Translation technobabble: An exploration of online discourse about machine translation and artificial intelligence using corpus-driven discourse analysis and appraisal theory

Gabriëlla Arntz, 1814710

2 February 2024

MA Thesis Professioneel Vertalen

Utrecht University

Supervisor: Haidee Kotze

15050 words

Universiteit Utrecht

Faculteit Geesteswetenschappen

Versie september 2014

VERKLARING KENNISNEMING REGELS M.B.T. PLAGIAAT

Fraude en plagiaat
Wetenschappelijke integriteit vormt de basis van het academisch bedrijf. De Universiteit Utrecht vat iedere vorm van wetenschappelijke misleiding daarom op als een zeer ernstig vergrijp. De Universiteit Utrecht verwacht dat elke student de normen en waarden inzake wetenschappelijke integriteit kent en in acht neemt.

De belangrijkste vormen van misleiding die deze integriteit aantasten zijn fraude en plagiaat. Plagiaat is het overnemen van andermans werk zonder behoorlijke verwijzing en is een vorm van fraude. Hieronder volgt nadere uitleg wat er onder fraude en plagiaat wordt verstaan en een aantal concrete voorbeelden daarvan. Let wel: dit is geen uitputtende lijst!

Bij constatering van fraude of plagiaat kan de examencommissie van de opleiding sancties opleggen. De sterkste sanctie die de examencommissie kan opleggen is het indienen van een verzoek aan het College van Bestuur om een student van de opleiding te laten verwijderen.

Plagiaat
Plagiaat is het overnemen van stukken, gedachten, redeneringen van anderen en deze laten doorgaan voor eigen werk. Je moet altijd nauwkeurig aangeven aan wie ideeën en inzichten zijn ontleend, en voortdurend bedacht zijn op het verschil tussen citeren, parafraseren en plagiëren. Niet alleen bij het gebruik van gedrukte bronnen, maar zeker ook bij het gebruik van informatie die van het internet wordt gehaald, dien je zorgvuldig te werk te gaan bij het vermelden van de informatiebronnen.

De volgende zaken worden in elk geval als plagiaat aangemerkt:

- het knippen en plakken van tekst van digitale bronnen zoals encyclopedieën of digitale tijdschriften zonder aanhalingstekens en verwijzing;
- het knippen en plakken van teksten van het internet zonder aanhalingstekens en verwijzing;
- het overnemen van gedrukt materiaal zoals boeken, tijdschriften of encyclopedieën zonder aanhalingstekens en verwijzing;
- het opnemen van een vertaling van bovengenoemde teksten zonder aanhalingstekens en verwijzing;
- het parafraseren van bovengenoemde teksten zonder (deugdelijke) verwijzing: parafrasen moeten als zodanig gemarkeerd zijn (door de tekst uitdrukkelijk te verbinden met de oorspronkelijke auteur in tekst of noot), zodat niet de indruk wordt gewekt dat het gaat om eigen gedachtengoed van de student;
- het overnemen van beeld-, geluids- of testmateriaal van anderen zonder verwijzing en zodoende laten doorgaan voor eigen werk;
- het zonder bronvermelding opnieuw inleveren van eerder door de student gemaakt eigen werk en dit laten doorgaan voor in het kader van de cursus vervaardigd oorspronkelijk werk, tenzij dit in de cursus of door de docent uitdrukkelijk is toegestaan;
- het overnemen van werk van andere studenten en dit laten doorgaan voor eigen werk. Indien dit gebeurt met toestemming van de andere student is de laatste medeplichtig aan plagiaat;
- ook wanneer in een gezamenlijk werkstuk door een van de auteurs plagiaat wordt gepleegd, zijn de andere auteurs medeplichtig aan plagiaat, indien zij hadden kunnen of moeten weten dat de ander plagiaat pleegde;
- het indienen van werkstukken die verworven zijn van een commerciële instelling (zoals een internetsite met uittreksels of papers) of die al dan niet tegen betaling door iemand anders zijn geschreven.

De plagiaatregels gelden ook voor concepten van papers of (hoofdstukken van) scripties die voor feedback aan een docent worden toegezonden, voorzover de mogelijkheid voor het insturen van concepten en het krijgen van feedback in de cursushandleiding of scriptieregeling is vermeld.

Universiteit Utrecht

In de Onderwijs- en Examenregeling (artikel 5.15) is vastgelegd wat de formele gang van zaken is als er een vermoeden van fraude/plagiaat is, en welke sancties er opgelegd kunnen worden.

Onwetendheid is geen excuus. Je bent verantwoordelijk voor je eigen gedrag. De Universiteit Utrecht gaat ervan uit dat je weet wat fraude en plagiaat zijn. Van haar kant zorgt de Universiteit Utrecht ervoor dat je zo vroeg mogelijk in je opleiding de principes van wetenschapsbeoefening bijgebracht krijgt en op de hoogte wordt gebracht van wat de instelling als fraude en plagiaat beschouwt, zodat je weet aan welke normen je je moeten houden.

Hierbij verklaar ik bovenstaande tekst gelezen en begrepen te hebben.
Naam: Gabriëlla Lisette Arntz Studentnummer: 1814710
Datum en handtekening: 02/02/2024

Dit formulier lever je bij je begeleider in als je start met je bacheloreindwerkstuk of je master scriptie.

Het niet indienen of ondertekenen van het formulier betekent overigens niet dat er geen sancties kunnen worden genomen als blijkt dat er sprake is van plagiaat in het werkstuk.

Abstract

This research explores the online discourse surrounding translation technology, specifically machine translation and artificial intelligence. This is done by comparing and contrasting different stakeholders, their attitudes and the linguistic resources they use to express these attitudes. This is accomplished by analysing discourse produced by the public, by language service providers, and by language software development companies using the methodologies of corpus-driven discourse analysis and appraisal theory. The corpus-driven discourse analysis shows partial overlap in the themes that the various stakeholders discuss within the discourse, although some themes unique to each stakeholder also emerge. Further investigation with discourse analysis and appraisal theory reveals that these overlapping themes are framed differently by the different stakeholders, through different associations and how these are expressed. In particular, the capabilities of machine translation and artificial intelligence are discussed by all stakeholders and especially by the software companies, who focus on incorporating these translation technologies as part of an overall business strategy. Language service providers, in contrast, focus on the role of humans as essential to the translation process and quality of the final product. The public focuses on the larger moral debate, taking into account potential consequences of the use of machine translation and artificial intelligence. The data from this thesis shows a general gradation of attitudes towards translation technology: from software companies as the most optimistic to the public as the most pessimistic.

Contents

1. Introduction

1.1 Motivation for this research

Research on machine translation has flourished since its inception, and the popularisation of each new iteration within industry and media leads to heightened interest in research as well. Informal observations of the current discourse in media and on social media seems to suggest a lively but also rather polarised debate surrounding artificial intelligence (AI) and specifically large language models (LLMs) like ChatGPT. The aim of this research is to investigate these informal observations further, and to determine in a systematic way whether the debate is indeed as polarised as it appears. In particular, this research aims to determine if and how the professional role of stakeholders especially influence their opinions and the linguistic resources they use to express these.

1.2 Research approach

This research describes the current online discourse surrounding translation technology through a comparison of different stakeholders, their attitudes, and the linguistic resources they use to express these attitudes. This is accomplished by focusing on evaluative language in web texts, specifically in blogs and newspaper articles. It concerns exploratory research where the main aim is to sketch an overview with general characteristics of the discourse. Additionally, attention is paid to the particulars of the data collection and analysis processes. To accomplish all this, firstly, previous research is discussed in chapter 2 to provide a context and initial ideas for data collection, including possible stakeholders and expectations for their attitudes. Data in the form of blogs and newspaper articles was collected according to the procedures described in section 3.2. This data was then analysed using two approaches. Corpus-driven discourse analysis (Baker, 2006), making use of frequency, keyness and concordance analysis in Sketch Engine (Kilgarriff et al., 2014), reveals the main themes for each of the stakeholders and associations with these themes. Subsequently, the framework appraisal theory (Martin & White, 2005) was applied to analyse the evaluative language. For this, texts were selected based on themes overlapping between the stakeholders and further analysed. This enables additional description of the attitudes expressed by the stakeholders and the linguistic evaluative resources used to express them. Through these analyses a general overview of the discourse can be derived in terms of the similarities and differences between the various stakeholders, the main themes they discuss, the associations attached to these themes and the evaluative resources used to express them.

1.3 Relevance

1.3.1 Scientific relevance

When considering the research done on attitudes towards translation technology, particularly machine translation, most research seems to concern the further technological development of the technology rather than paying much attention to its impact (Ragni & Nunes Vieira, 2022). This especially seems to be the case outside the field of translation studies (as will also be explored in section 2.3.3), but is still true for much research within translation studies as well. The main focus is often either the quality of the machine translation itself or the efficiency of post-editing, with perceptions of stakeholders hardly taken into account (Ragni & Nunes Vieira, 2022, 142, 145; Sakamoto & Yamada, 2020, 78). This would already make additional research on perception of translation technology, and particularly machine translation, a worthwhile endeavour.

The existing research on perceptions of machine translation within the field of translation studies mainly (and somewhat understandably) focuses on one stakeholder: the translator. (This is also quite clear when comparing the length and content of the different sections within section 2.3.) Research considering the attitudes of multiple stakeholders and possible connections between them could thus be especially interesting. I agree with Ragni & Nunes Vieira (2022), who argue that "…to understand the effect of [neural machine translation] on the industry, personal, perception-related data should be collected on a larger scale, and from the full spectrum of stakeholders involved, including textual consumers and project managers" (149). This research is a contribution to that goal.

1.3.2 Societal relevance

Aside from the general possibility to contribute to research in translation studies, this research also intends to capture an especially relevant and interesting moment in the online discourse surrounding translation technology spurred on by the newest hype surrounding the implementation of AI in many fields. Within the field of translation, an increasing emphasis on productivity and automation was already felt (Ragni & Nunes Vieira, 2022, 137). The main route for this automation within translation seems to be the use of machine translation and post-editing. However, an increased and increasingly accepted use of machine translation and other translation technology does not mean there is agreement on best practice (Sakamoto & Yamada, 2020, 79). I would argue that Sakamoto & Yamada (2020) are correct in describing the discourse surrounding machine translation as still very much in flux, especially with this latest iteration of new technological possibilities through AI. People and their opinions construct the discourse, but these people and their opinions are equally influenced in turn (Sakamoto & Yamada, 2020, 91). Mapping out the discourse by describing the various stakeholders and their attitudes gives insight into these possible influences and, additionally, possible power dynamics that might come into play as the discourse stabilises (Sakamoto & Yamada, 2020, 91). These insights might even be considered crucial when assessing the sustainability of the translation sector as a whole, which for example Moorkens & Rocchi (2020) argue is currently under threat due to the influence of (certain uses of) machine translation. They signal that the combination of continued growth and continued withdrawal by translators could become a serious issue and seems to be related to machine translation and perceptions of it (Moorkens & Rocchi, 2020, 5, 13).

1.4 Overview

The remainder of the thesis is structured as follows. Chapter 2 contains the literature review, where previous research is discussed, along with the theoretical grounding and concepts behind the methodology used for data analysis. It concludes with the formulation of research questions and some expected findings. Chapter 3 contains a description of the methodologies, starting with a description of how the data was collected and the characteristics of the resulting corpus and subcorpora. It then turns to a description of how the concepts of corpus-driven discourse analysis and appraisal theory are applied practically to the data. Chapter 4 discusses the results. It contains a description of the results of both methodologies for all stakeholders and a first comparison of the similarities and differences in themes, associations and evaluative resources. Chapter 5 comprises the discussion, where all data is combined to provide a more in-depth comparison is made between the various stakeholders and their attitudes and how they express them. This is connected and compared to previous literature, including some speculation about possible motivations behind the attitudes and the framing of them. This chapter concludes with a discussion of some of the limitations of the research. Chapter 6 contains a short conclusion, with a summary of the results and some suggestions for future research.

2. Literature review

2.1 Introduction

In this chapter previous research on the topic of perceptions of machine translation and related theoretical constructs are discussed. As a starting point to discuss previous research on machine translation within the field of translation, the meta-analysis of Ragni & Nunes Vieira (2022) is used. This provides a first insight into the current state of research on attitudes towards translation technology. Following this, potential stakeholders within the discourse are identified based on the research of Moorkens & Rocchi (2020) and Sakamoto & Yamada (2020). For each of these stakeholders an overview of their attitudes, as characterised through research, is given. Subsequent to this, the specific approaches to data analysis are discussed. Baker (2006) forms the basis of the discussion about the concepts and approach of corpus-driven discourse analysis. Martin & White (2005) is used as the basis for appraisal theory. The practical application of these approaches is discussed in chapter 3. To conclude the chapter, specific research questions and expected findings are formulated on the basis of the previously discussed research and theory.

