English for tourism in the non-native English classroom: Machine translation and corpora

Dominic Stewart
University of Trento, Italy

Abstract

This chapter focuses upon English for tourism in the foreign-language classroom, with particular attention to the translation of tourist texts from Italian to English, and considers some of the language resources available to learners. Machine translation (MT) can constitute an important aid in the English for tourism classroom, but despite recent improvements, the output of machine translators is not always reliable, notably within the domain of collocation. Since collocation is often beyond the expertise of even the most advanced L2 students (students of foreign languages), further assistance is needed in order to tidy up MT outcomes. Large monolingual corpora are particularly useful for the purpose of obtaining collocation data and thus of enhancing the quality of machine-generated texts. At the same time, it is shown that the corpus user must possess the necessary intuitions to interpret the data and make appropriate choices.

Keywords: machine translation, DeepL, language corpora, BNC, ukwac, translation into a foreign language, collocation

This paper focuses upon English for tourism in the non-native English classroom with particular attention to the translation of tourist texts from Italian to English. It will be argued that students of English for tourism, above all those translating from their native language into English, can be greatly assisted by machine translation (MT), but that despite the prodigious steps forward made in the recent past, the output of machine translators can be erratic, notably within the sphere of collocation. Since collocation is a mercurial phenomenon often transcending the expertise of even the most advanced L2 students (students of foreign languages), further assistance is required. While dictionaries can be useful in helping the user to decide in favour of or against a specific combination of words, it is clear that they can embrace collocational phenomena only to a limited degree. With this in mind, large monolingual corpora can be exploited for the purposes of retrieving empirical collocation data and ultimately enhancing the quality of machine-generated texts. However, as I hope to demonstrate above all with the final example analysed in this paper, mere

combinatorial possibility is not sufficient, so the corpus user must possess the necessary intuitive awareness to make appropriate choices.

Online translators

One way of confronting the issue of teaching English for tourism in the non-native English classroom is to ask students to compose tourist texts in their native language and then to machine-translate them into English, or, in translation training, to advise students simply to paste the foreign source text into a machine translator and then to click on the appropriate button. In other words, one could stop teaching English for tourism altogether. This scenario is not as futuristic as it sounds. Consider the following machine translation of a text taken from a website describing some caves in the north-east of Italy:

> The Caglieron Caves complex is located in Breda di Fregona, near Vittorio Veneto, and consists of a deep gorge carved into the limestone by the Caglieron stream and a series of artificial cavities that open on the walls of the canyon. These are quarries dating back to the 16th century that testify to the mining activity of the so-called 'piera dolza' (sandstone), used in the building of the time for jambs and frames. Today, an equipped path allows you to follow the course of the stream, immersing yourself in this unique and fascinating environment.
>
> This charming itinerary is suitable for adults and children, the only recommendation is to wear shoes suitable for slippery terrain. The route has wooden walkways suspended above the stream, which allow you to admire the depth of the gorge, the many waterfalls several meters high and the potholes created by the erosive action of water. Over time, deep natural tunnels have formed inside these cavities, which have been explored by speleologists. The low temperature and the lack of light inside the caves has favoured the development of a series of microenvironments of considerable botanical and zoological interest.
>
> At the end of the walk, we come out into the open air near a picturesque and ancient mill, now restored and used as a tavern. During the summer season the caves become an exciting setting for theatrical performances and cultural initiatives of various kinds.

This translation, not perfect but with a degree of accuracy and appropriateness practically unthinkable just a few years ago, was produced by the online translator *DeepL*,[1] first made available

[1] For details consult www.DeepL.com/Translator

in September 2017. As underlined by Heiss & Soffritti (2018), who examine the use of *DeepL* between Italian and German, upon its inception translation operators were struck by the consistent quality of its outputs, which compared favourably with MT applications adopted by Google, Microsoft, and Facebook.

I myself used the Italian source text of the version reproduced above for the purposes of a translation exam two years ago at the University of Trento, Italy, and at the time only a handful of my 30 advanced students of English were able to produce anything as accurate as the version generated by *DeepL*. This clearly invites the question of whether students working with English as L2 would not be better off adopting translation software rather than producing texts themselves. However, before analysing MT output in more depth, let us consider the more general issue of the translation of tourist texts into English as L2.

