UDC 81’1:81’37:81’246.2:81’27
DOI https://doi.org/10.26661/2414-1135-2026-103-30

Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0

CONTRASTIVE CORPUS ANALYSIS OF THE NOTION *DIGNITY* IN UKRAINIAN AND ENGLISH LINGUOCULTURES

Fabian M. P.
Doctor of Philology, Professor,
Professor at the English Philology Department
Uzhhorod National University
Universytetska str., 14, Uzhhorod, Ukraine
orcid.org/0000-0002-3351-1275
myroslava.fabian@uzhnu.edu.ua

Key words: *corpus linguistics, contrastive analysis, the notion of dignity, collocate, lemma, social value.*

The current paper provides a full-scale comparative corpus-based analysis of the notion *dignity* and its linguistic objectivization in modern Ukrainian and English linguistic cultures. With the help of significant quantitative data collected from large-size reference Web corpora using Sketch Engine, this paper achieves successful operationization of abstract social notions into lexical grammatical features which include adjectival modifications, verbal collocations, nominal collocates, genitive modifications, and synonymic sets. Advanced statistical association techniques, including the LogDice metric, as well as inter-coder reliability testing based on Cohen’s Kappa, guarantee the empirical soundness, methodological validity, and reliability of the results. Specifically, the study calculated distributional frequencies and association metrics for adjectival modifiers, extracted top verbal collocates categorized into violation and protection patterns, analyzed nominal heads and genitive possessors such as the Revolution of Dignity and social roles, and established structured synonymic fields for both lemmas. Analysis of the comparative nature of the lexemes *гідність* and *dignity* shows that although both languages share the anthropocentrism of their cultures, which means that there is a connection between the concept of dignity and human being, the differences are evident in terms of cultural models and cognitive strategies. Whereas in the Ukrainian linguistic tradition the notion *гідність* is dynamic, anthropocentric and conflictive in nature, connected with collective and national identity, as well as historical events such as the Revolution of Dignity, and is linked with such concepts as *honor* and *pride*, in the English linguistic tradition, the universal notion *dignity* is static in nature and reflects a legal and procedural category related to the protection of one’s status within the context of human rights. The differences in the lexical semantic nature of the notions under study show the influence of different historical and social experiences on the linguistic expression of universal concepts and can be used in practice of intercultural communication, sociolinguistics and professional translation studies.

ЗІСТАВНИЙ КОРПУСНИЙ АНАЛІЗ ПОНЯТТЯ *ГІДНІСТЬ* В УКРАЇНСЬКІЙ ТА АНГЛІЙСЬКІЙ ЛІНГВОКУЛЬТУРАХ

Фабіан М. П.
доктор філологічних наук, професор,
професор кафедри англійської філології
Ужгородський національний університет
вул. Університетська, 14, Ужгород, Україна
orcid.org/0000-0002-3351-1275
myroslava.fabian@uzhnu.edu.ua

Ключові слова: *корпусна лінгвістика, зіставний аналіз, поняття гідність, колокат, лема, соціальна цінність.*

У статті здійснено комплексний зіставний корпусний аналіз поняття *гідність* та його мовної об'єктивації в сучасних українській і англійській лінгвокультурах. На основі значного масиву кількісних даних, отриманих із великих еталонних вебкорпусів за допомогою платформи Sketch Engine, реалізовано успішну операціоналізацію абстрактного соціального поняття через його лексико-граматичні характеристики, зокрема прикметникові модифікатори, дієслівні колокації, іменникові колокати, генітивні модифікації та синонімічні ряди. Використання сучасних статистичних методів оцінювання асоціативної сили, зокрема метрики LogDice, а також перевірка міжекспертної узгодженості за коефіцієнтом Каппи Коена (Cohen's Kappa) забезпечили емпіричну обґрунтованість, методологічну валідність і надійність отриманих результатів.

Зокрема, у дослідженні було визначено частотність уживання та асоціативні показники прикметникових модифікаторів, виокремлено найтиповіші дієслівні колокати, класифіковані за моделями порушення та захисту гідності, проаналізовано іменникові вершини словосполучень і генітивні означення, зокрема «Революція Гідності» та позначення соціальних ролей, а також сформовано структуровані синонімічні поля для обох досліджуваних лем.

