JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 06/06 pp88−100
Volume 5, Number 1, March 2024
DOI: 10.23919/JSC.2024.0003

Multifaceted Disparities Associated with Translator Earnings: A Quantitative Study of Upwork Profiles

Tyler Horan*

Abstract: This study aims to quantitatively explore the multifaceted determinants that influence earnings in the translation industry. Using a dataset comprising 45 000 translator profiles, the study focusses on delineating disparities correlated with demographic variables such as age, gender, home country wealth, and language pairs. The study uses a random forest regression model to delineate the complex interaction between gender, the economic standing of a translator's domicile country, age, and linguistic proficiency, as they relate to earnings. Our findings substantiate and, in many ways, extend existing qualitative and anecdotal evidence that has shaped the discourse in this sector. The rigorous empirical framework employed here can be replicated or adapted to study other sectors within the gig economy, thus contributing to a more comprehensive understanding of labour dynamics in the digital age.

Key words: translation; online employment; gig economy; intercultural communication

1 Introduction

The role of translators goes beyond mere linguistic conversion to act as cultural mediators, contributing to the diversification of literary landscapes and facilitating intercultural communication[1, 2]. With the advent of digital platforms such as Upwork, the translation industry has seen significant growth, yet research on compensation structures and wage disparities within this digital landscape remains scarce[3]. Investigating these economic imbalances is crucial, as they could perpetuate existing power dynamics and reinforce dominant cultural narratives[4].

This investigation aims to fill the existing research gap by conducting an exhaustive analysis of the earnings and rates among translators on the Upwork platform. Using a dataset of 45 000 translator profiles, the focus is on identifying disparities related to demographic variables such as age, gender, educational attainment, and language pairs. The study also explores the variables that influence these earnings disparities, thus shedding light on how these factors can exacerbate existing cultural inequities[5–7].

Recognising the instrumental role of translation as a cultural broker, it becomes important to explore how income disparities within this sector could sustain existing power imbalances and dominant cultural narratives[8]. Through an examination of the Upwork earnings distribution, this research aims to provide a nuanced understanding of the remuneration dynamics among cultural intermediaries and the ramifications of such disparities on cultural inequities.

The results of this empirical investigation have substantial implications for both the translation industry and a wider understanding of the roles of translators as cultural facilitators[9]. By elucidating how wage inequalities contribute to perpetuating cultural imbalances, this study aims to galvanise efforts that are aimed at fostering equity within the translation industry. The ultimate objective is to facilitate a more egalitarian global landscape in which translators are compensated justly, allowing them to effectively fulfil their role as cultural brokers[10].

The study posits the following hypotheses for examination: H1: The economic standing of a

• Tyler Horan is with the Data Analytics and Computational Social Science Program, University of Massachusetts Amherst, Amherst, MA 01003, USA. E-mail: tjhoran@umass.edu.
∗ To whom correspondence should be addressed.
Manuscript received: 2023-09-29; revised: 2024-02-09; accepted: 2024-02-27

translator's domicile country exerts a significant influence on his earnings, notwithstanding the globalised nature of the workforce; H2: When controlled for languages, age, and country's economic status, gender still plays a consequential role in a translator's earnings, generally favouring males; H3: Age correlates positively with a translator's total earnings, countering the downward pressure on wages often observed on digital platforms; H4: The landscape of translation work is not monopolised by a solitary dominant country in terms of capital flows from core to periphery.

The rationale behind each hypothesis is deeply rooted in social science research and aims to fill gaps in our understanding of labour dynamics in digital platforms, specifically within the translation industry. H1 probes the influence of a country's economic position on translator earnings, a question that raises broader discussions of globalisation and income inequality[11]. It questions the extent to which platform labour markets are truly "borderless" and challenges the notion that digital platforms level the economic playing field across countries. H2 focusses on gender disparities, a persistent issue in labour economics, and seeks to quantify its impact within the translation sector[12]. This hypothesis aims to understand whether the digital economy replicates traditional gender imbalances or offers a new paradigm. H3 centres on the role of age, which serves as a proxy for experience and its correlation with earnings. Given that digital platforms are often seen to favour younger, tech-savvy individuals, this hypothesis challenges assumptions about age-based wage dynamics in online labour markets[13]. Finally, H4 investigates the distribution of earnings between countries, with the aim of dispelling or confirm theories about economic centralisation in international labour markets[11]. This hypothesis questions whether digital platforms democratise access to work or perpetuate existing imbalances. Therefore, each hypothesis not only addresses a specific gap in the literature, but also contributes to broader social science debates about labour, inequality, and the digital economy.

2 Background

Translation serves as a crucial vehicle for facilitating cross-cultural exchanges, positioning translators as vital intermediaries in the transfer of intellectual and informational resources across linguistic boundaries[10]. In the midst of the digital transformation of the labour markets, the translation sector has experienced significant growth, driven largely by the emergence of online freelancing platforms such as Upwork[6]. However, despite this growth, there is a noticeable dearth of empirical studies focussing on the distribution of earnings among translators active on these platforms[14].

Previous research efforts have underscored the pronounced inequalities that characterise the translation industry, including financial undercompensation, occupational precarity, and lack of formalised professional recognition[15]. More granular analyses have identified specific determinants of wage disparities, including, but not limited to, gender, language pairings, and educational attainment[16]. Such disparities have the potential to ossify pre-existing power hierarchies and reinforce the prevailing cultural narratives, thus having a profound impact on cultural equities.

To meaningfully engage with these issues, it becomes imperative to dissect the multifaceted determinants that affect earnings, along with elucidating the mechanisms through which these disparities could perpetuate existing cultural imbalances. Existing research has partially addressed these questions, revealing gender-based wage differences[17] and economic valuations associated with specific language pairings[18]. These insights suggest that structural cultural inequalities may influence remuneration within the translation industry.

