

Between Two Languages: Tourism Terminology in Albanian and English Based on a Comparative Corpus Analysis

Arjan Shumeli^{1*}, Erion Shehu¹ and Merita Isaraj²

¹Department of Foreign Languages, Agricultural University of Tirana, 1025 Pajsi Vodica Street, Koder Kamëz, Tirana, Albania.

²Department of Foreign Languages, “E. Çabej” University, “Studenti 30” Street, Gjirokaster, Albania.

*Corresponding author(s). E-mail(s): ashumeli@ubt.edu.al

Contributing authors: erion.shehu@ubt.edu.al; misaraj@uogj.edu.al

Abstract

This article examines Albanian-English tourism terminology through a purpose-built corpus of 300 term pairs drawn from professional tourism communication produced or circulated between 2018 and 2025. It addresses the still limited systematic treatment of Albanian tourism terminology in bilingual and digital contexts and analyzes the data by structural type, semantic field, terminological relation, and translation strategy. The results show that both languages rely heavily on multi-word terminology, although Albanian has more units composed of three or more words (26.7%) than English (14.0%), while English more often compresses concepts into single-word or two-word forms. Eight semantic fields are analyzed, with Cultural Heritage and Destination forming the largest group. Six translation strategies are identified: calque, functional equivalence, direct borrowing, descriptive translation, hybrid formation, and cultural substitution. The findings show that Albanian combines international borrowing with productive native patterns, especially the *turizëm* (Eng. tourism) + *modifier* pattern, while culture-bound terms such as *han* (Eng. inn), *bujtinë* (Eng. guesthouse), *pritje tradicionale* (Eng. traditional welcome), and *rrugë panoramike* (Eng. scenic route) require contextual interpretation. The study contributes a functional taxonomy of Albanian-English terminological correspondence and supports a standardized bilingual tourism glossary for translation, tourism education, destination marketing, and digital tourism communication.

Keywords: Albanian-English terminology, tourism terminology, structural asymmetry, calque, bilingual glossary, standardization

1 Introduction

Tourist encounters are organized through language from the outset. Before a traveller books a room, boards a plane, or follows a mountain trail, specialized vocabulary has already named the service, classified the place, and framed the cultural experience. Tourism language operates where economics, culture, geography, and professional communication meet; it is technical and promotional, regulatory and experiential, globally standardized and locally distinctive. It is therefore one of the most semantically dense forms of professional discourse (Maci & Spinzi, 2025).

A comparison of Albanian and English is particularly important in this context. In recent decades, Albania has become more visible as a destination, promoting its Adriatic and Ionian coastlines, mountain and canyon systems, Ottoman and Illyrian heritage, and rural hospitality. Yet the terminology used to present Albania to international visitors has received limited systematic attention. The gap is most apparent in digital and platform-mediated communication, where English interface terms, booking formulas, and destination-branding routines place pressure on Albanian naming practices (Buendía Castro, 2024).

Recent studies show that destination terminology no longer circulates only in brochures or official guides. It is also embedded in online platforms, social media, reputation systems, and immersive digital tools. Imran (2024) links digital communication with destination attractiveness through inspirational content, influencer collaboration, reputation management, and virtual/augmented/mixed reality (VR/AR/MR) promotion. Kuanishbaeva (2025) likewise shows that translation and multilingual services support safety, accessibility, marketing, and intercultural understanding in international tourism.

The present study draws on a bilingual corpus of 300 Albanian-English terminological pairs extracted from professional tourism communication. It examines structural distribution, semantic fields, terminological relations, and translation strategies, including calque, borrowing, hybrid formation, descriptive translation, functional equivalence, and cultural substitution. The study is guided by four questions: How do Albanian and English tourism terms differ structurally? Which fields dominate the dataset? Which strategies organize Albanian-English correspondence? How can the findings inform translation practice, tourism education, destination communication, and bilingual glossary design?

The argument is developed cumulatively. The literature review establishes the research gap, the theoretical framework defines the coding dimensions, the methodology section explains term collection and validation, the results analyze structure, semantic field, terminological relation, and translation strategy, and the discussion integrates these findings into a functional taxonomy and term-bank model.

2 Literature Review

The review focuses on four strands directly relevant to the corpus and coding procedure: tourism as specialized discourse, cross-linguistic tourism terminology, translation strategies, and corpus-based terminographic resources.

2.1 Tourism as a Language for Specific Purposes

Tourism language is a well-established language for specific purposes (LSP) because it has a dual task: to identify services and procedures precisely and to present destinations as attractive and culturally legible (Maci & Spinzi, 2025). Its terminology is shaped by multidisciplinary, internationalisms, Anglicisms, neologisms, synonymy, and movement between specialized and general discourse through terminologization and determinologization. Tourism is therefore a useful domain for corpus-based terminology, since its terms must remain technically reliable and readable for non-specialist travellers (Li et al., 2022; Maci & Spinzi, 2025).

2.2 Cross-Linguistic Studies of Tourism Terminology

Cross-linguistic work on tourism terminology has developed considerably, although smaller Balkan destination languages remain less documented. For instance, Li et al. (2022) provide a model for culturally

marked destination terms, Cheng and Bai (2018) demonstrate the value of corpus-based analysis in tourism translation, Buendía Castro (2024) shows why bilingual and multilingual tourism resources matter for translators, and Fijo León and Fuentes Luque (2013) offer a strategy-oriented model for compiling and translating rural tourism terms. Taken together, these studies link culture-bound naming, corpus evidence, bilingual resource design, and translator-oriented strategy classification. They do not, however, provide a systematically coded Albanian-English dataset, which leaves a practical and descriptive gap for translation, destination promotion, and LSP instruction.

2.3 Translation Strategies in Tourism Texts

Studies of tourism translation approach the issue from a complementary angle: translated terms help build destination image rather than merely transfer information. Castellano-Martínez and González-Rodríguez (2026) argue that translated tourism materials mediate between language, culture, and promotion, while Gandin (2013) shows that translated tourism English may differ from original English in discourse patterning, style, and persuasive effect. These findings support the treatment of Albanian-English tourism terms as choices that affect readability, image, and promotional force.

Malamatidou (2024) is useful here not merely as a recent reference; rather, the author's corpus-based approach shows that tourism translation also constructs cultural worldviews and audience-oriented representations. This perspective supports the present article's treatment of Albanian-English tourism terms as communicative and culturally situated choices, rather than as neutral lexical substitutions.

Tourism terminology therefore poses problems that ordinary dictionary equivalence cannot resolve (Fijo León & Fuentes Luque, 2013; Maci & Spinzi, 2025). Available strategies include transcription, calque, transference, gloss translation, functional equivalence, neutralization, zero translation, and addition. In this study, the strategic dimension is narrowed to the six procedures actually observed in the Albanian-English dataset, so that the labels remain evidence-based rather than decorative.

2.4 Corpus-Based Terminology and Digital Resources

Corpus-informed studies are relevant here because they clarify what evidence a tourism term bank should preserve. Méndez-Cendón and Pérez-Ruiz (2015) support the terminographic treatment of culture-bearing items such as *han*, *bujtinë*, and *shtëpi pritëse*, where a single gloss rarely captures the concept. Cenni and Goethals (2020) justify the inclusion of platform-mediated discourse, while Orel Frank (2022) treats tourism terminology as teachable LSP knowledge, a point that directly supports the pedagogical use of the Albanian-English glossary proposed in this paper. More specifically, Orel Frank (2022) connects tourism terminology with LSP syllabus design, learner needs, and domain communication, thereby supporting the glossary as a pedagogical terminographic resource.

Corpus-based terminography also shapes the method used in this study. Durán Muñoz (2022), Sagmeister (2010), and dos Santos Moreira (2012) show the value of specialized corpora, subject-field classification, candidate extraction, semantic relations, and detailed records in bilingual resources. Estopà Bagot (2001), Artstein and Poesio (2008), Rigouts Terryn et al. (2020), Calvi and Santos López (2020), Rojas-García (2022), Alali et al. (2024), Buendía Castro (2024) and Delvizio et al. (2025) further support the combined use of corpus tools, expert validation, agreement-based review, and digital terminography.

At a foundational level, Cabré Castellví (1999, 2023) provides the communicative basis for treating terms as usage-sensitive units; Pearson (1998) and Bowker and Pearson (2002) support corpus design and contextual reading; Temmerman (2000) accounts for fuzzy and culture-bound categorization; and Meyer and Mackintosh (1996), together with Bourigault et al. (2001), justify attention to phraseology and knowledge-rich contexts.

This section is also situated in a longer terminological tradition. Wüster (1979) remains useful as a point of departure for concept-oriented ordering and standardization, while L'Homme (2004) gives the practical terminographer a method for recording terms, contexts, definitions, and semantic relations. Daille (1994) is relevant to the extraction stage because it combines linguistic and statistical cues for identifying term candidates in corpora. Kageura (2012) further supports the treatment of term formation and terminological growth as systematic rather than accidental. Together, these sources justify recording

structural types, semantic fields, terminological relations, translation strategies, context, phraseology, source evidence, and process notes, so that the Albanian-English entries remain auditable corpus objects rather than detached bilingual labels (Daille, 1994; L'Homme, 2004; Kageura, 2012).

