

## Tourism Translation: from Corpus to Machine Translation (and back)

Patrizia Giampieri

School of Law - University of Camerino, Italy  
patrizia.giampieri@unicam.it

Martin Harper

Centro Linguistico di Ateneo - University of Macerata, Italy  
martin.harper@unimc.it

### Abstract

Tourism language is characterised by features which make it distinct from any other Language for Special Purposes (LSP). There are many examples of corpus-based studies and corpus-driven translation in the tourism sector but little regarding machine translation. Fewer still are the case studies or research papers dedicated to a comparison between machine-translated and corpus-based translated tourism texts. This paper aims to fill this gap by investigating whether, and to what extent, Google machine translation (from Italian into English) of a variety of tourism texts can be considered reliable or, at least, acceptable. To this end, it compares machine translations of tourism texts to their respective corpus-based translations. The paper's findings uncover issues which mostly concern lexical and collocational choices, as well as a neglect of certain English writing conventions, such as those relating to clause structures, ego-targeting and figurative language. MT appears to perform well with informative and descriptive tourism texts, where sentences are simpler and no vivid language is involved. These, however, could hardly be considered representative of tourism texts, as a whole. The paper calls for advancements in MT algorithms in order to address certain lexical and collocational issues. Moreover, it is the opinion of the authors that MT in the tourism field is best left to translators capable of discerning accurate word usage in context.

**Keywords:** machine translation; corpus-based translation; tourism texts; tourism translation; tourism language

Il linguaggio turistico è contraddistinto da elementi che lo rendono unico e particolare rispetto ad altri linguaggi settoriali o tecnici. Molti sono gli studi o le traduzioni in ambito turistico mediante corpora; pochi sono invece i casi di traduzione automatica. Ancor meno sono gli articoli accademici in ambito turistico dedicati ad un confronto tra testi tradotti automaticamente e testi tradotti con l'ausilio di corpora. Il presente articolo mira a colmare tale lacuna prefiggendosi di verificare se ed in che misura la traduzione automatica dall'italiano all'inglese di testi

turistici tramite Google Translate può essere considerata affidabile o, almeno, accettabile. A tal fine, il presente studio confronta la traduzione automatica di testi turistici con le rispettive traduzioni effettuate mediante consultazione di corpora settoriali. I risultati ottenuti mostrano che la traduzione automatica tende a trascurare aspetti lessicali della lingua di arrivo, oltre che convenzioni di scrittura, quali la struttura delle frasi, il linguaggio figurato e l'uso di pronomi personali. La traduzione automatica sembra invece dare risultati soddisfacenti laddove i testi assolvono a funzioni meramente descrittive ed informative, con frasi meno complesse ed assenza di linguaggio vivo. Ciò, tuttavia, potrebbe difficilmente ricondursi alla categoria di linguaggio turistico. Il presente studio evidenzia quindi la necessità di affinare gli algoritmi dei programmi di traduzione automatica affinché si affrontino e risolvano le lacune lessicali segnalate. Infine, gli autori sostengono che la traduzione automatica in ambito turistico può essere effettuata se si è in grado di discernere l'uso delle parole nel contesto.

**Parole chiave:** traduzione automatica; traduzione mediante corpora; testi turistici; traduzione turistica; linguaggio turistico

## Introduction

### Tourism language

The language of tourism is peculiar because there are many elements which make it different from other LSPs, such as the importance of the cultural factor in the source and target language; informative and appellative (or persuasive) functions; the presence of vivid and figurative expressions; *linguaging* (i.e., the use of foreign words); grammatical elements such as clause structure and the particular use of moods or tenses (e.g., the imperative or future), and ways to address readers (e.g., by using the second singular or plural pronoun).

With regard to the cultural aspect, not only are translators called upon to render a language but they also have to transfer elements of one culture into another ([25], 143). Cultural differences can be challenging because they are not always crystal clear ([20], 18). Even so, they must be taken into account when addressing tourism texts ([38], 332). A case in point would be culture-bound words, which are terms or expressions referring to the customs and traditions of the source culture. Scholars have dedicated several research papers to explaining how to tackle these features. For example, Laviosa and Cleverton ([23]) suggest translation strategies such as paraphrasing, combining borrowings with literal translations and/or explanations (*ibid.*, 5). Other researchers propose diverse approaches ranging from domesticating to foreignising ([34]).

The importance of cultural terminology should also be considered ([8];[9];[10];[14];[26]). Diki-Kidiri ([8];[10]) argues that there is a correlation between culture and terminology, which is referred to as “cultural terminology”. Cultural terminology has two main objectives: 1) developing a terminological approach that takes into account cultural diversity and safeguards the needs of different communities, and 2) implementing methods to effectively develop language and culture ([8], 27-28). In addition, he draws parallelisms between individuals and the community where they live, as people and their community develop a specific culture and identity ([9], 21). Individuals are cultural beings that develop culture by exploring their environment and

accessing the new ([10]). These elements must be taken into account in the translation process. Therefore, both a re-contextualisation of (i.e., the creation of a new cultural point of view) and a reformulation in the target language are necessary ([9], 132).

Another reason tourism discourse is peculiar is that tourism texts serve two main communicative purposes: informative and appellative or persuasive ([2]). Their informative function is to provide the reader with technical or factual information whereas their appellative function is to produce persuasive effects. Some scholars argue that the latter often prevails over the former ([1], 11) whereas others claim that certain tourism texts present a balance between the two ([6], 104-5). In some cases, their informative and appellative purposes may overlap, with one enhancing the other ([2], 263).

