The translation of tourism-related websites and localization: problems and perspectives.

Gloria Cappelli

Università di Pisa

L'articolo affronta alcune questioni poste dalla traduzione di siti internet, con particolare riferimento alla traduzione del contenuto testuale di siti turistici. Molti sono i problemi aperti in questo settore che ha iniziato a suscitare l'attenzione degli studiosi solo di recente e che spaziano dalla natura dei generi testuali del web fino alla distinzione fra traduzione e localizzazione, dal ruolo del traduttore di contenuti ipermediali al problema di valutare la qualità del prodotto finale. La complessa problematica è esemplificata attraverso lo studio delle parole chiave in un corpus di testi di argomento turistico che fornisce la base su cui fondare alcune proposte per la ricerca futura.

1. Introduction

Over the years, Translation Studies have evolved with the evolution of media. The rich scientific debate derived from the obstacles, with which practitioners and scholars are faced on a daily basis, has led to remarkable insights and positive advancements both in translation theory and practice. The technological advances in the computer sciences of the past few years, and the increasing circulation of multimedia and hypermedia have provided Translation Studies research with new thorny questions and new "knots to untie". Much attention has been devoted to the translation of audiovisual texts both in terms of dubbing and subtitling (Gambier 1994; Heiss and Bollettieri Bosinelli 1996; Pavesi 2005; Bruti 2007) and to the translation of CD-ROMs and other multimedia materials (Agorni 2000; Gambier and Gottlieb 2001). Computer-assisted translation is also widely discussed (Quah 2006), especially by reason of the growth of the Internet and the consequent, often-mentioned globalization, which makes geographical and cultural boundaries

thinner and thinner. In such a world, Bassnett (2002:10) claims, the significance of translators' work as well as the debate about crucial issues are destined to grow steadily. Despite the author's insightful remark, however, and despite the fact that the web has indeed expanded and that the number of non-English-speaking users is steadily growing, web content translation has not as yet received a systematic treatment.

Translators for the web cannot rely on the amount of theoretical support on which translators of literary or technical texts can rely. Nor does there seem to be any clear criteria to assess the quality of web translations, and yet, in a world where some businesses depend predominantly upon their presence on the World Wide Web, a good translation of the content of their websites is often responsible for their successes and failures. Tourism-related businesses certainly fall into this category of industries, but nevertheless many websites are badly translated.

The following sections tackle some questions relative to web translation, in particular to the translation of the textual content of tourist websites. The purpose of this article is not to provide clear-cut answers but to raise questions and to encourage the discussion on this topic.

2. From web promotion to linguistic research

It is widely acknowledged that the evolution of the Internet has brought about some significant changes in the way in which people interact and carry out their business. In most fields, and especially in the tourism industry, where the main goal is to "encourage" hundreds of thousands of people to travel throughout the globe, a good or a bad website can make the difference. Even more so, if we consider that, according to a research carried out by Global Market Insite, Inc. in 18 countries, in 2005, though, up to 70% of the informants affirmed that they booked their holidays online and these figures are destined to grow. Another research project carried out in Germany on several banks' websites (Stein 2006) has revealed that the linguistic quality of the website is perceived iconically in line with the supposed quality of the business or service offered.

For those interested in language and translation, these facts represent the tip of an iceberg, as they urge linguists to tackle some fundamental questions relative to the language used on websites; to the status of online texts, to the nature and quality of the translation for a particular medium – the web – and for a specialized domain – tourism – and, last but not least, to the role and training of translators.

3. How do we know when a translation is good?

House (2001) opens with this question, and continues writing that (p.243) it "[…] lies at the heart of all concerns with translation criticism". It is certain that objectifying the quality of translation is not easy. House (1997; 1998; 2001) distinguishes three main approaches to meaning, which in turn determine three different approaches to translation evaluation: mentalist approaches, presupposing the view of meaning as residing in the user's head; response-based approaches, that is, behaviouristic and functionalist approaches in which the most relevant criteria are either the reader's reaction to the translated text (e.g. Nida's "dynamic equivalence principle") or the respect of the function of the source text (e.g. Reiss and Vermeer's *Skopostheorie*); and text and discourse-based approaches such as Toury's Descriptive Translation Studies and other linguistically-oriented approaches, such as House's proposal itself, in which translation is, in the end, text re-contextualisation.

House's model adopts the concept of equivalence and presupposes the "analysis and comparison of an original text and its translation text on three different levels: Language/Text, Register and Genre, facilitating correlation of internal, linguistic textual features and external, contextual features" (2007). The translated text (TT) must have a function equivalent to that of the source text (ST) and this requires a "cultural filter" (House 2001), that is, a sort of "adaptation" of the translated text to make up for socio-cultural, stylistic and rhetorical differences among the ST and the TT.