Methodologically, this literature explains why the 300 Albanian-English pairs (see Table 3, Appendix B) are handled as corpus-attested terminological units rather than as isolated dictionary equivalents. Recurrence, contextual evidence, semantic-field membership, and expert validation are therefore used as filters, and the proposed resource records form, field, strategy, relation, source evidence, and culture-specific notes.

These studies also anticipate the results that follow. Structural typology draws on corpus evidence, semantic-field analysis follows concept-oriented terminography, and the discussion of equivalence depends on the same distinction between lexical matching and communicative function (Sager, 1990; Pearson, 1998; Orel Frank, 2022).

2.5 Recent Comparative Evidence on Tourism Terminology

Recent comparative studies confirm that tourism terminology is technical, cultural, and promotional. Prima (2020) treats tourism as a specialized terminological field in English discourse, Karimalieva (2024) links tourism discourse with translation difficulties, and Koval and Blaguta (2025) examine ethnospecific realia in cross-cultural tourist communication. Work on Uzbek and Karakalpak also shows borrowing, native description, metaphor, and evaluative destination language in less well-documented tourism languages (Akbarov, 2025; Yulchiyeva, 2025; Ziyotova, 2025; Seytkasimov & Seytbekov, 2026).

This comparative evidence is useful for Albanian because it shows a familiar pattern in smaller destination languages: international terms are adopted where the sector is global, while native structures remain productive where local culture or landscape is foregrounded. It supports the treatment of *han* (Eng. inn), *qendër historike* (Eng. historic centre), *prijtje tradicionale* (Eng. traditional welcome), and *rrugë panoramike* (Eng. scenic route) as genre-dependent and culturally marked units.

Terminographic work then gives the article its practical direction. Astorino (2013), Méndez-Cendón and Pérez-Ruiz (2015), Cenni and Goethals (2020), and Orel Frank (2022) all support a resource that is transparent enough for translation, glossary design, and tourism-language pedagogy.

The remaining gap is therefore well defined: there is still no integrated Albanian-English tourism dataset that combines corpus evidence, bilingual alignment, structural coding, semantic-field analysis, translation-strategy classification, and term-bank design.

3 Theoretical Framework

3.1 The Communicative Theory of Terminology

This study is grounded in the Communicative Theory of Terminology (CTT), which treats terms as units whose specialized value is activated in actual communicative settings rather than being fixed only in dictionaries (Cabré Castellví, 1999, 2023). This orientation suits tourism, where the same term may appear in institutional texts, booking platforms, hotel descriptions, brochures, guides, and visitor-facing digital discourse.

CTT is operationalized here through its cognitive, linguistic, and socio-communicative dimensions. The cognitive dimension supports semantic-field classification, the linguistic dimension supports structural coding, and the socio-communicative dimension accounts for source, register, and communicative setting. These three dimensions are then converted into coding values for each Albanian-English pair.

Faber Benítez's (2009) account of terminology and specialized translation strengthens this position because it treats terms as conceptual units that must be reconstructed in translation. Albanian descriptive forms such as *strukturë akomoduese* (Eng. accommodation facility) and culture-bound forms such as *bujtinë* (Eng. guesthouse/inn) or *han* (Eng. inn/historical roadside lodging) therefore require contextual interpretation, not automatic substitution.

Sager's (1990) model for identifying, processing, storing, and comparing terms, together with Temmerman's (2000) account of variation and gradual stabilization, also informs the coding procedure. Candidate terms were retained only when recurrence, source context, formal patterning, and communicative function supported their specialized value.

In this framework, terminologization, determinologization, and transterminologization are treated as observable changes in use. A unit may become specialized within tourism, move from expert or platform discourse into ordinary visitor-facing communication, or enter tourism from another field such as ecology, agriculture, management, or digital commerce. This interpretation links Cabré Castellví's communicative theory directly to the coding of the corpus and to the later discussion of Albanian-English asymmetry.

3.2 Operationalization of the Coding Procedure

To put the theoretical framework into practice, each Albanian-English term pair was coded through six interrelated dimensions. The procedure combines CTT, corpus-based terminography, and tourism translation studies, so that each item is treated as a corpus-attested terminological unit rather than as an isolated bilingual equivalent (Pearson, 1998; Cabré Castellví, 1999; Fijo León & Fuentes Luque, 2013).

The coding procedure was kept deliberately concrete. Each theoretical category was converted into an operational decision and applied to the 300 term pairs. Table 1 shows how concepts from CTT, corpus-based terminography, and specialized translation studies were translated into coding values. It also clarifies that each pair was examined through form, conceptual field, communicative setting, interlingual relation, translation strategy, and possible terminological movement. Operationally, a term-dynamics label was entered only when context showed specialization, visitor-facing generalization, or cross-domain transfer; the note field recorded that evidence and prevented intuitive labelling (Cabré Castellví, 1999; Meyer & Mackintosh, 1996).

As Table 1 indicates, the coding model links theory, method, and analysis. Structural type and semantic field support the quantitative findings on term formation and domain distribution; genre/register explains how source type and audience affect term choice; terminological relation and translation strategy account for equivalence, variation, and asymmetry (cases in which a term in one language has no equally specific equivalent in the other); and term dynamics connects the observed forms with terminologization, determinologization, and transterminologization. The model therefore makes the analysis explicit rather than impressionistic.

The same logic governed the treatment of term dynamics. Terminologization was recorded only when a general lexical item became stable in a tourism-specific combination, as in *rrugë panoramike* (Eng. scenic route). Determinologization was recorded when a technical or platform term circulated in visitor-facing discourse without losing its service function, as in *anulim falas* (Eng. free cancellation). Transterminologization was recorded when a term from another domain was refunctionalized in tourism, as in *ekoturizëm* (Eng. ecotourism) or *marketing turistik* (Eng. tourism marketing). These rules explain why term dynamics reappears in the results and discussion as evidence of Albanian descriptive expansion, English compactness, and cross-domain transfer.

For methodological consistency, a process label was entered only when three conditions coincided: a stable form, a tourism-specific context, and clear evidence of specialization, popularization, or cross-domain transfer. Items that were merely frequent, metaphorical, or stylistically attractive were not assigned a dynamics label unless the corpus showed this functional shift. The rule prevents overlap with translation strategy and explains why Section 9 treats term dynamics as a coded result rather than as a theoretical comment.

4 Methodology

This section describes how tourism texts were used to produce 300 validated Albanian-English term pairs. The process covered source selection, text conversion, concordance-supported extraction, candidate filtering, bilingual alignment, expert review, and coding. The aim was to keep the entries evidence-based,

Table 1 Operationalization of the theoretical framework in the coding procedure

Coding dimension	Operational coding in the dataset	Analytical function in the study
Structural type	Albanian and English forms were coded as single-word terms, two-word multi-word terms, or multi-word terms of three or more words, following CTT and corpus-based terminology principles (Pearson, 1998; Cabré Castellví, 1999).	Enables comparison of Albanian descriptive expansion with English terminological compactness.
Semantic field	Each pair was assigned to a tourism conceptual area, such as accommodation, transport, heritage, digital tourism, or outdoor activity, reflecting the cognitive organization of specialized knowledge (Sager, 1990; Cabré Castellví, 1999; Faber Benítez, 2009).	Shows which domain areas dominate the corpus and supports concept-oriented glossary design.
Genre/register	The source setting was coded as institutional, commercial, promotional, or visitor-facing, in line with CTT's socio-communicative dimension (Cabré Castellví, 1999; Faber Benítez, 2009).	Explains how source type, audience, and communicative purpose influence term choice.
Terminological relation	The conceptual relation between the Albanian and English items was coded as generic-specific (Sager, 1990; Temmerman, 2000).	Treats the corpus as a structured terminological system rather than as a bilingual wordlist.
Translation strategy	Albanian-English correspondence was coded as calque, functional equivalence, direct borrowing, descriptive translation, hybrid formation, or cultural substitution/zero equivalence (Fijo León & Fuentes Luque, 2013; Li et al., 2022; Maci & Spinzi, 2025).	Identifies the procedures through which equivalence is established, adapted, or problematized.
Term dynamics	Terminologization, determinologization, and transterminologization were coded only when supported by corpus evidence (Cabré Castellví, 1999; Temmerman, 2000; Meyer & Mackintosh, 1996).	Links formal and semantic change to tourism use while avoiding unsupported theoretical labelling.

not to produce unverified bilingual matches (Estopà Bagot, 2001; Artstein & Poesio, 2008; Rigouts Terryn et al., 2020).

4.1 Corpus Design and Source Profile

The corpus was built from professional tourism communication produced or circulated between 2018 and 2025. Sources were chosen to cover the eight semantic fields analyzed in the study, which are not presented as statistically representative of the whole tourism domain. The sample combines institutional, heritage-related, accommodation, platform-based, guidebook, and tour-operator materials, with priority given to sources available in both Albanian and English because they allow more reliable alignment.

Official and institutional sources included Albanian National Tourism Agency (Albanian National Tourism Agency, n.d.) sections on destinations, tourism types, attractions, festivals, eco-tourism, mountain tourism, and visitor information. Heritage and national-park materials included texts relating to Gjirokastrë (Visit Gjirokastra, n.d.), and Butrint (UNESCO World Heritage Center, n.d.). These sources supplied terms connected with institutional categories, cultural heritage, archaeological tourism, protected areas, and destination promotion.

