovenoemde hypothese is getest, is gekozen voor een dubbele opzet: participanten kregen de opdracht om de kwaliteit van automatische vertaling zowel in hun moedertaal (het Nederlands) als in hun vreemde taal (het Engels) te beoordelen. De metingen zijn uitgevoerd onder twee condities: één groep diende de kwaliteit van de output onmiddellijk te beoordelen, de andere groep beoordeelde de kwaliteit na een begeleide discussie over de eigenschappen van de brontekst.

1. THEORETISCH KADER

Ons onderzoek bevindt zich op het snijvlak van verschillende onderzoekstradities. Om de relevantie van het onderzoek over het voetlicht te kunnen brengen, is ervoor gekozen om kort vier kaders te schetsen. Het eerste is gericht op machinevertaling, het tweede op 'vertaalkwaliteit', het derde luik komt 'taalverandering' als maatschappelijk en wetenschappelijk fenomeen ter sprake en het vierde luik belicht ontwikkelingen binnen de vertaaldidactiek.

1.1. MACHINEVERTALING

Machinevertaling is niet nieuw. Na de Tweede Wereldoorlog boekte deinformatie- en communicatietechnologie grote vooruitgang, maar halverwege de jaren 1960 bleken de mogelijkheden van de spitstechnologie van toen, 'Rule-Based Machine Translation' (RBMT), beperkt. Zelfs nabewerking van machinevertaling (post-editing, PE) was toen nog niet effectief (voor een historisch overzicht, zie Van Egdom & Daems, 2021; Declercq & Van Egdom, 2025; Van Hee & Hoste, 2024). Vanaf de jaren 1980 en 1990 werd gewerkt aan twee nieuwe systemen: Statistical Machine Translation (SMT) en Neural Machine

Translation (NMT). SMT werd als eerste geïntroduceerd: de regelgeleide benadering werd vervangen door een statistische berekeningen uitgevoerd op taaldata. SMT kende een enorme publieke bijval door de succesvolle vermarkting van Google Translate. SMT werd in 2016 overvleugeld door NMT (Koehn, 2020). De opkomst van NMT werd mogelijk gemaakt door de ruime beschikbaarheid van taaldata en groter rekenvermogen, waardoor NMT taaldata kon verwerken op een manier die vergelijkbaar zou moeten zijn met de werking van de menselijke hersenen. Eigenlijk wil dit zoveel zeggen als dat de output niet zozeer het product is van een simpele waarschijnlijkheidsberekening, maar eerder van een belichting van bronmateriaal en doelmateriaal vanuit verschillende invalshoeken, invalshoeken bepaald door programmeringsparameters.

NMT kende gelijk al een groot voordeel: door machinelearningtechnieken kon het neurale systeem leren van nieuwe taaldata, zonder dat er programmering aan te pas hoefde te komen. Waar SMT een onbekende term gewoon onvertaald weergaf, daar deed NMT een 'educated guess' op basis van die multifocale analyse van het nieuwe materiaal (een 'zero-shottechniek') (zie Koehn, 2020). Een recente belangrijke impuls die NMT ontving, werd gegeven in de vorm van aandachtsmechanismen. Om procesvermogen te besparen en accuratere output te genereren kregen neurale systemen aangeleerd welke woorden in een bronzin zwaarder moesten doorwegen in de berekening van parameters.

Door de ontwikkeling van NMT vertegenwoordigt machinevertaling een enorme marktwaarde, mede omdat de kwaliteit ervan in menig onderzoek is aangeprezen (Bentivogli e.a., 2016; Way, 2018; Rivera-Trigueros, 2022). Toch kent NMT onvolkomenheden: neurale systemen zorgen nog altijd voor vrij letterlijke vertalingen en voor teksten met een minder rijk lexicon, beperkte nuance en een normalisering van stijl en andere tekstkenmerken (Dankers e.a., 2022; Freitag e.a., 2021; Vanmassenhove e.a., 2021). Bovendien is de vloeiende output uit NMT vaak misleidend: de systemen geven bronteksten niet altijd even accuraat weer, maar door de overtuigende formulering in de output valt die beperkte accuratesse nauwelijks op – er is sprake van *'false'* of 'fake fluency[1]'. Bovendien zijn er ontwikkelingen gaande om de relatieve getrouwheid aan de brontekst van NMT-modellen te combineren met Generatieve AI, hoewel die laatste technologie niet per se ontworpen is om te vertalen maar eerder om nieuwe inhoud te creëren (Van Egdom & Hartkamp, 2023). De eerste indicaties zijn er echter dat de hybridisering nieuwe problemen zoals

[1] Zie Flores e.a. (2003) voor beschrijving van de term in de oorspronkelijke context; zie ook Dahlkemper e.a. (2023).

hallucinatie meebrengt: met Generatieve AI geproduceerde teksten en vertalingen kunnen ongewenste tekst bevatten (zie ook Ji e.a., 2023). Door de verhoogde mate van vlotheid kunnen deze hallucinaties aan de aandacht van de vertaler, gebruiker of beoordelaar ontsnappen (zie ook Dahlkemper e.a., 2023). Niet alleen blijft kwaliteit van de automatische vertaling hierdoor problematisch, ook de beoordeling van kwaliteit wordt gecompliceerder.

1.2. Vertaalkwaliteit

'Vertaalkwaliteit' staat bekend als een heet hangijzer. In 1972 gaf James Holmes te kennen dat kwaliteitsevaluatie een subjectieve aangelegenheid was en dat een objectieve uitspraak over kwaliteit een onhaalbare kaart voor de vertaalwetenschap is (zie Holmes, 1988). Toch zijn er tal van concepten gemunt en methoden ontwikkeld om grip op kwaliteit te krijgen.

Zo zijn er verschillende methoden voorgesteld om kwaliteit meetbaarder te maken. Drie methoden genieten, hoewel ze in vele gedaanten voorkomen, de meeste bekendheid: de holistische, de analytische en de rubricmethode[2]. Bij holistische methoden vormt de tekst als geheel het uitgangspunt en wordt een uitspraak omtrent kwaliteit vaak gestaafd aan de hand van voorbeelden[3]. Bij analytische methoden wordt er doorgaans gekeken naar de wijze waarop eenheden in tekst worden weergegeven en worden punten in mindering gebracht voor fouten (die aan foutencategorieën zijn gekoppeld). Bij een rubricmethode wordt een gulden middenweg gekozen: de beoordelaar selecteert een rubrics (bijvoorbeeld: de rubrics 'betekenis' en 'stijl') en kent per rubric een globaal cijfer toe die de kwaliteit per aspect moet afdekken (zie bijvoorbeeld Angelelli, 2009; Lommel, 2018). Hoewel geen van deze methoden de sleutel tot objectieve beoordeling vormt, bieden deze verschillende benaderingen een basis voor intersubjectivitieit.

