Language and translation of Greek tourism websites: a corpus-based study

Giacomo Klein
Università degli Studi di Trieste
gklein@units.it

Kyriaki Kourouni
Aristotle University of Thessaloniki
kkourouni@enl.auth.gr

Abstract

This paper aims to analyse the main linguistic and translation features of Greek tourism websites. Tourism is the most flourishing branch of the Greek economy and a very large amount of language material is produced and frequently published on websites of tourism organizations and tourism- and travel-related agencies. Original texts were collected from Greek websites and analysed with corpus processing software programmes. A bottom-up approach was adopted, with particular emphasis on lexical, morphological, syntactic and rhetorical features as they arose from the automatic analysis of the collected corpora. The quantitative results were then used as the basis for hypothesis formation, leading to meaningful qualitative translation analysis.

Keywords

Tourism language, corpus linguistics, translation, Greek language, tourism websites.

1.   INTRODUCTION

Online tourism texts have been the focus of the first Greek language and translation course held at the Advanced School of Modern Languages for Interpreters and Translators of the University of Trieste. In this context, it became clear that certain lexical, morphological and syntactical characteristics repeat themselves creating a text type on its own. A quantitative analysis of Greek tourism texts provides us not only with a linguistic overview but it may also help translators dealing with this kind of texts by establishing some guidelines.

The present paper includes two main sections. In the first part, significant results of the quantitative and automatic analysis of Greek online tourism language are singled out. The second part presents a discussion over translation strategies implemented by translators– including explicitation, adaptation and omission –in order to present tourism-specific translation characteristics and a locus where translation, adaptation, rewriting, transcreation and advertising converge.

2.   INSTRUMENTS AND METHODS

2.1 THE CORPUS

The automatic analysis of Greek tourism websites was carried out with a corpus linguistic approach. A tourism corpus was compiled by extracting texts from the official websites of the Thessaloniki Tourism Organisation (T.T.O) (https://thessaloniki.travel/en/) and the Greek Tourism Organisation (G.T.O) (http://www.visitgreece.gr/). All sections of both websites were extracted and, as far as the G.T.O website is concerned, some editions of its newsletters have been included in the corpus as well.

In order to define the main features of Greek tourism language, a control corpus was also compiled, then automatically analysed and compared with the tourism corpus. The control corpus contains Greek Wikipedia texts that are thematically and quantitatively comparable with the content of the tourism websites. Another very important criterion is the total number of tokens of the corpus and the control corpus: both present the same size in order to allow an effective quantitative analysis.

Corpus description	Number of tokens (N)
Tourism corpus	129,143
Wikipedia Control corpus	133,512

2.2   THE AUTOMATIC ANALYSIS

Following compilation, the corpora were automatically analysed with TaLTaC2 and TreeTagger software. More specifically, TaLTaC2 was used in order to obtain the lexicometric measurements, namely information about the frequency of words occurring in the corpus. TreeTagger on the other hand provided specific lexical, morphological and syntactical information about the corpora. The quantitative corpus research focuses on some lexical, morphological and syntactic features that appear to be typical in tourism language (Cf. Dann, 2004) as for example loanwords, the use of the second person, the use of imperative or of hyperbolic language.

3.   DATA ANALYSIS

3.1   LEXICAL FEATURES

3.1.1   LEXICOMETRIC MEASUREMENTS

The first data obtained with the automatic analysis of the tourism and the control corpus are lexicometric measurements (table 1).

	Tourism corpus	Wikipedia control corpus
Tokens (N)	129,438	133,512
Types (V)	22,195	24,361
Type token ratio (V/N)	0.17	0.18
V/N%	17.15 %	18.25%
Hapax%	54.98 %	61.49%
N/V	5.83	5.48

Table 1, Lexicometric measurements.

While the type-token ratio is very similar in both corpora, there is an important difference as far as hapax legomena are concerned. Hapax legomena are words occurring just once in a corpus and a high hapax index indicates that we are dealing with a lexically rich and complex amount of texts. Tourism texts seem to present a poorer lexicon than the Wikipedia control corpus.