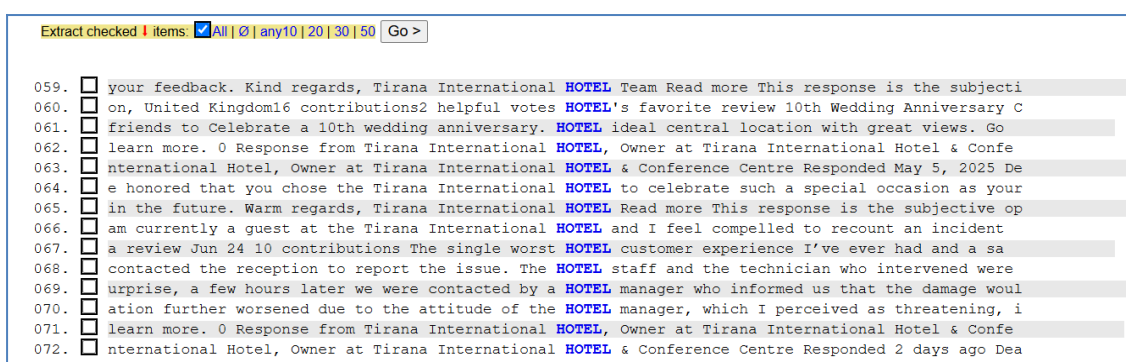


Fig. 1 Concordance output for identifying recurrent tourism terms and term combinations

Commercial and visitor-facing sources included Booking.com (Booking.com, n.d.) and TripAdvisor (TripAdvisor, n.d.) pages for Albanian destinations, especially those relating to the city of Vlorë, Albania tours and holidays (Intrepid Travel, n.d.), together with Lonely Planet Albania (Lonely Planet, n.d.), Tirana In Your Pocket (Tirana In Your Pocket, n.d.), and Responsible Travel (Responsible Travel, n.d.). Their consulted sections covered property descriptions, amenities, prices, booking and cancellation conditions, guest ratings, attractions, itineraries, activities, transport, responsible travel, and practical visitor information. Together, these materials provided accommodation, platform, service, itinerary, and travel-planning terminology.

Together, these sources give the corpus a practical profile: formal, promotional, commercial, platform-based, guidebook-based, and visitor-facing tourism discourse are all represented.

4.2 From Source Collection to Term Extraction and Validation

Relevant sections were first selected from tourism websites, guidebooks, brochures, and related travel materials. Content with no terminological value, including menus, advertisements, privacy notices, contact details, cookie notices, and legal or administrative boilerplate, was removed at the collection stage.

The selected materials were converted into editable text through manual copying into Notepad, extraction from HTML pages and digital guidebooks or brochures, and manual transcription of short printed sources where required. The cleaned material was stored as plain-text files (rawdata1.txt, rawdata2.txt, rawdata3.txt), which produced a uniform and searchable corpus.

The plain-text files were then imported into the concordance tool Concordancer-English-version 10 (Cobb, n.d.), for further analysis (Figure 1). The tool provided keyword-in-context access and made it possible to inspect recurrent lexical items and stable patterns, including *hotel ndërkombëtar* (Eng. international hotel), *dhomë me pamje nga deti* (Eng. sea-view room), and *peshë e lejuar e bagazhit* (Eng. baggage allowance). The tool did not select terms automatically; final selection remained manual and was based on terminological relevance, recurrence, contextual stability, and suitability for Albanian-English analysis.

The concordance lines were sorted, filtered, counted, and read manually. A semi-automatic filtering was therefore performed, which included recurrences searches and string-pattern sorting. Hence, semi-automatic filtering does not refer to automatic term extraction or automatic final selection.

4.3 Term Validation: Inclusion, Exclusion, and Filtering Criteria

Candidate terms were entered into a coding spreadsheet and reviewed by experts (see Appendix A, Table 2). For each item, the spreadsheet recorded the Albanian term, English equivalent, structural type, inclusion or exclusion decision, evaluator response, and explanatory note. The record helped keep the treatment of candidates consistent across the analytical categories.

Inclusion followed four pre-declared criteria: specialized sense activation, membership in a relevant semantic field, recurrence in at least two independent sources, and conceptual specificity. A candidate was retained only when it named a tourism concept and met all four criteria. Final inclusion also

required a clear Albanian-English alignment across the two languages after checking form, meaning, source evidence, and classification.

Retained items included terms naming accommodation, services, amenities, travel procedures, destinations, heritage objects, activities, or visitor practices. The terms also had to be structurally stable, contextually supported, terminologically precise, and alignable across languages, as in park *kombëtar* (Eng. national park), *dhomë familjare* (Eng. family room), *regjistrim hyrës* (Eng. check-in), *sigurim udhëtimi* (Eng. travel insurance), *kartë imbarkimi* (Eng. boarding pass), and *han* (Eng. inn/historic inn).

Items were excluded if they lacked reusable terminological value. Excluded items comprised:

- Isolated proper names, such as *Hotel Tirana* (Eng. Tirana Hotel), *Ksamil* (Eng. Ksamil), and *Saranda* (Eng. Saranda);
- General words without tourism-specific use, such as *rrugë* (Eng. road), *natyrë* (Eng. nature), and *pamje* (Eng. view);
- Website-administration vocabulary, such as *numri i telefonit* (Eng. telephone number), *adresë e-mail* (Eng. email address), and *formular kontakti* (Eng. contact form);
- Privacy or cookie-notice vocabulary, such as *politika e privatësisë* (Eng. privacy policy);
- Temporary slogans, such as *rezervo tani* (Eng. book now), *verë e paharrueshme* (Eng. unforgettable summer), and *zbuloni parajsën* (Eng. discover paradise);
- One-source promotional phrases, such as *ofertë javore* (Eng. weekly offer), *fundjavë relaksi* (Eng. relaxation weekend), and *arratisje verore* (Eng. summer escape);
- Near-duplicate variants without additional value, such as *dhomë me pamje nga deti / dhomë me pamje deti* (Eng. sea-view room) and *qendër informacioni turistike / qendra e informacionit turistik* (Eng. tourist information centre). Within each pair, the variants designate the same concept and were therefore recorded in a single row rather than as separate terms;
- Misleading literal alignments, such as *rrugë ecjeje* (Eng. road for walking). Alignment with English was established through the preferred equivalent “walking route” rather than the literal rendering “road for walking”;
- Unstable descriptive strings, such as *pamje e bukur* (Eng. beautiful view), *eksperiencë e rrallë* (Eng. rare experience), and *shërbim i mirë* (Eng. good service);
- Vague evaluative adjectives, such as *i mrekullueshëm* (Eng. wonderful), *i paharrueshëm* (Eng. unforgettable), and *ideal* (Eng. ideal). General words were retained only in stable tourism-specific combinations.

Borderline candidates were retained only when source evidence confirmed stable tourism use. For example, *apartament* (Eng. apartment/flat) was included only as a bookable short-stay unit; *dhomë familjare* (Eng. family room), *dhomë standarde* (Eng. standard room), and *qëndrim në familje* (Eng. homestay) were included as fixed accommodation categories; and *biletë vajtje-ardhje* (Eng. return ticket), *kartë imbarkimi* (Eng. boarding pass), and *dokument udhëtimi* (Eng. travel document) were retained as institutional mobility documents.

Seasonal terminology provides a useful borderline example. *Sezoni i ulët* (Eng. low season), *sezoni i lartë* (Eng. high season), and *sezon kulmor* (Eng. peak season) were retained because they encode fixed meanings in tourism pricing, booking cycles, destination planning, and visitor-demand management. Although *i ulët* (Eng. low) and *i lartë* (Eng. high) are general adjectives in isolation, in these combinations they function as specialized descriptors for periods of lower and higher tourist activity. Remaining borderline cases were resolved through expert discussion; consultation of specialist dictionaries (Akademia e Shqipërisë, 2006; Dictionary of leisure, travel and tourism, 2006; Nandhakumar, 2016) and bilingual glossaries; and judgement based on corpus evidence, specialized meaning, register, and stable Albanian-English alignment.

5 Structural Typology: Albanian and English Compared

5.1 Headline Distribution

An analysis of the term database shows that both Albanian and English tourism terminologies are strongly phrase-based, although they differ in compression. Single-word terms are the smallest category in both languages, with 52 Albanian entries and 60 English entries. Two-word multi-word terms form the largest group, especially in English, where they account for 198 entries compared with 168 Albanian entries. The sharpest contrast appears in units of three or more words: Albanian has 80 entries, while English has 42 entries. The pattern points to a clear structural tendency: Albanian often makes conceptual relations explicit through longer forms, whereas English tends to condense comparable concepts into shorter, lexicalized, or institutionally stabilized forms.

5.2 Single-Word Terms

Single-word terms form a compact core of 52 Albanian entries (17.3%) and 60 English entries (20.0%). They cover actors, places, services, activities, transport, and institutional concepts. Albanian combines native items such as *bujtinë* (Eng. guesthouse), *han* (Eng. inn), *kështjellë* (Eng. castle), and *fluturim* (Eng. flight) with borrowed or internationalized items such as *hotel*, *resort*, *spa*, *terminal*, *buking*, and *aeroport*.

The contrast between *han* (Eng. inn), which carries cultural and historical associations of lodging and hospitality, and *resort* (Eng. resort), a direct international borrowing, illustrates the coexistence of local semantic depth and global tourism vocabulary.

