

# CORPUS ANALYSIS

Antonella Serra

## CORPUS ANALYSIS

### *ITINEARIES ON LINE: SARDINIA, CAPRI AND CORSICA*

TOTAL NUMBER OF WORD TOKENS	13.260
TOTAL NUMBER OF WORD TYPES	3188

## QUANTITATIVE ANALYSIS

### THE MOST SIGNIFICATIVE ITEMS IN THE LIST:

#### NOUNS

1	[16]	corsica	67
2	[28]	town	46
3	[27]	sea	46
4	[26]	island	46
5	[30]	coast	36
6	[36]	sardinia	32
7	[35]	cala	32
8	[34]	area	32
9	[44]	tourist	25
10	[43]	porto	25
11	[52]	port	22
12	[50]	grotto	22
13	[48]	century	22
14	[57]	north	21
15	[54]	genoa	21
16	[69]	years	19
17	[66]	centre	19
18	[65]	beaches	19
18	[72]	san	18
20	[70]	beach	18
21	[80]	punta	17
22	[79]	part	17

23	[88]	water	16
24	[87]	territory	16
25	[100]	world	15
26	[96]	name	15
27	[93]	costa	15
28	[92]	city	15
29	[91]	church	15
30	[110]	south	14
31	[147]	thanks	11

### **VERBS**

1	[7]	is	214
2	[17]	will	67
3	[20]	are	64
4	[22]	be	54
5	[29]	was	44
6	[37]	has	31
7	[47]	can	22
8	[67]	have	19
9	[75]	find	17
10	[85]	known	16
11	[84]	built	16
12	[108]	situated	14
13	[105]	made	14
14	[112]	were	14
15	[126]	visit	13
16	[133]	like	12
17	[146]	stretch	11
18	[150]	been	10
19	[158]	being	9
20	[189]	due	8
21	[181]	became	8

### **Auxiliaries**

1	[7]	is	214
2	[22]	be	54
3	[67]	have	19
4	[20]	are	64
5	[37]	has	31
6	[29]	was	44
7	[112]	were	14

### **Modals**

1	[47]	can	22
---	------	-----	----

### **Present tenses**

1	[7]	is	214	Simple present
2	[20]	are	64	Simple present

3	[22]	be	54	Simple present
4	[37]	has	31	Simple present
5	[67]	have	19	Simple present
6	[75]	find	17	Simple present
7	[126]	visit	13	Simple present
8	[147]	thanks	11	Simple present
9	[146]	stretch	11	Simple present
10	[133]	like	12	Simple present
11	[158]	being	9	Present continuous

### **Past tenses**

1	[29]	was	44	Simple past	
2	[85]	known	16	Past participle	
3	[112]	were	14	Simple past	
4	[84]	built	16	Simple past	Past participle
5	[108]	situated	14	Simple past	Past participle
6	[105]	made	14	Simple past	Past participle
7	[150]	been	10	Past participle	
8	[189]	due	8	Past participle	
9	[181]	became	8	Simple past	

### **PASSIVE FORMS**

1	is situated in
2	is still to be determined,
3	is to be found
4	is located on
5	is a well known seaside resort
6	is easily reached
7	is found in
8	is made up
9	is due to
10	is positioned on
11	is overlooked
12	is named in
13	is surrounded by
14	is directly connected with
15	is guaranteed by
16	is also renowned for
17	is held every year
18	is the only inhabited island of
19	is also referred to
20	is overlooked by
21	is one of the best preserved
22	is subject to
23	is taken up by
24	is placed on
25	is principally concerned with
26	is dominated by
27	is derived from
28	is known as

29	is followed by
30	is expressed by
31	is undertaken
32	is reached by
33	is divided into
34	is transformed into
35	is managed to
36	is struck in
37	is attached to
38	is occupied by
39	is tortured

The passive form is largely diffused with other auxiliaries too: are, was, were.