From the studies in translation quality several suggestions about the relevant aspect for the analysis of (web) translation emerge: the *genre* and the *function* of our texts, the *subject* and the *linguistic features* of the text itself, the *channel* and the nature and relations of the *participants*. Far from being an easy task, any attempt to apply the model to tourism-related web content immediately unveils further complex questions that await an answer.

4. Web content: new genres or just a new medium for the 'old' genres?

For the purpose of the present discussion, I endorse the definition of genre put forward by Taylor Torsello (1999), which sees it as a "culture-dependent abstraction", a generalization made by people who belong to the same culture on the basis of their experience of texts and contexts. For a long time, web pages have been considered as sorts of "online brochures" or the electronic version of traditional printed genres (Cappelli 2006). The lively debate of the past few years though, has contributed to bring to the fore several distinctive structural and linguistic features of the textual contents published online (Biber 2004; Santini 2006; Stein 2006) which make it possible to consider websites as the *locus* of new genres and not just as "a new way to present old information".

4.1 The structure and textuality of web pages

If genres are "socially ratified text-types in a community" (Kress and Threadgold 1988, quoted in Taylor Torsello 1999), which come with a conventional structure, easily recognizable by the members of that culture, web pages can definitely be considered as a macro-genre. Despite the structural differences between the different sub-genres of web documents (e.g. blogs, portals, company websites, discussion forums, etc.) and the differences derived from the creative work of web developers and web designers, all websites and their pages have common and recognizable features (e.g. title bar, navigation, main content area, etc.) which are now conventionalized[1] and easily recognized by web users.

Websites are mutlidimensional, and language plays a fundamental role at all levels of their structure. Text is not only present in the main body of the page but also in the navigation systems and in "source code", which lies beneath the surface. Some of the textual elements included in the latter are not perceivable at the surface level, but are indeed visible to search engines (e.g. *Google*).

Web documents defy any traditional view of the text. Much has been said about hypertextuality and its characteristics. The "linearity" of text in its literal sense dissolves, since the reader chooses his path and builds his "subjective linearity" *online*. *Entry* and *exit pages*, solely determined by the user's choices, replace beginning and end. Coherence relations become the fundamentals in which the whole hypertext is grounded; links are created following conceptual connections relating different texts and act as cohesive devices which keep web discourse together.

Web textuality is dynamic and interactive. Textual content can be automatically adjusted to the preferences of the readers (e.g. multilingual websites can present the content in a preferred language) and continually modified by multiple authors so that it can be considered as a collective product to which different professionals contribute by reason of its multidimensionality.

The technological aspects contribute to the complexity of web textuality: since language plays a fundamental role in the source code as well, this latter includes "linguistic hints" about the content of the web pages in the form of keywords, key phrases, and full sentences or short texts. These elements contribute to determine the website ranking in the results of search engines and consequently the visibility of the information presented.

For this reason, each web page should be considered as a harmonious whole, since it is "a sort of container from where the reader picks up information s/he needs", and therefore "artificially separating what is considered to be the main body from the rest is an arbitrary operation and it would not make sense in many cases […]. In sum, web pages tend to be more complex and more mixed than traditional paper or electronic documents" (Santini 2006:67-68).

4.2 The language of web promotion in tourist websites

Taylor Torsello (1999) points out how the different roles of the participants, the communicative goals and the medium cannot but result in significant differences in the way in which texts are built. Given the importance of quality content and the technical requirements for an effective functioning of websites, some of the features of "promotional English" become more pronounced, as is evident in the choice and distribution of keywords (Cappelli 2006). The limited space allowed to advertisements by some portals, for instance, results in increasing lexical and expressive conciseness and in the widespread use of premodification as a form of relative clause reduction (Gotti 2006).

Several verbal techniques typical of promotional tourism discourse in order to "persuade, lure, woo, and seduce millions of human beings" (Dann 1996:2) are commonly found, such as an abundant use of adjectives and of emphatic language (*language euphoria*); the frequent use of the imperative mood and of the formulae of direct address to the reader (*ego-targeting*); common collocations meant to satisfy the personal and cultural expectations of potential customers and to describe an attraction by resorting to certain sociolinguistic perspectives (e.g. authenticity, strangerhood, etc.) and topics (*key words*).

The importance of keywords in the language of the web is one of the most interesting phenomena for linguists and translators alike and the lexical level of online content provides a good vantage point for some observations on the nature and quality of web translation.

5. How is this of interest for linguists and translators?

If, as House (2001) says, when we translate we need to consider the *genre* and the *function*, the *subject* and the *linguistic features* of the text, the *channel* and the nature and relations of the *participants*, web translation gives us food for thought.