Translating Tourist Texts into English as L2

Although references to translation into L2 have traditionally been scarce in studies on translation, the situation has changed over the last twenty years with the publication of several works on the topic, including conference proceedings (Grosman, Kadric, Kovaĉiĉ, Snell-Hornby, 2000; Kelly, Martín, Nobs, Sánchez, Way, 2003) and monographs (Campbell, 1998; Pokorn, 2005). These have revealed largely favourable attitudes to translation into L2 among translation scholars, though it has been observed, notably by Kearns (2007), that this positive orientation is not mirrored in professional ambiences, where attitudes to translation into a foreign language remain circumspect, with translation tasks generally assigned, availability permitting, to native speakers of the target language. Such attitudes are often reflected in pedagogical environments; for example in Italy translation trainers generally teach translation into their own native language (Stewart, 2000).

The situation obtaining in professional and pedagogical translation contexts thus implies an enduring qualitative distinction between the production of native speakers and that of non-native speakers, a distinction which has been problematised in recent years in applied linguistics and in ELT in general, particularly when English is the encoded language (for example Widdowson, 1994; Davies, 2003). While some translation scholars have underlined that the native / non-native dichotomy has become more blurred when English is the target language (for example Pokorn, 2005), it is a dichotomy which remains conspicuous in professional situations and in translation training. What has been better recognised in recent years is that on a global level it is completely implausible to insist upon translation into L1. As Kearns (2007) underlines, the directionality (that is into or from the translator's native language) adopted in professional environments is contingent

upon availability. With regard to translation assignments involving so-called languages of limited diffusion, for example Polish, there are so few non-Polish translators with an acceptable command of Polish that questions of directionality are practically redundant. In other words, it would be unrealistic to decree that only native speakers of, for instance, English or Spanish should translate from Polish to English or Spanish. A macroscopic example of this is translation from Chinese to other languages. Thus the question of who actually carries out or revises translation tasks is very language-dependent. See Stewart (2011; 2012, pp. 6-13) for further details. In consideration of this, it may not matter if tourist literature (as opposed, for example, to literary texts), as well as some other text typologies, is not stylistically perfect or does not demonstrate native-like proficiency, as long as it delivers the source text message clearly and persuasively. Thus a rather clumsy sequence such as "near a picturesque and ancient mill" in the translation reproduced above, would by this reasoning be acceptable, whether produced by human or machine.

If we accept this line of reasoning, then it holds important implications both for vocational translation training, i.e., training which takes place in an academic environment but with an eye to the professional world of translation, and for the teaching of English for tourism in general. The focus would be on an unambiguous, persuasive delivery of the source-text message rather than on elegance and native-like proficiency, especially since the majority of the target readers of English-language tourist texts are in any case likely to be non-native speakers of English.

The question that arises at this point is an intriguing one. If we agree that in the teaching of English as L2 for tourism the emphasis should be on clarity of information and persuasiveness rather than native-like elegance, and if machine translation compares favourably with advanced students' production in English, then why teach English for tourism at all? Translation trainees could simply machine-translate the source texts with which they are confronted, while students preparing original tourist texts could simply compose the text in their own language and then adopt MT.

The human touch: Assessing MT outcomes

In reality there are counter-arguments to the notion that English for tourism in the non-native English classroom should be left to a machine. Firstly, not all tourist texts can be machine-translated as successfully as the one reproduced above regarding the Caglieron Caves. The output quality of *DeepL* and other online translators can vary considerably from one source text to the next. Since the average user is largely unaware of the criteria utilised by online translators, there can be surprising qualitative differences between translations of different source texts which the human translator might judge to be of an analogous level of difficulty. For this reason, end-users of

machine translation, particularly those for whom the output language is not their own, need to have not only considerable expertise in the target language but also a solid grounding in the use of translation resources in order to be able to evaluate and edit the texts generated (Zanettin, 2018).
With this in mind let us consider a further translation produced by *DeepL*. The Italian source text, which describes a path with 52 tunnels built on Mount Pasubio near Vicenza in the north-east of Italy, is included in the Appendix. The *DeepL* translation of this text is reproduced here, and the parts underlined will be discussed below.