Sinds de ontwikkeling van de eerste machinevertaalsystemen is er al kwistig gestrooid met uitspraken over de kwaliteit van de automatische vertaling. De kwistigheid vloeide voort uit de wens investeringen in technologie te rechtvaardigen. De laatste jaren is er vooral veel bedrijvigheid geweest in de computationele hoek, waar technieken en algoritmes werden ontwikkeld om kwaliteit te kwantificeren, dat wil zeggen: om de kwaliteit van een tekst

[2] Voor een overzicht van evaluatiemethoden, zie Van Egdom e.a. (2018).

[3] Voor een vergelijking van een error-based analysis en een holistische aanpak zie bijvoorbeeld ook Waddington (2001).

cijfermatig uit te drukken (voor een overzicht, zie Vanroy e.a., 2023). Een kwantitatief perspectief op kwaliteit is echter niet vanzelfsprekend. Daarom wordt het nut van computationele technieken en algoritmes vooral afgemeten aan het aantal variabelen dat wordt ondervangen en, in belangrijkere mate, de overeenstemming tussen de kwantitatieve kwaliteitsmeting en het menselijk oordeel. Onder andere bij een 'concours' tussen systemen is het gunstig als de kwaliteitsmaatstaven helder zijn. Hoewel er lange tijd blind werd gevaren op het BLEU-model, dat de vormelijke, sequentiële gelijkenis tussen een machinevertaling en een referentievertaling als maatstaf hanteertis er sinds NMT een sterke nood ontstaan aan modellen die niet alleen afgaan op gelijkenis met een referentievertaling maar die ook aandacht genereren voor contextgebonden variabelen (Kenny, 2018; Vanroy e.a., 2023; Van Egdom e.a., 2023; Stasimioti, 2024; Gulla-Kowalik, 2024).

Ondanks alle bedrijvigheid om kwaliteit van machinevertaling te beoordelen, kan er niet worden beweerd dat kwaliteit echt veel tastbaarder is geworden. Vaak hangt kwaliteitsmeting onbedoeld samen met de perceptie van kwaliteit. Elk oordeel over kwaliteit zegt iets over onze eigen visie op kwaliteit. Bovendien kan worden aangenomen dat perceptie nog wordt beïnvloed door andere factoren die niet zuiver cognitief van aard zijn. Te denken valt aan variabelen als: de context waarin de meting wordt uitgevoerd, het concrete doel van de vertaling, de psychofysiologische staat waarin iemand verkeert, de tijdsdruk waaronder iemand werkt, de hoeveelheid tekst die beoordeeld moet worden, de talenkennis van de beoordelaar en diens vertaalexpertise. Zo blijft vertaalkwaliteit een vlietend begrip.

1.3. Taalverandering

Taal is voortdurend in beweging. Toch worden er initiatieven genomen om de dynamiek op zijn minst af te remmen (zie De Sutter, 2017; Ooms, 2020). Dit effect wordt bijvoorbeeld beoogd met al dan niet centraal georganiseerd taalbeleid (denk hierbij aan handboeken ter bevordering van standaardisering en onderwijspakketten). Toch zien we dat er de laatste tijd meer 'onrust' in de taal is geslopen. Voor een flink deel lijkt deze onrust terug te leiden tot de toename van taalcontact, een toename die mogelijk is gemaakt doordat er nieuwe technologieën zijn ontwikkeld. Technologie heeft niet alleen geleid tot een sterkere vertegenwoordiging van vreemde taal in de eigen cultuur, maar ook tot de vorming van taalvarianten die door subculturen in het leven zijn

geroepen (zie Natsir e.a., 2023). Daarnaast zien we dat de taalbeweging dynamischer wordt doordat initiatieven ter bevordering van de standaardtaal minder effect lijken te sorteren, doordat nieuwe technologieën zeer nadrukkelijk zorgen voor ontlezing – door toegenomen nadruk op beeldtaal – en ontschrijving – onder andere door de opkomst van spraaktechnologie (Schaper e.a., 2019; Stichting Lezen, 2025). Op het snijvlak van taalbeheersing, technologisering en kwaliteit is er dus een weelde aan onderzoeksmogelijkheden.

Een taalverschijnsel dat pas recent bij taalbeheersers in de kijker is gekomen, is de 'synthetische taal'. Door de vooruitgang die de laatste jaren is geboekt binnen de natuurlijketaalverwerking (Natural Language Processing, NLP) worden mensen vaker blootgesteld aan tekst die niet (volledig) door mensen is geproduceerd. NLP is gericht op de technologische bewerking van taal (zie Van Hee & Hoste, 2024). De technologische toepassingen van NLP zijn legio. NLP is lang niet altijd gericht op de productie van nieuwe (bruikbare) tekst, toch zijn de meest zichtbare toepassingen hier wel op gericht. De meest zichtbare toepassing van NLP is vertaling: machinevertaling is al jaren vrij verkrijgbaar en vindt gretig aftrek. Sinds 2022 is er een nieuwe buitengewoon zichtbare toepassing gepresenteerd aan het grote publiek: Generatieve AI (Bahrini e.a., 2023). Typerend aan de teksten die met taaltechnologieën als machinevertaling en Generatieve AI zijn geproduceerd, is dat ze, hoewel ze gericht zijn op de (re)creatie van natuurlijk taal, geheel eigen kenmerken hebben. Er kan worden gesteld dat NLP taalvarianten in het leven roept (Vanmassenhove e.a., 2021). De kenmerken van deze talen staan in de computerlinguïstiek en de vertaalwetenschap beschreven. Toch worden ze vooral 'technisch' benaderd: er is in onderzoek beperkte aandacht voor deze synthetische taalvarianten als epifenomeen van natuurlijke taal. Taalgebruik vloeit immers direct voort uit taalverwerving. Taalgebruikers verwerven taal onder andere door teksten te lezen. Op het moment dat synthetische taal vaker wordt gebruikt in alledaagse on- en offline-communicatie zullen taalgebruikers ook deze taal assimileren en incorporeren in de eigen taal (zie ook Kotze, 2023).