2.2 Recent research into human factors in machine translation

The meta-analysis of Ragni & Nunes Vieira (2022) has already been briefly referenced in chapter 1. It provides a good and, to my knowledge, the most up-to-date overview of the current state of research into the human aspects of machine translation. To accomplish this, Ragni & Nunes Vieira (2022) specifically selected studies that included both the use of neural machine translation and a human element. This means the research was either addressing the implications of neural machine translation for humans, there was direct human participation in the study or there was a (somewhat) detailed description of human participation in a prior stage of the research (Ragni & Nunes Vieira, 2022, 138-139). They conclude that only about a quarter of the research involved perception and that most studies were concerned with quality evaluation (Ragni & Nunes Vieira, 2022, 142, 145). Additionally, end-users were not often involved (Ragni & Nunes Vieira, 2022, 146). When translators or linguists participated in the research, this was usually in the role of evaluator rather than end-user, meaning their contribution mainly benefited another stakeholder instead of themselves (Ragni & Nunes Vieira, 2022, 146-147). Ragni & Nunes Vieira (2022) conclude from this that in the research they analysed a commercial perspective is employed, which focuses on the development of machine translation with the aim to save time and money rather than an interest in potential consequences of the technology or its integration into existing translation workflows (151). This meta-analysis provides a starting point for further exploration and an overall impression of attitudes towards machine translation, one with mixed opinions and a certain asymmetry in the positions and treatment of stakeholders.

2.3 Stakeholders

To build upon the first impression provided by Ragni & Nunes Vieira (2022), the various stakeholders discussed in previous literature should be identified more precisely and their opinions should be examined in more detail. For this research, of particular interest is the professional role of a stakeholder and its influence on their attitude. This is also the approach quite often taken in other research: defining the stakeholders by their professional role. Moorkens & Rocchi (2020) thus approach the question of stakeholders by taking the language service provider (LSP) as a central point and then identifying its internal and external stakeholders. The internal stakeholders would be the owners, project managers and in-house translators. The external stakeholders are the language software developers, freelance translators, clients, end-users and society (Moorkens & Rocchi, 2020, 8). Sakamoto & Yamada (2020) take a somewhat similar approach, by taking project managers as their central point, but limiting their further discussion to clients, the management of LSPs and (freelance) translators. Based on this previous literature, the initial categories of stakeholders investigated are: LSPs (as one complete unit), translators (both freelance and in-house), language software developers, and the "public" (considering clients and end-users together).

2.3.1 Attitudes of language service providers

Sakamoto (2019) describes a polarisation of approaches evident in LSPs, with some LSPs offering post-editing services and others commenting negatively on it (207). Whether such services are offered may be influenced by the location and size of the language service provider and the profiles of their already employed staff (Sakamoto & Yamada, 2020, 87). This split in approaches is also mirrored in the attitudes expressed by the project managers, who often represent the LSPs both in business and research. Sakamoto & Yamada (2020) discuss a divide between enthusiasm and scepticism and link this to the amount of contact the project managers have had with machine translation (84). Nunes Vieira & Alonso (2020) report that project managers have relatively balanced views, especially when compared to translators, which may be due to their professional position allowing them a broader overview of translation business and activities (174, 176). In general, the project managers seem mainly concerned with balancing the various relationships they hold and streamlining the workflow of the project as a whole (Nunes Vieira & Alonso, 2020, 171). This would suggest that the attitudes these project managers express are indeed clearly connected with their professional role, which allows them more of an overview but also forces them to balance the expectations of various parties involved. So, Sakamoto (2019) finds project managers to have an overall negative view of post-editing specifically, mainly due to concerns about the reaction of translators (206). On the other hand, LSPs might fear losing clients if they do not meet the service and price demands of clients who only seem interested in lowering costs through technology, which might result in offering the services regardless and a race-to-the-bottom pricing strategy to stay profitable (Moorkens & Rocchi, 2020, 3-4; Sakamoto & Yamada, 2020, 86). The current main issue for project managers and LSPs then comes down to the unpredictability of machine translation as it exists today, since this connects closely to such issues as communication and expectation management, especially on the client side (Nunes Vieira & Alonso, 2020, 172-173). Overall, research suggests that project managers have generally split but also mild attitudes towards machine translation, guided by their need (and struggle) to balance the attitudes and expectations of other stakeholders.

2.3.2 Attitudes of translators

As mentioned in chapter 1, translators are clearly the most researched group when it comes to attitudes towards machine translation and translation technology in general. This is especially evident when professional translators, both in-house and freelance, are considered in combination with translation students. The overall conclusion than can be drawn from the previous literature seems to be that translators lean more towards the negative side and mixed views are often reported. In an early paper on this topic, Fulford (2002) finds that a translator’s view depends on their level of exposure to machine translation (similar to Rossi & Chevrot (2019, 14)) and that scepticism is often not due to perceiving machine translation as a threat but rather due to doubt whether the machine is capable of such a complex task as translation (120). Although the quality of machine translation has certainly improved over time, this scepticism is still partly present since translators are generally much less willing to compromise on translation quality compared to clients and managers (Nunes Vieira & Alonso, 2020, 175).

Aside from the quality of the machine translation itself, many papers discuss related issues. Läubli & Orrego-Carmona (2017) argue that translators not feeling included in the development of machine translation is an important factor in their scepticism (61). Sakamoto (2021) adds that for translators this exclusion from development can create the feeling of furthering the competition between human and machine translation and the feeling that translators are only helping along their replacement, rather than truly benefitting from further developments (250). Nunes Vieira & Alonso (2020) conclude current uses of machine translation mainly exacerbate existing issues (172-173). This is reflected in translators’ concerns surrounding the rates and organisation of post-editing, especially the imposition of such work, issues which translators all link to (unrealistic) expectations (e.g. Guerberof Arenas, 2013, 93; Moorkens & Rocchi, 2020, 12; Vidal et al., 2020, 54-55; Vieira, 2020, 13-14).

The common thread linking these issues can thus be identified as agency or a sense of agency, a topic only briefly discussed here but quite often investigated and discussed within translation studies in general. Cadwell et al. (2018) indeed find the sense of and actual level of agency to be important for translators' perception of machine translation (317). This is echoed by both Moorkens & Rocchi (2020), who find a negative perception of machine translation by translators to be due to imposition of automation, digital dispossession and predictions of unemployment (21), and Nunes Vieira & Alonso (2020), who find translators are distanced from business aspects as well as planning and estimation stages of translation projects, which might leave them frustrated with clients' misguided assumptions in later stages (173, 177). Such circumstances can leave translators feeling unheard in industrial workflows in general, not just in relation to the use of machine translation, although they may have little recourse to push back due to power disparities (Moorkens & Rocchi, 2020, 24; Vidal et al., 2020, 64). Indeed, research on the attitudes of translators towards machine translation reports mostly mixed feelings and overall, a feeling of anxiety. However, this seems to be mostly caused by surrounding factors that affect the (sense of) agency of translators as professionals and exacerbate existing issues in industrial workflows.

2.3.3 Attitudes of language software developers

As already discussed, Ragni & Nunes Vieira (2022) conclude that there is relatively little research into perception, signalling a potential rift between language software developers and others involved with translation technology. For instance, Kenny (2011) argues that developers of machine translation often downplay the source of their data and the role of translators in the creation of this data (7) – a statement that seems to ring true for developers of generative AI as well and which is publicly being investigated and debated at the time of writing (e.g. Zirpoli (2023)). When considering recent research, this observation of what could be characterised as a lack of attention for other stakeholders seems to hold. Most research does indeed seem to focus on either the quality or application of machine translation. Some examples include research by Gao et al. (2023) on using ChatGPT to enhance machine translation compared to general commercial engines, Guo (2022) on using deep learning to optimise machine translation, Pham et al. (2023) on using AI to augment data for machine translation of low resource languages, and Zhang (2023) on the development of reference free machine translation evaluation. Especially these last two examples are interesting, because their explicit aim is to turn machine translation into a fully autonomous system, no longer reliant on external (human) data. Recent examples of direct reflection on the relationship between human and machine translation do exist as well. One example is Ai (2022), which reflects on the current problems of machine translation in the age of AI. The conclusion drawn is that, due to existing problems, machine translation and human translation should coexist. These examples illustrate that, at least in the broader academic research context outside of translation studies, machine translation is (understandably) treated as a goal in its own right, rather than a complement to or tool for translators. A tentative conclusion is thus that from the perspective of language software developers, the main aim is indeed to develop an (autonomous) machine translation system that is able to fully replace (parts of) human translation. This is also currently, based on the number of available papers on the topic, considered a worthwhile and realistic pursuit.

2.3.4 Attitudes of the public

Ragni & Nunes Vieira (2022) conclude that very little research involves end-users directly. Some recent research can be found, investigating the use of machine translation in healthcare, a traditionally high-risk translation area (Zappatore & Ruggieri, 2024), in education, specifically language learning (Cotelli Kureth et al., 2023; Yang et al., 2023), and in what could be designated "personal" applications, specifically travel (Carvalho et al., 2023). These examples seem to suggest that currently the main interest in this area of research is to explore the boundaries of the usability of machine translation for end-users. In terms of attitude, most research concludes that, overall, end-users seem to find the quality of machine translation acceptable enough (Castilho & O'Brien, 2017, 134). They are at least less critical than translators (Kasperė et al., 2023, 13), although this does correlate with the level of education of the users (Kasperė et al., 2021, 15). Thus, the "professional role" of the end-users as public does indeed seem to affect their attitudes, but rather through a sense of absence. Depending on their circumstance (such as level of education), these users might not be aware of what possible alternatives to machine translations might look like. In this sense, users are not just affected in a binary sense, of whether the machine translation was good or not, rather, this unawareness of and possible lack of access to alternatives could come with additional risk attached (Guerberof-Arenas & Moorkens, 2023; Kasperė et al., 2023, 13).

For translation clients, Sakamoto & Yamada (2020) again describe a division between enthusiasm and scepticism (85). From an industry perspective, the perception is that demand for machine translation is indeed very much client-driven (Sakamoto & Yamada, 2020, 85; Sakamoto, 2021, 245). General trends in client attitude as signalled by Sakamoto & Yamada (2020) and Sakamoto (2021) are a greater tolerance for sub-optimal machine translation output, increased price pressure and pressure for faster delivery, which might either manifest itself as clients expecting fast delivery regardless of quality or the usual level of quality with a machine translation discount (85; 244, 248). At least from an industry perspective then, clients really do first and foremost take a business-oriented approach to improve their own bottom line (Moorkens & Rocchi, 2020; Ragni & Nunes Vieira, 2022), very much acting in the interest of their professional role. Simultaneously, there might be possible frustrations and risks for end-users that remain invisible to these users, due to their particular position in the web of stakeholders.

2.3.5 Initial conclusions from previous research

Previous research thus suggests that the professional role of stakeholders does indeed influence their attitude, mainly in respect of what aspects are emphasised under influence of the overall (professional) goals. In terms of their methodologies, these attitudes were mainly investigated directly, using experiments, interviews or focus groups to examine the attitudes of one or, only sometimes, multiple stakeholders. To my knowledge, a comparison of unsolicited data on the attitudes of multiple stakeholders towards machine translation is non-existent or at the very least extremely rare.

Based on this overview, it can also be argued that the current state of research itself might play a not insignificant role in the perceptions of machine translation by creating a kind of feedback loop. Within existing research there is quite a strong focus on quality and productivity due to the commercial perspective often taken (as mentioned by Ragni & Nunes Vieira, 2022). This large amount of commercially focused research can in turn serve as a legitimisation of the idea that good machine translation will make workflows more cost-effective and this type research is thus a worthy pursuit (as mentioned by Sakamoto & Yamada, 2020). As Moorkens & Rocchi (2020) state: “[w]ithout an obvious return of investment, the arguments to tailor tools or workflows for translators tend not to filter through the production network” (12). Especially when taking into account the volume (and main direction) of current research done by language software developers, this then seems to promote the notion that machine translation and human translation are mutually exclusive, as Vieira (2020) observes (3).

This previous literature forms the first part of the basis used to formulate specific research questions and expected findings. In the next section, the theoretical framework for data analysis is discussed, forming the second part of the basis. This theoretical framework can then be used for guiding the specific shape of the research questions and the operationalisation of these questions through the methodology discussed in chapter 3.

2.4 Theoretical approaches

Before the two analytical approaches used in this study can be discussed in more detail, it is important to understand their foundational connections and disparities. Both corpus-driven discourse analysis and appraisal theory are intimately connected to the concepts of discourse and discourse analysis (or discourse studies). Corpus-driven discourse analysis relates to these by very directly connecting methodologies and concepts of corpus linguistics and discourse analysis. Appraisal theory relates to these by casting itself as part of discourse semantics, where the linguistic resources that can be described with the framework of appraisal theory contribute to the creation of interpersonal meaning at the discourse-semantic level. (Martin & White, 2005, 10, 33). Discourse in the context of this thesis can then be understood to refer to the collection of specific sets of meanings about a particular topic (in this case machine translation and artificial intelligence) constructed by particular groups (in this case the various stakeholders) through the use of particular forms (Richardson & Flowerdew, 2017, 2). This discourse is underpinned by the ideologies of the particular groups, which are actually recreated through the discourse (Richardson & Flowerdew, 2017, 2, 3). This in turn means that the ideologies of these groups, in this particular context their attitudes towards translation technology, can be reconstructed and described through analysis of the discourse. The main disparity between the two analytical frameworks is in the approach to the analysis: which forms of expression are analysed and how.

