Between Norm and Narration: A Corpus-Based Comparative Analysis of German and Uzbek Legal Press Language

[1]Yokubov Ulugbek Yokubkhon , [2]Mamadayupova Vasila Shonazarovna, [3]Sodikov Akmal Shavkat ugli, [4]Sultonova Umeda Istatovna, [5]Firuz Tuygunovich Rashidov, [6]Gaffarov Usmon Bobonazarovich,

[1]Lecturer in English for Specific Purposes, Department of Foreign Languages, Tashkent State University of Law ORCID: 0009-0004-3858-5240 u.yokubov@tsul.uz

[2]Acting Associate Professor of English for Specific Purposes Department of Foreign Languages, Tashkent State University of Law v.mamadayupova@mail.ru ORCID: 0000-0001-7912-6295

[3]PhD in Law, Associate professor of Tashkent State University of Law E-mail. a.sodikov@tsul.uz ORCID 0000-0001-6453-1327

[4]umedasultonova303@gmail.com usultanova@uwed.uz Jahon Iqtisodiyoti va Diplomatiya Universiteti(JIDU) ORCID: 0009-0009-9748-5780

[5]PhD in Law, Associate Professor Department of Social Sciences and Law, International Islamic Academy of Uzbekistan, Tashkent, Uzbekistan ORCID: 0000-0003-3181-9847 E-mail: firuzrashidov7@gmail.com

[6]Head of the Department of Maxillofacial Surgery at Samarkand State Medical University, PhD, docent, https://orcid.org/0009-0009-7559-4046 ORCID ID Email: dr.gaffarovusmov@gmail.com

HOW TO CITE:

Yokubov Ulugbek Yokubkhon , Mamadayupova Vasila Shonazarovna, Sodikov Akmal Shavkat ugli, Sultonova Umeda Istatovna, Firuz Tuygunovich Rashidov, Gaffarov Usmon Bobonazarovich, (2026). Between Norm and Narration: A Corpus-Based Comparative Analysis of German and Uzbek Legal Press Language. International Journal of Special Education, 41(12s), 1186-1199

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors. Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:

Despite the expanding body of cross-linguistic corpus studies in legal discourse, the language of legal press, a hybrid genre mediating between juridical normativity and journalistic narration, remains comparatively understudied, particularly in language pairs that span typologically distant families. No prior corpus-based study has systematically compared German and Uzbek legal press language, a gap that is notable given their divergent morphological profiles (inflectional vs. agglutinative), distinct legal traditions (Romano-Germanic civil law vs. Uzbek legal reform), and markedly different levels of representation in international corpus linguistic research. The present study addresses this gap through a purpose-built bilingual corpus of 120,000 tokens (60,000 per language), drawn from established German legal press outlets (Legal Tribune Online, beck-aktuell) and Uzbek legal periodicals (Huquq va Hayot, Yurist), covering the period 2019–2023. Using AntConc 4.2.0 and Sketch Engine, the analysis examines lexical density, standardized type-token ratio (STTR), nominalization frequency, mean sentence length, keyword profiles, and hedging devices. The findings reveal that German legal press exhibits significantly higher nominalization density (M = 18.7 per 100 tokens) and longer mean sentence length (M = 24.3 words) than its Uzbek counterpart (M = 11.2 and M = 16.8, respectively), while Uzbek legal press displays greater lexical variation and a higher proportion of epistemic hedges. These patterns are interpreted through register theory (Halliday & Matthiessen, 2004) and genre theory (Swales, 1990; Bhatia, 1993). The study contributes the first

quantitative baseline for Uzbek legal press language and offers implications for legal translation pedagogy, terminology standardization, and corpus-assisted language teaching in Central Asian legal contexts.

Keywords: legal discourse; corpus linguistics; German; Uzbek; comparative analysis; register; legal press language

1. Introduction

Legal language has long attracted scholarly attention as a register characterized by lexical precision, syntactic complexity, and institutional embeddedness. Yet the genre of legal press, that is, journalism specifically oriented toward legal affairs, legislative developments, and judicial commentary, occupies an analytically ambiguous position: it is neither strictly statutory in its normativity nor entirely journalistic in its register. Legal press texts negotiate the boundary between the specialist discourse of the legal community and the communicative demands of a broader, semi-specialist audience, producing what Bhatia (1993) terms a "hybrid genre" whose linguistic properties have not been systematically investigated across languages with markedly different typological and cultural profiles.

German legal journalism represents a well-institutionalized tradition, anchored in publications such as the Neue Juristische Wochenschrift (NJW) and, in the digital era, outlets including Legal Tribune Online and beck-aktuell. These sources address audiences of legal professionals, law students, and informed citizens, and they operate within the Romano-Germanic civil law tradition that has historically exerted influence across continental Europe and beyond, including in Central Asian states undergoing legal reform. Uzbekistan, having declared independence in 1991, has since engaged in an ongoing process of legal modernization that incorporates elements of German-inspired codification (Saidov, 2012), making the German-Uzbek pairing not merely typologically interesting but historically and institutionally relevant.

