“The complexity burden in transfer pricing compliance: A computational assessment of Ukrainian tax law and its implications for accounting”

AUTHORS	Serhii Lehenchuk Dmytro Zakharov Viktoriia Gryn Viktoriia Makarovych Gabriella Loskorikh

ARTICLE INFO	Serhii Lehenchuk, Dmytro Zakharov, Viktoriia Gryn, Viktoriia Makarovych and Gabriella Loskorikh (2026). The complexity burden in transfer pricing compliance: A computational assessment of Ukrainian tax law and its implications for accounting. *Accounting and Financial Control*, *7*(1), 49-65. doi:10.21511/afc.07(1).2026.05
DOI	http://dx.doi.org/10.21511/afc.07(1).2026.05
RELEASED ON	Friday, 27 February 2026
RECEIVED ON	Friday, 12 December 2025
ACCEPTED ON	Tuesday, 24 February 2026

JOURNAL	"Accounting and Financial Control"
ISSN PRINT	2543-5485
ISSN ONLINE	2544-1450
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	Sp. z o.o. Kozmenko Science Publishing

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
29	0	2

BUSINESS PERSPECTIVES

LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine

www.businessperspectives.org

Serhii Lehenchuk (Ukraine), **Dmytro Zakharov** (Ukraine), **Viktoriia Gryn** (Ukraine), **Viktoriia Makarovych** (Ukraine), **Gabriella Loskorikh** (Ukraine)

THE COMPLEXITY BURDEN IN TRANSFER PRICING COMPLIANCE: A COMPUTATIONAL ASSESSMENT OF UKRAINIAN TAX LAW AND ITS IMPLICATIONS FOR ACCOUNTING

Type of the article: Research Article

Received on: 12th of December, 2025
Accepted on: 24th of February, 2026
Published on: 27th of February, 2026

Serhii Lehenchuk, Doctor of Economics, Professor, Head of the Department of Information Systems in Management and Accounting, Faculty of Business and Service Sector, Zhytomyr Polytechnic State University, Ukraine. (Corresponding author)

Dmytro Zakharov, Ph.D., Associate Professor, Department of Information Systems in Management and Accounting, Faculty of Business and Service Sector, Zhytomyr Polytechnic State University, Ukraine.

Viktoriia Gryn, Doctor of Economics, Associate Professor, Faculty of Economics, Department of Accounting and Taxation, Zaporizhzhia National University, Ukraine.

Viktoriia Makarovych, Doctor of Economics, Professor, Department of Accounting and Auditing, Ferenc Rakoczi II Transcarpathian Hungarian College of Higher Education, Ukraine.

Gabriella Loskorikh, Ph.D., Associate Professor, Department of Accounting and Auditing, Ferenc Rakoczi II Transcarpathian Hungarian College of Higher Education, Ukraine.

Abstract

Ukrainian transfer pricing legislation demonstrates a significantly higher level of regulatory complexity than international OECD standards, creating a disproportionate burden on the accounting and reporting system. The study aims to quantify the regulatory burden generated by the complexity of Ukrainian transfer pricing legislation through a computational linguistic analysis of its algorithmic characteristics in comparison with international OECD standards. The research methodology is based on Halstead metrics to calculate the algorithmic complexity of regulatory texts, considered as formal structures with the distribution of lexical units into operators and operands. The computational assessment reveals that Ukrainian transfer pricing regulations demonstrate algorithmic complexity 10 to 37 times higher than OECD standards, as the complexity index (L) for Article 39 of the Tax Code of Ukraine equals 2.742 percent versus 0.148 percent for OECD Transfer Pricing Guidelines, while Law of Ukraine No. 4536-IX (Verkhovna Rada of Ukraine, 2023) reaches 5.455 percent, exceeding international benchmarks by 37 times. This excessive complexity directly affects accounting practices, requiring additional resources for recordkeeping, increasing internal control requirements, and increasing the risk of financial reporting errors. The empirical findings demonstrate that excessive algorithmic density directly increases compliance costs for accounting departments, requiring additional resources for interpretation, documentation, and internal control. The study provides quantitative evidence supporting the necessity of systematic simplification of national transfer pricing regulations.

Keywords transfer pricing, taxation, regulatory policy, Pillar One, Pillar Two

JEL Classification F23, H25, K34, M41

INTRODUCTION

Transfer pricing (TP) is a cornerstone of taxation for multinational companies as it directly affects the allocation of tax bases between jurisdictions. However, regulation in this area worldwide faces a fundamental dilemma of balancing the pursuit of tax fairness and the prevention of base erosion against the need for administrative simplicity and predictability for businesses – a tension first articulated by Adam Smith in his principles of convenience and economy (Smith, 1776), which remains pertinent today, often manifesting as multi-layered regulation that increases overall tax complexity.

TP rules are widely recognized as a key driver of overall tax system complexity (Hoppe et al., 2021; Schipp et al., 2024). This complexity, while sometimes seen as a means to achieve greater fairness or attract investment (Hoppe et al., 2021; Esteller-Moré et al., 2021; Euler et al.,

Conflict of interest statement:
Author(s) reported no conflict of interest

2024), carries high economic costs. It can increase compliance and control costs, distort competition, stifle investment, and ultimately may lead to reduced tax revenues (Saad, 2014; Abeler & Jäger, 2015; Zwick, 2021). Consequently, simplifying TP administration has become a priority, as evidenced by its prominence in international initiatives such as the OECD's BEPS project and Pillar One/Two frameworks (Colliard et al., 2021).

The Ukrainian TP system presents a critical and revealing case for studying regulatory complexity as it combines two primary complexity factors: the need to implement intricate international OECD standards and the presence of distinctive national regulatory features that often deviate from these global benchmarks. This combination manifests itself in specific challenges such as legal uncertainty and ambiguous enforcement practices, as observed by Hryshchuk and Ponomarenko (2025), the significant administrative burden and financial costs placed on corporate accounting departments, detailed by Kvasovskyi et al. (2024), and the consequent need for businesses to rely heavily on external consultants to mitigate compliance risks, a pattern consistent with the findings of Pomeranz and Serrato (2025).

Despite the acknowledged severity of these issues, there is a notable research gap. The Ukrainian TP landscape lacks comprehensive, quantitative assessments of its inherent complexity. Without a clear, measurable understanding of the problem's scale and structure, any regulatory reforms aimed at simplification are likely to be based on anecdotal evidence and may have limited effectiveness.

Therefore, this study aims to quantify the regulatory burden generated by the complexity of Ukrainian TP legislation through a computational linguistic analysis of its algorithmic characteristics in comparison with international OECD standards. The findings are intended to contribute to more efficient tax administration, reduce compliance costs for businesses, and inform the ongoing evolution of TP systems in economies undergoing similar tax integration processes.

1. LITERATURE REVIEW

The issue of tax complexity has a deep theoretical foundation in modern economic science. The modern understanding of this phenomenon is largely based on the fundamental taxonomy developed by Bradford (1986, 2005), which decomposes complexity into three key, interrelated components:

1) rule complexity, reflecting the problems of interpreting formal and informal tax norms;

2) compliance complexity, associated with the procedural burden of record-keeping, form selection, and settlement; and

3) transactional complexity, arising during tax planning strategies. This classification remains dominant and forms a critical conceptual basis for empirical analysis.

