

A Corpus-based Study of Tourism Destinations and Attractions in Indonesia

Eva Tuckyta Sari Sujatna^{1*}, Kasno Pamungkas¹,
Meita Lukitawati Sujatna²

¹ Universitas Padjadjaran, Indonesia

² Universitas Widyatama, Indonesia

*Corresponding author: eva.tuckyta@unpad.ac.id

DOI: <https://doi.org/10.24922/eot.v12i1.126018>

Article Info

Submitted:
January 17th 2025
Accepted:
March 25th 2025
Published:
March 30th 2025

Abstract

While tourism has been widely studied across various disciplines, there is still limited research on the connection between tourism and corpus linguistics. The purpose of this study is to examine the types of tourism destinations and attractions in Indonesia by analyzing linguistic patterns in a large-scale corpus. The data source is the Leipzig Corpora Collection, comprising Indonesian language texts published between 2008 - 2018 (excluding 2014). Using the Master Plan of Indonesian Tourism Development 2010 - 2025 as a classification framework, this study identifies and categorizes tourism destinations and attractions based on occurrences of the word “*wisata*” ‘tourism’. The classification includes three categories: nature, culture, and man-made attractions. It is further divided into 10 classifications and 39 sub-classifications. The findings reveal that nature-based tourism is the most frequently mentioned category, comprising 499 instances (52%), followed by man-made attractions (38%), and cultural tourism destinations and attractions (10%). The most common reference made when talking about nature tourism is land-based tourism destinations and attractions, followed by marine waters, island, and geoparks. Through this research, we gain an understanding of how tourism destinations and attractions in Indonesia are represented linguistically and demonstrate the utility of corpus linguistics as a tool in tourism studies.

Keywords: culture; Indonesia; man-made product; nature; tourism destination attractions

INTRODUCTION

Background

The Republic of Indonesia, commonly known as Indonesia, is a nation that covers over 17,000 islands in Southeast

Asia, making it the largest archipelago nation in the world. The geographical scale of it is balanced by the rich cultural, linguistic, and ecological diversity. Indonesia is home to hundreds of ethnic groups, languages, and natural ecosystems, which

have great potential for tourism development. Over the last few decades, tourism has come to be seen as a crucial driver of economic growth, employment, and cultural exchange in the country. As global travel has become more common, Indonesia has become a prominent tourist destination in Southeast Asia, with its UNESCO World Heritage sites, tropical beaches, ancient temples, and lively urban centers. As Mwinuka et al. mentioned that Indonesia is one of the Asian countries that are abundant in cultural attractions, natural beauty, and a variety of dishes with high culinary tourism (Mwinuka et al., 2024).

Indonesia is focusing on diversifying its economy away from traditional sectors like oil and gas, which means the tourism sector is playing an increasingly important role in its national development strategy. Tourism's benefits are not limited to national revenue; they also encompass regional development, infrastructure improvement, and international visibility. In recognition of its strategic importance, the Indonesian government launched the Master Plan of Tourism Development 2010–2025 (Regulation No. 55/2011), which outlines a comprehensive framework for classifying and promoting various types of tourism destinations and attractions, including nature-based, cultural, and man-made attractions.

Although the industry is becoming more important, research on Indonesian tourism has primarily concentrated on topics such as destination branding, tourism competitiveness, marketing strategies, sustainable development, and human resource capacity. While these studies have offered valuable insights, they tend to rely on qualitative case studies, survey-based analyses, or policy reviews. Few studies have explored how tourism is represented in the language of public discourse or how linguistic data can reveal prevailing narratives and trends in tourism development.

By combining tourism studies with corpus linguistics, this study bridges the gap by analyzing patterns in natural

language use through the use of large-scale language data. Specifically, this research utilizes the Leipzig Corpora Collection, a publicly available linguistic resource that includes a substantial compilation of texts in Bahasa Indonesia collected over a ten-year period (2008–2018, except for 2014). Through the analysis of occurrences of the keyword 'wisata' (meaning 'tour,' 'travel,' or 'tourism'), the study explores how various types of tourism and attractions are depicted in Indonesian texts through language.

The study's interdisciplinary nature and methodological innovation are the reasons why it is important. In the beginning, a corpus-based approach to tourism research is introduced, a method that is still underutilized in the Indonesian context. The second step is to utilize a decade's worth of genuine language data to provide empirical evidence of how tourism and attractions are represented in written discourse. The third point is that it complies with national tourism development priorities by aligning linguistic patterns with the official classifications set forth in the government's tourism master plan.

By uncovering the frequency and context in which various tourism destination types, such as nature, culture, and man-made, are mentioned in the corpus, this study contributes to a deeper understanding of tourism discourse in Indonesia. Furthermore, it provides a base for future research that could investigate how language affects perceptions of place, identity, and national branding in the tourism industry.

Research Objectives

This study is aimed at bridging the gap by employing corpus linguistic methods to analyze representations of tourism in Indonesia. Specifically, the study has two main objectives: (1) to identify and categorize Indonesian tourism destinations and attractions in the Leipzig Corpora Collection using the keyword “*wisata*” (tourism); and (2) to examine the frequency and

distribution of nature, culture, and man-made attractions within Indonesian texts over the ten-year period. By mapping linguistic patterns of the national tourism classification system, this research provides empirical insight into how different types of destinations and attractions are portrayed in everyday language, contributing to both tourism and corpus-based studies.