Uzbek, an agglutinative Turkic language, differs from German not only morphologically, in its use of suffixal morphology to encode grammatical relations that German expresses through inflection, but also in the extent to which its legal press has been examined through corpus linguistic methods. To the authors' knowledge, no peer-reviewed corpus study has focused specifically on Uzbek legal press language, and the broader field of Central Asian legal linguistics remains underrepresented (cf. Aminova, 2018; Nazarov, 2021). This lacuna has both theoretical and practical consequences: it limits the applicability of genre-based accounts of legal discourse to the full range of world legal languages, and it leaves practitioners of legal translation between German and Uzbek without an empirically grounded description of register-specific contrasts.

The present study aims to address this gap by conducting a corpus-based comparative analysis of German and Uzbek legal press texts, drawing on frameworks from register theory, genre theory, and corpus-based contrastive linguistics. The study constructs a purpose-built bilingual corpus and applies quantitative and qualitative corpus methods to investigate three dimensions of comparison: lexical complexity, syntactic structure, and discourse organization.

The following research questions guide the investigation:

RQ1: What differences in lexical density and standardized type-token ratio characterize German and Uzbek legal press texts?

RQ2: How do nominalization patterns and syntactic complexity (as measured by mean sentence length and clause embedding) differ across the two corpora?

RQ3: What genre-specific discourse markers and hedging devices are characteristic of each language's legal press, and what do these

distributions suggest about the rhetorical orientation of each tradition?

2. Theoretical Framework and Literature Review

2.1 Register Theory and Legal Press

The concept of register, as developed within the systemic functional linguistic tradition by Halliday (1978) and elaborated in Halliday and Matthiessen (2004), provides a foundational framework for understanding how language choices vary systematically with situational context. Register is analyzed along three metafunctional dimensions: field (the domain of activity), tenor (the relationship between participants), and mode (the channel and role of language). Applied to legal press, these dimensions reveal a distinctive configuration: the field is legal-institutional (legislative, judicial, or regulatory), the tenor is asymmetrical but bridging (a specialist author addressing a semi-specialist audience), and the mode is written-to-be-read rather than written-to-be-acted-upon, as in statutory texts.

Biber's (1988) multidimensional analysis of register variation provides an empirical complement to functional approaches, demonstrating that registers cluster along statistically derived dimensions that cut across traditional genre boundaries. Legal language consistently patterns toward high informational density, complex nominal phrases, and low interactional involvement — a profile that distinguishes it from journalistic prose (Biber & Conrad, 2009). Legal press, as a hybrid genre, may be expected to display intermediate or variable patterns along these dimensions, making it a theoretically productive object of comparative study.

2.2 Genre Theory and the Hybrid Status of Legal Press

Genre theory, as developed by Swales (1990) and extended by Bhatia (1993, 2004) in the context of professional discourse, conceptualizes genres as recognizable communicative events with shared purposes, structured conventions, and community-specific rhetorical moves. Legal texts have been analyzed as genres with highly conventionalized move structures: statutes follow definitional and prescriptive moves; judicial decisions follow case-summary, ratio-decidendi, and obiter dictum sequences (Bhatia, 1993).

Legal press texts, by contrast, are less amenable to a single move structure. They may combine elements of legal commentary (interpretive moves), news reporting (narrative moves), and opinion journalism (evaluative moves). This generic hybridity has been noted in analyses of English-language legal journalism (Conley & O'Barr, 2005) and in studies of EU legal press (Clay, 2026), but it has not been systematically examined in non-European legal press languages. The present study treats German and Uzbek legal press texts as instantiations of a shared genre that may be realized differently according to typological and cultural constraints.

2.3 Corpus-Based Contrastive Linguistics and Legal Language

Corpus-based contrastive linguistics, as consolidated by McEnery and Hardie (2012) and previously by Biber, Conrad, and Reppen (1998), employs comparable corpora, texts of similar type and domain in different languages, to identify systematic cross-linguistic differences in frequency, distribution, and co-occurrence of linguistic features. This approach has been applied productively in legal contexts: studies have compared English and Italian legal phraseology (Biel, 2014), German and Finnish court judgments (Lindroos, 2015), and the EU Eurolect across multiple national varieties (Biel, Wasilewska, & Kozbial, 2026).

The methodology of keyword analysis, which identifies lexical items significantly overrepresented in a target corpus relative to a reference corpus (Scott, 1997; Rayson, 2008), has proven particularly valuable in legal corpus linguistics. Log-likelihood statistics provide a robust measure of keyness that avoids the sensitivity of chi-square to corpus size (Dunning, 1993). Collocational analysis, which examines statistically significant co-occurrence patterns (Sinclair, 1991), reveals the phraseological

patterning characteristic of specialized registers — a dimension of legal language emphasized by Biel (2014) and Gozdz-Roszkowski (2011).

Nominalization, the process by which verbs or adjectives are converted to nouns, has been consistently identified as a hallmark of formal written registers, and legal language in particular (Tiersma, 1999; Gibbons, 2003). Halliday (1994) characterizes nominalization as "grammatical metaphor," arguing that it increases lexical density by packaging entire processes as nominal entities. German is particularly prone to nominalization, given its productive derivational morphology and its tradition of Nominalstil in administrative and legal writing (Eroms, 2008). Whether Uzbek legal press exhibits analogous patterns, given different morphological strategies for nominalization (the suffix -ish/-lik and related forms), represents a central empirical question for the present study.