Building on this foundation, Slemrod (2005) provided an operational definition, conceptualizing tax complexity as the total resource costs of tax collection incurred by taxpayers, authorities, and third parties. His research highlighted specific causes of complexity, such as poor coordination between regulatory levels and frequent legislative changes.

In the context of taxing multinational enterprises, Eden (2023) frames the challenge as a "wicked problem", characterized by complex interdependencies between national sovereignties, capital mobility, and corporate profit maximization. This systemic perspective underscores three fundamental challenges for tax authorities: determining jurisdiction, allocating income, and valuing cross-border transactions, with the latter constituting the core domain of TP.

Empirical studies confirm the significant economic burden of tax complexity, particularly pronounced in international taxation. Blumenthal and Slemrod (1995) found that compliance costs for foreign income regulations were disproportionately high for multinational corporations, ac-

http://dx.doi.org/10.21511/afc.07(1).2026.05

counting for about 40% of their total compliance burden, and directly linked these costs to regulatory complexity. This burden is not limited to compliance costs but extends to behavioral responses, as a comprehensive cross-country investigation by Saptono et al. (2024) provides robust evidence that complex tax systems directly increase the propensity for firm tax evasion. Analyzing data from over 46,000 firms in 83 countries, the study finds these effects to be more pronounced in lower-income countries and specific sectors. This behavioral impact of complexity is further detailed at the level of individual taxpayer psychology. Research by Paleka and Vitezić (2023), utilizing factor and cluster analysis, segments taxpayers into distinct behavioral profiles (e.g., "compliant", "adversarial", and "opportunistic"). Their work reveals that high perceived complexity is a key driver pushing taxpayers into less cooperative clusters, directly linking subjective regulatory burden to measurable declines in compliance morale and intention.

This erosion of voluntary compliance at the individual level ultimately translates into a measurable deficit at the macroeconomic level. Research by Aguirre and Del Villar (2024) on Latin American customs systems quantifies this link, demonstrating that higher procedural and regulatory complexity directly reduces a state's "tax effort" – its efficiency in collecting potential revenue. Their findings provide crucial evidence that complex systems can become self-defeating, eroding the very fiscal capacity they are designed to protect.

Recent research consistently identifies TP rules as a primary driver of tax code complexity. Hoppe et al. (2021) highlight that TP regulations create substantial complexity for businesses, a finding corroborated by their cross-country Tax Complexity Index. This complexity has tangible consequences, as Benzarti and Wallossek (2023) demonstrate a correlation between tax code complexity, which is often proxied by word count, and the tax gap, thereby illustrating how intricate regulations can undermine compliance and revenue collection.

Recent methodological advances using textual analysis further support this view, as studies like Swenson et al. (2024) demonstrate that tax-reducing provisions in legislation are intentionally crafted with greater verbosity and cognitive complexity to limit their application, thereby linking structural textual features directly to regulatory intent and compliance burden.

A significant methodological breakthrough in quantifying regulatory complexity came from adapting approaches from computer science. The foundation lies in Halstead's (1977) metrics, originally designed to assess the psychological complexity of software algorithms by analyzing operators and operands. This framework was adapted for legal analysis by rethinking norms as algorithms that transform taxpayer data into tax liabilities.

Katz and Bommarito (2014) pioneered the application of computational linguistics and network analysis to legal texts, proving that the structural complexity of legislation could be measured with quantitative indicators. This approach was directly applied to tax documentation by Colliard et al. (2021), who used adapted Halstead metrics to analyze OECD TP guidelines. Their comparative study revealed that newer international initiatives (Pillar One and Two) exhibit higher lexical density and uniqueness, signaling rising complexity.

Building on these foundations, Hoppe et al. (2021) developed a comprehensive Tax Complexity Index that integrates various dimensions of corporate tax systems. Alongside such composite indices, researchers employ alternative proxies like word count and readability indices while acknowledging their limitations, as a longer text has the potential to either obscure or clarify the rules it contains.

An important conceptual contribution to the field is the institutional distinction between necessary and unnecessary complexity, advanced by the Office of Tax Simplification (2015). Necessary complexity represents the minimum level of detail required to achieve a policy objective, whereas unnecessary complexity adds no normative value and only hinders application. This distinction is crucial for evaluating the quality of legislation and targeting simplification efforts effectively.

Within Ukrainian academic discourse, the study of TP has focused primarily on qualitative analysis of legal adaptation and practical enforcement. Scholars such as Pyroha (2024) provide critical

analysis of legislative reforms, highlighting conceptual shortcomings and resulting ambiguities in the law. Similarly, Hryshchuk and Ponomarenko (2025) examine issues of legal uncertainty and enforcement. The practical burden on businesses is emphasized by Kvasovskyi et al. (2024), who identify the high complexity of compliance procedures as a fundamental determinant of tax risks, leading to significant resource costs for data processing and specialist engagement.

Despite significant advances in studying tax system complexity at the international level, Ukrainian academic discourse lacks systematic quantitative studies of national TP legislation complexity using formalized metrics. Existing works are valuable but limited to qualitative analysis or descriptions of contradictions. This methodological gap is critical, as without a clear, empirical understanding of the sources and levels of complexity, any attempts at legislative reform in Ukraine remain intuitive and potentially ineffective. This study aims to address this gap by applying adapted computational metrics to quantitatively assess the regulatory complexity of Ukraine's TP system.

2. AIMS

The study aims to quantify the regulatory burden generated by the complexity of Ukrainian transfer pricing legislation through a computational linguistic analysis of its algorithmic characteristics in comparison with the international OECD standard.

3. METHODOLOGY

This research employs a formal quantitative methodology for the comparative analysis of normative legal texts, grounded in the principles of computational linguistics and algorithmic complexity theory (Halstead, 1977). The core analytical premise is the conceptualization of tax regulations as deterministic algorithms – formalized sequences of logical and procedural instructions that define the transformation of inputs (e.g., financial data, transactional parameters) into compliance determinations and tax liabilities. This conceptual model facilitates the translation of qualitative textual attributes into objective, measurable indicators of structural complexity, enabling a replicable empirical assessment of regulatory burden (Colliard et al., 2021).

To ensure methodological rigor and cross-document comparability, a systematic, multi-stage procedure was developed and applied uniformly to a corpus of foundational international and national legal acts. The methodology is designed to be transparent and reproducible, comprising three principal phases:

1) the construction of standardized, functionally equivalent classification dictionaries;

2) the automated lexical processing and syntactic annotation of the document corpus;

3) the calculation and comparative interpretation of formal complexity metrics derived from the annotated texts.