Websites can be described in terms of genres, given (at least) the structural, textual and linguistic peculiarities described above. The subject can vary considerably, of course. As far as tourism-related websites are concerned, this point raises further questions. First, is there such a thing as the language of tourism? Dann (1996) seems to prove that indeed there is, and that it has specific features and modes. Second, is there only one type of language of tourism? Everyday experience tells us that the language of tourism is a multifaceted, multidisciplinary type of discourse which draws on several specific domains such as art, architecture, eco-talk, etc. (Fodde and Denti 2005, Nigro 2006; Vestito 2005), yet another aspect that adds complexity to the translation process. Third, can tourism English be considered as a type of specialized discourse such as medical English or computer English? Recent studies (Gotti 2006; Nigro 2006; Cappelli 2006) seem to agree that tourism English can be recognized as a type of specialized discourse, displaying different levels of specialization associated with different types of texts that address a more or less specialized audience.

As for the linguistic features of web content, there is still much work to be done in order to analyze all the aspects of interest for the linguist. Certainly the structural constraints of web genres bring about interesting phenomena at the linguistic level. Some of them have been touched upon in section 4.2, a few more will be considered in the next section, where the question of keywords will be dealt with. Translators need of course to be familiar with the most relevant linguistic features of websites and web pages, not only at the discursive and pragmatic level, but also at the semantic, syntactic and morphological level. As far as the channel is concerned, the web is in a continuous evolution and challenges every clear-cut distinction between the oral and written system. Much research is still needed on the influence of virtual media on textuality.

As for the "participants", questions arise relative to the nature of the source text (e.g. was it meant to be translated?) and of its author (e.g. just one author or many?). The receivers are "global readers": they come from all over the world, they have very diverse cultures, interests, values and, for most of them,

English is not an L1. Moreover, visitors arrive at a website by searching for specific keywords or key phrases, building their own unique path through the web and, as potential tourists, they come with dreams that need to be "wrapped up and sold back to them", which is what promotion does. Translators also need to take into account this complex interplay of attitudes (Bertuccelli Papi 2000) and language idiosyncrasies when they sit down to translate for the web.

6. Is web translation a type of specialized translation? Focus on the lexicon.

As a direct consequence of the multidimensionality of the documents to be translated, website content translation is a type of specialized translation that requires special skills and competences at various levels. On the linguistic level, translators have to deal both with the technicalities (in the broadest possible sense) of the language of tourism and with the strategies required by the promotional aim of the text. On the technical level, translators must keep in mind the requirements imposed by the need to optimize web content for search engine rankings.

6.1 The fundamental role of keywords in promotional tourist websites

Keywords and key phrases are essential in search engine ranking, and, as a consequence, on occasion, the text is almost exclusively made up of them, to the extent that the actual message is repetitive and inadequately informative. Lists of synonyms are provided, with slight variations to ensure the presence on the page of all the variants of the same search phrase for a given item. Thus, web content has generally a higher concentration of keywords and key phrases than printed promotional material:

> Italian villa rentals, hotels and vacation accommodation, self catering homes, holiday rentals, apartments, farmhouse holidays, bed and breakfasts - they're all here, with maps, images and local information. Direct contact - we're not an agency - a wealth of detail, and an enormous choice. Whether you want to rent an Italian villa, apartment, castle or holiday rental cottage, or you're looking for a hotel or guest house, start at knowital.com.

Websites which target English-speaking countries will probably include – as in this case – key phrases which are identical in meaning, but belong to different varieties of English, such as, *vacation house* (Am.) vs. *holiday home* (Br)[2]. The usage of keywords and keying are common techniques of the language of tourism (Dann 1996). Keywords are meant to fire imagination while mirroring the consumer, and therefore they must be part of the consumer's language, so as to allow the potential tourists to recognize themselves. In order to be effective they must be short, clear and current. This applies even more so when the language has to fulfill the web medium requirements (i.e. search engine optimization), that is to say when they must actually be *the very same ones* tourists would use to describe what is being sought after.

Keywords are generally concentrated at the beginning of the page, as the first paragraph of a web page content is valued more than the rest of the text. Several factors concur to this. First of all, the English-speaking world writing-style is traditionally grounded in the so-called "five paragraph essay", and the first one is the most important, in that it has to grab the reader's attention by informing him or her of the main points of the essay. Since search engines are biased towards the English language, as the vast majority of online texts are in English, they also assume that people who write in English will follow this style, and therefore, that the first few sentences will be the most relevant. Second, Website users tend to scan, not read (Morkes and Nielsen 1997), and therefore, more often than not, the only passage that they read is the beginning of a page.