Hoe subtiel synthetische taalvarianten zich een weg in de natuurlijke taal banen, blijkt als we de aandacht vestigen op PE (O'Brien, 2010). Bij PE corrigeert de vertaler (post-editor) een machinevertaling tot de doeltekst voldoet aan de verwachtingen van de klant en de eindgebruiker. Vanuit onderzoek is bekend dat machinevertaling binnen deze context een sturende rol vervult: doordat de vertaler vertrekt vanuit de machinevertaling treedt er cognitieve 'priming' op, waardoor de vertaler sneller zal kiezen voor een vertaling die in de meeste gevallen niet te zeer afwijkt van de machinesuggestie (Bangalore

e.a., 2016; Declercq & Van Egdom, 2023). Dat betekent dat de aanwezigheid van machinevertaling tijdens het hele PE-proces een effect heeft op het taalgebruik van de vertaler. Maar daarmee is het laatste woord over het effect van synthetische taal op taalgebruik niet gezegd. Om ervoor te zorgen dat PE efficiënt gebeurt, geldt als regel dat de vertaler nooit mag 'over-editen' (Hu & Cadwell, 2016; Daems & Macken, 2019). Dat betekent dat de richtlijnen van PE, ondanks een onderscheid tussen *light PE* en *full PE* (O'Brien, 2010), zonder uitzondering aansturen op een maximaal behoud van de synthetische kenmerken van de doeltekst. Zelfs als er dus taalkundig een keus is om af te wijken is er een economische noodzaak om de synthetische tekst in de mate van het mogelijke te behouden. De echte impact op het vlak van taalgebruik wordt pas ervaren als ook de omvang van deze onderneming goed in kaart wordt gebracht. PE is als onderdeel van het vertaalproces en zelfs als vertaaldienst duidelijk aan een opmars bezig (zie *European Language Industry Survey*, 2025). Dit betekent dat vertalers steeds vaker in een positie worden geplaatst waarin ze zoveel mogelijk synthetische tekst dienen te behouden. Die toenemende blootstelling aan synthetische tekst leidt onvermijdelijk tot gewenning en geleidelijke assimilatie in de eigen taal[4].

1.4. Vertaaldidactiek

De laatste jaren zijn er binnen de vertaaldidactiek veel ontwikkelingen geweest (Hurtado Albir, 2007; Valero-Garcés & Corrochano, 2018). De belangrijkste ontwikkeling is misschien wel de introductie van vertaalcompetentiemodellen. Deze modellen zijn in eerste instantie gericht op het bepalen van de expertise van professionele vertalers, maar ze hebben vooral hun nut bewezen binnen het vertaalonderwijs, omdat ze de afstemming tussen onderwijs en praktijk mogelijk maken. Door scherp te stellen op de (kern)vaardigheden van vertalers kunnen we het vertaalonderwijs zo inrichten dat alle vereiste vaardigheden tijdens de studie aan bod komen. Competentiemodellen dragen dus bij aan de professionalisering van het vertaalonderwijs en vergroten de slaagkansen van alumni om een carrière te kunnen uitbouwen op de vertaalmarkt. De laatste jaren heeft technologie, zoals gezegd, een flinke stempel op de vertaalmarkt gedrukt. Hierdoor is er de afgelopen jaren ook een nood ontstaan

[4] Een volgend onderzoek kan dit proberen koppelen aan taalverarming en -vervlakking.

aan updates van competentiemodellen, waarbinnen taaltechnologie een prominentere rol krijgt toebedeeld. Die updates hebben hun weerslag op vertaalopleidingen niet gemist. Steeds luider weerklinkt de roep om een snelle en vroege integratie van taaltechnologie in het vertaalonderwijs (Kenny, 2019).

Toch is er weinig bekend over het effect van vroege en intensieve blootstelling van taaltechnologie binnen het vertaalonderwijs. Het lijdt weinig twijfel dat technologieonderwijs kan bijdragen aan de technologische vaardigheden die het vertalerschap vereist. Toch blijft het de vraag of technologie-integratie, zoals Mossop (2000) al opwierp, geen sta-in-de-weg zal vormen voor 'traditionele' vaardigheden. In het licht van de vorige paragraaf lijkt het niet ondenkbaar dat het vermogen van vertalers om alternatieve oplossingen in de doeltaal te bedenken en die doelgericht in de vertaling te verwerken verstramt door de alomtegenwoordigheid van automatische oplossingen.

2. METHODE

In het voorjaar van 2024 is er aan de Universiteit Utrecht een onderzoek uitgevoerd naar de invloed van technologisering op de perceptie van vertaalkwaliteit. Daarbij kregen 2 groepen studenten de vraag om de kwaliteit van segmenten die met de machine waren vertaald te beoordelen. Daarbij is er gebruikgemaakt van segmenten die in de moedertaal van de studenten, het Nederlands, waren vertaald en segmenten die in hun vreemde taal, het Engels, waren vertaald. Hieronder wordt de opzet toegelicht.

2.1. DEELNEMERS

De participanten waren 28 studenten die waren ingeschreven voor het bachelorvak Vertalen Engels-Nederlands, een keuzevak binnen de opleiding Engelse Taal en Cultuur. Het onderzoek was ingebed in een werkcollege van het vak. De participanten hadden vergelijkbare ervaring op het gebied van vertaling: ze beschouwden zichzelf als tweetalig of nagenoeg tweetalig, met noties van vertaling. Op een 10 puntsschaal schatten ze hun moedertaalbeheersing (7,4) opmerkelijk genoeg iets lager in dan hun taalbeheersing Engels

(7,6[5]). In een vertaalwaardehiërarchie plaatsten ze stilistische getrouwheid aan de brontekst (7,9) boven vlotheid (7,6) en correctheid (7,4). De leeftijd van de deelnemers lag tussen de 20 en 30 jaar. Hierdoor mag bij beide groepen een min of meer vergelijkbaar taalniveau en een vergelijkbare woordenschat worden verondersteld. De participanten werden voor het onderzoek opgedeeld in 2 groepen. Groep 1 telde 15 deelnemers, groep 2 13 deelnemers. .

2.2. MATERIAAL

Voor het college hadden de docenten 2 bronteksten geselecteerd. Het ging om een brontekst in het Engels en een brontekst in het Nederlands. Bij de selectie werd er gemikt op literaire teksten met een relatief lage moeilijkheidsgraad. Uiteindelijk werd er gekozen voor de volgende teksten: 'The Great Hug' [1975] van Donald Barthelme (1175 woorden) en 'Poep' [1993] van Manon Uphoff (2323 woorden). Het bronmateriaal werd met behulp van de vertaalmachine DeepL vertaald (snapshot april 2024), in het Nederlands (eerstetaalvertaling (hier: T1)) en in het Engels (tweedetaalvertaling (hier: T2)). Vervolgens werden er door de docenten 10 machinevertaalde segmenten per doeltaal aangewezen. Het ging om een willekeurige selectie van segmenten (zie annex 1 en 2). Wel werd er aandacht gevestigd op segmenten die, hoewel ze zelden flagrante fouten bevatten, zonder uitzondering voor verbetering vatbaar waren. Ook werd uitgegaan van variatie in zinslengte. Om die verbeteringsruimte te duiden hebben de beoordelaars gebruikgemaakt van een eigen modelvertaling, aangevuld met de bestaande vertalingen. De selectie van 10 segmenten is in beide gevallen naderhand nog aangevuld met 2 segmenten uit de bestaande vertalingen die vervaardigd waren door professioneel literair vertalers (zie annex 1 en 2).