Vivid and figurative language should certainly be considered when addressing the appellative function of this kind of writing. Euphoric language, too, has become a feature of tourism texts ([7], 65), as figurative expressions aim to reduce the sense of strangeness and transform something unfamiliar into something beautiful and familiar ([7], 173). In this respect, some scholars warn against an excessive use of vivid language, especially when translating into (British) English. The reasons are manifold: on the one hand, languages such as Spanish or Italian are more florid *per se* ([6]), whereas other languages are not. Therefore, translators from Spanish or Italian should not indulge in figurative language too much, otherwise they may make the target language sound unauthentic or artificial. On the other hand, the English language tends to be vivid only when promoting faraway destinations ([22]). As a result, using vivid expressions for continental destinations may be perceived as excessive.

Another strategy used in tourism texts is *linguaging*, which consists of using foreign terms to evoke an exotic feeling of adventure, or to make cultural references by resorting to culture-bound words ([7], 183-5). *Linguaging* is an important technique which translators must not disregard or underestimate since, otherwise, the target text may expose a cultural gap and they would have failed to create a connection between the source and target cultures ([5]).

Another thing that translators should take into account is clause structure, which follows a fixed order in English, with a subject, a verb and an object ([25], 151). This might not be the case for other languages, where clause structure may vary for reasons of emphasis and style ([25]). For example, adverbial phrases of place and time are rather infrequent at the beginning of a sentence in English tourism texts ([3], 143). In Italian, by contrast, adverbials of time and place, as well as the thematisation (or fronting) of objects and complements are very frequent ([25], 151).

Also, ego-targeting ([7], 187-92) is a technique whereby the tourist is addressed with the second singular person pronoun. Its aim is to make the addressee feel at ease. In languages other than English, the writer addresses the holiday-maker more formally, e.g. through the passive voice or with impersonal pronouns. In the Italian language, for instance, the impersonal *si* is fairly frequent ([20], 23). Only in particularly empathic situations is the tourist addressed with the second plural person *voi*.

The imperative and the *will* future are frequently used in English to encourage tourists to enjoy their destination ([31], 193; [13], 337) or to exert control over them ([7], 79-84; [31], 192).

Text genre should also be taken into account when addressing tourism translation. Text genre is argued to be intrinsically connected with terminology ([29], 263). In particular, Pecman and Kübler ([29]) posit that genre is a dimension that allows users to analyse terminology, especially

through textual corpora ([29], 264). Tourism language, in fact, has its own specific genres, which can be investigated via corpus analysis. Corpora help focus on specific terminological genre-related features and are useful to carry out contrastive analyses on registers, text types and domains ([29]). As textual sources are used to compile corpora, it is important to consider genre and its correlation with domain, corpus and text when approaching specific terminology ([29], 284).

### **Corpora**

Nowadays scholars argue that translation into a foreign language has become widespread and acceptable ([36]), especially in the tourism sector ([36]). In fact, students of Translation Studies are very likely to be confronted with tourism texts in the course of their careers ([36], 4-5). In this respect, corpora can be useful as they provide users with samples of language patterns and collocations (meaning both general and terminological collocations). Sinclair ([35], 170) defines collocations as “the occurrence of two or more words within a short space of each other in a text”. General collocations are words with a general meaning that frequently appear together in any genre of a given language ([24], 200). Terminological (or specialized) collocations, on the other hand, are lexical combinations that occur in specialised discourse ([24], 201). Within terminological collocations, a further distinction is made between genre-specific and domain-specific collocations ([28]).

Consulting corpora reduces first language interference ([18], 20) as well as semantic, idiomatic and syntactical errors ([21], 321). Amongst other things, corpora also allow users to find idiomatic expressions and figurative language ([18], 17) which are relevant to tourism texts ([7], 2).

For these reasons, the advantages of corpus-based translation in the tourism sector have been long highlighted by numerous scholars ([17];[12];[16];[18];[19];[39];[20]). Gatto ([17]), for example, considers the Web as corpus and warns against the many pitfalls of Internet searches in the tourism field and in other sectors. In particular, she suggests ways to retrieve reliable documents and materials. Durán Muñoz ([12]) discusses how corpora have become indispensable when tackling the language of tourism and develops reference comparable corpora for translations from Spanish into German and vice-versa. In her opinion, it is possible to overcome the existing lack of language resources in this way. Gandin ([16]) advocates the importance of corpus linguistics applied to translation training and practice, and reports the benefits of consulting parallel and monolingual corpora for tourism translations. [18] (2019 and 2020) explains how an *ad hoc* corpus can help students of Translation Studies find acceptable equivalences. She argues that *ad hoc* corpora reduce mistranslations in the case of tourism texts describing nature or adventurous walks. Turci and Aragrande ([39]) present a pilot study on the corpus-based translation (from Italian into English) of texts dedicated to heritage tourism. In their study, they highlight the benefit of corpus linguistics applied to ESP (English for Specific Purposes).