5.3 Two-Word Terms

Two-word terms are the most productive category. Albanian uses *noun + adjective* structures such as *destinacion turistik* (Eng. tourist destination) and *park kombëtar* (Eng. national park), as well as *noun + complement* structures such as *kartë imbarkimi* (Eng. boarding pass) and *sigurim udhëtimi* (Eng. travel insurance). English alternates between compounds and *adjective + noun* phrases, so equivalence cannot be reduced to simple word-order reversal.

The *rrugë* (Eng. route) group shows terminologization: a general word becomes specialized through modifiers, as in *rrugë bregdetare* (Eng. coastal route), *rrugë kulturore* (Eng. cultural route), *rrugë panoramike* (Eng. scenic route), and *rrugë ecjeje* (Eng. walking route). These forms require contextual interpretation because literal renderings may obscure their status as tourism route products (Faber Benítez, 2009; Li et al., 2022).

5.4 Multi-Word Terms of Three or More Words

Longer multi-word terms remain more frequent in Albanian (26.7%) than in English (14.0%). This difference points to an Albanian preference for explicit prepositional, genitive, or adjectival expansion, while English more often relies on compact compounds or institutional labels.

Examples such as *peshë e lejuar e bagazhit* (Eng. baggage allowance), *dhomë me pamje nga deti* (Eng. sea-view room), *qendër informacioni turistike* (Eng. tourist information centre), and *paketë turistike gjithëpërfshirëse* (Eng. all-inclusive tourist package) show that extended Albanian terms are not merely longer forms; they clarify semantic relations and make the tourism concept more transparent to users.

6 Semantic Field Analysis

The semantic-field distribution was calculated from the integrated Albanian-English tourism dictionary (N = 300) by grouping the dictionary's short-note labels into eight macro-fields. Cultural Heritage and Destination is the largest field, representing 25.3% of the corpus. It is followed by Accommodation and Hospitality (19.7%), Customer Service and Digital Tourism (14.0%), and Transport and Travel Documentation (13.3%). The smaller fields are Types of Tourism (8.3%), Nature and Outdoor Activities (7.3%), Tourism Organization and Management (6.3%), and Gastronomy and Leisure (5.7%). The distribution shows that the dataset is centered on destination identity, heritage representation,

accommodation services, visitor-facing communication, and travel procedures, while still preserving specialized vocabulary from management, outdoor activities, gastronomy, and leisure.

6.1 Accommodation and Hospitality

Accommodation and Hospitality is a layered terminological field rather than a purely hotel-based category. The corpus shows three related patterns:

- International hotel and platform vocabulary: *hotel* (Eng. hotel), *resort* (Eng. resort), *spa* (Eng. spa), *suite* (Eng. suite), *buking* (Eng. booking), and *regjistrim hyrës / check-in* (Eng. check-in). These items reflect global accommodation platforms, hotel classification systems, and guest-facing digital communication.
- Native and culturally embedded Albanian designations: *bujtinë* (Eng. guesthouse), *shtëpi pritëse* (Eng. guesthouse), *odë* (Eng. traditional reception room), *han* (Eng. inn), *qëndrim në familje* (Eng. homestay), and *pritje tradicionale* (Eng. traditional welcome). These terms show that the field is both technical and cultural.
- Descriptive or calqued multi-word formations: *hotel me pesë yje* (Eng. five-star hotel), *shërbim në dhomë* (Eng. room service), *politikë anulimi* (Eng. cancellation policy), *mëngjes i përfshirë* (Eng. breakfast included), *dhomë me pamje nga deti* (Eng. sea-view room), and *strukturë akomoduese* (Eng. accommodation facility).

6.2 Transport and Travel Documentation

Transport and travel-documentation terminology combines adapted internationalisms with stable Albanian calques. The field is represented by two recurrent patterns:

- International or borrowed mobility terms: *aeroport* (Eng. airport), *terminal* (Eng. terminal), *vizë* (Eng. visa), *kroçerë* (Eng. cruise), *traget* (Eng. ferry), and *fluturim charter* (Eng. charter flight). These forms reflect practical tourism usage in airport, airline, guidebook, booking-platform, and visitor-information contexts.
- Calqued or descriptive Albanian formations: *fluturim direkt* (Eng. direct flight), *fluturim i vonuar* (Eng. delayed flight), *bagazh dore* (Eng. hand luggage), *dokument udhëtimi* (Eng. travel document), *sigurim udhëtimi* (Eng. travel insurance), and *portë imbarkimi* (Eng. boarding gate). These terms indicate tourism-oriented standardization rather than direct dependence on IATA, ICAO, or IMO nomenclature.

6.3 Cultural Heritage and Destination Identity

Cultural Heritage and Destination is the most prominent field and the clearest point of contact between technical naming and national cultural identity. The corpus shows two main patterns:

- Institutional heritage and visitor-information labels: *trashëgimi kulturore* (Eng. cultural heritage), *trashëgimi natyrore* (Eng. natural heritage), *sit arkeologjik* (Eng. archaeological site), *qendër historike* (Eng. historic centre), *qytet muze* (Eng. museum city), *park arkeologjik* (Eng. archaeological park), *muze arkeologjik* (Eng. archaeological museum), and *amfiteatër* (Eng. amphitheatre). These items usually follow transparent Albanian calque patterns, while English provides internationally stabilized equivalents used in heritage promotion.
- Culture-specific and historically marked terms: *kështjellë* (Eng. castle), *fortesë* (Eng. fortress), *han* (Eng. inn), *odë* (Eng. traditional reception room), and *pritje tradicionale* (Eng. traditional welcome). These entries require interpretation beyond direct lexical matching because they preserve meanings linked to defence, hospitality, memory, and local identity.

6.4 Types of Tourism: Terminological Productivity

The Types of Tourism field is especially regular because it follows the productive Albanian pattern *turizëm* (Eng. tourism) + *modifier*, normally corresponding to the English pattern modifier + “tourism”. This pattern

builds an internal taxonomy through calque, native modification, and hybrid formation, rather than through borrowing alone. Examples include *turizëm kulturor* (Eng. cultural tourism), *turizëm malor* (Eng. mountain tourism), *turizëm bregdetar* (Eng. coastal tourism), *turizëm rural* (Eng. rural tourism), *turizëm aventuror* (Eng. adventure tourism), *turizëm gastronomik* (Eng. gastronomic tourism), *turizëm biznesi* (Eng. business tourism), *turizëm shëndetësor* (Eng. health tourism), *turizëm dimëror* (Eng. winter tourism), and *turizëm veror* (Eng. summer tourism). Hybrid and transferred forms such as *agroturizëm* (Eng. agritourism), *ekoturizëm* (Eng. ecotourism), *turizëm i qëndrueshëm* (Eng. sustainable tourism), and *turizëm i jetës së egër* (Eng. wildlife tourism) connect tourism with agriculture, ecology, sustainability, and environmental discourse.

6.5 Genre/Register

Genre and register were coded because tourism terminology appears in distinct communicative settings: official websites, booking platforms, hotel descriptions, tour-operator pages, brochures, guidebooks, and visitor-oriented online materials. This follows the socio-communicative dimension of CTT, according to which terms acquire part of their specialized value from the users involved, the communicative purpose, the situation, and the register (Cabr  Castellv , 1999; Faber Ben tez, 2009).

The corpus shows clear register variation. *Zon  e mbrojtur* (Eng. protected area), *park komb tar* (Eng. national park), and *trash gimi kulturore* (Eng. cultural heritage) are mainly associated with formal institutional discourse. Booking and hotel sources favour practical service expressions such as *anulim falas* (Eng. free cancellation), *m ngjes i p rfshir * (Eng. breakfast included), *disponueshm ri dhomash* (Eng. room availability), *dhom  familjare* (Eng. family room), and *koment i verifikuar* (Eng. verified review). Promotional brochures, travel guides, and tour-operator texts favour more descriptive forms, including *itinerar kulturor* (Eng. cultural itinerary), *udh tim ditor* (Eng. day trip), *guid  lokale* (Eng. local guide), *qend r historike* (Eng. historic centre), and *shteg natyror* (Eng. nature trail). Genre/register coding therefore shows why institutional sources favour formal standardization, booking platforms favour service-based terminology, and promotional sources favour culturally marked description.

7 Terminological Relations

7.1 Generic-Specific Hierarchies

Generic-specific relations show that the corpus functions as a conceptual system rather than as a simple bilingual list. The broad Albanian term *akomodim* (Eng. accommodation) includes *hotel* (Eng. hotel), *hostel* (Eng. hostel), *bujtin * (Eng. guesthouse), *apartament* (Eng. apartment / flat), *q ndrim n  familje* (Eng. homestay), and *struktur  akomoduese* (Eng. accommodation facility). Transport terminology follows the same pattern: *bilet * (Eng. ticket) includes *bilet  avioni* (Eng. flight ticket), *bilet  vajtje-ardhje* (Eng. return ticket), and *bilet  elektronike* (Eng. e-ticket), while *fluturim* (Eng. flight) includes *fluturim direkt* (Eng. direct flight), *fluturim lidh s* (Eng. connecting flight), and *fluturim i vonuar* (Eng. delayed flight).