> Extraordinary work of military engineering that leads from Bocchetta Campiglia to the 'Porte del Pasubio' (1935 m) (refuge gen. Achille Papa) and that allowed the supply of troops perched on the Pasubio with a daring path protected from enemy actions.
> It is an excursion of great satisfaction, a historical route that all lovers of hiking, and especially those who are interested in history, should at least once attend. Along the way, numerous interesting educational signs illustrate the history and construction details of the road. This is probably the most beautiful and exciting of the excursions in the Venetian Pre-Alps. Always crowded with hikers, as always crowded with walkers and cyclists is the refuge Pope of Cai di Schio, especially on summer weekends.
> The road is covered in 3:00/3:30 hours starting from the pass of Xomo or reaching (from the latter) the mouth Campiglia. The difference in height is about 800 meters and you go through very bold tunnels, even with curves and bifurcations, and it is essential to have a reliable flashlight. Especially in tunnels, some sections are sometimes slippery due to the dripping of water. The ceiling, in some cases, is quite low, so very tall people consider using a helmet. In particular, pay attention not to lean over the daring road and not to enter secondary tunnels, often semi-microlled or dangerous. The excursion should not be underestimated and is not a tourist walk, even if it does not present particular difficulties or dangers.

It is striking that the quality of the English of this translation is vastly inferior to the one reproduced earlier, with problems on several different levels. Since the creators of *DeepL* have not disclosed with any precision the criteria adopted in order to yield results, one can only surmise what the complicating factors might be, and I shall try to do this below. Many of the shortcomings highlighted in the translation could, however, be corrected without too much effort by advanced students of English, as will be outlined in the following section.

Correcting grammar, syntax, lexis, and place names

It will be postulated in the sub-section entitled "Correcting Collocation" below that online language aids can help students of L2 to tidy up machine translations into the L2 concerned. However, in my experience many of the flaws in *DeepL* versions can be eradicated fairly quickly by advanced students without recourse to such resources. Here are some examples, divided for convenience into the categories of grammar, syntax, lexis, and place names.

Grammar. The sequence *An extraordinary work* is preferable to the literal and unidiomatic *Extraordinary work* without a determiner**,** while *tall people should consider* is a better rendering than *tall people consider*. In this latter case the software has not recognised the meaning of the subjunctive form *prendano in considerazione*, but the human translator would recognise this difference at once. One could therefore hypothesise that machine translation runs into trouble when confronted with those instances of the Italian subjunctive which convey a different meaning by comparison with the Italian indicative.[2]

Syntax. The word order of *at least once attend* is not ideal (nor is the collocation *attend an excursion*, for which see below), but can easily be rearranged. As regards *Always crowded with hikers, as always crowded with walkers and cyclists is the…*, the syntax is clumsy and would be reformulated.

Lexis. The excessively formal *bifurcations* (the formal nature of this term is flagged in monolingual dictionaries) would be corrected to *forks***,** and the non-existent *semi-microlled* would be converted without too much effort to *partially collapsed*. These suggest that MT applications (i) struggle to distinguish formal from informal contexts, and (ii) have difficulty with attached prefixes such as *semi*.

Place names. There is manifest uneasiness here, as there often is when place names are machine-translated. The two references in the source text to the mountain refuge called Generale Achille Papa are rendered in a confusing way, and indeed give the impression that General Papa was a pope (*papa* in Italian). We also find the bizarre translation *the mouth Campiglia* instead of the original name *Bocchetta Campiglia* (*bocchetta* is a technical word for a type of mountain pass, but it can

[2] A similar case is the *DeepL* version of the Italian news headline "Pedofilia: la chiesa collabori" (which could be rendered as "Paedophilia: the church must cooperate"), where the subjunctive form *collabori* denotes the need for cooperation precisely because in reality cooperation has been in short supply. Yet *DeepL*, again unable to differentiate subjunctive and indicative forms, renders it with the factive sequence "Paedophilia, the church collaborates".

also function as the diminutive form of *bocca*, meaning *mouth*), as well as *the pass of Xomo* rather than the conventional word order *the Xomo Pass* (along the lines of *the Khyber Pass*). When confronted with place names, machine translators are frequently unable to determine when or when not to render those constituents of the name which have lexical meaning, in the present instances *papa* and *bocchetta*.

