Utilizing Machine Learning Techniques for Classifying Translated and Non-Translated Corporate Annual Reports

Zhongliang Wang, Ming Liu & Kanglong Liu

View supplementary material

Published online: 10 Apr 2024.

Submit your article to this journal

View related articles

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Utilizing Machine Learning Techniques for Classifying Translated and Non-Translated Corporate Annual Reports

Zhongliang Wang [iD][a], Ming Liu [iD][b], and Kanglong Liu [iD][b]

[a]School of Foreign Languages and Commerce, Guangzhou Railway Polytechnic, Guangzhou, China; [b]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

ABSTRACT

Globalization has led to the widespread adoption of translated corporate annual reports in international markets. Nonetheless, it remains largely unexplored whether these translated documents fulfill the same function and communicate as effectively to international investors as their non-translated counterparts. Considering their significance to stakeholders, differentiating between these two types of reports is essential, yet research in this area is insufficient. This study seeks to bridge this gap by leveraging machine learning algorithms to classify corporate annual reports based on their translation status. By constructing corpora of comparable texts and employing thirteen syntactic complexity indices as features, we analyzed the reports using eight different algorithms: Naïve Bayes, Logistic Regression, Support Vector Machine, k-Nearest Neighbors, Neural Network, Random Forest, Gradient Boosting and Deep Learning. Additionally, ensemble models were created by combining the three most effective algorithms. The best-performing model in our study achieved an Area Under the Curve (AUC) of 99.3%. This innovative approach demonstrates the effectiveness of syntactic complexity indices in machine learning for classifying translational language in corporate reporting, contributing valuable insights to text classification and translational language research. Our findings offer critical implications for stakeholders in multilingual contexts, highlighting the need for further research in this field.

Introduction

Corporate annual reports are comprehensive documents that deliver an in-depth analysis of a corporation's financial performance, activities, plans, and strategies over the preceding year. They function as a pivotal communication medium, sharing information about a company's financial statements, corporate governance, and social responsibility activities with its shareholders,

CONTACT Kanglong Liu ✉ kl.liu@polyu.edu.hk 🖃 Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

investors, and other stakeholders (Ren and Lu 2021). The importance of narrative disclosures in annual reports has been underscored by the US Securities and Exchanges Commission (SEC 1998). These reports play a crucial role in aiding investors in their decision-making process (Bhatia 2010). They provide key insights through which investors and other stakeholders can evaluate a company's financial health, performance, and prospects. They aid investors in making informed decisions about purchasing, holding, or selling a company's shares and also assist regulators in monitoring compliance with financial reporting standards. The impetus behind homing in on corporate annual reports as the specific focal point for this research is the significant role these reports play in the global business ecosystem and the unique challenges they present in the context of translation studies (Wang, Liu, and Moratto 2023). Corporate annual reports are more than just financial documents; they are the principal mode of formal communication between a corporation and the global market (Beattie, McInnes, and Fearnley 2004). They serve as a critical interface for a myriad of stakeholders, including investors, analysts, customers, employees, and regulators, each with diverse linguistic and cultural backgrounds.

The readability and language functions of annual reports have received considerable scrutiny in academia due to their significance (Bhatia 2008; Clarke, Hrasky, and Tan 2009; Courtis and Hassan 2002; Garzone 2004; Li 2008; Liao 2021; Li, Luo, and Deng 2019; Ren and Lu 2021). The majority of readability studies on annual reports have employed classic formula-based readability measures to obtain readability scores and to conduct analyzes and comparisons. These formula-based readability measures have been critiqued for their simplicity, focusing primarily on word and sentence difficulty (Benjamin 2012; Crossley et al. 2022; Lu, Gamson, and Eckert 2014). Syntactic complexity, encompassing the diversity and degree of complexity of the syntax used in a document (Housen and Kuiken 2009; Pallotti 2015), represents an important and multi-faceted aspect of text complexity (Frantz, Starr, and Bailey 2015; Lu 2011). Previous studies on annual reports have primarily relied on formula-based readability measures that only use average sentence length to assess syntactic complexity, thereby overlooking other aspects of sentence complexity. Furthermore, most of the prior research has focused on original annual reports, underscoring the need for scrutiny of translated annual reports as a subset of company annual reports. As English serves as the lingua franca in the international business and finance sectors, corporations in non-English-speaking countries often release a second annual report in English to enhance the accessibility of their messages to international investors, thereby yielding economic benefits (Jeanjean, Lesage, and Stolowy 2010; Jeanjean, Stolowy, and Erkens 2010).

This study endeavors to address existing research gaps by employing syntactic complexity metrics as features for classifying both translated and non-

translated annual reports. Through the application of text classification techniques, the study aims to determine the efficacy of syntactic complexity as features in effectively distinguishing between these two categories. We utilized four dimensions (comprising 13 indices) of syntactic complexity as features and applied eight machine learning models, including Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Neural Network, Random Forest, Gradient Boosting and Deep Learning, to distinguish translated annual reports from non-translated ones. Translated texts often exhibit certain features known as "translation universals." Measures of syntactic complexity can help identify these features (Liu and Afzaal2021; Wang, Liu, and Moratto 2023; Xu and Liu 2023). Syntactic complexity measures, such as sentence length, subordination level, depth of syntactic trees, and the variety and frequency of syntactic structures, can serve as valuable features in machine learning models (Lin and Liang 2023). By quantifying these aspects, classifiers can be trained to recognize patterns and discrepancies that are indicative of translated or non-translated texts.

The rest of this paper is organized as follows: Section 2 offers a brief review of the related research on annual reports and text classification. Section 3 introduces the research questions, as well as the two corpora compiled and used in this study for obtaining the values of syntactic complexity features. Section 4 presents the results of the eight machine learning models. The discussion of results and conclusions are presented in Sections 5 and 6, respectively.

Related Work

Previous Studies on Annual Reports

In the accounting literature, scholars have paid attention to the readability of annual reports. It is defined as a text's capacity to be read with speed and ease (Schroeder and Gibson 1990). Regulated parties in finance and business support accountability and integrity, and have accepted these fundamental concepts of readability. For example, the US Securities and Exchange Commission (SEC 1998) published "Plain English Disclosure Guidelines" to encourage issuers to improve the effectiveness of their financial and business disclosures. Similarly, the Financial Reporting Council (2009) argued that enterprises should use plain language. These requirements increase the importance of information in financial disclosures by making them readable in order to increase the accuracy of analytical forecasts and ultimately benefit shareholder choices. Jones (1994) highlighted the importance of readability to the effectiveness of narratives, and Lehavy, Li, and Merkley (2011) found that the readability of corporate texts of communication impacts financial analysts.

Schroeder and Gibson (1990) compared the footnotes and chairman's statements, and found that the footnotes to be more complex and complicated in terms of words and syntax. They compared 40 samples of the MD&A section with the president's letter section and footnotes. The authors found that the MD&A narratives were significantly lower than the footnote narratives in the use of passive voice. They found no significant differences in word length, sentence length, or readability score. The MD&A had significantly longer word length and lower readability than the president's letter section. Clarke, Hrasky, and Tan (2009) compared corporate annual report narratives with local government annual report narratives in Australia. They found corporate annual reports to be significantly less readable than their local governmental counterparts, as reflected by the Flesch Reading Ease scores.

The relationship between annual report readability and various variables, such as company performance, stock price, company size, business strategies, and earnings management, have received significant attention from scholars. Smith and Taffler (1992) conducted a systematic investigation into the correlation between the measures of readability of the chairman's statement in the corporate annual report and financial performance, using three readability formulas: Lix, Flesch, and Cloze. They found a positive correlation between the readability of the chairman's statements and corporate profitability, financial gearing, and liquidity. Subramanian, Insley, and Blackwell (1993) echoed these findings and further revealed the presence of statistical variations between companies with good and poor performance in terms of annual report readability. However, no statistical difference was found in the use of jargon or modifiers between the good and poor performers. On the contrary, both Baker and Kare (1992) and Courtis (1995) reported inconsistent results. Rutherford (2003) examined the relationships between the readability of operating and financial reviews with different indices of corporate performance. The study found no significant correlation between them. Smith et al. (2006) investigated the relationship between the readability of the chairman's statement in annual reports and the financial performance of Malaysian firms, taking into account such factors as corporate size and board membership. Their results found little evidence to support the obfuscation hypothesis. Li (2008) investigated the correlation between the readability of annual reports and companies' performance and earnings using the Gunning Fog index to measure readability. He found that companies with poorer earnings tended to produce less readable and longer annual reports, whereas those companies whose annual reports were more readable had higher persistent earnings. Ajina, Laouiti, and Msolli (2016) investigated the connection between annual report readability and earnings management in the French stock market. They gathered annual reports from 163 listed French companies between

2010–2013. They used the Gunning Fog index to measure the readability of annual reports, and discovered a statistical and positive relationship between annual report readability and the amount of discretionary accounting adjustments. Businesses that manage their revenues tend to produce less readable and comprehensible annual reports. Moreover, significant correlations between the readability of annual reports and other variables, such as profitability and company size, were also identified. Their discovery may be helpful for investors or businesses looking to better understand how profit management may impact annual reports on the French stock market. Lim, Chalmers, and Hanlon (2018) discovered that company strategy significantly influences annual report readability, with prospector strategy-pursuing organizations producing more readable reports than companies with defender strategy. The study also found that the annual reports of companies of technology sectors and those with decentralized organizational structures are more difficult to understand. Habib and Hasan (2020) examined how a company's unique business strategy may be affected by the readability of narrative disclosures in its annual reports. The researchers evaluated the readability of annual reports through using the Flesch Reading Ease and Flesch-Kincaid Grade Level indices. They found that businesses with various strategies generate annual reports of varying degrees of readability. Specifically, companies that employ prospector-type business strategies provide less understandable annual reports than those who employ defender-type business strategies. However, the study also claimed that annual reports are often difficult to comprehend.

Some scholars have also explored bilingual annual reports. Courtis and Hassan (2002) examined the bilingual versions of annual reports in Hong Kong and Malaysia, which use English and Chinese, and English and Malaysian, respectively. In the study, the Flesch Reading Ease index was employed to gauge the readability of English annual reports, the Yang index for Chinese texts, and the Yunus index for Malaysian reports. As the scores from those indices cannot be compared directly, Courtis and Hassan (2002) connected the scores to their indices' prearranged benchmarks of text difficulty and investigated "the distributions of reading ease degree of difficulty' interpretations per formula" (p. 404). They found that the English annual reports were less readable than the Malaysian and Chinese versions. Studies on the readability of annual reports have mainly used formula-based readability indices. However, critics of these indices have argued that such formulas only involve simplistic measures of word and sentence difficulty (Benjamin 2012; Lu, Gamson, and Eckert 2014). Syntactic complexity is an essential, and multidimensional, component of text complexity (Frantz, Starr, and Bailey 2015; Lu 2011). Readability formulas only use the average sentence length to measure the syntactic aspect, thus entirely neglecting other measures of sentence

complexity. Therefore, there is a need for a systematic investigation of the syntactic complexity of annual reports.