LITERATURE REVIEW

Indonesia which is officially called as the Republic of Indonesia and as one of the Southeast Asian countries, is an archipelago and it is the world's largest archipelagic state (Cribb & Ford, 2009). It has around 17,504 islands in 34 provinces with around 718 local languages and in 2010 it was reported that Indonesia has around 1300s tribes. In 2020, it was announced that the tourists visited ASEAN countries in 2019 were about 133 million, including Indonesia and this number was increased from the 2018 that was 130 million (Puspardini, 2020) while in 2015 Indonesia achieved the target of welcoming 10 million foreign visitors (Anonymous, 2016). It is in line with the argumentation that in 2019 the tourism revenue would be the largest contributor since oil and gas, coal and palm oil production continued to decline, while tourism continues to grow (Nugroho & Sihite, 2018). It is supported by Sari that tourism is one of the primary industries that contribute to the country's foreign exchange earnings in Indonesia (Sari, 2022).

Tourism

Tourism is defined as a temporary movement (Buckart & Medlik, 1974; Mathieson & Wall, 1982) and it is also defined as a combination of activities that could involve the relationship or interaction among tourists, tourism suppliers, host (government or communities) and surrounding environment in captivating and welcoming the visitors (Goeldner &

Ritchie, 2003) while others argue that tourism also has a close relationship with mobility (Shaw & Williams, 2004). Since tourism plays an important role in getting country's revenue or impact to the country's economy, it could be said that tourism could influence the economic activity of the country, including Indonesia especially for the tourists who visit their countries (Barros & Sousa, 2023; Dwyer et al., 2004; Hall & Page, 2002; Sujatna et al., 2019, 2020, 2021; Todd, 2001).

Many researchers did their studies on tourism destination, such as the concept of tourism destination (Saraniemi & Kylänen, 2011); the complexity of the tourism destination branding (Pike, 2005); the competitiveness of tourism destination (Enright & Newton, 2004); the application and usefulness of tourism destination loyalty (Oppermann, 2000). Beside the research mentioned earlier, several studies on Indonesian tourism have been done, such as Indonesian Tourism Human Resources Development in the Era of Society 5.0 (Darsana & Sudjana, 2022), Competitive Advantage Model for Tourism Destinations in Indonesia (Lesmana & Sugiarto, 2021), the Indonesian Tourism Destination via Indonesia. Travel @indtravel (Nugroho & Sihite, 2018), the development of transport and tourism destination based on Indonesian perspective (Henderson, 2009), and Challenges and Opportunities Facing Indonesian Tourism (Wall & Nuryanti, 1997). Besides the research mentioned, it is still limited research on tourism related to corpus linguistics, especially the tourism destination attractions in Indonesia as described in this paper.

Tourism, in this case leisure, has a big impact on people's quality of life, as mentioned by Auger (2020) who studied leisure and recreation have an important role in the quality of life of individuals and communities (Auger, 2020); Badia, et al. (2013) who did their study of the relationship between leisure and people with developmental disability quality of life (Badia et al., 2013); Brajsa-Zganec, Merkas,

and Sverko (2011) who did their research on leisure activities contribute to subjective wellbeing (Brajsa-Zganec et al., 2011). As an archipelago, Indonesia has many nature tourism destinations and Indonesia is also listed as one of UNESCO Global Geoparks (UGGPs) with six geoparks (Pamungkas et al., 2021). The six geoparks are Ciletuh-Palabuhanratu Geopark, Batur Geopark, Gunung Sewu Geopark, Rinjani Geopark, Toba Geopark, and Belitong Geopark while the first geopark recognized by UGGPs was Ciletuh-Palabuhanratu UNESCO Global Geoparks (CPUGGP) in 2018.

Corpus Linguistics

Corpus Linguistics (CL) is concerned as more frequent and popular strategies of the language (Baker & McEnery, 2005). Corpus is considered as data that could be connected to an empirical research and natural texts (Sujatna et al., 2019). Corpus as a language database could be identified as a repository of examples (McEnery & Hardie, 2012) and it is an electronically stored to make the reader easier in searching both spoken and written clauses (Jones & Waller, 2015). The corpus also could describe the potentially useful for linguistic research in different point of view (Baker, 2009). The collection of corpora which were collected by Leipzig Corpora Collection (LCC) coverages 20 languages from around the world in 2007 including Bahasa Indonesia (Biemann et al., 2007) and today has become 200 languages (Goldhahn et al., 2012) and today has become 293 languages. It was reported that Leipzig corpora collection - Indonesian in 2013 collected 74,329,815 Bahasa Indonesia clauses which consist of 7,964,109-word types and 1,206,281,985 tokens (*Leipzig Corpora Collection: Indonesian Mixed Corpus Based on Material from 2013*, n.d.).

The number of corpora in Leipzig Corpora Collection is identified as the newest and the biggest (Kwary, 2019). Some of researches on CL such as the

research trends in tourism and hospitality from 1991-2020 (Wang et al., 2022), the description of the word “tourism” in English based on Corpus of Contemporary American English (COCA) (Sujatna et al., 2020), loan words and expression in tourism discourse based on the BBC-travel corpus (Gandin, 2014), and translating the language of tourism as a corpus based study on Translational Tourism English Corpus (T-TourEC) (Gandin, 2013) but it is still limited to discussing Indonesian tourism related to corpus linguistics. This paper tries to figure out the Indonesia tourism destination attractions related to corpus, in this case Leipzig Corpora Collection - Indonesian in 2008 up to 2018 excluding 2014 since the data were not found.

From the collected of Bahasa Indonesia corpus, the present writers tried to identify the word “*wisata*” as the key word to find out the Indonesian tourism destination attractions. It was signified that the word “*wisata*” 27,755 times from 148,090,329 tokens in the data collected. From the 27,755 times of the word “*wisata*” it is only 983 corpora that describe the tourism destination attractions in Indonesia. The word “*wisata*” is discussed in this paper is Bahasa Indonesia word which could be translated into English as the word ‘tour’. The word “*wisata*” has a close relation to the word tourism or ‘*tourisme*’ in Bahasa Indonesia.