The comparative analysis is based on a curated corpus of fundamental regulatory acts governing TP. The selection includes cornerstone international guidelines and key Ukrainian legislative instruments to enable a robust cross-jurisdictional comparison. The corpus consists of the OECD Transfer Pricing Guidelines from 2017 as the established international benchmark, the OECD Pillar One Blueprint on the Tax Certainty Framework for Amount A from 2020 representing a recent and complex evolution in international tax rulemaking, Article 39 of the Tax Code of Ukraine in its current version as the core national statute, Law of Ukraine No. 466-IX from 2020 (Verkhovna Rada of Ukraine, 2020) as a significant legislative reform package, and Law of Ukraine No. 4536-IX from 2023 (Verkhovna Rada of Ukraine, 2023) as the most recent major legislative update.

The operationalization of Halstead`s framework necessitated the development of two language-specific, functionally equivalent dictionaries for the classification of lexical units as either operators or operands. This foundational step was executed through a structured, four-stage procedure to ensure validity and reliability:

Stage 1: Taxonomy Adaptation. The initial categorical framework for operators and operands was

http://dx.doi.org/10.21511/afc.07(1).2026.05

adapted from the seminal work of Colliard et al. (2021), who first applied software complexity metrics to OECD tax documentation.

Stage 2: Criteria Application. Lexical units were identified and classified based on dual criteria. The first criterion was semantic relevance – the term's functional role in prescribing action, establishing conditions, or defining subjects/objects. The second criterion was frequency of recurrence across the analytical corpus and within the broader legal-taxonomic domain.

Stage 3: Establishment of Functional Equivalence. For the analysis of Ukrainian-language legislation, a rigorous process was undertaken to establish functional equivalence. This ensured that terms performing identical normative, logical, or procedural functions in Ukrainian were accurately mapped to their corresponding operator or operand categories within the original English-based analytical framework.

Stage 4: Expert Validation. The preliminary dictionaries underwent iterative expert validation by researchers specialized in tax law and computational linguistics to assess and ensure completeness, categorical consistency, and terminological accuracy (see Appendix A for the finalized dictionaries).

The resulting taxonomies categorize textual elements as follows:

1. Operators – logical and procedural constructs that define actions, conditions, permissions, and relationships. Primary categories include:

- Logical/Conditional: e.g., "if", "provided that", "unless", "in case";
- Normative/Deontic: e.g., "shall", "must", "may", "should", "is required to";
- Mathematical/Comparative: e.g., "exceeds", "equals", "at least", "not less than";
- Procedural: e.g., "calculate", "document", "submit", "verify", "apply".

2. Operands – the objects, subjects, values, and concepts upon which operators act. Primary categories include:

- Legal and Tax Subjects: e.g., "taxpayer", "controlled entity", "reporting person";
- Economic Quantities and Indicators: e.g., "income", "price", "profit margin", "cost".;
- Methodological Concepts: e.g., "comparable uncontrolled price", "arm's length principle", "transfer pricing method";
- Jurisdictional and Procedural Terms: e.g., "reporting period", "tax authority", "controlled transaction".

The entire document corpus was processed computationally using the Natural Language Toolkit within a Python programming environment. A standardized pre-processing pipeline was uniformly applied to all texts, comprising tokenization (segmenting text into individual words and punctuation), lemmatization (reducing words to their canonical dictionary form), and the removal of language-specific stop-words (e.g., common articles, prepositions, conjunctions without normative function). Subsequently, an automated annotation algorithm parsed each processed document, classifying every lexical token as an operator, an operand, or a neutral linguistic element based on its presence and categorization in the predefined dictionaries.

Following the automated annotation, primary counts were extracted for each document to serve as inputs for the calculation of Halstead (1977) metrics, adapted here as indicators of regulatory complexity:

$$V = N_1 + N_2, \quad (1)$$

where V is the total volume of the regulatory text; N_1 is the total number of operators; N_2 is the total number of operands.

$$V^* = 2 + \eta_2, \quad (2)$$

where V^* is the potential minimum volume; η_2 is the number of unique operands.

To further assess structural efficiency, the Level of complexity (L) is calculated as the ratio of the program's potential minimum volume ($V^* = 2 + \eta_2$) to its actual volume (V), expressed as a percentage:

$$L = \frac{V^*}{V} \cdot 100\%, \quad (3)$$

where L is the level of complexity of the regulatory text, %.

The L indicator is interpreted as a proportion of required elements in the total volume of the text; it characterizes the degree of compactness, repeatability, and logical richness of the regulatory document. Metrics measure not the number of words, but the density of unique operations and concepts relative to the total volume of text. The L indicator is the ratio of the minimum required elements to the actual ones, making it independent of the absolute language verbosity. To ensure comparability of results, unified criteria for classifying lexical units were applied, adapted to the structural features of each legal system, but functionally equivalent in purpose.

In summary, the employed methodology constitutes a quantitative, algorithmically oriented approach to legal text analysis that extends beyond traditional doctrinal interpretation. By transforming regulatory texts into formalized, calculable structures, this method enables the study of legislation through both qualitative and quantitative lenses. It allows for the assessment of normative density, the identification of multi-layered logical dependencies, and the systematic comparison of regulatory complexity across different legal systems using uniform, reproducible indicators. The resulting metrics provide an empirical basis for evaluating the structural coherence, cognitive load, and interpretative transparency of TP legislation, thereby strengthening the analytical foundation for conclusions regarding its regulatory efficacy and potential avenues for simplification.

4. RESULTS

Table 1 presents the results of the empirical analysis of the algorithmic complexity of key regulatory and legal acts governing TP at the national and international levels. Based on automated text processing for the five studied documents, lexical units were identified and classified into operator (N_1, η_1) and operand (N_2, η_2) categories. The application of adapted Halstead (1977) metrics allowed for the calculation of integral characteristics of the text's structural organization and the resulting complexity level (L). As the data in Table 1 demonstrate, the algorithmic density indicators show significant variability. The coefficient L reflects the relationship between the potential minimum volume (V*) – theoretically required to convey the logical structure – and the actual text volume (V). The higher the proportion of unique elements per unit of total text, the higher its algorithmic richness and interpretation complexity.

The interpretation of normative complexity indicators is based on the assumption that the value of the coefficient L reflects the relationship between the potential minimum volume and actual volume of the normative text. The greater the proportion of unique elements per unit of general text, the higher its algorithmic richness and interpretation complexity. To interpret the L values obtained, an empirical classification is proposed based on statistical analysis of the distribution of data in the sample.

Based on the data distribution, three complexity categories were defined:

- Low complexity (L < 3%). This category includes international OECD guidelines (Transfer Pricing Guidelines, Pillar One Blueprint) and

Table 1. Results of the analysis of the normative complexity level

No.	Regulatory act	N_1	N_2	η_1	η_2	V	V*	L
1	Transfer Pricing Guidelines (2017)	14,704	7,556	36	31	22,260	33	0.148
2	Pillar One Blueprint (2020)	7,682	10,517	32	29	18,199	31	0.17
3	Law of Ukraine No. 466-IX (2020)	723	326	27	14	1,049	16	1.525
4	Article 39 of the Tax Code of Ukraine (2010)	408	139	23	13	547	15	2.742
5	Law of Ukraine No. 4536-IX (2023)	51	4	9	1	55	3	5.455

http://dx.doi.org/10.21511/afc.07(1).2026.05

certain provisions of Ukrainian legislation (Law of Ukraine No. 466-IX, Article 39 of the Law of Ukraine). Such texts are characterized by relative simplicity, consistency, and standardized constructions with minimal duplication (Verkhovna Rada of Ukraine, 2020);

- Moderate complexity (3% ≤ L < 6%). Documents demonstrate a multi-level logical structure with conditional connections, but retain a systematic presentation;

- High complexity (L ≥ 6%). The analysis revealed a natural break in the distribution between values of 2.74% and 5.46%, confirming the qualitative difference between documents with high complexity. Such regulatory and legal acts are characterized by increased terminological and syntactic density, a large number of exceptions, references, and clarifying constructions.