Corpus-driven discourse analysis, in general, does not involve a set methodology as neither corpus linguistics nor discourse analysis involves a set methodology (Taylor & Marchi, 2018, 2). Rather, this approach attempts to combine the strengths of both frameworks, by providing a more neutral starting point for analysis through corpus linguistics without ignoring the context so crucial to discourse analysis (Taylor & Marchi, 2018, 4). This approach does not lead to some elusive "objectivity" (Taylor & Marchi, 2018, 7). However, as Baker (2006) describes, the corpus data itself forms the starting point of the analysis rather than that the corpus merely supporting a prior hypothesis with data, which is especially crucial for less obvious, even unconscious discursive patterns (16, 175). This approach means that the text as essential unit of analysis for discourse analysis is not considered as such, but rather that the concept of patterns as essential to identifying ideologies as well as power dynamics is highlighted (Richardson & Flowerdew, 2017, 1). The use of a corpus is especially appropriate for this since it can reveal both conscious and unconscious meanings associated with a word or word cluster by presenting many instances of use at once (Hunston & Thompson, 2000, 16-18). In this study, the main lexical themes and their associations thus make up the patterns that allow for the (re)constructing of the attitudes of the various stakeholders towards machine translation.

Appraisal theory uses the lens of evaluative language to describe the discourse. Hunston & Thompson (2000) describe how evaluation in a text serves to reflect the values of a person or community, to foster a relationship with the reader and to organise the discourse within the text (6). The evaluative elements are thus the linguistic resources (or particular forms) that are used to express attitudes or ideologies through the discourse. These evaluative resources can be found on various levels: in lexis, grammar and the text as a whole (Hunston & Thompson, 2000, 13-14). By labelling these resources with appraisal theory, it will be possible to describe the forms in which the underlying attitudes are expressed in detail, along with a description of the content of the labelled elements.

2.4.1 Corpus-driven discourse analysis

Baker (2006) gives a concise but complete overview of how corpus-driven discourse analysis can be approached and is taken as the main guide for discourse analysis in this thesis. The analysis in this thesis is performed using the analytical tools of frequency, keyness and concordance. Additionally, collocation can give a measure of "nearness" between words so that the most typical meanings and associations of a specific word can be understood (Baker, 2006, 96). However, in this study, exploratory collocation analysis did not seem to provide much additional information, so further analysis was not pursued, due to time constraints.

Frequency, as an analytical tool, can be utilised by simply generating a list of the most commonly used words in a corpus or subcorpus. Baker (2006) argues that since word use is not random (or, is ideologically motivated in discourse-analytical terms), frequency can give a first indication of potentially interesting aspects within a corpus (47, 68). Following Baker et al. (2013), collections of these words can also be grouped together into themes. To further identify potentially interesting themes, frequency can then be combined with the notion of keyness, which produces another word list based on the saliency of words and word clusters by comparing two data sets (Baker, 2006, 125). This works especially well when comparing multiple data sets, as in this thesis, because through a comparison of the whole corpus with a reference corpus similarities in themes between various stakeholders can be discovered, and through a comparison of subcorpora of the discourse of each stakeholder with the whole corpus differences in themes among the stakeholder groups can be discovered (Baker, 2006, 138, 146). Frequency and keyness together thus give an overview of the main themes of the corpus and some first indication of the overall positive or negative slant associated with these. To investigate specific associations further, concordance analysis can then be used. The concordance shows all occurrences of a specific word (cluster), within the selected corpus or subcorpus, in the form of lines where the word is embedded in its (limited) context (Baker, 2006, 72). Through concordance analysis, the specific context of and associations with themes can then be examined. It can not only give a sense of which attitudes are expressed (through semantic nearness) but also whether these are frequent (Baker, 2006, 86-89). This allows for a more qualitative interpretation of the negative or positive attitude associated with both the specific theme words and their surrounding associations.

2.4.2 Appraisal theory

Martin & White (2005) give a thorough description of their appraisal theory, which can be used to identify the linguistic resources that are used to encode evaluation in a (naturalistic) text (25, 31). It is then possible to label these resources to reconstruct the underlying attitude of a text and describe the ways in which it is expressed. The framework for labelling these resources is constituted by the concepts of *attitude*, *engagement* and *graduation*. An overview of this is provided in figure 1. In this thesis, mainly due to time and space constraints (although style considerations are also discussed in section 4.2.1), only the dimension of *attitude* will be considered in detail.

The concept of *engagement* describes the interpersonal position of the author of a text, especially in relation to the reader (Martin & White, 2005, 93). The concept of *graduation* describes the up- or down-scaling of an opinion (Martin & White, 2005, 135). *Attitude* is the concept that describes feelings, which can be general (internal) feelings (categorised under *affect*) as well as feelings towards behaviour (categorised under *judgement*) or feelings towards phenomena (categorised under *appreciation*) (Martin & White, 2005, 42-44). Each of these categories is further subdivided into specific "feelings" (e.g. *happiness*, *normality*, *valuation*, etc.), each with both a positive and negative variant, which can then be used as labels for stretches of text. The specific definitions and usage of these labels in this study is discussed in more detail in section 3.5.2. Through this labelling, the associations of the various stakeholders with the themes found through corpus-driven discourse analysis, and especially how these are expressed, can be examined in detail. In addition, it is also possible to compare *attitude* (and *engagement*) across a range of texts and identify the common evaluative resources between them. Martin & White (2005) refer to this kind of generalisation as "key", a coherent register where specific from the total of possibilities of the system of appraisal have been selected and are used together (163-164). From this it would be possible to describe an evaluative style, or key, for each stakeholder, describing not only their attitude but also how they use specific evaluative resources to express this attitude (Martin & White, 2005, 166).

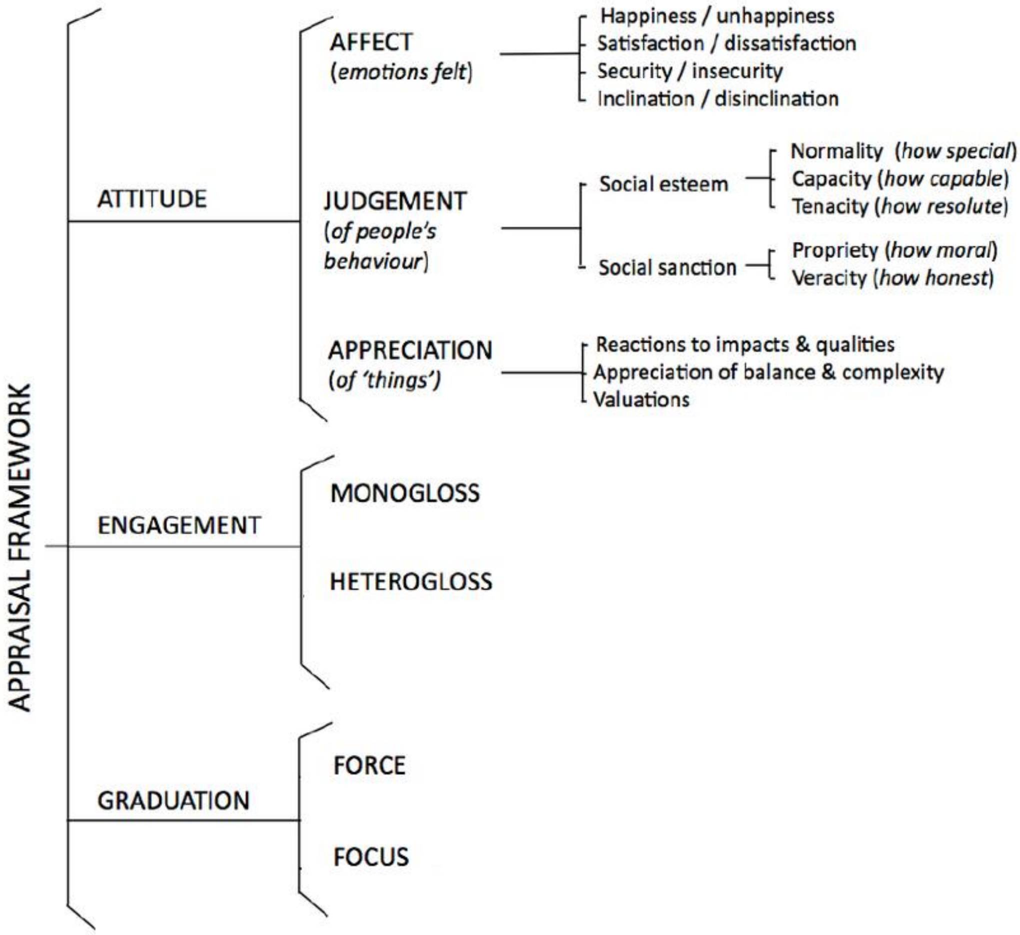

Figure 1. Overview of appraisal framework. Appraisal can be divided into attitude, engagement *and* graduation*, with each category containing further subdivisions. Adapted from (Stewart, 2015).*

2.5 This research

2.5.1 Research questions

The overall aim of this research is thus to add to the existing research about the perception of machine translation (and also AI), at an interesting moment dominated by discussion about LLMs and AI. To accomplish this, the main research question can be formulated as follows: which similarities and differences can be identified in how various stakeholders within the online discourse about machine translation and other translation technology express themselves? To answer this main question, each of the concepts contained within it need to be addressed. Firstly, multiple stakeholders will need to be identified so they can be compared. Which stakeholders can be identified and in which ways can these be characterised? Next, the expression of attitudes needs to be analysed, specifically using the framework of corpus-driven discourse analysis and appraisal theory. Which themes do various stakeholders address related to machine translation and other translation technology as identified through the use of frequency and keyness? What associations with these themes can be identified through the use of concordance and appraisal theory? Lastly, overall conclusions about the similarities and differences need to be drawn, which can be divided up into two steps. The first question to answer is: what attitudes can be identified for the various stakeholders separately? And then finally, to what extent do the various themes, associations and attitudes of various stakeholders overlap or differ?

Which similarities and differences can be identified in how various stakeholders within the online discourse about machine translation and other translation technology express themselves?

- Which stakeholders can be identified and in which ways can these be characterised?

- Which themes do various stakeholders address related to machine translation and other translation technology as identified through the use of frequency and keyness?
- What associations with these themes can be identified through the use of concordance and appraisal theory?

- What attitudes can be identified for the various stakeholders separately?
- To what extent do the various themes, associations and attitudes of various stakeholders overlap or differ?

2.5.2 Expected findings

Based on the existing literature, some expected findings can already be formulated. Firstly, it should indeed be possible to define general differences and similarities between stakeholders. The previous literature in chapter 2 seems to indicate that it should be possible to describe the overall sentiment of a professional stakeholder group, but that some internal disparity can also be expected. A clear example of this is the combination of overall negative sentiment and mixed opinions found for translators in section 2.3.2. For the first sub-question, about the identification of the stakeholders, it should be possible to distinguish these by their professional role (as discussed in section 2.3) and this should indeed be the main distinguishing characteristic, minimising the influence of for example location and other factors. A bottom-up approach may be used to identify other stakeholders in addition to the ones discussed in previous research. For the other sub-questions, themes that might be identified through frequency and keyness analysis would be expected to include some direct discussion of machine translation, for example related to quality or evaluation. However, based on the previous literature discussed in section 2.3, related factors are also likely to arise. For example, LSPs might include mention of costs and productivity (as discussed in section 2.3.1). At least partial overlap in themes would also be expected, for example a general inclusion of the aforementioned themes of costs and productivity. Most difference might be found within the associations. The same themes might be described with an overall more positive or more negative tone or framed differently, based on professional needs. Costs could be linked to concerns for LSPs (as in section 2.3.1) and opportunities for clients (as in section 2.3.4).

3. Methodology

3.1 Introduction

This thesis reports on exploratory research, focusing on the collection of unsolicited data (discourse on machine translation and AI) from various stakeholders, and the application of corpus methodologies and appraisal theory to identify themes and attitudes associated with different stakeholder groups. First, this data collection process is reported in detail, since the outcomes of this process played a significant role in shaping the research overall. The profiles and frameworks used for the data collection are described, along with the technical process of collecting and processing the texts collected from blogs and newspapers. The limitations of the data collection process are discussed throughout. This is followed by a description of the data in the corpus and various subcorpora. Then, the practical methods for analysis of the data are discussed. It is shown how Sketch Engine (Kilgarriff et al., 2014) allows the identification of themes and associations for corpus-driven discourse analysis (Baker, 2006). The final section describes how specific texts for analysis using appraisal theory (Martin & White, 2005) were selected, how the labelling for analysis was done, and how the quantitative and qualitative trends were identified using the labelled data set.

3.2 Data collection

3.2.1 Framework for data collection

To establish a framework for data collection, a bottom-up and iterative approach was employed. As a first step, a search for the keyword "machine translation" was done via the Google search engine and on various (public) social media websites to map the available data on the online discourse surrounding machine translation and AI. This served the dual purposes of presenting the possibility to expand on the stakeholders discussed in section 2.3 and of assessing the actual availability of potential data. Through this process, the initial phase of the data collection itself ended up determining the final framework for data collection. Based on the sub-question from section 2.5.1 (which stakeholders can be identified and in which ways can these be characterised?), professional role was taken as the main distinguishing characteristic within the framework. Additional factors were attempted to be kept equal or similar. It was then possible to increasingly narrow down the following parameters for the search framework: the profiles for the stakeholders based on professional role, the time frame, the region/language of origin of the texts, the platforms to collect the texts on and the keywords used for the various searches.