Порівняльний аналіз лексем *гідність* і *dignity* засвідчив, що, попри спільний антропоцентричний характер обох досліджуваних мов та культур, який проявляється у взаємозв'язку поняття *гідності* з людиною, між ними існують істотні відмінності, зумовлені специфікою культурних моделей і когнітивних стратегій. Якщо в українській мовній традиції поняття *гідність* має динамічний, антропоцентричний і конфліктогенний характер та пов'язується з колективною та національною ідентичністю, історичними подіями, зокрема Революцією Гідності, а також концептами *честь* і *гордість,* то в англійській мовній традиції універсальне поняття *dignity* має переважно статичний характер і репрезентує насамперед правову та процедурну категорію, пов'язану із захистом статусу особи в контексті прав людини.

Виявлені лексико-семантичні відмінності між цими поняттями демонструють вплив різного історичного й соціального досвіду на мовне вираження універсальних цінностей і можуть бути використані у практиці міжкультурної комунікації, соціолінгвістики та перекладознавства.

Problem statement. The language is not just an image of reality; it creates it, placing the values orientation of the people in its lexical grammatical structures. At the present stage of development, when the world is becoming increasingly interconnected, the analysis of the value system is especially urgent under the conditions of intercultural communication and translation, when the incorrect reproduction of the semantic value of socially important terms distorts the content of sociopolitical and philosophical texts. One of the central terms that serves as the basis for the formation of the moral values of European culture is *dignity*; hence, the objective of modern linguistics is to objectify this notion.

The methodological framework of this research relies on the transition from introspective linguistics

to empirical corpus-based analysis [Stefanowitsch, 2020, p. 15]. Working with large text arrays from platforms such as Sketch Engine ensures the validity and reliability of findings, moving research beyond subjective assumptions. Axiological linguistics is closely connected with the theory of intercultural communication, as interaction between bearers of different traditions requires an understanding of core values that structure a nation's worldview [Dinu, 2025, p. 5]. Regarding the notion itself, researchers distinguish between inherent dignity, which belongs to a person by right of birth, and acquired dignity, formed through social interaction [Yıldırım, Akın, 2026, p. 110].

Despite an extensive corpus of literature devoted to theoretical and philosophical perspectives, there is an extremely limited number of contrastive corpus-based investigations regarding the issue of *dignity* in Ukrainian and English linguistic cultures. Earlier monographic investigations have already demonstrated that such values as homeland, peace, family, or work can be considered as the accurate markers of cultural identity, but, despite its constitutional and public importance, the notion of *dignity* has not been investigated quantitatively and systematically from the perspective of contrasting the two cultures.

The **aim** of this paper is to carry out a corpus-based comparative analysis of *dignity* in the Ukrainian and English linguistic cultures and to build nation-specific models of its conceptualization.

In order to achieve the goal, the following tasks are to be performed:

1) to operationalize the theoretical notion of the social value via a certain number of measurable lexico-grammatical features;

2) to develop and verify a method of automatic extraction and annotation of data from the corpora with estimation of inter-coder reliability (using Cohen's Kappa coefficient);

3) to quantitatively analyze the distribution of modifiers, collocations and synonym rows of the lexemes *гідність* and *dignity*;

4) to interpret the discovered quantitative differences from the perspective of culturally specific models;

5) to make some recommendations concerning translation in the light of proper transfer of the notion in the interlingual communication.

The preliminary assumption is that there are significant differences in distributional properties of the Ukrainian lexeme *гідність* and the English *dignity* which indicate substantial differences in their conceptual content.

The **object** of the research is the notion of *dignity* as a verbalized social value in Ukrainian and English linguistic cultures.

The **subject** of the study comprises the lexico-semantic and lexico-grammatical features of the functioning of the lexemes *гідність* and *dignity* in contemporary Ukrainian and English web corpora, specifically: frequency and association strength (LogDice) of adjectival modifiers, verbal collocations (lemma in object and subject positions of action), genitive modifiers, and synonym rows of both lexemes.