Despite the aforementioned advances, a lacuna persists in research specifically probing the financial ecology of translators operating on platforms such as Upwork. Given the prominent role of translation as a cultural intermediary, understanding how financial disparities within this field can amplify existing power dynamics and mainstream cultural frameworks becomes critical. By rigorously examining the earnings distribution on Upwork, this study aims to contribute to an enriched understanding of remunerative policies that affect cultural intermediaries and the broader implications of such disparities on cultural inequalities. It seeks to enhance the sociology of translation through the examination of the cultural variation of remuneration across the sector.

2.1 Sociology of translation

Since the cultural turn in the 1990s, the translation field has seen a shift in the way that the profession and the practice of translation has unfolded. For example, Lefevere and Bassnett[19] demonstrated that translation was and continues to be a social practice embedded in social institutions. The sociological turn in the field of translation studies has emerged as a leading area of scholarly investigation, characterised by its incorporation of a wide range of themes and theoretical frameworks[20]. This intellectual trajectory has witnessed the application of various sociological theories to the domain of translation studies, with notable prominence given to Bourdieu's theory of social fields, Luhmann's social systems theory, and Callon and Latour's actor network theory[21–26].

2.2 Gender

The gender disparity within the translation sector represents an ongoing challenge, particularly manifested in the wage differences between male and female practitioners. Empirical investigations substantiate that, regardless of educational attainment or professional experience, translators are systematically remunerated less than their male counterparts[17]. Such inequities are observed more acutely within male-centric sectors, such as translation, where translators face formidable obstacles related to professional recognition, representation, and career progression opportunities[18].

Sociological research reveals that wage disparities are not limited to differences in experience or qualifications; women are paid less than men even for analogous tasks[27]. This remuneration gap is clearly amplified among female translators in recognition of their literary work[28]. Furthermore, the sectoral representation of women is disproportionally skewed, with an insubstantial proportion ascending to leadership roles within translation-centric organisations[18].

Studies in sociology and economics have illuminated how earnings are influenced by a multitude of factors, including gender, age, cohort, and period effects, often highlighting the complex interplay between these variables. For example, the gender wage gap is a well-documented phenomenon in labour economics, and research indicates that it is influenced by factors such as occupational segregation, gender norms, and family-related career interruptions[29]. Similarly, the effects of age on earnings have been extensively studied, often focussing on the life cycle of income and how it is affected by factors such as work experience, skill obsolescence, and retirement planning[30].

However, the gig economy, characterised by its online platforms, presents a distinct landscape from traditional labour markets. It operates distinctly from conventional employment structures, offering a more globalised workforce distribution and often a democratisation of work opportunities[31, 32]. This online activity however, is embedded within pre-existing relationships, where opportunities are often tied to social capital[33]. This novel environment challenges existing labour market theories, as traditional barriers to entry are lowered and the nature of work becomes more fluid. In this context, it is plausible to hypothesise that the predictors of earnings within online platforms like Upwork may differ significantly from those in traditional labour markets.

For example, in a conventional setting, age and experience are closely linked, often leading to higher earnings with increasing age. However, in the gig economy, where skills and deliverables are more project specific and less tied to tenure, this relationship may be less pronounced or even non-existent[34]. Similarly, gender dynamics can play out differently on online platforms. While gender discrimination is a persistent issue in traditional labour markets, the relative anonymity of online work and the focus on skills and deliverables could mitigate some of these biases[35]. Therefore, while acknowledging the rich insights from existing labour market research, it is also crucial to recognise the unique attributes of the gig economy and consider how these might influence earnings disparities differently.

Such wage and representational asymmetries have considerable repercussions on the translation industry. They reinforce established power structures, perpetuating the cycle of gender inequality[18]. This negative feedback loop has the consequence of constraining female translators' career trajectories, thereby further magnifying industry-wide gender disparities.

In summary, existing scholarship unambiguously demonstrates that gender inequality remains a systemic impediment within the translation industry, significantly impacting both remuneration and professional representation for female translators. These dynamics not only reinforce preexisting power hierarchies, but also have broader implications for cultural inequalities. This evidence base accentuates the need for additional research to clarify the contributing factors to these disparities, with the objective of formulating effective remedial strategies.

2.3 Nationality

Inequality based on national origin constitutes an ongoing challenge in the translation sector, characterised by notable remuneration and representational disparities affecting translators from specific countries[36]. Empirical analyses indicate that translators from developed countries pay higher wages for their services, whereas their counterparts from developing countries often receive lower remuneration, independent of educational, or experiential qualifications[18]. This difference is especially noticeable within specialised translation domains, such as legal or technical translation, where market control is predominantly exercised by developed countries[37].

The existing scholarly literature illustrates that translators from developed countries are often perceived as more reliable and competent, in contrast to the comparatively negative perception of translators from developing countries[37]. Such biased perceptions perpetuate existing power hierarchies, conferring greater market influence on developed countries[18]. This dynamic manifests itself in a strengthening cycle, in which translators from developed countries receive higher compensation and are therefore more likely to advance professionally, thus establishing disparities based on national origin within the industry[38].

These imbalances in remuneration and representation have far-reaching implications for the translation sector. Not only do they solidify existing cultural inequities, but also contribute to perpetuating nationality-based disparities, culminating in an industry biased towards specific cultural viewpoints and values[37]. Depending on the level of urbanization, opportunities can be limited and lead to forms of isolation on both a social and economic level[39].

In summation, existing research robustly substantiates that nationality-based inequality remains a systemic concern within the translation sector. Translators from particular countries face substantial inequities in both payment and professional representation, which has broader implications for cultural disparities. These findings underscore the imperative for more research to elucidate the underlying determinants of these inequalities and develop effective remedial strategies.

2.4 Age

Demographic analyses demonstrate that the preponderance of professional translators is clustered within the mid-30 to mid-50 age range with an average age of 41[40]. This demographic segment likely has acquired field-specific experience and expertise, contributing to elevated earnings. However, this demographic concentration invokes questions related to potential age discrimination and concomitant marginalisation of younger industry entrants, who may lack equitable access to professional opportunities and recognition.