3.1.2 LEXICAL CATEGORIES

Lexical categories are the grammatical categories we can find in a corpus. In this section, we will compare the tourism corpus with the Wikipedia control corpus

in order to detect some of the most important lexical features of Greek tourism texts. This analysis takes place through an automatic lemmatisation process with the TreeTagger software.
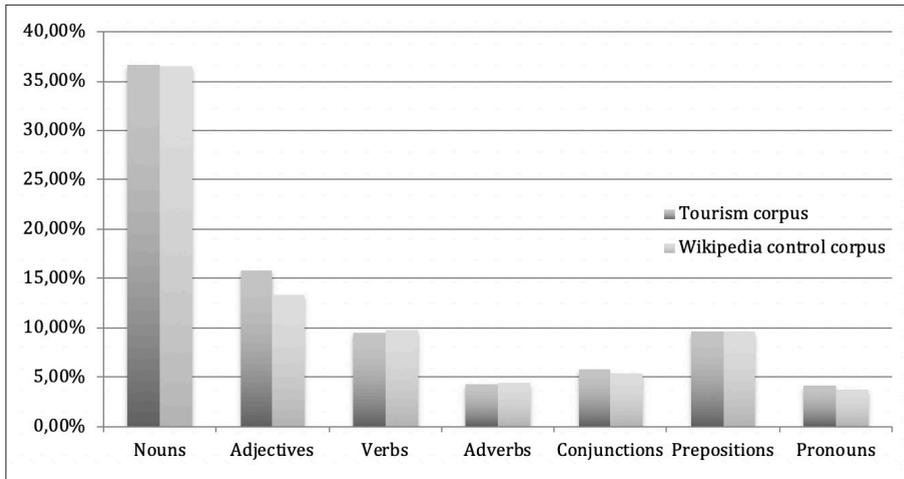

Table 2, content words.

Both corpora seem to have very similar lexical characteristics. However, it is possible to spot two main trends: the predominance of nouns in both the tourism corpus and the Wikipedia control corpus and the difference between the two corpora as far as adjectives are concerned. As Nigro (2006) points out in her analysis of tourism language, nominal style and nominalisation are key elements of any sectoral languages.

> Un fenomeno sintattico molto frequente nelle varietà specialistiche è la nominalizzazione, ovvero la trasformazione di un sintagma verbale in sintagma nominale. Tale processo garantisce oggettività al pensiero scientifico e risponde ai requisiti di sinteticità e concisione dei linguaggi specialistici (Nigro 2006: 57)

This tendency seems to be valid also in Greek tourism language where a nominal style is predominant. On the other hand, it is possible to notice an interesting adjectival trend. A higher concentration of adjectives has been spotted in the tourism corpus and this feature confirms the important role of adjectives in tourism language.

> The selection of adjectives, together with their pragmatic meaning, grammatical patterns and collocational behaviour play a fundamental role in the construction of the promotional discourse of tourism (Pierini 2009: 113)

3.1.3 LEXICAL DENSITY

Another measurement, which can be carried out based on lexical categories, is lexical density. In order to do so it is necessary to consider the ratio between content and function words. The former are words carrying a meaning on their own (nouns, adjectives, verbs and adverbs) while the latter carry a weak or no meaning at all when they are decontextualized (conjunctions, articles, prepositions etc.). As seen in table 3, there is no relevant difference between the two corpora as far as lexical density is concerned.
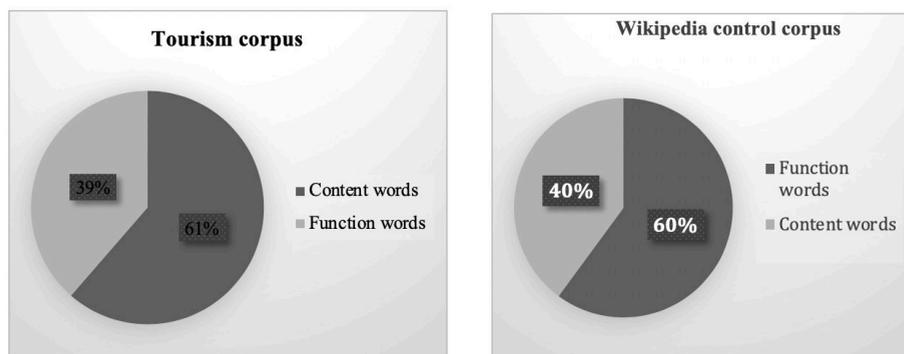

Table 3, lexical density

3.2 MORPHOLOGICAL FEATURES

In this section, some of the main morphological characteristics of Greek tourism language are analysed and compared with the results obtained through the automatic analysis of the Wikipedia control corpus. The analysis focuses on four features: verb tenses, second person, verb mood and adjective-adverb value (basic, comparative and superlative).