They are many ways of classifying tourism. Related to the type of accommodation used, tourism could be classified, such as resort tourism, campground tourism, and other (Fletcher et al., 2018). The other scientists categorize tourism based on primary purpose of travel such as business, visiting friends or relatives (VFR), other personal business, and pleasure (Hall & Page, 2002).

To classify the tourism destination attractions found in Leipzig Corpora Collection – Indonesia 2008 -2018 (excluding 2014), the taxonomy of the tourism product is needed. This paper was inspired by McKercher (2016) who offered five

classifications including Pleasure, Personal Quest, Human Endeavour, Nature, and Business with 27 Product Family and 90 Product Classes which was described in the Table 1 (McKercher, 2002).

Table 1. Tourism Product Taxonomy (adopted from (McKercher, 2002))

Pleasure	Personal Quest	Human Endeavour	Nature	Business
Food and Drink - Drink - Food - Hybrid - Knowledge enhancement	Personal History - Diaspora - genealogy/ roots - personal memory	Industrial - Science - Factories - Historic - Transport - Primary Industry	Winter participatory - Play - animal based - Motorised	Meetings
Leisure (general) - Shopping - Sightseeing/ exploring - Second homes - Photography	Religious - Pilgrimage - Sacred travel - Religious - Missionary Safaris - Spiritual Retreats	Built Heritage - Archaeology - Historic sites - Religious heritage - Military - Contemporary	Place based - Protected areas - Places of outstanding beauty - Unique land and water forms	Conventions - Congress - Symposia - Conferences - Summits
Indulgent - Sex - Nightclubs, bars and discos - Drug tourism - Party tourism - Kleptotourism	Medical/ Wellness - Procedures - Health - Wellness	People and intangible heritage - The 'other' - The 'same' - Cultural celebrations	Consumptive - Hunting - Fishing	Exhibitions - Trade shows - Consumer show
Personal Events - Family - Friends	Learning - Cultural exchange - Study tour - Service learning - Third age - Immersion Programmes - Personal development	Creative - Literary - Performance - Arts - Festival and Events - Creative Places	Adventure - Aerial - Aquatic - Terrestrial	
Built Attractions - Gaming - Purpose built attractions		Dark - Dark fun factories - Dark exhibitions - Dark dungeons - Dark resting places - Dark shrines - Dark conflict sites - Dark camps of genocide	Natural area and wildlife appreciation and learning - Ecotourism - Flora and fauna - Geotourism - Palaeontology	
Sport - Passive - Active		Museums and interpretative centres - Science and technology - Museums celebrating human endeavours		
Recreation - Passive - Active - Facility based				

To classify the tourism destination attractions in Indonesia, the government regulations of the Republic of Indonesia No. 55 Year 2011 Master Plan of Indonesian Tourism Development year 2010-2025 is referred. The regulation divides the development of tourist attraction into three major classifications. The three classifications are the attraction of nature, the attraction of culture, and the attraction of man-made product.

METHOD

This research applied a mix method, that is the combination of qualitative and quantitative methods to produce a

comprehensive approach in doing the research (Creswell & Creswell, 2018).

The first step applied in this research is data filtering. In this step, the writers chose the sources of the data related to time range and source type selections. The data used in this research was obtained from Leipzig Corpora Collection – Indonesia in the year 2008 up to 2018 (excluding 2014). It was collected from several sources which were categorized into news (including news tufs), newscrawl (including newscrawl tufs), web (including web tufs), blogs (mixed and mixed tufs), web public, and Wikipedia (including Wikipedia tufs). The composition of the sources is described in the following Figure 1.

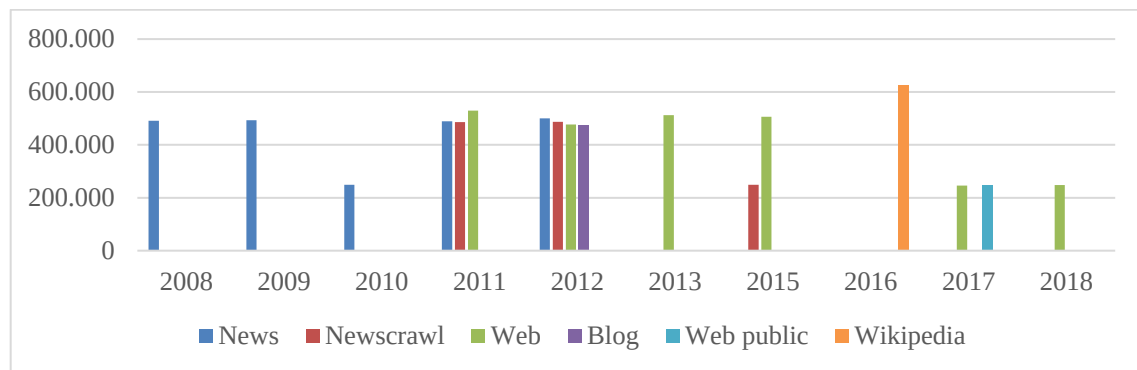


Figure 1. Leipzig Corpora Collection – Indonesian 2008 -2018 (excluding 2014)

From the corpus collected, the total number of Leipzig corpora collection – Indonesia 2008 -2018 (excluding 2014) used

as data is 7,309,185 corpora as described in the following Table 2.