Moreover, previous studies have primarily focused on original annual reports, disregarding the significance of translated annual reports as an important category of company annual reports. As English is a lingua franca in the international business and financial domain, corporations in non-English-speaking nations and regions often disclose a second annual report in English to enhance the accessibility of their messages to international investors, thus yielding economic benefits (Jeanjean, Lesage, and Stolowy 2010; Jeanjean, Stolowy, and Erkens 2010). Translated annual reports have a significant impact on the communication between companies and their external investors (Jeanjean, Lesage, and Stolowy 2010). Indeed, studies have shown that translated annual reports have different linguistic features from the originals, and that these differences may affect the effectiveness of communication (Huang and Wang 2020; Liao 2021). Therefore, a comprehensive and comparative investigation of both translated and non-translated annual reports' syntactic complexity features is crucial for a thorough understanding of their language functions. This study distinguishes itself from previous research on annual reports by undertaking a comprehensive and granular analysis of the nuanced differences in syntactic complexity between translated and non-translated annual reports. The study's innovative approach to syntactic analysis, at a fine-grained level, elucidates the intricate linguistic nuances that differentiate these financial documents, thereby contributing novel insights to the field.

Syntactic Complexity As Features

The concept of syntactic complexity, as defined by the diversity and intricacy of a text's syntax (Ortega 2003), plays a pivotal role in quantifying text readability and evaluating written work (Jagaiah, Olinghouse, and Kearns 2020; Lei and Shi 2023; Ortega 2015). It is a key construct tied to the production units of language and grammatical frameworks. Syntactic complexity finds its main application in the domain of language learning, particularly within reading and writing exercises. It acts as a mirror, reflecting a learner's capacity to construct complex sentences. The research interest lies not only in mapping the evolution of learners but also in discerning the myriad factors that could influence syntactic complexity. These factors encompass the genre of the text, the learner's mother tongue, their academic level, and educational history (Bulté and Housen 2014; Jagaiah, Olinghouse, and Kearns 2020; Lu and Ai 2015). In the study of language reading, syntactic complexity has been scrutinized for its vital role as an indicator of reading difficulty. It is perceived as an essential component in modifying teaching materials and reading resources (Jin, Lu, and Ni 2020; Lei and Shi 2023). Recently, the investigation of syntactic

complexity extends to translation studies, which frequently involves identifying the unique properties of translated and interpreted languages (Chen, Li, and Liu 2024; Liu et al. 2022; Liu and Afzaal 2021; Liu, Liu, and Lei 2022; Xu and Liu 2023). For example, Liu and Afzaal (2021) found that translated texts exhibit significantly reduced syntactic complexity, supporting the simplification theory. Similarly, Wang, Liu, and Moratto (2023) identified partial simplification in translated chairman's statements. Regarding the explicitation universal, Al-Jabr (2006) noted that translations often become more explicit through syntactical changes. Xu and Li (2021) linked this explicitation to the formality level in translations. These studies collectively highlight how translation affects syntactic complexity and explicitness.

Machine Learning and Classification Between Translated and Non-Translated Texts

Text classification is a pivotal technique in the domain of natural language processing, enabling the systematic categorization of textual data into predefined labels. This can vastly improve information retrieval and content organization (Manning, Raghavan, and Schütze 2008). Text classification encompasses a variety of tasks where the goal is to assign one or more categories to a piece of text. This can be done based on the content, style, sentiment, purpose, or other criteria. Some of the key tasks in text classification include sentiment analysis (e.g., Mouthami, Devi, and Bhaskaran 2013), topic categorization (e.g., Zhou, Li, and Liu 2009), spam detection (e.g., Lau et al. 2012), translationese identification (e.g., Liu et al. 2022), and intent detection (e.g., Balodis and Deksne 2019). By applying machine learning techniques, such as Naive Bayes classifiers or SVM, textual data can be effectively classified into predefined labels (Ilisei et al. 2010; Liu et al. 2022). In the specialized task of translationese identification, these models offer a robust approach to distinguishing between original and translated texts because they can discern complex and non-linear relationships between features and labels. Machine learning algorithms can be trained on extensive datasets of labeled texts, where each text is identified as translated or original. These classifiers can then leverage these labeled datasets to predict the status of new texts. Furthermore, machine learning allows for the incorporation of various statistical and linguistic features extracted from the text, thereby enhancing the accuracy and resilience of the classifiers.

Baroni and Bernardini (2006) utilized a machine learning approach to classify translated and non-translated Italian geopolitical journal articles. They employed word or part-of-speech n-grams as features and used the SVM as their machine learning algorithm. Their findings reveal that an ensemble of SVMs could achieve an accuracy rate of 86.7% in the classification task, reinforcing the notion that machine learning models can identify

linguistic variations between translated and non-translated texts. Ilisei et al. (2010) undertook a study on the universals of simplification in translated Spanish, using various machine learning algorithms to classify between translated and non-translated Spanish texts. They found that all eight classifiers showed improvement when the simplification features were included. Notably, the SVM exhibited the most significant improvement, increasing from 73.65% to 81.76% accuracy on the test set. Similarly, Volansky, Ordan, and Wintner (2015) applied a supervised machine learning classifier to test hypotheses of translation universals, finding that the use of mean word rank as a feature yielded the highest accuracy (77%) among the simplification features.

Hu and Kübler (2021) explored the unique characteristics of translated Chinese texts and how they differ from texts originally written in Chinese. They used machine learning techniques to investigate the nuanced relationship between translations and their source languages, as well as variations within the translations. The researchers utilized the SVM classifier and 10-fold cross-validation for their analysis. The segmentation, POS tagging, and parsing of the Chinese texts were performed using the Stanford CoreNLP toolkit and its default models. The study found that translated Chinese texts exhibit distinct features that set them apart from non-translated texts, constituting a unique variety of Chinese. The authors underscored the novelty of their approach, which provided valuable empirical results from a non-Indo-European target language and examined a typologically diverse group of source languages. Similarly, Liu et al. (2022) explored the potential of machine learning techniques to distinguish between original and translated Chinese texts. In their study, the authors employed seven entropy-based metrics and four machine learning models – SVM, Linear Discriminant Analysis, Random Forests, and Multilayer Perceptron – to classify a balanced Chinese comparable corpus. The study's findings suggest that the combination of Shannon's entropy information-theoretic indicators with machine learning techniques offers a unique and effective method for analyzing translation as a distinct communicative activity. The results support the hypothesis that translational language differs qualitatively from original language, with the SVM identified as the most effective model, reaching an AUC of 90.5%. This study provides valuable insights for corpus-based investigations of the translationese phenomenon within the field of translation studies.

Rubino, Lapshinova-Koltunski, and Van Genabith (2016) conducted a study to address the challenge of distinguishing between human translations and original texts. The research specifically aimed to discriminate between the outputs of novice and professional translators. The authors proposed a feature set inspired by Quality Estimation (QE) and information density, and assessed its effectiveness in differentiating between non-translated texts, professional translations, and translations by translator trainees. The authors used SVM and four types of features – surface features, surprisal features, complexity

features, and distortion features – for the classification task. The study found that the combination of all four feature types yielded the highest accuracy, with distortion and complexity sets following closely. However, surprisal features were ineffective in differentiating between professional and student translations. Although the research suggests a close proximity between the two types of translations according to the feature types, with accuracy rates slightly above the 50% baseline, the authors argue that the proposed feature set shows promise in distinguishing between translations and non-translated texts. However, they also acknowledge the need for further research to improve the accuracy of the classification task. Building on the theme of language differentiation, Habic, Semenov, and Pasiliao (2020) applied deep learning techniques to ascertain whether texts were authored by native or non-native speakers. They collected a variety of labeled datasets and devised a specialized deep neural network (DNN) tailored to this specific classification challenge. Their methodology resulted in a notable enhancement of accuracy, achieving a peak of 88.75%. This significant achievement not only underscores the efficacy of deep learning in the realm of language processing but also sets a new standard for subsequent research endeavors aiming to precisely detect nuanced differences in language use and translation.

While the majority of studies on classifying between translated and non-translated texts have employed supervised learning, a few studies have utilized unsupervised learning. For instance, Nisioi and Dinu (2013) used a clustering approach to identify translationese in the novels of Vladimir Nabokov, a bilingual author. They employed the hierarchical clustering method and found that function words – common words serving grammatical purposes rather than conveying meaning – were the most informative features for identifying translationese. Similarly, Rabinovich and Wintner (2015) used an unsupervised machine learning approach to identify translationese, and their findings suggest that their proposed method outperforms supervised methods in this task. According to the authors, their two-phase clustering approach is a reliable method for distinguishing original texts from translated counterparts in highly heterogeneous datasets. Natural Language Processing (NLP) models have also been used for text classification tasks. A salient example of this is FinBERT, a large language model specifically designed for extracting information from financial texts (Huang, Wang, and Yang 2023). FinBERT builds upon the BERT (Bidirectional Encoder Representations from Transformers) architecture, which has revolutionized the understanding of context in language by considering the bidirectional nature of text. According to their study, FinBERT has demonstrated superior performance in sentiment classification, outperforming both the Loughran and McDonald dictionary – a widely recognized lexicon tailored for financial context analysis – and other conventional machine

learning algorithms. In a similar vein, Dogra et al. (2021) explored advanced machine learning and statistical techniques, including NLP models, to analyze sustainability in banking stocks, a task that entails the classification of events based on their relevance and impact on stock performance. Their study demonstrates how the integration of NLP with event study methodologies can uncover latent patterns within financial news and disclosures that traditional quantitative analyzes might overlook.

Although the existing literature predominantly relies on SVMs for classification, Liu et al. (2022) argue that it is essential to consider alternative classifiers to comprehensively evaluate the performance of classification techniques from a machine learning perspective. An ensemble model is also needed to improve accuracy and reduce overfitting. As such, the present study will apply eight machine learning models – Naïve Bayes, Logistic Regression, SVM, kNN, Neural Networks, Random Forest, Gradient Boosting and Deep Learning – for the classification of translated and non-translated annual reports and create ensemble models integrating the most effective models among the eight machine learning models.