Correcting collocation

As stressed above, the flaws discussed could be tidied up by means of some rapid editing. However, it is the collocational oddities of the passage which are potentially more problematic, because collocation – a notorious stumbling-block for anyone dealing with a foreign language – is more difficult to assess for the (non-native) post-editor. Back in 1988, the translation theorist Peter Newmark observed that the translation process involves "a continual struggle to find appropriate collocations" (1988, p. 213). Since dictionaries are perforce of limited assistance for the investigation of collocation, at this point post-editors can profitably turn to language corpora in order to seek further data and thus improve the machine-translated version (Heiss & Soffritti, 2018). There are many online interfaces enabling users to extract data from corpora, and one of the most efficient and user-friendly of these is Sketch Engine[3] (Kilgarriff et al. 2014), a corpus management software offering a battery of search strategies. For the purposes of investigating collocation, Sketch Engine offers three main query strategies: (i) concordancing, (ii) collocational profiles, and (iii) Word Sketch. I shall consider the degree of usefulness of each of these with reference to some of the problematic co-occurrences generated by *DeepL*. The corpora adopted here are the *British National Corpus*, a recent corpus of 100 million words of British English (90% written texts and 10% spoken) and the *British Web 2007* (also known as *ukwac*) containing around 1 billion 300 million words, whose texts are extracted from uk-domain websites.

(1) *It is an excursion <u>of great satisfaction</u>, a historical route that* […]

This sequence is a fairly literal translation of the Italian *di grandissima soddisfazione* (literally *of very great satisfaction*). Here the most obvious initial corpus search for the non-native English post-editor is by means of the simple search 'great satisfaction',[4] which retrieves a concordance of 74 occurrences in the *BNC*, including the combinations *greater satisfaction* (17) and *greatest satisfaction* (10) because the simple search option is lemmatised. However, having sorted the results alphabetically to the left of the key words, the user is struck both by the number of times we find (i)

[3] Sketch Engine is a corpus manager and analysis software created by Lexical Computing Ltd in 2003, now with over 500 corpora in over 90 languages. See https://www.sketchengine.eu for further details.

[4] Single inverted commas will be adopted for corpus queries.

a preceding verb governing the phrase (*GIVE*, *PROVIDE*, *DERIVE*, *BRING*[5]), and (ii) the sequence *with great satisfaction* (Figure 1).