Regarding the putative influence of age on translational quality, the existing literature is inconclusive. Some studies suggest that increasing age is positively correlated with superior translation quality, as older translators presumably capitalise on accumulated experience and expertise[38]. However, counter-evidence indicates a negligible age-dependent variation in translation quality[41]. These conflicting perspectives emphasise the need for more comprehensive research to elucidate the interplay between age and translational quality, as well as age-related impacts on remuneration and professional opportunities within the industry.

In summary, preliminary scholarly insights suggest age-related inequalities within the translation industry, with older people ostensibly enjoying remunerative and professional advantages. However, additional empirical research is warranted to unravel the complexities surrounding age-dependent disparities in both earnings and opportunities.

3 Method

3.1 Objective

The primary objective of this investigation was to rigorously examine the distribution of remuneration and rates among translators operating on the Upwork platform. Furthermore, the study sought to elucidate the relationships between these economic results and various demographic indicators, including age, sex, nationality, and linguistic ability. Data were obtained from all publicly accessible Upwork profiles through the Application Programming Interface (API) of the platform.

3.2 Data procurement and preprocessing

To gather the necessary dataset for this study, Upwork's public API was used to extract data from 45 000 translator profiles, geographically distributed in a multitude of nations. These profiles were all publicly available translator profiles on the platform. This comprehensive dataset encompassed various attributes, including the total salary of each translator, hourly rates, demographic age, gender, country of residence, and language proficiency.

3.2.1 Data cleaning

The initial dataset presented several challenges that required rigorous data cleaning. Incomplete or missing data were prevalent in fields such as age, gender, and total earnings. To address this, rows containing missing or anomalous data in key attributes were removed from the dataset. This step was crucial to ensure the integrity and reliability of subsequent analyses.

In addition, hourly rates were initially presented in various currencies, which could potentially introduce inconsistencies in the earnings data. To standardise this, all the currency values were converted to US dollars using the prevailing exchange rates at the time of data collection.

3.2.2 Feature engineering

In preparation for inferential analyses, several new features were engineered. One such feature was the Gross Domestic Product (GDP) corresponding to the translator's country of residence. This economic indicator was sourced from the data in Ref. [42] and merged into the dataset to serve as a proxy for the economic standing of each translator's home country.

Another engineered feature was age, calculated as the mean of the inferred minimum and maximum ages for each translator. This was done to provide a more nuanced representation of age, capturing its potential influence on earnings more effectively than using discrete age brackets.

3.2.3 Demographic inferencing through machine learning

For the inferential derivation of demographic attributes such as age and gender, a machine learning based facial recognition API provided by Inferdo (https://www.inferdo.com) was used. This API algorithmically ascertained age and gender based on profile images. Precision and recall scores for the effectiveness of the detection API have been evaluated and examined to be consistent with other major providers[43]. Although traditional demographic inferencing methods often rely on self-reported data, which may be subject to various biases or inaccuracies, the machine learning model offered a more objective and automated approach[44].

This method uses a Convolutional Neural Network (CNN) trained on a large dataset of annotated facial images. The model outputs a probabilistic range for age and a gender classification with associated confidence scores. These inferences were then incorporated into the dataset, providing a robust demographic layer to the existing economic and linguistic variables.

3.2.4 Outlier detection and elimination

To further refine the dataset, outlier detection algorithms were deployed. Specifically, the Z-score method was used to identify entries that deviated significantly from the mean in key attributes such as earnings and age. Entries with Z-scores that exceeded a predetermined threshold were considered outliers and subsequently removed from the dataset. This step was designed to mitigate the influence of extreme values that could skew the results and interpretations.

3.3 Statistical analysis

Subsequent to data acquisition and demographic inferencing, the study used a combination of statistical techniques to rigorously analyse the relationships between earnings, age, sex, nationality, and language. Descriptive statistics were used to provide a comprehensive summary of the dataset, giving insight into the central tendencies and variability of the variables under scrutiny.

For inferential analysis, a random forest regression model was used instead of traditional multiple regression. The random forest model is a non-parametric ensemble learning method that provides a robust means of capturing complex relationships among variables. This model offers the advantage of handling high-dimensional data and can account for non-linearities and interactions among variables, making it well suited for analysing the multifaceted determinants of earnings in the translation industry[45].

The random forest model was trained to predict total earnings based on a set of demographic and economic variables, including the GDP of the translator's home country, age, sex, and language proficiency. The performance of the model was evaluated using the value R^2 and the importance of the characteristics was extracted to quantify the impact of each variable on the earnings[46].

Using this machine learning approach, the study was able to delineate the relative importance of each demographic and economic factor while simultaneously accounting for their interdependencies, thus offering a nuanced understanding of earnings disparities in the translation sector.

The choice to employ a random forest model in our analysis is guided by its inherent strengths and suitability for our specific data characteristics and analytical goals. Random forest, as an ensemble learning method, integrates multiple decision trees to produce a more robust, accurate, and generalisable model. This method excels in handling large datasets with numerous features, including both numerical and categorical variables, without the necessity for extensive pre-processing or scaling. Its nonparametric nature allows it to capture complex and nonlinear relationships between features and the target variable, which might be missed by traditional linear models like Ordinary Least Squares (OLS) regression.

While traditional models like the Age-Period-Cohort (APC) are adept at elucidating specific types of structured relationships, they often require stringent assumptions about data distribution and relationships. In contrast, random forest can accommodate various data distributions and complex interactions without imposing rigid functional forms. This flexibility makes it particularly valuable in the exploratory stages of analysis or when the underlying data relationships are not well understood.

Furthermore, compared to other machine learning algorithms like Support Vector Machines (SVMs), random forest is generally more interpretable due to its decision tree based structure. It also provides insight into the importance of characteristics, helping to identify key drivers in the data. Although SVMs can be powerful for certain applications, they often require careful tuning and are less interpretable, which can be a disadvantage in social science research where understanding the influence of specific variables is crucial.