For the profiles of the stakeholders, since their professional role was the main distinguishing characteristic, the most practical approach was to compile a list of particular organisations belonging to each stakeholder group and gather data from them directly rather than attempting to assign these labels after some other procedure of data collection. Language software development companies exist in a relatively limited number, at least as accessible via listings on the internet, especially when focusing specifically on machine translation/AI translation applications. This made it relatively uncomplicated to compile a list of software companies based in the United States and United Kingdom that run both a blog and LinkedIn page (since LinkedIn was still considered at this point; see below). LSPs explicitly identify themselves as part of that stakeholder group by participating in public directories like those of the Institute of Translation and Interpreting (*Find A Language Service Provider*, n.d.) and the American Translators Association (*ATA's Language Services Directory*, n.d.). Both these directories contain a large number of members, but (similar to the software companies) each member had to be checked for both a blog and LinkedIn page, producing a relatively limited final list. For the category of public (taking clients and end-users together), it was quite difficult to define boundaries since these stakeholders rarely speak directly on the topic of machine translation or AI in a professional capacity. Rather, there are informal jokes (for example on the social media Reddit) or clients and end-users are spoken about (for example in news reporting). Newspapers were eventually chosen as a proxy. This makes it possible to gain at least a general overview of sentiments surrounding AI and machine translation as presented to and expressed by individuals who are neither translators or software developers, although the opinions of translators or developers may of course be quoted in newspaper articles. These articles could be gathered through the use of the newspaper database Nexis Uni. Additional stakeholders identified in the online discourse were professional associations (or interest groups) both for translation and machine translation separately, individual translators and individual software developers. However, it turned out to be impossible to gather data from software developers through establishing a list of individuals or organisations first (from publicly available material at least, as described above), thus making direct comparisons between these additional translation and software stakeholders impossible for this thesis. It seems the stakeholders on the software side apparently present and organise themselves differently online, thus making it impossible to apply the methodology that was employed for collecting data on the translation side.

The time frame had to be relatively recent, so that current discourse fuelled by developments in AI could be captured. However, the frame also had to be wide enough to allow for sufficient data for each stakeholder to be collected to enable corpus-driven discourse analysis. The exact date of publication was not immediately apparent for all data, so the timeframe is not extremely strictly defined, ranging from approximately the end of November 2021 to the end of November 2023.

Through the use of the directories of the Institute of Translation and Interpreting (*Find A Language Service Provider*, n.d.) and the American Translators Association (*ATA's Language Services Directory*, n.d.) to find English language LSPs, the regions were quite naturally restricted to the United States of America and the United Kingdom. This restriction was then in turn applied to the newspapers (used to collect discourse representing public opinion) and language software development companies.

In terms of platforms used for data collection, most social media platforms do not seem to be that inviting to or popular with language professionals. X (formerly Twitter) might qualify but is not usable for research at the moment. Therefore, in terms of spaces where professional attitudes on machine translation might be expressed, the initial focus was on LinkedIn and blogs. For the public, comments on newspaper articles (as parallel for LinkedIn) and the articles themselves (as parallel for blogs) could then be used as a proxy. However, due to technical difficulties, LinkedIn data was not included in this research.

To gather specific texts of interest from each of these stakeholders, keywords had to be further refined through testing. These keywords also varied by stakeholder. For language software development companies, it would be expected that they would discuss the technical side by default, so translation-specific keywords were used to identify discourse about translation technology, namely "translat*" and "locali*", including wild cards to account for various forms of "translation"/"translate"/"translator" and "localisation"/"localise". For LSPs, the opposite would hold true (i.e., that they discuss translation by default), so technology-specific keywords were used, namely "machine translation" and "artificial intelligence". For newspapers, as proxy for the public, a combination was necessary since any number of topics might be discussed. An initial search was done for "translat*", to identify articles that discuss or at least mentioning translation, and within this selection, articles with "machine" were isolated.

As a final piece of the framework, a manual check and selection was necessary to determine whether these texts contained enough evaluative content and were not merely informative. The final parameters for the data collection framework were thus established through iterative testing during the process of data collection itself. The stakeholders investigated are LSPs, language software development companies and the public; the timeframe ranges from November 2021 to November 2023; the locations are restricted to the United States and United Kingdom; the text types are blogs and newspaper articles, manually selected for evaluative content; and the set of specific search keywords varies for each stakeholder. An overview of the final corpus of data is presented in section 3.3.

3.2.2 Collecting and processing texts

As already suggested in the previous section, the available data was rather limited and at least some measure of manual selection was necessary. This precludes the possibility of any true randomness. After determining the framework, specific texts were collected and processed for use in Sketch Engine (Kilgarriff et al., 2014) and Nvivo 14 (*Nvivo*, 2023). For the selected LSPs and software companies, their blogs were searched using the pre-defined keywords for each stakeholder to find blogposts within the timeframe of November 2021 to November 2023. In Sketch Engine, the function "find texts on web" can be utilised to import the text of webpages directly using their URLs. This function was employed in two ways for the data collection, either to import the whole blog including all its posts (which could be sorted; see below) or to import only the specific pages containing the desired blog posts. Minimum file size was set to 0 kB and otherwise standard settings were used. The texts of the webpages are then automatically imported, and part-of-speech tagged by Sketch Engine. When a whole blog was imported, blogposts outside the appropriate timeframe or without the appropriate keywords (either in the file name or in the title within the file itself) were be discarded. As mentioned in the previous section, for all texts a manual check needed to be done to ensure that the texts collected were, in fact, evaluative rather than only informative, which was done mainly by scanning the title and a first section of the text if necessary.

The newspaper articles were found through the database Nexis Uni, with the query "translat*" and within that "machine", in the category news and subcategory papers, with a time frame of within the last 2 years. The results generated from this query were sorted on date from newest to oldest, duplicates were grouped, and then individual American and British newspapers could be selected as a source, ranging from most to fewest hits. From there, articles that seemed appropriate were manually selected and downloaded for each source. For newspaper articles the selection of texts was thus completed first, since directly importing the texts into Sketch Engine was not possible from the particular database used for this research. Selection was mainly based around the piece containing enough evaluative language (as with the blogposts) and discussing as its main topic the impact of AI or machine translation on society (often related to work). These texts could then be manually uploaded into Sketch Engine using the function "I have my own texts".

3.3 Description of corpus and (sub)corpora

The final corpus consists of approximately 211,257 words divided over 153 texts. Even though the subcorpus of the public stakeholder contains the most words, these are divided over the fewest number of texts. The number of sources in the subcorpus of LSPs is the highest, indicating that LSPs might only discuss the topic of machine translation or artificial intelligence once in reaction to recent developments. This also seems to be reflected in the distribution of texts across the chosen time frame. Almost all texts by LSPs were published in either 2022 or 2023. Some newspapers do return to the topic in multiple years and with multiple articles, which is also reflected in the number of texts compared to the number of sources. The language software development companies return to the topic most often, both in terms of timing and number of articles. These distributions connect quite well to the different professional roles of the stakeholders and the relation each has to machine translation and artificial intelligence. The statistics of the corpus and subcorpora are summarised in table 1 below.

Stakeholder	**Text type**	**Number of texts**	**Number of words**	**Region**
LSPs	Blogposts	53 (19 sources)	50,627	Mostly United Kingdom, some United States
Software companies	Blogposts	57 (4 sources)	65,002	Mostly United States, some United Kingdom
Public	Newspaper articles	43 (7 sources)	95,628	United States and United Kingdom
Total corpus		153 (30 sources)	211,257	

Table 1. Characteristics of the complete corpus and each of the subcorpora, including the stakeholder, text type, number of texts and the number of sources contributing these texts (i.e. individual companies), the number of words and the region.

3.4 Corpus-driven discourse analysis

3.4.1 Identification of themes

To identify general themes within each subcorpus, the function "wordlist" was used within Sketch Engine to uncover the most frequently used nouns, verbs and adjectives in each stakeholder subcorpus. Since the number of total words, texts and sources for each subcorpus varies, Average Logarithmic Distance Frequency (ALDF) was used as the standard for the sorting of each list. This measure takes into account the overall distribution of each item within the (sub)corpus, mitigating the variety in words, texts and sources somewhat. From each list, potentially interesting terms were selected within the top 50. This was a large enough number to allow for the identification of themes not only through individual words but the grouping of these words. The function "n-grams" mainly generated lists with terms combining the individual words also contained in the wordlists.

To identify similarities in the themes discussed by the various stakeholders the function "keywords" within Sketch Engine was used to compare the corpus as a whole to the reference corpus enTenTen21 (the most recent corpus of the complete English web in Sketch Engine, gathered according to the methodology described in (Jakubíček et al., 2013)). For these lists the option Score was selected as the standard for sorting, since this is the standard measure of keyness that Sketch Engine provides. Once again, potentially interesting terms were selected from the top 50. For this function, the individual words and word clusters only had partial overlap (unlike with the n-grams), so both individual words and word clusters were taken into account. To identify differences in the themes discussed by the stakeholders, a similar approach was taken, comparing each subcorpus to the corpus as a whole. The specific overlapping and unique themes were then identified by grouping together some of the interesting terms into collections of words and word clusters, similar to Baker et al. (2013). Associations with these themes could then be further investigated through the use of concordance and appraisal theory.

3.4.2 Identification of associations

To examine the associations with the similar and unique themes, the concordancing function in Sketch Engine was used to compile the contexts for each term within a theme. These contexts were then analysed for the overall sentiment around each theme and terms associated with it. The attitude of a particular stakeholder towards translation technology, based on the corpus-driven discourse analysis, thus consists of the similar and unique themes they discuss, made up of specific terms (that may already give an indication of the overall sentiment), and the associations with these themes, in the form of specific words and the overall sentiment associated with the terms within each theme.

3.5 Appraisal analysis

3.5.1 Selection of texts

To enable further analysis of the themes and sentiments with appraisal theory, a limited number of texts needed to be selected. Due to time constraints, only the overlapping theme of business was selected, with the three terms of tool, strategy and business (as described in section 4.1). The texts were then selected from the concordance of these three terms in Sketch Engine. Depending on the number of concordance lines, it was either possible to directly select a text or to generate a random sample of 10 lines using the function "Get a random sample" to select from. For each term one text per stakeholder was manually selected based on the concordance line and title of the text, leading to a total of 9 texts that were used for the analysis with appraisal theory.

3.5.2 Labelling of *attitude*

Each of the selected texts was imported into the qualitative analysis software Nvivo 14 (*Nvivo*, 2023) for coding and further analysis. Each of the appraisal theory labels contained within the category *attitude* (feelings) could be added as a "code" in Nvivo 14 within a hierarchical structure, taking into account the division of *attitude* into *affect* (general feelings), *appreciation* (feelings towards phenomena) and *judgement* (feelings towards behaviour) and their further subdivisions (as illustrated in figure 1). All texts were coded for *attitude* in two rounds. In the first round, the explanations and examples for each label as given by Martin & White (2005) were followed as closely as possible (48-56). After this first round of coding, it became clear that some labels could become extremely dominant through this method and a decision was made to apply a second round of coding with some adjustments, as discussed below. Finally, a general check for consistency was performed by assessing the coded words and word clusters for each label and adjusting these as needed.

There were several reasons why some of the labels could have become very dominant. Firstly, many of the evaluations in most of the texts were indirectly realised or invoked rather than inscribed (Martin & White, 2005, 61-62). Most texts also employed what could be called an objectively presenting style, which involved mainly monoglossic description of things, actions and their impact (represented mainly by the labels *capacity* and also *valuation* in this instance) rather than people and their feelings (represented more by labels under e.g. *affect*) (Martin & White, 2005, 98-102). This also often involved potentially ambiguous evaluation or hypothetical or advisory scenarios. Evaluations that were felt to be too ambiguous were left unmarked, but hypothetical scenarios and advice were coded to labels as expressions of *attitude*. Secondly, the objectively presenting style additionally led to an overabundance of what Martin & White (2005) characterise as "considered opinions" (57), which in this context would mostly be linked to the labels of *capacity* and *valuation.* These characteristics of the texts under consideration would thus flatten any potential differences between the various stakeholders considerably. To somewhat counter this effect, several specifications and considerations were added for the second round of coding.

For this second round, the division between affect, judgement and appreciation was kept as clear as possible. *Affect* is reserved for (internal) emotions, *judgement* for (ethical) evaluation of behaviour and *appreciation* for (aesthetic) evaluation of phenomena, making the object of judgement central to this distinction (Martin & White, 2005, 43-45). However, rather than employing a strict proposition by proposition analysis, a more fine-grained labelling was purposely employed to allow capture of additional data on the evaluative resources used. Thus, in a structure like: "AI can enhance accuracy and efficiency", which might be labelled as a single proposition under *capacity*, each element was instead labelled separately: "AI can enhance (*judgement: capacity: positive*) accuracy (*appreciation: quality: positive*) and efficiency (*appreciation: valuation: positive*)".