2.4 Legal Language in Central Asian and Uzbek Contexts

Research on Uzbek legal language occupies a marginal position in international linguistics journals, though a modest body of scholarship exists in Uzbek and Russian. Saidov (2012) traces the historical development of Uzbek legal terminology, noting the layered influence of Arabic-Persian borrowings (pre-Soviet), Russian legal lexis (Soviet period), and more recent processes of terminological nativization and partial Latinization following the shift from Cyrillic to Latin script in the 1990s. Aminova (2018) examines lexical borrowing in Uzbek legal texts, and Nazarov (2021) discusses register features of Uzbek legislative prose, though neither study employs corpus methods or addresses the legal press genre specifically. The absence of a corpus-based account of Uzbek legal press represents both a scholarly gap and a practical limitation, given the growing importance of Uzbekistan as a legal reform context and the demand for trained legal translators working between German, Russian, Uzbek, and English (Tursunova, 2020). The present study seeks to establish a preliminary empirical baseline that may inform future translation-oriented and pedagogically oriented research in this domain.

2.5 The German–Uzbek Pairing: Typological and Cultural Dimensions

German and Uzbek differ along several dimensions relevant to the present study. Typologically, German is a fusional (inflectional) language with a relatively free constituent order constrained by the verb-second rule and the sentential bracket (Satzklammer), which places finite and non-finite verbal elements at the peripheries of the clause (Eisenberg, 2013). This structural feature has implications for sentence length and syntactic complexity, as embedded clauses and extended nominal groups can be intercalated between the verbal elements. Uzbek, by contrast, is an agglutinative, verb-final SOV language in which grammatical relations are expressed through transparent, sequential morphological suffixes rather than through fusion or inflection (Johanson & Csató, 1998). These typological properties may be expected to produce different patterns of syntactic complexity even when controlling for domain and register.

Culturally, Germany's civil law tradition and its associated Rechtsstaat ideology have shaped a legal press that emphasizes systematic legal interpretation, doctrinal commentary, and reference to case law. The Uzbek legal press, operating in a transitional context, reflects a different tension: between inherited Soviet-era legal discourse norms, nativizing tendencies in legal terminology, and aspirations toward alignment with international legal standards. These contextual factors motivate the expectation that German and Uzbek legal press will differ not only in formal linguistic features but in their rhetorical and epistemic orientations.

3. Methodology

3.1 Corpus Design

The present study is based on a purpose-built bilingual comparable corpus of 120,000 tokens, divided equally into two subcorpora of 60,000 tokens each. The German subcorpus (GLC: German Legal Corpus) comprises articles

published between January 2019 and December 2023 drawn from two established German-language legal press outlets: Legal Tribune Online (lto.de) and beck-aktuell (beck-aktuell.de). Both sources target a professional legal readership (attorneys, judges, law students) and cover developments in legislation, case law, and legal policy commentary. Texts were selected to represent three thematic domains: criminal law (Strafrecht), civil procedural law (Zivilprozessrecht), and constitutional affairs (Verfassungsrecht), with approximately equal representation across domains.

The Uzbek subcorpus (ULC: Uzbek Legal Corpus) comprises articles published in the same five-year period, drawn from Huquq va Hayot ("Law and Life"), the digital portal of the Uzbek Ministry of Justice (justice.gov.uz/uz/news), and the periodical Yurist ("The Jurist"). These sources similarly address a semi-specialist readership and cover analogous thematic domains: criminal law (jinoyat huquqi), civil procedure (fuqarolik protsessual huquqi), and constitutional matters (konstitutsiyaviy masalalar). Domain balance was maintained proportionally.

Text selection followed three criteria: (1) texts had to belong to the commentary or analytical sub-genre (excluding purely news-wire reports or legislative texts); (2) texts had to be authored by legally trained contributors (editorial board members, practicing attorneys, or legal academics); and (3) texts had to be available in digital form with stable URLs for documentation purposes. Metadata for each text, including source, date of publication, domain, and approximate author affiliation, were recorded in a corpus metadata spreadsheet.

Table 1 provides a summary of corpus composition.

Table 1. Corpus Composition Summary

Feature	German Subcorpus (GLC)	Uzbek Subcorpus (ULC)
Total tokens	60,247	59,831
Total texts	78	82
Mean text length (tokens)	772	730
Time span	2019–2023	2019–2023
Primary sources	LTO, beck-aktuell	Huquq va Hayot, Yurist, justice.gov.uz
Thematic domains	Criminal, Civil Procedure, Constitutional (equal thirds)	Criminal, Civil Procedure, Constitutional (equal thirds)
Author affiliation	Legal professionals, academics	Legal professionals, academics

Note. Token counts were generated using AntConc 4.2.0 word list function with default tokenization settings.

3.2 Analytical Tools and Procedures

The primary tool for corpus analysis was AntConc 4.2.0 (Anthony, 2022), which was used for word list generation, keyword analysis, concordance searches, and collocation calculations. Sketch Engine (Kilgarriff et al., 2014) was used for Uzbek corpus management (given its support for Uzbek language resources) and for computing standardized type-token ratio (STTR) using a moving window of 1,000 tokens, which normalizes for corpus size effects. All

analyses were conducted on lowercased, minimally cleaned texts (removal of HTML markup and non-linguistic metadata), without lemmatization, to preserve morphological surface forms relevant to the typological comparison.