### **Machine Translation**

Machine translation (MT) is a process whereby computers replace humans and translate texts automatically. Less than a decade ago, MT was still not qualitatively reliable (see, for example, the harsh remarks on MT by Pierini [30], 99). Owing to rapid developments in technology, however, MT has now become widespread in many sectors and fields ([15]). It has performed

increasingly better and has been applied in many sectors, including tourism (see [11]). Rivera-Trigueros ([32]), for example, focuses on the English-Spanish language combination and claims that of all the MT programmes available, Google Translate has become the most used. On the other hand, Roig Allué ([33]) laments the many lexico-grammatical and syntactic shortcomings of Google Translate when tackling the English and Spanish language combination in the tourism field. However, these apparently different perspectives might be considered consistent with one another. Several scholars have, in fact, highlighted the importance of human intervention in MT for quality control purposes ([4]). Stewart ([37]), for example, remarks that MT cannot always be relied upon, especially with regard to collocations. As dictionaries do not provide a wide range of collocations, a certain amount of assistance is necessary in order to enhance translation quality. Hence, he posits that consulting sector-specific corpora can help identify collocations and tackle MT issues. Nonetheless, users must be trained to distinguish between word meanings and to recognize relevant and non-relevant usage. His final advice is to focus and rely upon the users' language intuition, especially in tourism texts, where "persuasiveness is a defining feature of the discourse" ([37], 128). As for the use of dictionaries in tourism translation training, general (bilingual) dictionaries together with corpora (or the Web as corpus) have been advocated by some scholars (see [36]) as a way of teaching students how to tackle tourism texts using tools that are easily accessible. Furthermore, they provide a general approach to the language that students must further elaborate (see [17]). Alternative language resources could be domain-specific terminographic resources (e.g., YourTerm CULT within the framework of the European project Terminology without Borders).

### **Research question**

Given the latest developments in MT and the asserted usefulness of corpora, this paper is aimed at exploring to what extent the machine translation of tourism texts (from Italian into English) is reliable and qualitatively satisfactory. To do so, the machine-translations of a variety of tourism texts are compared with the corpus-based translations into English of the same source texts.

The research questions that this paper wishes to pose are, therefore, the following: is MT reliable in the tourism field? To what extent can it be considered qualitatively satisfactory *vis-à-vis* corpus-based translation?

### **Methodology**

In order to investigate the reliability and quality of MT in the tourism field, a few texts translated automatically from Italian into English by Google Translate are compared with the corpus-based translations of the same source texts.

The original texts and the related corpus-assisted translations are sourced from the work by [20] (2021), where a variety of genres are dealt with. In particular, this paper evaluates the translation quality of five tourism texts dealing with the following topics: the description of a beach resort; a city train tour; a nature walk; tourism trends and rules for urban tourism. The genres addressed

are varied, and different writing strategies and styles might be expected. The five texts mentioned above are selected on the basis of an in-depth language analysis carried out in the book by [20] (2021), as well as on the basis of the excellent results obtained through corpus analysis. Moreover, the varied types and genres of the texts analysed are arguably sufficient to generalise the findings, at least as an initial study.

This paper, then, seeks to verify whether the machine translation of the five source texts might be equally satisfying. To this end, it explores not only the lexical and grammatical choices but also the persuasive techniques, the clause structures, the ego-targeting strategies and other writing devices.

### Analysis

The first text to analyse is the description of a beach resort. [Table 1](#) shows the source text (ST), the MT, and the corpus-based translation (CbT) of the source text. The source text and the translations are divided into sections in order to allow comparisons.

The corpus-assisted translation was carried out by consulting the Leeds Collection of Internet Corpora ([40]) and the Hoepli online bilingual dictionary (see [20] 2021, 155-65). Differences between the CbT and the MT are underlined and discussed at the bottom.

No.	ST	Google MT of ST	CbT of ST
1	LIDO SABBIAADORO <i>Situato sul litorale più suggestivo della costa barese, il Capitolo, il nostro stabilimento balneare è immerso in un paesaggio senza tempo: la Costa degli Ulivi.</i>	LIDO SABBIAADORO Located on the most <u>suggestive</u> coast of the Bari coast, the Capitolo, our <u>bathing establishment</u> is <u>immersed</u> in a timeless landscape: the Costa degli Ulivi.	LIDO SABBIAADORO Lido Sabbiadoro is located along the <u>picturesque</u> coast of the city of Bari named <i>Il Capitolo</i> .  Our <u>beach resort</u> is <u>surrounded</u> by the stunning landscape of Costa degli Ulivi where olive trees thrive in a timeless atmosphere.
2	<i>Questo lembo meridionale d'Italia accompagna ogni viaggiatore in un percorso fatto di colori e suggestioni dall'anima mediterranea, che culmina in spiagge sabbiose e dorate ed un mare limpido e cristallino.</i>	This southern strip of Italy accompanies <u>every traveler</u> on a <u>path made of colors and suggestions with a Mediterranean soul, which culminates</u> in sandy and golden beaches and a clear and crystalline sea.	This strip of land in the South of Italy will take <u>you</u> on a <u>colourful and fascinating journey across the Mediterranean Sea to</u> golden sandy beaches and the clearest sea water.

Table 1

The ST, MT and CbT of the description of a beach resort  
ST sourced and abridged from: <https://www.sabbiadorobeach.com/>

Note that the machine-translated text presents not only lexical issues but also a certain disregard

for collocations, not to mention the writing strategies, pertaining to the language of tourism in English.

Lexical issues include, for example, the following: “suggestive” (translating *suggestivo*) and “bathing establishment” (translating *stabilimento balneare*) in line 1, and “suggestions” (translating *suggerzioni*) on line 2. As for the modifier “suggestive”, the literature has discussed this frequently mistaken word at length (see [17], 93-5) and has proposed valid equivalences, such as “spectacular” or “breathtaking”. The CbT, for instance, proposes “picturesque”. This translation option might be considered satisfying, to a certain extent, although the terms proposed in the literature may seem more appropriate. Also, (*un percorso fatto di...*) *suggerzioni* is figurative and rendered as “fascinating (journey)” in the CbT, in contrast to the MT’s more literal “(a path made of ...) suggestions”.