Table 2. Total Number of Leipzig Corpora Collection 2008 -2018 (excluding 2014)

No.	Year	Number of Corpora
1	2008	491,099
2	2009	493,535
3	2010	249,344
4	2011	1,504,170
5	2012	1,936,880
6	2013	512,245
7	2015	754,935
8	2016	627,356
9	2017	491,680
10	2018	247,941
Total		7,309,185

From the Leipzig Corpora Collection 2008 – 2018 (excluding 2014), it was described that the highest number of corpora is in 2012, while the smallest one is 2018.

The second step is keyword filtering. This step concerned identifying and classifying the data. Based on the total number of 7,309,185 corpora, then the corpora were identified into the clauses containing the word “*wisata*”. It was informed that the word “*wisata*” found 27,755 times in the 7,309,185 corpora which is applied as the data of the research.

The third stage engaged the categorization. In categorizing the tourism

destination attractions in Indonesia, the three major classifications in the Republic of Indonesia’s Master Plan of Indonesian Tourism Development which was written in Government Regulation no 50 Year 2011 is applied. The three categories scripted in the regulation are the attraction of nature, the attraction of culture, and the attraction of man-made product. To propose a comprehensive categorization, the modification of the three major categories is offered as portrayed in the following Table 3 that represents the three major categorizations with ten modified classifications.

Table 3. Tourism Taxonomy in Indonesia

Tourism Destination Attractions	
Nature	Marine Waters Land Areas Island Geopark
Culture	Tangible Intangible
Man-made	Historical Sites Religious Places Pleasure Industry/ Technology

RESULTS AND DISCUSSION

Republic of Indonesia is known as an archipelagic country located between Pacific and Indian Oceans. Indonesia has around 17,000 islands and United Nation (UN) has officially accepted and registered 14,752 islands in Indonesia. UN has recognized Indonesia as an archipelagic state which has national territory that includes both land and water areas as a unified. The recognition only given by UN to the six nations: Indonesia, Fiji, Trinidad and Tobago, the Bahamas, Papua New Guinea, and the Philippines. As one of the six nations as an archipelagic state was proven

by the data found in the Leipzig Corpora Collection 2008 – 2018 (excluding 2014).

From the 27,755 tokens of the word “*wisata*”, it was only 966 words refer to tourism destination attractions in Indonesia. After identifying the 966 data, the three major categories (the attraction of nature, culture, and man-made product) are applied. The 499 data are sorted as the attraction of nature as the dominant one and 367 data are restricted into the attraction of man-made product while 100 data are categorized into the attraction of culture as the least number as illustrated in Figure 2.

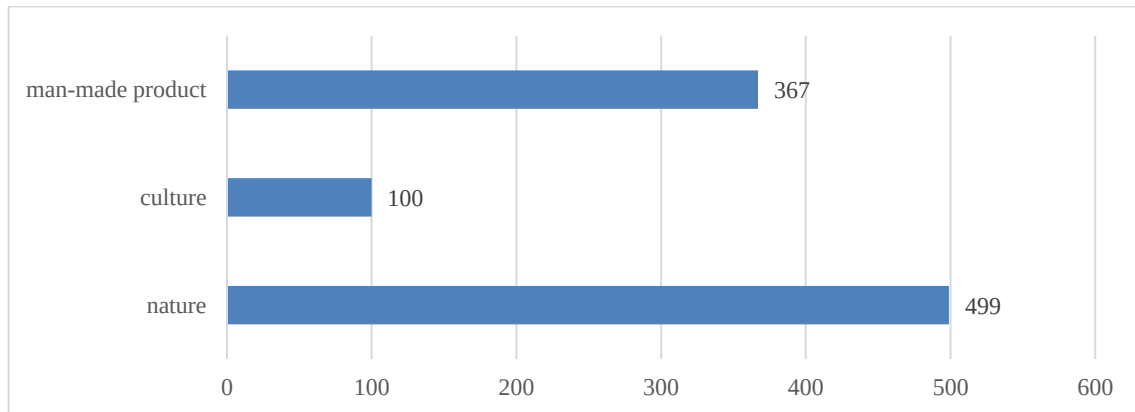


Figure 2. Number of Tourism Destination Attractions in Indonesia

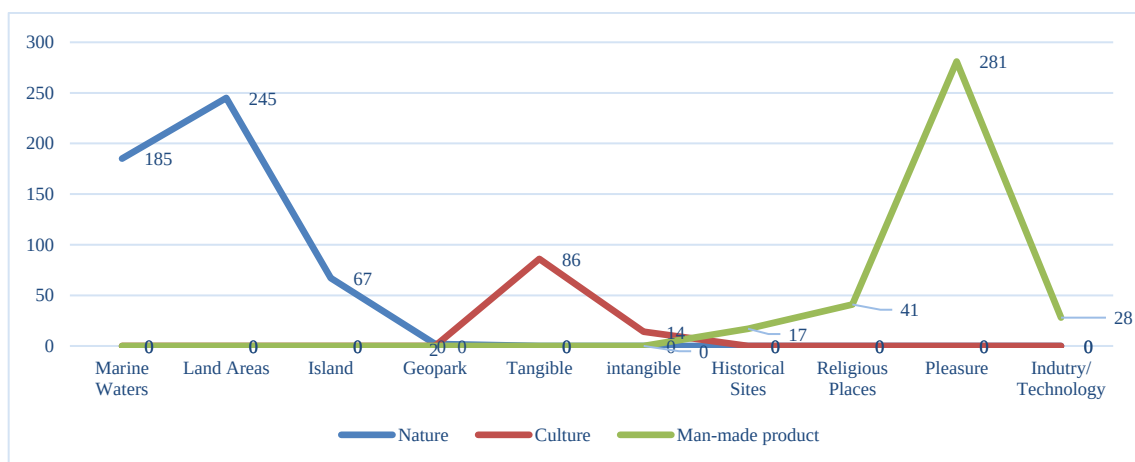


Figure 3. The Composition of Tourism Destination Attractions Corpora

The attraction of nature which has the highest number is divided into marine waters, land area, island, and geopark. From the collected data, the 499 found as the attraction of nature and it is dominated by land areas (245 corpora), marine waters (185 corpora), island (67), geopark (2 corpora) as described in Figure 3. To describe the nature tourism destination attractions found in detail, it could be seen in Table 4. The group of nature tourism destination attractions is the highest number of attractions, it is 499 nature attractions or 52% from the total data of tourism destination attractions.