Methods

Research Questions

The present study aims to use machine learning techniques to examine syntactic complexity of annual reports of companies from the Chinese Mainland listed on the Hong Kong Stock Exchange and those of American companies listed on the American stock exchanges. For the text analysis, two comparable corpora have been compiled, including the annual reports of companies from Mainland China (listed on the Hong Kong Stock Exchange, HKEX), and US companies listed on American stock exchanges. The HKEX was chosen due to the region's sheer importance as a business and financial center, as well as the abundant linguistic resources the HKEX provides for Chinese and English translation studies because of its special historical status and geographic location (Huang and Wang 2020; Jeanjean, Lesage, and Stolowy 2010). Many Mainland Chinese enterprises in the international market create their annual reports in Chinese and then translate them into English (Wang 2014). The study aims to address the following research questions:

RQ1: Can machine learning techniques differentiate between translated English annual reports and non-translated English annual reports using syntactic complexity features?
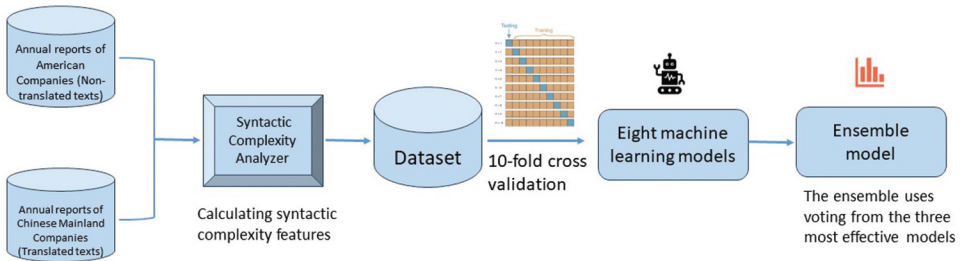
Figure 1. Overview of methodological approach employed in the study.

RQ2: If the answer to the first question is affirmative, which syntactic complexity features perform best in the classification?

RQ3: Do ensemble models outperform individual models in the classification task?

To quantify the syntactic complexity of corporate annual reports, we utilized the L2 Syntactic Complexity Analyzer developed by Lu (2010), which provided us with a comprehensive set of thirteen syntactic complexity measures. For the classification task, we first deployed eight diverse machine learning models, leveraging the syntactic complexity measures as predictive features while using the translation status as the target variable. We then assembled the three most effective algorithms to develop a final model. This approach was designed to harness the strengths of different algorithms, providing a robust assessment of the reports' syntactic nuances in relation to their translational status. Figure 1 illustrates the workflow of the methods in our study.

Data

The linguistic data used in this study consists of three corpora, namely the Corpus of American Annual Reports (USAR), the Corpus of Chinese Mainland Annual Reports (CMAR).

This study extracted data from companies listed on both US stock exchanges and HKEX. As of April 1, 2022, 5,295 stocks were listed on US stock exchanges and 2,251 on the HKEX. In order to ensure balanced representation in the corpora, the stock information was sorted by respective industry sectors. Given that the research focus was on the translated annual reports of Mainland Chinese companies listed on the HKEX, the distribution of stocks on the HKEX was used as a reference to design a sector-based scheme for company selection. Companies were chosen from ten distinct sectors, with a proportional number of companies selected from each sector (refer to Table 1 for details). The base locations of the companies were screened to ensure that 100 companies for USAR were based in the US, and 100 companies for CMAR were based in Mainland China. They

Table 1. The distribution of the companies by sectors in the corpora.

Sector	Number of companies
Finance	26
Consumer Services	10
Industrial Services	9
Electronic Technology	9
Consumer Non-Durables	9
Commercial Services	8
Technology Services	8
Process Industries	7
Producer Manufacturing	7
Health Technology	7
Total	100

Table 2. Description of USAR and CMAR corpora and their sub-corpora.

Corpus	Sub-corpus	Section of annual reports	Number of texts	Total tokens	Year	Average tokens
USAR	US-C	Chairman's statements	100	183,851	2020	1,839
	US-M	MD&A	100	1,375,585	2020	13,756
	US-N	NCFS	100	1,927,131	2020	19,271
CMAR	CM-C	Chairman's statements	100	154,965	2020	1,550
	CM-M	MD&A	100	680,926	2020	6,809
	CM-N	NCFS	100	1,992,268	2020	19,923

annual reports were converted from PDF format to plain text, and subsequently saved as text files for computational analysis. Non-textual content, such as graphs and images, was excluded. Sections such as the chairman's statements, Management Discussion and Analysis (MD&A), and Notes to Consolidated Financial Statements (NCFS) were isolated into separate text files. The USAR and CMAR were further subdivided into three sub-corpora each, namely US-C, US-M, US-N, CM-C, CM-M, and CM-N. Detailed information regarding the annual reports corpora and their sub-corpora can be found in Table 2.

Data Analysis

A machine learning approach was applied in this study to evaluate the performance of several dimensions of syntactic complexity in the classification of translated and non-translated annual reports. The study used the syntactic complexity measures proposed by Lu (2010). Given that the fifth dimension of syntactic complexity is characterized by a singular feature, the research primarily focused on quantifying the classification performance of the remaining four dimensions. These dimensions encompassed: 1) Length of production unit; 2) Degree of subordination; 3) Degree of coordination; 4) Degree of phrasal sophistication (see Table 3). Length of production unit refers to the average length of spoken or written text units, which could be sentences, T-units (main clauses plus any subordinate clause attached to them), or clauses. A higher average length typically indicates more complex production.

Table 3. Syntactic complexity indices used in the study (from Lu and Ai 2015, 18).

Feature	Code	Definition
Dimension 1: Length of production unit		
Mean length of clause	MLC	# of words/# of clauses
Mean length of sentence	MLS	# of words/# of sentences
Mean length of T-unit	MLT	# of words/# of T-units
Dimension 2: Amount of subordination		
Clauses per T-unit	C/T	# of clauses/# of T-unit
Complex T-units per T-unit	CT/T	# of complex T-units/# of T-units
Dependent clauses per clause	DC/C	# of dependent clauses/# of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses/# of T-units
Dimension 3: Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/# of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per sentence	T/S	# of T-units/# of sentences
Dimension 4: Degree of phrasal sophistication		
Complex nominals per clause	CN/C	# of complex nominals/# of clauses
Complex nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/# of T-units

This is often measured in terms of words per sentence/T-unit/clause, with the understanding that longer units may contain more information and more complex syntactic structures. Amount of subordination measures the use of subordinate clauses within larger sentence structures. Subordinate clauses are clauses that provide additional information to the main clause but cannot stand alone as a complete sentence. They are often introduced by subordinating conjunctions such as "because," "although," "if," etc. A higher amount of subordination would be an indicator of greater syntactic complexity because it shows the writer's ability to embed clauses within clauses, creating more nuanced and detailed sentences. Amount of coordination refers to the syntactic construction that links clauses or phrases of equal syntactic importance with coordinating conjunctions like "and," "but," "or," "nor," "for," "so," and "yet." The amount of coordination is a measure of the frequency and complexity of these structures. In syntactic complexity analysis, a high amount of coordination can reflect the writer's ability to connect ideas and maintain structural parallelism, contributing to the overall complexity of the text. Degree of phrasal sophistication measures the complexity within phrases, including complex nominals and verb phrases. A higher degree of phrasal sophistication indicates that the writer can manipulate phrase structure to convey more precise and complex meanings. Each dimension was critically analyzed to gain a deeper understanding of its role in syntactic complexity and, consequently, its influence on the classification of texts as either translated or non-translated.

In this study, the L2 Syntactic Complexity Analyzer (Lu 2010) was utilized to generate scores for thirteen indices of syntactic complexity. This computational tool, specifically designed for assessing the syntactic complexity of written English, was selected for several compelling reasons. According to Lu and Ai (2015), its

advantages include its free availability, a comprehensive suite of syntactic complexity measures, the ability to process large batches of files, and its demonstrated high reliability (p.18). The L2 Syntactic Complexity Analyzer has gained widespread recognition in the field of applied linguistics, as evidenced by its use in numerous studies. Researchers have employed it to investigate the syntactic complexity of texts produced by English learners (Lu and Ai 2015; Wu, Mauranen, and Lei 2020) and to address the differences between translation and non-translation (Liu and Afzaal 2021). The tool's robust design facilitates detailed analyzes of syntactic complexity, and its capacity to handle extensive datasets makes it an invaluable resource for linguistic research.

This study applied eight machine learning algorithms – Naïve Bayes, Logistic Regression, SVM, kNN, Neural Network, Random Forest, Gradient Boosting and Deep Learning – to classify annual reports as either translated or non-translated. For the kNN algorithm, we adhered to the common heuristic of setting k equal to the square root of the number of samples in the dataset. Given a dataset size of 200, we selected an approximate k value of 14. As for the SVM, we compared the performance of linear and RBF kernels, finding the linear kernel to yield superior results. To evaluate the models' performance and generalizability, a 10-fold cross-validation procedure was implemented. This method involves partitioning the dataset into 10 equal-sized subsets. The model is then trained on nine subsets and tested on the remaining one, a process repeated 10 times with each subset serving as the test set once. This validation approach promotes efficient data utilization, stable performance estimates, improved parameter tuning, and representative test outcomes, thus offering a more reliable measure of model performance than a single train-test split. The goal of binary classification is to categorize an ambiguous object into a specific class (Lei and Shi 2023). If the object falls into the specified category, it is labeled as "positive;" otherwise, it is labeled as "negative." For example, in a task to classify a fruit as an apple, a "positive" outcome would indicate that the model identifies it as an apple, while a "negative" outcome would suggest the opposite. These performance measures allow for a comprehensive evaluation of the machine learning models employed in this study. Model performance was quantified using several metrics: AUC, Accuracy, F1 score, Precision, and Recall. These metrics provide a comprehensive evaluation of the models' classification performance (Hand 2012).