Keyword analysis employed the log-likelihood (LL) statistic with a threshold of LL ≥ 6.63 ($p < 0.01$, $df = 1$), following Rayson (2008). The German subcorpus was compared against a 500,000-token reference corpus of general German journalistic prose (drawn from the Leipzig Corpora Collection, Goldhahn, Eckart, & Quasthoff, 2012), and the Uzbek subcorpus was compared against a 200,000-token reference corpus of general Uzbek newspaper prose compiled for this study. The use of general journalistic reference corpora follows the rationale of Hardt-Mautner (1995): keywords relative to general journalistic prose isolate features specific to the legal press genre rather than to written language generally.

Nominalization frequency was operationalized as the proportion of tokens belonging to morphologically identifiable nominalizations per 100 tokens. For German, deverbal nominalizations (e.g., Entscheidung "decision", Verurteilung "conviction") and deadjectival nominalizations (e.g., Rechtswidrigkeit 'illegality') were identified using a morphological pattern list based on the suffixes -(ung), -(heit), -(keit), -(schaft), -(tum), and -(nis), cross-referenced against a reference word list. For Uzbek, deverbal nominalizations were identified by the suffixes -ish (e.g., qaror qilish "decision-making"), -uv (e.g., ko'rib chiqiluv "review"), and -lik (deadjectival, e.g., qonuniylik "lawfulness"). Manual verification of a 10% random sample confirmed a tagging precision of 94.3% for German and 91.7% for Uzbek.

Hedging and epistemic modality markers were identified through a predefined seed list, subsequently expanded by concordance inspection. German hedges included modal verbs (könnte, dürfte, scheint), evidential adverbials (anscheinend, offenbar), and approximators (etwa, ungefähr). Uzbek hedges included modal particles (ehtimol "perhaps", ko'rinadi "it seems"), modal verb constructions (-sa kerak "should/must"), and epistemic adverbials (aftidan "apparently"). Items were counted per 1,000 tokens to normalize for subcorpus size differences.

Mean sentence length (MSL) was calculated as total tokens divided by total sentences, with sentence boundaries defined by terminal punctuation (period, exclamation mark, question mark). German compound sentences and those with embedded relative clauses were also analyzed for clause depth, defined as the number of embedded clausal levels per sentence, using a stratified random sample of 200 sentences per subcorpus annotated manually.

3.3 Ethical Considerations

All texts analyzed in this study are publicly available on the internet and are not subject to privacy restrictions. No human participants were involved in the research. The use of copyrighted journalistic texts for linguistic analysis constitutes scholarly fair use under applicable German (§51 UrhG) and Uzbek (Law on Copyright and Related Rights, 2006, Article 18) copyright frameworks, consistent with established practice in corpus linguistics research (Hunston, 2002). Full bibliographic references for all source texts are maintained in an archived metadata file available from the corresponding author.

4. Results

4.1 RQ1: Lexical Density and Type-Token Ratio

Table 2 presents the mean lexical density (LD) and standardized type-token ratio (STTR) for both subcorpora, calculated across all texts and all three thematic domains.

Table 2. Lexical Density and Standardized Type-Token Ratio by Subcorpus

Measure	German (GLC)	Uzbek (ULC)	Difference	Statistical Test
Mean Lexical Density (%)	52.4 (SD = 3.1)	44.7 (SD = 4.2)	+7.7%	$t(158) = 13.42$, $p < .001$
STTR (per 1,000 tokens)	38.6 (SD = 2.8)	47.3 (SD = 3.5)	-8.7	$t(158) = -16.21$, $p < .001$
Total word types	12,847	15,234	—	—
Total word tokens	60,247	59,831	—	—

Note. STTR computed using a 1,000-token moving window (Sketch Engine). Lexical density = content words / total words × 100.

The German subcorpus exhibits markedly higher lexical density than the Uzbek subcorpus (52.4% vs. 44.7%), a difference that is statistically highly significant ($t(158) = 13.42$, $p < .001$). This finding is consistent with the characterization of German legal prose as a high-density register shaped by nominalization and complex nominal groups, which concentrate multiple content words within compact syntactic structures. By contrast, the Uzbek subcorpus shows a higher STTR (47.3 vs. 38.6), indicating greater lexical variety relative to text size. This pattern may reflect the agglutinative morphology of Uzbek, in which a single lexical root generates a large number of morphologically distinct surface forms, each counted as a separate type. However, it may also reflect a more accessible stylistic orientation in Uzbek legal press, which addresses readers less socialized into a specialized legal lexicon.

Domain analysis reveals that differences in lexical density are consistent across all three thematic domains, with criminal law texts showing the highest densities in both subcorpora (German: 54.1%; Uzbek: 46.8%) and constitutional texts the highest STTR values (German: 40.2%; Uzbek: 49.1%). These domain-internal patterns suggest that the LD and STTR differences are register-level rather than topic-specific phenomena.

4.2 RQ2: Nominalization Patterns and Syntactic Complexity

Table 3 presents key measures of morphosyntactic complexity for both subcorpora.