From the 499 attractions, land area is the dominant one from the four classifications of nature attractions. The total number of land areas (245 corpora or 49%

from the total data of nature attractions) includes lake (25.3%), mountain (19.6%), river (8.6%), waterfall (8.6%), park (7%), cave (7%), Hill (7%), plantation (4.5%), forest (4%), crater (3.7%), agriculture (1.6%), karst (0.8%), cliff (0.8%), and pond (0.4%). The marine waters (185 corpora or 37% from the total data of tourism destination attractions) as the second total number of nature destination attraction containing beach (94% of the total data of marine waters), seascapes (2.1%), bay (2.1%), and marine park (1.6%). The third rank mentioned in the nature attraction is island tourism (67 corpora or 13,4% from the total of nature attractions) and the last is geopark tourism (2 corpora or 1% from the total of nature attractions).

Table 4. Nature Tourism Destinations and Attractions

Nature Tourism Destination Attractions			
Marine Waters	Land Areas	Island	Geopark
- Bay (4): Wisata Teluk Atong, Wisata Teluk Gelam, Wisata Teluk Karang, Wisata Teluk Montego - Beach (174): Wisata Pantai Kuta, Wisata Pantai Anyer-Carita, Wisata Pantai Carocok Painan, Wisata Pantai Watu Ulo etc. - Marine Park (3): Wisata Wakatobi, Wisata Bunaken, Wisata Raja Ampat - Seascapes (4): Wisata laut Bali, Wisata Laut Nusa Permai, Wisata laut Palabuhanratu, Wisata laut Tanjung Pasir Total Data: 185	- Agriculture (4): Wisata buah Kabupaten Sambas, Wisata Buah Naga Desa Tawang Sari, Wisata Tirta Arum Baru Kendal, wisata Sondokoro - Cave (18): Wisata Goa Garba, Wisata Goa Gudawang, Wisata Goa Jatijajar, Wisata Gua Lawa, etc. - Cliff (2): Wisata Tebing Breksi di Prambanan, Wisata Tebing Koja - Crater (9): Wisata Gunung Anak Krakatau, Wisata Kawah Gunung Galunggung, Wisata kawah purba di Gunung Patuha, Wisata Lava Tour di Tlogo Putri Kaliurang, etc. - Forest (10): Wisata Bening Sintang, Wisata hutan Logending, Wisata Deles Indah, Wisata Mangunan, etc. - Hill (18): Wisata Bukit Gundaling Brastagi, Wisata Bukit Jonggrang, Wisata Bukit Lawang Bahorok, Wisata Punthuk Setumbu, etc. - Karst (2): Wisata Lubang Sewu di tepi Waduk Wadaslintang, Wisata Lubang Tambang Mbah Soero - Lake (62): Wisata Danau Belibis Pulo Sarok, Wisata Danau Rawa Taliwang, Wisata Telaga Permai Batu Besaung, Wisata Bondowoso Rawa Indah Almors, etc. - Mountain (48): Wisata alam Tangkuban Parahu, Wisata Gunung Bengkuk Desa Cepedak, Wisata Gunung Mas Puncak Bogor, Wisata Gunung Papandayan Cisarupan, etc. - Park (18): Wisata Taman Krida Kudus, Wisata Taman Nasional Betung Kerihun, Wisata Taman Nasional Danau Sentarum, Wisata Taman Nasional Gunung Leuser, etc.	Wisata Kepulauan Derawan, Wisata Kepulauan Togean, Wisata pulau Bususen, Wisata Pulau Cangkir Kronjo, etc. Total Data: 67	Wisata Geopark Ciletuh-Palabuhanratu, Wisata Geopark Wonocolo Total Data: 2

Nature Tourism Destination Attractions			
Marine Waters	Land Areas	Island	Geopark
	- Plantation (11): Wisata Kebon Agung Imogiri, wisata kebun Kampoeng Kopi Banaran, Wisata Kebonrojo, Wisata Kebun Gowa, etc. - Pond (1): Wisata Cikuya Belawa - River (21): Wisata Kapuas Mahakam, Wisata Jembatan Akar Bayang, Wisata Lubuk Ikan Larangan, Wisata Kali Besar, etc. - Water fall (21) Wisata Air Terjun Akang Siwah, Wisata Desa Air Terjun Ciupang, Wisata Grojogan Lepo Imogiri, Wisata Roro Kuning, Desa Bajulan, etc. Total Data: 245		
Total Data of Nature Destinations & Attractions: 499			

The second category of tourism related to Government Regulation no 50 Year 2011 is the attraction of culture which is divided as tangible, and intangible is the least number (100 corpora or 10.4% from the total data of tourism destination attractions) found in the data. From the 100 corpora, the tangible (86 corpora) as the

dominant one found as the attractions of culture in Leipzig Corpora Collection 2008 – 2018 (excluding 2014) while 14 corpora as the intangible attractions of culture as described in Diagram 4. To illustrate the culture tourism destination attractions found in detail, it could be seen in Table 5.