Results

Figures 2, 3 and 4 depict the syntactic complexity measures differentiating between the two sub-corpora within chairman's statements (US-C and CM-C), MD&A (US-M and CM-M), and NCFS (US-N and CM-N), respectively. Figure 2 illustrates that CM-C is generally higher in the mean length of clauses, mean length of T-units, coordinate phrases per clause, coordinate phrases per T-unit, complex
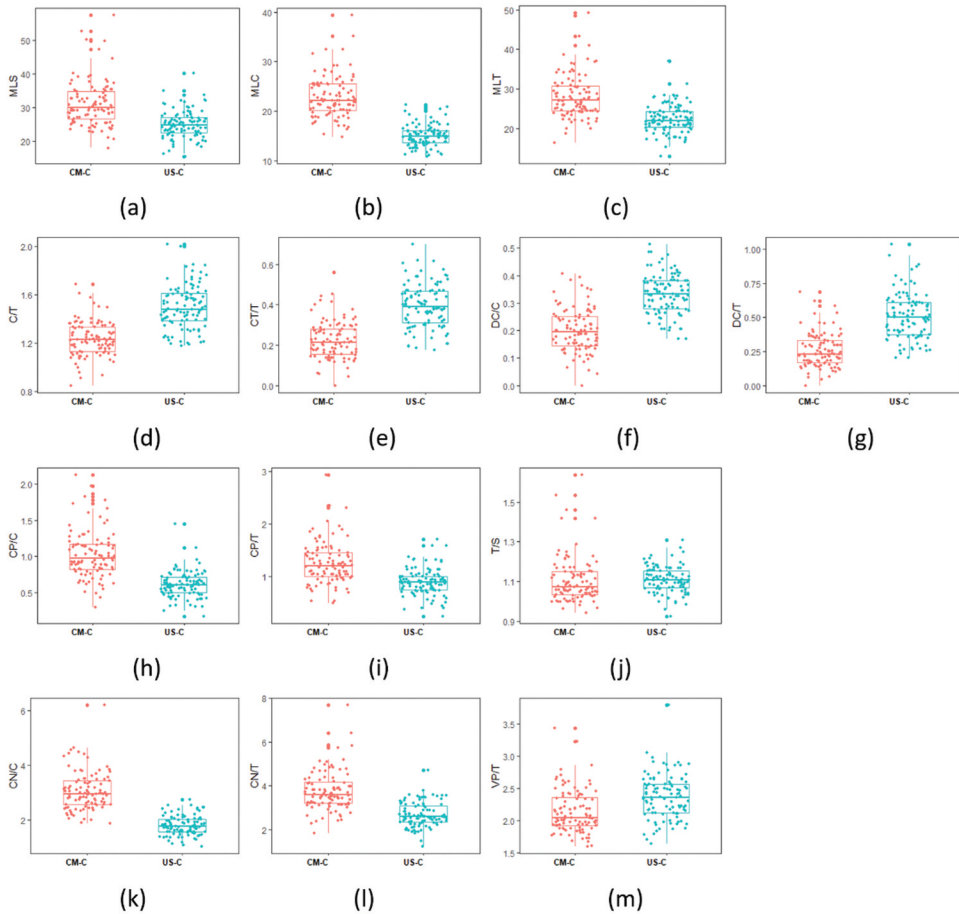
Figure 2. Syntactic complexity measures between the two sub-corpora of chairman's statements: (a) mean length of clause; (b) mean length of clause; (c) mean length of T-unit; (d) clauses per T-unit; (e) complex T-units per T-unit; (f) dependent clauses per clause; (g) dependent clauses per clause; (h) coordinate phrases per clause; (i) coordinate phrases per T-unit; (j) T-units per sentence; (k) complex nominals per clause; (l) complex nominals per T-unit; (m) verb phrases per T-unit.

nominals per clause, and complex nominals per T-unit but is lower in complex T-units per T-unit, dependent clauses per clause, T-units per sentence, and verb phrases per T-unit than US-C. Figure 3 shows that CM-M is generally higher in the mean length of clauses, mean length of T-units, coordinate phrases per clause, T-units per sentence, complex nominals per clause, and complex nominals per T-unit but is lower in complex T-units per T-unit, dependent clauses per clause, and verb phrases per T-unit than US-M. Figure 4 indicates that CM-N is generally higher in the mean length of clauses, mean length of T-units, clauses per T-unit, complex T-units per T-unit, dependent clauses per clause, coordinate phrases per T-unit, T-units per sentence, complex nominals per T-unit, and verb phrases per T-unit but is lower in the mean length of clauses than US-N. Descriptive statistics for the syntactic complexity measures within the chairman's statements, the
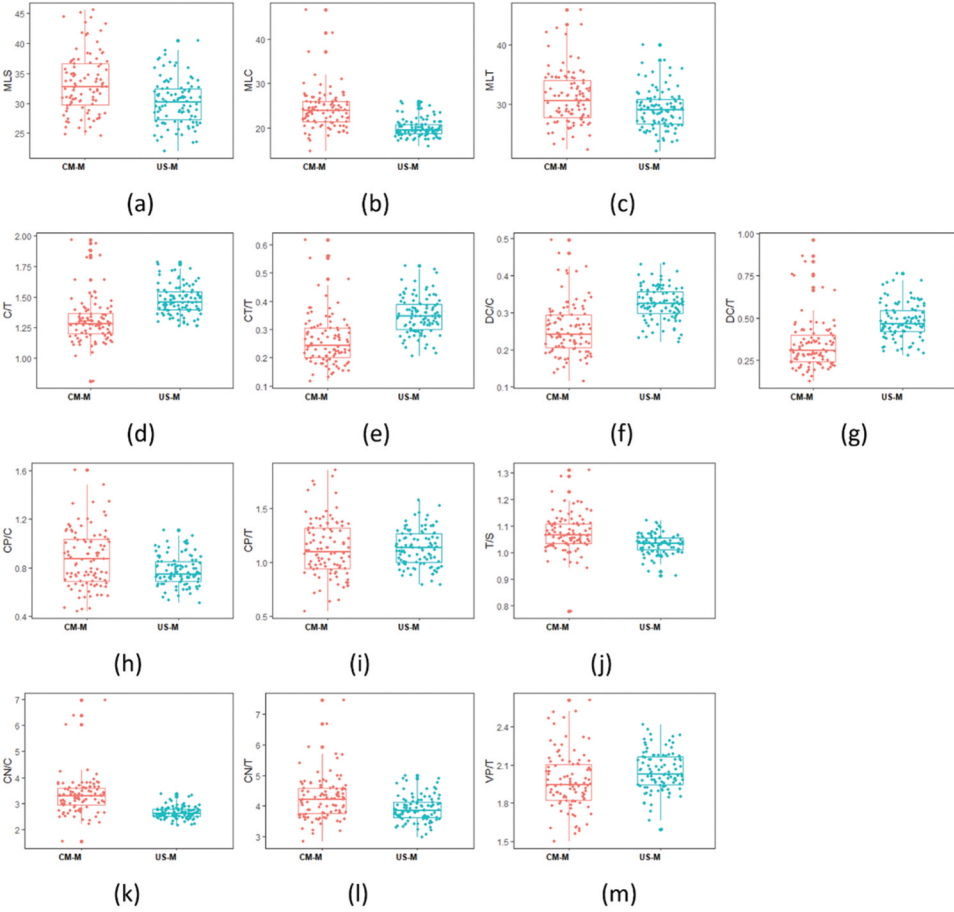
Figure 3. Syntactic complexity measures between the two sub-corpora of MD&A: (a) mean length of clause; (b) mean length of clause; (c) mean length of T-unit; (d) clauses per T-unit; (e) complex T-units per T-unit; (f) dependent clauses per clause; (g) dependent clauses per clause; (h) coordinate phrases per clause; (i) coordinate phrases per T-unit; (j) T-units per sentence; (k) complex nominals per clause; (l) complex nominals per T-unit; (m) verb phrases per T-unit.

Management Discussion and Analysis (MD&A), and Notes to the Consolidated Financial Statements (NCFS) across the two corpora are presented in Tables 4, 5 , and 6, respectively. Table 4 reveals that CM-C exhibits higher standard deviations than US-C in most syntactic complexity features. Table 5 indicates that CM-M is more dispersed than US-M across all 13 features. Table 6 shows that CM-N has lower standard deviations than US-N in seven features. These tables provide a comprehensive overview of the syntactic complexity observed within each type of corporate documentation and highlight the differences between the two corpora.

Table 7 presents the performance results of the machine learning models applied to chairman's statements. In the context of measuring the length of production units within chairman's statements, Logistic Regression outperforms
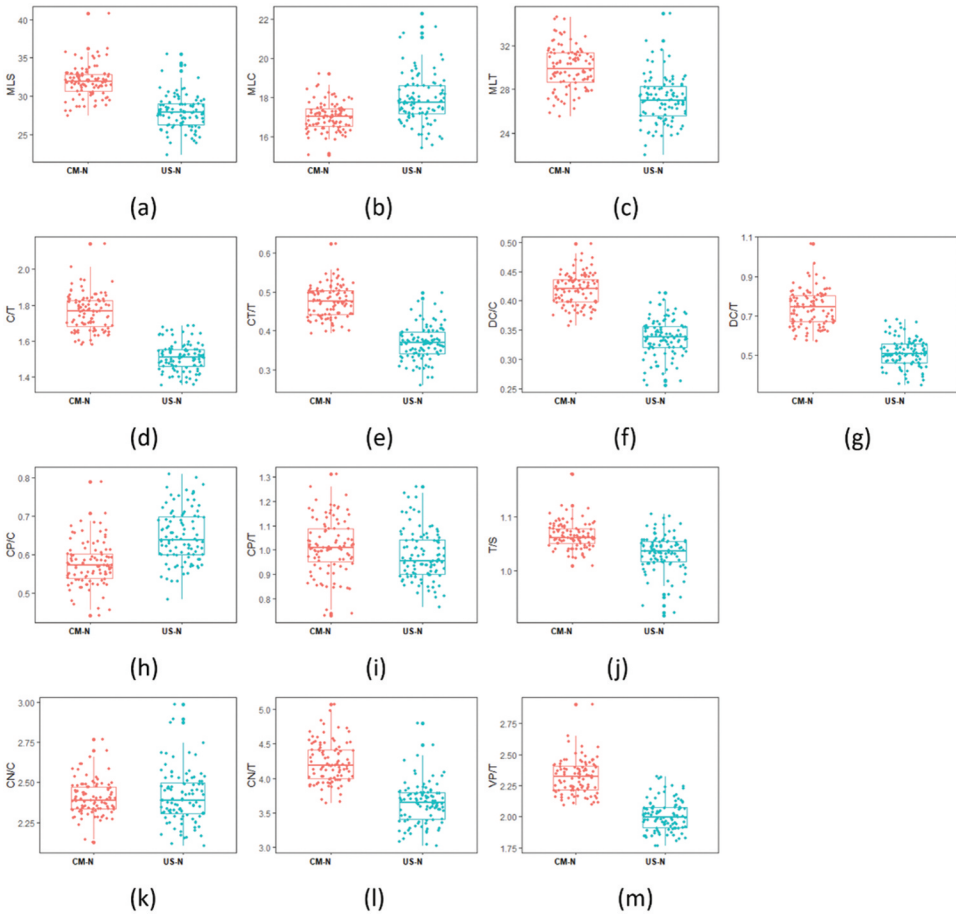
Figure 4. Syntactic complexity measures between the two sub-corpora of NCFS: (a) mean length of clause; (b) mean length of clause; (c) mean length of T-unit; (d) clauses per T-unit; (e) complex T-units per T-unit; (f) dependent clauses per clause; (g) dependent clauses per clause; (h) coordinate phrases per clause; (i) coordinate phrases per T-unit; (j) T-units per sentence; (k) complex nominals per clause; (l) complex nominals per T-unit; (m) verb phrases per T-unit.