## **ADJECTIVES**

1	[12]	its	90
2	[23]	one	51
3	[24]	this	50
4	[25]	that	49
5	[32]	most	34
6	[41]	small	28
7	[51]	more	22
8	[55]	little	21
9	[58]	their	21
10	[63]	natural	20
11	[61]	beautiful	20
12	[68]	other	19
13	[71]	large	18
14	[81]	rich	17
15	[77]	mediterranean	17
16	[76]	first	17
17	[86]	last	16
18	[97]	numerous	15
19	[95]	important	15
20	[94]	his	15
21	[113]	white	14
22	[111]	well	14
23	[117]	genoese	13
24	[115]	famous	13
25	[114]	ancient	13
26	[136]	roman	12
27	[135]	possible	12
28	[134]	long	12
29	[131]	great	12
30	[128]	corsican	12
31	[142]	your	12
32	[138]	these	12
33	[145]	old	11
34	[144]	french	11
35	[151]	extraordinary	10
36	[179]	wild	9
37	[164]	international	9
38	[199]	largest	8
39	[194]	high	8

40	[184]	calcareous	8
41	[183]	blue	8

### **Qualifying adjectives**

1	[32]	most	34
2	[41]	small	28
3	[51]	more	22
4	[55]	little	21
5	[63]	natural	20
6	[61]	beautiful	20
7	[71]	large	18
8	[81]	rich	17
9	[77]	mediterranean	17
10	[97]	numerous	15
11	[95]	important	15
12	[113]	white	14
13	[111]	well	14
14	[117]	genoese	13
15	[115]	famous	13
16	[114]	ancient	13
17	[136]	roman	12
18	[135]	possible	12
19	[134]	long	12
20	[131]	great	12
21	[128]	corsican	12
22	[145]	old	11
23	[144]	french	11
24	[151]	extraordinary	10
25	[179]	wild	9
26	[164]	international	9
27	[199]	largest	8
28	[194]	high	8
29	[184]	calcareous	8
30	[183]	blue	8

### **Possessive adjectives**

1	[12]	its	90
2	[58]	their	21
3	[94]	his	15
4	[142]	your	12

### **Numeral adjectives**

1	[23]	one	51
2	[76]	first	17
3	[86]	last	16

### **Indefinite adjectives**

1	[68]	other	19
---	------	-------	----

### **Demonstrative adjectives**

1	[24]	this	50
2	[25]	that	49
3	[138]	these	12

### **PRONOUNS**

1	[8]	which	142
2	[11]	it	103
3	[38]	we	30
4	[64]	they	20
5	[73]	you	18
6	[90]	who	16
7	[175]	them	9