In the [Table 1](#) above, there are also collocational shortcomings. For example, it seems improbable that a beach resort can actually be “immersed” in a “timeless landscape” (first line). In fact, if the phrase “*immersed in a timeless landscape*” is googled, only 9 hits are retrieved and these are mainly from Italian websites. Collocations, of course, are not established simply by their frequency of appearance on a Google results page; it could be argued that “immersed” does not frequently collocate with “timeless landscape” in a native English context. The CbT proposes the phrase “surrounded by a stunning landscape”. Another collocational issue is apparent in the expression “a path (...) which culminates” (second line). If the phrase “*path \* culminating|culminates*” is googled, webpages dealing with spiritual healing, enlightenment and Buddhism come to the fore, raising the question of whether this expression may be deemed appropriate in a text describing a beach resort. The CbT proposes “(this strip of land) will take you to”, which is more in keeping with the less figurative English writing style. The CbT also divides the long sentence of the first line into two shorter sentences, which is more typical of English clause structures.

In line two, there are some other figurative expressions that are translated literally, such as “a path made of colors” (which is rendered as “a colourful... journey” in the CbT) and “with a Mediterranean soul” (toned down to “across the Mediterranean Sea” in the CbT). It can be argued that moderating vivid expressions, as the CbT does, is more in line with the English writing style.

Finally, the CbT takes ego-targeting into account and proposes “will take you (on a ... journey)” to translate *accompagna ogni viaggiatore (in un percorso)*. On the contrary, the MT suggests the more literal “accompanies every traveller (on a path)”, neglecting the ego-targeting strategies of English tourism language.

The next text to be analysed concerns a city train tour. [Table 2](#) reports the ST and the related MT and CbT. The corpus-assisted translation was effected using the Web as corpus and Google advanced search techniques. The Cambridge monolingual online dictionary was also consulted (see [20], 143-7). The texts are divided into sections and the MT’s shortcomings are underlined in order to make the analysis clearer.

No.	ST	Google MT of ST	CbT of ST
1	<i>Il trenino turistico Explorer è il modo ideale per tutti, dai più grandi ai più piccoli per visitare e scoprire la città in modo del tutto nuovo e divertente.</i>	The Explorer tourist train is the ideal way for everyone, <u>from the largest to the smallest</u> , to visit and discover the city in a completely new and fun way.	The Explorer tourist train is the best way to look around and discover the city in an interesting and amusing way. It's for everyone – <u>from the youngest to the oldest</u> .
2	<i>Comodamente seduti sul fantastico trenino e accompagnati da una chiara e completa audio guida in italiano ed inglese.</i>	<u>Comfortably seated</u> on the fantastic train and accompanied by a clear and complete audio guide in Italian and English.	<u>You will be</u> comfortably seated in the sightseeing tourist train, and accompanied by an audio guide in Italian and English.

Table 2  
The ST, MT and CbT of the description of a tourist train city tour  
ST sourced and abridged from: <https://senigalliaexplorer.it/>

One of the first things that stands out in the first line of Table 2 above, is the fact that the long sentence of the ST is split into two in the CbT, where the phrases are better organised. They have, in fact, a subject, a verb and an object. There are no subordinate clauses or embeddings which would make the discourse somewhat intricate. It is evident that the MT does not comply with these writing techniques. In addition, there are collocational issues in the first line, such as the expression “from the largest to the smallest”, which should render *dai più grandi ai più piccoli*. The meaning of the source phrase is actually “from the youngest to the oldest”, as evidenced in the CbT. In the second line, there are also clause structure issues. The fronting of the adverbial phrase “comfortably seated” makes the MT rather awkward in an English-speaking context. As the literature remarks, such constructions should be avoided in the English language ([20], 21-2). The CbT, in fact, proposes a standard subject + verb construction at the beginning of the sentence.

The next text describes a nature walk. Table 3 reports the ST, the MT and the CbT. The corpus-assisted translation was carried out by consulting an *ad hoc* corpus (see [20], 2021, 213-23). MT shortcomings are underlined.



No.	ST	Google MT of ST	CbT of ST
1	<i>Partenza dall'ingresso principale del Parco della Chiusa.</i>	<u>Departure from</u> the main entrance of the Parco della Chiusa.	<u>The trail starts from</u> the main entrance of the Parco della Chiusa.
2	<i>In questo primo tratto è da segnalare Villa Sampieri Talon, ora oggetto di lavori di manutenzione.</i>	In this first section it is worth mentioning Villa Sampieri Talon, <u>now the subject of</u> maintenance works.	On this stretch, you will note Villa Sampieri Talon, <u>currently</u> undergoing repair.
3	<i>Proseguendo oltre ci si addentra tra la vegetazione attraverso uno stretto sentiero ("Sentiero delle Montagnole"), che, descrivendo un arco, ci porta sulla sommità del pianoro erboso e poco dopo alla Montagnola di Sopra, sede del Parco.</i>	<u>Continuing further we</u> enter <u>the vegetation</u> through a narrow path ("Sentiero delle Montagnole"), which, <u>describing an arch</u> , takes <u>us</u> to the top of the grassy plateau and shortly after to <u>Montagnola di Sopra, seat of the Park.</u>	<u>Follow</u> the narrow path through <u>the wood</u> ( <i>Sentiero delle Montagnole</i> ). The path <u>curves around and climbs up</u> to reach the top of the plateau. After a short distance, <u>you</u> reach <u>Montagnola di Sopra, the Park Headquarters.</u>