3.2.1 VERB TENSES

The following table shows some features about verb tenses used in Greek tourism language. As Nigro (2006) underlines, present tense plays a key role in this direction.

> Il tempo verbale maggiormente usato nel discorso specialistico è il simple present, in quanto richiesto dale funzioni comunicative quali la definizione, la descrizione, l'osservazione, l'esplicitazione, l'affermazione. Il simple present è usato per esprimere una verità scientifica generica (...) (Nigro 2006: 59)
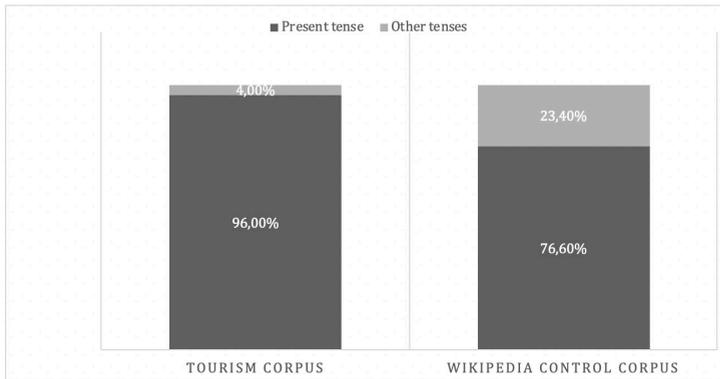
Table 4, verb tenses

This result reflects our expectations: both tourism and Wikipedia texts are main-ly informative but references to the past or the future seem to be rare in the for-mer while in the latter a higher number of different tenses coexist.

3.2.2   THE USE OF SECOND PERSON AND IMPERATIVES

The use of the second person (plural and singular) and of the imperative mood are typical strategies in narration since it is used as a means to involve the reader making them become an active participant of what is described in the text.

> Il linguaggio turistico fa un uso particolare dei pronomi 'we', 'our', 'us' rappresentanti l'industria turistica, e 'you', 'your' che indicano il, oppure, i soggetti riceventi. Chi scri-ve, infatti, si indirizza ad un gruppo collettivo con l'obiettivo di raggiungere e attrar-re il maggior numero di potenziali clienti. Il turista, invece, percepisce il messaggio come se fosse indirizzato esclusivamente a lui, una tecnica quest'ultima conosciuta come ego-targetting. (Nigro, 2006: p.61)

The use of second person in our tourism corpus and in our control corpus was examined.

	Second person sg. (%)	Second person pl. (%)	Total use of second person (%)
Tourism corpus	15.25%	14.87	30.12%
Wikipedia control corpus	1.05%	0.29%	1.34%

Table 5, the use of second person

	Imperative mood (tokens)	Imperative mood (%)
Tourism corpus	331	2.92%
Wikipedia control corpus	37	0.31%

Table 6, the imperative mood

The results stemming from the automatic analysis in tables 5 and 6 reveal an interesting trend in both the use of the second person and in the use of imperative mood: the comparison between the tourism corpus and the Wikipedia control corpus indicates that Greek tourism language abounds with phatic elements:

> As to the 'phatic' function, whereas in oral communication it is easier to keep the channel and psychological contact open, in a written/pictorial context a conversational style has to be consciously adopted in order to maintain the interest of the reader/ beholder (…). (Dann 1996: p.37)

3.2.3   Value of adjectives and adverbs

It is often claimed that comparative and superlative adjectives play an important role in tourism language. The idea behind this theory is that tourism texts move on the edge between information and advertisement.