De enquête werd aangevuld met een lijstje met zes vragen die betrekking hadden op taalbeheersing (T1 en T2), literair inzicht en prioritering van correctheid, vlotheid en stijl in vertaling (zie het kopje 'Deelnemers'). Daarnaast bevatte de vragenlijst twee commentaarruimtes waarin studenten konden aangeven wat ze van de vertalingen vonden.

[5] Een van de redenen toegekend aan de ontlezing van het Nederlands als moedertaal is de opvallende opkomst van niet-vertaalde Engelse literatuur en het *global English* van posts op sociale media.

2.3 CONDITIES

De participanten moesten de kwaliteit van de twee sets segmenten beoordelen. Beide groepen kregen het bronmateriaal voor de les aangereikt en moesten de teksten grondig lezen. De 2 groepen voerden de beoordeling vervolgens uit onder verschillende omstandigheden. Groep 1 kreeg aan het begin van de les de doelsegmenten voorgelegd. Pas na afloop van de beoordeling gingen zij in gesprek over het bronmateriaal. Bij groep 2 werd de beoordeling van de segmenten voorafgegaan door een begeleide groepsdiscussie die erop gericht was plot en stijl van de literaire teksten te doorgronden en een algemeen beeld te vormen van de punten waar vertalers op moeten letten bij de vertaling van de bronteksten. Opvallend was dat studenten uit zichzelf vertaalproblemen klassikaal bespraken en geneigd waren om mogelijke oplossingen aan te dragen. Pas aan het einde van de les kregen ze het verzoek om de kwaliteit van de segmenten die met de machine waren vertaald te beoordelen. De volgorde T1-T2 werd in beide gevallen aangehouden.

2.4. ANALYSE

De participanten moesten de kwaliteit van de twee sets segmenten beoordelen. De beoordelingen werden verzameld met behulp van Qualtrics. In de tool kregen de studenten de sets met segmenten (met bronmateriaal) te zien en daarbij kregen ze de vraag om een zogenaamde ‘confidence score’ toe te kennen. De score wordt vaak toegekend in de context van machinevertaling en vormt de neerslag van het vertrouwen van een respondent in de kwaliteit van machineoutput (zie Specia, ; Specia e.a., 2010). Het cijfer ligt altijd tussen de 1 en de 5, waarbij 1 duidt op een volledig gebrek aan vertrouwen in de machineoutput en 5 duidt op een groot vertrouwen in de output. De scores per conditie en per doeltaal werden met behulp van beschrijvende statistiek vergeleken. De aanname was dat zowel de blootstelling aan het door de machine vertaalde materiaal als de mate van vertrouwdheid met de doeltaal (T1 versus T2) een effect zouden hebben op de waardering voor machineoutput.

3. RESULTATEN

Hieronder worden de resultaten van het onderzoek voorgesteld. Daarbij wordt eerst gekeken naar de perceptie van kwaliteit bij een verschillende instructie en vervolgens naar de perceptie van kwaliteit bij de doeltalen.

3.1. INSTRUCTIE

Een blik op tabellen 1 en 2 leert ons dat er verschillen zijn op te merken tussen de waardering van groep 1 en de waardering van groep 2. Zoals verwacht, is groep 1 veel meer te spreken over de kwaliteit van de vertaling. Zo waardeert groep 1 de DeepL-vertaling van 'The Great Hug' gemiddeld met een 3,38, waar groep 2 een magerdere score van 2,96 toekent (zie Tabel 1), en kent groep 1 de DeepL-vertaling van 'Poop' gemiddeld een 3,49 toe, terwijl de waardering van groep 2 blijft hangen op een 3 (zie Tabel 2). Opvallend is dat nagenoeg alle segmenten die met de machine zijn vertaald een positievere confidence score krijgen van groep 1. De enige uitzondering wordt gevormd door 'Het staat in… van heilige dieren' (6), een segment dat door beide groepen gemiddeld met een 3,3 wordt beoordeeld. Dat wil dus zeggen dat zowel de Nederlandse als de Engelse doeltekst door groep 1 positiever is beoordeeld. Daarbij moet de kanttekening worden geplaatst dat de verschillen in waardering niet enorm uitgesproken lijken.

Tabel 1: Confidence scores 'The Great Hug'

	Groep 1:			***Groep 2:***		
Segment	**Gemiddelde**	**Mediaan**	**Afwijking**	**Gemiddelde**	**Mediaan**	**Afwijking**
1. Ze kon niet … het ook probeerde.	3,5	3,0	1,0	2,8	3,0	0,8
2. "Dit is het… ben," zei hij."	2,9	3,0	1,5	1,8	2,0	1,2
3. En de kinderen.. hebben het geld.	4,1	4,0	0,9	3,7	4,0	1,0
4. Het zit allemaal...ondoordachte juiste beweging.	3,1	3,0	1,2	2,9	3,0	1,0
5. Sommige mensen gaan...geen ballon kopen.	3,5	4,0	1,4	3,1	3,0	1,2
6. Het staat in... van heilige dieren.	3,3	4,0	1,3	3,3	3,0	1,3
7. De omhelzing tussen...om te zien.	2,5	3,0	1,5	2,4	3,0	1,1
8. Ik wil er niet aan denken.	3,9	4,0	1,2	3,3	4,0	1,7
9. Toen hij onze...we ons beter.	3,8	4,0	0,9	3,5	3,0	1,2
10. Sommige mensen vinden... het leuk vinden.	3,2	3,5	1,2	2,8	3,0	1,2
DeepL (gemiddelde)	***3,38***			***2,96***		
11. De Ballonnenman wil… op de foto.	2,5	3	1,0	2,8	3,0	1,4
12. Holle balonnen verkoop… in hun sok.	3,4	3,5	0,9	3,5	4,0	1,4
Onno Kosters (gemiddelde)	***2,95***			***3,15***		