Table 3  
The ST, MT and CbT of the description of a nature walk  
ST sourced and abridged from: <http://www.trackguru.net/>

The first line in [Table 3](#) above is interesting as the MT, as expected, provides a literal translation. In a similar way to the previous text, the machine-translated sentence starts with infrequent fronting (i.e., “departure from”). The CbT, by contrast, re-phrases the source text by placing a subject and a verb at the beginning of the sentence (“the trail starts”). The second line of the MT presents lexical issues, since the phrase “now the subject of (maintenance works)” is a mistake. In fact, if “*the subject of maintenance works*” is googled, very few hits are obtained, and mainly in non-English (or non-American) domains. However, it could be argued that “currently under maintenance” is more frequent than “currently undergoing repair”. In fact, if the expressions “*currently under maintenance*” [site:.gov.uk](http://www.gov.uk) and “*currently undergoing repair*” [site:.gov.uk](http://www.gov.uk) are googled, more hits are obtained in the first search. The same results are obtained if the US Government’s domain is queried (command: [site:.gov](http://www.gov)). The commands [site:.gov.uk](http://www.gov.uk) and [site:.gov](http://www.gov) help obtain results only from the British and North American Government’s domains, respectively.

The third line is relevant for many reasons. Firstly, the MT sentence starts (again) with an infrequent gerund (“continuing further”). Moreover, the automatic translation system proposes the use of the first plural person pronoun (“we”, “us”) to render the impersonal “*si*” and the atonic pronoun “*ci*”. In these cases, as suggested by the literature, ego-targeting or the imperative should have been applied ([7];[31];[20], 23). The CbT, in fact, proposes either an imperative (“follow”) or the second person pronoun (“you”).

The MT presents other lexical shortcomings, such as the word “vegetation” to translate *vegetazione* (which is best rendered in the CbT as “the wood”), and the “seat of the Park”, which wrongly

translates *sede del Parco*. In this case, the correct words to use are “Park Headquarters” (see Author 1 2019, 17-9).

Finally, the MT makes a serious mistake by translating the figurative and idiomatic expression *descrivendo un arco* literally. The CbT, on the other hand, renders its meaning effectively with the phrase “curves around and climbs up” (see also [18], 16-7).

Finally, the CbT divides the long text of line three into three sentences in order to comply with English writing standards by making shorter, simpler phrases.

The next text deals with trends in the tourism sector. [Table 4](#) reports the ST, the MT and the CbT. The corpus-assisted translation was carried out by consulting an *ad hoc* corpus (see [20], 234-44).

No.	ST	Google MT of ST	CbT of ST
1	<i>Nel 2018, l'Italia ha raggiunto il record storico di oltre 428 milioni di turisti (+1,8% rispetto al 2017).</i>	In 2018, Italy reached an all-time high of over 428 million tourists (+ 1.8% compared to 2017).	In 2018, Italy achieved a new record of 428 million tourists (+1.8% compared with 2017).
2	<i>Le attività economiche connesse a questo comparto, sottolinea il Rapporto annuale Istat pubblicato oggi, generano il 6% del valore aggiunto dell'economia: una quota simile al comparto delle costruzioni.</i>	The economic activities related to this sector, underlines the Istat Annual Report published today, generate 6% of the added value of the economy: a similar share to the construction sector.	The annual report of the <u>Italian National Institute of Statistics (Istat)</u> published today remarks that economic activities related to tourism account for 6% of the value added of the economy – a similar percentage is noticed for the building sector.

Table 4  
The ST, MT and CbT of the description of tourism trends  
ST sourced and abridged from: <https://www.ilsole24ore.com>

It is remarkable that there are no lexical, grammatical or collocational issues in the MT in [Table 4](#) above. The only noticeable shortcoming concerns the acronym “Istat” which should have been explained by a descriptive term ([27], 148), as in the CbT.

The same results are obtained in the last text, which is an extract from a journal article focusing on new tourism regulations in the city of Amsterdam.

No.	ST	Google MT of ST	CbT of ST
1	<i>La sindaca di Amsterdam Femke Halsema sta valutando la possibilità di vietare ai turisti l'ingresso ai coffeeshop e ai locali a luci rosse della città.</i>	Amsterdam mayor Femke Halsema is considering banning tourists enter the city's coffeeshops and red-light bars.	Amsterdam Mayor Femke Halsema is assessing the possibility of prohibiting tourists from entering coffee shops and red light district bars.
2	<i>Questa iniziativa avrebbe l'obiettivo di contrastare i comportamenti molesti e di dirottare il turismo verso altre attrazioni.</i>	This initiative would have the aim of combating harassing behavior and diverting tourism to other attractions.	This initiative is aimed at curbing aggressive behaviour and diverting tourists to other tourist destinations.

Table 5

The ST, MT and CbT of a journal article on city tourism rules  
ST sourced and abridged from: <https://video.repubblica.it/>

Clearly, there are no problems with MT, here. In fact, the MT is well-structured and smooth. Probably, this is due to the fact that the text is purely informative and the clause structure is simple. The article, in fact, more closely resembles a statement of fact than a persuasive or figurative tourism text.