Table 2. Descriptive statistics of the complexity level indicator

No.	Indicator	Value, %
1	Average value (μ)	2.008
2	Median	1.525
3	Standard deviation	2.189
4	First quartile	0.159
5	Third quartile	2.742
6	Interquartile range	2.583
7	Min	0.148
8	Max	5.455
9	Range of variation	5.307
10	Coefficient of variation	109.0

Statistical analysis revealed significant variability in the complexity indicator. The mean value is 2.008% with a standard deviation of 2.189%, corresponding to a high coefficient of variation (109.0%). This indicates the presence of qualitatively different groups of documents in the sample. The median (1.525%) is lower than the mean, indicating a right-skewed distribution caused by the extremely high value for Law of Ukraine No. 4536-IX (Verkhovna Rada of Ukraine, 2023). The interquartile range (2.583%) shows that the central 50% of observations are concentrated in the range from 0.159% to 2.742%. Based on the distribution analysis, threshold values for categorizing complexity were defined: low (L < 3%); moderate (3% ≤ L < 6%); high (L ≥ 6%). The 3% threshold approximately corresponds to the 80th percentile of the distribution, providing an empirical justification for the boundaries between categories.

Within the OECD documentation, an objective differentiation in the level of detail is observed. The Pillar One Blueprint demonstrates 14.9% higher complexity compared to the Transfer Pricing Guidelines (0.170% vs. 0.148%), reflecting the more technically intricate structure of the new profit allocation initiatives.

The national legal system exhibits significantly higher and uneven internal variability. The complexity indicator ranges from 1.525% (Law of Ukraine No. 466-IX) to 5.455% (Law of Ukraine No. 4536-IX), representing a 3.6-fold difference (Verkhovna Rada of Ukraine, 2020; 2023). This indicates substantial differences in the structural organization even among acts from the same jurisdiction.

A comparison of Ukrainian and international regulatory acts reveals a critical disproportion, as Ukrainian TP regulations exceed the international benchmark of the OECD Transfer Pricing Guidelines in terms of structural complexity by significant multiples. Specifically, Law No. 466-IX is 10.3 times more complex, Article 39 of the Tax Code of Ukraine is 18.5 times more complex, and Law No. 4536-IX is 36.8 times more complex (Verkhovna Rada of Ukraine, 2020; 2023).

The case of Law No. 4536-IX is particularly illustrative. Despite having a significantly smaller total volume ($V = 55$) compared to the International Guidelines ($V = 22{,}260$), it demonstrates an almost 37-fold increase in the L indicator. This disproportion is difficult to explain solely by an objective need for detail.

The analysis of the distribution of operators and operands (η_1, η_2) confirms that the high complexity of Ukrainian acts is caused not only by volume but also by increased terminological and syntactic density. Documents with high L are characterized by a relatively large number of unique operators (conditional references, exceptions, clarifying constructions) and operands (unique terms) per text unit, complicating their interpretation.

Significant intra-group differences in complexity indicators, both among English-language documents (a 14.9% difference) and among Ukrainian-language documents (a 3.6-fold difference), provide a strong argument that the L indicator reflects normative complexity rather than linguistic features of the language. If the differentiation were caused solely by linguistic specificity, documents within the same language would demonstrate homogeneous results, which is not observed. Therefore, the L indicator is a relative measure that minimizes the impact of absolute verbosity and allows for comparative analyses.

5. DISCUSSION

This study introduces a novel application of algorithmic complexity analysis to TP regulations, providing a quantitative lens through which to examine structural density. Our core finding is a significant and systematic disparity in complexity between international best practices (OECD standards) and Ukrainian national legislation. This section interprets these findings by comparing them with prior research, reconciling methodological differences, and situating the results within the broader discourse on tax law efficacy and reform.

The most salient result is the order-of-magnitude higher complexity (L) of Ukrainian regulations compared to the OECD Transfer Pricing Guidelines (TPG). For instance, Article 39 of Ukraine's Tax Code shows an L value 18.5 times higher than the TPG. This quantitative evidence provides a structural explanation for the persistent qualitative critiques of Ukrainian tax law. Scholars have long highlighted issues of ambiguity, contradiction, and administrative burden (e.g., Kvasovskyi et al., 2024; Hryshchuk & Ponomarenko, 2025). Our analysis pinpoints the mechanism behind these issues: an excessive density of unique operands and operators (high η_1, η_2 relative to text volume), manifesting as a proliferation of cross-references, exceptions, and qualifying clauses that increase interpretation risk without clear regulatory benefit. This directly validates the observations of Kvasovskyi et al. (2024) regarding the abstract nature of provisions increasing tax risks, and explains the mechanism leading to the legal disputes noted by Hryshchuk and Ponomarenko (2025).

This aligns with Bradford (1986) taxonomy, where high "rule complexity" directly inflates "compliance complexity", The extreme case of Law No. 4536-IX (L = 5.455%), a very short but hyper-dense text, quantitatively validates observations by Pyroha (2024) that legislative amendments often add layers of procedural intricacy that obscure the core regulatory objective, a phenomenon akin to "legislative inflation" where "the real problem is completely lost sight of".

A direct comparison with the seminal work of Colliard et al. (2021) is crucial for validating our approach. Their reported L values for the TPG (1.75%) and Pillar One Blueprint (3.06%) are an order of magnitude higher than ours (0.148% and 0.170%, respectively). This discrepancy is not a contradiction but a reflection of methodological choice in dictionary construction. Colliard et al. (2021) employed extensive, inclusive dictionaries (157 operators, 139 operands) to achieve comprehensive coverage of all textual elements. Our study, conversely, uses focused dictionaries of core regulatory constructs (36 operators, 31 operands), deliberately filtering out common language to isolate normative complexity from general verbosity.