Keywords and key phrases are also fundamental in titles and as text in active links. Because it would not be possible to have overly long textual links or titles, the keywords chosen must be at the same time informative and short. These requirements prompt some very interesting effects at the linguistic level (Gotti 2006), such as an increase in conciseness. Thus, at the morphological level, it is common to find *blending* (e.g. *campsite* vs. *camping site*); acronyms and abbreviations; reductions (user id vs. *user identity*; *psw* vs. *password*). At the

syntactic level, relative clauses are avoided in favour of lighter constructions such as lexemes obtained by means of affixation with affixes endowed with a clear semantic value (*self-catering accommodation* vs. *accommodation where you can cook your own meals*), simplified passive relative clauses (*pre-paid room* vs. *a room which has been paid in advance*), or nouns specified by a complex expression (*a three-night getaway* vs. *a getaway that lasts three nights*). These are quite easy solutions in English, since it allows right-to-left constructions so that premodification can easily be used to create very dense noun phrases to use as key words (e.g. *tour operator*, *hotel room*, *car rental*, *arrivals gate)*. These processes obey the criteria of conciseness and transparency, in that the resulting expressions are compact and definitely shorter than the original ones and this is a feature that puts web English and tourism English in general on a par with other forms of specialized discourse (Gotti 2006).

7. Translation, localization, copywriting or "all-in-one"?

The previous sections should have made clear that translating for the web is indeed a difficult task and that many different skills are required. The technological and linguistic nature of the web and of its new genres forces linguists and translators to reconsider some central concepts of Translation Studies, such as that of source and target text, equivalence, acceptability, and even the concept of translation itself.

As we have said, the source text of a promotional website is strongly reader-oriented, in the sense that it must include what readers are most likely to expect. The author, with few rare exceptions, is not known, and many people have modified, expanded or optimized the source text. From the linguistic point of view, web content is often written to be translated, and in the most specialized cases, the language used is simplified in order to make the job easier for the translator: this process is known as *internationalization*. The linguistic analysis of the source text advocated by House (1997) might not be as relevant as in the translation of other types of text, whereas a thorough knowledge of the target

system, of the principles of search engine optimization and of the technical functioning of websites seems to be very important.

We might conclude that the equivalence that the translator has to achieve in web translation is close to the concept of equivalence found in response-based approaches such as the Skopostheorie. A good web translation has a precise final *skopos*: it must render the core message and, at the same time, make a website optimal for its intended audience, and therefore for search engines. Receivers' expectations become the most relevant factor in the translation process. Translating becomes then "moving texts" (Pym 2004b) from one cultural context to the other and the translator assumes a strong decisional power.

All this, however, seems to require more than the ability to translate in a professional way, and this consideration is at the heart of a lively, but unfortunately still marginal discussion, as yet more of interest for translation training than for translation studies. Can we speak of translation, or is it rather adaptation? Does the translator have to be also an expert in copywriting and localization, or are these three different professional domains that must be brought together towards a common goal?

7.1 Localization and translation

Localization is that process by which a product (in the broadest possible sense) is made linguistically and culturally appropriate to the target "locale", terms which indicate "both a language variety and a set of cultural preferences" (Pym 2001) of the target market where the product must be promoted or sold. Localization is a word that has its origin in computer technology where it referred – and it still largely does – to the adaptation of software (e.g. Microsoft Word menu) or the translation of help files for technological products of various sort. Whereas in the early 80s, localization was a completely separate step from the production of the original document, which was handed over to localizers later, today, the source documents are created with localization in mind.

Localization may take place at different levels and in different degrees, ranging from mere "enabling", to real adaptation (Brooks 2000). The relationship between translation and localization, and the problem of whether one encompasses the other or not are still debated and there is no agreement on the matter. Pym (2001) seems to think that translation is a part of lexicalization and laments that, if Translation Studies refuse to include localization in the theoretical debate, the succeeding generations of translators will be doomed to remain "'just translators', while the big money and more interesting work flows the way of marketing specialists and information engineers".

Because internationalization and localization affect the concept of "source text", the concept of equivalence is also affected. Pym (2004a; 2004b) claims that internationalization produces a new kind of equivalence, since "when we no longer have a source […] equivalents are artificially determined in the place of internationalization". According to this view, translation would be relegated to the mere transposition of few strings of text, whereas this new, market-driven equivalence would be left to the people who adapt the text: "the engineers, marketing experts, publishers, etc.".

7.2 A lost cause or a new challenge for translators?

If we look at web translation as one type of specialized translation, which requires skills at both the linguistic and technical level, there is no reason why we cannot conclude that translators can bridge the gap between technical know-how and language skills. In all kinds of specialized translation (Scarpa 2002) translators and "experts" need to cooperate, but translators who work in specialized sectors have developed tools (e.g. glossaries, terminological databases, etc.) that allow them to overcome the major problems.