Materials and methods. For the purposes of this study, a corpus-based contrastive analysis is used as a tool for the investigation of the conceptualisation of *dignity* in modern Ukrainian and English linguistic communities. As opposed to an intuitive introspective methodology of lexical semantics, the study makes use of empirical data drawn from the reference corpora available on the internet in order to ensure high methodological validity and reproducibility. Overall methodology combines corpus data retrieval, association measures, lexico-grammatical operationalisation, semantic annotation and inter-coder reliability testing in order to construct a precise cognitive model of the notion under discussion.

The main comparative framework of the research includes the study of two target lexemes: *гідність* in Ukrainian and *dignity* in English. In order to ensure high cross-linguistic validity of the results obtained, the study considers five different syntagmatic and paradigmatic aspects presented in Table 1 below. The combination of frequency and association allows to evaluate socio-cultural values embedded in linguistic structures.

The empirical basis of the present study is built on large web corpora available via the Sketch Engine. The Ukrainian corpus is taken from *uktenten22_rft2*, while the English one from *ententen21_tt31*. Both corpora have millions of running words with examples of modern usage of the language in different styles, such as journalistic, scientific, legal, and public.

In order to create a statistically solid base, frequency and dispersion values were determined for each of

Table 1

Summary of Corpus Extraction Parameters and Overall Frequencies

Linguistic culture	Target lemma	Tokens	Primary grammatical relations analyzed
Ukrainian	гідність	107,037	Adjectival modifiers, nominal heads, object/subject verbs, genitive modifiers, synonyms
English	dignity	174,830	Adjectival modifiers, nominal heads, object/subject verbs, genitive modifiers, synonyms

the examined lemmas. Collocations in the specified grammatical relations were extracted through Word Sketch searches. In order to exclude insignificant collocation pairs, the strength of association was computed with the help of the LogDice measure, which is not affected by the size of the corpus and has a value from 0 to 14 [Rychlý, 2008, p. 6].

Abstract social values can be operationalized by the translation of philosophical and sociological notions into measurable lexico-grammatical variables. In this research, the dependent variable is understood as the contextual lexical environment of the target lemma, operationalized in terms of unique syntactic slots, whereas the independent variable is the concrete linguistic culture. Data extraction process involved the automated processing of Word Sketch algorithms and manual exclusion of any parsing errors like proper names and other parsing noise. After the automatic data extraction, all ambiguous or polysemantic concordance lines were manually annotated. A coding system was created in order to establish mutually exclusive semantic categories. For example, verbal collocations could be categorized into the following two macro-semantic classes: (1) Violation / Infringement (*принижувати, ображати, violate, undermine*) and (2) Protection / Maintenance (*поважати, відстоювати, respect, preserve, restore*). In order to minimize researcher bias and provide for internal validity, each concordance line was independently coded by two linguist experts. The classification rubric was piloted with a test sample of 1,000 concordances per language to further define category boundaries and create exhaustive operational definition. Inter-coder reliability was formally assessed with the help of Cohen's Kappa coefficient (k), calculated with the following formula:

$$\kappa = \frac{P_o - P_e}{1 - P_e}$$

where P_o represents the relative observed agreement among raters, and P_e denotes the hypothetical probability of chance agreement.

The inter-rater reliability coefficients obtained in all categories varied from 0.82 to 0.89, showing very high to perfect concurrence between the two coders based on standard methodological thresholds [O'Connor & Joffe, 2020]. Any disagreements between the coders were negotiated, and the codes re-examined using the criteria outlined in the coding manual.

The final coded data set was arranged into a matrix based on the lexico-grammatical context and measured in terms of frequency and association as well as semantics categories. The statistical analysis involved the comparison of proportions in the different categories using the chi-square test of independence in order to confirm the statistical significance of the observed differences between cultures, thereby making the empirical results completely objective, reproducible, and verifiable.

Presentation of the main material. The empirical results obtained as the outcome of the contrastive corpus analysis allow for presenting quantitative evidence on the structural, semantic and pragmatic differences in the conceptualization of *dignity* in the Ukrainian and English linguistic cultures. Exploring the adjectival modification, co-occurrence of verbs, collocation with nouns and synonym relations, the research provides an objective basis for evaluation of the linguistic entrenchment of cultural values.

The quantitative distribution of adjectives modifying the word under investigation points to both anthropocentric universals and culture-specific priorities (see Table 2).