Table 5. Culture Tourism Destinations and Attractions

Culture Tourism Destination Attractions	
Tangible	Intangible
- Cultural Heritage (41): Wisata Bantimurung, Wisata Kampoeng Wayang, Wisata kuliner makanan tradisional Sambas, Wisata Peninggalan Leluhur Kota Singkawang, etc. - Museum (6): Wisata Edukasi Museum Tekstil Indonesia, Wisata Museum Maritim Indonesia, Wisata Museum Sangiran, Wisata Huta Bolon Simanindo, etc. - Palace (7): Wisata sejarah Kerajaan Majapahit, Wisata Keraton Ngayogyakarta Hadiningrat, Wisata sejarah Keraton Kasepuhan Cirebon, Wisata sejarah Kesunanan Surakarta - Temple (3): Wisata Rambut Monte, Wisata Ratu Boko, Wisata Sejarah Candi Gunung Kawi	- Art (6): Wisata Batik Laweyan Solo, Wisata Batik Nakula Sadewa` di Sleman, Wisata Batik Pesindon, Wisata kerajinan gerabah Kasongan, etc. - Indigenous Life (8): Wisata Budaya Baduy, Wisata Kampung Budaya Betawi Setu Babakan, Wisata kampung seni Ubud - Kabupaten Gianyar, Wisata kampung adat Pulo, etc. Total Data: 14

Culture Tourism Destination Attractions	
Tangible	Intangible
- Traditional Village (29): Wisata Desa Brayut, Wisata Cinangneng Desa Cihideung Udik, Wisata Pentingsari Doto Yogantoro, Wisata Sade di Lombok Tengah, etc. Total Data: 86	
Total Data of Culture Destinations & Attractions: 100	

The cluster of cultural tourism destination attractions is the smallest number of attractions. It has 100 corpora containing 86% of tangible as the dominant and 14% of intangible cultural tourism destination attractions. The tangible contains of cultural heritage (41 corpora or 47.7%), traditional village (29 corpora or 33.8%), palace (7 corpora or 8.1%), museum (6 corpora or 6.9%), and temple (3 corpora or 3.5%).

The last classification of tourism is the attraction of man-made product. This classification is the second highest number of corpora found in the Leipzig Corpora Collection 2008 – 2018 (excluding 2014). The man-made product is distributed into four classifications, they are historical sites, religious places, pleasure, and

industry/ technology as described in Diagram 4. To identify the classification more details, Table 6 illustrates the sub-classifications of man-made product.

It was reported that 367 corpora (38% from the 966 data of tourism destination attractions) found as man-made product attractions. From the four categorizations, the pleasure (including leisure and recreation/ education) is the dominant number found in the data; it is 281 corpora (76.6% from 367 man-made product attractions). The second one is religious places (41 corpora or 11.2%) followed by industry/ technology (28 corpora or 7.6%), and the last is historical sites (17 corpora or 4.6%), as described in Table 6 as man-made products.

Table 6. Man-made Product Tourism Destinations and Attractions

Man-made Product Attractions			
Historical Sites	Religious Places	Pleasure	Industry/ Technology
- Building (3): Wisata Balai Kota DKI Jakarta, Wisata Linggarjati, Wisata Lawangsewu kota Semarang - Fort (9): Wisata Benteng Anoi Itam, wisata Benteng Kuto Besak (BKB), Wisata sejarah benteng Indra Patra, Wisata sejarah Benteng Otanaha peninggalan Portugis - Memorial/ Dark (2): Wisata tsunami di Banda Aceh, Wisata Lapindo	- Cave (1): Wisata religi di Sendangsono - Cemetery (12): Wisata religi Wali Songo, Wisata Religi Pesarean Aer Mata Ebu di Arosbaya, Madura, Wisata rohani Makam Raden Wangsa Muhammad, Wisata religius Makam Syech Burhanuddin, etc. - Church (2): Wisata religi di Kota Probolinggo, Gereja Merah, Wisata religi Gereja Katolik Hati Santa Maria	- Leisure (187): Wisata Balikpapan, Wisata Banda Aceh, Wisata Provinsi Riau, Wisata Provinsi Sulawesi Barat, etc. - Recreation/ Education (94): Wisata Edukasi Gondang Outbound (WEGO), Wisata Pemandian Air Panas Cipanas Buraran, Wisata Floating Market Lembang, Wisata loko uap Kota Sawahlunto, etc. Total Data: 281	- Reservoir/ Dam (28) Wisata Waduk Cengklik di Desa Ngargorejo, Wisata Waduk Penjalin - Desa Winduaji, wisata waduk Wonorejo - Kabupaten Tulungagung, Wisata Waduk Situ Gintung, etc. Total Data: 28

Man-made Product Attractions			
Historical Sites	Religious Places	Pleasure	Industry/ Technology
- Monument (3): Wisata Tugu Divisi ALRI, Wisata Tugu Khatulistiwa Pontianak, Wisata Monkasel Total Data: 17	- Mosque (11): Wisata religi Masjid Raya Baiturrahman, Wisata Masjid Menara Kudus di Kudus, Wisata religi Masjid Tuban, Wisata religi masjid Tua Kayu Jao, etc. - Statue (1): Wisata Rohani Patung Kristus Raja, Maluku - Well (2): Wisata sumur Dewi Kwan Yin, Wisata religi Sumur Pitu - Other (12): Wisata religi Gunung Kawi, Wisata Rohani ke Pulau Mansinam, Wisata religi di Jepara, wisata ritual Gunung Kawi, etc. Total Data: 41		
Total Data of Man-made Product Destinations & Attractions: 367			

CONCLUSION

The limitations of this study must be acknowledged. The categorization of the tourism destination attractions in Indonesia, as applied in this paper, is based on the framework outlined in the Indonesian government's Master Plan of Tourism Development (2010–2025), which classifies destinations into three major categories: nature, culture, and man-made attractions. Despite its usefulness for organizing and analyzing linguistic data, the framework has fixed boundaries that may not fully encompass the fluid, overlapping, or emerging forms of tourism that exist in practice. This structural limitation should be considered when interpreting the findings.