The most productive hierarchy is organized around *turiz m* (Eng. tourism), which generates a clear taxonomy of tourism types: *turiz m kulturor* (Eng. cultural tourism), *turiz m malor* (Eng. mountain tourism), *turiz m bregdetar* (Eng. coastal tourism), *turiz m rural* (Eng. rural tourism), *agroturiz m* (Eng. agritourism), *ekoturiz m* (Eng. ecotourism), *turiz m i q ndruesh m* (Eng. sustainable tourism), and *turiz m fetar* (Eng. pilgrimage tourism). These clusters show that Albanian has internalized international tourism taxonomies through productive native or hybrid patterns, while English often expresses the same concepts through compact compounds.

7.2 Synonymy

Synonymy is a recurrent feature of tourism terminology and a potential source of ambiguity in professional communication (Maci & Spinzi, 2025). In this corpus, synonymy operates within Albanian, within English, and across the Albanian-English pairing. It is therefore treated as controlled variation, not as free interchangeability.

Several cases illustrate this controlled variation. In digital booking discourse, *rezervim* (Eng.

booking / reservation) functions as the standard Albanian term, whereas *buking* (Eng. booking) appears as a direct platform borrowing. In hospitality vocabulary, *han* (Eng. inn) is historically and culturally marked, while *shtëpi pritëse* (Eng. guesthouse) is a transparent descriptive alternative. In visitor-service terminology, *udhërrëfyes turistik* (Eng. tourist guide / guidebook) and *guidë turistike* (Eng. tour guide) overlap but differ in register and referential scope. The distinction between *guidë turistike* (Eng. tour guide) and *agjent turistik* (Eng. tourist agent) is more restrictive, since the first denotes a guide who accompanies visitors and the second denotes a professional intermediary. Seasonal terms such as *sezon kulmor* (Eng. peak season) and *sezoni i lartë* (Eng. high season) are near-equivalent, but they should still be standardized according to source type and intended audience.

7.3 Polysemy and Terminological Ambiguity

Polysemy appears in both Albanian and English, particularly in visitor-service and activity terminology. Albanian *guidë* (Eng. guide) may refer to a person, as in *guidë turistike* (Eng. tour guide), to a written or digital source, as in *udhërrëfyes turistik* (Eng. guidebook / tourist guide), or to a locally based tourism actor, as in *guidë lokale* (Eng. local guide). English “guide” shows similar ambiguity. These examples confirm that equivalence depends on context, genre, users, and communicative purpose, in line with the Communicative Theory of Terminology and cognitive approaches to specialized translation (Cabré Castellví, 1999; Faber Benítez, 2009).

Ambiguity is especially important when it affects the description of tourism products. The distinction between *zhytje* (Eng. diving) and *zhytje me maskë* (Eng. snorkeling) is not only lexical; it signals different activities, equipment, levels of difficulty, safety expectations, and visitor experiences. A literal or vague equivalent may therefore mislead tourists, operators, or digital platforms. For this reason, the study treats ambiguous terms through corpus evidence, expert review, explanatory notes, and translation-strategy coding, in line with research that sees tourism discourse as technical, practical, cultural, and promotional at the same time (Li et al., 2022; Maci & Spinzi, 2025).

8 Translation Strategy: A Taxonomic Analysis

Translation-strategy labels are used as mutually exclusive coding categories for quantitative reporting, not as claims that a term has only one possible interpretation. When categories overlapped, a priority rule was applied: direct borrowing overrides hybrid formation when the whole unit is imported; hybrid formation overrides calque when an international element combines with Albanian morphology; descriptive translation overrides functional equivalence when the target form is explanatory; functional equivalence applies to lexicalized but structurally independent terms; and calque applies only when the internal components of the source expression are substantially preserved.

The six-category scheme used here draws on Fijo León and Fuentes Luque (2013) and Li et al. (2022), with refinements for Albanian-specific patterns:

- Calque (structural): Component-by-component translation that preserves the source structure, with Albanian word-order adjustment where necessary. Examples include *destinacion turistik* (Eng. tourist destination), *atraksion turistik* (Eng. tourist attraction), *park kombëtar* (Eng. national park), *rezervat natyror* (Eng. nature reserve), *plazh privat* (Eng. private beach), *sezon turistik* (Eng. tourist season), *kuzhinë lokale* (Eng. local cuisine), and *trashëgimi kulturore* (Eng. cultural heritage).
- Functional equivalent: The target-language term performs the same communicative function as the source term, although it is not a literal structural translation. Examples include *shteg ecjeje* (Eng. hiking trail), *pikë kufitare* (Eng. border crossing), *fluturim lidhës* (Eng. connecting flight), *pistë* (Eng. runway), *bujtinë* (Eng. guesthouse), *turizëm fetar* (Eng. pilgrimage tourism), *qëndrim në familje* (Eng. homestay), and *pikë referimi* (Eng. landmark).
- Direct borrowing: An international or English form adopted into Albanian with minimal or no modification. Examples include *hotel* (Eng. hotel), *resort* (Eng. resort), *terminal* (Eng. terminal), *spa* (Eng. spa), *rafting* (Eng. rafting), *check-in* (Eng. check-in), *check-out* (Eng. check-out), and

buking (Eng. booking).

- Descriptive translation: A longer Albanian phrase that makes the meaning of a compact English term explicit. Examples include *lodhje nga ndryshimi i orës* (Eng. jet lag), *peshë e lejuar e bagazhit* (Eng. baggage allowance), *dyqan pa taksa* (Eng. duty-free shop), and *dhomë me pamje nga deti* (Eng. sea-view room).
- Hybrid formation: A formation that combines Albanian lexical material with international prefixes, stems, or compounding elements. Examples include *agroturizëm* (Eng. agritourism), *ekoturizëm* (Eng. ecotourism), *mbi-rezervim* (Eng. overbooking), *biletë elektronike* (Eng. e-ticket), *turizëm gastronomik* (Eng. gastronomic tourism), *turizëm aventuror* (Eng. adventure tourism), *marketing turistik* (Eng. tourism marketing), and *menaxhim turizmi* (Eng. tourism management).
- Cultural substitution/zero equivalence: The Albanian concept exceeds what the English gloss conveys, or vice versa. Examples include *han* (Eng. inn), *odë* (Eng. guest room / traditional reception room), *kullë* (Eng. tower house), *shtëpi muze* (Eng. museum house), and *sofra tradicionale* (Eng. traditional dining setting).

Ambiguous cases were not counted as belonging to several categories at once. They were discussed during expert review and assigned to one reporting category according to the priority rule above. For example, *biletë elektronike* (Eng. e-ticket) was treated as a hybrid formation because an international element combines with an Albanian head noun, while *rrugë ecjeje* (Eng. walking route) was treated as a functional equivalent because “walking route” is not a literal component-by-component rendering of the Albanian phrase.

This priority-based coding also makes the standardization argument clearer. In practice, a term can invite more than one explanation, especially when an international stem is fitted into an Albanian syntactic frame or when a culturally marked Albanian form is rendered by a familiar English equivalent. Assigning one primary strategy keeps the quantitative interpretation stable, while the explanatory notes preserve borderline complexity. The approach is therefore both analytical and terminographic.

9 Term Dynamics

Term dynamics were coded to explain how lexical items acquire, broaden, or transfer specialized value in Albanian-English tourism discourse. In line with CTT, terms were treated as usage-based units whose specialized status emerges in communicative, social, and disciplinary contexts (Cabr  Castellv , 1999). Coding relied on corpus evidence: recurrence, stable form, semantic-field membership, and a clear function in institutional, accommodation, visitor-facing, or promotional communication. Items were labelled as terminologized, determinologized, or transterminologized only when the evidence supported the process; weak cases remained unmarked or were handled as borderline during expert review.

Terminologization was recorded when a general lexical item acquired a tourism-specific sense through stable modification or repeated use as the name of a route, service, facility, product, visitor practice, or destination feature. Thus, *rrug * (Eng. road / route) was not coded as a term in isolation, whereas *rrug  panoramike* (Eng. scenic route), *rrug  ecjeje* (Eng. walking route), *rrug  kulturore* (Eng. cultural route), and *rrug  turistike* (Eng. tourist route) were retained because they name recognized route types in tourism discourse. Similar cases include *zon  turistike* (Eng. tourist area), *qaf  mali* (Eng. mountain pass), *kapacitet akomodes* (Eng. accommodation capacity), *tarif  hyrjeje* (Eng. entrance fee), *hart  turistike* (Eng. tourist map), *qend r historike* (Eng. historic centre), and *kapacitet mbajt s* (Eng. carrying capacity). In these units, modifiers restrict the meaning of a general word and stabilize it within the tourism domain.

Determinologization was coded when a specialized, institutional, or platform-based tourism term moved into broader visitor-facing or everyday travel communication. This corresponds to the movement of terms beyond expert use into lay discourse (Meyer & Mackintosh, 1996). Items such as *akomodim* (Eng. accommodation), *dhom  standarde* (Eng. standard room), *dhom  familjare* (Eng. family room), *politik  anulimi* (Eng. cancellation policy), *anulim falas* (Eng. free cancellation), and *m ngjes i p rfshir * (Eng. breakfast included) retain their technical service value but now circulate in hotel descriptions, booking confirmations, platform interfaces, travel advertisements, and guest reviews.