the other models, achieving an AUC of 96.88%. The SVM, Neural Network, and kNN models follow closely, each achieving an AUC above 95%. Although Deep Learning ranks lowest among the models, it still attains an AUC exceeding 91%. The high AUC values suggest that the length of production units serves as a highly discriminative feature. A comparison of means reveals that the CM-C sub-corpus typically has longer production units than the US-C sub-corpus. In terms of coordination within chairman's statements, SVM ranks highest among the models, achieving an AUC of 92.11%, with the Neural Network model trailing closely at 90.5% AUC. For the dimension of subordination within chairman's statements, Logistic Regression again stands out, achieving the highest AUC of 88.82% among the eight models. Finally, when assessing phrasal sophistication, Deep learning tops the list with an AUC of 98.1%. Logistic Regression, SVM, Neural Network,

Table 4. Descriptive statistics for the syntactic complexity measures of chairman's statements.

Measure	US-C				CM-C			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Dimension 1: Length of production unit								
MLC	15.05	2.18	10.94	21.36	22.92	4.23	14.86	39.40
MLS	24.88	4.12	15.43	40.10	31.17	6.84	17.93	57.45
MLT	22.45	3.61	13.07	37.08	28.22	5.74	16.33	49.25
Dimension 2: Amount of subordination								
C/T	1.50	0.18	1.18	2.02	1.24	0.14	0.85	1.69
CT/T	0.40	0.11	0.18	0.70	0.23	0.10	0.00	0.56
DC/C	0.33	0.07	0.17	0.51	0.20	0.08	0.00	0.41
DC/T	0.51	0.17	0.20	1.04	0.26	0.13	0.00	0.69
Dimension 3: Amount of coordination								
CP/C	0.61	0.18	0.17	1.45	1.03	0.34	0.30	2.13
CP/T	0.90	0.26	0.25	1.71	1.25	0.40	0.50	2.93
T/S	1.11	0.07	0.93	1.31	1.11	0.12	0.94	1.64
Dimension 4: Degree of phrasal sophistication								
CN/C	1.79	0.35	1.03	2.75	3.06	0.71	1.87	6.20
CN/T	2.67	0.56	1.24	4.72	3.76	0.92	1.84	7.69
VP/T	2.37	0.34	1.64	3.79	2.14	0.33	1.60	3.44

Table 5. Descriptive statistics for the syntactic complexity measures of MD&A.

Measure	US-M				CM-M			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Dimension 1: Length of production unit								
MLC	19.88	2.00	15.77	25.94	24.09	4.51	14.78	46.63
MLS	30.16	3.78	22.05	40.48	33.31	5.00	24.61	45.63
MLT	29.20	3.32	22.22	40.04	31.11	4.54	22.43	45.92
Dimension 2: Amount of subordination								
C/T	1.47	0.12	1.26	1.78	1.31	0.19	0.82	1.96
CT/T	0.35	0.07	0.21	0.52	0.26	0.09	0.12	0.62
DC/C	0.32	0.05	0.22	0.43	0.26	0.07	0.12	0.50
DC/T	0.48	0.10	0.28	0.76	0.35	0.15	0.13	0.96
Dimension 3: Amount of coordination								
CP/C	0.77	0.13	0.51	1.11	0.88	0.24	0.44	1.61
CP/T	1.13	0.17	0.79	1.58	1.13	0.26	0.55	1.86
T/S	1.03	0.04	0.91	1.12	1.07	0.07	0.78	1.31
Dimension 4: Degree of phrasal sophistication								
CN/C	2.66	0.23	2.18	3.37	3.30	0.73	1.57	6.96
CN/T	3.90	0.42	2.98	5.00	4.26	0.72	2.86	7.46
VP/T	2.05	0.16	1.59	2.41	1.97	0.22	1.50	2.61

kNN, Random Forest, and Naïve Bayes models also perform strongly, each achieving an AUC above 95%, with respective values of 97.64%, 97.53%, 97.32%, 96.38%, 95.47%, and 95.15%. Gradient Boosting, although ranking last, still maintains a relatively high AUC of 94.52%. We developed an ensemble model combining Deep Learning, SVM, and Neural Network algorithms to analyze chairman's statements. The model utilizes the "Degree of Phrasal Sophistication" dimension, which includes three features: CN/C, CN/T, and VP/T. It demonstrated robust performance, achieving an AUC of 97.7%, an accuracy of 92%, an F1-score of 92.22%, a precision of 91.59%, and a recall of 94% (See Table 10).

Table 6. Descriptive statistics for the syntactic complexity measures of NCFS.

Measure	US-N				CM-N			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Dimension 1: Length of production unit								
MLC	17.91	1.26	15.43	22.28	17.02	0.71	15.11	19.23
MLS	27.91	2.29	22.36	35.51	31.88	2.05	27.46	40.80
MLT	27.03	2.15	22.02	34.95	29.95	2.01	25.51	34.63
Dimension 2: Amount of subordination								
C/T	1.51	0.08	1.36	1.69	1.76	0.10	1.58	2.14
CT/T	0.37	0.04	0.26	0.50	0.48	0.04	0.39	0.62
DC/C	0.34	0.03	0.26	0.41	0.42	0.03	0.36	0.50
DC/T	0.51	0. 07	0.35	0.68	0.74	0.09	0.57	1.07
Dimension 3: Amount of coordination								
CP/C	0.65	0.07	0.48	0.81	0.58	0.06	0.44	0.79
CP/T	0.98	0.10	0.76	1.26	1.01	0.11	0.73	1.31
T/S	1.03	0.03	0.92	1.10	1.06	0.02	1.01	1.18
Dimension 4: Degree of phrasal sophistication								
CN/C	2.41	0.17	2.11	2.99	2.40	0.11	2.13	2.77
CN/T	3.64	0.30	3.03	4.80	4.23	0.29	3.64	5.07
VP/T	2.01	0.12	1.77	2.33	2.32	0.14	2.09	2.90

Table 8 provides the performance results of the machine learning models applied to the Management Discussion and Analysis reports (MD&A). In terms of the length of production units within MD&A, the Neural Network model achieves the highest AUC among all models, reaching 89.25%. The Logistic Regression and SVM models follow, ranking second and third with AUC values of 88.24% and 88.06% respectively. Naïve Bayes, on the other hand, exhibits less optimal performance, achieving an AUC of 75.66%. When considering coordination within the MD&A narratives, the Neural Network model again ranks highest, achieving an AUC of 84.99%. The kNN and SVM models are closely behind, with AUC values of 83.19% and 81.24% respectively. For the dimension of subordination within MD&A narratives, the Gradient Boosting model surpasses the others, reaching an AUC of 86.21%. The Random Forest model follows closely with an AUC of 85.97%. In assessing phrasal sophistication, Random Forest tops the list with an AUC of 89.8%. The kNN, Neural Network, Gradient Boosting, and SVM models rank second, third, fourth, and fifth respectively, each achieving an AUC above 85%. We created an ensemble model that merges the analytical strengths of Logistic Regression, SVM, and Neural Network techniques for scrutinizing MD&As. This model employs the "Length of production units" dimension, encompassing three features: MLS, MLC, and MLT. The model's efficacy was good, with a 91.1% AUC, 83.5% accuracy, 84.1% F1-score, 83.28% precision, and 86% recall (See Table 10).

Table 9 shows the performance results of the machine learning models applied to the Notes to Consolidated Financial Statements (NCFS). In the context of the length of production units within the NCFS, Deep Learning has the highest AUC at 99.3%, the SVM model, Neural Network model and Logistic Regression closely follow, reaching an AUC of 99.05% and 99%, 98.92% respectively. Notably, the

Table 7. Results of machine learning models on chairman's statements.

Model	AUC	Accuracy	F1	Precision	Recall
Length of production units					
Logistic Regression	96.88%	91.00%	91.00%	91.07%	91.00%
SVM	96.82%	91.00%	91.00%	91.07%	91.00%
Neural Network	96.52%	89.50%	89.49%	89.60%	89.50%
kNN	95.26%	90.00%	90.00%	90.00%	90.00%
Gradient Boosting	94.27%	87.50%	87.49%	87.59%	87.50%
Random Forest	94.18%	91.50%	91.50%	91.50%	91.50%
Naïve Bayes	91.75%	86.00%	85.99%	86.06%	86.00%
Deep Learning	91.50%	84.50%	84.03%	87.71%	82.00%
Subordination					
Logistic Regression	88.82%	80.50%	80.48%	80.65%	80.50%
Neural Network	88.66%	79.50%	79.49%	79.57%	79.50%
Deep Learning	87.75%	77.50%	77.98%	76.81%	81.00%
SVM	87.51%	80.00%	79.97%	80.19%	80.00%
kNN	87.18%	80.50%	80.48%	80.65%	80.50%
Naïve Bayes	86.64%	80.50%	80.48%	80.65%	80.50%
Random Forest	86.22%	81.00%	81.00%	81.01%	81.00%
Gradient Boosting	84.75%	78.50%	78.50%	78.50%	78.50%
Coordination					
SVM	92.11%	84.50%	84.39%	85.53%	84.50%
Neural Network	90.50%	84.00%	83.99%	84.12%	84.00%
kNN	89.04%	82.00%	81.91%	82.64%	82.00%
Logistic Regression	87.95%	82.50%	82.48%	82.66%	82.50%
Deep Learning	87.50%	80.00%	80.65%	79.61%	83.00%
Random Forest	86.30%	83.00%	82.97%	83.21%	83.00%
Gradient Boosting	86.01%	82.00%	82.00%	82.00%	82.00%
Naïve Bayes	84.41%	78.50%	78.49%	78.57%	78.50%
Phrasal sophistication					
Deep Learning	98.10%	93.00%	93.42%	92.44%	95.00%
Logistic Regression	97.64%	91.50%	91.49%	91.60%	91.50%
SVM	97.53%	91.00%	91.00%	91.02%	91.00%
Neural Network	97.32%	92.50%	92.50%	92.61%	92.50%
kNN	96.38%	89.50%	89.49%	89.69%	89.50%
Random Forest	95.47%	89.50%	89.50%	89.50%	89.50%
Naïve Bayes	95.15%	89.50%	89.50%	89.50%	89.50%
Gradient Boosting	94.52%	90.00%	90.00%	90.02%	90.00%