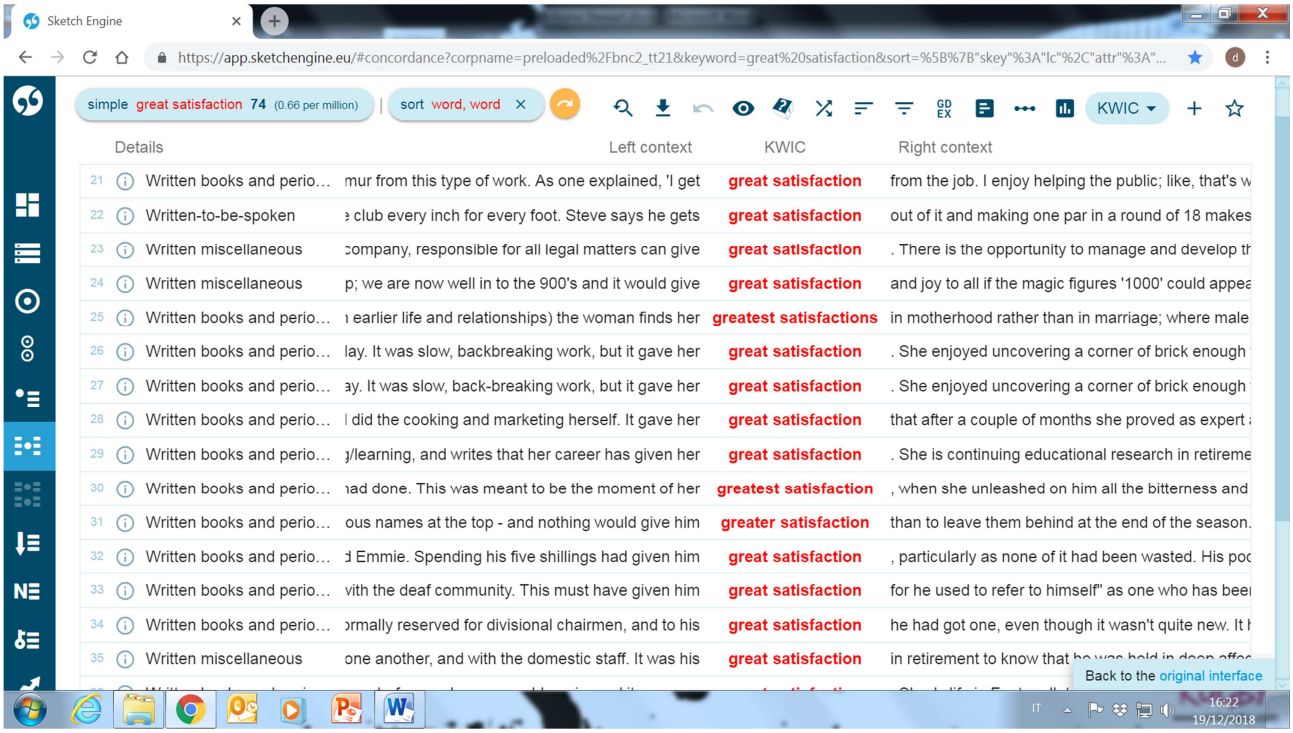

Figure 1. Extract of *BNC* concordance 'great satisfaction' with alphabetical sorting L1/L2: page 2

If the user makes the query more specific by entering 'of great satisfaction' (see Figure 2), we discover that this sequence occurs just twice in the *BNC*, while of the 33 occurrences in the *ukwac*, 20 are part of the phrase *source of great satisfaction* and three are part of *matter of great satisfaction* (interestingly, this time none of the results features either the comparative – *greater* – or superlative – *greatest* – forms).

Simple search:

	BNC	*ukwac*
of great satisfaction	2	33

	BNC	*ukwac*
source of great satisfaction	1	20

[5] Word-forms in capital letters are to be intended as lemmas: therefore *GIVE* = *give, gives, giving, gave, given*.

	BNC	*ukwac*
matter of great satisfaction	0	3

Figure 2. Frequency outcomes of the searches 'of great satisfaction', 'source of great satisfaction', 'matter of great satisfaction' in the *BNC* and *ukwac*

Other L1 occurrences (that is, forms one word to the left of the search word / words) in the *ukwac* are (one of each) *feeling, expression, cause*, and *sense.* In R1 position the most frequent finding is *to* followed immediately by a pronoun, for example *a matter of great satisfaction to me*. In Sinclair's (1996) terms, the broader lexical item of the phrase *of great satisfaction* looks like this:

"NOUN *BE a source/matter of great satisfaction to* PRONOUN"

The corpus data would therefore suggest that the combination *excursion of great satisfaction* is unidiomatic.

(2) *It is an* excursion […] *that all lovers of hiking* […] *should at least once* attend

The advanced student of English will certainly be already aware of collocations such as *attend a lesson, attend university, attend a meeting* etc., but may well feel the need to the check the frequency of the less familiar *attend an excursion*. One way of doing this is to perform the simple search 'attend', which you will find has over 9,000 hits in the *BNC*, and then adopt the collocation facility, perhaps using the default span of three words to the left and three words to the right of the key word. The result is a long list of collocates of *attend*, listed in order of salience rather than in order of frequency (Figure 3).