In the terms of sub-classifications, the study identifies three types of tourism destinations and attractions (nature, culture, and man-made products), including four specifics to nature-based tourism (499): land areas (245), marine waters (185),

islands (67), and geoparks (2). The culture, destinations and attractions could be defined as two specific things (100): tangible (86) and intangible (14). The third sub-classification, man-made products, including four distinct products (367): historical sites (17), religious places (41), pleasure (281), and Industry/ Technology (28). The data indicates that nature tourism is prevalent due to Indonesia's vast and diverse natural landscape, which includes rainforests, mountains, coral reefs, and volcanoes, which are deeply ingrained in the country's tourism identity. Besides the descriptive findings, this study makes several theoretical and methodological contributions.

First, it addresses a significant research gap by applying corpus linguistics to the study of tourism discourse in Indonesia. This is an interdisciplinary approach that remains underutilized in both fields. This study utilizes large-scale, natural language data to provide a data-driven method for identifying the representation of

tourism destinations and attractions in public discourse over time. The addition of this to existing qualitative research on tourism image and perception provides a scalable and replicable analytical framework.

Second, the study contributes to destination image theory by revealing which types of attractions dominate linguistic representations. The prevalence of nature-based tourism suggests a national tourism narrative that is anchored in environmental and ecological themes. This insight has implications for tourism branding, policy development, and regional planning, especially as Indonesia seeks to balance economic development with environmental sustainability.

Third, this research contributes to the broader field of discourse analysis in tourism studies. It underscores the impact of language on perceptions of place, identity, and value. The corpus data shows not just the types of destinations and attractions mentioned, but the relative importance of each category, which suggests underlying socio-cultural and policy-driven priorities.

In sum, while the study is limited by its reliance on a predefined classification system and a single keyword (“wisata”), it opens new avenues for research by integrating linguistic analysis with tourism studies. Additional keywords, exploring multi-word expressions, or analyzing discourse across various media platforms could be explored in future studies to expand on this work. Furthermore, a comparison of various regions or time periods could offer a deeper understanding of how tourism discourse evolves in response to changing national and global dynamics.

ACKNOWLEDGEMENT

The authors wish to thank the Rector of Universitas Padjadjaran, Indonesia to support the process of the research until this article is published. This research is funded through the Academic Leadership Grant (ALG) scheme with contract 1452/UN6.3.1/PT.00/2024.

REFERENCES

- Anonymous. (2016). *Tourism Industry Indonesia*. Indonesia Investments. <https://www.indonesia-investments.com/business/industries-sectors/tourism/item6051>
- Auger, D. (2020). Leisure in everyday life. *Society and Leisure*, 43(2), 127–128.
- Badia, M., Orgaz, M. B., Verdugo, M. A., Ulan, A. M., & Martinez, M. (2013). Relationships between Leisure Participation and Quality of Life of People with Developmental Disabilities. *Journal of Applied Research in Intellectual Disabilities*, 26, 533–545.
- Baker, P. (2009). The BE06 Corpus of British English and recent language change. *International Journal of Corpus Linguistics*, 14(3), 312–337. <https://doi.org/10.1075/ijcl.14.3.02bak>
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 4(2), 197–226.
- Barros, C. L., & Sousa, B. B. (2023). Relationship Management in SPA Tourism: An Approach to Cross-Border Regions. In M. A. Khan, N. Kumar, B. B. Sousa, & S. Sharma (Eds.), *Tourist Behavior: Past, Present, and Future*. Apple Academic Press, Inc.
- Biemann, C., Heyer, G., Quasthoff, U., & Richter, M. (2007). Monolingual corpora of standard size. *Corpus Linguistics* 2007, 1–13.
- Brajsa-Zganec, A., Merkas, M., & Sverko, I. (2011). Quality of Life and Leisure Activities: How Do Leisure Activities Contribute to Subjective Well-Being? *Social Indicators Research*, 102(1), 81–91.