The main differences between the texts in [Table 1](#), [Table 2](#) and [Table 3](#), and those of [Table 4](#) and [Table 5](#) above can be found in the fact that the former are both descriptive and persuasive. Hence, the language is more elaborate: a certain amount of figurative language is used; the collocational and lexical choices are consistent, accurate and relevant, and sentences are longer as they contain subordinate clauses or embeddings. The latter, by contrast, are merely informative, the clauses are simpler and neither persuasiveness nor vivid language are involved.

Given the above, it might be deduced that purely informative and descriptive tourism texts can be successfully translated by MT. One wonders, however, to what extent such texts fully represent the variety of tourism discourse.

## Discussion

The analyses carried out in this paper highlight the fact that, at present, MT is not particularly useful or reliable in the tourism sector. Problems remain which MT does not yet address. The paper has brought to the fore the most relevant, which range from lexical and collocational shortcomings to writing style issues.

Some examples of lexical shortcomings are “suggestive” to translate *suggestivo* (CbT: “picturesque”), or “bathing establishment” to translate *stabilimento balneare* (CbT: “beach resort”) ([Table 1](#)), as well as “now the subject of (maintenance works)” to render *ora oggetto di (lavori di manutenzione)* (CbT: “currently undergoing repair”), and “seat of the Park” to translate *sede del Parco* (CbT: “Park Headquarters”) ([Table 3](#)).

Collocational issues include the erroneous selection of language pairs, such as “immersed” referring to a beach resort and “culminates” describing to a path (Table 1), or “from the largest to the smallest” when translating *dai più grandi ai più piccoli* (CbT: “from the youngest to the oldest”) (Table 2).

Moreover, MT does not take into account the English writing style of tourism discourse, as it proposes literal translations. Hence, sentences tend to be long with subordinate clauses or embeddings (like the source texts). For example, the non-defining relative clause in line 4 of Table 3 is too long: “which, describing an arch, takes to the top of the grassy plateau and shortly after to *Montagnola di Sopra*, seat of the Park”. The CbT splits this sentence into three parts to make the meaning clearer. In this way, the phrases are smoother and easier to follow.

MT is also prone to infrequent fronting. Sometimes there are adverbial phrases at the beginning of a sentence (e.g., “comfortably seated”, in Table 2); nouns (e.g., “departure from”), or gerunds (e.g., “continuing”) (both in Table 3) whereas the English language tends to be characterised by sentences starting with a subject and a verb.

Another writing style issue is ego-targeting. The Italian texts make use of the impersonal *si*, or the first plural person. English tourism texts, on the other hand, use the second singular or plural person “you” and imperatives. This is something that MT obviously has difficulty with (see line two of Table 1 and line three of Table 3).

Given that MT is mainly characterised by literal translations, it is evident that idiomatic expressions and figurative descriptions continue to be inadequately expressed. Although “a path made of colors” might arguably translate *un percorso di colori* (Table 1), “describing an arch” does not render the idiomatic *descrivendo un arco* referring to a path (Table 3).

Purely descriptive tourism texts such as those in Tables 4 and 5, however, are tackled successfully. The clauses are well-organised and the lexical choices are correct; they do not present any particular shortcomings, apart from an unaddressed acronym (Table 4). For these reasons, it could be deduced that MT is reliable as long as the text is merely descriptive. This, however, would lessen, if not challenge, the nature and *raison d'être* of tourism discourse, which is mostly appellative or euphoric (see [37], 128, in this regard).

Appendix 1 summarises the MT shortcomings presented and discussed above, together with the source phrases and the related corpus-based translations.

## Conclusions

This paper has sought to explore the reliability and effectiveness of machine translation for the tourism industry. To this end, a variety of tourism texts were translated automatically from Italian into English by using Google Translate. The machine translations were then compared with the corpus-based translations of the source texts.

The analysis exposed several problems in the MT, ranging from lexical and collocational shortcomings to unaddressed English writing style conventions. Automatically translated sentences were, in fact, often too long as they merely reflected the source text’s sentence structure. The same can be said of fronting, where nouns, adverbial phrases and gerunds wrongly appeared

at the beginning of the sentence. Also, idiomatic expressions and vivid language were rendered literally, thereby producing inauthentic or awkward language.

Surprisingly, MT performed well with highly descriptive and informative texts, where no issues were noted (apart from an unaddressed acronym).

In light of the above, it is evident that MT in the tourism field should be employed by experienced translators, or at least by those who can discern relevant word usages from non-relevant word usages, as already suggested by the literature ([37]). Hence, it is the opinion of the authors that MT still has a long way to go before it can be considered completely reliable in the tourism sector. Therefore, this paper calls for further developments in MT algorithms, which should take into account collocational uses and/or particular lexical choices characterising specific language pairs.

At present, corpora seem to be the best option for handling tourism texts, as they permit translators to find and corroborate collocations, samples of language patterns and word usages in context. For these reasons, corpora should be used as reliable language reference tools, even when users opt for MT.

The limits of this paper lie in the fact that only two languages were involved. Perhaps, more insightful results could have been obtained by analysing a greater number. Also, the analyses concerned only translations from Italian into English, without addressing translations from English into Italian. Future research might therefore embrace the machine translation of tourism texts from and into a wider variety of languages. Moreover, further studies may encompass a greater number of tourism texts, so that these initial findings might be either corroborated or challenged.