AUC values of the top seven models exceed 98%, indicating high performance. Regarding accuracy, Logistic Regression leads with an impressive 96.5% accuracy rate. It is closely followed by SVM, Neural Networks, and Random Forest, which achieve accuracy rates of 96%, 95.5%, and 95.5%, respectively. For the dimension of subordination within the NCFS, the Gradient Boosting model outperforms the others, achieving an AUC of 98.58%. All eight models demonstrate strong performance, with AUC values exceeding 96%. In terms of coordination within the NCFS, the Neural Network model ranks highest, achieving an AUC of 98.81%. The kNN, SVM, and Logistic Regression models follow, with respective AUC values of 98.71%, 98.66%, and 98.44%. Notably, six out of the eight models have AUC values higher than 96%. Finally, when assessing phrasal sophistication, the SVM model ranks highest among the eight models, achieving an AUC of 97.61%. The Neural Network, kNN and Deep Learning models rank second, third and fourth respectively, each achieving an AUC above 97%. Remarkably, all eight models display strong performance in this dimension with AUC values exceeding 95%. We designed and compared two ensemble models to analyze NCFS. The first ensemble

Table 8. Results of machine learning models on MD&A.

Model	AUC	Accuracy	F1	Precision	Recall
Length of production units					
Neural Network	89.25%	83.50%	83.49%	83.58%	83.50%
Logistic Regression	88.24%	83.50%	83.45%	83.91%	83.50%
SVM	88.06%	83.50%	83.45%	83.91%	83.50%
kNN	87.31%	81.50%	81.48%	81.66%	81.50%
Random Forest	84.34%	80.00%	80.00%	80.01%	80.00%
Gradient Boosting	84.06%	80.00%	80.00%	80.01%	80.00%
Deep Learning	77.40%	73.00%	74.23%	74.42%	78.00%
Naïve Bayes	75.66%	72.00%	71.96%	72.14%	72.00%
Subordination					
Gradient Boosting	86.21%	81.00%	81.00%	81.01%	81.00%
Random Forest	85.97%	81.50%	81.50%	81.53%	81.50%
kNN	84.61%	80.00%	79.93%	80.44%	80.00%
Neural Network	84.01%	77.50%	77.34%	78.32%	77.50%
Deep Learning	83.50%	76.00%	76.69%	73.87%	81.00%
Logistic Regression	82.62%	78.00%	78.00%	78.01%	78.00%
Naïve Bayes	81.41%	76.50%	76.49%	76.57%	76.50%
SVM	69.45%	70.50%	70.50%	70.50%	70.50%
Coordination					
Neural Network	84.99%	78.50%	78.43%	78.85%	78.50%
kNN	83.19%	75.00%	74.91%	75.37%	75.00%
SVM	81.24%	78.50%	78.16%	80.40%	78.50%
Gradient Boosting	79.78%	74.00%	73.91%	74.35%	74.00%
Logistic Regression	78.90%	73.00%	72.73%	73.96%	73.00%
Random Forest	77.81%	74.00%	73.98%	74.09%	74.00%
Deep Learning	77.70%	68.50%	64.62%	66.06%	66.00%
Naïve Bayes	75.89%	71.50%	71.47%	71.61%	71.50%
Phrasal sophistication					
Random Forest	89.80%	80.50%	80.50%	80.53%	80.50%
kNN	88.20%	80.00%	79.84%	81.00%	80.00%
Neural Network	87.65%	82.50%	82.43%	83.06%	82.50%
Gradient Boosting	86.99%	79.50%	79.50%	79.50%	79.50%
SVM	85.13%	82.00%	81.82%	83.33%	82.00%
Logistic Regression	84.51%	82.50%	82.43%	83.06%	82.50%
Naïve Bayes	82.85%	74.50%	74.39%	74.92%	74.50%
Deep Learning	80.00%	75.00%	78.26%	69.23%	90.00%

model, which encompasses Deep Learning, SVM, and Neural Networks, was optimized based on the AUC metrics. The second ensemble model combined Logistic Regression, SVM, and Neural Network algorithms, with a primary focus on maximizing accuracy. Both ensemble models incorporated the "Length of Production Units" dimension, encompassing three key features: MLS, MLC, and MLT. Results show that the first ensemble model demonstrated exceptional results, achieving an AUC of 99.2%, accuracy of 95.5%, an F1-score of 95.42%, precision of 97.09%, and recall of 94%. Meanwhile, the second ensemble model also showed strong performance, with an AUC of 99.2%, an enhanced accuracy of 96%, an F1-score of 95.94%, precision of 97.18%, and a slightly better recall of 95%. Given the comparative analysis, the latter ensemble model, which yielded a higher accuracy and recall, was selected as the more suitable model for analyzing NCFS (See Table 10). This model's balance of high precision and recall, coupled with its marginally superior accuracy, suggests that it is the better-performing model for our specific application in analyzing NCFS.

Table 9. Results of machine learning models on NCFS.

Model	AUC	Accuracy	F1	Precision	Recall
Length of production units					
Deep Learning	99.30%	94.00%	92.88%	96.78%	92.00%
SVM	99.05%	96.00%	96.00%	96.02%	96.00%
Neural Network	99.00%	95.50%	95.50%	95.54%	95.50%
Logistic Regression	98.92%	96.50%	96.50%	96.54%	96.50%
Random Forest	98.60%	95.50%	95.50%	95.50%	95.50%
kNN	98.40%	94.00%	93.98%	94.44%	94.00%
Gradient Boosting	98.08%	93.50%	93.50%	93.50%	93.50%
Naïve Bayes	92.65%	83.00%	83.00%	83.01%	83.00%
Subordination					
Gradient Boosting	98.58%	93.00%	93.00%	93.00%	93.00%
kNN	97.94%	90.50%	90.47%	91.00%	90.50%
Random Forest	97.91%	89.50%	89.50%	89.54%	89.50%
SVM	97.88%	91.50%	91.49%	91.60%	91.50%
Logistic Regression	97.83%	91.50%	91.50%	91.50%	91.50%
Deep Learning	97.80%	87.00%	85.76%	94.64%	80.00%
Neural Network	97.56%	92.00%	92.00%	92.07%	92.00%
Naïve Bayes	96.29%	92.50%	92.50%	92.50%	92.50%
Coordination					
Neural Network	98.81%	97.00%	97.00%	97.02%	97.00%
kNN	98.76%	95.00%	95.00%	95.16%	95.00%
SVM	98.66%	97.00%	97.00%	97.02%	97.00%
Logistic Regression	98.44%	96.00%	96.00%	96.02%	96.00%
Gradient Boosting	97.13%	93.50%	93.50%	93.54%	93.50%
Random Forest	96.38%	92.00%	92.00%	92.02%	92.00%
Deep Learning	90.10%	83.00%	82.99%	88.04%	82.00%
Naïve Bayes	89.71%	83.00%	82.99%	83.05%	83.00%
Phrasal sophistication					
SVM	97.61%	90.00%	90.00%	90.06%	90.00%
Neural Network	97.58%	92.00%	92.00%	92.00%	92.00%
kNN	97.14%	91.50%	91.48%	91.84%	91.50%
Deep Learning	97.00%	90.00%	89.66%	92.05%	88.00%
Random Forest	96.77%	91.50%	91.50%	91.50%	91.50%
Logistic Regression	96.36%	91.50%	91.50%	91.50%	91.50%
Naïve Bayes	95.61%	87.50%	87.49%	87.59%	87.50%
Gradient Boosting	95.28%	92.00%	92.00%	92.07%	92.00%

Table 10. Results of ensemble models.

	Features	AUC	Accuracy	F1	Precision	Recall
Chairman's Statements	CN/C, CN/T, VP/T	97.70%	92.00%	92.22%	91.59%	94.00%
MD&As	MLS, MLC, MLT	91.10%	83.50%	84.10%	83.28%	86.00%
NCFS	MLS, MLC, MLT	99.20%	96.00%	95.94%	97.18%	95.00%

Discussion

Translation is a dual-faceted phenomenon, encompassing both a mental process within the translator's psyche and a cross-cultural social behavior (House 2014). The Hypothesis of Gravitational Pull, as proposed by Halverson (2003), offers a widely accepted explanation for this phenomenon by merging bilingual theory and cognitive linguistics (Liu and Afzaal2021). This hypothesis proposes that the language of translation is molded by three interconnected forces. The first force, known as the magnetism effect, would entice translators to mirror the conspicuous linguistic characteristics of the target language in

their translations. The second force, termed the gravitational pull effect of the source language, acts in opposition to the magnetism effect and represents a challenge the translator must navigate. The third and final force, the connection effect, arises from the frequent co-occurrence of translation equivalents in both the source and target languages. These three forces operate collectively to shape the language of translation (Halverson 2017). The magnetism effect has primarily been employed to elucidate unique features in translated texts at the lexical level. However, Liu and Afzaal(2021) and Wu et al. (2023) also suggest that it can elucidate unique syntactic features in translated texts. In our study, it has been observed that translated chairman's statements and MD&A narratives contain more coordinate phrases than their non-translated counterparts. This pattern is indicative of the magnetism effect, demonstrating how a translator tends to incorporate the prominent linguistic features of the target language into their translation, and may even utilize these features more abundantly than they appear in non-translated texts of the target language.

On the other hand, the complexity levels of annual reports are also shaped by societal factors. A notable societal influence is regulatory guidance. In 1998, the U.S. Securities and Exchange Commission (SEC) initiated the Plain English Movement, advocating for the use of clear and straightforward language in financial disclosures. This initiative was designed to facilitate efficient and effective communication between companies and their investors. The SEC provided guidelines to evade pitfalls in communication such as lengthy sentences, weak verbs, and abstract nominalizations (SEC 1998). The SEC identified the prevalent issue of long sentences in disclosure documents and provided strategies to avoid them. It also emphasized the problem of using weak verbs, which often involve substituting verbs with nouns derived from them, typically ending in "−tion." These nominalizations, while less active, tend to be more abstract than their verb counterparts (Loughran and McDonald 2014). In line with SEC guidelines, U.S. companies usually aim to decrease the use of nominalizations and shorten the average sentence length in their annual reports. This method enhances the comprehensibility of these reports for investors by promoting the use of verbs or verb phrases over their equivalent nouns or nominal phrases.