Certainly, random forest is not without its limitations, including potential overfitting with very noisy data and less effectiveness with very sparse data. However, in the context of our study, where our objective is to capture complex relationships in a rich dataset and value interpretability alongside predictive power, random forest offers a balanced and effective approach. As with any analytical method, it is vital to consider the model choice in relation to the specific research question, data characteristics, and the need for interpretability in the results.

3.4 Ethical adherence

The investigation was carried out strictly in accordance with the prevailing ethical standards, including those defined by the Institutional Review Board (IRB). All data obtained from Upwork profiles were de-identified to ensure anonymity and confidentiality, thereby adhering to the ethical directives established by the IRB and disciplinary standards[47]. Additionally, the use of the Inferdo facial detection API was carried out in full compliance with relevant privacy regulations and data protection statutes[48].

3.5 Concluding remark

To recapitulate, the methodological framework of this research included the extraction of data from Upwork profiles through the use of its public API, inferential demographic analysis through the Inferdo facial detection API, and the deployment of descriptive and inferential statistical paradigms to investigate relationships between earnings, age, sex, nationality, and language proficiency. The study was carried out in strict compliance with the applicable ethical guidelines, and measures were implemented to ensure the confidentiality of all data according to IRB protocols[47].

4 Result

4.1 Hypothesis H1: Influence of economic standing on translator earnings

A random forest regression model was used to investigate the impact of a country's economic standing, represented by its GDP, on translator earnings. As shown in Fig. 1, the model yielded a value of R^2 of 0.542, suggesting that approximately 54.2% of the variance in total earnings could be accounted for by the model predictors, which included GDP, age, and gender. Among these, GDP stood out as a significant predictor, with a feature importance score of 39.3. When GDP is removed from the model, R^2 drops to 0.170, suggesting that economic standing of the country is heavily predictive.

The results indicate that the economic standing of a country exerts a significant influence on translator earnings. This aligns with the existing literature that has consistently shown a wage gap between translators from developed and developing countries[36, 38]. Our findings further refine this understanding by quantifying the degree to which GDP impacts earnings, thus filling a gap in existing research that often lacks such quantitative assessments. All selected characteristics had a positive effect on the target variable, which was confirmed by running a Pearson correlation to detect the direction of the characteristics.

4.2 Hypothesis H2: Gender effect on translator earnings

In our random forest model, the variable (gender) was used to represent the gender, where a binary encoding was applied: 0 for women and 1 for men. The model revealed that this characteristic (gender) represented approximately 15.0% of the importance of the characteristic in predicting the total income of translators.

This is a striking figure considering that the model

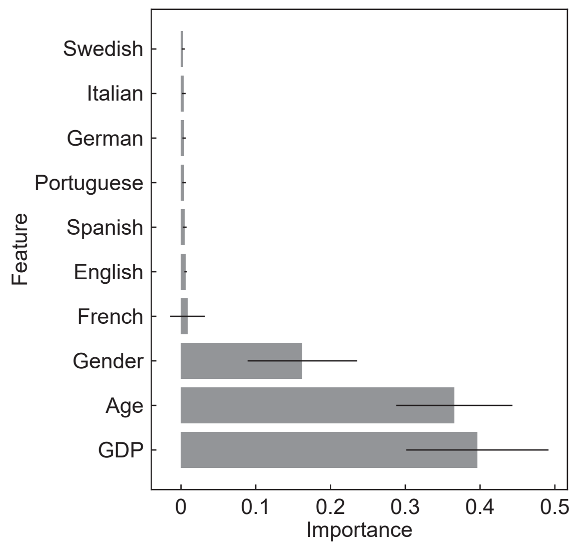

Fig. 1 Feature importances in predicting translator earnings (top 10).

also included other potentially influential variables such as GDP, age, and language proficiency. An importance of 15.0% characteristics does not simply suggest a peripheral influence; it indicates a substantial and statistically significant role that gender plays in the determination of translator earnings. This is noteworthy because it quantifies the extent to which gender, even in a presumably neutral online marketplace, can impact one's earning potential.

The implication is twofold. First, it provides empirical support for what has long been qualitatively observed: that gender disparity in earnings persists in various sectors, including the translation industry[17, 18]. Our results go beyond anecdotal or qualitative evidence to offer robust and quantitative support for this claim.

Second, the significant importance of the characteristics attributed to gender raises critical questions about the underlying mechanisms that perpetuate this disparity. Given that the model controls for other factors, including GDP and age, the significant impact of gender suggests that there may be systemic biases, either within the platform's algorithms or within client behaviour, that contribute to this observed gender wage gap.

It is worth mentioning that the gender variable, being binary, offers a simplified representation of a complex social construct. However, this simplified representation produces significant predictive power to determine earnings, highlighting the prevalence of gender-based income disparities.

By quantifying the impact of gender on earnings in this manner, our study adds an empirical layer to the existing literature, providing actionable insights that could form the basis for policy interventions aimed at addressing these disparities.

4.3 Hypothesis H3: Age and translator earnings

The importance of the characteristics attributed to the average age (age) in our random forest model was remarkably high, accounting for 36.1% of the predictive power of the model. This percentage is not only statistically significant but also practically meaningful, especially when other variables such as GDP, gender, and language proficiency are controlled in the model.

The prominence of age as a predictive characteristic suggests that age, often considered a proxy of experience, is the most significant factor in determining the earnings of a translator on on-line platforms. While age alone is not a perfect measure of experience or skill, its strong correlation with earnings suggests that accumulated expertise, professional networks, and possibly reputation, which often come with age, play a critical role in earning potential.

This finding aligns with, but also amplifies, the existing literature on the role of age and experience in the labour market[38, 40]. Whereas previous studies have shown mixed results regarding age as a determinant of earnings, our results offer a compelling argument in favour of age's significant influence, particularly in the translation sector. This has important implications for how we understand career progression and wage dynamics in the gig economy.