As far as the translation of promotional tourist websites is concerned, where localization is obviously essential, translators should know that there are technical tools that can help them find the best linguistic solutions, such as keyword generators, keyword popularity calculators and website statistics

analyzers. Once translators know "where" language matters (e.g. active links, meta-descriptions in the source code, title bars, etc.), they can treat web translation just as another type of specialized translation. Their linguistic background too can provide them with some useful tools, such as the use of corpora to evaluate translation equivalents (Wilkinson 2005; Nigro 2006).

In the next few sections, whenever the term "translation" is used, it subsumes both translation and localization, as I take them to be essentially two inseparable processes as far as promotional tourist websites are concerned.

8. A case study: promoting Siena and surroundings

Guidère (2005), writing about the translation of advertisements, speaks of "the intrinsic economic value of a translation", in that it enables a service provider to stand out in a highly competitive market. There remains the question as to how it can be ascertained that a translation ensures the best communication possible, and the author concludes that as far as advertising is concerned, the quality cannot but be evaluated according to the results. That visibility on the web depends on factors other than translation has been pointed out; however, a web translation must provide optimized content. Computer science – thanks to keyword popularity analysers – and (corpus) linguistics can help the translator, as is evidenced by the following case study.

8.1 The corpus

The contributions of the methods of corpus linguistics to translation studies and specialised discourse are widely acknowledged (Castello 2002; Ulrych 2002; Bowker and Pearson 2002; Nigro 2006; Partington, Morley and Haarman2004; Tognini Bonelli 2004; Laviosa 2006). For the present study two small corpora were collected: an Italian-English parallel corpus and a comparable English corpus.

The parallel corpus includes the textual content of several informational-promotional websites promoting Siena and surroundings originally created in Italian (SIENA_PAR_IT) and then translated into English (SIENA_PAR_EN) by native speakers. The comparable corpus includes the textual content of several informational-promotional websites promoting Siena and surroundings originally created in English for an English-speaking audience (SIENA_EN). Table 1 shows some statistics obtained with WordSmith Tools (Scott 1998): tokens are the actual number of words, thus each of the two corpora in English contains over 100.000 words.

Corpus name	SIENA_PAR_IT	SIENA_PAR_EN	SIENA_EN
Tokens	105605	100978	108766
Types	13832	10428	11267
Ave word length	5,2	4,87	4,58
Sentences	2719	2746	4151
Sentence length	28,46	28,06	18,41

Table 1.

8.2 Keywords analysis

These corpora provide the linguistic material for the study of the usage of keywords in translated web content and in original English web content. The keywords in each corpus can be individuated with WordSmith Keywords Tool. The program identifies 500 keywords in each corpus, that is, all those words "whose frequency is unusually high in comparison with some norm" (Scott 1998:71). In order to identify them, it was first necessary to compile frequency lists for both SIENA_PAR_EN and SIENA_EN[3], i.e. a list of the most frequently occurring words in the corpora. The two frequency lists were compared to the frequency list compiled from the written part of the British National Corpus, which was used as the reference corpus, i.e. a representative and non-specialized corpus of the language under scrutiny (Bowker et al. 2002).

From the original 500 keywords, some elements were taken out, such as proper names of restaurants and streets, abbreviations such as "tel.", "www", etc.

and successively the two lists were compared. Here is a list of the first 50 items of SIENA_PAR_EN, ordered according to their keyness and frequency:

LIST 1

1. Sienese	18. productive	35. campo
2. production	19. mediaeval	36. fountain
3. contrada	20. tradition	37. province
4. constancy	21. itinerary	38. flavour
5. wine	22. docg	39. medieval
6. Tuscan	23. contradas	40. woods
7. century	24. vernaccia	41. route
8. palazzo	25. senesi	42. periods
9. persistence	26. tools	43. area
10. palio	27. façade	44. itineraries
11. wines	28. olive	45. pici
12. Etruscan	29. brunello	46. ancient
13. doc	30. grapes	47. dop
14. cinta	31. observations	48. fermentation
15. pecorino	32. fonte	49. truffle
16. museum	33. sangiovese	50. processes
17. classico	34. regulations	

The following is a list of the first 50 items of SIENA_EN, ordered according to their keyness and frequency:

LIST 2

1. Siena	18. medieval	35. tourist
2. Tuscany	19. terzo	36. terme
3. palio	20. cypress	37. etruscan
4. piazza	21. road	38. olive
5. chianti	22. wines	39. horses
6. tuscan	23. towns	40. classico
7. contrada	24. pubblico	41. restaurant
8. campo	25. fortezza	42. walls
9. wine	26. tour	43. restaurants
10. Sienese	27. horse	44. springs
11. contrade	28. countryside	45. sangiovese
12. town	29. center	46. mangia
13. duomo	30. city	47. walk
14. brunello	31. vineyards	48. hill
15. Italian	32. race	49. beautiful
16. palazzo	33. hills	50. senese