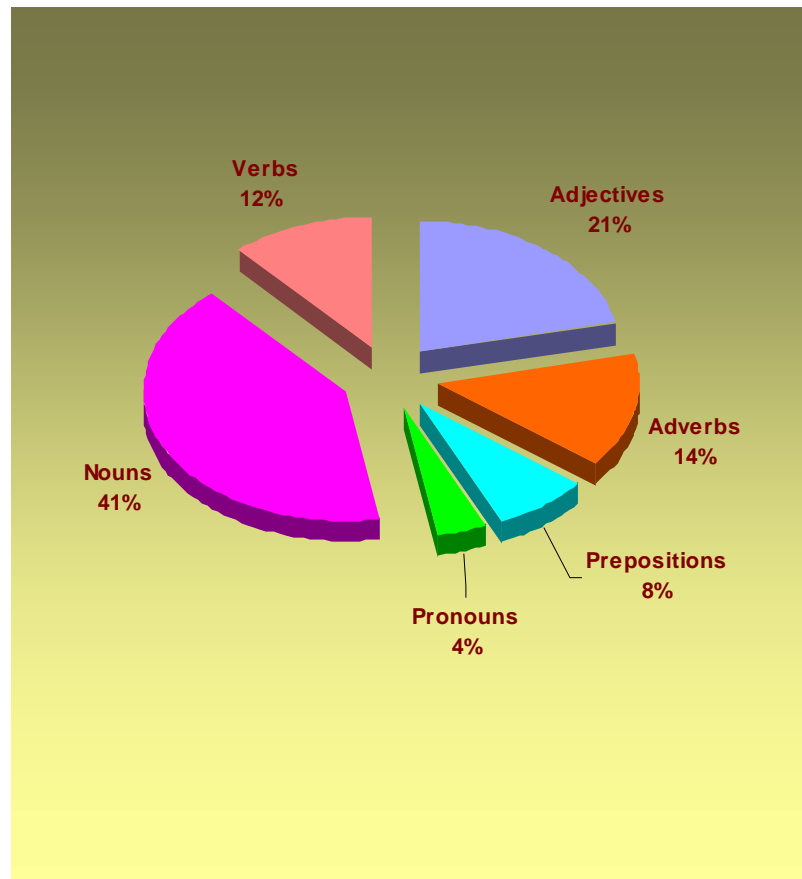
### **PREPOSITIONS**

1	[2]	of	758
2	[4]	in	321
3	[5]	to	270
4	[9]	by	121
5	[10]	with	116
6	[13]	from	86
7	[14]	on	79
8	[15]	as	70
9	[18]	for	66
10	[21]	at	59
11	[59]	up	21
12	[62]	between	20
13	[109]	some	14
14	[104]	into	14
15	[168]	over	9

## **ADVERBS**

1	[31]	there	35
2	[40]	also	28
3	[39]	all	28
4	[45]	not	24
5	[46]	then	23
6	[56]	many	21
7	[53]	but	21
8	[60]	after	20
9	[78]	out	17
10	[74]	back	17
11	[89]	where	16
12	[83]	around	16
13	[82]	along	16
14	[99]	still	15
15	[98]	only	15
16	[103]	during	14
17	[101]	about	14
18	[118]	middle	13
19	[141]	towards	12
20	[130]	few	12
21	[148]	very	11
22	[156]	without	10
23	[155]	so	10
24	[178]	when	9
25	[161]	even	9
26	[193]	here	8
27	[182]	before	8

<b>Nouns</b>	<b>41%</b>
<b>Adjectives</b>	<b>21%</b>
<b>Adverbs</b>	<b>14%</b>
<b>Verbs</b>	<b>12%</b>
<b>Prepositions</b>	<b>8%</b>
<b>Pronouns</b>	<b>4%</b>





## MINIMAL CORE VOCABULARY

NOUNS	ADJECTIVES	VERBS	ADVERB	PREPOSITIONS
architecture	ancient	are	about	As
area	beautiful	be	after	At
areas	blue	became	all	Between
arzachena	calcareous	been	along	By
beach	corsican	being	also	For
beaches	extraordinary	built	around	From
NOUNS	ADJECTIVES	VERBS	ADVERB	PREPOSITIONS
cala	famous	can	back	In
castle	first	due	before	Into
caves	french	find	but	Of
centre	genoese	has	during	On
centuries	great	have	even	Over
century	high	is	few	Some
cervo	his	known	here	To
church	important	like	many	Up
city	international	made	middle	With
coast	its	situated	not	
coastline	large	stretch	only	
colour	largest	thanks	out	
corsica	last	visit	so	
costa	little	was	still	
east	long	were	then	
emerald	mediterranean	will	there	
end	more		towards	
festival	most		very	
france	natural		when	
gallura	numerous		where	
genoa	old		without	
granite	one			
grotto	other			
gulf	possible			
harbour	rich			
history	roman			
houses	small			
inhabitants	that			
interest	their			
island	these			
islands	this			
italy	well			
land	white			
life	wild			
maddalena	your			
maquis				
marina				
name				
nature				
north				
olbia				

paoli
part
pascal
people
place
places
point
port
porto
punta
road
<b>NOUNS</b>
rock
rocks
san
sand
sardinia
sea
smeralda
south
territory
time
today
tourism
tourist
tower
town
traditions
village
war
water
world
years

## **QUALITATIVE ANALYSIS**

This kind of analysis is very involving and useful to show how tourism language is different from everyday language.