Importantly, both methodologies yield consistent comparative conclusions. First, they confirm that Pillar One is structurally more complex than the base Transfer Pricing Guidelines, with our study calculating it as 14.9 percent higher and Colliard et al. observing a similar directional trend. Second, both approaches demonstrate that national legislation is dramatically more complex than international standards. The key finding – that Ukrainian norms are disproportionately complex – remains true regardless of the absolute scale, as the relative ratios between documents remain robust across methodological paradigms. This convergence strengthens the validity of the core claim that, as noted by Colliard et al. (2021), regulatory complexity is increasing in new initiatives, but it also crucially highlights that the complexity inherent in Ukrainian legislation is of a different and excessive order altogether.

http://dx.doi.org/10.21511/afc.07(1).2026.05

Our findings resonate with and extend a wide range of global studies on tax complexity. For instance, regarding the Tax Complexity Index (TCI), Hoppe et al. (2021) identified TP as the foremost driver of complexity globally, and our results suggest Ukraine likely ranks alongside jurisdictions like Australia and Brazil at the high end of this spectrum, providing a specific, text-based metric to supplement survey-based indices like the TCI. On the relationship between complexity and efficacy, Benzarti and Wallossek's (2024) finding of a positive correlation between code complexity and the tax gap (non-compliance) supports our interpretive caution that the high L values in Ukrainian law may not enhance precision but rather create loopholes, uncertainty, and compliance errors, ultimately undermining policy goals and confirming that excessive detail is counterproductive. Finally, concerning accessibility, our algorithmic measure of structural density comes to a conclusion consistent with research on tax code readability: dense regulatory texts become inaccessible to non-specialists, effectively requiring expert interpretation. This demand for expertise directly raises compliance costs, a link quantitatively established by Blumenthal and Slemrod (1995), who found that complex international rules accounted for 40 percent of multinationals` total compliance burden.

The results allow us to operationalize the conceptual distinction between "necessary" and "excessive" complexity (Office of Tax Simplification, 2015). Necessary complexity is exemplified by the Pillar One Blueprint, which shows how technically detailed rules for novel challenges like digital economy profit allocation can maintain a moderately low L value of 0.170 percent. This indicates complexity that arises from substantive regulatory need and is managed through systematic drafting. In contrast, the extreme L value of 5.455 percent for Law No. 4536-IX, which governs a narrow procedural rule, signals complexity that is divorced from regulatory substance. This excessive complexity likely stems from accumulative drafting, defensive legalism, and a lack of systemic editorial review, thereby confirming the thesis of Kvasovskyi et al. (2024) that such regulatory density forces businesses into either costly outsourcing or heavy internal investment. Thus, the L indicator serves as a diagnostic tool. A low L suggests efficient, principle-based drafting where actual volume approaches the theoretical minimum. A high L indicates textual "bloat" where redundant structures dominate.

For corporate accounting departments, high algorithmic density (L) directly translates into a measurable compliance overhead that can influence taxpayer behavior. The constant need to decipher intricate conditional logic, exceptions, and specialized terminology not only diverts critical resources but also systematically elevates the risk of reporting errors, fostering an environment of frustration and adversarial attitudes identified in taxpayer typologies (Paleka & Vitezić, 2023). Consequently, the L indicator quantifies the specific excessive complexity within a regulation – the textual "bloat" that imposes deadweight costs on accounting workflows without enhancing regulatory clarity or fairness, thereby undermining voluntary compliance. This metric provides a tangible benchmark for assessing the direct burden regulatory design places on accounting efficiency and overall tax morale.

The internal corporate friction and inefficiency measured by the L metric directly fuels the broader structural issue where systemic complexity erodes a state's tax effort, as shown by Aguirre and Del Villar (2024). Therefore, optimizing the algorithmic density of legislation is a critical step for both reducing business operational costs and restoring the efficiency of tax administration and fiscal capacity at the macroeconomic level.

This quantifiable burden on accounting functions exacerbates the negative implications for all stakeholders, substantiating earlier research. For businesses and advisors, as noted by Kvasovskyi et al. (2024), high complexity translates directly into significantly increased financial costs for collecting and processing data sets, gaining access to information sources, and attracting qualified specialists, thereby forcing a choice between costly internal expertise or outsourcing. For tax authorities, excessive complexity complicates administration, increases the likelihood of erroneous assessments, and fuels disputes, a phenomenon observed in practice by Hryshchuk and Ponomarenko (2025). For the broader economy, following the logic of Blumenthal and Slemrod (1995), if even simpler international rules impose high costs, the disproportionate complexity of Ukrainian legislation creates a disproportionately heavy compliance burden, which negatively impacts investment attractiveness.

CONCLUSION

This study aimed to quantify the regulatory burden generated by the complexity of Ukrainian TP legislation through a computational linguistic analysis of its algorithmic characteristics in comparison with international OECD standards.

The core empirical finding reveals a profound disparity. Analysis using adapted Halstead metrics demonstrates that the algorithmic density (L) of key Ukrainian acts ranges from 1.525% to 5.455%. In stark contrast, foundational OECD documents exhibit significantly lower values of 0.148% and 0.170%. This quantifies that Ukrainian norms are structurally 10 to 37 times more complex than their international counterparts, moving the debate from qualitative critique to measurable fact.

Based on this result, a principal conclusion must be drawn that the excessive complexity of Ukrainian TP law is not a matter of stylistic verbosity but a systemic flaw in regulatory design. This structural density, manifested as a high concentration of unique operators and operands (cross-references, exceptions, qualifying clauses), directly generates the administrative burdens and compliance risks widely acknowledged in the professional community. For corporate accounting departments, this translates into a measurable compliance overhead, where deciphering intricate rules diverts critical resources, elevates reporting risks, and fosters adversarial compliance attitudes, thereby undermining both operational efficiency and voluntary tax morale. It substantiates the argument that the current regulatory framework imposes disproportionately high cognitive and operational costs on businesses, hindering both voluntary compliance and effective state administration.

Therefore, the findings necessitate a fundamental policy shift. The goal of legislative reform should be redefined from merely adding new rules to actively reducing the algorithmic complexity index (L) of existing texts. This implies a systemic simplification project focused on architectural clarity, elimination of redundant constructs, and adoption of principle-based drafting techniques observed in OECD standards.

It is crucial to acknowledge the study's limitations, which define the context for interpreting its results and outline directions for future research. First, while the analysis demonstrates significant intra-linguistic variability, suggesting the metric captures normative rather than purely linguistic features, a direct comparison of texts across different languages and legal traditions may introduce subtle methodological biases. Second, while the focus on a limited sample of primary legislative acts is illustrative, it does not encompass the full regulatory landscape, including secondary bylaws and explanatory guidance, which also contribute to the overall compliance burden. Third, Halstead's methodology, though robust, represents one specific approach to measuring structural complexity; triangulation of these results with other methods – such as network analysis of cross-references, expert surveys on perceived complexity, or readability indices would strengthen the validity and comprehensiveness of the findings.