The high importance of age also raises questions about the inclusion of online platforms. If age is indeed acting as a proxy for experience, then these platforms may be less accessible as lucrative earning avenues for younger or less experienced translators. This could potentially exacerbate generational income disparities within the industry.

Moreover, the prominence of age in determining earnings challenges the notion that digital platforms are entirely democratising forces. Although they offer unprecedented access to global markets, these platforms may also perpetuate traditional hierarchies and disparities, such as those related to age and experience.

By quantitatively substantiating the role of age in determining earnings, our study provides nuanced insights that could influence both industry practises and policy formulations aimed at creating more equitable labour markets. It calls for a more complex understanding of how traditional factors, such as age and experience, intersect with the new opportunities and challenges presented by digital labour platforms.

4.4 Hypothesis H4: Distribution of translation work by country

In evaluating the distribution of total earnings by country, we sought to investigate the degree to which translation work is concentrated in a single nation. The data reveal a more distributed landscape, which lends empirical support to Hypothesis H4.

Contrary to the core-periphery theory that posits labour markets, particularly digital ones, as being monopolised by a few dominant or "core" countries, our findings suggest a much more nuanced scenario[49]. As shown in Fig. 2, the cumulative percentage of earnings does not reach 80% until roughly 30 countries. Although there is evidence that a few countries have a large part of the total earnings pie, the distribution is not winner-take-all. This is significant because it shows that earnings in digital labour markets are not primarily directed to individuals residing in a few countries.

However, given that when the GDP variable is omitted from the random forest model, the overall predictive rate drops significantly, the model suggests that while the distribution of countries is not consolidated into a single or small few, economic standing of the nation does play a significant impact in total earnings. This would indicate that capital flows are not spreading to less wealthy nations in terms of hourly earnings at similar rates to richer countries.

Our study thus adds a layer of complexity to the understanding of international labour markets in the age of digital platforms. It suggests that these platforms might be democratising forces to some extent, breaking down traditional geographical barriers and allowing for a more even distribution of work and earnings across countries.

However, it is important to note that this does not negate the need to examine labour practices on these platforms more closely. Although the distribution may be wide, it is crucial to investigate whether it is also equitable, especially in relation to factors such as GDP, gender, and age, among others.

By highlighting the dispersed nature of earnings

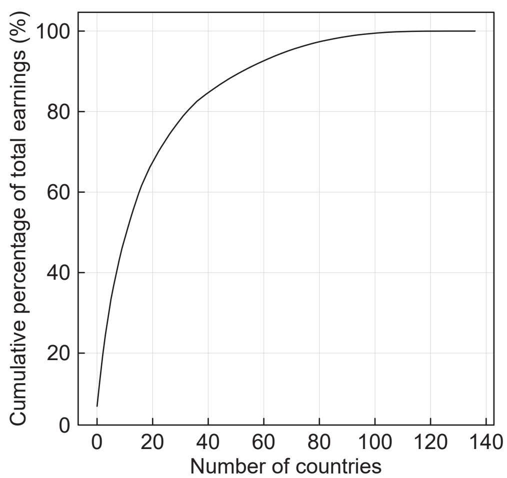

Fig. 2 Cumulative percentage of total earnings by country.

across countries, our findings offer a counternarrative to theories that emphasise economic centralisation and concentration. They invite policymakers and industry stakeholders to reconsider assumptions about how digital labour markets function and to explore mechanisms that could make these markets even more inclusive and equitable.

5 Discussion

5.1 Nationality

The role of a translator's nationality, often proxied by the GDP of their country of residence, emerges as a significant predictor in our random forest model. This aligns with existing scholarship that has previously explored the influence of nationality or country of origin on translators' earnings[36, 38].

What distinguishes our study from previous work is the use of a robust machine learning model to quantitatively dissect this relationship. Unlike traditional statistical methods, our model accounts for multiple variables simultaneously, allowing for a more nuanced understanding of how nationality intersects with other demographic and economic factors to influence earnings.

This nuanced understanding disrupts simplified narratives that posit translators from wealthier or core countries as universally out-earning those from peripheral or less wealthy countries[49]. Although GDP is a significant factor, its influence is moderated by other variables such as age, gender, and language proficiency. This suggests a more complex interplay of factors than is often acknowledged in discussions about wage inequality in international labour markets, including the translation sector.

The implications of these findings are multiple. From a policy perspective, understanding the nuanced role of GDP and nationality in determining earnings could inform initiatives aimed at reducing wage disparities between translators from different countries. Such insights could be particularly valuable in the context of online platforms, which have the potential to mitigate or exacerbate existing inequalities[36].

Furthermore, our findings contribute to ongoing debates about the "democratising" potential of digital labour platforms. Although these platforms offer unprecedented access to global markets, our results suggest that access alone is not sufficient to level the playing field. Structural factors, such as the economic standing of a translator's country of origin, continue to exert significant influence, corroborating earlier studies that question utopian narratives often associated with digital labour markets[38].

Using a machine learning model to quantify the influence of GDP and nationality, our study adds empirical weight to theoretical discussions about wage inequality in the translation industry. It also offers actionable insights that could serve as the basis for policy interventions aimed at creating more equitable labour markets both within and beyond the translation sector.

5.2 Gender

Despite the ostensibly democratising nature of online platform marketplaces, our findings indicate that gender disparities are not only prevalent but also persistent. The attraction of online platforms often lies in their promise of meritocracy, a space where skills and expertise are the primary determinants of earnings, supposedly devoid of the traditional biases that plague offline economies[50]. However, the data suggest otherwise.

The significance of gender in our random forest model stands as a statistical testament to the enduring nature of gender-based wage gaps. Even when considering variables such as age, GDP, and language proficiency, gender emerges as a significant predictor of earnings. This is particularly concerning, given that online platforms are often touted as avenues for equal opportunity and financial empowerment[51].

Our findings not only corroborate, but also extend existing qualitative research that has long argued for the existence of a gender wage gap in the translation industry[17, 18]. Although previous studies have provided valuable qualitative information on gender disparities[41], our research adds a robust quantitative dimension to these discussions. Quantified importance of gender in the determination of earnings provides empirical evidence that cannot be easily dismissed and could serve as the basis for policy interventions aimed at ameliorating these disparities.