Transterminologization was recorded when a term from another specialized field entered tourism

and was refunctionalized through tourism context, modifiers, or phraseological patterns. This process refers to the transfer of a term from one knowledge field into another, where it acquires a new domain-specific role (Gorokhova, 2021). In the corpus, *turizëm i qëndrueshëm* (Eng. sustainable tourism), *ekoturizëm* (Eng. ecotourism), *agroturizëm* (Eng. agritourism), *marketing turistik* (Eng. tourism marketing), *menaxhim turizmi* (Eng. tourism management), and *platformë rezervimi* (Eng. booking platform) retain traces of sustainability, ecology, agriculture, business, management, or digital commerce, while becoming operational in tourism communication.

This process layer made coding more transparent and helped explain Albanian-English asymmetry. It also prevented over-classification, since terms were not assigned to a process category simply because they appeared theoretically interesting. Albanian frequently forms explicit descriptive units, while English often uses compact, internationalized, or institutionally stabilized equivalents. Term dynamics therefore connects selection, structural analysis, and translation strategy in one analytical frame, and shows how tourism vocabulary develops through repeated use, cross-domain contact, and visitor-facing simplification.

10 Discussion

10.1 A Functional Taxonomy of Albanian-English Correspondence

Drawing together the structural, semantic, relational, and translational findings, the study proposes a functional taxonomy of Albanian-English terminological correspondence. The taxonomy is derived from the 300 coded pairs, and is not presented as a universal typology for all tourism language. Each type combines structural form, translation strategy, cultural specificity, and source-use evidence, which makes it useful in translation, tourism communication, terminology education, and glossary design (Buendía Castro, 2024; Maci & Spinzi, 2025).

- **Type I** covers stable international equivalents, where forms are identical or minimally adapted because they circulate through global mobility, hospitality, or institutional practice: *terminal* (Eng. terminal), *aeroport* (Eng. airport), *pasaportë* (Eng. passport), *hotel* (Eng. hotel), *UNESCO* (Eng. UNESCO), *vizë* (Eng. visa), *resort* (Eng. resort), and *spa* (Eng. spa).
- **Type II** covers systematic calques such as *park kombëtar* (Eng. national park), *kontroll sigurie* (Eng. security check), *destinacion turistik* (Eng. tourist destination), *trashëgimi kulturore* (Eng. cultural heritage), *qendër historike* (Eng. historic centre), *sezon turistik* (Eng. tourist season), and *sigurim udhëtimi* (Eng. travel insurance). These are structurally transparent and support standardization.
- **Type III** includes functional equivalents, where Albanian and English name the same concept through different structures: *rrugë ecjeje* (Eng. walking route), *mosparaqitje* (Eng. no-show), *pikë kufitare* (Eng. border crossing), *pistë* (Eng. runway), *fluturim lidhës* (Eng. connecting flight), *qëndrim në familje* (Eng. homestay), *turizëm fetar* (Eng. pilgrimage tourism), and *pikë referimi* (Eng. landmark).
- **Type IV** includes contested synonyms in which an Albanian form and an English borrowing circulate together, as in *regjistrim hyrës / check-in* (Eng. check-in), *regjistrim daljeje / check-out* (Eng. check-out), *udhërrëfyes / guidë* (Eng. guide), *rezervim / buking* (Eng. booking), and *shtëpi pritëse / guesthouse* (Eng. guesthouse). Such cases need standardization guidance rather than automatic replacement.
- **Type V** includes culture-specific terms whose Albanian meaning exceeds a simple English gloss: *han* (Eng. inn), *kështjellë* (Eng. castle), *odë* (Eng. traditional reception room), *kullë* (Eng. tower house), *sofër tradicionale* (Eng. traditional dining setting), and *zanat lokal* (Eng. local craft). These terms require concise explanation or glossing because literal equivalence may erase historical, cultural, or symbolic meaning.

The value of the taxonomy lies in making decisions explicit. It indicates when a calque is sufficient, when a functional equivalent is preferable, when an English borrowing has already entered Albanian practice, and when a cultural explanation is needed. It also shows Albanian-English tourism terminology moving along a continuum from international standardization to culturally marked local

expression.

10.2 Implications for Destination Promotion and Translation Practice

The findings have practical implications for Albanian tourism communication. Consistent Type II calques such as *park kombëtar* (Eng. national park) and *rrugë ecjeje* (Eng. walking route) can improve professional coherence and international discoverability. Where Type IV variants coexist, as in *rezervim / buking* (Eng. booking) and *regjistrim hyrës / check-in* (Eng. check-in), a bilingual standard can balance Albanian terminological identity with international usability.

For Type V terms, the study supports partial foreignization with concise explication (Li et al., 2022; Maci & Spinzi, 2025). A heritage text may retain *Kështjella e Gjirokastrës* (Eng. Gjirokastra Castle, the historic fortress of Gjirokastrë). This preserves cultural specificity while making the term accessible to international visitors.

10.3 Implications for Digital Terminology Resources

The results also point to the need for a structured Albanian-English tourism term bank. Each entry should record the Albanian form, English equivalent, semantic field, structural type, translation strategy, source type, register note, and usage comment. These fields are necessary because many tourism terms combine technical reference, service function, destination promotion, and cultural specificity.

The culture-specific note field is especially important for entries such as *amfiteatër* (Eng. amphitheatre), *fortesë* (Eng. fortress), *sit arkeologjik* (Eng. archaeological site), *qytet muze* (Eng. museum city), *kullë tradicionale* (Eng. traditional tower house), *zonë e mbrojtur* (Eng. protected area), *rezervat natyror* (Eng. nature reserve), *itinerar kulturor* (Eng. cultural itinerary), *turizëm fetar* (Eng. pilgrimage tourism), *sofra tradicionale* (Eng. traditional dining setting), and *kuzhinë lokale* (Eng. local cuisine). Their meaning depends on cultural setting, destination discourse, source context, and visitor-facing use rather than direct lexical equivalence alone.

Figure 2 illustrates the proposed structure of a digital Albanian-English tourism term-bank entry. Using *bujtinë* (Eng. guesthouse) as an example, it shows how linguistic, cultural, source-based, and usage information can be combined within a single record.

Figure 2 translates the proposed model into an entry for *bujtinë* rather than repeating the list of data fields. Its first section, Identification, assigns the record ID ALB-TOUR-002, presents *bujtinë* as the Albanian lemma (dictionary form) and preferred headword, gives “guesthouse” as the English equivalent, and separates the attested variants *bujtina*, *bujtinë tradicionale*, and *bujtinë malore*. The second section, Source Information, grounds this choice in accommodation listings, rural-tourism materials, and destination-promotion brochures produced between 2018 and 2025. These sources constitute the source group, while the register note - accommodation, cultural, promotional - corresponds to genre/register and identifies the communicative settings in which the term is used.

The third section, Linguistic and Terminological Classification, explains the relation between the two forms. Both *bujtinë* and “guesthouse” are single-word terms in Accommodation and Hospitality, but the strategy is functional equivalence/cultural substitution and their relation is partial equivalence. Guesthouse identifies the lodging function without carrying the Albanian associations of domestic hospitality, local food, family-based accommodation, and traditional reception. Figure 2 thus makes this cross-linguistic asymmetry visible instead of presenting the terms as fully interchangeable.

The fourth section, Validation and Usage, converts that analysis into guidance. The validation note identifies *bujtinë* as a reusable tourism term with cultural value, while the usage comment limits it to small-scale lodging linked to local reception and rural or traditional tourism. For contexts where “guesthouse”

DIGITAL TOURISM TERM BANK ENTRY	
Term ID: ALB-TOUR-002	
	1. Identification Albanian lemma/form: <i>bujtinë</i> English equivalent: guesthouse Variant forms: <i>bujtina; bujtinë tradicionale; bujtinë malore</i>
	2. Source Information Source type: accommodation listings; rural tourism materials; destination-promotion brochures Source year: 2018–2025 Register note: accommodation, cultural, promotional
	3. Linguistic and Terminological Classification Semantic field: Accommodation and hospitality Structural type in Albanian: single-word term Structural type in English: single-word term Translation strategy: functional equivalent / cultural substitution Terminological relation: partial equivalence
	4. Validation and Usage Validation note: reusable tourism term with cultural value Usage comment: refers to small-scale lodging connected with local reception, family hospitality, and rural or traditional tourism Culture-specific explanation: <i>Bujtinë</i> denotes more than a lodging facility. In Albanian tourism discourse, it may evoke domestic hospitality, local food, family-based accommodation, and traditional reception. The English term <i>guesthouse</i> is close but culturally less specific. Suggested English rendering: <i>bujtinë (traditional guesthouse)</i> Example: <i>bujtinë malore (eng. mountain guesthouse)</i>

Fig. 2 Sample digital term-bank entry: *bujtinë* (Eng. guesthouse)

alone would be too general, the entry recommends *bujtinë* (traditional guesthouse); the example *bujtinë malore* (Eng. mountain guesthouse) shows how a modifier can specify the accommodation type without removing the cultural base. Taken together, the four sections make the entry traceable and usable: translators can see where the term occurs, why the English equivalent is only partial, and when an explanatory rendering is required.