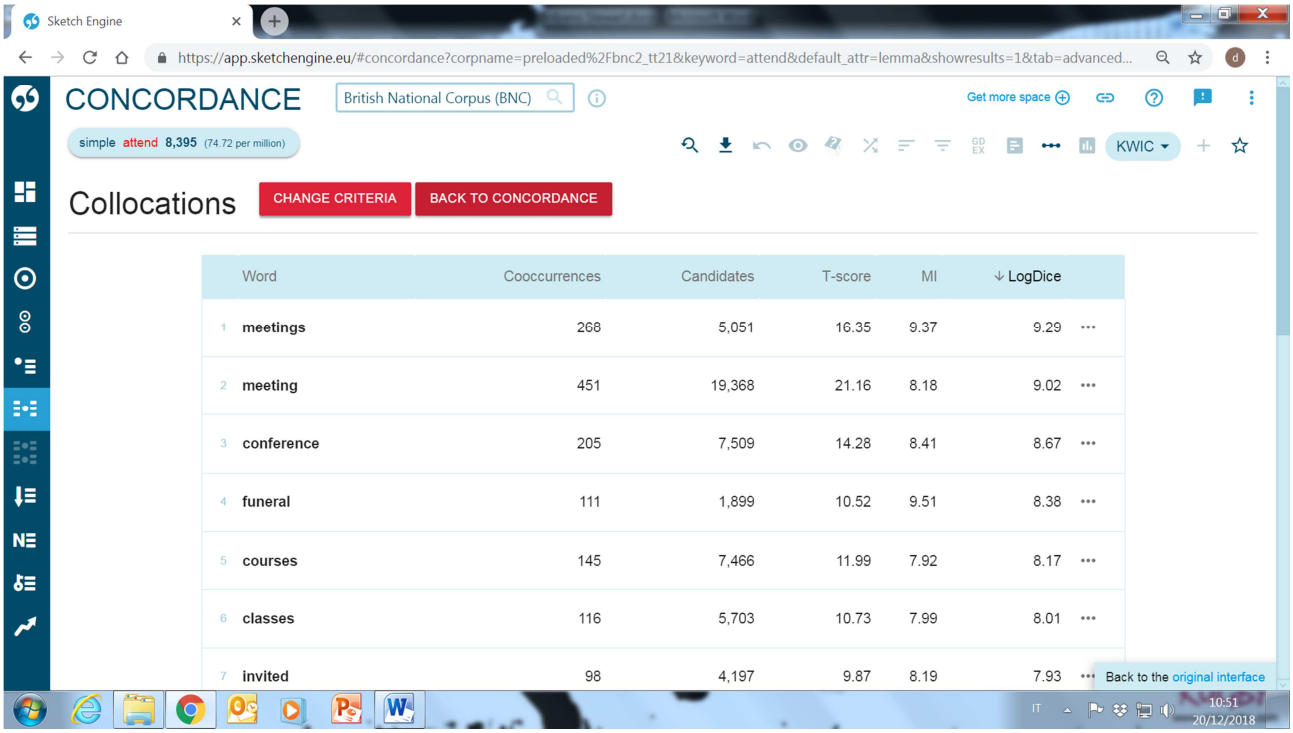

Figure 3. Extract of *BNC* collocation profile of 'attend': span L3 / R3

By clicking on individual collocates, the user can visualize the type of structures in which the combinations occur. Among other things, this will reveal that most of the items in this collocational profile are grammatical objects of the active verb, for instance "60,000 people attended a Sandinista rally", or the subject of a passive structure, for instance "at a rally attended by 100,000 people". However, some of the items are irrelevant to the current context, either because they are grammatical subjects of the active structure, for example *delegates* ("about 150 delegates attended the Grozny meeting") and *members* ("eighty-six members attended committee meetings"), or because they are not nouns (*invited, unable, annual, regularly*).