Furthermore, the persistence of gender-based wage gaps on on-line platforms implicates the platforms themselves. It raises questions about whether these platforms, consciously or unconsciously, perpetuate existing social biases. For example, algorithmic recommendation systems could be biased in ways that disadvantage female translators or social biases could be replicated in client choices on these platforms.

The implications of these findings extend beyond the translation industry. They contribute to a growing body of literature that scrutinises the so-called “gig economy” through a critical lens, challenging the notion that digital platforms are neutral marketplaces. Rather, these platforms appear to mirror and perhaps even magnify existing social inequalities, including gender-based wage disparities.

This evidence base accentuates the need for additional research to clarify the mechanisms through which online platforms can perpetuate gender disparities. It also highlights the need for policy interventions that address these systemic biases to ensure equitable remuneration throughout the industry.

5.3 Age

The role of age in determining translator earnings has long been a topic of interest, although one that has often been discussed anecdotally or through qualitative methodologies. Our study substantially advances this discourse by furnishing empirical evidence that age is not merely a peripheral factor but a significant predictor of translator earnings. In our random forest model, the importance of characteristics attributed to age was markedly high, averaging 36.1%.

This empirical validation serves to substantiate the long-held notion that age, commonly viewed as a proxy of experience, plays a pivotal role in the translation industry. While previous demographic analyses have shown a concentration of professional translators within the mid-30 to mid-50 age range[40], our findings go beyond simple observation to provide quantitative confirmation that age is significantly correlated with earnings.

The high importance of age in the determination of earnings aligns with the broader research of the labour market, suggesting that seniors can accrue benefits from the accumulated experience and expertise, thus earning higher wages[52]. However, our findings extend these insights by quantifying the extent to which age impacts earnings, specifically within the context of the digital platform economy.

These findings have several far-reaching implications. First, they raise questions about the equitable distribution of opportunities across age groups. If age indeed acts as a proxy for experience, then younger translators might find it challenging to compete effectively on these platforms, potentially exacerbating generational income disparities within the industry.

Second, the strong correlation between age and earnings challenges the popular narrative that online platforms are fully democraticizing forces. While they offer unparalleled access to global markets, these platforms seem to mirror traditional labour market hierarchies, where experience, often measured by age, is highly valued.

Lastly, the quantifiable impact of age on earnings provides actionable insights for both industry stakeholders and policy makers. Understanding the nuances of how age impacts earnings could inform targeted interventions to support younger translators or those entering the field later in life, thus promoting a more equitable industry.

Using robust statistical methodologies to quantify the impact of age, our study not only corroborates existing qualitative observations, but also improves them. It adds a layer of empirical rigour to the discussions surrounding age-related wage disparities in the translation industry, offering valuable insights for future research and policy formulation.

5.4 Limitation

The generalisability of our findings in the translation sector is an important consideration in understanding labour dynamics in the digital age. Although our study provides valuable information on earnings disparities, it is essential to acknowledge limitations regarding the representativeness of our sample and the generalisability of the results to the entire translation industry.

It should be noted that our data were collected from Upwork profiles, which may not fully capture the diversity and dynamics of the entire translation sector. While Upwork is a widely used platform, it represents primarily a specific segment of the industry. Therefore, caution should be exercised when making broad generalisations based solely on the findings of this dataset.

To further enhance the generalisability of the findings, future research efforts should strive to incorporate a wider range of data sources and sampling techniques that encompass various platforms, institutions, and geographical regions within the translation industry. This would enable a more comprehensive understanding of the nuances and complexities that may exist in different parts of the sector.

Furthermore, it is important to recognise that income disparities can be influenced by a multitude of factors beyond those explored in this study. Factors such as educational background, work experience, language pairs or specialisations, and platform-specific dynamics are potential areas for further investigation. Examining these factors in conjunction with economic indicators can provide a more comprehensive understanding of the underlying dynamics and complexities that contribute to income disparities in the translation sector.

By taking these considerations into account and incorporating a more diverse range of data sources and variables, future research efforts can contribute to a more robust and comprehensive understanding of labour dynamics, not just in the translation sector but also across different sectors within the broader gig economy.

Additionally, it is crucial to emphasise that the analysis conducted using the random forest model, while insightful, does not imply causation. The predictive ability of the model helps identify relationships and patterns within the data, but it does not establish causal links between the features and the target variable (translator earnings). This distinction is important in the context of social science research, where understanding the causal mechanisms behind observed phenomena is often a primary objective.

The inability to infer causality from our model means that, while we can identify factors that are associated with higher or lower earnings in the translation sector, we cannot conclusively state that changes in these factors will lead to changes in earnings. For instance, while the model might find a strong association between certain languages and higher earnings, this does not necessarily mean that learning these languages will causally lead to higher income for all translators. There could be underlying factors, such as market demand, proficiency levels, or even unobserved variables such as networking skills, that play a significant role.

To establish causality, methodologies such as controlled experiments or causal inference models would be required. These approaches, however, often demand more stringent data requirements and experimental or quasiexperimental conditions that may not be feasible in all research contexts. Therefore, while the findings of our study are valuable for identifying potential areas of interest and generating hypotheses about the translation market, they should be interpreted as correlational rather than causal insights. Future research might build on these findings employing causal analysis methods to more thoroughly explore the dynamics of translator earnings.

6 Conclusion

This study embarked on an ambitious journey to quantitatively explore the multifaceted determinants that influence earnings in the translation industry. Using a random forest model, we probed the complex interplay between gender, the economic standing of a translator's domicile country, age, and language proficiency, as they relate to earnings.

Our findings substantiate and, in many ways, extend the existing qualitative and anecdotal evidence that has shaped the discourse in this sector. For example, although gender disparity in translator earnings has often been qualitatively discussed, our study provides empirical evidence that gender remains a significant predictor, even when other variables are controlled. This aligns with previous work but adds a quantitative rigour that can serve as a robust basis for future research and policy interventions.