Figure 2 shows that a tourism term-bank entry should not stop at the Albanian term and its English equivalent. The example of *bujtinë* (Eng. guesthouse) is especially useful because it shows that some Albanian tourism terms carry meanings that cannot be fully transferred through a simple English equivalent. In such cases, the entry needs a validation note, a culture-specific explanation, and a suggested rendering that helps translators, students, tourism professionals, and destination communicators use the term consistently. The figure therefore supports the main argument of the study: Albanian-English tourism terminology requires an evidence-based, contextual, and culturally informed term-bank model rather than a simple bilingual list. Such a model can improve terminological consistency while preserving the cultural meanings that contribute to the identity and distinctiveness of Albanian tourism destinations.

10.4 Limitations

Three limitations should be kept in view. The validation procedure relied on expert co-annotation and consensus adjudication, not on a blind multi-annotator experiment with formal agreement statistics. The corpus is also text-based; spoken interactions in hotels, travel agencies, tourism offices, and guiding contexts may reveal further variation. Finally, claims about the Albanianized borrowing *buking* (Eng. booking) and the coexisting forms *check-in* and *regjistrim hyrës* are grounded in the selected sources rather than in a full frequency-by-genre analysis.

11 Conclusions

The findings show that Albanian and English share a strong reliance on multi-word tourism terminology, although the 300-pair corpus also points to different degrees of structural compression. Within the analyzed data, Albanian more often uses longer descriptive units, while English tends to express comparable concepts through shorter, often two-word or lexicalized forms. This pattern should be understood as a corpus-based tendency, not as a general rule for the two languages as a whole.

The semantic field results further suggest that, in the selected corpus, Albanian-English tourism terminology is particularly concentrated in Cultural Heritage and Destination, Accommodation and Hospitality, Customer Service and Digital Tourism, and Transport and Travel Documentation. This distribution indicates that the dataset is oriented not only toward hotel services and travel procedures, but also toward heritage, hospitality, destination identity, outdoor activities, and digital visitor services. Terms such as *han* (Eng. inn), *odë* (Eng. traditional reception room), and *pritje tradicionale* (Eng. traditional welcome) show that some equivalents need contextual interpretation when historical, cultural, or symbolic meanings are involved.

With regard to terminological relations and translation strategies, the analysis identified generic-specific hierarchies, controlled synonymy, polysemy, and six main translation strategies: calque, functional equivalence, direct borrowing, descriptive translation, hybrid formation, and cultural substitution. Taken together, these findings suggest that the Albanian-English pairs in this corpus are positioned along a continuum between international standardization and local cultural expression. Stable international equivalents and calques appear to support clarity and consistency, whereas culture-specific terms generally require explanation, glossing, or contextual interpretation.

In conclusion, the study offers a methodological, descriptive, and practical contribution within the limits of the present corpus. Methodologically, it applies a corpus-informed coding procedure to an under-described Balkan tourism language pair. Descriptively, it documents structural, semantic, relational, and translational patterns in Albanian-English tourism terminology. Practically, it proposes a functional taxonomy and term-bank model that may support translation, ESP instruction, tourism education, destination marketing, and digital tourism communication. Future research could test these findings across larger datasets, additional genres, spoken tourism encounters, and user-based evaluations of the proposed glossary.

References

- Akademia e Shkencave e Shqipërisë. (2006). *Fjalor i Gjuhës Shqipe*. <https://fjalorthi.com/>
- Akbarov, A. (2025). Linguistic and cultural features of tourism terminology: An analysis in a cross-cultural context. *Current Research Journal of Pedagogics*, 6(10), 34–39. <https://doi.org/10.37547/pedagogics-crjp-06-10-05>
- Alali, B. A., Ali, A. M., Azmin, A., & Sahar, R. (2024). The use of corpus analysis in analysing tourism texts. *South Asian Journal of Social Sciences and Humanities*, 5(6), 13–25. <https://doi.org/10.48165/sajssh.2024.5602>
- Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Astorino, C. M. (2013). *Viajando pela terminologia de agenciamento de viagens e turismo: Reflexões e proposta de dicionário multilíngue* [Doctoral dissertation, Universidade de São Paulo]. <https://doi.org/10.11606/T.8.2013.tde-13112013-122704>
- Bourigault, D., Jacquemin, C., & L’Homme, M.-C. (Eds.). (2001). *Recent advances in computational terminology*. John Benjamins Publishing Company. <https://doi.org/10.1075/nlp.2>

- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. Routledge. <https://doi.org/10.4324/9780203469255>
- Buendía Castro, M. (2024). Recursos bilingües y multilingües del dominio del turismo: Estudio comparativo con fines traductológicos. *Hikma*, 23(3), 1–27. <https://doi.org/10.21071/hikma.v23i3.16979>
- Cabré Castellví, M. T. (1999). *Terminology: Theory, methods and applications*. John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.1>
- Cabré Castellví, M. T. (2023). *Terminology: Cognition, language and communication*. John Benjamins Publishing Company. <https://doi.org/10.1075/ivitra.36>
- Calvi, M. V., & Santos López, L. J. (2020). From the Linguaturismo glossary to the Dictionary of Food and Nutrition: Proposal for a new electronic multilingual lexicography. In M. J. Domínguez Vázquez, M. Mirazo Balsa, & C. Valcárcel Riveiro (Eds.), *Studies on multilingual lexicography* (pp. 93–112). De Gruyter. <https://doi.org/10.1515/9783110607659-006>
- Castellano-Martínez, J.-M., & González-Rodríguez, R. M. (2026). The projection of tourist destination image through translation: A multilingual and qualitative analysis. *Tourism Culture & Communication*, 26(2), 143–160. <https://doi.org/10.3727/194341425X17518426871449>
- Cenni, I., & Goethals, P. (2020). Positive reviews on TripAdvisor: A cross-linguistic study of contemporary digital tourism discourse. *Onomázein*, Special Issue VII, 18–40. <https://doi.org/10.7764/onomazein.ne7.02>
- Cheng, T., & Bai, H. (2018). The enlightenment of corpus-based translation studies for Chinese-English translation of tourism texts in Xi'an. In *of the 3rd International Conference on Contemporary Education, Social Sciences and Humanities (ICCESSH 2018)* (Advances in Social Science, Education and Humanities Research, Vol. 233, pp. 932–935). <https://doi.org/10.2991/iccessh-18.2018.214>
- Cobb, T. (n.d.). Concordancer—English (Version 10) [Web-based concordance tool]. The Compleat Lexical Tutor. Retrieved July 16, 2026, from <https://www.lex tutor.ca/conc/eng/>
- Dictionary of leisure, travel and tourism. (2006). A & C Black.
- Daille, B. (1994). Study and implementation of combined techniques for automatic extraction of terminology. In D. D. McDonald (Ed.), *The balancing act: Combining symbolic and statistical approaches to language* (pp. 29–36). Association for Computational Linguistics. <https://aclanthology.org/W94-0104/>
- Delvizio, I. A., Silva, L. B. G. da, Jordão, M. N., & Kalaoum, F. (2025). Glossário e corpus de textos científicos do Turismo: Proposta metodológica de atualização. *Caderno Virtual de Turismo*, 25(2), 76–97. <https://doi.org/10.18472/cvt.25n2.2025.2276>
- dos Santos Moreira, A. C. (2012). A translation-oriented tourism term bank. *Tourism & Management Studies*, 2, 1066–1068. <https://www.tmstudies.net/index.php/ectms/article/view/291>
- Durán Muñoz, I. (2022). El trabajo terminográfico basado en corpus: El caso del recurso DicoAdventure. *Estudios de Traducción*, 12, 109–118. <https://doi.org/10.5209/estr.80584>

- Estopà Bagot, R. (2001). Les unités de signification spécialisées: Élargissant l'objet du travail en terminologie. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 7(2), 217–237. <https://doi.org/10.1075/term.7.2.06est>
- Faber Benítez, P. (2009). The cognitive shift in terminology and specialized translation. *MonTI: Monografías de Traducción e Interpretación*, 1, 107–134. <https://doi.org/10.6035/MonTI.2009.1.5>
- Fijo León, M. I., & Fuentes Luque, A. (2013). A corpus-based approach to the compilation, analysis, and translation of rural tourism terms. *Meta: Journal des traducteurs / Meta: Translators' Journal*, 58(1), 212–226. <https://doi.org/10.7202/1023817ar>
- Gandin, S. (2013). Translating the language of tourism. A corpus based study on the Translational Tourism English Corpus (T-TourEC). *Procedia - Social and Behavioral Sciences*, 95, 325–335. <https://doi.org/10.1016/j.sbspro.2013.10.654>
- Gorokhova, N. V. (2021). Determinologization and transterminologization processes in modern oil and gas discourse. In I. Savchenko (Ed.), *National interest, national identity and national security. Proceedings of the International Scientific Forum (NININS 2020)* (Vol. 102, pp. 329–335). European Publisher. <https://doi.org/10.15405/epsbs.2021.02.02.42>
- Imran, M. (2024). Digital tourism communication to strengthen the attractiveness of destinations. *Galore International Journal of Applied Sciences and Humanities*, 8(3), 55–66. <https://doi.org/10.52403/gijash.20240307>
- Kageura, K. (2012). *The quantitative analysis of the dynamics and structure of terminologies*. John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.15>
- Karimalieva, G. E. (2024). Analyzing tourism discourse: Linguistic dimensions and translation issues. *American Journal of Applied Scientific Research*, 10(4), 65–70. <https://doi.org/10.11648/j.ajars.20241004.11>
- Koval, N., & Blaguta, O. (2025). Pragmatic dimensions of tourism terminology: Cross-cultural analysis of ethno-specific reality terms in tourist discourse. *Bulletin of Science and Education*, (4)34, 105–116. [https://doi.org/10.52058/2786-6165-2025-4\(34\)-105-116](https://doi.org/10.52058/2786-6165-2025-4(34)-105-116)
- Kuanishbaeva, A. M. q. (2025). The impact of translation and multilingual services on international tourism. *American Journal of Philological Sciences*, 5(12), 72–74. <https://doi.org/10.37547/ajps/Volume05Issue12-18>
- L'Homme, M.-C. (2004). *La terminologie: Principes et techniques*. Presses de l'Université de Montréal. <https://doi.org/10.4000/books.pum.10693>
- Li, Q. J., Wu, R. R., & Ng, Y. (2022). Developing culturally effective strategies for Chinese to English geotourism translation by corpus-based interdisciplinary translation analysis. *Geoheritage*, 14(1), Article 6. <https://doi.org/10.1007/s12371-021-00616-1>
- Maci, S. M., & Spinzi, C. G. (2025). *Translating tourism*. Routledge. <https://doi.org/10.4324/9781003465331>
- Malamatidou, S. (2024). *Translating tourism: Cross-linguistic differences of alternative worldviews*. Palgrave Macmillan Cham. <https://doi.org/10.1007/978-3-031-49349-2>