Table 3. Nominalization Frequency, Mean Sentence Length, and Clause Embedding

Measure	German (GLC)	Uzbek (ULC)	Difference	Statistical Test
Nominalization freq. (per 100 tokens)	18.7 (SD = 2.4)	11.2 (SD = 2.1)	+7.5	$t(158) = 19.84$, $p < .001$
Mean Sentence Length (words)	24.3 (SD = 6.2)	16.8 (SD = 4.7)	+7.5	$t(158) = 8.91$, $p < .001$
Mean clause depth	2.4 (SD = 0.7)	1.6 (SD = 0.5)	+0.8	$t(398) = 8.06$, $p < .001$

Measure	German (GLC)	Uzbek (ULC)	Difference	Statistical Test
% sentences with 3+ embedded clauses	34.7%	11.3%	+23.4%	$\chi^2(1) = 47.6$, $p < .001$

Note. Clause depth based on stratified random sample of 200 sentences per subcorpus (N = 400). Nominalization frequency computed per 100 tokens.

The German subcorpus demonstrates substantially higher nominalization frequency than the Uzbek subcorpus (18.7 vs. 11.2 per 100 tokens), a difference that is statistically significant ($t(158) = 19.84$, $p < .001$). The most frequent German nominalizations include Entscheidung ("decision", f = 847), Regelung ("regulation", f = 712), Verletzung ("violation", f = 634), and Zuständigkeit ("jurisdiction", f = 521). Concordance analysis reveals that these nominalizations frequently appear in complex nominal groups with multiple prenominal modifiers and postmodifying genitive phrases — a construction type that is characteristic of German legal and administrative Nominalstil and that contributes directly to the elevated lexical density observed in RQ1.

The following concordance lines illustrate this pattern in the German subcorpus (GLC):

(1) ...die Verletzung des allgemeinen Persönlichkeitsrechts durch die rechtswidrige Veröffentlichung...

('...the violation of the general right of personality through the unlawful publication...')

(2) ...die Unzulässigkeit der rückwirkenden Anwendung verschärfter Strafvorschriften...

('...the impermissibility of the retroactive application of more stringent penal provisions...')

In Uzbek, nominalization is morphologically distinct. The suffix -ish is the most productive deverbal nominalizer, appearing in high-frequency forms such as ko'rib chiqish ("review/examination", f = 423), tahlil qilish ("analysis", f = 387), and belgilash ("determination/establishment", f = 342). These forms are structurally more transparent than their German counterparts, as the underlying verb is recoverable from the nominalized form without morphophonological opacity. The following Uzbek concordance lines illustrate the contrast:

(3) ...jinoyat ishini qayta ko'rib chiqish uchun asoslar mavjud bo'lganda...

('...when there are grounds for reviewing the criminal case again...')

(4) ...huquqiy me'yorlarni to'g'ri qo'llash masalasi dolzarb bo'lib qolmoqda...

('...the issue of correct application of legal norms remains urgent...')

Mean sentence length is significantly higher in German (M = 24.3 words) than in Uzbek (M = 16.8 words), consistent with the German tendency toward complex, hypotactic sentence construction enabled by the Satzklammer. The proportion of sentences containing three or more embedded clausal levels is markedly higher in German (34.7%) than in Uzbek (11.3%), a difference that achieves high statistical significance ($\chi^2(1) = 47.6$, $p < .001$). These findings converge on a picture of German legal press as morphosyntactically denser and structurally more complex than its Uzbek counterpart.

4.3 RQ3: Discourse Markers and Hedging Devices

Table 4 presents the frequency of hedging and epistemic modality devices per 1,000 tokens in each subcorpus.

Table 4. Hedging Devices per 1,000 Tokens by Subcorpus and Category

Hedging Category	German (GLC)	Uzbek (ULC)	LL Statistic	Direction
Modal verbs (epistemic)	12.4	6.8	LL = 87.3**	GLC > ULC
Epistemic adverbials	8.7	14.2	LL = 112.6**	ULC > GLC
Approximators	3.2	5.6	LL = 44.1**	ULC > GLC
Attribution formulae	6.1	9.3	LL = 61.4**	ULC > GLC
Total hedge density	30.4	35.9	LL = 29.8**	ULC > GLC

Note. ** $p < .001$ (LL threshold ≥ 10.83). Frequencies per 1,000 tokens, normalized for subcorpus size.

The results reveal an interesting asymmetry in hedging strategies. While German legal press makes greater use of epistemic modal verbs (könnte "could", dürfte "may/must", scheint "seems": 12.4 per 1,000 tokens vs. 6.8 in Uzbek), Uzbek legal press shows higher frequencies of epistemic adverbials, approximators, and attribution formulae (total hedge density: 35.9 vs. 30.4 per 1,000 tokens). The German pattern of modal-verb-mediated hedging is consistent with the convention in German legal commentary of framing interpretive uncertainty through deontic-epistemic modal constructions — a practice that has been linked to the tradition of doctrinal legal writing and its preference for argument-internal qualification (Luttermann, 2010).

The higher frequency of epistemic adverbials and attribution formulae in Uzbek legal press, including forms such as mutaxassislar fikricha ("according to experts"), bir qator tadqiqotlar ko'rsatishicha ('as several studies indicate'), and aftidan ('apparently'), may reflect a different rhetorical strategy for managing epistemic authority: rather than hedging through grammatical means, Uzbek legal press authors appear more likely to attribute claims to external sources, thereby distributing epistemic responsibility. This pattern resonates with observations from studies of journalistic evidentiality in Turkic languages (Johanson, 2000), and it may also reflect the transitional nature of Uzbek legal discourse, in which citing official and expert sources provides institutional legitimacy that individual authorial voice alone may not confer.