> Un'altra caratteristica che distingue il testo turistico è l'uso dei superlativi (…). Il linguaggio turistico parla solo in termini brillanti e positivi (…), è una forma di "extreme language" in cui l'uso del superlativo è rigoroso. (Nigro 2006: 59)

Basic, comparative and superlative adjectives in our two corpora (table 7) were analysed accordingly:

	Basic	Comparative	Superlative
Tourism corpus	92.67%	3.78%	0.38%
Wikipedia control corpus	92.93%	4.14%	0.47%

Table 7, value of adjectives and adverbs

The automatic analysis points out that the use of comparative and superlative adjectives and adverbs in the tourism corpus is actually less dense than in the Wikipedia control corpus. The use of this feature seems not to play a strategic role in Greek tourism website language as it probably does in other languages or in different contexts.

3.3 SYNTAX

In this section, certain textuality and syntactic features that emerged during the analysis of Greek tourism language were examined. In particular, focus was on the average length of sentences and on subordinate and coordinate preposition data. The analysis through the TreeTagger software permitted a classification thereof.

3.3.1 SENTENCE DATA

The total number of sentences as well as the average length of sentences in both corpora (table 8).

	Total number of sentences
Tourism corpus	5,363
Wikipedia control corpus	5,176

Table 8, total number of sentences

	Average length of sentences (words)
Tourism corpus	24.1
Wikipedia control corpus	25.7

Table 9, average length of sentences

The automatic analysis shows a slightly higher average length of sentences in the Wikipedia control corpus (control with average number of words per sentence in Greek).

3.3.2 COORDINATE AND SUBORDINATE PREPOSITIONS

The automatic analysis with TreeTagger permitted the extraction of information related to the number of coordinate and subordinate prepositions in both corpora. It must be taken into account that the number of given prepositions does not always correspond with the number of equivalent coordinate and subordinate prepositions. Nevertheless, prepositions can lead to some insights in this sense.

	Coordinate prepositions	Subordinate prepositions
Tourism corpus	93.98%	6.02%
Wikipedia control corpus	90.14%	9.86%

Table 10, coordinate and subordinate prepositions

The results show a very high percentage of coordinate clauses indicating a large predominance of parataxis in both the tourism corpus and in the Wikipedia control corpus. The data also presents a trend towards even more coordinate prepositions in Greek tourism texts. A possible explanation to this trend is that in tourism web writing there is the need to create a simple and direct message in order to approach as many readers/consumers as possible.

4. The translation perspective

4.1 The idea-the setting

The main idea was to use material from the Greek corpus analysis as a springboard and guide for targeted analysis of the translations from the corresponding tourism websites into English. By adopting an essentially bottom-up approach, the aim was to shift focus and find information of use to professional translators (cf. Aston 1999, 2001) and, consequently, strengthen links between (linguistic) theory and (translation) practice. Corpus-linguistics data led the way into forming sets of assumptions in view of future practical application. Representative examples were selected accordingly and are presented in the same sequence as corpus analysis data in the previous part above. The main question asked in all cases was "how is this finding translated into translation strategies?"

4.2. Features, observations and selected examples

4.2.1. Lexical features

Regarding lexicometric measurements, the high Hapax figure in the Greek corpus is already encouraging research-wise because the high number of unique items implies material of a highly informative nature. The assumption for the English corpus was that the Hapax figure would be relatively lower partly because vis-à-vis original material in Greek there is a number of authors whereas the tendency is to have a single translator work on a website. Samples point to a similar picture.

Regarding lexical categories-function words, the assumption was that focus would be on adjectives, on the basis that the intended function of the website is the same and the translator would wish to synthesize an equivalent effect.

EL. Εντυπωσιακοί χώροι με ιδιαίτερη αισθητική, τα κλαμπ της Θεσσαλονίκης αναλαμβάνουν τη διασκέδαση σας μέχρι τις πρώτες πρωινές ώρες με ξένες και ελληνικές μουσικές επιλογές, διοργανώνοντας *special* events και φιλοξενώντας συχνά γνωστούς καλλιτέχνες και Dj's.

EN. *Impressive* spaces of *special* aesthetic, the clubs of Thessaloniki take your entertainment off until the *early* morning hours with *foreign* and *Greek* music selections by organizing *special* events and frequently by hosting artists and *famous* Dj's.