Richten we de aandacht op losse items, dan blijkt dat nuancering van de resultaten steeds op zijn plaats is. Binnen de selectie segmenten uit de vertaling van ‘The Great Hug’ krijgt het segment ‘En de kinderen… hebben het geld’ van beide groepen de hoogste score toegekend: een 4,1 van groep 1 en een 3,7 van groep 2. Binnen de selectie segmenten uit de vertaling van ‘Poep’ krijgt het segment ‘Once upon a… much to spend’ van groep 1 een 4 en van groep 2 een 3,9 toegekend. Deze waarderingen staan in contrast met items als ‘De omhelzing van… om te zien’ en ‘Dit is het… goed in ben’. Die twee items worden minder goed gesmaakt (respectievelijk een 2,5 en een 2,9 bij groep 1 en een 2,4 en een 1,8 bij groep 2). In de vertaling van ‘Poep’ is het segment ‘He threw them… and unaffected again’ zeer problematisch volgens beide groepen, met scores van respectievelijk 2,8 bij groep 1 en 1,8 bij groep 2. Het is hierbij verleidelijk om te kijken naar de concrete vertalingen: flagrante interferentie (de letterlijke weergave van idioom) lijkt bij de slechtst gewaardeerde items de grondslag van uitgesproken negatieve oordelen te vormen. In de commentaarruimte wordt er ook door beide groepen opgemerkt dat de vertaling ‘soms te letterlijk is’ en dat ‘het […] misloopt in veel zinnen’. Ter illustratie van die letterlijkheid: de Engelse vertaling van ‘Poep’ bevat ‘He trew them [the pebbles] as flat as possible over the water…’ als vertaling van ‘Die wierp hij zo vlak mogelijk over het water…’. Positief waren studenten volgens de docenten vooral als het betekenisverschil genuanceerder was: vergelijk bijvoorbeeld ‘we hebben het geld’ (DeepL-vertaling) met ‘we hebben er het geld voor’ (standaardvertaling) of ‘we hebben genoeg geld (op zak)’ (modelvertaling). Zo geeft een student uit groep 1 aan: ‘naar mijn mening was deze vertaling opzich [sic!] best wel goed.’ Er is uitgebreider onderzoek nodig om meer diepgang te kunnen verlenen aan de interpretatie van oordelen van de studenten.

Tabel 2: Confidence scores 'Poep'

	Groep 1:			*Groep 2:*		
Segment	**Gemiddelde**	**Mediaan**	**Afwijking**	**Gemiddelde**	**Mediaan**	**Afwijking**
1. Once upon a… much to spend	4,0	4,0	0,9	3,9	4,0	0,9
2. Having reached the.. pick up pebbles.	3,6	4,0	1,3	3,2	3,0	0,9
3. He threw them...and unaffected again.	2,8	3,0	1,1	1,8	2,0	1,0
4. Such a house,...money to spend.	2,8	3,0	1,7	2,6	2,0	1,2
5. Just at that...dogs came out.	3,5	3,0	1,2	3,2	3,0	1,1
6. It's a very... to pay for it.	3,9	4,0	1,2	3,3	3,0	1,1
7. As he said… up the bank.	3,7	4,0	1,1	3,4	3,0	0,8
8. His stomach, esophagus… it was over.	3,5	4,0	1,3	3,3	3,0	1,0
9. Is that acceptable to you?	3,8	4,0	1,1	3,7	4,0	0,9
10. Off!" called the woman… high and shrill.)	3,3	3,0	1,1	3,2	2,5	1,2
Gemiddelden DeepL	***3,49***			***3***		
11. Her mother-of-pearl nail… piles of dog poop.	3,5	3,5	0,9	3,4	4,0	1,1
12. If you eat… along with it.	3,5	3	1,5	3,5	4,0	1,0
Gemiddelden Sam Garrett	***3,5***			**3,45**		

Naast het verschil tussen positief en negatief beoordeelde items, is er ook het verschil in waardering tussen de groepen. Soms zitten de groepen op een lijn. Bij 6 van de 10 items uit de vertaling van 'The Great Hug' is het verschil in oordeel kleiner dan 0,5. Het verschil in oordeel is het grootst bij de items: 'Dit is het… ben," zei hij' (1,1 punt) en 'Ze kon het… het ook probeerde' (0,7 punt). Deze twee zinnen kwamen overigens ook tijdens de begeleide discussie ter sprake. Hetzelfde patroon valt te ontwaren als we kijken naar de waardering voor de segmenten uit de vertaling van 'Poep': bij 6 van de 10 items is het verschil in oordeel kleiner dan 0,5. Grote verschillen worden opgemerkt tussen: 'He threw them…and unaffected again' (1 punt) en 'Off, called the… high and shrill' (0,7 punt). Bij de interpretatie van de verschillen in waardering is het belangrijk op te merken dat idioom, toon en register als grootste aandachtspunten werden aangemerkt tijdens de begeleide discussie van groep 2. Als voorbeeld kan hierbij het verschil in idioom worden aangehaald in de weergave van 'I don't want to think about it', waar 'Ik wil er niet aan denken' (DeepL-vertaling) sterk contrasteert met 'Ik moet er niet aan denken' (standaardvertaling). Kenmerkend voor deze waardering is ook de reactie van een student uit groep 2: 'Prima [vertaling], maar soms te letterlijk en niet idiomatisch genoeg.'

Ten slotte verschaft tabel 1 ook nog inzicht in de waardering voor de standaardvertaling. Bij 'The Great Hug' zien we dat de waardering voor de twee items sterk uiteenloopt (bij elke groep is het verschil groter dan 0,5 punt). Toch is er bij beide items een effect op te merken in de waardering van de

twee groepen. De segmenten uit de standaardvertaling ontvangen een iets positievere waardering van groep 2: bij ‘De Ballonnenmeneer wil niet op de foto’ valt de score 0,3 punt hoger uit, bij ‘Holle ballonnen verkoop… in hun sok’ is er een verschil van 0,1. Kenmerkend voor het negatievere oordeel van groep 1 is de reactie van een student die aangeeft dat de vertaling bij het segment ‘wil niet op de foto’ ‘teveel [sic!] vanuit het nederlands bekeken’ is. Betreffende student komt met de verbeteringssuggestie: “weigert [op de foto te gaan].” Bij ‘Poep’ blijven de verschillen beperkt: bij een segment (‘Her mother-of-pearl nail… of dog poop’) is er zelfs geen verschil in waardering (3,5), terwijl er bij ‘If you eat…. along with it’ een minimaal verschil optreedt (0,1). Hoewel in dit geval ook de verleiding om harde uitspraken te doen moet worden weerstaan, heeft het er zeker de schijn van dat de waardering voor creatievere oplossingen, de oplossingen die we opmerken in de ‘menselijke’ vertalingen, toeneemt als studenten meer inzicht in de brontekst hebben verworven en een betere kijk hebben op de specifieke aard van vertaalproblemen en het doel van vertaling.

3.2. DOELTALEN

Er is ook gekeken naar het effect van de doeltaal op de waardering van machinevertaling. Uit Tabel 3 valt af te leiden dat de hypothese tot op zekere hoogte geschraagd wordt: studenten laten zich positiever uit over de kwaliteit van de machinevertalingen in de vreemde taal. Hoewel de letterlijkheid in de commentaarruimte door beide groepen wordt gethematiseerd, is er meer lof voor de vertaling: ‘De vertaling is hier een stuk beter’ (student groep 1) en ‘Goede vertaling, lopende zinnen’ (groep 2). Ook valt het op dat er minder vaak wordt teruggegrepen naar concrete voorbeelden. Alleen de flagrante interpretatiefout “off” (als variant van “af”, gericht tot een hond) wordt aangehaald. Toch moet worden toegegeven dat het effect niet enorm groot is: de T1 worden gemiddeld met een 3,19 beoordeeld en de T2 worden gemiddeld met een 3,27 beoordeeld. Ondanks kleine verschillen kan er worden gesteld dat er met deze kleine beperkte data een zeer bescheiden trend te ontwaren valt. De taalrichting lijkt de perceptie van kwaliteit van machinevertaling in enige mate te beïnvloeden.