For the categories *affect* and *judgement*, the descriptions of Martin & White (2005) were applied as strictly as possible. For *affect*, the label *satisfaction* then relates specifically to feelings of achievement or frustration in relation to activities, *security* specifically to feeling peaceful or anxious in reaction to one's environment, *happiness* to either (internal) feelings of happiness and sadness or (external) feelings of like and dislike and *inclination* relates specifically to intentions towards a hypothetical situation (Martin & White, 2005, 48-51). For *judgement*, the label *normality* represents whether something is considered usual or not, *capacity* represents a judgement of the ability to do something and, especially in the context of this thesis, also the ability to assist in the capacity of others (e.g. AI supports…), *tenacity* represents mindset or inclination, *veracity* represents truthfulness or probability (especially in terms of outcome, rather than mindset), *propriety* represents a judgement of morality in the general sense, a plain ethical judgement of behaviour.

For the category of *appreciation*, Martin & White (2005) indicate how the subcategories contained within can be related to mental processes, where *impact* and *quality* are related to affection, *balance* and *complexity* are related to perception and *valuation* is related to cognition (57). Since the objectively presented style is quite closely connected to considered opinions, i.e. its presentation strongly suggests a process of cognition underlying all statements, the label of *valuation* could be extremely broadly applied. This once again risks a flattening of potential differences between texts and stakeholders. Thus, the labels were encoded mostly with reference to the table of provided examples (Martin & White, 2005, 56), rather than a real consideration of the underlying mental processes. Within this thesis therefore, a term like "accuracy", which could be considered valuable or helpful (thus given the label of positive *valuation*), is instead considered for its general definition and assigned the label *quality* because it can thus be considered a desirable thing in and of itself.

3.5.3 Identification of trends

Trends in the coding and the attitudes of the stakeholders were then identified and are described both quantitatively and qualitatively. The labels indicating the evaluative resources used are first discussed quantitively and graphically represented for each stakeholder. Additionally, since it is possible to group the coded stretches of text by label, the contents for these are described for the various stakeholders to give a qualitative description. The attitude of a particular stakeholder towards translation technology, based on the analysis with appraisal theory, then consists of associations grouped by label and a general description of their most used evaluative resources; in other words, how they express their attitudes. An overall use of the evaluative resources by all stakeholders is also further examined. In chapter 5, synthesis of the various strands can take place and the overall attitude of each stakeholder, based on all data, as well as their key, or evaluative style profile, is discussed. Additionally, a more in-depth comparison between the various stakeholders is performed.

4 Results

4.1 Introduction

In this chapter, the results of the corpus-driven discourse analysis and appraisal analysis are described. Firstly, the results of the corpus-driven discourse analysis are discussed. These include the themes based on frequency and keyness analysis, giving an indication of the overall (positive or negative) tone each stakeholder uses in their discussions as well as both overlapping and unique themes. The associations with these themes, based on analysis of the concordance, are then summarised. By combining these findings, initial impressions of the attitudes of the stakeholders are formulated. For the appraisal analysis, the quantitative distribution of labels (and thus evaluative resources) of each stakeholder is presented and generally compared to the other stakeholders to draw up a profile of evaluative resources unique to each stakeholder. Then, the contents of the labelled text are considered to give an additional qualitative description of associations for each stakeholder, but ordered by evaluative resource rather than theme. These findings can again be combined to describe the attitudes of each stakeholder based specifically on the appraisal analysis.

4.2 Corpus-driven discourse analysis

4.2.1 Themes

For each subcorpus, a general frequency analysis was performed first. Wordlists for verbs, nouns and adjectives were generated for each subcorpus and the words that signal discussion of a certain aspect related to machine translation or AI, or some direct judgement of machine translation or AI, were identified. These selected words are presented in the tables below. Table 2 contains the verbs, nouns and adjectives identified for the public. Table 3 contains the verbs, nouns and adjectives identified for the LSPs. Table 4 contains the verbs, nouns and adjectives identified for the language software companies.

Verbs	**ALDF**	**Nouns**	**ALDF**	**Adjectives**	**ALDF**
think	107	people	167	human	108
write	73	company	104	new	85
work	64	time	88	good	83
create	60	world	87	social	33

learn	57	human	73	different	33
build	56	work	58	hard	30
want	55	industry	52	bad	28
understand	50	end	49	possible	26
become	49	job	48	able	24
believe	46	problem	42	important	20
need	46	person	33	great	20
develop	43	copyright	30	powerful	19
help	35	power	30	wrong	18
train	31	tool	29	global	14
try	31			public	14
produce	25			political	14
require	24				
feel	23				

Table 2. Selected words for the public from the wordlists of verbs, nouns and adjectives. These words seem to signal discussion of a theme related to translation technology or to indicate a direct judgement. Each of these wordlists were sorted using ALDF and the ALDF for each word is presented here.

Verbs	ALDF	Nouns	ALDF	Adjectives	ALDF
translate	123	service	92	human	203
ensure	72	tool	75	accurate	59
provide	54	time	66	good	52
understand	54	process	65	cultural	48
need	49	quality	62	specific	46
work	47	accuracy	56	professional	39
help	47	industry	54	new	34
improve	46	business	52	essential	29
require	43	nuance	42	global	23
learn	38	company	42	significant	23
know	35	expertise	38	high-quality	22
allow	30	world	36	crucial	22
create	27	communication	34	appropriate	18
produce	26	cost	29	complex	18
rely	25	solution	29	necessary	18
reduce	23	human	27	potential	17
enhance	22	benefit	26	important	17
adapt	20	audience	25	certain	16
increase	18	results	25	available	16
deliver	17	source	23	natural	16

remain	17			quick	16
evolve	16			great	14
involve	15			fast	14
result	15			efficient	13
				able	13
				critical	13
				unique	12
				original	12

Table 3. Selected words for LSPs from the wordlists of verbs, nouns and adjectives. These words seem to signal discussion of a theme related to translation technology or to indicate a direct judgement. Each of these wordlists were sorted using ALDF and the ALDF for each word is presented here.

Verbs	**ALDF**	**Nouns**	**ALDF**	**Adjectives**	**ALDF**
use	145	time	111	new	87
need	88	process	84	good	86
provide	78	quality	79	human	55
help	69	business	60	important	54
create	64	service	57	global	44
work	50	customer	55	able	34
improve	46	result	54	accurate	40
learn	39	market	50	specific	29
allow	38	solution	46	easy	27
want	33	tool	42	great	24
increase	33	company	41	possible	22
require	29	experience	38	available	20
enable	28	work	36	additional	18
deliver	28	industry	33	relevant	16
add	26	accuracy	32	traditional	15
ensure	25	workflow	31	efficient	14
change	22	strategy	30	significant	13
reduce	21	audience	28	effective	13
achieve	20	organization	28	crucial	12
		need	27		
		user	27		
		cost	27		
		people	26		

Table 4. Selected words for software companies from the wordlists of verbs, nouns and adjectives. These words seem to signal discussion of a theme related to translation technology or to indicate a direct judgement. Each of these wordlists were sorted using ALDF and the ALDF for each word is presented here.

Based only on this frequency analysis, a first impression of possible themes and the overall framing within each subcorpus can already be formulated. The newspaper articles seem to provide a more general discussion (often not aimed at translation), which is not particularly surprising. The articles highlight both positives and negatives, indicated by very general but direct words such as "good" and "bad", as can be seen in table 2. Also partially as a result of the focus in data collection, these articles seem to mostly discuss the impact of AI on work and the relation of AI to people, indicated by words such as "people" and "job" in table 2. LSPs seem to interpret the use of AI and machine translation as a possible opportunity, by employing words like "potential", "solution" and "benefit", as can be seen in table 3. Overall, the tone in discourse from LSPs seems quite positive, at least through the absence of overtly negative words. There is also a clear focus on quality, using words like "quality", "accuracy" and "appropriate", visible in table 3. Language software development companies also employ words like "solution" or "tool", as seen in table 4, to seemingly discuss the potential of AI and machine translation but seem more focused on the topic of efficiency with both the words "efficient" and "effective" present in table 4. The topic of business is also more emphasised, with words like "industry", "strategy" and "workflow" in table 4.

To further investigate possible themes, similarities and differences within the corpus were examined using the keyness analysis discussed in section 3.4.1. The selected results from this are shown in table 5. Three overall themes that characterise the whole corpus were identified from the keywords, by grouping them thematically (Baker et al., 2013, 262). The first theme concerns humans and their various roles, illustrated by the multiple appearances of "human" within the multiterm keywords in table 5, especially in combinations such as "human translator", "human touch" and "human expertise". The second theme concerns the use of AI or machine translation as a tool for business, implied in keywords such as "translation process" and quite literally stated in keywords such as "translation tool" and "translation solution", visible in table 5. The final theme is that of quality, connected to keywords in table 5 such as "nuance", "translation quality" and "accurate translation". The specific terms connected to each theme, through which the associations with each theme were gathered, were subsequently identified through concordance analysis, as discussed in the next section.

Keywords	Score	Multiterm keywords	Score
chatbot	238,039	human translator	608,615
translator	226,106	human translation	407,021
linguist	191,323	translation tool	266,829
chatbots	186,536	translation solution	208,838

ai-powered	163,338	professional translator	184,757
multilingual	153,216	translation quality	183,504
ai-driven	96,159	accurate translation	158,348
nuance	74,996	cultural nuance	156,402
idiomatic	66,555	localization strategy	153,667
post-editor	58,563	translation process	149,483
		localization process	147,163
		professional translation	145,097
		translation industry	135,954
		human touch	114,522
		high-quality translation	111,432
		human evaluation	106,355
		human expertise	105,171
		human linguist	101,197

Table 5. Selected words from the keyness analysis, indicating similarities in themes between the stakeholders. Both single word keywords and multiterm keywords are included. Each of these wordlists were sorted using the score, indicating the saliency of the various words, and the score for each word is presented here.

The differences between the themes discussed by each stakeholder were obtained through a similar process, but in this case comparing a specific subcorpus to the corpus as a whole (rather than comparing the corpus as a whole to a reference corpus), as described in section 3.4.1. Table 6 shows that the newspaper articles have surprisingly negative unique keywords. They seem to really focus on the potential of AI to be an “existential risk”, or at least a risk in general, which includes a potential for “misinformation” or “disinformation”, a potential for a “bad actor” to misuse AI and the potential to be used in or as a “weapon”, as can all be seen in table 6. With the LSPs, table 7 illustrates that their focus seems to be on the uniqueness of translation and the role of human skill within this. The human aspect is also one of the overlapping themes identified from table 5. However, the LSPs specifically seem to promote the idea of translation quality as something more than “simply” accuracy by employing concepts like “intricacy”, “intended meaning” and “emotional intelligence”, visible in table 7, which also strongly point towards the role of humans as possibly “irreplaceable”. Language software development companies indeed seem mostly focused on the overlapping theme of business, as visible from table 8. One unique aspect that can be isolated from this is customisability, based on the use of “hyper-localization”, “customizable” and “adaptable MT” listed in table 8, a concept which seems to be absent from the discussions by the other stakeholders.

Keywords	Score	Multiterm keywords	Score
weapon	2,295	turing test	2,279
misinformation	2,294	bad actor	2,271
disinformation	2,29	science fiction	2,268
warn	2,288	existential risk	2,268
autonomous	2,277	human reason	2,268
conspiracy	2,277	technological worker	2,263
		climate change	2,263
		conspiracy theory	2,263
		nuclear weapon	2,258

Table 6. Selected words from the keyness analysis, indicating terms unique to the public. Both single word keywords and multiterm keywords are included. Each of these wordlists were sorted using the score, indicating the saliency of the various words, and the score for each word is presented here.

Keywords	Score	Multiterm keywords	Score
human-aided	4,292	idiomatic expression	4,31
multi-skilled	4,282	large project	4,299
intricacy	4,268	human-aided machine translation	4,292
human-powered	4,248	awesome discount	4,287
appropriateness	4,233	website localization process	4,287
bespoke	4,233	multi-skilled content wizard	4,282
misinterpretation	4,214	large volume of text	4,268
irreplaceable	4,214	intended meaning	4,268
natural-sounding	4,187	human professional	4,268
industry-specific	4,187	cultural reference	4,259
refined	4,187	human-powered translation	4,248
round-the-clock	4,187	linguistic accuracy	4,233
regulated	4,149	ethical consideration	4,233
discrepancy	4,149	language services industry	4,233
readability	4,149	limitation of machine translation	4,233
pitfall	4,149	specific need	4,233
		human translation service	4,214
		appropriate translation	4,214
		cultural appropriateness	4,214
		specialized field	4,214
		skilled translator	4,214
		level of quality	4,214
		emotional intelligence	4,214

Table 7. Selected words from the keyness analysis, indicating terms unique to LSPs. Both single word keywords and multiterm keywords are included. Each of these wordlists were sorted using the score, indicating the saliency of the various words, and the score for each word is presented here.

Keywords	Score	Multiterm keywords	Score
ROI	3,35	localization strategy	3,375
locale	3,35	human evaluation	3,368
formality	3,335	contextual ai	3,366
latency	3,335	context awareness	3,361
in-context	3,33	mt evaluation	3,352
self-service	3,33	localization workflow	3,347
evaluator	3,323	localization tool	3,343
language-specific	3,323	quality estimation	3,34
hyper-localization	3,315	adaptable mt	3,34
customizable	3,305	mt quality	3,33
retention	3,305	global experience	3,33
velocity	3,305	automatic metric	3,33
		content owner	3,315
		new audience	3,305
		human feedback	3,305
		translation market	3,305
		ai strategy	3,305
		end user	3,305

Table 8. Selected words from the keyness analysis, indicating terms unique to language software development companies. Both single word keywords and multiterm keywords are included. Each of these wordlists were sorted using the score, indicating the saliency of the various words, and the score for each word is presented here.