Keyword analysis reinforces these patterns. Table 5 presents the top 20 keywords (by LL value) for each subcorpus relative to its general journalistic reference corpus.

Table 5. Top 10 Keywords by Log-Likelihood for Each Subcorpus (Relative to General Journalistic Reference)

Rank	German Keyword	LL	Uzbek Keyword	LL
1	Entscheidung ('decision')	1,842.3	huquq ('law/right')	1,621.4
2	Regelung ('regulation')	1,704.1	qonun ('statute')	1,587.2
3	Verletzung ('violation')	1,621.7	sud ('court')	1,432.6
4	Zuständigkeit ('jurisdiction')	1,543.2	jinoyat ('crime')	1,287.4

Rank	German Keyword	LL	Uzbek Keyword	LL
5	Tatbestand ('legal fact/element')	1,489.6	belgilash ('determination')	1,214.7
6	Anwendbarkeit ('applicability')	1,312.4	ko'rib chiqish ('review')	1,198.3
7	Rechtswidrigkeit ('unlawfulness')	1,288.7	qoidabuzarlik ('violation')	1,087.2
8	Grundsatz ('principle')	1,241.3	tartibga solish ('regulation')	1,042.1
9	Verfassungsmäßigkeit ('constitutionality')	1,187.6	vakolat ('authority/competence')	987.4
10	Rechtsfolge ('legal consequence')	1,143.2	amaliyot ('practice')	921.6

Note. LL = log-likelihood value ($p < .001$ threshold: LL ≥ 10.83). Keywords ranked by LL value descending.

The German keyword profile is dominated by abstract deverbal and deadjectival nominalizations — all top keywords are morphologically complex nouns that package entire legal concepts within a single lexical unit. The Uzbek keyword profile, by contrast, includes a higher proportion of morphologically simplex nouns (huquq, qonun, sud) alongside verbal nominalizations (belgilash, ko'rib chiqish). This distributional difference reinforces the finding from RQ2 that nominalization in German legal press tends toward derivational complexity, while Uzbek nominalization more often preserves transparent verbal morphology.

5. Discussion

5.1 Typology, Register, and the Shape of Legal Press Language

The results of this study offer several converging lines of evidence that German and Uzbek legal press differ systematically in lexical, morphosyntactic, and rhetorical dimensions. German legal press is characterized by higher lexical density, more frequent and morphologically complex nominalization, longer and more deeply embedded sentences, and a preference for grammatical hedging through epistemic modal verbs. Uzbek legal press exhibits greater lexical variety, more morphologically transparent nominalization, shorter and less complex sentences, and a higher proportion of source-attribution hedging through adverbials and attribution formulae.

These patterns can be understood at multiple levels. At the typological level, the contrast between German's inflectional morphology and Uzbek's agglutinative morphology has direct consequences for the surface forms taken by nominalization and for the structural possibilities for sentence-internal embedding. German's sentential bracket (Satzklammer) allows and even encourages the insertion of extended nominal groups and subordinate clauses between the auxiliary and the main verb, producing the long, multi-clause structures observed in the present corpus. Uzbek's verb-final structure and postpositional marking of grammatical relations distribute syntactic information differently, discouraging the kind of prenominal stacking that characterizes German nominal groups (cf. Eisenberg, 2013; Johanson & Csató, 1998).

At the register level, the findings are consistent with Biber's (1988) characterization of formal written registers as tending toward high informational density and nominal complexity. German legal press patterns closer to this pole than Uzbek legal press, a finding that may partly

reflect the longer institutionalization of German legal journalism and the deep socialization of German legal professionals into a writing tradition that values systematic, abstract, and densely nominalized exposition. The comparatively lower lexical density of Uzbek legal press may reflect a different audience assumption, one in which clarity and accessibility are weighted more highly relative to terminological precision, or it may reflect the transitional state of Uzbek legal terminology, which continues to negotiate among Arabic-Persian, Russian, and nativized lexical resources (Saidov, 2012).

5.2 Hedging, Epistemic Authority, and Legal Culture

The divergent hedging strategies observed in the two subcorpora merit particular interpretive attention. The German preference for modal-verb hedging reflects a well-documented feature of German legal commentary: the epistemic modal dürfte, in particular, functions in German legal discourse as a marker of doctrinal plausibility rather than factual uncertainty, encoding the author's assessment of what the law probably is or what courts are likely to hold (Luttermann, 2010). This use of grammatical hedging allows authors to maintain interpretive authority while signaling that the claims offered are scholarly assessments rather than binding pronouncements.

The Uzbek preference for source-attribution hedging may be understood against a different background of epistemic norms. In a legal press tradition that is still establishing its voice and conventions, attributing claims to experts, official bodies, or comparative legal sources serves a legitimizing function that individual authorial hedging alone cannot fully provide. This finding converges with Johanson's (2000) observations on evidentiality in Turkic languages, where the grammatical expression of information source is a pervasive and conventionally marked phenomenon. The higher frequency of attribution formulae in Uzbek legal press may thus reflect the intersection of a general Uzbek discourse norm (marking information source) with a specifically legal press norm (managing epistemic authority in a transitional discursive context).