Similarly, the role of a country's economic position, often mediated by GDP, has been qualitatively studied[36, 38]. Our quantitative analysis not only corroborates these findings, but also adds nuance by showing that, while GDP is a significant factor, it does not monopolise the earnings landscape. This challenges some of the traditional core-periphery theories that characterise international labour markets[49].

Perhaps one of the most striking revelations of our study is the pronounced role of age in determining translator earnings. Although age has been anecdotally linked to earnings, our study offers empirical validation, suggesting that experience, often correlated with age, is highly valued in the translation industry.

One of the critical takeaways from our study is the realisation that online work marketplaces like Upwork are far less meritocratic than they appear. Far from being platforms that democratise and level the playing field, these platforms often mirror, and sometimes even exacerbate existing social, economic, and demographic disparities. This is an important insight that goes against the optimistic narratives often associated with the gig economy.

Our study contributes significantly to the ongoing academic and policy discourse on wage disparities in the translation industry. By providing a quantitative basis to assess the impact of various factors on translator earnings, we not only validate previous qualitative work, but also open new avenues for future research. The empirical framework employed here can be replicated or adapted to study other sectors within the gig economy, thus contributing to a more comprehensive understanding of labour dynamics in the digital age.

References

[1] I. H. Senstad, Literary diversity through translation: The role of independent cultural mediators in the selection of Cuban literature in Norway, in *Literatura Latinoamericana Mundial*, G. Guerrero, J. J. Locane, B. Loy, and G. Müller, eds. Berlin, Germany: De Gruyter, 2020, pp. 97–122.

[2] D. La Penna and S. Sullam, Introduction: Literary exchanges between the Italian and anglo-american publishing markets: Readers, translators, mediators (1945–1970), *Italianist*, vol. 41, no. 3, pp. 339–347, 2021.

[3] E. Alangari, Cultural mediation in tourism translation: Saudi Arabia as a case study, *Arab World Engl. J. Transl. Lit. Stud.*, vol. 6, no. 3, pp. 82–98, 2022.

[4] S. A. Shirinzadeh and T. S. T. Mahadi, Translators as cultural mediators in transmitting cultural differences, *Procedia Soc. Behav. Sci.*, vol. 208, pp. 167–174, 2015.

[5] F. Malik, R. Heeks, S. Masiero, and B. Nicholson, Digital platform labour in Pakistan: Institutional voids and solidarity networks, *Information Technology & People*, vol. 34, no. 7, pp. 1819–1839, 2021.

[6] J. Horton, W. R. Kerr, and C. Stanton, Digital labor markets and global talent flows, https://www.nber.org/papers/w23398, 2017.

[7] E. Foong, N. Vincent, B. Hecht, and Gerber E. M, Women (still) ask for less: Gender differences in hourly rate in an online labor marketplace, *Proc. ACM Hum. Comput. Interact.*, vol. 2, no. CSCW, pp. 1–21, 2018.

[8] S. Simon, *Gender in Translation*. London, UK: Routledge, 1996.

[9] M. Tymoczko, *Enlarging Translation, Empowering Translators*. London, UK: Routledge, 2014.

[10] M. Baker, *In Other Words: A Coursebook on Translation*. London, UK: Routledge, 2018.

[11] S. Tashiro and S. Choi, Labor market outcomes under digital platform business models in the sharing economy: The case of the taxi services industry, *Bus. Econ.*, vol. 56, no. 4, pp. 240–251, 2021.

[12] C. Abidin, T. Barbetta, and J. Lee, Influencers, brands, and pivots in the time of COVID-19: A look at Australian, Japanese, and Korean issues, *M/C Journal*, doi: https://doi.org/10.5204/mcj.2729.

[13] E. Kocher, Reshaping the legal categories of work: Digital labor platforms at the borders of labor law, *Weizenbaum Journal of the Digital Society*, vol. 1, no. 1, p. w1.1.2, 2021.

[14] K. Kuhn, J. Meijerink, and A. E. Keegan, Human resource management and the gig economy: Challenges and opportunities at the intersection between organizational HR decision-makers and digital labor platforms, http://hdl.handle.net/10197/24617, 2021.

[15] J. Moorkens, D. Lewis, W. Reijers, E. Vanmassenhove, and A. Way, Translation resources and translator disempowerment, presented at the International Conference on Language Resources and Evaluation, Portorož, Slovenia, 2016.

[16] J. Milton, The figure of the factory translator: University and professional domains in the translation profession, https://www.mendeley.com/catalogue/b43a5e74-b92e-3635-b92a-f65af77d2dae/, 2004.

[17] H. Låndqvist, Gender equality and/or inequality? Female and male translators in a Swedish digital encyclopaedia of translators, *STRIDON: Journal of Studies in Translation and Interpreting*, vol. 3, no. 1, pp. 93–114, 2023.

[18] R. Sela-Sheffy, Elite and non-elite translator manpower: The non-professionalised culture in the translation field in Israel, *The Journal of Specialised Translation*, vol. 25, pp. 54–73, 2016.

[19] A. Lefevere and S. Bassnett, Introduction: Proust's grandmother and the thousand and one nights: The 'cultural turn' in translation studies, in *Translation, History, and Culture*, S. Bassnett and A. Lefevere, eds. London, UK: Cassell, 1990, pp. 1–13.

[20] M. Baker, *Critical Readings in Translation Studies*. New York, NY, USA: Routledge, 2010.

[21] M. Inghilleri, The sociology of bourdieu and the construction of the 'object' in translation and interpreting studies, *The Translator*, vol. 11, no. 2, pp. 125–145, 2005.

[22] H. Buzelin, Unexpected allies, *The Translator*, vol. 11, no. 2, pp. 193–218, 2005.

[23] S. Tyulenev, *Applying Luhmann to Translation Studies: Translation in Society*. New York, NY, USA: Routledge, 2011.