17. frescoes   34. favorite

At first sight these are just lists, but if we compare the two lists some important differences emerge. The first consideration is that keywords immediately give us an idea of the preferred sociolinguistic perspective used to present and promote Tuscany: the perspective used is authenticity (Dann 1996), and particular importance is given to the culinary tradition and to art. If we classify the 500 keywords in thematic groups, eight main topics emerge:

1. History: e.g. *medieval, Etruscan…*
2. Buildings and infrastructures: e.g. *palazzo, duomo, tower…*
3. Siena: e.g. *Campo, Banchi, sienese, palio…*
4. Territory: e.g. *natural, wooden, hills, vineyards…*
5. Food and drink: e.g. *doc, red, wine…*
6. Activities: e.g. *thermal, walk, tasting…*
7. Art: e.g. *gothic, frescoed, archaeological…*
8. Services: e.g. *restaurant, enoteca, villa…*

Even though at first sight the two corpora, SIENA_PAR_EN, i.e. translated English, and SIENA_EN, i.e. native English, seem to be very similar, at a deeper look, some important differences emerge. First of all, even just by reading List 1 and List 2 it is evident that the most frequent elements in List 1 are adjectives and technical denominations (*doc, docg, dop, igp*), whereas reading List 2 we have almost a visual image of what Siena and surroundings are like: the keywords list is mostly made up of nouns, several of which are Italian. Scanning the full list one notices immediately that this is not just because in the comparison of frequency lists foreign words stick out, but because the English corpus is full of words that have been left in Italian, a sign of a significant use of the *languaging technique* (Dann 1996; Cappelli 2006), i.e. the use of foreign words to give an authentic feel.

Comparing themes and keywords for the two corpora, SIENA_PAR_EN seems to privilege food and drink above all, but also history and art. SIENA_EN,

on the other hand is more balanced but we find only few references to food and drink (*wine, olive, Chianti Classico, Brunello*) and history (*medieval*) in favour of lexical items describing the city, the territory, and the activities. Once again, this first impression is confirmed by the scrutiny of the full list of keywords. Even though both corpora include keywords from all these topics, there are some interesting differences.

1. **Difference in keywords**. Certain keywords are present in one list but not in the other. This information becomes relevant if some of the top keywords of one list are missing from the other. Thus, the second most frequent keyword in SIENA_PAR_EN, *production*, is not listed among the keywords of SIENA_EN nor is *tradition*. Surprisingly the translated corpus does not have *Siena* or *Tuscany,* or *cypress* as keywords.

2. **Difference in distribution**. Certain keywords which are very frequent in one corpus are less frequent in the other, e.g. *museum* occurs 128 times in SIENA_PAR_EN but only 52 times in SIENA_EN and the reverse happens with *countryside* which occurs 84 times in the second corpus and only 49 in the first.

These considerations assume a remarkable value if we keep in mind that keywords are essential in order to rank highly in search engine results as described in detail in section 6.1. SIENA_EN should give us an idea of the expectations (both at the linguistic and cultural level) of the target readers. Even keeping in mind the size limits of the corpus, we can however consider it as a representative sample of the way in which native speakers of English (Am, Br, Au) think about Siena and surroundings. Presumably, they will use some of those common keywords and key phrases to search the web for Siena-related websites. If the keywords which they would naturally produce are not present in the website, then, the website is not optimized[4] and we must conclude that the translation would not help the website fulfill its function.

This assessment does not refer to the quality of the translation *per se*, as it does not take into consideration matters of style and intercultural rhetoric, but focuses exclusively of one of the factors – maybe the most important – that allow us to evaluate a translation for the web. Thus, a good starting point for a web translator would be to use the instruments that corpus linguistics provides and analyze "what is out there". Computer technology comes to the rescue of the translator with keyword popularity tools.

8.3 Keyword popularity tools

Keyword popularity tools are programs that generate keywords and visualize statistics relative to specific ones. *Wordtracker*, for instance, is one such program which can give a valuable help to translators because it can provide an idea as to the most searched keywords and key phrases relative to a certain topic, and therefore the ones that should supposedly be included in web content.

By searching for *Siena*, we first obtain a list of related keywords: *travel, information, vacation, Tuscany, lodging, Italy, hotel, inn, accommodations, language.* Some of them are not included in the keywords list for SIENA_PAR_EN and SIENA_EN, but most interestingly some of them are only present in SIENA_EN (*travel, vacation, Tuscany* and *hotel*), which supports the claims made in 8.2.