Given these considerations and the established need for simplification, future research should be directed towards transforming diagnostic metrics into actionable reform tools. The immediate prospect is to develop specific recommendations to reduce both rule complexity and compliance complexity, taking into account best international practices. This involves several concrete steps. First, the analysis should be expanded to a broader corpus, including subordinate legislation, to identify the most burdensome provisions. Second, a detailed typology of complexity-driving elements (e.g., specific types of cross-references, nested exceptions) must be created to prioritize targets for revision. Third, the potential reduction in compliance costs should be modeled through optimized procedures for documentation and tax reporting. Finally, model legislative clauses with lower algorithmic density need to be drafted. Ultimately, this applied research trajectory aims to provide policymakers with a clear, evidence-based roadmap for legislative redrafting, which will help reduce compliance costs, mitigate tax risks, and improve the operational efficiency of corporate accounting departments.

http://dx.doi.org/10.21511/afc.07(1).2026.05

AUTHOR CONTRIBUTIONS

Conceptualization: Serhii Lehenchuk, Dmytro Zakharov.
Data curation: Dmytro Zakharov, Viktoriia Makarovych, Gabriella Loskorikh.
Formal analysis: Viktoriia Gryn, Gabriella Loskorikh.
Funding acquisition: Viktoriia Gryn, Viktoriia Makarovych.
Investigation: Serhii Lehenchuk, Viktoriia Gryn, Viktoriia Makarovych, Gabriella Loskorikh.
Methodology: Serhii Lehenchuk, Dmytro Zakharov, Viktoriia Makarovych.
Project administration: Serhii Lehenchuk.
Resources: Dmytro Zakharov, Viktoriia Gryn, Gabriella Loskorikh.
Software: Gabriella Loskorikh.
Supervision: Serhii Lehenchuk, Dmytro Zakharov.
Validation: Viktoriia Gryn, Viktoriia Makarovych.
Visualization: Viktoriia Makarovych, Gabriella Loskorikh.
Writing – original draft: Serhii Lehenchuk, Dmytro Zakharov.
Writing – review & editing: Serhii Lehenchuk, Dmytro Zakharov, Viktoriia Gryn, Viktoriia Makarovych.

REFERENCES

1. Abeler, J., & Jäger, S. (2015). Complex tax incentives. *American Economic Journal: Economic Policy, 7*(3), 1-28. https://doi.org/10.1257/pol.20130137
2. Aguirre, J. G., & Del Villar, A. (2022). Effect of the complexity of the customs tax system on the tax effort. *Economies, 10*(3), 55. https://doi.org/10.3390/economies10030055
3. Benzarti, Y., & Wallossek, L. (2023). *Rising income tax complexity* (NBER Working Paper No. W31944). https://doi.org/10.2139/ssrn.4660269
4. Blumenthal, M., & Slemrod, J. B. (1995). The compliance cost of taxing foreign-source income: Its magnitude, determinants, and policy implications. *International Tax and Public Finance, 2*(1), 37-53. https://doi.org/10.1007/bf00873106
5. Bradford, D. F. (1986). *Untangling the income tax*. Harvard University Press eBooks. https://doi.org/10.4159/harvard.9780674733930
6. Bradford, D. F. (2005). A tax system for the twenty-first century. In Auerbach, A. J., & Hassett, K. A. (Eds.), *Toward Fundamental Tax Reform* (pp. 11-33). AEI Press. Retrieved from https://www.aei.org/wp-content/uploads/2013/12/-toward-fundamental-tax-reform-book_130711622278.pdf
7. Colliard, J., Eden, L., & Georg, C. (2021). Tax Complexity and Transfer Pricing Blueprints, Guidelines, and Manuals. *Tax Management International Journal, 50*(2), 76-89. https://doi.org/10.2139/ssrn.3749740
8. Eden, L. (2023). Taxing multinationals: Three lenses on international tax cooperation (Comments on UN Resolution 77/244, Promotion of inclusive and Effective international tax governance at the United Nations). *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4394624
9. Esteller-Moré, A., Rizzo, L., & Secomandi, R. (2021). The role of tax system complexity on foreign direct investment. *Applied Economics, 53*(45), 5208-5220. https://doi.org/10.1080/00036846.2021.1922587
10. Euler, I., Harst, S., Schanz, D., Sureth-Sloane, C., & Voget, J. (2024). *Tax complexity and foreign direct investment* (WU International Taxation Research Paper Series No. 2024-10). https://doi.org/10.2139/ssrn.4990616
11. Halstead, M. H. (1977). *Elements of Software Science*. Elsevier.
12. Hoppe, T., Schanz, D., Sturm, S., & Sureth-Sloane, C. (2021). The Tax Complexity Index – a Survey-Based Country measure of tax code and framework complexity. *European Accounting Review, 32*(2), 239-273. https://doi.org/10.1080/09638180.2021.1951316
13. Hryshchuk, H. V., & Ponomarenko, A. S. (2025). Assessing the effectiveness of tax control of transfer pricing in Ukraine in the context of digital transformation. *Aktualni Pytannia Ekonomichnykh Nauk,* (3), 1-15. (In Ukrainian). https://doi.org/10.5281/zenodo.17063236
14. Katz, D. M., & Bommarito, M. J. (2014). Measuring the complexity of the law: the United States Code. *Artificial Intelligence and Law, 22*(4), 337-374. https://doi.org/10.1007/s10506-014-9160-8
15. Kvasovskyi, O., Stetsko, M., & Nipialidi, O. (2024). Tax risks of business entities in the field of transfer pricing and ways to prevent and minimise them. *Visnik Ekonomiki – Herald of National Economics, (1),* 47-66. (In Ukrainian). https://doi.org/10.35774/visnyk2024.01.047
16. Office of Tax Simplification (OTS). (2015). *The OTS Complexity Index*. London: HM Treasury. Retrieved from https://assets.publishing.service.gov.uk/media/5a816d6ae5274a2e87dbd91f/OTS_complexity_index_methodology_June_2015.pdf

17. Paleka, H., & Vitezić, V. (2023). Tax Compliance Challenge through Taxpayers' Typology. *Economies, 11*(9), 219. https://doi.org/10.3390/economies11090219

18. Pomeranz, D., & Serrato, J. C. S. (2025). Do transfer pricing reforms lead to a boom in tax consultants? *National Tax Journal, 78*(2), 513-533. https://doi.org/10.1086/735462

19. Pyroha, S. S. (2024). Transfer pricing reforms in the context of BEPS: challenges remain. *Pravo i Bezpeka – Law and Safety, (1),* 163-172. (In Ukrainian). https://doi.org/10.32631/pb.2024.1.15

20. Saad, N. (2014). Tax knowledge, tax complexity and tax Compliance: Taxpayers' view. *Procedia – Social and Behavioral Sciences, 109,* 1069-1075. https://doi.org/10.1016/j.sbspro.2013.12.590

21. Saptono, P. B., Mahmud, G., Salleh, F., Pratiwi, I., Purwanto, D., & Khozen, I. (2024). Tax Complexity and firm Tax Evasion: A Cross-Country Investigation. *Economies, 12*(5), 97. https://doi.org/10.3390/economies12050097

22. Schipp, A., Siahaan, F., & Sureth-Sloane, C. (2024). *Determinants of Tax Complexity in Tax Regulations and Tax Procedures -Evidence from a Developing Country* (WU International Taxation Research Paper Series No. 2024-05). https://doi.org/10.2139/ssrn.4924632

23. Slemrod, J. (2005). The Etiology of Tax Complexity: Evidence from U.S. State Income Tax Systems. *Public Finance Review, 33*(3), 279-299. https://doi.org/10.1177/1091142105275003

24. Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations* (Book V, Ch. 2). Retrieved from http://gesd.free.fr/smith76bis.pdf

25. Swenson, C., Qu, H., & Song, X. (2024). Complexity and the Tax Code: A textual analysis. *Public Finance Review, 52*(4), 466-538. https://doi.org/10.1177/10911421241233110