5.3 Implications for Legal Translation Pedagogy and Terminology

The findings have several practical implications. For legal translators working between German and Uzbek, the study identifies specific asymmetries that may generate translation difficulty: the dense prenominal nominal groups of German have no direct structural equivalent in Uzbek, requiring strategies of syntactic reorganization and nominalization-to-verb conversion. Conversely, the source-attribution formulae characteristic of Uzbek legal press may require expansion or transformation when rendered in German, where modal verb hedging is the expected convention.

For legal terminology standardization in Uzbek, the keyword analysis suggests that Uzbek legal press relies on a relatively small set of high-frequency simplex terms (huquq, qonun, sud) alongside a growing inventory of verbal nominalizations. This pattern may reflect an unfinished terminologization process, in which complex legal concepts have not yet been lexicalized into single morphologically complex nominal units (as they are in German). Corpus data of the type assembled in this study can contribute directly to terminology planning by identifying the forms that are already in productive use in the legal press register.

For corpus-assisted language teaching (CALL) in legal English or German courses at Uzbek law institutions, a context directly relevant to the authors' institutional setting, the present corpus data could be adapted for data-driven learning (DDL) activities (Johns, 1991; Boulton & Cobb, 2017), in which learners use concordance output to identify genre-appropriate lexical and syntactic patterns. The contrast between German and Uzbek nominalization profiles, in particular, provides a productive point of departure for raising learner awareness of grammatical metaphor and its role in academic and legal register.

5.4 Convergence, Divergence, and the Internationalization of Legal Discourse

A theoretically significant question raised by this study is whether Uzbek legal press shows evidence of convergence toward international (i.e., predominantly Western European or Anglo-American) norms of legal discourse. The findings present a mixed picture. On the one hand, the adoption of German-derived terminology (e.g., vakolat, huquqiy me'yor) and the increasing use of structural argumentation moves in Uzbek legal commentary suggest some degree of intertextual borrowing from European legal discourse. On the other hand, the rhetorical conventions of source attribution, the morphological transparency of nominalization, and the shorter, less embedded sentence structures suggest that Uzbek legal press retains distinctive features that resist full convergence.

This ambivalent convergence is consistent with the theoretical account of globalization in legal discourse offered by Bhatia (2004), who argues that globalization produces not homogenization but rather "glocalization" — the adaptation of global discourse norms to local institutional, linguistic, and cultural contexts. The present findings suggest that Uzbek legal press is engaged in precisely this kind of glocalization process, selectively incorporating features of international legal discourse while maintaining a distinctively Uzbek rhetorical and morphosyntactic profile.

6. Conclusion

This study has presented the first corpus-based comparative analysis of German and Uzbek legal press language, drawing on a purpose-built bilingual corpus of 120,000 tokens and applying a range of quantitative and qualitative corpus methods. The findings address the three research questions as follows.

In response to RQ1, German legal press exhibits significantly higher lexical density (52.4%) than Uzbek legal press (44.7%), while Uzbek shows higher lexical variety as measured by standardized type-token ratio (47.3 vs. 38.6). These differences are consistent across all three thematic domains examined. In response to RQ2, German legal press is characterized by substantially higher nominalization frequency (18.7 vs. 11.2 per 100 tokens), longer mean sentence length (24.3 vs. 16.8 words), and greater clause depth, reflecting typological properties of German morphosyntax and the tradition of Nominalstil in German legal and administrative writing. In response to RQ3, German legal press favors grammatical hedging through epistemic modal verbs, while Uzbek legal press relies more heavily on source-attribution formulae and epistemic adverbials, a pattern interpreted in terms of differing epistemic norms and the transitional institutional context of Uzbek legal journalism.

The principal contribution of this study is the establishment of a quantitative baseline for Uzbek legal press language — a register that has received virtually no attention in international corpus linguistics. The study also demonstrates the analytical productivity of the German–Uzbek comparison as a typologically motivated, historically grounded cross-linguistic pairing. The findings have implications for legal translation between German and Uzbek, for Uzbek legal terminology standardization, and for corpus-assisted legal language teaching in Central Asian legal education contexts.

The study has several limitations that should be acknowledged. The corpus, though internally balanced, is modest in size (120,000 tokens), and results should be interpreted as preliminary. The analysis does not address diachronic variation: whether Uzbek legal press has been converging toward German-like nominalization patterns since the 1990s legal reforms, or diverging from them, remains an open empirical question. The restriction to print/digital press commentary excludes other legal press genres (letters to the editor, court reporting, opinion columns) that may show different profiles.

Future research could extend this work in several directions: a diachronic study tracking changes in Uzbek legal press language across the post-independence period; a trilingual comparison incorporating Russian, which remains a significant language of legal communication in Uzbekistan; and the pedagogical application of the present corpus data in data-driven learning

activities for legal English and legal German courses at Uzbek law schools. Such developments would build on the empirical foundation established here and contribute to the further integration of Central Asian legal linguistics into the broader international field.

References

Aminova, M. (2018). Lexical borrowing in Uzbek legal terminology: Russian and international influences. Toshkent Davlat Yuridik Universiteti Axborotnomasi, 2(1), 34–49.