[24] S. Tyulenev, Systemics and lifeworld of translation, in *Translation and Philosophy*, L. Foran, ed. Frankfurt am Main, Germany: Peter Lang, 2012, pp. 139–155.

[25] S. Tyulenev, *Translation and the Westernization of Eighteenth-Century Russia: A Social-Systemic Perspective*. Berlin, Germany: Frank & Timme, 2012.

[26] S. Tyulenev, Translating in the public sphere: Birth pangs of a developing democracy in today's Russia, *South. Afr. Linguist. Appl. Lang. Stud.*, vol. 31, no. 4, pp. 469–479, 2013.

[27] D. Acemoglu and P. Restrepo, Tasks, automation, and the

rise in U. S. wage inequality, *Econometrica*, vol. 90, no. 5, pp. 1973–2016, 2022.
[28] F. Fernández, Assessing masculine domination in a cultural field: Women translators and book translation awards in Spain (1984–2012), *The Translator*, vol. 20, no. 2, pp. 162–177, 2014.
[29] F. D. Blau and L. M. Kahn, The gender wage gap: Extent, trends, and explanations, *Journal of Economic Literature*, vol. 55, no. 3, pp. 789–865, 2017.
[30] J. J. Heckman, A life-cycle model of earnings, learning, and consumption, *J. Polit. Econ*, vol. 84, no. 4, pp. S9–S44, 1976.
[31] V. De Stefano, The rise of the “just-intime workforce”: On-demand work, crowdwork, and labor protection in the “gig-economy”, *Comparative Labor Law & Policy Journal*, vol. 37, no. 3, pp. 471–504, 2016.
[32] T. Horan and I. Bullen-Smith, Online ranking system effects on perceived fairness: Gender, income and education, *Journal of Social Media in Society*, vol. 12, no. 2, pp. 44–61, 2023.
[33] T. Horan, Meeting up together: Economic embeddedness of social capital in America, *Soc. Sci.*, vol. 11, no. 4, p. 158, 2022.
[34] A. Agrawal, J. Horton, N. Lacetera, and E. Lyons, Digitization and the contract labor market: A research agenda, in *Economic Analysis of the Digital Economy*, A. Goldfarb, S. M. Greenstein, and C. E. Tucker, eds. Chicago, IL, USA: University of Chicago Press, 2015, pp. 219–256.
[35] A. R. Barzilay and A. Ben-David, Platform inequality: Gender in the gig-economy, *Seton Hall Law Review*, vol. 47, no. 2, p. 393, 2017.
[36] J. Munday, *Introducing Translation Studies: Theories and Applications*. London, UK: Routledge, 2009.
[37] J. Heilbron, Towards a sociology of translation: Book translations as a cultural world system, *European Journal of Social Theory*, vol. 2, no. 4, pp. 429–444, 1999.
[38] E. Gentzler, *Contemporary Translation Theories*. London, UK: Routledge, 1993.
[39] T. Horan, The effects of urbanization and social media use on individuals’ perceived social isolation, *Online J. Commun. Medium. Technol.*, vol. 14, no. 1, p. e202411, 2024.
[40] Zippia, Translator demographics and statistics in the US, https://www.zippia.com/translator-jobs/demographics/#, 2023.
[41] K. K. Zethsen and I. Askehave, Talking translation: Is gender an issue? *Gend. Lang.*, vol. 7, no. 1, pp. 117–134, 2013.
[42] The World Bank, GDP (current USD), https://data.worldbank.org/indicator/NY.GDP.MKTP.CD, 2023.
[43] Z. Li, P. Y. Chen, and T. Y. Ho, Great score: Global robustness evaluation of adversarial perturbation using generative models, arXiv preprint arXiv: 2304.09875, 2023.
[44] M. S. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel, and J. Movellan, Machine learning methods for fully automatic recognition of facial expressions and facial actions, in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, The Hague, the Netherlands, 2005, pp. 592–597.
[45] L. Breiman, Random forests, *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[46] A. Liaw and M. Wiener, Classification and regression by random forest, *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[47] C. K. Varnhagen, M. Gushta, J. Daniels, T. C. Peters, N. Parmar, D. Law, R. Hirsch, B. S. Takach, and T. Johnson, How informed is online informed consent? *Ethics & Behavior*, vol. 15, no. 1, pp. 37–48, 2005.
[48] B. Mittelstadt, Ethics of the health-related Internet of Things: A narrative review, *Ethics Inf. Technol.*, vol. 19, no. 3, pp. 157–175, 2017.
[49] S. Sassen, *The Global City: New York, London, Tokyo*. Princeton, NJ, USA: Princeton University Press, 1991.
[50] S. C. Kuek, C. Paradi-Guilford, T. Fayomi, S. Imaizumi, P. Ipeirotis, P. Pina, and M. Singh, The global opportunity in online outsourcing, Washington, DC: World Bank Group, http://documents.worldbank.org/curated/en/138371468000900555/The-global-opportunity-in-online-outsourcing, 2015.
[51] A. J. Wood, M. Graham, V. Lehdonvirta, and I. Hjorth, Good gig, bad gig: Autonomy and algorithmic control in the global gig economy, *Work. Employ. Soc.*, vol. 33, no. 1, pp. 56–75, 2019.
[52] A. Klevmarken and J. M. Quigley, Age, experience, earnings, and investments in human capital, *J. Polit. Econ.*, vol. 84, no. 1, pp. 47–72, 1976.

Tyler Horan received the bachelor degree from Boston University, USA in 2006, and the PhD degree in sociology from The New School for Social Research, USA in 2018. He is currently a lecturer at the Data Analytics and Computational Social Science Program, University of Massachusetts, Amherst, USA. His research areas include social inequalities, social networks, and computational social science methodologies. He is a member of the American Sociological Association, Complex Systems Society, Association for the Advancement of Artificial Intelligence, ACM Special Interest Group on Computer-Human Interaction, and International Communication Association.