Incorporating measures of syntactic complexity has resulted in a substantial enhancement of classification performance, elevating the Area Under the Curve (AUC) to an impressive 99.3% and achieving an accuracy of 97%. This represents a marked improvement over previous methodologies, such as those reported by Liu et al. (2022), which attained an AUC of 90.5% and an accuracy rate of 84.3% by utilizing entropy values. Similarly, it outperforms the approach of Baroni and Bernardini (2006), which relied on the use of word or part-of-speech n-grams and achieved an accuracy rate of 86.7%. These comparisons underscore the efficacy of syntactic complexity measures in

refining the precision and reliability of classification outcomes. The high AUC and accuracy achieved by our machine learning models substantiate the categorical distinction between translated and original annual reports in terms of syntactic complexity. By evaluating eight distinct machine learning models, we further establish the unequal performance of different algorithms in differentiating between translated and non-translated annual reports. The findings of this study reveal that the Naive Bayes algorithm frequently underperforms in comparison to the other models. This suboptimal performance can potentially be attributed to Naive Bayes' inherent assumption of feature independence (Domingos and Pazzani 1997). This assumption may not hold true in real-world datasets, where features often exhibit correlation, thereby leading to issues of underfitting or overfitting. The subpar performance of Naive Bayes has been observed in some datasets, particularly when the assumption of independence is violated and the features are significantly correlated (e.g. Nguyen and Kim 2021). Given the tendency of syntactic complexity features to be highly correlated (Lei and Shi 2023), this could partially account for the subpar performance of the Naive Bayes model. These findings underscore the necessity of judiciously selecting the appropriate machine learning algorithm based on the data's characteristics and feature relationships.

In this study, we developed an ensemble model by amalgamating the three most effective machine learning algorithms identified from a pool of eight candidates. Conventional wisdom and existing literature often posit that ensemble models, through their integration of diverse algorithms, are likely to outperform individual models due to their ability to capture a broader range of patterns and reduce overfitting (Opitz and Maclin 1999; Rokach 2010). Contrary to these expectations, our findings indicate that the ensemble approach did not necessarily surpass the predictive performance of the best individual model in our experiment. This outcome echoes the observations made by Kuncheva and Whitaker (2003), who argue that the success of an ensemble is contingent upon the diversity and accuracy of the individual classifiers; if the constituent models are too similar or if one model is significantly more accurate than the others, the ensemble may not offer any substantial benefit. Moreover, the phenomenon where ensemble models fail to improve results has been acknowledged in studies such as (Sagi and Rokach 2018), which suggest that the complexity of combining models does not always translate to higher accuracy, particularly when the individual models are already highly optimized.

The present study makes valuable contributions to financial reporting by utilizing computational tools to analyze translated texts, a method that is typically not used in traditional translation studies. The application of machine learning techniques to assess syntactic complexity provides a quantitative way to analyze translation, which could supplement the more

traditional methods often used in translation studies. Machine learning models applied in the study improve with better data analysis and a better understanding of the task at hand. This interdisciplinary approach could pave the way for future research that combines these two fields, leading to more comprehensive and nuanced insights. In addition, by comparing the syntactic complexity of translated and non-translated corporate annual reports, this study provides empirical evidence about how translation affects syntactic structures. These insights hold the potential to significantly contribute to the evolution of new theories that delve into the ways translation affects text structures. Understanding these dynamics may, in turn, have noteworthy implications for the actual practice of translated financial reporting.

Our study also has practical implications for text classification and translation practices. Text classification models often rely on features like word frequency and document length. This study introduces syntactic complexity as a new kind of feature that can be used to improve the accuracy of these models. By incorporating syntactic complexity measures into the feature extraction stage, it might be possible to distinguish between different kinds of texts more accurately. Syntactic complexity is a key factor affecting how easy a text is to read and understand. By providing insights into the syntactic complexity of corporate annual reports, this study could help organizations improve the readability and comprehensibility of both their original and translated annual reports, making them more accessible to a wider audience, and thus yield more positive economic outcomes.

Conclusion

The aim of this study was to examine whether features of syntactic complexity could effectively distinguish between translated English annual reports and their non-translated counterparts. The comparison encompassed four dimensions of syntactic complexity, revealing distinctive characteristics in translational language, which, as a mediated form of communication, sets it apart from non-translated language in the context of annual reports. This study's findings not only successfully demonstrate the applicability of syntactic complexity in discerning between translated and non-translated text but also highlight the efficacy of advanced methodologies, such as machine learning models, in contrast to traditional statistical procedures.

The significance of this research lies in its contribution to new empirical evidence and enhanced understanding, specifically within the realm of translated English. It advances financial reporting by applying computational analysis to translated texts, introducing a quantitative approach to assess translation's impact on syntactic complexity. This novel method enhances traditional translation studies and can improve machine learning model

accuracy in text classification. By exploring the syntactic differences between translated and non-translated texts, the research offers insights that could influence translation theory and practice. Furthermore, it suggests that syntactic complexity serves as an important variable for text classification. Consequently, financial reporting can adjust the complexity of their reports to improve readability, benefiting organizations and stakeholders through clearer communication. However, it is important to acknowledge a limitation: the study's focus is restricted to the English language. As cautioned by prior research (Volansky, Ordan, and Wintner 2015), investigating the translationese phenomenon solely through comparable corpora may pose challenges in extrapolating the findings. Future research can broaden the investigations by exploring diverse language corpora. This inclusive approach allows for a more thorough examination of the complexities associated with translational language. Understanding how translation impacts text structures can then offer practical insights for financial reporting in multicultural and multilingual contexts.

Disclosure statement

Funding

ORCID

Zhongliang Wang ⓘ http://orcid.org/0000-0001-6007-1589
Ming Liu ⓘ http://orcid.org/0000-0002-3815-4705
Kanglong Liu ⓘ http://orcid.org/0000-0003-3962-2563

References

Ajina, A., M. Laouiti, and B. Msolli. 2016. Guiding through the fog: Does annual report readability reveal earnings management? *Research in International Business and Finance* 38:509–16. doi:10.1016/j.ribaf.2016.07.021.

Al-Jabr, A. F. 2006. Effect of syntactic complexity on translating from/into English/Arabic. *Babel* 52 (3):203–21. doi:10.1075/babel.52.3.01alj.

Baker, H. E., and D. D. Kare. 1992. Relationship between annual report readability and corporate financial performance. *Management Research News* 15 (2):1–4. doi:10.1108/eb028188.

Balodis, K., and D. Deksne. 2019. Fasttext-based intent detection for inflected languages. *Information* 10 (5):161. doi:10.3390/info10050161.

Baroni, M., and S. Bernardini. 2006. A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing* 21 (3):259–74.

Beattie, V., B. McInnes, and S. Fearnley. 2004. A methodology for analysing and evaluating narratives in annual reports: A comprehensive descriptive profile and metrics for disclosure quality attributes. *Accounting Forum* 28 (3):205–36. doi:10.1016/j.accfor.2004.07.001.

Benjamin, R. G. 2012. Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review* 24 (1):63–88. doi:10.1007/s10648-011-9181-8.

Bhatia, V. K. 2008. Genre analysis, ESP and professional practice. *English for Specific Purposes* 27 (2):161–74. doi:10.1016/j.esp.2007.07.005.

Bhatia, V. K. 2010. Interdiscursivity in professional communication. *Discourse & Communication* 4 (1):32–50. doi:10.1177/1750481309351208.

Bulté, B., and A. Housen. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing* 26:42–65. doi:10.1016/j.jslw.2014.09.005.

Chen, J., D. Li, and K. Liu. 2024. Unraveling cognitive constraints in constrained languages: A comparative study of syntactic complexity in translated, EFL, and native varieties. *Language Sciences* 102 (5):101612. doi:10.1016/j.langsci.2024.101612.

Clarke, D. P., S. L. Hrasky, and C. G. Tan. 2009. Voluntary narrative disclosures by local governments: A comparative analysis of the textual complexity of mayoral and chairpersons' letters in annual reports. *Australian Journal of Public Administration* 68 (2):194–207. doi:10.1111/j.1467-8500.2009.00630.x.

Courtis, J. K. 1995. Readability of annual reports: Western versus Asian evidence. *Accounting Auditing & Accountability Journal* 8 (2):4–17. doi:10.1108/09513579510086795.

Courtis, J. K., and S. Hassan. 2002. Reading ease of bilingual annual reports. *Journal of Business Communication* 39 (4):394–413. doi:10.1177/002194360203900401.

Crossley, S., S. Skalicky, C. Berger, and A. Heidari 2022. Assessing readability formulas in the wild. In *Polyphonic Construction of Smart Learning Ecosystems: Proceedings of the 7th Conference on Smart Learning Ecosystems and Regional Development*, ed. M. Dascalu, P. Marti, and F. Pozzi, 91–101. Singapore: Springer Nature Singapore.

Dogra, V., A. Singh, S. Verma, A. Alharbi, and W. Alosaimi. 2021. Event study: Advanced machine learning and statistical technique for analyzing sustainability in banking stocks. *Mathematics* 9 (24):3319. doi:10.3390/math9243319.

Domingos, P., and M. Pazzani. 1997. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning* 29 (2/3):103–30. doi:10.1023/A:1007413511361.

Frantz, R. S., L. E. Starr, and A. L. Bailey. 2015. Syntactic complexity as an aspect of text complexity. *Educational Researcher* 44 (7):387–93. doi:10.3102/0013189X15603980.

Garzone, G. 2004. Annual company reports and CEO's letters: Discoursal features and cultural markedness. In *Intercultural aspects of specialized communication*, ed. C. Candlin and M. Gotti, 311–43. US: Peter Lang.

Habib, A., and M. M. Hasan. 2020. Business strategies and annual report readability. *Accounting & Finance* 60 (3):2513–47. doi:10.1111/acfi.12380.

Habic, V., A. Semenov, and E. L. Pasiliao. 2020. Multitask deep learning for native language identification. *Knowledge-Based Systems* 209:106440. doi:10.1016/j.knosys.2020.106440.

Halverson, S. L. 2003. The cognitive basis of translation universals. *Target International Journal of Translation Studies* 15 (2):197–241. doi:10.1075/target.15.2.02hal.