Each of the related keywords can be expanded to see their "collocations", i.e. the most searched key phrases. Expanding Siena, we obtain key phrases such as *St. Catherine of Siena, Siena college, Siena Italy, Siena villa, Siena hotel*, etc. Once again, keywords present in SIENA_EN but not in SIENA_PAR_EN, which confirms that by using these technological tools, translators could in effect bridge the gap between their linguistic education and the more technological requirements of web-translation/localization.

9. Conclusions

The expansion of the web brings about new challenges, questions and brain teasers for linguists, translation scholars and practitioners. Translating for the web is indeed a complex task, which involves many different levels of analysis, and urges translators to develop skills which, traditionally, were not part of their professional profile. Their linguistic education, though, through the suggestions of Corpus Linguistics and Descriptive Translation Studies provide them with some valuable tools to progress in the integration of their competences with those required by web content SEO (Search Engine Optimization), a crucial part of web content translation.

Web translation can be considered a type of specialized translation, and the translation of tourist-related webpages are no exception. The specialization of this type of translation derives from a multiplicity of factors of which the specificities of the language of tourism and of the technical features of the media are but two aspects. Many questions at the theoretical level still need to be answered, also in terms of the evaluation of what a good web translation is, since, as House (1998:197) points out "[…] different views of translation itself lead to different concepts of translation quality, and different ways of assessing it".

As far as the present discussion is concerned, I think that keyword analysis is an important part of the translation process for the web. Whether this process should be called translation, localization or adaptation is a question for further research with the premise in mind that web content is not just an online version of a printed or printable document, but a brand new land for Translation Studies and linguistics to explore.

References

Agorni, M., 2000, "Quale teoria per la pratica della traduzione multimediale?", in R.M. Bollettieri Bosinelli, C. Heiss, M. Soffritti and S. Bernardini (eds.), *La traduzione multimediale. Quale traduzione per quale testo?*, CLUEB, Bologna, p.395-406.

Askehave, I. and Nielsen, A. E., 2006, "Digital genres: a challenge to the traditional genre theory", Information Technology and People 18-2, pp.120-141. Bassnett, S., 2002, *Translation Studies*, 3rd Edition, Routledge, London.

Bertuccelli Papi, M., 2000, *Implicitness in Text and Discourse*, ETS, Pisa.

Biber D., 2004, "Towards a typology of web registers: A multi-dimensional analysis", Invited lecture, *Conference on Corpus Linguistics: Perspective for the future*, University of Hidelberg Germany.

Bowker, L. and Pearson, J., 2002, *Working with Specialized Language. A practical guide to using corpora*, Routledge, London.

Bruti, S., 2007, "Complexity in Audiovisual Translation. Some Reflections on the Speech Act of Complimenting", in M. Bertuccelli Papi, G. Cappelli and S. Masi (eds.), *Lexical complexity: theoretical assessment and translational perspectives*, Plus Pisa University Press, Pisa.

Cappelli, G., 2006, *Sun, Sea, Sex and the Unspoilt Countryside. How the English language makes tourists out of readers*, Pari Publishing, Pari.

Castello, E., 2002, *Tourist-information Texts: a Corpus Based Study of Four Related Genres*, Unipress, Padova.

Dann, G., 1996, *The Language of Tourism*, CAB International, Oxford.

Fodde, L. and Denti, O., 2005, "Cross-cultural representations in tourist discourse: The image of Sardinia in English tourist guides", in M. Bondi and N. Maxwell (eds.), *Cross-cultural encounters – linguistic perspective*, Officina Edizioni, Roma, pp.116-129.

Gambier, H., 1994, "Subtitling: People Translating People," in C. Dollerup and A. Lindegaard (eds.), *Teaching Translation and Interpreting*, John Benjamins, Amsterdam/Philadelphia, p. 261-274.

Gambier, H. and Gottlieb, H., eds., 2001, *(Multi) Media Translation. Concepts, practices, and research*, John Benjamins, Amsterdam/Philadelphia.

Gotti, M., 2006, "The Language of Tourism as Specialized Discourse", in O. Palusci and S. Francesconi (eds.), *Translating Tourism. Linguistic/cultural representations*, Editrice Università degli Studi di Trento, Trento, pp.15-34

Guidère, M., 2005, "Translation Process in International Advertising", Translation Journal 11-1, http://accurapid.com/journal/15advert.htm, last accessed January 12, 2007

Heiss, C. and Bollettieri Bosinelli, R.M., eds., 1996, *Traduzione multimediale per il cinema, la televisione, la scena*, Clueb, Bologna.

House, J., 1997, *Translation Quality Assessment: A Model Revisited*, Günter Narr, Tübingen.