For this reason it would be more efficient in the present case to adopt the Word Sketch facility. Sketch Engine's Word Sketch is "an automatic, corpus-derived summary of a word's grammatical and collocational behaviour" (Kilgarriff, Kovář, Krek, Srdanovic, Tiberius 2010), originally conceived in order to meet some new corpus query requirements. Concordances had been a ground-breaking invention but they often prove hard to read and can therefore be time-consuming (Kilgarriff & Kosem, 2012, pp. 9-13; Atkins & Rundell, 2008, p. 109), even with alphabetical sorting. Collocational profiles too are useful in disclosing the immediate environment of a word, but, as noted above, tend to be in the form of simple lists deriving from a relationship of frequency or salience between the keyword and its co-occurrences, with no distinction of the grammatical relations at work within the sentence. In short, this type of collocation retrieval "is grammatically

blind. It only considers proximity" (Kilgarriff & Kosem, 2012, p. 14; see also Thomas, 2015). Word Sketch, on the other hand, furnishes lists of collocates divided according to the grammatical construction in which they appear:

> The Word Sketch […] provides one list of collocates for each grammatical relation the word participates in. For a verb, the subject, the objects, the conjoined verbs (*stand and deliver, hope and pray),* modifying adverbs, prepositions, and prepositional objects, are all presented in different lists (Kilgarriff, Rychlý, Smrz, Tugwell, 2008, p. 298).

Let us now examine a Word Sketch query for 'attend' (Figure 4), but before analysing this it should be emphasised that Word Sketch is lemmatised *in toto*. Indeed it is not possible to retrieve single word-forms which are not lemmas. For example, a *BNC* Word Sketch of the past-tense form 'gave' is unproductive, despite the fact that *gave* appears nearly 22,000 times in the corpus.