In view of the absolute predominance of modifiers referring to humanness (*людський / human*), we may conclude about the existence of a common universal semantic core, but hierarchical orders give grounds for identifying the specific conceptual orientations. Thus, while the Ukrainian *гідність* prioritizes possession and collectivity (*власний, національний, особистий*), the English *dignity* tends to emphasize transcendence and innateness (*inherent, god-given, innate*) along with the legal-egalitarian aspect (*equal*). Moreover, the Ukrainian top tier is characterized by the prominence of the adjective *принижуючий*, which demonstrates the essential experiential connection of the notion with its devaluation contexts, a phenomenon absolutely foreign to the core adjectival structure of the English *dignity*.

It is noteworthy that the absolute leader in frequency in both languages is the adjective denoting 'human' (*людський / human*), which confirms the shared anthropocentric basis of the notion. Divergence emerges at the second rank: the Ukrainian lexeme tends toward modifiers of ownership and collective belonging (*власний, національний, особистий*), whereas its English counterpart gravitates toward modifiers of innateness and transcendence (*inherent, god-given, innate* – further down the list) alongside the modifier equal, indicating a parallel link between

Table 2

Adjectival modifiers

Rank	гідність (LogDice)	dignity (LogDice)
1	людський – 26420 (9.50)	inherent – 3869 (7.84)
2	принижуючий – 737 (7.80)	human – 58574 (7.46)
3	власний – 16257 (7.24)	god-given – 598 (6.51)
4	національний – 11760 (6.84)	quiet – 3156 (6.45)
5	особистий – 436 (6.78)	equal – 2708 (5.94)

the notion and the legal category of equality. Notably, the inclusion of the adjective *принижуючий* in the top-2 Ukrainian list serves as a direct lexical trace of the frequent collocational pairing of *гідність* with the semantic field of degradation, which is absent from the English material.

The qualitative distribution of the top-20 verbal collocates of each language into two semantic groups ('violation / infringement', including *принижувати, ображати, топтати, зневажати*, *violate, offend, undermine, insult*, etc., and 'protection / maintenance', including *поважати, відстоювати, зберегти, respect, preserve, uphold, restore, etc*.) yielded the following quantitative results (see Table 3).

Table 3

Verbal collocates (X in the object position)

Rank	гідність (freq, LogDice)	dignity (freq, LogDice)
1	принижувати – 6904 (12.21)	respect – 5856 (7.63)
2	ображати – 1052 (9.51)	uphold – 2830 (7.42)
3	поважати – 1164 (8.72)	affirm – 1471 (6.92)
4	принизити – 420 (8.65)	restore – 5609 (6.74)
5	принижуючи – 237 (8.01)	preserve – 4045 (6.43)

This is the most statistically convincing result of the study. It empirically confirms the hypothesis regarding the dynamically conflictual nature of the Ukrainian notion, where *гідність* is primarily mentioned in contexts of threats to it, as opposed to the statically protective nature of the English notion. Specifically, *dignity* is predominantly maintained or restored, with verbs such as *regain*, *reclaim*, and *restore* pointing to a legal and procedural model of 'restoring a lost status' that differs both from purely 'innate immutability' and from the Ukrainian model of 'continuous struggle'.

The "nouns modified by X" category revealed the most culturally significant outcome for the Ukrainian lexeme (see Table 4). The absolute leader in terms of LogDice (12.11) and the second highest in frequency (21,477 token occurrences) is the lexeme *революція*, which directly refers to the official designation of the 2013–2014 events in Ukraine known as the Revolution of Dignity (*Революція Гідності*). This is followed by *приниження* (3,694; LogDice 10.28), *почуття* (15,516; 9.26), *повага* (806; 7.13), *марш* (243; 6.55, referring to the March of Dignity / Марш Гідності), *утвердження*, and *відстоювання*. The English material demonstrates no comparable phenomenon, namely, a specific historical event entrenched within the very name of the notion, where abstract categories such as *self-respect, equality, fairness, compassion, decency, autonomy,* and *privacy* clearly dominate.