## Appendix 1

ST, MT shortcomings, types of shortcoming and CbT

Text Reference	ST	MT	Type of Shortcoming	CbT
Table 1 line 1	<i>Suggestivo</i>	Suggestive	Lexical	Picturesque
Table 1 line 1	<i>Stabilimento balneare</i>	Bathing establishment	Lexical	Beach resort
Table 1 line 1	<i>Immerso</i>	Immersed	Collocation	Surrounded
Table 1 line 2	<i>Ogni viaggiatore</i>	Every traveler	Writing style: Ego-targeting	You
Table 1 line 2	<i>Un percorso fatto di colori</i>	A path made of colors	Figurative language	A colourful journey
Table 1 line 2	<i>(Un percorso fatto di ...) suggestioni</i>	(A path made of ...) suggestions	Lexical + Figurative language	Fascinating (journey)
Table 1 line 2	<i>Dall'anima mediterranea</i>	With a Mediterranean soul	Figurative and vivid language	Across the Mediterranean Sea

Table 1 line 2	<i>(Un percorso ...) che culmina</i>	(A path ...) which culminates	Collocation + Figurative language	(A journey across ...) to
Table 2 line 1	<i>Dai più grandi ai più piccoli</i>	From the largest to the smallest	Collocation	From the youngest to the oldest
Table 2 line 2	<i>Comodamente seduti</i>	Comfortably seated	Writing style: Fronting	You will be comfortably seated
Table 3 line 1	<i>Partenza da</i>	Departure from	Writing style: Fronting	The trail starts from
Table 3 line 2	<i>Ora oggetto di (lavori di manutenzione)</i>	Now the subject of (maintenance works)	Lexical	Currently (undergoing repair)
Table 3 line 3	<i>Proseguendo oltre</i>	Continuing further	Writing style: Fronting	Follow
Table 3 line 3	<i>Ci si addentra</i>	We enter	Writing style: Ego-targeting	(Follow ...) through
Table 3 line 3	<i>La vegetazione</i>	The vegetation	Lexical	The wood
Table 3 line 3	<i>Descrivendo un arco</i>	Describing an arch	Figurative language	(The path) curves around and climbs up
Table 3 line 3	<i>Ci porta</i>	Takes us	Writing style: Ego-targeting	You reach
Table 3 line 3	<i>Sede del Parco</i>	Seat of the Park	Lexical	Park Headquarters
Table 4 line 1	<i>Istat</i>	Istat	Lexical	Italian National Institute of Statistics (Istat)

## References

- [1]. Agorni, Mirella. 2012. "Introduzione". In *Prospettive linguistiche e traduttologiche negli studi sul turismo*, edited by Mirella Agorni, 7-21. Milan: Franco Angeli.
- [2]. Agorni, Mirella. 2018. "Cultural Representation Through Translation: an Insider-Outsider Perspective on the Translation of Tourism Promotional Discourse." *Altre Modernità* 20 (11): 253-275.
- [3]. Baker, Mona. 1992. *In Other Words*. New York: Routledge.
- [4]. Chan, Sin-wai. 2018. *The Human Factor in Machine Translation*. New York: Routledge.
- [5]. Cesiri, Daniela. 2017. "Representing Venice's Local Culture to International Tourists - The Use of the 'Languaging' Technique in Websites in English." *Annali di Ca' Foscari. Serie occidentale* 51: 195-216.

- [6]. Cómite Narváez, Isabel and José María Valverde Zambrana. 2014. “How to translate culture-specific items: a case study of tourist promotion campaign by Turepaña.” *The Journal of Specialised Translation* 21: 71-112.
- [7]. Dann, Graham. 1996. *The Language of Tourism. A Sociolinguistic Perspective*. Oxon: CAB international.
- [8]. Diki-Kidiri, Marcel. 2000. “Une approche culturelle de la terminologie.” *Terminologies Nouvelles* (21): 27-31.
- [9]. Diki-Kidiri, Marcel. 2008. *Le vocabulaire scientifique dans les langues africaines. Pour une approche culturelle de la terminologie*. Paris: Karthala.
- [10]. Diki-Kidiri, Marcel. 2022. “Cultural Terminology.” In *Theoretical Perspectives on Terminology: Explaining terms, concepts and specialized knowledge*, edited by Pamela Faber and Marie-Claude L’Homme, 197-216. Amsterdam: John Benjamins.
- [11]. Dimitroulia, Titika. 2022. *The use of eTranslation in the tourism industry*. Accessed June 28, 2022. <https://blogs.ec.europa.eu/emt/the-use-of-etranlation-in-the-tourism-industry/>
- [12]. Durán Muñoz, Isabel. 2010. “A Corpus-based Ontoterminological Tool for Tourist Translations.” *International Journal of Translation* 22 (1-2): 149-165.
- [13]. Durán Muñoz, Isabel. 2012. “Analysing common mistakes in translations of tourist texts (Spanish, English and German).” *OnOmázein* 26 (2): 335-349.
- [14]. Fijo León, María I. & Fuentes Luque, Adrián. 2013. “A Corpus-based Approach to the Compilation, Analysis, and Translation of Rural Tourism Terms.” *Meta* 58 (1): 212–226. <https://doi.org/10.7202/1023817ar>
- [15]. Gambier, Yves. 2016. “Rapid and Radical Changes in Translation and Translation Studies.” *International Journal of Communication* 10: 887–906.
- [16]. Gandin, Stefania. 2016. “Teaching and learning the language of tourism as an LSP: corpus-based approaches.” In *New Insights into Corpora and Translation*, edited by Daniel Gallego-Hernández, 93-110. Newcastle upon Tyne: Cambridge Scholars Publishing.
- [17]. Gatto, Maristella. 2008. *From body to Web*. Bari: Laterza.
- [18]. Giampieri, Patrizia. 2019. *Corpus-based translation in the tourism sector: a case study with final year bachelor students*, *mediAzioni* 24: 1-32.
- [19]. Giampieri, Patrizia. 2020. *Volcanic experiences: comparing non-corpus-based translations with corpus-based translations in translation training*, *Perspectives* 29 (1): 46-63. DOI:10.1080/0907676X.2019.1705361
- [20]. Giampieri, Patrizia and Harper, Martin. 2021. *Tourism Translation: from dictionary to corpus*. Brienza: Le Penseur.
- [21]. Hansen, Gyde. 2009. “A classification of errors in translation and revision.” In *CIUTI-Forum 2008: Enhancing Translation Quality: Ways, Means, Methods*, edited by Martin Forstner, Hannelore Lee-Jahnke and Peter A. Schmitt, 313-326. Bern/Berlin: Peter Lang.
- [22]. Jaworska, Sylvia. 2017. “Metaphors we travel by: a corpus-assisted study of metaphors in promotional tourism discourse.” *Metaphor and Symbol* 32 (3): 161-177.