4.2.2 Associations

To continue the analysis, the specific associations with the various themes were investigated next. For each stakeholder, their unique theme is discussed in more detail. This is followed by an examination of the associations each stakeholder attaches to the overlapping themes of quality, the human aspect and the business aspect.

Firstly, the results from the concordance analysis for the unique themes are discussed. To investigate the theme of AI as existential risk in newspapers, the terms “weapon”, “warn”, “bad actor” and “existential risk” present the most interesting results in the concordance. A picture emerges where the risk of misuse of AI by bad actors is often presented as a risk to be taken seriously, along with the potential of AI to create autonomous weapons, especially due to a lack of regulation. “He believes if AI becomes more intelligent than humans, it could be exploited by bad actors, including authoritarian leaders” (fileID file30724858). Warnings linking AI directly to existential risks are also reported to be given by both “insider” and “outsider” experts. Simultaneously, when looking at the concordance for the term “existential risk”, the newspapers present an unresolved (ethical) discussion about what the “true” issues and risks of AI are, both current and in the future. “He had always believed that this existential risk was way off, but had recently recalibrated his thinking on its urgency” (fileID file30724733) can clearly be contrasted with “was worried the summit would dwell too much on existential risk and not “real and current issues”” (fileID file30724843).

The uniqueness of translations for LSPs was investigated using the terms “intricacy”, “appropriateness”, “idiomatic expression”, “intended meaning” and “limitation”. Starting the examination from “limitation”, LSPs seem to acknowledge some of the possibilities of AI and machine translation, but mainly stress the fact that these should be combined with human input: “Machine translation isn't without its limitations. Enter human-aided machine translation, a powerful combination of artificial intelligence and human expertise” (fileID file31093009). It is suggested that idiomatic expressions and intended meaning may sometimes be resolved by the translation technology, but that these are quite closely linked to the intricacies inherent in language and that humans are essential for resolving these appropriately: “Technology is still behind human translation in one crucial area: its inability to pick up on the idioms, cultural references , and nuances inherent to any language. This is where adding the human touch to machine translation becomes essential” (fileID file31093803).

The final unique theme, the customisability of machine translation as discussed by language software development companies, was investigated through the terms “hyper-localization”, “customizable” and “adaptable”. These companies seem to be presenting customisability as the next big step in the development of machine translation, for example by applying this in the process of hyper-localisation: “The global demand for content is only going to increase, and customizable tools like our Context Awareness models will become increasingly critical” (fileID file30696374). Especially in the concordance of the term “adaptable”, this is also explicitly linked to the use of pre-existing company resources and the ability of businesses to do their own customisation using these resources: “Adaptable MT will give you the ability to more accurately translate this content with your existing SMEs and datasets” (fileID file30945620).

For the overlapping themes, the focus of this investigation was mainly the clear differences between the associations of the various stakeholders, to highlight how the same theme might receive different framings. The specific terms for each theme were determined on the basis of the concordance lines from the newspaper articles (representing the public stakeholder). These texts are much less likely to contain specialised terms that might appear in the texts of the LSPs or software companies. Based on this initial concordance analysis, some terms were left out or simplified to make direct comparison across all stakeholders possible.

The first of the overlapping themes is the aspect of quality, which was examined for each stakeholder using the terms “quality”, “nuance” and “accuracy”. In the newspaper articles, the main focus seems to be on morality rather than “pure” quality. The discussion does not end with the (potential) capabilities of AI, rather, these are placed within the broader debate of whether they should be utilised. LSPs do acknowledge the existing accuracy and potential to improve of machine translation, although humans are clearly presented as the standard for quality. Especially when considering the concordance of “nuance”, humans are usually presented as superior to machine translation and thus essential to translation quality. One LSP argues that “[h]uman translators are able to pick up on cultural nuances , connotations, and emotions in the text, which [neural machine translation] systems can still sometimes miss” (fileID file30947643). Language software development companies clearly focus on accuracy over nuance, which is visible in the number of concordance lines for both keywords (101 and 18 respectively). They emphasise the current level of quality and especially the potential for improvement of machine translation, but overall do focus on the efficiency of machine translation, which then allows businesses to strike an appropriate balance between quality, speed and costs: “When measuring the success of your localization programs and initiatives [sic], keep your eyes on cost, speed, and quality of translation” (fileID file30711360).

The overlapping theme focusing on the human aspect was investigated using the terms “translator”, “human touch” and “human translation”, thus linking it firmly to the specific topic of translation. Newspaper articles here mainly reflect an anxiety about either outright replacement or the need to adjust to a new role: “Increasingly, says Bone, she and her colleagues are acting as editors of a machine's first pass, rather than translators of the raw material. For some, that's fine”(fileID file30724882). LSPs mainly argue that human translators remain important due to their translation specific skills (e.g. creativity, interpretation and comprehension), with only few LSPs more directly opposing machine translation and others speaking of a combination of human and machine translation. For language software development companies, human translation is clearly presented as the standard, for example by declaring “accuracy of up to 96%, which emulates the best of human translation” (fileID file30695875). Simultaneously, there seems to be a general inclination to reduce the human touch. Examining the concordance of “translator” specifically, even within the only 4 software companies included in the corpus, there seems to be split opinions on whether the goals is to “reduce […] reliance on human translators” (fileID file30696026) or to build relationships with translators so the companies do not "risk alienating the very group of people that make this industry work" (fileID file30945627).

The final overlapping theme is that of business, investigated with the terms "tool", "strategy" and "industry". Newspapers highlight the potential of AI, both for good and bad, both in its usage in general and its impact on various industries. LSPs mainly focus on the limitations of AI and machine translation, suggesting that both translation technology and human translators can be incorporated into an overall business strategy: "A hybrid translation strategy uses the best features of both machine translation (MT) and human translation" (fileID file31093803). Language software development companies present AI and machine translation as a "tool" in the most literal sense, which can be used to accomplish certain objectives and should incorporated as part of an overall business strategy: "If you don't already have an AI strategy in place, now might be the time to conduct an AI strategy evaluation" (fileID file30707768).

4.2.3 Attitudes of stakeholders based on corpus-driven discourse analysis

On the basis of the corpus-based discourse analysis, a very general impression of the attitudes of the various stakeholders can now be characterised through its themes and associations. The attitude of the public as expressed through newspaper articles could be presented as somewhat balanced, in that it represents both good and bad aspects of translation technology, although also somewhat anxious, especially in relation to the job market and the potential use of AI in weapon development. Overall, there seems to be a fairly strong focus on morality, centred around the question whether potential benefits will outweigh potential risks, both in the short and long term. The attitude of LSPs as expressed through their blogposts seems to suggest a cautious optimism regarding the potential of AI and machine translation. The main focus for the LSPs is to stress that human translators are essential to the process of translation and that they can alleviate (some of) the limitations of translation technology. The attitude of the language software development companies as expressed through their blogposts can be characterised as by far the most positive. Drawbacks are acknowledged but also often framed in terms of efficiency, where quality should be in balance with costs. Overall, there is a strong focus on business, which can be described as the aim of enabling businesses to use translation technology as an effective part of their overall business strategy.

4.3 Appraisal analysis

4.3.1 Quantitative description of findings

In this section, the distribution of labels for each stakeholder is briefly considered and compared quantitatively. For LSPs, there is a strong presence of positive labels overall, as is visible in figure 2. Table 9 in turn shows, within the category *affect*, *satisfaction* (positive) is the most used resource, but this is due to one of the texts containing 18 counts of a total of 19. Within the category *appreciation*, the resource *valuation* is used most by LSPs, complemented by *quality*. Compared to the other stakeholders, the counts of *balance* and *quality* are relatively high, as can be seen in table 9. *Capacity* is by far the most used resource within the category *judgement*, but relatively comparable to the instances of *capacity* used by the language software development companies (126 compared to 116 by the software companies).

The public presents a much more balanced view, with a more equal distribution between positive and negative evaluations clearly visible in figure 2. Within the category *affect*, the resource of *insecurity* stands out most with 15 counts in table 9. Overall, the resources *security* and *insecurity* are most often used by the public as well as more negative *affect* in general. Like the LSPs, *valuation* is also the most used resource within the category *appreciation*. However, the difference between positive and negative *valuation* is much smaller compared to both other stakeholders. *Impact* is also used relatively often within this category, but comparable to the use by LSPs (23 compared to 26 by LSPs). Within the category of *judgement*, *capacity* and *propriety* seem to be the main concerns. Especially *capacity* and *impropriety* are often used resources by the public. The overall division between resources is much more spread out compared to the other stakeholders and also once again more balanced between negative and positive.

The language software developments companies present a strongly positive view, similar to the LSPs, as visible in figure 2. Within the category *affect*, *inclination* is the most used resource due to one of the texts containing 15 counts. Within the categories of *appreciation* and *judgement*, table 9 shows the language software development companies are quite comparable to the LSPs, only with an even stronger focus on *valuation*.

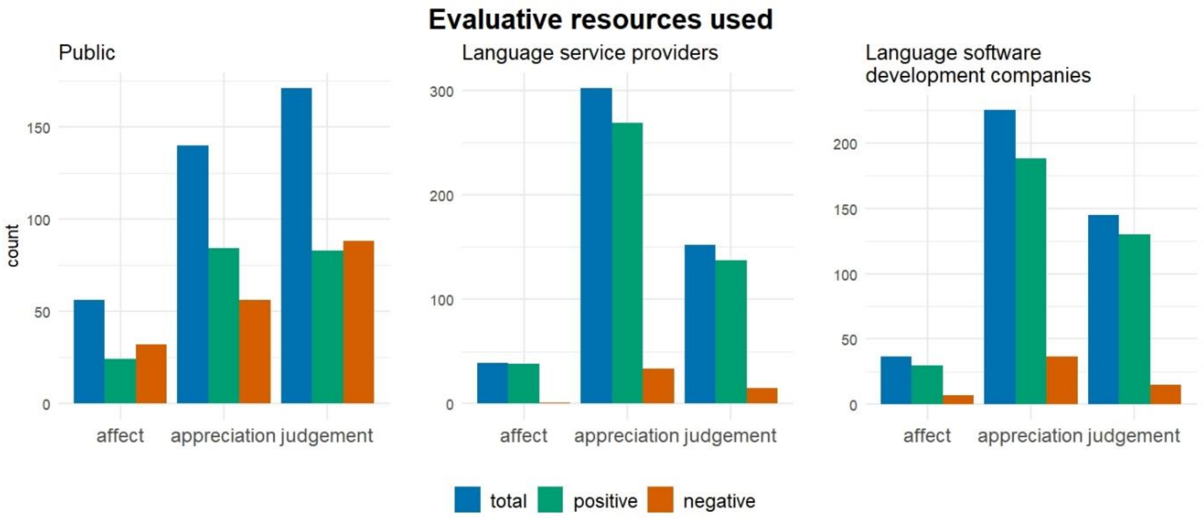

Figure 2 Evaluative resources used by each of the stakeholders. This figure shows the number of labels assigned (count, on the y-axis) in each category of attitude*:* affect, appreciation *and* judgement *(indicated on the x-axis). It indicates both the total count of labels used and the count of positive and negative labels each.*

	Total	LSP	Public	Software		Total	LSP	Public	Software		Total	LSP	Public	Software
Affect	**132**	**39**	**56**	**37**	**Appreciation**	**667**	**302**	**140**	**225**	**Judgement**	**468**	**152**	**171**	**145**
Happiness					Balance					Capacity				
Positive	10	6	1	3	Positive	59	39	0	20	Positive	273	115	54	104
Negative	6	0	6	0	Negative	4	1	1	2	Negative	37	11	14	12
Inclination					Complexity					Normality				
Positive	43	12	9	22	Positive	25	12	4	9	Positive	44	16	9	19
Negative	7	0	6	1	Negative	7	3	0	4	Negative	10	2	5	3
Satisfaction					Impact					Propriety				
Positive	27	19	5	3	Positive	56	24	20	12	Positive	12	2	10	0
Negative	6	1	5	0	Negative	8	2	3	3	Negative	40	2	38	
Security					Quality					Tenacity				
Positive	12	1	9	2	Positive	129	81	9	39	Positive	18	3	9	6
Negative	21	0	15	6	Negative	20	7	12	1	Negative	15	0	15	0
					Valuation					Veracity				
					Positive	272	113	51	108	Positive	3	1	1	1
					Negative	87	20	40	27	Negative	16	0	16	0

Table 9. Counts of the labels encoded and thus evaluative resources used by each stakeholder for each "feeling" contained in attitude*. Overall totals are also included. Each vertically divided section represents one of the subcategories of* attitude*:* affect, appreciation *and* judgement.