On the basis of relevant bibliography (Zanettin 2012: 15), related to lexical density as a feature of simplification,

> "translated texts are hypothesized to have a higher ration of grammatical words and more repetition, and thus a simplified lexicon. Findings consistently pointed to a higher lexical density in non-translated texts, ie. grammatical and most frequent words take up a larger share in the corpus of translation vis-à-vis the corpus of non-translations".

4.2.2. MORPHOLOGICAL FEATURES

As far as use of tenses is concerned, the initial impression that the translator would stick to the present tense of the original is confirmed by the translated material, a fact evident also from the examples used throughout the present section.

Assumptions related to the use of the 2nd person in the English corpus were modified since English does not differentiate between 2nd person singular and 2nd person plural, unlike Greek and many other languages. The question in this case was whether the translator would keep the 2nd person or prefer an alternative solution. As demonstrated by the second example that follows, other solutions were also chosen, such as the use of a gerund or the use of an infinitive:

> EL. Όταν σκέφτεστε θάλασσα και ήλιο, σκέφτεστε την Ελλάδα. [2nd p.p; 2nd p.p].
> EN. When *you* think of sun and sea, *you* think of Greece.
>
> EL. (...) σύγχρονες συλλογές τέχνης που παρουσιάζουν ιστορικά εκθέματα και, βεβαίως, σου προσφέρει τη δυνατότητα να δεις τη μοναδική παλιότερη και σύγχρονη ιστορία της [2nd p.s.; 2nd p.s.].
> EN. amazing landscapes, modern art galleries exhibiting historical artifacts and, of course, *offering* the visitor the opportunity *to see* its unique both older and modern history."

Translation of the imperative mood was examined separately as it is a feature of interest to many other languages cultures, the assumption being that there would be less use of the imperative in the English corpus due to the directness involved and that perhaps would alarm the translator to avoid risking sounding too forward.

> EL. "**Ξεκινήστε** από την αρχαία Αγορά και περιπλανηθείτε σε εποχές μνημειώδεις: αρχαία ελληνική, ελληνιστική, ρωμαϊκή. Το ταξίδι σας στους αιώνες ξεκινά – **γνωρίστε** τις σημαντικότερες στάσεις του (...)"

EN. "*Start* with the Ancient Agora and its monuments from different eras; the Golden Age of Pericles, Hellenistic, Roman. Over everything presides the Sacred Rock of the Acropolis. The cradle of civilisation beckons, so up you go. Your journey through the ages has begun – you now *can get to know* the most significant stops (...)"

Contrary to initial assumption albeit in harmony with the Greek corpus, the initial assumption that there would be increased use of adjectives and adverbs in the comparative and superlative form is not confirmed.

4.2.3. SYNTAX

Sentence data in the English corpus, including total number of sentences as well as the average length of sentences resembles corresponding data from the Greek corpus, an indication that the translator adopted original sentence structure.

The high frequency of coordinate clauses in the Greek tourism corpus seems to occur in the English corpus as well. This emphatic use of language implies that in both the original and the translated material are considered to be or presented as important and worthy of attention, that every single item should make an impression to the perspective tourist. The assumption was that the translator would react in manner similar to that of the original author(s), ie. copying the sentence structure, when translating a touristic website. The material at hand confirms the assumption. It should be noted, however, that emphasis might have been rendered in different ways and translators would have to respond to it.

EL. Ανεβαίνετε προς τη Βόρεια Ελλάδα. Εκεί θα σας υποδεχτεί η «νύμφη» της, η Θεσσαλονίκη, μια πόλη με μακραίωνη ιστορία και μοναδική ζωντάνια. Από τον Λευκό Πύργο ως το Γεντί Κουλέ, και την Αγια-Σοφιά ως την Άνω Πόλη, μνημεία βυζαντινά και μουσουλμανικά, μουσεία σημαντικά σας συστήνουν τη ζωή της πόλης εδώ και αιώνες. Στο λιμάνι, κάθε χρόνο γίνεται το Διεθνές Φεστιβάλ Κινηματογράφου Θεσσαλονίκης.