CONCLUSIE

De resultaten van dit onderzoek lijken erop te wijzen dat de vorm van blootstelling aan machinevertaling een zekere invloed heeft op de wijze waarop machinevertaling door vertaalstudenten wordt gepercipieerd en gewaardeerd. Studenten die niet gelijk werden blootgesteld aan een voorgekauwde oplossing en die op voorhand in discussie waren getreden over de eigenschappen van de bronteksten en de mogelijke problemen bij het vertalen van de teksten, verhielden zich iets kritischer tot de machinevertaling. Studenten die gelijk werden blootgesteld, lieten zich positiever uit over de kwaliteit van de machinevertaling. Door dit verschil in perceptie zou het ook niet ondenkbaar zijn dat vroege blootstelling aan machinevertaling in de vertaalopleiding de waarschijnlijkheid verkleint dat er in sterke mate zal wordt afgeweken van machinesuggesties. Dit idee wordt ondersteund door de bevinding dat de confidence scores in het onderzoek, ondanks duidelijke gebreken in de geselecteerde segmenten, opvallend hoog zijn uitgevallen.

Naast de vorm van blootstelling aan machinevertaling lijken de taalrichting en de taalbeheersing overigens mogelijk ook een rol te spelen in de perceptie van en waardering voor machinevertaling. Vooral bij de groep studenten die geen analyse van de bronteksten had uitgevoerd, was het vertrouwen in machinevertalingen in de vreemde taal tamelijk groot. Deze resultaten zouden er wellicht op kunnen wijzen dat participanten een minder kritische omgang met vertaling en machinevertaling in hun tweede taal hebben. Toch moet er bij deze observaties steeds een kritische kanttekening worden geplaatst: hoewel er steeds een bescheiden trend is opgemerkt, kunnen er geen uitspraken worden gedaan over de statistische significantie van verschillen.

Dit brengt ons bij de beperkingen van ons onderzoek. Het onderzoek is ten uitvoer gebracht binnen een bachelorvak. Dat wil zeggen dat de ervaring die studenten met ‘ouderwets’ vertalen hebben eerder beperkt is en dat zij daardoor relatief gezien misschien minder inzicht hebben in de aanpassingen die nodig zijn om de doelteksten te verbeteren. Daarnaast is de steekproef ook beperkt gebleven. Aan het onderzoek hebben 28 studenten deelgenomen die slechts twee (korte) teksten hebben beoordeeld. Uiteraard is het wenselijk om de steekproef te vergroten om statistische bewerking mogelijk te maken en de significantie van verschillen te onderzoeken: beoordeling van meerdere teksten door meer deelnemers zou de generaliseerbaarheid van de uitkomsten alleen maar ten goede komen. Daarnaast zou dit onderzoek baat kunnen hebben bij verfijning van het instrumentarium: zo zou er kunnen worden nagegaan

welk soort fouten vooral een effect hebben op de kwaliteitsperceptie van participanten en zou een categorisering van aangetroffen fouten alsmede een evenwichtige spreiding van fouten in de itemset een meerwaarde hebben voor dit onderzoek. Verfijndere meetinstrumenten zouden ook meer diepgang kunnen verlenen aan verschillen tussen perceptie van kwaliteit bij T1- en T2-vertaling.

Hoe dan ook lijken de resultaten van dit onderzoek ons te sterken in de overtuiging dat de integratie van taaltechnologie niet te lichtvaardig mag gebeuren. Er is meer onderzoek nodig, onder andere vorsing naar onnatuurlijke-taalgewenning door vroege en intensieve blootstelling aan taaltechnologie, om een beeld te vormen van de effecten van taaltechnologie op cognitieve processen bij vertaalstudenten en op de mogelijke weerslag van die effecten op competentieverwerving binnen het vertaalonderwijs.

BIBLIOGRAFIE

Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelelli & H. E. Jakobson (Eds.), *Testing and assessment in translation and interpreting studies* (pp. 13–47). John Benjamins Publishing Company.

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, opportunities, and threats. *2023 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 274–279). https://arxiv.org/pdf/2304.09103

Bangalore, S., Behrens, B., Michael, C., Ghankot, M., Heilmann, A., Nitzke, J., Schaeffer, M., & Sturm, A. (2016). Syntactic Variance and Priming Effects in Translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New directions in empirical translation process research*, New Frontiers in Translation Studies (pp. 211–238). Springer.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural *versus* phrase-based machine translation quality: A case study. In J. Su, K. Duh, & X. Carreras (Eds.), *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 257–267). Association for Computational Linguistics. https://aclanthology.org/D16-1025.pdf

Cabrero-Daniel, B. & Sanagustín Cabrero, A. (2023). Perceived trustworthiness of natural language generators. In K. Devlin & J. Fischer, *TAS '23: Proceedings of the First International Symposium on Trustworthy Autonomous Systems* (pp. 1–9). Association for Computing Machinery. https://doi.org/10.1145/3597512.3599715

Colina, S. (2008). Translation quality evaluation: Empirical evidence for a functionalist approach. *The Translator*, *14*(1), 97–134. https://doi.org/10.1080/13556509.2008.10799251

Daems, J. & Macken, L. (2019). Van CAT tot TenT: van computerondersteund vertalen tot vertaalomgevingen. In G. De Sutter & I. Delaere (Eds.),. *In balans: een inleiding tot vertaal- en tolkwetenschap* (pp. 262–283). Acco.

Dahlkemper, M. N., Lahme, S. Z., & Klein, P. (2023). How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT. *Physical Review Physics Education Research*, *19*(1), 010142. https://doi.org/10.1103/PhysRevPhysEducRes.19.010142

Dankers, V., Bruni, E., & Hupkes, D. (2022). The paradox of the compositionality of natural language: A neural machine translation case study. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Vol. 1. Long papers* (pp. 4154–4175). Association for Computational Linguistics. https://aclanthology.org/2022.acl-long.286.pdf

Duarte, F. (2024). *Google Translate – friend or foe?*. LanguageWire. https://www.languagewire.com/en/blog/google-translate

De Sutter, G. (Ed.). (2017). *De vele gezichten van het Nederlands in Vlaanderen. Een inleiding tot de variatietaalkunde*. Acco.