4.3.2 Qualitative description of findings

In addition to the quantitative distribution of resources, the contents of the labelled textual elements can be summarised to give a qualitative overview. For each stakeholder, the main focus will be their (comparatively) most prominent resources. For LSPs, within the resource of *balance* are contained textual elements that cluster around the concepts of consistency and coherence (in line with the example "consistent" of (Martin & White, 2005), describing internal consistency (56)), and ones that cluster around the concepts of appropriateness, customisation and adaptation (in line with the examples "consistent", "harmonious" and "considered", describing consistency with something else (Martin & White, 2005, 56)). The resource *quality* contains textual elements clustering around the concepts of accuracy, fluency and quality, which can be seen as outright mentions of quality, and elements clustering around the concepts of understanding and expertise (as a measure of internal quality), which is mostly linked to humans. Comparatively, in the newspaper articles representing the public the "quality" of humans and machines is compared more, and within the texts of the software companies quality and accuracy are most closely linked. For the public, the resource *insecurity* is mostly used to express concern. The resource of specifically negative valuation contains mostly expressions of risks or issues related to AI (which could be related to the examples of "shoddy" and "ineffective" (Martin & White, 2005, 56)), compared to more of a focus on issues that might be resolved through AI in the texts of the other stakeholders. Through the resource of *propriety*, concerns of morality are quite directly expressed by contrasting terms like "responsible" (fileID file30724703) and "morally" (fileID file30724735) with terms like "cause harm" (fileID file30724703) and "misuse" (fileID file30724735). For the language software development companies (and really all stakeholders), several clusters can also be identified in the resource of *capacity*. One set of textual elements clusters around the concept of facilitating capacity, usually in the form of AI or machine translation enabling/empowering/facilitating for a human. Another set clusters around what could be termed "quality assurance", ensuring or guaranteeing some element of quality. The last set clusters around added competence, in the form of improvement or optimisation for example.

4.3.3 Attitudes of stakeholders based on appraisal theory

On the basis of appraisal theory, another impression of the attitudes of the various stakeholders, and especially of the resources they use to express these, can be compiled. LSPs seem to mainly focus on *capacity* but relate this more than the other stakeholders to a combination of *balance*, *quality* and *valuation*. Especially with the resource of *quality*, a clear link is made to human expertise. The public seems to provide a more balanced account overall. Here too, *capacity* and *valuation* are often discussed, but more focus is also on what could be called the "human" side, by employing the resources *security* and *insecurity* and also (negative) *tenacity*, as well as the moral side, by employing the resources of *propriety* and *veracity* in rather direct form. The language software companies can be considered quite comparable to LSPs in their profile of evaluation resources, but with an even stronger emphasis on *capacity* and *valuation*.

4.4 Evaluative resources used by all stakeholders

To round out this results chapter, the overall distribution of various resources for all stakeholders together as well as various connections between the resources are briefly discussed. The most employed resources fall within the categories of *appreciation* and *judgement,* as can be seen in figure 3. Within the category of *appreciation*, these are mostly *quality* and *valuation*, and within the category of *judgement* this is mostly *capacity*, as could already be concluded from the previous discussions of the individual stakeholders in section 4.3. In all texts then, there is a central discussion surrounding *capacity*, which is also often linked to *quality* or *valuation* to create the following structures: able/unable/enabling to do/cause this positive thing or able/unable/enabling to resolve this negative thing. This structure especially applies to the *capacity* of AI and machine translation. There is some general discussion of what translation technology is not able to do (yet), but overall, *capacity* is very much used in the positive sense. This holds especially true for the texts of the language software development companies. For the LSPs, considerations of *balance* and the human aspect are added to this discussion of *capacity*. Interestingly, in both texts from LSPs and language software companies, a connection is sometimes made between human as the standard for quality (positive quality) and the desire to "free" them from manual tasks (manual taken as negative valuation). "Our system allows skilled linguists to review AI-generated translation suggestions instead of starting from scratch, guaranteeing quality" (fileID file30707987). For the public, considerations of (the morality of) the consequences of such capabilities are added to the discussion. This is made visible especially in instances where the *capacity* of AI is linked to (human) *insecurity* or negative *propriety*, *tenacity* or *veracity*, indicating that the capacities of AI can lead to undesirable consequences.

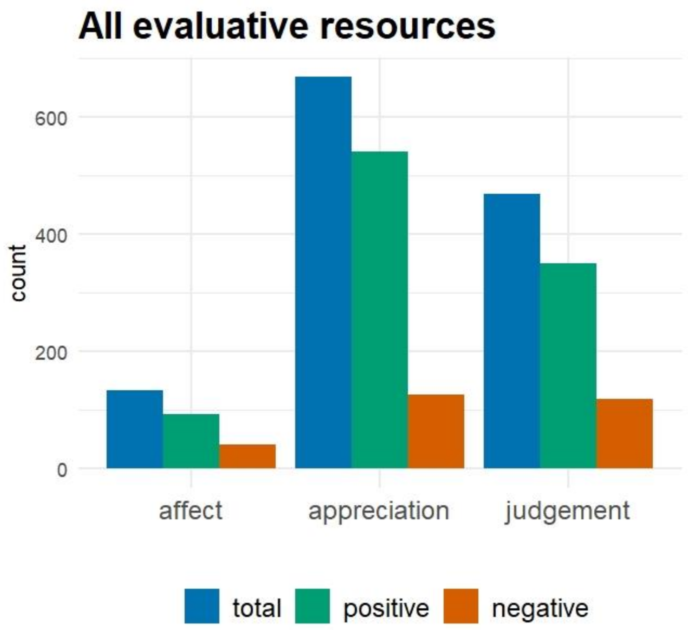

Figure 3 Evaluative resources used by each of the stakeholders. This figure shows the number of labels assigned (count, on the y-axis) in each category of attitude*:* affect*,* appreciation *and* judgement *(indicated on the x-axis). It indicates both the total count of labels used and the count of positive and negative labels each.*

5. Discussion

5.1 Introduction

In this chapter, the material from previous chapters is synthesised to arrive at final conclusions. The attitudes and resources used by the stakeholders are compiled into one complete overview. Then, these attitudes and resources can be compared between the various stakeholders as well as with the previous literature outlined in chapter 2. To end the chapter, some of the limitations of the research will be discussed.

5.2 Attitudes and resources used by stakeholders

In this section, the findings from the corpus-driven discourse analysis and analysis with appraisal theory can be combined and further interpretated to synthesise a final description of the attitude of each stakeholder towards machine translation. Both methodologies seem to complement each other in this. Additionally, a key or evaluative style for each stakeholder can be described by generalising from the evaluative profile and highlighting the most characteristic differences with the other stakeholder.

5.2.1 Attitude and key of public

The attitude of the public was investigated through the proxy of newspaper articles, which present a variety of opinions. What emerges overall seems to be a relatively balanced although somewhat sceptical picture. This is especially visible through the use of a mixture of positive and negative evaluative resources and lexical items. It is clearly reflected, both through the use of evaluative resources and the findings from the corpus analysis, that much attention is paid to both the capabilities of translation technology and potential unwanted consequences that may arise from these. This leads to a much broader and more morally focused discussion compared to the other stakeholders. The question whether benefits will outweigh risks, both short-term and long-term, does seem to be answered rather sceptically. This especially relates to issues of job security and the potential (mis)use of AI in weapon development. Overall, although optimism is admitted and discussed, the current attitude of the public towards translation technology and AI in particular seems to be rather on the anxious and pessimistic side.

Since the public was only represented by proxy through newspaper articles, the key of the public is very much related to this text type. This key first and foremost seems to be characterised by the need to present a balanced account, employing positive and negative evaluations near equally (191 positive against 176 negative evaluative instance). This desire of a balanced account is further reflected through the co-occurrence of positive *capacity* with *insecurity* or negative *propriety*, *tenacity* or *veracity*. Thus, the public as represented through newspaper articles is in the first place an outside observer, who weighs the negative and positive aspects of AI and machine translation carefully against each other. This effect is added to through the fact that the newspaper articles are often at least partially a compilation of others' opinions rather than purely the author's own. As has been a running thread throughout the analysis, the main intent of newspaper articles seems to be to engage in a broader and generally ethically focused discussion, an aim which can be facilitated by taking on the role of an outside observer.

5.2.2 Attitude and key of LSPs

The general attitude of the LSPs can be defined as optimistic, although in some respects cautious. The optimism is mainly shown through the focus on the capabilities of translation technology, visible in both evaluative resources and lexical items. Caution is mainly reflected in the insistence on human expertise. LSPs link human involvement explicitly to quality and insist that humans are essential for this aim. They combine this with a discussion of the uniqueness of translation to illustrate some of the shortcomings of (current) translation technology. This approach of insistence on human involvement and overall attitude of cautious optimism are not surprising given the stakeholder position of LSPs, which will be discussed in more detail in section 5.3.

The texts of the LSPs can be broadly described as marketing texts, which are presented more or less overtly. Through the strong focus on *capacity* and various resources within *appreciation*, the main message of these texts is centred around possibilities. The variety of resources then indicates the multiple benefits that may arise from these possibilities. The key of the LSPs can thus described as a marketing voice attempting a broader appeal. When combined with the previous discussions of uniqueness of translation and necessity of humans, the main aim of the LSPs for these texts seems to be to assure themselves of a (continued) role in the translation process by appealing to the fact that they can provide additional benefits outside of concrete added value in terms of e.g. efficiency.

5.2.3 Attitude and key of MT companies

The attitudes of the language software development companies towards translation technology are by far the most positive. The strong emphasis on the specific resources *capacity* and *valuation* can now be linked to the strong focus on business found in the corpus analysis. The software companies are indeed mainly focused on the opportunities of increased efficiency and how these technological products can be brought to companies to become part of their overall business strategy. Although shortcomings are generally somewhat acknowledged, there is much emphasis on how this can be part of a strategy to balance quality, speed and costs and how there is a large potential for improvements. The software companies thus mainly display confidence in translation technology.

Similarly to the texts of the LSPs, the texts of the language software development companies can be described as marketing texts, often even more overt than the LSP texts. The key of the language software development companies can thus be defined as a marketing voice with a narrow appeal. Rather than discussing multiple benefits, the evaluative resources are less spread out and thus more targeted. It can be argued that software company marketeers might be convinced that the concrete added value of their products is convincing enough on its own. The aim of the blogposts would thus in the first place be to present this added value simply, making more focused use of fewer evaluative resources.

5.3 Comparison of stakeholders, their attitudes and relationships

When considering the attitudes of the various stakeholders together, they indeed give some insight into the current discourse surrounding translation technology. In this moment, the language software development companies display confidence in their position of power as providers of translation technology. The LSPs can use these developments as an opportunity to retain their position, but only if they can successfully incorporate these technologies into their own processes without rendering their involvement obsolete. The public is rather torn about the potential benefits of these technologies, especially the more general use cases for AI – an anxiety that is quite understandable since the advent of AI is also very much presented as inevitable and as already having arrived in the texts of all stakeholders. Interestingly, the conclusions from this thesis seem to suggest most resistance to AI and machine translation coming from the public and quite little from the LSPs, who put a much more prominent emphasis on hybrid solutions.

Compared to the literature discussed in chapter 2, the split approach of LSPs is thus not reproduced in this research. This could be a product of the data collection procedure inadvertently excluding the more negative voices or indicate a general shift in attitudes within the LSPs. The generally mild attitude found in previous research is also not entirely reproduced, although this could be argued to be in line with the cautious optimism identified here. Different research methodologies would surely have an effect on how particular attitudes are observed to be expressed, with these public and outward presenting texts taking on a more assured tone. The internal issues and struggles of LSPs in regard to machine translation, as found in the previous literature, would not present itself in these texts. If this data does indeed indicate a general shift in attitude, this might be linked to an even further increase in pressure from clients to provide machine translation services (in line with the discussions by Sakamoto (2021) and Sakamoto & Yamada (2020)). In section 4.2.2, it was briefly addressed that there does seem to be somewhat of a split between some language software development companies on whether humans, specifically translators, should be (more) involved with translation technology. Overall, however, the emphasis seems to be very much on empowering businesses to employ various translation technology tools directly. This then would be in line with the general trend of software developers focusing on the development of (autonomous) machine translation systems, as found in section 2.3.3. For the public, there seems to be more of a trend of anxiety than was signalled in the previous literature. This could very well be linked to the use of the newspapers articles as proxy, with the newspapers taking a more critical stance in general compared to the "average" end-user or client. However, it could also once again indicate a general shift in sentiment, in which a generally more critical attitude is employed specifically by the public. Since the less critical attitude was suggested to possibly be related to a lack of awareness of and access to alternatives in section 2.3.4, this more critical attitude could be related to gaining more awareness and access and thus a broader view of machine translation and AI. Overall, at the least the broad strokes of the findings in this thesis seem in line with the findings from previous research.

5.4 Limitations of this research

Finally, some limitations of the research will be discussed to end this chapter. Within the context of this thesis, it was only possible to give a very general impression of the discourse surrounding machine translation and AI and the influence of their professional role on the attitudes of the stakeholders. Especially the process of data collection was rather experimental and necessarily quite limited. Firstly, due to the restrictions of the thesis, both in terms of time and other practicalities, the data collection was very much based on convenience sampling, rather than being able to apply any true randomness. This was especially forced by the fact that the starting point of data collection had to be the stakeholders themselves in order to properly organise the data for analysis and this identification of the stakeholders was kept rather simple out of necessity. Additionally, since the data collection was based on the use of specific and rather restricted search keywords, certainly much material that only refers indirectly to machine translation or artificial intelligence was not taken into account. For the final step of data collection, the texts had to be manually assessed for the amount of evaluative language they contained, to determine whether the text would yield enough material for analysis.