- Méndez-Cendón, B., & Pérez-Ruiz, L. (2015). Compilation of a multilingual (Spanish / English / French / Portuguese) lexicon of rural tourism terms of Castile and León. In T. Poibeau, & P. Faber (Eds.), *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence (TIA 2015)*, Universidad de Granada, Granada, Spain, November 4–6, 2015. (pp. 197–200). CEUR.org. https://ceur-ws.org/Vol-1495/paper_25.pdf
- Meyer, I., & Mackintosh, K. (1996). Refining the terminographer's concept-analysis methods: How can phraseology help? *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(1), 1–26. <https://doi.org/10.1075/term.3.1.02meyer>
- Nandhakumar, B. (2016). *Dictionary of tourism* (1st ed.). Himalaya Publishing House.
- Orel Frank, T. (2022). Model of tourism terminology implementation in language for tourism purposes university courses. *JOLLT Journal of Languages and Language Teaching*, 10(2), 148–160. <https://doi.org/10.33394/jollt.v10i2.4778>
- Pearson, J. (1998). *Terms in context*. John Benjamins Publishing Company. <https://doi.org/10.1075/scl.1>
- Prima, V. (2020). Terminological field «Tourism» in English tourism discourse. *Language Teaching Research Quarterly*, 14, 69–79. <https://doi.org/10.32038/ltrq.2020.14.05>
- Rigouts Terryn, A., Hoste, V., & Lefever, E. (2020). In no uncertain terms: A dataset for monolingual and multilingual automatic term extraction from comparable corpora. *Language Resources and Evaluation*, 54(2), 385–418. <https://doi.org/10.1007/s10579-019-09453-9>
- Rojas-García, J. (2022). Semantic representation of context for description of named rivers in a terminological knowledge base. *Frontiers in Psychology*, 13, Article 847024. <https://doi.org/10.3389/fpsyg.2022.847024>
- Sagmeister, E. (2010). *Eine Korpusanalyse tourismusspezifischer Textsorten und Extraktion von Termkandidaten am Beispiel von Tourismustexten über die Stadt Wien und ihre Umgebung* [Master's thesis, University of Vienna]. <https://doi.org/10.25365/thesis.12374>
- Sager, J. C. (1990). *A practical course in terminology processing*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.44>
- Seytkasimov, D., & Seytbekov, N. S. (2026). Tourism terms in the Karakalpak language. *American Journal of Philological Sciences*, 6(4), 64–66. <https://doi.org/10.37547/ajps/Volume06Issue04-14>
- Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive approach*. John Benjamins Publishing Company. <https://doi.org/10.1075/tlrp.3>
- Wüster, E. (1979). *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Springer.
- Yulchiyeva, D. A. (2025). Metaphorical structures of tourism terminology in English and Uzbek. *American Journal of Philological Sciences*, 5(9), 93–94. <https://doi.org/10.37547/ajps/Volume05Issue09-24>
- Ziyotova, R. (2025). Lingua-cultural peculiarities of English and Uzbek tourism discourse. *Foreign Languages in Uzbekistan*, 11(3), 176–192. <https://doi.org/10.36078/1751035950>

Appendix A Coding spreadsheet used for the validation of Albanian–English tourism terminology candidates.

Table 2 Coding spreadsheet used for the validation of Albanian-English term candidates

ID	Albanian term	English term	Structural type (Alb.)	Incl./Excl.	Eval.			Explanatory notes
					1	2	3	
T00 1	<i>Dhomë me pamje nga deti</i>	Sea-view room	N+Prep+N	Included	+	+	+	Included as stable accommodation term
T00 2	<i>Ofertë javore</i>	Weekly offer	N+Adj	Excluded	-	-	-	Excluded as a generic promotional expression with no terminological value.
T00 3	<i>Hotel Tirana</i>	Tirana Hotel	N+N	Excluded (proper-name filter)	+/-	+/-	+/-	Proper name excluded; the base term hotel is coded separately when it functions as an accommodation term.
T00 4	<i>Udhëtim me guidë</i>	Guided tour	N+Prep+N	Included	+	+	+	Included as stable tourism term

Appendix B Integrated Albanian-English Tourism Terminology Dictionary

Table 3 Sample entries from the integrated Albanian-English tourism terminology dictionary

No.	Albanian term (terminological unit)	English equivalent	Albanian structural type	English structural type	Source of term	Short note
1	<i>aeroport</i>	airport	Borrowing/ N	N	Airport / airline info	Included - denotes transport/air-travel tourism.
2	<i>agjenci turistike</i>	tourism agency	N + Adj	N + N	National Tourism Agency	Included - denotes institutional/ promotion tourism.
3	<i>agjent turistik</i>	tourist agent	N + Adj	N + N	Tour operators	Included - denotes activity/itinerary tourism.
4	<i>agroturizëm</i>	agritourism	Hybrid formation/N	N	National Tourism Agency of Albania	Included - denotes destination- promotion tourism.
5	<i>akomodim</i>	accommodati on	N	N	Hotel websites	Included - denotes accommodation tourism.
6	<i>akomodim alternativ</i>	alternative accommodati on	N + Adj	Adj + N	Hotel websites	Included - denotes accommodation tourism.
7	<i>akses në plazh</i>	beach access	N + PP	N + N	Booking.com	Included - denotes beach/location and facility tourism.
8	<i>aktivitet me muzikë live</i>	live music event	N + PP + Borrowing	Adj + N + N	Tirana in Your Pocket	Included - denotes nightlife and events tourism.
9	<i>aktivitete kulturore</i>	cultural activities	N + Adj	Adj + N	Tirana in Your Pocket	Included - denotes cultural tourism.

Table 3 provides a representative sample of the Albanian–English tourism terminology dataset. The full dataset can be accessed by clicking the link below.

<https://www.mediafire.com/file/mobzn5l94ymrsua/300+tourism+terminology+Albanian-English.docx/file>

Note on terms: Inclusion principle: terms, terminological units, and multi-word terminological units with clear tourism value have been retained; loose non-terminological items have been excluded. Structural-type key: N = noun; Adj = adjective; PP = prepositional phrase; Gen = genitive/complement relation; Conj = coordination; Borr = borrowing/loan element; Hyb = hybrid Albanian/borrowed formation; Acr = acronym.

Appendix C Supplementary materials for term extraction

- Albanian National Tourism Agency. (n.d.). *National Tourism Agency*. Retrieved July 3, 2026, from <https://akt.gov.al/en>
- Booking.com. (n.d.). *Search hotels in Vlorë*. Retrieved July 3, 2026, from <https://www.booking.com/city/al/vlore.en-gb.html>
- Intrepid Travel. (n.d.). *Albania tours & holidays*. Retrieved July 3, 2026, from <https://www.intrepidtravel.com/de/albania>
- Lonely Planet. (n.d.). *Europe. Albania*. Retrieved July 3, 2026, from <https://www.lonelyplanet.com/destinations/albania>
- Responsible Travel. (n.d.). *Responsible travel*. Retrieved July 3, 2026, from <https://www.responsibletravel.com/holidays/albania/small-group>
- Tirana In Your Pocket. (n.d.). *Tirana guide*. Albania Experience Sh. P. K. Retrieved July 3, 2026, from <https://www.inyourpocket.com/albania/tirana>
- TripAdvisor. (n.d.). *Vlorë, Albania*. Retrieved July 3, 2026, from https://www.tripadvisor.com/Tourism-g678774-Vlore_Vlore_County-Vacations.html
- UNESCO World Heritage Centre. (n.d.). *Butrint*. Retrieved July 3, 2026, from <https://whc.unesco.org/en/list/570>
- Visit Gjirokastra. (n.d.). *Archaeological park of Antigonea*. Retrieved July 3, 2026, from https://visit-gjirokastra.com/article/archaeological_park_of_antigonea