In order to discover something in this direction, I have chosen some meaning words and I have analyzed them and how they are used.

### **MEANINGFUL WORDS AND HOW THEY ARE USED**

#### **TOURIST**

The word TOURIST appears in the corpus 25 times and it is used 22 times as an adjectives. For ex: tourist port, tourist attraction, tourist point of view, tourist complex, tourist centre, t. airport and so on

Only 3 times as a noun:

- authentic rarities which attracts tourist...
- today the tourist may say...
- essential service to satisfy the tourist..

#### **CENTRE**

The word CENTRE appears in the corpus 19 times and it is used in two ways. Only one time it is used as an adjective : the old town centre houses.

The other occurrences are nouns.

#### **THANKS**

The noun Thanks occurs 11 time and its use is not very common in every day language. On the contrary, in Tourism field there is a very large use of it expressing the cause-effect relationship.

So, instead of using Because of or since, the writer express themselves with "thanks" to underline the exciting effects something or somebody has on the places. It is a trick to call tourists' attention and to switch on their curiosity.

#### **VISIT**

Visit can be used as a verb and as a noun.

While in normal lexicon it is much more used as a verb "to visit" persons and often it is substituted with the verb "to see", in ESP English Tourist is almost a technical word used as a verb and as a noun to indicate a visit in a specific place.

In the corpus it is often found with the expression: come to visit, return to visit, willing to visit, exciting place to visit, have a long visit, pay a visit, throughout your visit.

As we can see the word is used to persuade people to go to see a place.

## STRETCH

Stretch can be used as a verb and as a noun.

In this field it is used only as a noun followed by the preposition of.

I think in common English it is largely used as a verb. In tourism field is very much used as a noun and in this corpus, the writer wants to underline the vast extensions of coastlines, white sandy beaches and so on.

**In my opinion, in this corpus there is not the use of technical words, as I didn't find terms which belong strictly to Tourism Language.**

**These 5 words I have analyzed in fact, don't occur so often (and some of them are not in the more used first 200 words) but I picked up them because I think they are used in different ways if compared in every day English.**

## TYPICAL LINGUISTIC PATTERNS

### IS SITUATED

I think that "is situated" is not so common in every day language as it is often replaced by "you can find it in...", "it is placed in" etc.

So, in Tourist language is a typical verb used to describe the location of places.

It occurs 14 times and with these various forms:

- ✓ is situated in..
- ✓ is situated at
- ✓ is situated on
- ✓ is situated along

### IS KNOWN (THAT / AS)

The verb to know in every day English is used to express the functions of "having information", be aware of something etc. In this field it express "reputation" in the sense that it is used to underline the notoriety of the places mentioned and in the meanwhile describe them.

The verb appears 16 times in the passive form with expressions such as: "is Known as ...", "it is known that" followed by the name of the locality .

It is very diffused with these forms too: "is well a known.." or "known as one of the best places..."; contribute to make known throughout the world...", better known as", "Known by" etc.

## NOMINALIZATIONS

This corpus is full of nominalizations. This is a process by which we transform verbs, adjectives and adverbs into names to abbreviate the discourse so, for concision.

***These are the nominalizations ending with "ness":***

"The walkway meanders in the coolness which emanates from the sea ....

" this time of darkness which were the first century...."

"must then face Genoese covetousness..."

***But much more diffused are nominalizations in "ing":***

thriving, relaxing, striking, the landing; setting; the fishing, the coming; surrounding; foodlighting; etc

***As for nominalizations in "ation":***

Location, accommodation, stagnation, civilization, relaxation, concentration, inhalation, etc.

***As for nominalizations in "ment":***

Sediment, settlement, refreshment, entertainment, tournament, retirement.

## **COMPOUND NAMES**

In my corpus there is a large use of Compound names usually formed by an adjective and a name. The name is qualified by the previous word.

Examples:

Surrounding countryside; beautiful sandy beaches; gothic-catalanian style, tranquil landing place, old town centre, fishing centre, seaside centre, administrative centre, very striking kind of scenes, Mediterranean cultures Festivals, Jazz festivals..., richly smelling underbrush, ever changing colors, multi-millennial history etc.