For the corpus-driven discourse analysis, only a limited number of words from the various wordlists could be considered to make (manual) analysis manageable. Each step of narrowing down the list from the general wordlist to the list of potentially interesting terms to the list of terms contained in the themes was done manually based on personal judgement. For the analysis with appraisal theory, the number of concordance lines to choose from was already quite limited for most terms and from there the number of texts to be analysed had to be narrowed down even further, which was once again done through manual selection based on personal judgement. It was sadly not possible to perform the analysis together with another coder, which would have decreased the subjectivity of the coding. Especially for the invoked evaluation, as mentioned in section 3.5.2, the weight given to particular elements and their context can very much influence what label is assigned to a particular piece of text by a coder. This issue is also mainly what led to the adjustments after the first round of coding and clarification of these adjustments in section 3.5.2. Analysis with appraisal theory still yielded interesting results that complemented the ones from the corpus-driven discourse analysis well. However, it is clearly best suited to texts that express sentiments more directly or are at least more centred around human sentiments and reactions.

6. Conclusion

6.1 Summary

To conclude this thesis, the research questions can be definitively answered before offering some suggestions for future research. Firstly, the stakeholders were identified and characterised through the process of data collection. Language software development companies were identified and characterised through internet listings pertaining to translation technology. LSPs were identified and characterised by their presence in public LSP registries, essentially self-declared identification as LSPs. The public, end-users and clients, were hard to identify precisely and eventually characterised through the proxy of newspaper articles to gain a general insight. For these stakeholders, overlapping themes could be identified as quality, the human aspect and the business aspect. Unique themes included existential risk for the public, uniqueness of translation for LSPs and customisability for software companies.

Associations with the themes of quality for the public included a focus on morality, questioning whether the capabilities of machine translation and AI would truly be (mostly) beneficial. For LSPs this included an acknowledgement of the capabilities of translation technology but also an insistence on humans as a quality standard. For language software development companies, the focus was mainly on accuracy as well as the idea of balancing quality with speed and cost. Associations with the theme of the human aspect were mostly coloured by anxiety for the public, especially in relation to the job market. For LSPs, this included arguments about the importance of human translators. For software companies, human translation remains a standard of quality, but some also argue human involvement in the process of translation should be reduced. Associations with the theme of business were investigated both through the use of corpus-driven discourse analysis and analysis with appraisal theory. This showed the public painting a relatively balanced picture, although with once again a strong moral focus, explicitly highlighting both good and bad sides of both the usage and impact of AI and machine translation. For LSPs, although limitations were discussed, overall possibilities were highlighted more. Software companies focused most on the integration of translation technology tools in business strategies as a means to promote efficiency.

In the overall attitude of the public, the focus on the broader context and more moral debate is quite clear. The public is mainly concerned with weighing the benefits and risks of machine translation and AI in a (somewhat) balanced way but seems to overall come to a more pessimistic conclusion. The overall attitude of the LSPs can be characterised as cautious optimism. The optimistic aspect is represented by the strong focus on capabilities, both of translation technology and human translator. The cautious aspect is represented by the insistence that those human translators are still essential in the translation process. Language software development companies have by far the most positive attitude. Although shortcomings are acknowledged, the main focus is on improving efficiency by embedding translation technology in the overall business strategy of companies.

Thus, it is indeed possible to identify both similarities and differences between the different stakeholders. As expected, the similarities are mostly present in the themes and the differences mostly in the associations. Some similarities are also present in the evaluative resources used, so in the ways the attitudes are expressed. The main similarity is that the capabilities of AI and machine translation are generally much discussed and also acknowledged in the attitudes of all stakeholders. However, differences arise in the further associations with AI and machine translation. Some of the quality is thrown into doubt by the LSPs, insisting humans need to be involved in the process of translation to reach the desired level of quality. The public, on the other hand, is mainly concerned with the (potential) consequences of these capabilities. This leads to an overall gradation of attitudes with software companies as the most optimistic, the public as rather more pessimistic and the LSPs somewhere in the middle with a cautiously optimistic attitude towards machine translation and AI. Through this thesis research it was possible to provide at least a general answer to all research questions.

6.2 Suggestions for future research

As the final part of this thesis, some suggestions for future research are offered. First of all, a more in-depth continuation of this research might yield interesting results. This might be done either by considering the various elements of corpus-driven discourse analysis and appraisal theory more completely, for example by adding an analysis of *engagement* in addition to the analysis of *attitude*, or by recording more in depth what the object of each evaluation was, to construct an impression of what resources and words and word clusters each stakeholder associated with that object. This might have afforded a more direct description of attitudes towards for example AI or machine translation through the qualitative analysis with appraisal theory. Additionally, more data could be added. One way to accomplish this would be through a broader use of keywords. An attempt could be made to involve more stakeholders, although an appropriate procedure for data collection would need to be chosen. Other text types, including from social media, could be added as well. Especially an opportunity to add more "feeling" focused texts could make the combination of corpus-driven discourse analysis and appraisal theory even more productive. Finally, investigating this discourse on machine translation and AI through an explicitly non-professional lens, truly investigating the public sentiment as expressed in media and on social media in more detail, might be a worthwhile endeavour, especially if there is indeed a shift in public opinion taking place as seems to be (tentatively) indicated by the results of this thesis.

Bibliography

Ai, Q. (2022). Reflection on the machine translation and human translation in the age of artificial intelligence. *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 1–6. https://doi.org/10.1109/ICERECT56837.2022.10060519

ATA's Language Services Directory. (n.d.). American Translators Association (ATA). Retrieved 2 February 2024, from https://www.atanet.org/directory/.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). Sketching Muslims: A corpus driven analysis of representations around the word 'Muslim' in the British press 1998–2009. *Applied Linguistics*, *34*(3), 255–278. https://doi.org/10.1093/applin/ams048

Cadwell, P., O'Brien, S., & Teixeira, C. S. C. (2018). Resistance and accommodation: Factors for the (non-) adoption of machine translation among professional translators. *Perspectives*, *26*(3), 301–321. https://doi.org/10.1080/0907676X.2017.1337210

Carvalho, I., Ramires, A., & Iglesias, M. (2023). Attitudes towards machine translation and languages among travelers. *Information Technology & Tourism*, *25*(2), 175–204. https://doi.org/10.1007/s40558-023-00253-0

Castilho, S., & O'Brien, S. (2017). Acceptability of machine-translated content: A multi-language evaluation by translators and end-users. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, *16*. https://doi.org/10.52034/lanstts.v16i0.430

Cotelli Kureth, S., Delorme Benites, A., Haller, M., Noghrechi, H., & Steele, E. (2023). "I looked it up in DeepL": Machine translation and digital tools in the language classroom. *Studie e Richerche: Human Translation and Natural Language Processing Towards a New Consensus?* https://doi.org/10.30687/978-88-6969-762-3/006

Find A Language Service Provider. (n.d.). Institute of Translation and Interpreting (ITI). Retrieved 2 February 2024, from https://www.iti.org.uk/find-professional-translator-interpreter/find-a-language-service-provider.html.

Fulford, H. (2002, November 14). Freelance translators and machine translation: An investigation of perceptions, uptake, experience and training needs. *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*. EAMT 2002, Manchester, England. https://aclanthology.org/2002.eamt-1.13

Gao, Y., Wang, R., & Hou, F. (2023). *How to design translation prompts for ChatGPT: An empirical study* (arXiv:2304.02182). arXiv. https://doi.org/10.48550/arXiv.2304.02182

Guerberof Arenas, A. (2013). What do professional translators think about post-editing. *The Journal of Specialised Translation (JoSTrans)*, *19*, 75–95.

Guerberof Arenas, A., & Moorkens, J. (2023). Ethics and machine translation: The end user perspective. In H. Moniz & C. Parra Escartín (Eds.), *Towards Responsible Machine Translation* (pp. 113–133). Springer. https://doi.org/10.1007/978-3-031-14689-3_7

Guo, X. (2022). Optimization of English machine translation by deep neural network under artificial intelligence. *Computational Intelligence and Neuroscience*, *2022*, e2003411. https://doi.org/10.1155/2022/2003411

Hunston, S., & Thompson, G. (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press. http://catdir.loc.gov/catdir/enhancements/fy0640/00701290-d.html

Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). *The TenTen corpus family*. https://is.muni.cz/publication/1120431/en/The-TenTen-Corpus-Family/Jakubicek-Kilgarriff-Kovar-Rychly

Kasperė, R., Horbačauskienė, J., Motiejūnienė, J., Liubinienė, V., Patašienė, I., & Patašius, M. (2021). Towards sustainable use of machine translation: Usability and perceived quality from the end-user perspective. *Sustainability*, *13*(13430), 13430. https://doi.org/10.3390/su132313430

Kasperė, R., Motiejūnienė, J., Patasienė, I., Patašius, M., & Horbačauskienė, J. (2023). Is machine translation a dim technology for its users? An eye tracking study. *Frontiers in Psychology*, *14*. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1076379

Kenny, D. (2011, June 5). *The ethics of machine translation*. New Zealand Society of Translators and Interpreters Annual Conference 2011, Auckland, New Zealand. https://doras.dcu.ie/17606/

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, *1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Läubli, S., & Orrego-Carmona, D. (2017). When Google Translate is better than Some Human Colleagues, those People are no longer Colleagues. *Läubli, Samuel; Orrego-Carmona, David (2017). When Google Translate Is Better than Some Human Colleagues, Those People*

Are No Longer Colleagues. In: Translating and the Computer, London, 16 November 2017 - 17 November 2017. AsLing, 59-69., 59–69. https://doi.org/10.5167/uzh-147260

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan. http://catdir.loc.gov/catdir/enhancements/fy0624/2005043360-t.html

Moorkens, J., & Rocchi, M. (2020). Ethics in the translation industry. In K. Koskinen & N. K. Pokorn (Eds.), *The Routledge Handbook of Translation and Ethics* (1st ed., pp. 320–337). Routledge. https://doi.org/10.4324/9781003127970-24

Nunes Vieira, L., & Alonso, E. (2020). Translating perceptions and managing expectations: An analysis of management and production perspectives on machine translation. *Perspectives*, *28*(2), 163–184. https://doi.org/10.1080/0907676X.2019.1646776

Nvivo (Version 14). (2023). [Computer software]. Lumivero. https://lumivero.com/products/nvivo/

Pham, V. H., Pham, T. M., Nguyen, G., Nguyen, L., & Dinh, D. (2023). *Semi-supervised neural machine translation with consistency regularization for low-resource languages* (arXiv:2304.00557). arXiv. https://doi.org/10.48550/arXiv.2304.00557

Ragni, V., & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, *30*(1), 137–158. https://doi.org/10.1080/0907676X.2021.1889005

Richardson, J., & Flowerdew, J. E. (2017). Introduction. In J. Richardson & J. E. Flowerdew (Eds.), *The Routledge Handbook of Critical Discourse Studies* (1st ed., pp. 1–10). Routledge.

Rossi, C., & Chevrot, J.-P. (2019). Uses and perceptions of machine translation at the European Commission. *The Journal of Specialised Translation (JoSTrans)*, *31*.

Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu’s concepts. *The Journal of Specialised Translation (JoSTrans)*, *31*, 201–216.

Sakamoto, A. (2021). The value of translation in the era of automation: An examination of threats. In R. Desjardins, C. Larsonneur, & P. Lacour (Eds.), *When Translation Goes Digital: Case Studies and Critical Reflections* (pp. 231–255). Springer International Publishing. https://doi.org/10.1007/978-3-030-51761-8_10

Sakamoto, A., & Yamada, M. (2020). Social groups in machine translationpost-editing: A SCOT analysis. *Translation Spaces*, *9*(1), 78–97. https://doi.org/10.1075/ts.00022.sak

Stewart, M. (2015). The language of praise and criticism in a student evaluation survey. *Studies in Educational Evaluation*, *45*, 1–9. https://doi.org/10.1016/j.stueduc.2015.01.004

Taylor, C., & Marchi, A. (2018). *Corpus Approaches to Discourse: A Critical Review*. Routledge.

Vidal, S. A., Badia, T., & Oliver, A. (2020). Post-editing for professional translators: Cheer or fear? *Revista Tradumatica*, *18*. https://doi.org/10.5565/REV/TRADUMATICA.275

Vieira, L. N. (2020). Automation anxiety and translators. *Translation Studies*, *13*(1), 1–21. https://doi.org/10.1080/14781700.2018.1543613

Yang, Y., Wei, X., Li, P., & Zhai, X. (2023). Assessing the effectiveness of machine translation in the Chinese EFL writing context: A replication of Lee (2020). *ReCALL*, *35*(2), 211–224. https://doi.org/10.1017/S0958344023000022

Zappatore, M., & Ruggieri, G. (2024). Adopting machine translation in the healthcare sector: A methodological multi-criteria review. *Computer Speech & Language*, *84*. https://doi.org/10.1016/j.csl.2023.101582

Zhang, M. (2023). Implicit cross-lingual word embedding alignment for reference-free machine translation evaluation. *IEEE Access*, *11*, 32241. https://doi.org/10.1109/ACCESS.2023.3260835

Zirpoli, C. T. (2023). *Generative artificial intelligence and copyright law* (LSB10922; CRS Legal Sidebar, pp. 1–6). Congressional Research Service. https://crsreports.congress.gov/product/pdf/LSB/LSB10922