Declercq, C. & van Egdom, G.-W. (2023). No more buying cats in a bag? Literary translation in the age of language automation. *Revista Tradumàtica. Tecnologies de la Traducció*, (21), 49–62.

Declercq, C. & van Egdom, G.-W. (2025). *Beyond the strictest computation of the general proportion: Language automation, machine translation and their complicated possibilities for cultural exchange*. In D. Nguyen, B. Mutsvairo, & J. Zeng (Eds.), *Technology, power and society* (pp. 273–299). Brill.

Davis, D. (2024). *Number of ChatGPT Users.* Exploding Topics. Geraadpleegd op 21 oktober 2024 van https://explodingtopics.com/blog/chatgpt-users

European Language Industry Survey 2025. Trends, expectations and concerns of the European language industry (2025). https://elis-survey.org/wp-content/uploads/2025/03/ELIS-2025_Report. pdf

Flores, G., Laws, M. B., Mayo, S. J., Zuckerman, B., Abreu, M., Medina, L., & Hardt, E. J. (2003). Errors in medical interpretation and their potential clinical consequences in pediatric encounters. *Pediatrics*, *111*(1), 6–14. https://doi.org/ 10.1542/peds.111.1.6

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, *9*, 1460–1474. https://doi.org/10.1162/tacl_a_00437

Gulla-Kowalik, O. (2024, 17 april). *New ISO Standard 5060 focuses on human evaluation to ensure translation quality.* Slator. https://slator.com/new-iso-standard-5060-focuses-on-human-evaluation-to-ensure-translation-quality/

Holmes, J. S. (1988). *Translated! Papers on literary translation and translation studies*. Rodopi.

House, J. (1997). *Translation quality assessment: A model revisited*. Gunter Narr Verlag.

House, J. (2001). Translation quality assessment: Linguistic description versus social evaluation. *Meta: Journal des traducteurs / Translators' Journal*, *46*(2), 243–257. https://doi.org/10.7202/003141ar

Hu, K. & Cadwell, P. (2016). A comparative study of post-editing guidelines. *Baltic Journal of Modern Computing*, *4*, 346–353.

Hurtado Albir, A. (2007). Competence-based curriculum design for training translators. *The Interpreter and Translator Trainer*, *1*(2), 163–195. https://doi.org/10.1080/1750399X.2007.10798757

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), article 248. https://doi.org/10.1145/3571730

Kenny, D. (2018). Sustaining disruption? The transition from statistical to neural machine translation. *Revista Tradumàtica. Tecnologies de la Traducció*, (16), 59–70. https://doi.org/10.5565/rev/tradumatica.221

Kenny, D. (2019). Technology and translator training. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 498–515). Routledge.

Koehn, P. (2020). *Neural Machine Translation*. Cambridge University Press.

Kotze, H. (2023, 1-3 februari). *Beyond patterns: How corpus-linguistic comparisons of human and machine translation can help us understand how language works* [presentatie]. Convergence Conference: Human-machine integration in translation and interpreting, Guildford, United Kingdom.

Lommel, A. (2018). Metrics for translation quality assessment: A case for standardising error typologies. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 109–127). Springer.

Massey, G. & Ehrensberger-Dow, M. (2017). Machine learning: Implications for translator education. *Lebende Sprachen*, *62*(2), 300–312. https://doi.org/10.1515/les-2017-0021

Moorkens, J., Castilho, S., Gaspari, F., & Doherty, S. (Eds.). (2018). *Translation quality assessment: From principles to practice*. Springer.

Mossop, B. (2003). What should be taught at translation school? In A. Pym, C. Fallada, J. R. Biau, & J. Orenstein (Eds.), *Innovation and e-learning in translator training* (pp. 20-22). Universitat Rovira i Virgili. http://www.intercultural.urv.cat/media/upload/domain_317/arxius/Innovation/innovation_index.pdf

Natsir, N., Aliah, N., Zulkhaeriyah, Amiruddin, & Esmianti, F. (2023). The impact of language changes caused by technology and social media. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, *7*(1), 115–124. https://jurnal.uisu.ac.id/index.php/languageliteracy. https://doi.org/10.30743/ll.v7i1.7021

O'Brien, S. (2010). Introduction to post-editing: Who, what, how and where to next? In *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas: Tutorials*, Denver, USA. Association for Machine Translation in the Americas. https://aclanthology.org/2010.amta-tutorials.1.pdf

Ooms, M. (2020). *Buurtaal: praktische gids voor het Nederlands in België en Nederland*. Sterck & De Vreese.

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Language Resources and Evaluation*, *56*(2), 593–619. https://doi.org/10.1007/s10579-021-09537-5

Schaper, J., Wennekers, A., & de Haan, J. (2019). *Trends in Media: Tijd*. Sociaal en Cultureel Planbureau.

Specia, L. (2011). Exploiting objective annotations for measuring translation post-editing effort. In M. L. Forcada, H. Depraetere, & V. Vandeghinste (Eds.), *Proceedings of the 15th International Conference of the European Association for Machine Translation* (pp. 73–80). Leuven.

Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. *Machine Translation*, *24*(1), 39–50. https://doi.org/10.1007/s10590-010-9077-2

Stasimioti, M. (2024, 2 april). *Unbabel presents a new evaluation metric for chat translation*. Slator. https://slator.com/unbabel-presents-new-evaluation-metric-for-chat-translation/

Stichting Lezen – Leesmonitor (2025). *Een ander perspectief op ontlezing*. https://www.lezen.nl/onderzoek/een-ander-perspectief-op-ontlezing/

Valero-Garcés, C. & Corrochano, C. C. (2018). Approaches to didactics for technologies in translation and interpreting. *Special Issue of trans-kom*, *11*(2), 154–161. https://www.trans-kom.eu/bd11nr02/trans-kom_11_02_01_Valero_Cedillo_Introduction.20181220.pdf

Van Egdom, G.-W. & Daems, J. (2021, 24 januari). Ontwikkelingen rond literair vertalen en technologie:een inleiding. *Filter. Tijdschrift over vertalen*. https://www.tijdschrift-filter.nl/webfilter/dossier/literair-vertalen-en-technologie/januari-2021/ontwikkelingen-rond-literair-vertalen-en-technologie-een-inleiding/

Van Egdom, G.-W. & Hartkamp, E. (2023). *Generatieve AI en machinevertaling. Praktische handleiding voor het onderwijs*. https://netherlands.representation.ec.europa.eu/system/files/2024-02/Lesmodule%20Generatieve%20AI%20en%20MT.NL_.pdf

Van Egdom, G.-W. & Segers, W. (2019). *Leren vertalen: een terminologie van de vertaaldidactiek*. Pelckmans Pro.