26. Verkhovna Rada of Ukraine. (2010, December 2). *Podatkovyi kodeks Ukrainy* [Tax Code of Ukraine] (Act No. 2755-VI). Legislation of Ukraine. (In Ukrainian). Retrieved from https://zakon.rada.gov.ua/laws/show/2755-17

27. Verkhovna Rada of Ukraine. (2020, January 16). *Pro vnesennia zmin do Podatkovoho kodeksu Ukrainy shchodo vdoskonalennia administruvannia podatkiv, usunennia tekhnichnykh ta lohichnykh neuzgodzhenostei u podatkovomu zakonodavstvi [On Amendments to the Tax Code of Ukraine Regarding the Improvement of Tax Administration, Elimination of Technical and Logical Inconsistencies in Tax Legislation]* (Act No. 466-IX). Legislation of Ukraine. (In Ukrainian). Retrieved from https://zakon.rada.gov.ua/laws/show/466-20

28. Verkhovna Rada of Ukraine. (2023, March 9). *Pro vnesennia zmin do Podatkovoho kodeksu Ukrainy ta inshykh zakonodavchykh aktiv Ukrainy u zviazku z pryiniattiam Zakonu Ukrainy «Pro intehrovane zapobihannia ta kontrol promyslovoho zabrudnennia» ta z metoiu udoskonalennia okremykh polozhen podatkovoho zakonodavstva [On Amendments to the Tax Code of Ukraine and Other Legislative Acts of Ukraine in Connection with the Adoption of the Law of Ukraine "On Integrated Prevention and Control of Industrial Pollution" and for the Purpose of Improving Certain Provisions of Tax Legislation]* (Act No. 4536-IX). Legislation of Ukraine. (In Ukrainian). Retrieved from https://zakon.rada.gov.ua/laws/show/4536-20

29. Zwick, E. (2021). The costs of corporate tax complexity. *American Economic Journal Economic Policy, 13*(2), 467-500. https://doi.org/10.1257/pol.20180406

http://dx.doi.org/10.21511/afc.07(1).2026.05

APPENDIX A

Ukrainian legislation

```
import os
import pandas as pd
from docx import Document
from nltk.tokenize import word_tokenize
import nltk

# --- Preparation ---
try:
    nltk.data.find("tokenizers/punkt")
except LookupError:
    nltk.download("punkt")

# OPERATORS AND OPERANDS FOR EACH UKRAINIAN NORMATIVE ACT

# --- Law of Ukraine No. 466-IX ---
operators_466 = {
    # Logical
    "if", "in case", "provided", "except", "besides", "when", "however"
    # Normative
    "shall", "may", "must", "should", "not allowed",
    "subject to", "defined", "established", "allowed"
    # Mathematical
    "calculate", "accrued", "exceeds", "less than", "more than", "rate", "interest", "pro rata"
    # Procedural
    "submit", "inform", "register", "verify", "report", "applied",
    "considered", "control", "keep",
    # Coordination
    "interact", "agree", "agreed", "transmit", "give", "receive", "provide"
}

operands_466 = {
    # Subjects
    "taxpayer", "regulatory authority", "authorized body",
    "resident", "non-resident", "representative",
    # Economic items
    "income", "profit", "costs", "price", "cost", "royalty",
    "tax rate", "object of taxation", "tax base",
    # Documents
    "declaration", "tax reporting", "report", "inspection report",
    "report on controlled transactions", "TP documentation",
    # Indicators
    "annual income", "tax liability", "coefficient", "share", "interest", "risk",
    # Methods
    "method", "comparative analysis", "market approach", "cost calculation", "pricing", "arm's length principle",
    # Jurisdiction
    "state", "Ukraine", "country", "tax jurisdiction", "international treaty"
}

# --- Law of Ukraine No. 4536-IX ---
operators_4536 = {
    "if", "in case", "provided", "except", "when",
    "shall", "must", "should", "not allowed", "subject to", "determined", "defined", "applies", "may",
    "calculates", "charged", "exceeds", "less than", "more than", "interest",
    "submits", "informs", "registers", "validates", "reports", "considered", "keeps",
    "interacts", "agree", "transmits", "gives", "receives", "provides"
}

operands_4536 = {
   "taxpayer", "regulatory authority", "resident", "non-resident", "enterprise",
   "tax", "income", "profit", "costs", "object of taxation", "cost", "tax rate",
   "tax base", "tax liability", "environmental tax",
   "declaration", "tax reporting", "inspection report", "message",
   "emission level", "marginal norm", "coefficient", "interest", "risk", "share",
   "method", "calculation procedure", "methodology", "approach",
   "state", "Ukraine", "territory", "international treaty", "foreign country"
}
```

```
# --- Article 39 of the Tax Code of Ukraine ---
operators_pku39 = {
  “if”, “in case”, “provided”, “excluding”, “except”, “when”, “and also”, “simultaneously”,
  “shall”, “must”, “should”, “may”, “subject to”, “determined”, “established”, “applies”, “not allowed”, “allowed”,
  “exceeds”, “equals”, “at least”, “no more than”, “rate”, “share”, “interest”, “difference”, “pro rata”, “coefficient”,
  “submit”, “informs”, “registers”, “calculates”, “tests”, “keeps”, “reports”, “applies the method”, “considers”, “makes a calculation”,
  “interacts”, “agreed”, “coordinates”, “transmits”, “provides”, “receives”, “informs”, “cooperates”, “exchanges information”
}

operands_pku39 = {
  “taxpayer”, “regulatory authority”, “related parties”, “non-resident”, “resident”,
  “authorized representative”, “legal entity”, “enterprise”, “participant in a controlled operation”,
  “price”, “market price”, “profit”, “income”, “costs”, “tax base”,
  “cost”, “royalty”, “asset”, “product”, “service”, “object of taxation”,
  “report on controlled transactions”, “transfer pricing documentation”, “declaration”,
  “inspection report”, “information about operations”, “request”, “justification”,
  “profitability level”, “interest”, “share of expenses”, “profitability indicator”,
  “correlation”, “price range”, “market value corridor”, “volume of transactions”,
  “comparative uncontrolled price method”, “resale price method”, “expense-plus method”,
  “net profit method”, “profit distribution method”, “arm’s length principle”, “country”, “Ukraine”,
  “foreign country”, “treaty”, “tax jurisdiction”, “competent authority”, “territory of Ukraine”
}

# Universal Halstead Analysis

def read_text(file_path):
  if file_path.endswith(“.docx”):
    doc = Document(file_path)
    text = “\n”.join(p.text for p in doc.paragraphs)
  else:
    with open(file_path, “r”, encoding=“utf-8”) as f:
      text = f.read()
  return text.lower()

def classify_tokens(text, operators, operands):
  tokens = [t for t in word_tokenize(text) if t.isalpha()]
  ops = [t for t in tokens if t in operators]
  oprs = [t for t in tokens if t in operands]
  return ops, oprs

def halstead_metrics(ops, oprs):
  N1 = len(ops)
  N2 = len(oprs)
  η1 = len(set(ops))
  η2 = len(set(oprs))
  V = N1 + N2
  V_star = 2 + η2
  L = round(V_star / V * 100, 3) if V > 0 else 0
  return {“N1”: N1, “N2”: N2, “η1”: η1, “η2”: η2, “V”: V, “V*”: V_star, “L (%)”: L}

# --- MAIN ---
if __name__ == “__main__”:
  print(“=== Halstead’s analysis of Ukrainian regulations===“)
  print(“Select a document to analyze:”)
  print(“1 - Law No. 466-IX”)
  print(“2 - Law No. 4536-IX”)
  print(“3 - Article 39 of TCU)

  choice = input(“Enter the number: “).strip()
  mapping = {
    “1”: (“ЗУ 466-IX”, operators_466, operands_466),
    “2”: (“ЗУ 4536-IX”, operators_4536, operands_4536),
    “3”: (“ПКУ ст.39”, operators_pku39, operands_pku39)
  }

  if choice not in mapping:
    print(“❌ Incorrect choice.”)
    exit()
```