Table 4

Nominal collocates (nouns modified by X)

Rank	гідність (freq, LogDice)	dignity (freq, LogDice)
1	Революція – 21477 (12.11)	self-respect – 246 (7.25)
2	приниження – 3694 (10.28)	respect – 2759 (6.86)
3	почуття – 15516 (9.26)	equality – 1412 (6.60)
4	повага – 806 (7.13)	fairness – 252 (6.31)
5	Марш – 243 (6.55)	compassion – 445 (6.30)

A similar trend is confirmed by the category of genitive modifiers / possessors (see Table 5). For the Ukrainian lexeme, the second-highest item by LogDice is *українець* (1,515; 6.45), reflecting the fixed collocation *гідність українця*, preceded only by the more general terms *людина* (9,527; 6.38) and *особа* (3,384; 6.24). In contrast, the English category of possessors of X is characterized by a legal and human rights register featuring items such as *detainee*, *refugee*, *resident*, and *migrant*, alongside general terms like *being* (5.92), *person* (993; 5.47) and *individual* (331; 5.00). Some results, such as proper names, represent parsing noise that requires manual filtration prior to final analysis.

Table 5

Genitive modifiers / possessors

Rank	гідність (freq, LogDice)	dignity (freq, LogDice)
1	поводження – 448 (7.01)	being – 38 (5.92)
2	українець – 1,515 (6.45)	detainee – 23 (5.66)
3	людина – 9,527 (6.38)	person – 993 (5.47)
4	особа – 3,384 (6.24)	refugee – 22 (5.23)
5	особистість – 1,015 (6.11)	resident – 121 (5.22)

The empirical data introduces a substantial correction to the previously stated thesis (see Table 6). For the English lexeme, the leading synonym is not *integrity* (LogDice 7.37, ranking only 15th), but *respect* (LogDice 10.14, freq 33,005), which significantly outperforms the rest of the series (worth 8.91; honor 8.67; grace 8.23; equality 8.22; freedom 8.20). For the Ukrainian lexeme, the primary synonym is *честь* (LogDice 13.09, freq 29,568), substantially surpassing *свобода* (10.56), *гордість* (8.59), *самоповага* (8.43), and *повага* (7.96).

Thus, both languages deploy a synonymic field around the category of 'respect' (*честь/повага* ↔ *respect*), yet with varying association strength. The Ukrainian link 'гідність ↔ честь' (LogDice 13.09) is statistically stronger than the English 'dignity ↔

Table 6

Synonymic field (category "X and/or…")

Rank	гідність (freq, LogDice)	dignity (freq, LogDice)
1	честь – 29568 (13.09)	respect – 33005 (10.14)
2	свобода – 6002 (10.56)	worth – 5624 (8.91)
3	гордість – 836 (8.59)	honor – 6013 (8.67)
4	самоповага – 580 (8.43)	grace – 5544 (8.23)
5	повага – 931 (7.96)	equality – 6151 (8.22)

respect' (LogDice 10.14), which confirms a higher degree of semantic cohesion between the Ukrainian notion and the category of honor.

The quantitative word sketch data unequivocally support the core hypothesis of the study, providing a significantly more precise empirical foundation than intuitive observations. The Ukrainian conceptualization of *dignity* is statistically and lexically closely tied to eventfulness, humiliation, and collective-national identity (Revolution of Dignity, dignity of a Ukrainian, dominance of verbs of humiliation), whereas the English conceptualization is associated with the legal-procedural maintenance and restoration of an individual's status within human rights discourse (respect, preserve, restore, detainee, refugee).

Conclusions. The conducted study confirms the high efficiency of applying corpus linguistics methodology to the study of abstract social values. In accordance with the principles of objective linguistic research, systematic quantitative analysis of the distribution of linguistic phenomena in texts makes it possible to avoid the subjectivity of introspective judgments and obtain empirically reliable results. The operationalization of the theoretical construct of value through specific lexico-grammatical patterns has proven its complete validity, transforming philosophical categories into measurable linguistic parameters suitable for statistical evaluation.

The contrastive analysis of the lexemes *гідність* and *dignity* using corpus tools revealed fundamental divergences in their conceptual content despite a shared denotative basis. The quantitative analysis shows that the notion of dignity in the Ukrainian language culture is characterized by a very anthropocentric and dynamic nature. The frequent co-occurrence of this word with the adjectives *national* and *own* and the verbs *defend*, *protect*, or *preserve* reveals that the Ukrainians perceive this notion as something to constantly fight for, as well as an important attribute of collective identity. Close synonymic relations with the notions of *honour* and *pride* also add an emotional aspect to *dignity*.