Anthony, L. (2022). AntConc (Version 4.2.0) [Computer software]. Waseda University. https://www.laurenceanthony.net/software

Bhatia, V. K. (1993). Analysing genre: Language use in professional settings. Longman.

Bhatia, V. K. (2004). Worlds of written discourse: A genre-based view. Continuum.

Biber, D. (1988). Variation across speech and writing. Cambridge University Press.

Biber, D., & Conrad, S. (2009). Register, genre, and style. Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press.

Biel, Ł. (2014). Lost in the Eurofog: The textual fit of translated law. Peter Lang.

Biel, Ł., Wasilewska, K., & Koźbiał, D. (2026). Dimensions of variation across institutional legal and administrative registers: An MDA analysis of the Polish Eurolect and the national variety. International Journal of Corpus Linguistics, 31(1), 64–93.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. Language Learning, 67(2), 348–393.

Busse, D. (1992). Recht als Text: Linguistische Untersuchungen zur Arbeit mit Sprache in einer gesellschaftlichen Institution (Reihe Germanistische Linguistik, Bd. 131). Tübingen: Niemeyer. 366 S.

Clay, E. (2026). Measuring divergence in migration-related terminology between EU legal discourse and press articles in English and French. International Journal of Corpus Linguistics, 31(1), 13–35.

Conley, J. M., & O'Barr, W. M. (2005). Just words: Law, language, and power (2nd ed.). University of Chicago Press.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. Computational Linguistics, 19(1), 61–74.

Eisenberg, P. (2013). Grundriss der deutschen Grammatik: Der Satz (4th ed.). Metzler.

Eroms, H.-W. (2008). Stil und Stilistik: Eine Einführung. Erich Schmidt.

Felder, E., & Vogel, F. (Eds.). (2017). Handbuch Sprache im Recht (Handbücher Sprachwissen, Bd. 12). Berlin/Boston: De Gruyter. XIX, 570 S.

Gibbons, J. (2003). Forensic linguistics: An introduction to language in the justice system. Blackwell.

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), 759–765.

Gozdz-Roszkowski, S. (2011). Patterns of linguistic variation in American legal English: A corpus-based study. Peter Lang.

Halliday, M. A. K. (1978). Language as social semiotic. Arnold.

Halliday, M. A. K. (1994). An introduction to functional grammar (2nd ed.). Arnold.

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). An introduction to functional grammar (3rd ed.). Arnold.

Hardt-Mautner, G. (1995). 'Only connect': Critical discourse analysis and corpus linguistics. UCREL Technical Paper 6. University of Lancaster.

Hunston, S. (2002). Corpora in applied linguistics. Cambridge University Press.

Johanson, L. (2000). Evidentiality in Turkic. In L. Johanson & B. Utas (Eds.), Evidentials: Turkic, Iranian and neighbouring languages (pp. 1–24). Mouton de Gruyter.

Johanson, L., & Csató, É. A. (Eds.). (1998). The Turkic languages. Routledge.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. English Language Research Journal, 4, 1–16.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. Lexicography, 1(1), 7–36.

Lerch, K. D. (Ed.). (2004–2005). Die Sprache des Rechts. Vol. 1: Recht verstehen (2004, 466 S.); Vol. 2: Recht verhandeln (2005, 598 S.); Vol. 3: Recht vermitteln (2005, 540 S.). Berlin/New York: De Gruyter.

Lindroos, E. (2015). The language of lower court judgments in German and Finnish: A formulaic approach. In M. Gotti & C. Williams (Eds.), Insights into specialized translation (pp. 191–212). Peter Lang.

Luttermann, K. (2010). Juristische Kommunikation im internationalen Kontext: Deutsch–Englisch–Latein. LIT Verlag.

McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge University Press.

Nazarov, B. (2021). Register features of Uzbek legislative prose: A functional approach. O'zbekiston Milliy Universiteti Xabarlari, 3(2), 121–134.

Rayson, P. (2008). From key words to key semantic domains. International Journal of Corpus Linguistics, 13(4), 519–549.

Saidov, A. K. (2012). Comparative law and national legal systems: Uzbekistan in context. Fan.

Scott, M. (1997). PC analysis of key words and key key words. System, 25(2), 233–245.

Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press.

Swales, J. M. (1990). Genre analysis: English in academic and research settings. Cambridge University Press.

Tiersma, P. (1999). Legal language. University of Chicago Press.

Tursunova, Z. (2020). Legal translation in Uzbekistan: Between Russian, Uzbek and English. Translation Studies in the New Millennium, 18(1), 87–104.

Vogel, F., Hamann, H., & Gauer, I. (2018). Computer-assisted legal linguistics: Corpus analysis as a new tool for legal studies. Law & Social Inquiry, 43(4), 1340–1373.

Vogel, F. (2012). Linguistik rechtlicher Normgenese: Theorie der Rechtsnormdiskursivität am Beispiel der Online-Durchsuchung (Sprache und Wissen, Bd. 9). Berlin/Boston: De Gruyter. 485

S.B.M Rahimovna .Oriental Art and Culture, 143-146.The lexicographical aspects of legal terms in german and uzbek languages (2023)

Baxriddinova, M. (2023). Nemis va o‘zbek tillari yuridik terminlarining leksikografik tadqiqi. Toshkent: Fan va texnologiya.