Halverson, S. L. 2017. Gravitational pull in translation: Testing a revised model. In *Empirical translation studies: New methodological and theoretical traditions*, ed. G. D. Sutter, M. Lefer, and I. Delaere, 9–46. US: Mouton de Gruyter.

Hand, D. J. 2012. Assessing the performance of classification methods. *International Statistical Review* 80 (3):400–14. doi:10.1111/j.1751-5823.2012.00183.x.

House, J. 2014. *Translation quality assessment: Past and present.* US: Routledge.

Housen, A., and F. Kuiken. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30 (5):461–73. doi:10.1093/applin/amp048.

Huang, D., and Y. Wang. 2020. A CGA-Based study on translation characteristics of chairman speeches in company annual reports. *Journal of Literature and Art Studies* 10 (5):126–35. doi:10.17265/2159-5836/2020.02.005.

Huang, A. H., H. Wang, and Y. Yang. 2023. FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research* 40 (2):806–41. doi:10.1111/1911-3846.12832.

Hu, H., and S. Kübler. 2021. Investigating translated Chinese and its variants using machine learning. *Natural Language Engineering* 27 (5):339–72. doi:10.1017/S1351324920000182.

Ilisei, I., D. Inkpen, G. C. Pastor, and R. Mitkov. 2010. Identification of translationese: A machine learning approach. In *Computational linguistics and intelligent text processing. CICLing* 2010, Vol. 6008 of *lecture notes in computer science*, ed. A. Gelbukh, 503–11. US: Springer.

Jagaiah, T., N. G. Olinghouse, and D. M. Kearns. 2020. Syntactic complexity measures: Variation by genre, grade level, students' writing abilities, and writing quality. *Reading and Writing* 33 (5):2577–638. doi:10.1007/s11145-020-10057-x.

Jeanjean, T., C. Lesage, and H. Stolowy. 2010. Why do you speak English (in your annual report)? *The International Journal of Accountin* 45 (2):200–23. doi:10.1016/j.intacc.2010.04.003.

Jeanjean, T., H. Stolowy, and M. Erkens. 2010. Really "Lost in translation"? The economic consequences of issuing an annual report in English. *Crises et nouvelles problématiques de la Valeur* 1–36.

Jin, T., X. Lu, and J. Ni. 2020. Syntactic complexity in adapted teaching materials: Differences among grade levels and implications for benchmarking. *The Modern Language Journal* 104 (1):192–208. doi:10.1111/modl.12622.

Jones, M. J. 1994. A comment to contextualize "Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *The Journal of Business Communication* 31 (3):225–30. doi:10.1177/002194369403100305.

Kuncheva, L. I., and C. J. Whitaker. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning* 51 (2):181–207. doi:10.1023/A:1022859003006.

Lau, R. Y., S. Y. Liao, R. C. W. Kwok, K. Xu, Y. Xia, and Y. Li. 2012. Text mining and probabilistic language modeling for online review spam detection. *ACM Transactions on Management Information Systems* 2 (4):1–30. doi:10.1145/2070710.2070716.

Lehavy, R., F. Li, and K. Merkley. 2011. The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review* 86 (3):1087–115. doi:10.2308/accr.00000043.

Lei, L., and Y. Shi. 2023. Syntactic complexity in adapted extracurricular reading materials. *System* 113:103002. doi:10.1016/j.system.2023.103002.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45 (2–3):221–47. doi:10.1016/j.jacceco.2008.02.003.

Liao, S. 2021. A multidimensional analysis of letters to the shareholders by Chinese and American advanced equipment manufacturing companies. *International Journal of Language and Linguistics* 9 (4):161–68. doi:10.11648/j.ijll.20210904.14.

Li, J., W. Luo, and X. Deng. 2019. The effect of chairman's statement tone changes in annual reports from Hong Kong. *Journal of Physics Conference Series* 1168 (3):032024. doi:10.1088/1742-6596/1168/3/032024.

Lim, E. K., K. Chalmers, and D. Hanlon. 2018. The influence of business strategy on annual report readability. *Journal of Accounting and Public Policy* 37 (5):65–81. doi:10.1016/j. jaccpubpol.2018.01.003.

Lin, Y., and J. Liang. 2023. Informativeness across interpreting types: Implications for language shifts under cognitive load. *Entropy* 25 (2):243.

Liu, K., and M. Afzaal. 2021. Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. *Public Library of Science ONE* 16 (6):e0253454. doi:10. 1371/journal.pone.0253454.

Liu, K., Z. Liu, and L. Lei. 2022. Simplification in translated Chinese: An entropy-based approach. *Lingua* 286 (5):103364. doi:10.1016/j.lingua.2022.103364.

Liu, K., R. Ye, Z. Liu, and R. Ye. 2022. Entropy-based discrimination between translated Chinese and original Chinese using data mining techniques. *Public Library of Science ONE* 17 (3):e0265633. doi:10.1371/journal.pone.0265633.

Loughran, T., and B. McDonald. 2014. Measuring readability in financial disclosures. *The Journal of Finance* 69 (4):1643–71. doi:10.1111/jofi.12162.

Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics* 15 (4):474–96. doi:10.1075/ijcl.15.4.02lu.

Lu, X. 2011. A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly* 45 (1):36–62. doi:10.5054/tq. 2011.240859.

Lu, X., and H. Ai. 2015. Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing* 29:16–27. doi:10.1016/j.jslw.2015.06.003.

Lu, X., D. A. Gamson, and S. A. Eckert. 2014. Lexical difficulty and diversity in American elementary school reading textbooks: Changes over the past century. *International Journal of Corpus Linguistics* 19 (1):94–117. doi:10.1075/ijcl.19.1.04lu.

Manning, C. D., P. Raghavan, and H. Schütze. 2008. *Introduction to information retrieval*. Cambridge, UK: Cambridge University Press.

Mouthami, K., K. N. Devi, and V. M. Bhaskaran. 2013. Sentiment analysis and classification based on textual reviews. In *2013 International Conference on Information communication and embedded systems (ICICES)*, 271–76. Chennai.

Nguyen, B. Q. V., and Y. T. Kim. 2021. Landslide spatial probability prediction: A comparative assessment of naïve Bayes, ensemble learning, and deep learning approaches. *Bulletin of Engineering Geology and the Environment* 80 (5):4291–321. doi:10.1007/s10064-021-02194-6.

Nisioi, S., and L. P. Dinu. 2013. A clustering approach for translationese identification. In Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, 532–38.

Opitz, D., and R. Maclin. 1999. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research* 11:169–98. doi:10.1613/jair.614.

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics* 24 (5):492–518. doi:10. 1093/applin/24.4.492.

Ortega, L. 2015. Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing* 29 (5):82–94. doi:10.1016/j.jslw.2015.06.008.

Pallotti, G. 2015. A simple view of linguistic complexity. *Second Language Research* 31 (5):117–34. doi:10.1177/0267658314536435.

Rabinovich, E., and S. Wintner. 2015. Unsupervised identification of translationese. *Transactions of the Association for Computational Linguistics* 3 (5):419–32. doi:10.1162/ tacl_a_00148.

Ren, C., and X. Lu. 2021. A multi-dimensional analysis of the management's discussion and analysis narratives in Chinese and American corporate annual reports. *English for Specific Purposes* 62 (5):84–99. doi:10.1016/j.esp.2020.12.004.

Rokach, L. 2010. Ensemble-based classifiers. *Artificial Intelligence Review* 33 (1–2):1–39. doi:10.1007/s10462-009-9124-7.

Rubino, R., E. Lapshinova-Koltunski, and J. Van Genabith. 2016. Information density and quality estimation features as translationese indicators for human translation classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, San Diego, California, USA, 960–70.

Rutherford, B. A. 2003. Obfuscation, textual complexity and the role of regulated narrative accounting disclosure in corporate governance. *Journal of Management & Governance* 7 (5):187–210. doi:10.1023/A:1023647615279.

Sagi, O., and L. Rokach. 2018. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8 (4):e1249. doi:10.1002/widm.1249.

Schroeder, N., and C. Gibson. 1990. Readability of management's discussion and analysis. *Accounting Horizons* 4 (December):78–87.

SEC (The U.S. Securities and Exchange Commission). A plain English handbook: How to create clear SEC disclosure documents. 1998. Accessed July 16, 2023. http://www.sec.gov/pdf/handbook.pdf

Smith, M., A. Jamil, Y. Chik Johari, and S. Ahmar Ahmad. 2006. The chairman's statement in Malaysian companies: A test of the obfuscation hypothesis. *Asian Review of Accounting* 14 (1–2):49–65. doi:10.1108/13217340610729464.

Smith, M., and R. Taffler. 1992. The chairman's statement and corporate financial performance. *Accounting & Finance* 32 (4):75–90. doi:10.1111/j.1467-629X.1992.tb00187.x.

Subramanian, R., R. G. Insley, and R. D. Blackwell. 1993. Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *Journal of Business Communication* 30 (1):49–61. doi:10.1177/002194369303000103.

Volansky, V., N. Ordan, and S. Wintner. 2015. On the features of translationese. *Digital Scholarship in the Humanities* 30 (1):98–118. doi:10.1093/llc/fqt031.

Wang, P. 2014. A comparison between Chinese version and English version of company annual reports. *The Journal of Chinese Sociolinguistics* 2:33–44.

Wang, Z., K. Liu, and R. Moratto. 2023. A corpus-based study of syntactic complexity of translated and non-translated chairman's statements. *Translation & Interpreting* 15 (1):135–51. doi:10.12807/ti.115201.2023.a07.

Wu, J., K. Liu, R. Hu, and W. Zhou. 2023. A comparative study of the syntactic complexity of translated Chinese and original Chinese. *Foreign Language Teaching and Research* 55 (2):264–75.

Wu, X., A. Mauranen, and L. Lei. 2020. Syntactic complexity in English as a lingua franca academic writing. *Journal of English for Academic Purposes* 43:100798. doi:10.1016/j.jeap.2019.100798.

Xu, J., and J. Li. 2021. A syntactic complexity analysis of translational English across genres. *Across Languages and Cultures* 22 (2):214–32. doi:10.1556/084.2021.00015.

Xu, H., and K. Liu. 2023. Syntactic simplification in interpreted English: Dependency distance and direction measures. *Lingua* 294 (5):103607. doi:10.1016/j.lingua.2023.103607.

Zhou, S., K. Li, and Y. Liu. 2009. Text categorization based on topic model. *International Journal of Computational Intelligence Systems* 2 (4):398–409. doi:10.1080/18756891.2009.9727671.