English-language discourse demonstrates a fundamentally different cultural model. The lexeme *dignity* functions predominantly as a universal, philosophical, and abstract-legal category. The statistical significance of the modifiers *inherent* and *innate* in combination with the verbs *preserve* and *maintain* shapes the perception of *dignity* as an innate static quality that merely needs to be maintained and safeguarded. The direct connection of the English lexeme with concepts such as *integrity*, *liberty*, and *compassion* testifies to the integration of this value into the general system of liberal democratic ideals, devoid of the acute emotional colouring of active resistance.

The results of our work prove that the universal human values are objectified in different languages in entirely different ways, directly reflecting the unique historical and social experience of a specific nation. Language does not simply transmit information about abstract values; rather, it forms the specific motivational infrastructure of society through the allocation of attention and the prioritization of certain semantic fields. The identified lexical semantic discrepancies require special attention in the fields of intercultural communication and translation studies, since the adequate transfer of meanings is possible only under the condition of a deep understanding of underlying cultural models. Prospects for further research are envisioned in engaging a wider spectrum of social, vital, and political values in contrastive corpus analysis, which will allow for the creation of a comprehensive semantic map of value orientations across different linguistic cultures.

BIBLIOGRAPHY

1. Dinu E. Exploring the Role of Intercultural Communication and Cultural Intelligence in Enhancing Stakeholder Engagement and ESG Practices in International Organisations, Insights from the UN. *Sustainability*. 2025. Vol. 17, No. 22. Article 10388. DOI: https://doi.org/10.3390/su172210388
2. O'Connor, C., & Joffe, H. Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, 2020. 19, 1–13. https://doi.org/10.1177/1609406919899220
3. Rychlý P. A A Lexicographer-Friendly Association Score. *Proceedings of the 6th Conversational Workshop on Recent Advances in Slavonic Natural Languages Processing.* 2. vyd. Brno: Masarykova Univerzita, 2008, p. 6–9.
4. Sketch Engine. (n.d.). English web corpus ententen21_tt31 [Data file and software]. Lexical Computing. https://www.sketchengine.eu
5. Sketch Engine. (n.d.). Ukrainian web corpus uktenten22_rft2 [Data file and software]. Lexical Computing. https://www.sketchengine.eu

6. Stefanowitsch A. Corpus linguistics, A guide to the methodology. Berlin : Language Science Press, 2020. 508 p.
7. Yıldırım D., Akin E. The Core of Caring, Human Dignity in Nursing Through a Constructivist Grounded Theory. *Nursing Ethics*. 2025. Vol. 32, No. 1. DOI: https://doi.org/10.1111/scs.70222

REFERENCES

1. Dinu, E. (2025). Exploring the role of intercultural communication and cultural intelligence in enhancing stakeholder engagement and ESG practices in international organisations, insights from the UN. *Sustainability*, Vol. 17, No. 22. Article 10388. https://doi.org/10.3390/su172210388
2. O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods*, *19*, 1–13. https://doi.org/10.1177/1609406919899220
3. Rychlý, P. (2008). A lexicographer-friendly association score. In *Proceedings of the 6th Conversational Workshop on Recent Advances in Slavonic Natural Languages Processing* (pp. 6–9). Masaryk University.
4. Sketch Engine. (n.d.). *English web corpus ententen21_tt31* [Data file and software]. Lexical Computing. https://www.sketchengine.eu
5. Sketch Engine. (n.d.). *Ukrainian web corpus uktenten22_rft2* [Data file and software]. Lexical Computing. https://www.sketchengine.eu
6. Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press. https://doi.org/10.5281/zenodo.3735824
7. Yıldırım, D., & Akin, E. (2025). The core of caring, human dignity in nursing through a constructivist grounded theory. *Nursing Ethics*, Vol. 32, No. 1. https://doi.org/10.1111/scs.70222

Дата першого надходження статті до видання: 13.07.2026
Дата прийняття статті до друку після рецензування: 16.08.2026
Дата публікації (оприлюднення) статті: 18.11.2026