I think that in every day English they are not so common. In this specific field they are devices through which the writers want to catch visitors' attention striking them with short but involving clues with exciting descriptions.

## **PASSIVE FORMS**

The corpus is built on passive forms. They are very much spread. We find them in the third person singular of Simple present "is" but they are frequent with the other forms of auxiliaries too, such as: are, was, were.

I think that in common language the use of this forms is not very common, so perhaps I can conclude that this is one of the most important features of this corpus and in this branch of Tourism English (Describing renowned localities).

Why is the use of passive forms a very important feature of the corpus?

To underline the attention of the texts on the places and the effects time and human hand have had on them.

## **COLLOCATION OF CAUSE-EFFECT ITEMS:**

**Since, because of, as a result of, therefore.**

Since 1778, since april ..

As we can see, since is used to specify dates not to underline an effect.

So called because of the walls

Because of its small...

So named because of the dark....

Because of the dominant influence of Pisa...

Because is used to identify cause and effect relationship and it is the most used term to underline this aspect.

The town and the province are practically uncontaminated therefore maintaining rich...

The form therefore is used only one time.

## **CONCLUSIONS:**

What I have noticed analyzing the corpus has been already said but in this section, I'd want to add something more and in the meanwhile summarize:

In my corpus there are a wide use of qualifier adjectives as through these text the authors want to describe the beauty of the places on which they focus on. There are a lot of superlatives as the places are described as one of the most famous localities all over the world.

The most common verbal tense is the Simple Present by which the writers describe the characteristics of places. Simple Past is very used to describe architectural characteristics and the period they belong to.

As I have already said there are many Passive forms to describes the influence the events and the human hand had on the environments. The verbs are all in the third person (singular and plural) by which the writers show objectivity keeping distance from what written.

There is a large use of prepositions associated with the passive forms.

There are a large use of adverbs (of time, of place etc.) because writers go forwards and backwards through time and places to describe the changes the territories have taken or the movies tourists have to do to reach the localities described.

There is an absence of forms such as can, may and should (modals) except for some sentences but, only normal verbs to express certainty and sureness in describing the beauty of the localities. So, the tone of the speech is never in terms of uncertainty and possibility.

There is an absence of the use of Cause – Effect items as this function is expressed with some verb in the passive forms such as: is due to... is determined by, is directly connected with, is derived from...

As I have already said “thanks to ..” it is a common word in my corpus as it is the noun that is mostly used to express the cause-effect function.

I think that in this corpus there are not Technical words to reflect on but only general lexicon used in a specific field (context). Words are those of normal English only used with a different meaning.

Being a corpus based on Tourism, the majority of the words are nouns which are very often preceded by adjectives as the writers want to underline the characteristics of the cities, beaches, ports etc. I think it is only the way by which language is used that differentiate this branch of ESP English Tourism.

The language in the corpus doesn't aim to be concise but it is an emphatic language often with long sentences full of superlatives. In fact, my corpus is a combination of three Tourist Itineraries on line, by which writers try to convince people to visit these places. So, the genre of the corpus is a promotional one where the writers show a kind of involvement in their words!

## **INFLUENCES OF THIS INVESTIGATION ON MY SYLLABUS.**

Only after this short investigation I have realized and reflected how can change English if studied in a different field rather than in everyday life. So, in my Syllabus on Tourism ESP I'll take into consideration the results I have catch in this analysis :

- first of all I'll take into consideration this Corpus, using it as an important tool in the third Module of the Syllabus in order to let know students the kind of language to be used .
- The analysis reveals a large use of passive forms, a lot of adjectives, prepositions and adverbs. It is why in my syllabus I am going to consolidate these grammatical features or enlarge vocabulary.
- The corpus is a valid example how Tourist Itineraries are structured, so it will be an important tool to show students how to build them, if they are asked to, in their future jobs.

Antonella Serra