House, J., 1998, "Quality of Translation", in M. Baker (ed.), *Routledge Encyclopedia of Translation Studies*, Routledge, London, pp.197-200.

House, J., 2001, "Translation Quality Assessment: Linguistic Description versus Social Evaluation", *Meta* XLVI 2, pp.243-257.

House, J., 2007, "Translation Criticism: From Linguistic Description and Explanation to Social Evaluation", in M. Bertuccelli Papi, G. Cappelli and S. Masi (eds.), *Lexical complexity: theoretical assessment and translational perspectives*, Plus Pisa University Press, Pisa.

Kress, G, and Threadgold, T., 1988, "Towards a Social Theory of Genre", *Southern Review* 21/3, pp.215-243.

Laviosa, S., 2002, *Corpus-based Translation Studies. Theory, Findings, Applications*, Rodopi, Amsterdam/New York.

Morville, P. and Rosenfeld, L., 2006, *Information Architecture for the World Wide Web*, Third Edition, O'Reilly, Sebastopol, CA.

Morkes, J., and Nielsen, J.,1997, "Concise, SCANNABLE, and Objective. How to Write for the Web", http://www.useit.com/papers/webwriting/writing.html, last accessed January 12, 2007.

Nigro, M. G., 2006, *Il linguaggio specialistico del turismo. Aspetti storici, teorici e traduttivi*, Aracne Editrice, Roma.

Partington, A., Morley, J. And Haarman, L. eds, 2004, Corpora and Discourse, Peter Lang, Bern.

Pavesi, M., 2005, *La traduzione filmica. Aspetti del parlato doppiato dall'inglese all'italiano*, Carocci, Roma.

Pym, A., 2001, "Localization and Linguistics", Paper presented to the *SLE Conference*, August 2001, http://www.tinet.org/~apym/on-line/loclinguistics.pdf, 14pp, 321kb, last accessed on January 12, 2007.

Pym, A, 2004a, "Localization from the Perspective of Translation. Overlaps in the Digital Divide?", Paper presented to the *SCALLA Conference*, Kathmandu, January 2004, http://www.elda.org/en/proj/scalla/SCALLA2004/Pymv2.pdf, 7pp, 152kb, last accessed on January 12, 2007.

Pym, A., 2004b, *The Moving Text: Localization, translation, and distribution*, John Benjamins, Amsterdam.

Quah, C.K., 2006, *Translation and Technology*, Pelgrave Macmillan, London.

Santini, M., 2006, "Web pages, text types, and linguistic features: Some issues", *ICAME Journal*, Vol. 30, pp.67-86.

Stein, D., 2006, "The website as a domain-specific genre", Language@internet, Special Volume 1-6, http://www.languageatinternet.de, last accessed January 12, 2007.

Taylor Torsello, C., 1999, "Sul Concetto di Genere", in M. Peschieri, (ed.), *La radiocronaca calcistica: dall'Analisi di un corpus alla descrizione di un genere*, SSLMIT, Trieste, pp.vii-xxiii.

Tognini Bonelli, E., 2004, "Welcoming Children, Pets and Guests: towards Functional Equivalence in the Languages of Agriturismo and Farmhouse Holidays", in K. Aijmer and B. Altemberg, (eds.), *Advances in Corpus Linguistics*, Rodopi, Amsterdam-New York, pp.50-70.

Ulrych, M., 2002, "An evidence-based approach to applied translation studies", in A. Riccardi (ed.), *Translation Studies. Perspectives on an Emerging Discipline*, CUP, Cambridge, pp.198-213.

Vestito, C., 2005, "The Tourist Representation of the Italian South", *Textus* XVIII, pp.369-380

Wilkinson, M., 2005, "Discovering Translation Equivalents in a Tourism Corpus", Translation Journal 9-4, http://accurapid.com/journal/34corpus.htm, last accessed January 12, 2007.

Notes

[1] Information architects and web developers work continually to find the best way to display information on web pages (see Morville and Rosenfeld 2006), and to make these latter most accessible and usable. Their research, however, consists in finding the optimal way to display conventional and standardised elements of web pages and not to "invent" new ones.

[2] Some big companies have multiple English versions of their pages to meet the needs of North American, British, Australian and New Zealand guests.

[3] The comparison here focuses on the English translation vs. the original English web content. SIENA_PAR_IT will not be considered in this discussion, but during the analysis it was considered in order to determine whether the discrepancies between the two English corpora were due to the influence of the Italian original and in most cases it was.

[4] Even so, it could still rank high in search engines because of other factors which contribute to ranking but this does not change the evaluation of the translation-localization.

January 25, 2007. Version n.1.
cappelli@cli.unipi.it
Domicilio privato
Via degli Olivi 6
58045 Civitella Marittima (GR)