Van Egdom, G.-W., Kosters, O., & Declercq, C. (2023). The riddle of (literary) machine translation quality: Assessing automated quality evaluation metrics in a literary context. *Revista Tradumàtica: Tecnologies de la Traducció*, (21), 129–159. https://doi.org/10.5565/rev/tradumatica.345

Van Egdom, G.-W., Verplaetse, H., Schrijver, I., Kockaert, H. J., Segers, W., Pauwels, J., Wylin, B., & Bloemen, H. (2018). How to put the translation test to the test? On preselected items evaluation and perturbation. In E. Huertas-Barros, S. Vandepitte, & E. Iglesias-Fernández (Eds.), *Quality assurance and assessment practices in translation and interpreting* (pp. 26–56). IGI Global.

Van Hee, C. & Hoste, V. (2024). *Taaltechnologie ontrafeld. Hoe taal en technologie hand in hand gaan*. Pelckmans.

Vanmassenhove, E, Shterionov, D., & Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in Machine Translation. In P. Merlo, J. Tiedermann, & R. Tsarfaty (Eds.), *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume* (pp. 2203–2213). Association for Computational Linguistics.

Vanroy, B., Tezcan, A., & Macken, L. (2023). MATEO: Machine Translation Evaluation Online. In M. Nurminen, J. Brenner, M. Koponen, S. Latomaa, M. Mikhailov, F. Schierl, T. Ranasinghe, E. Vanmassenhove, S. Alvarez-Vidal, N. Aranberri, M. Nunziatini, C. Parra Escartín, M. Forcada, M. Popovic, C. Scarton, & H. Moniz (Eds.), *Proceedings of the 24th Annual Conference of The European Association for Machine Translation* (pp. 499–500). European Association for Machine Translation.

Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta: Journal des traducteurs / Translators' Journal*, *46*(2), 311–325. https://doi.org/10.7202/004583ar

Way, A. (2018). Quality expectations of machine translation. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 159–178). Springer.

Yuxiu, Y. (2024). Application of translation technology based on AI in translation teaching. *Systems and Soft Computing*, *6*, article 200072.

ANNEX 1

Brontekstsegment	***Doeltekstsegment***
1. She couldn't cry, she tried.	1. Ze kon niet huilen, hoe ze het ook probeerde.
2. "This is the kind of thing I do so well," he said.	2. "Dit is het soort dingen waar ik zo goed in ben," zei hij.
3. And the kids will say, "C'mon, Balloon Man, give us a red balloon and a green balloon and a white balloon, we got the money."	3. En de kinderen zullen zeggen, "Kom op, Ballonnenman, geef ons een rode ballon en een groene ballon en een witte ballon, we hebben het geld."
4. It's all in the gesture - the precise, reunpremeditated right move.	4. Het zit allemaal in het gebaar - de precieze, onvoorbedachte juiste beweging.
5. Some people don't go to the circus and so don't meet the Balloon Man and don't get to buy a balloon.	5. Sommige mensen gaan niet naar het circus en ontmoeten dus de Ballonnenman niet en krijgen geen ballon te kopen.
6. It's in the cards, in the stars, in the entrails of sacred animals.	6. Het staat in de kaarten, in de sterren, in de ingewanden van heilige dieren..
7. The embrace of Balloon Man and Pin Lady will be something to see.	7. De omhelzing van Ballonnenman en Pinnendame zal iets zijn om te zien.
8. I don't want to think about it.	8. Ik wil er niet aan denken.
9. When he created our butter-colored balloon, we felt better.	9. Toen hij onze boterkleurige ballon creëerde, voelden we ons beter.
10. Some people won't like it. Some people will like it.	10. Sommige mensen vinden het niet leuk. Sommige mensen zullen het leuk vinden.
11. The Balloon Man won't let you take his picture.	11. De Ballonnenman wil niet op de foto.
12. "I don't sell the Balloon Jejune," the Balloon Man will say, "let them other people sell it, let them other people have all that wet and nasty kid-money mitosising in their sock.	12. ‘ Holle ballonnen verkoop ik niet’, zegt de Ballonnenmeneer, ‘die moeten anderen maar verkopen, kunnen die rondlopen met dat natte onsmakelijke voortschimmelende kindergeld in hun sok. (Onno Kosters)

ANNEX 2

Brontekstsegment	**Doeltekstsegment**
1. Er was eens een aardige lange man die heel arm was en nooit veel te besteden had.	1. Once upon a time, there was a nice tall man who was very poor and never had much to spend.
2. Helemaal aan het eind gekomen - daar waar de singel een scherpe bocht maakte en in zichzelf terugkeerde - op de plek waar een aantal met mossig groen overdekte oude bomen stonden, hield hij stil om zich te bukken en steentjes te rapen.	2. Having reached the very end - where the canal made a sharp bend and returned into itself - at the place where a number of old trees covered with mossy green stood, he stopped to stoop and pick up pebbles.
3. Die wierp hij zo vlak mogelijk over het water, keek ze na tot ze onder water verdwenen en het wateroppervlak weer strak en onaangedaan achterlieten.	3. He threw them as flat as possible over the water, watched them until they disappeared under the water and left the water surface tight and unaffected again.
4. Zo'n huis, dacht de man die maar zo weinig geld te besteden had.	4. Such a house, thought the man who had so little money to spend.
5. Juist op dat moment ging de deur van het schitterende huis open en kwam er een dame met twee grote honden naar buiten.	5. Just at that moment, the door of the splendid house opened and a lady with two large dogs came out.
6. Het is een heel duur huis, maar ik heb het er graag voor over.	6. It's a very expensive house, but I'm willing to pay for it.
7. Terwijl hij dit zei sprongen de Deense doggen met een jolig enthousiasme het water van de singel in, om er een paar minuten later uitgelaten blaffend weer uit te kruipen, de kant op.	7. As he said this, the Great Danes jumped with jovial enthusiasm into the water of the canal, to emerge a few minutes later exuberantly barking, climbing up the bank.
8. Zijn maag, slokdarm en darmstelsel hadden het heft in handen en namen eenzelfde beslissing. Afgelopen was het.	8. His stomach, esophagus, and intestines had taken over and made the same decision. It was over.
9. Is dat acceptabel voor u?	9. Is that acceptable to you?
10. ‘Af!’ riep de vrouw. Haar stem was hoog en schel geworden.	10. "Off!" called the woman. Her voice had become high and shrill.
11. Haar parelmoerkleurige nagel wees vooruit en stuurde zijn blik naar wat ze bedoelde: de twee diepbruine, glanzende hopen hondepoep.	11. Her mother of pearl nail directed his gaze towards the two gleaming, dark brown piles of dog poop. (Sam Garrett)
12. Als u die twee hopen opeet, krijgt u mijn huis. Het huis van uw dromen, met de tuin en alles erbij.	12. If you eat both of those, I will give you my house The house of your dreams, with the garden and everything along with it. (Sam Garrett)