EN. You head up to Northern Greece *to get* acquainted with 'The Bride of the North,' Thessaloniki; a city with a long history, unique vitality and so many engaging images and things to do; the famous White Tower, Ano Poli, Aristotelous Avenue, shopping on Tsimiski Street, bar hopping on Valaoritou Avenue, the International Film Festival held in the warehouses of the port, live music at Ladadika's eateries, walks through the vaulted arches and the open air markets, visits to museums and Byzantine monuments.

4.3. CONCLUDING THOUGHTS AND REMARKS

The automatic analysis of the tourism corpus and the Wikipedia control corpus has allowed the demonstration of certain main features of Greek tourism texts. As far as the lexicon is concerned, a lower lexical richness and a systematic use

of adjectives have been pointed out. The morphological analysis showed that the present tense, imperative mood, 2nd person play a key role in Greek tourism texts. Regarding syntax, sentence length appears to be higher in the Wikipedia control corpus than in the tourism corpus and a higher number of coordinate prepositions and clauses are used in tourism texts.

It has to be acknowledged that additional fine-grained analysis is necessary in order to validate findings related to the translated corpus. The addition of more language pairs, would be another step toward this direction as well.

It would have also been interesting to be able to check a level further whether it is

"possible to dinstinguish between what is specific to translation and what is more generally a feature of 'intercultural writing'. Should systematic similarities be detected between translation and interlanguage, it is hypothesized that these could be accounted for by the notion of mediation rather than translation universals (Gaspari and Bernardini in Zanettin 2012 28).

The use of corpora and creation of corpora among professional translators seems to be less documented (Zanettin 2000: 32). Reasons include that "corpus consultation remains time consuming, and corpus construction enormously more so" (Aston 2009) and, as a result, "[n]ot all translators appreciate that corpus use may have a medium of long-term effect payoff which can override what they often perceive as short-term disadvantages" (ibid.). We hope that our translation glimpse into corpus analysis, including the selection of checked features itself, has opened up a platform for further discussion of how use of corpora may benefit translators in terms of expectations and readiness and improve the quality of the translation product itself.

References

Aston G. (1999) *Corpus Use and Learning to Translate" in Textus XII (2)*, Amsterdam/ Philadelphia, John Benjamins Publishing Company, pp.289-314.

Aston G. (2001) (ed) *Learning with Corpora*, Houston TX, Athelstan, pp.177-97.

Biber D. (1998) *Corpus Linguistics: Investigating Language Structure and Use*, New York, Cambridge University Press.

Bolasco S. (1999) *Analisi multidimensionale dei dati*, Roma, Carocci.

Bolasco S. (2010) TaLTaC2.10. *Sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi*, Roma, LED.

Dann G. (1996) *The language of tourism. A sociolinguistic Perspective*, Wallingford, CAB INTERNATIONAL.

Lebart L., Salem A. (1994), *Statistique textuelle*, Paris, Dunod.

Nigro M.G. (2006) *Il linguaggio specialistico del turismo. Aspetti storici, teorici e traduttivi*, Roma, ARACNE editrice.

Pierini P. (2009) *Adjectives in tourism english on the web: a corpus-based study*, in *CÍRCULO de Lingüística Aplicada a la Comunicación (clac) 40/2009*, 93-116, Madrid, Universidad Complutense de Madrid

Ondelli S. & Viale M. (2010) *L'assetto dell'italiano delle traduzioni in un corpus giornalistico. Aspetti qualitativi e quantitativi,* in International Journal of Translation, 12, Trieste, EUT pp. 1-62.

Tuzzi A. (2003) *L'analisi del contenuto: introduzione ai metodi e alle tecniche di ricerca*, Roma, Carocci.

Zanettin F. (2000) *Parallel Corpora in Translation Studies*, in *Maeve Olohan (ed.) Intercultural Faultlines*, Manchester. pp 105-118.

Zanettin F. (2012) *Translation-driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies,* Manchester, UK, Kinderhook (NY), USA, St. Jerome Publishing.

Online sources

Greek Tourism Organization official website, http://www.visitgreece.gr/, last consulted: 20/12/2018.

Thessaloniki Tourism Organization official website, https://thessaloniki.travel/en/, last consulted: 20/12/2018.

Corpus extracted from https://el.wikipedia.org, last consulted: 20/12/2018.