http://dx.doi.org/10.21511/afc.07(1).2026.05

```
    name, ops_dict, oprs_dict = mapping[choice]
    file_path = input("Enter the path to the regulatory act text(.docx або .txt): ").strip()

    text = read_text(file_path)
    ops, oprs = classify_tokens(text, ops_dict, oprs_dict)
    metrics = halstead_metrics(ops, oprs)

    print(f"\n      Halstead analysis results: {name}")
    df = pd.DataFrame(metrics, index=[name])
    print(df.to_string())

    out_file = os.path.join(os.path.dirname(file_path), f"halstead_{name.replace(' ', '_')}.csv")
    df.to_csv(out_file, encoding="utf-8-sig")
    print(f"\n ✅ Results saved in: {out_file}")
```

OECD legislation

```
import os
import pandas as pd
from docx import Document
from nltk.tokenize import word_tokenize

# 1. WORD DICTIONARIES

# --- Operators and operands for OECD TPG (2017)
operators_tpg = {
    # Logical
    "if", "where", "when", "except", "provided", "unless", "and", "or", "not", "without",
    # Normative
    "shall", "must", "should", "may", "can", "deemed", "considered", "treated", "defined",
    # Mathematical
    "equals", "greater", "less", "exceeds", "percentage", "ratio",
    # Procedural
    "apply", "calculate", "determine", "verify", "report", "review", "assess", "adjust",
    # Coordination
    "exchange", "provide", "cooperate", "implement"
}

operands_tpg = {
    # Subjects
    "taxpayer", "enterprise", "associated", "entity", "administration",
    # Economic items
    "income", "profit", "loss", "cost", "price", "transaction",
    # Documents
    "documentation", "report", "file", "record", "data",
    # Indicators
    "value", "range", "arm's", "length", "ratio", "benchmark",
    # Methods
    "cup", "tnmm", "resale", "cost", "plus", "profit", "split", "method",
    # Jurisdictions
    "country", "state", "jurisdiction", "treaty"
}

# --- Operators and operands for Pillar One (2020)
operators_pillar = {
    # Logical
    "if", "when", "where", "unless", "notwithstanding", "provided", "and", "or", "not",
    # Normative
    "shall", "must", "should", "may", "can", "deemed", "determined", "defined",
    # Mathematical
    "equals", "exceeds", "threshold", "percentage", "share", "margin",
    # Procedural
    "compute", "assess", "allocate", "review", "reconcile",
    # Coordination
    "coordinate", "exchange", "implement", "report", "provide"
}

operands_pillar = {
    # Subjects
    "mne", "group", "market", "jurisdiction", "lead", "administration", "panel",
```

```
    # Economic items
    "profit", "residual", "routine", "return", "revenue", "sales",
    # Documents
    "statement", "filing", "report", "dataset",
    # Indicators
    "nexus", "threshold", "revenue", "profitability", "ratio",
    # Methods
    "amount", "a", "b", "allocation", "key", "segmentation",
    # Jurisdictions
    "market", "residence", "jurisdiction", "economy"
}

# 2. HELPER FUNCTIONS

def read_text(file_path: str) -> str:
    """Read text from .txt or .docx file"""
    if file_path.endswith(".docx"):
        doc = Document(file_path)
        text = "\n".join(p.text for p in doc.paragraphs)
    else:
        with open(file_path, "r", encoding="utf-8") as f:
            text = f.read()
    return text.lower()

def classify_tokens(text: str, operators: set, operands: set):
    """Tokenize and classify operators and operands"""
    try:
        tokens = word_tokenize(text)
    except LookupError:
        import nltk
        nltk.download("punkt")
        tokens = word_tokenize(text)
    ops = [t for t in tokens if t in operators]
    oprs = [t for t in tokens if t in operands]
    return ops, oprs

def halstead_metrics(ops, oprs):
    """Calculate Halstead metrics"""
    N1, N2 = len(ops), len(oprs)
    η1, η2 = len(set(ops)), len(set(oprs))
    V = N1 + N2
    V_star = 2 + η2
    L = round(V_star / V * 100, 3) if V > 0 else 0
    return {
        "N1 (operators)": N1,
        "N2 (operands)": N2,
        "η1 (unique operators)": η1,
        "η2 (unique operands)": η2,
        "V = N1 + N2": V,
        "V* = 2 + η2": V_star,
        "L = V*/V (%)": L
    }

def analyze_file(file_path, doc_type):
    """Analyze text using the chosen dictionary"""
    text = read_text(file_path)
    if doc_type == "TPG":
        ops, oprs = classify_tokens(text, operators_tpg, operands_tpg)
    else:
        ops, oprs = classify_tokens(text, operators_pillar, operands_pillar)
    metrics = halstead_metrics(ops, oprs)
    df = pd.DataFrame(metrics, index=[f"{doc_type}_{os.path.basename(file_path)}"])
    print(f"\n=== HALSTEAD ANALYSIS: {doc_type} ===")
    print(df.to_string())
    return df

# 3. MAIN SCRIPT

if __name__ == "__main__":
    import nltk
```


http://dx.doi.org/10.21511/afc.07(1).2026.05

```
try:
    nltk.data.find("tokenizers/punkt")
except LookupError:
    nltk.download("punkt")

print("Select the document type to analyze:")
print("1 - OECD Transfer Pricing Guidelines (TPG, 2017)")
print("2 - OECD Pillar One Blueprint (2020)")
choice = input("Enter 1 or 2: ").strip()

if choice == "1":
    doc_type = "TPG"
elif choice == "2":
    doc_type = "PillarOne"
else:
    print("❌ Incorrect choice.")
    exit()

file_path = input("Enter the path to the file(.txt або .docx): ").strip()
if not os.path.exists(file_path):
    print(f"\n❌ File not found: {file_path}")
else:
    results = analyze_file(file_path, doc_type)
    output_file = os.path.join(os.path.dirname(file_path), f"halstead_{doc_type}_results.csv")
    results.to_csv(output_file, encoding="utf-8-sig")
    print(f"\n✅ Results saved to file: {output_file}")
```