- Buckart, A. J., & Medlik, S. (1974). *Tourism, Past, Present and Future*. Heinemann.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach*. SAGE Publications Ltd.
- Cribb, R., & Ford, M. (2009). Indonesia as an Archipelago: Managing Islands, Managing the Seas. In *Indonesia beyond the Water's Edge: Managing an Archipelagic State* (pp. 1-27). ISEAS–Yusof Ishak Institute. (pp. 1–27). ISEAS–Yusof Ishak Institute.
- Darsana, I. M., & Sudjana, I. M. (2022). A Literature Study of Indonesian Tourism Human Resources Development in the Era of Society 5.0. *Al-Ishlah: Jurnal Pendidikan*, 14(3), 2691–2700.
<https://doi.org/10.35445/alishlah.v14i1.2014>
- Dwyer, L., Forsyth, P., & Spurr, R. (2004). Evaluating tourism's economic effects: new and old approaches. *Tourism Management*, 25, 307–317.
- Enright, M. J., & Newton, J. (2004). Tourism destination competitiveness: a quantitative approach. *Tourism Management*, 25, 777–788.
<https://doi.org/10.1016/j.tourman.2004.06.008>
- Fletcher, J., Fyall, A., David, G., & Warhill, S. (2018). *Tourism: Principles and Practice* (sixth). Pearson.
- 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>
- Gandin, S. (2014). Investigating loan words and expressions in tourism Discourse: a corpus driven analysis on the BBC-travel corpus. *European Scientific Journal*, 10(2), 1–17.
<https://doi.org/10.19044/esj.2014.v10n2p%25p>
- Goeldner, C. R., & Ritchie, J. R. B. (2003). *Tourism: Principles, Practices, Philosophies*. Wiley-Blackwell.
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. *8th International Language Resources and Evaluation*, 759–765.
- Hall, C. M., & Page, S. J. (2002). *The Geography of Tourism and Recreation* (Second). Routledge.
- Henderson, J. (2009). Transport and tourism destination development: An Indonesian perspective. *Tourism and Hospitality Research*, 9(3), 199–208.
- Jones, C., & Waller, D. (2015). *Corpus Linguistics for Grammar*. Routledge.
- Kwary, D. A. (2019). A corpus platform of Indonesian academic language. *SoftwareX*, 9, 102–106.
<https://doi.org/10.1016/j.softx.2019.01.011>
- Leipzig Corpora Collection: Indonesian mixed corpus based on material from 2013*. (n.d.). https://corpora.uni-leipzig.de?corpusId=ind_mixed_2013.
- Lesmana, H., & Sugiarto. (2021). Formulating A Competitive Advantage Model for Tourism Destinations in Indonesia. *Journal of Asian Finance, Economics and Business*, 8(3), 237–250.
<https://doi.org/10.13106/jafeb.2021.vol8.no3.0237>
- Mathieson, A., & Wall, G. (1982). *Tourism: Economic, Physical and Social Impacts*. Longman Cheshire.

- McEnery, T., & Hardie, A. (2012). *Corpus Linguistics*. Cambridge University Press.
- McKercher, B. (2002). Towards a Classification of Cultural Tourists. *International Journal of Tourism Research*, 4, 29–38. <https://doi.org/https://doi.org/10.1002/jtr.346>
- Mwinuka, N. N., Phabiano, E., & Sumarwan, A. (2024). How Does the Official Indonesian Tourism Website Apply the Dimensions of The Experiential Marketing? *E-Journal of Tourism*, 109. <https://doi.org/10.24922/eot.v11i1.112542>
- Nugroho, A., & Sihite, J. (2018). Exploring the Indonesian Tourism Destination via Indonesia.Travel @indtravel. *Advances in Economics, Business and Management Research (AEBMR)*, 52, 29–32.
- Oppermann, M. (2000). Tourism Destination Loyalty. *Journal of Travel Research*, 39, 78–84.
- Pamungkas, K., Sujatna, E. T. S., Heriyanto, Haron, R., & Rosana, M. F. (2021). Brand Association of Ciletuh - Palabuhanratu Geopark Towards COVID-19 Pandemic and Sustainable Tourism. *International Journal of Sustainable Development and Planning*, 16(8), 1563–1574. <https://doi.org/10.18280/ijstdp.160817>
- Pike, S. (2005). Tourism destination branding complexity. In *Journal of Product & Brand Management* (Vol. 14, Issue 4). <https://doi.org/10.1108/10610420510609267>
- Pusparini, I. G. (2020). *Tourist Arrival to ASEAN Countries 2019 Reaches 133 million*. Seasia. <https://new.seasia.co/2020/02/11/revealed-tourist-arrival-to-asean-countries-2019>
- Saraniemi, S., & Kylänen, M. (2011). Problematizing the Concept of Tourism Destination: An Analysis of Different Theoretical Approaches. *Journal of Travel Research*, 50(2), 133–143. <https://doi.org/10.1177/0047287510362775>
- Sari, D. P. (2022). Slow Tourism Possibilities in Cimande Tourism Village – West Java. *E-Journal of Tourism*, 31. <https://doi.org/10.24922/eot.v9i1.82895>
- Shaw, G., & Williams, A. M. (2004). *Tourism and Tourism Spaces*. SAGE Publications Ltd.
- Sujatna, E. T. S., Heriyanto, Krisnawati, E., Amalia, R. M., & Pamungkas, K. (2020). Portraying the Word ‘Tourism’ in English: A Corpus Linguistic Study. *Sosiohumaniora*, 22(2), 181–189.
- Sujatna, E. T. S., Pamungkas, K., & Heriyanto. (2019). Names as branding on nature tourism destinations in Pangandaran, Jawa Barat - Indonesia: A linguistic perspective. *Pertanika Journal of Social Sciences and Humanities*, 27(2).
- Sujatna, E. T. S., Pamungkas, K., & Heriyanto. (2021). Tourism sites naming in Ciletuh geopark: Toponymic and morphological studies. *Kasetsart Journal of Social Sciences*, 42, 147–152. <https://doi.org/10.34044/j.kjss.2021.42.1.23>
- Sujatna, M. L., Sujatna, E. T. S., & Pamungkas, K. (2019). Exploring the Use of Modal Auxiliary Verbs in Corpus of Contemporary American English (COCA). *Sosiohumaniora*, 21(2), 166–172.

- Todd, G. (2001). World travel and tourism today. In A. Lockwood & S. Medlik (Eds.), *Tourism and Hospitality in the 21st Century*. Elsevier.
- Wall, G., & Nuryanti, W. (1997). Marketing Challenges and Opportunities Facing Indonesian Tourism. *Journal of Travel & Tourism Marketing*, 6(1), 69–84. https://doi.org/10.1300/J073v06n01_06
- Wang, G., Wang, H., & Wang, L. (2022). Research trends in tourism and hospitality from 1991 to 2020: an integrated approach of corpus linguistics and bibliometrics. *Journal of Hospitality and Tourism Insights*, ahead-of-p(ahead-of-print). <https://doi.org/10.1108/JHTI-09-2021-0260>