- [23]. Laviosa Sara and Valerie Cleverton. 2006. “Learning by Translating: A Contrastive Methodology for ESP Learning and Translation.” *Scripta Manent* 2 (1): 3-12.
- [24]. Lorente, Mercè, Martínez-Salom, Àngels, Santamaría, Isabel and Vargas-Sierra Chelo. 2016. “Specialized collocations in specialized dictionaries.” In *Collocations and other lexical combinations in Spanish*, edited by Sergi Torner Castells and Elisenda Bernal Gallen, 200-222. New York / London: Routledge.
- [25]. Maci, Stefania M. 2019. “‘The church was built on a basilica plan.’ Translating and mistranslating Italian churches’ panels.” *Altre Modernità* 21: 140-157.
- [26]. Malenkina, Nadezhda and Ivanov, Stanislav. 2018. “A Linguistic Analysis of the Official Tourism Websites of the Seventeen Spanish Autonomous Communities”. *Journal of Destination Marketing & Management* 9: 204-233. <https://doi.org/10.1016/j.jdmm.2018.01.007>
- [27]. Newmark, Peter. 1988. *A textbook of translation*. Hertfordshire: Prentice Hall.
- [28]. Pecman, Mojca. 2012. “Etude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques.” *The Journal of Specialised Translation* 18: 113-138.
- [29]. Pecman, Mojca and Kübler, Natalie. 2022. Text genres and Terminology. In *Theoretical Perspectives on Terminology – Explaining terms, concepts and specialized knowledge*, edited by Pamela Faber and Marie-Claude L’Homme, 263-290. Amsterdam: John Benjamins.
- [30]. Pierini, Patrizia. 2007. “Quality in Web Translation: An Investigation into UK and Italian Tourism Web Sites.” *The Journal of Specialised Translation* 8: 85-103.
- [31]. Pierini, Patrizia. 2008. “A warm welcome. Guaranteed. Aspetti dell’Inglese nei siti degli Enti Nazionali per il Turismo.” *Studi Linguistici e Filologici Online* 6: 163-202.
- [32]. Rivera-Trigueros, Irene. 2021. “Machine translation systems and quality assessment: a systematic review.” *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-021-09537-5>
- [33]. Roig Allué, Blanca. 2017. “The Reliability And Limitations Of Google Translate: A Bilingual, Bidirectional And Genre-Based Evaluation.” *Entreculturas* 9: 67-80.
- [34]. Sanning, He. 2010. “Lost and Found in Translating Tourist Texts Domesticating, Foreignising or Neutralising Approach.” *The Journal of Specialised Translation* 13: 124-137.
- [35]. Sinclair, John. 1991. *Corpus concordance collocation*. Oxford: Oxford University Press.
- [36]. Stewart, Dominic. 2012. *Translating Tourist Texts from Italian to English as a Foreign Language*. Naples: Liguori.
- [37]. Stewart, Dominic. 2019. “English for tourism in the non-native English classroom. Machine translation and corpora.” In *Teaching English for Tourism: Bridging Research and Praxis*, edited by Michael Ennis and Gina Petrie, 114-130. Abingdon: Routledge.
- [38]. Sumberg, Carolyn. 2004. “Brand leadership at stake: Selling France to British tourists.” *The Translator* 10 (2): 329-353.
- [39]. Turci, Monica and Gaia Aragrande. 2020. “On Translating Art and Heritage



Discourse from Italian into English: From a Learner Corpus to a Specialized Corpus.” In *The Language of Art and Cultural Heritage*, edited by Ana Pano Alamán and Valeria Zotti, 12-38. Cambridge: Cambridge Scholar Publishing.

[40]. Wilson, James, Anthony Hartley, Serge Sharoff and Paul Stephenson. 2010. “Advanced corpus solutions for humanities researchers.” In *Proceedings of the 24<sup>th</sup> Pacific Asian Conference on Language, Computation and Information*, 769-778. <http://anthology.aclweb.org/Y/Y10/Y10-1089.pdf>

### Online resources

[41]. Cambridge monolingual dictionary: <https://dictionary.cambridge.org>

[42]. Hoepli bilingual dictionary: <https://www.dizionari.repubblica.it>

[43]. Leeds: <http://corpus.leeds.ac.uk/internet.html>

[44]. YourTerm CULT: <https://yourterm.eu/yourterm-cult/> Occumque cori abore-rio quistrum explabo ratatec aborum fugit aut hari ilitas santur soluptia si voluptio vella comnit eossequatem etur?