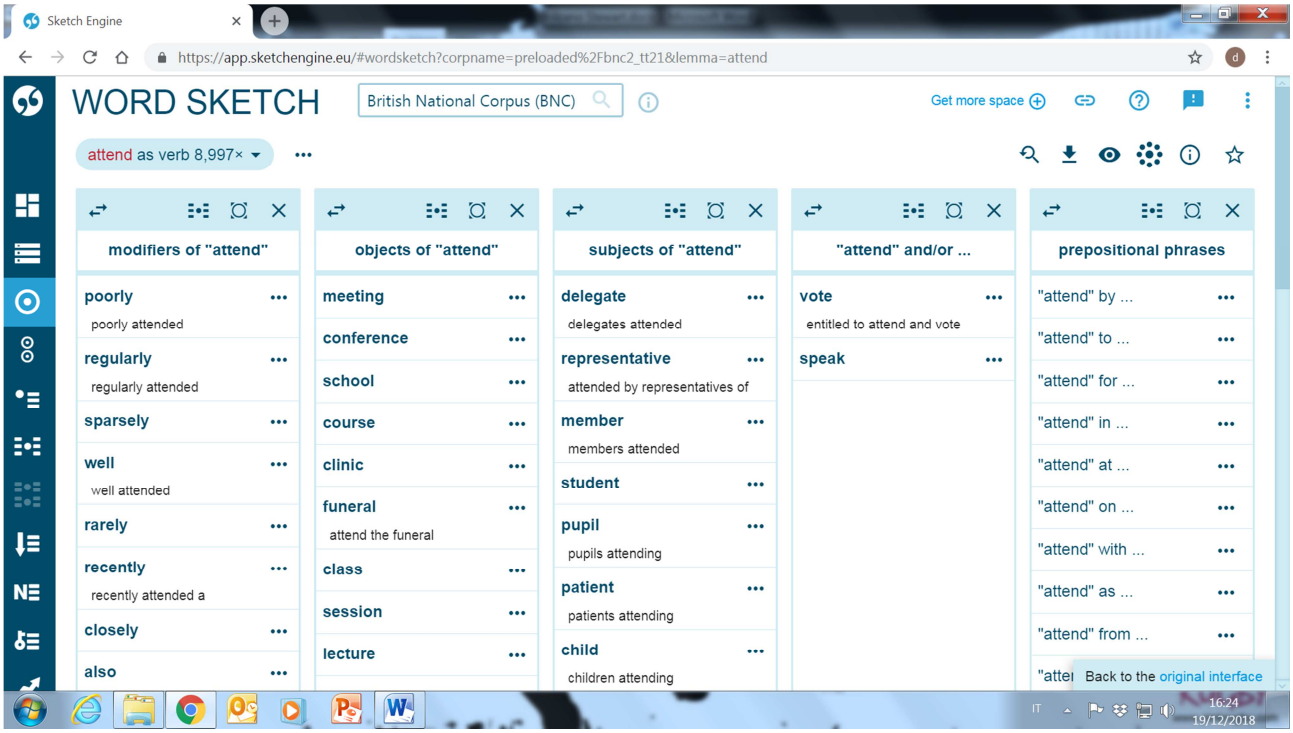

Figure 4. Extract of *BNC* Word Sketch query 'attend'

For present purposes we are concerned with the second column 'objects of *attend*'. If we try to divide these grammatical objects into approximate semantic clusters, we find that things attended tend to be academic/scholastic (for example *school, lessons, lectures*), ceremonies (*weddings, mass,*

funerals), meetings (*summits, reunions*), medical institutions (*hospitals, clinics*), legal situations (*courts, hearings*), and performances (*shows, matches*). There is no trace, however of any member of a cluster corresponding to the notion of 'journey' (*trip, journey, excursion, hike, ride, voyage*, etc.), and the same query in the *ukwac* produces similar results. Again, it would seem, the online translator has generated a very unlikely outcome.

(3) *Along the way, numerous interesting <u>educational signs</u> illustrate the history and construction details of the road.*

Once again this is a literal rendering of the source text phrase. Bilingual dictionaries provide various equivalents of both *cartello* (*sign, panel, board, signboard*) and *didattico* (*educational, instructional, didactic, information, informative*), but not for *cartello didattico*, so the task for the learner of English here is to identify an appropriate combination from the range of options supplied. One way to achieve this is by means of a series of simple two-item searches (as underlined above the simple search option is lemmatised, and therefore in this case it will retrieve both the singular and plural form of the noun). Consider the frequency rates for the two-word combinations listed in Figure 5.

Simple search	*BNC*	*ukwac*
information board	12	744
information panel	6	322
information sign	2	145
informative sign	0	14
educational sign	0	6
educational signboard	0	1
instructional board	0	1
didactic panel	0	0
illustrative panel	0	0
didactic sign	0	0

Figure 5. Frequency outcomes of some two-word combinations in the *BNC* and *ukwac*

Naturally the analysis of these outcomes needs to be both quantitative and qualitative. It goes without saying that not all of the occurrences retrieved are appropriate to the context under

discussion – for example three of the six occurrences of *educational sign* in the *ukwac* are part of the sequences *educational sign language interpreting* and *educational sign language interpreters*, while another is part of *educational sign systems*. However, there is no doubt that the data offer some very useful pointers, and that *information board* and others are likelier outcomes than the one generated by *DeepL*.

(4) *The road is covered in 3:00/3:30 hours*

An extension of the simple search option in Sketch Engine is what has sometimes been referred to as a query builder, which enables the user to elaborate the simple query by adding in specifications of co-text. For instance in the present case one can retrieve occurrences involving the combination *COVER* (verb) + *ROAD* within a span of five words. This search produces 5 hits in the BNC, but 4 of these occurrences are inappropriate to our context ("the actions in the Plan cover all road users"; "road users cover their track costs"; "television cameras will also be on site to cover two more unusual road races"; "Optional rubber tip to cover metal tip for road use"). However, the remaining example is very relevant, particularly as it occurs in a tourist guide (about Madeira): "you will need three to four days to cover most of the main roads and to savour some of the breathtaking views".

Aside from the fact that one occurrence is insufficient for cogent conclusions, this search option is not optimum for the information we require, firstly because it again relies on simple juxtaposition without considering parts of speech (we need instances where *road* is the grammatical object of *cover*), and secondly it would be useful to know if *cover* can govern terms in the same semantic field as *road* such as *way, path, track, trail, itinerary*. Once again, therefore, the most productive search would appear to be a Word Sketch of *cover* as a verb (Figure 6).

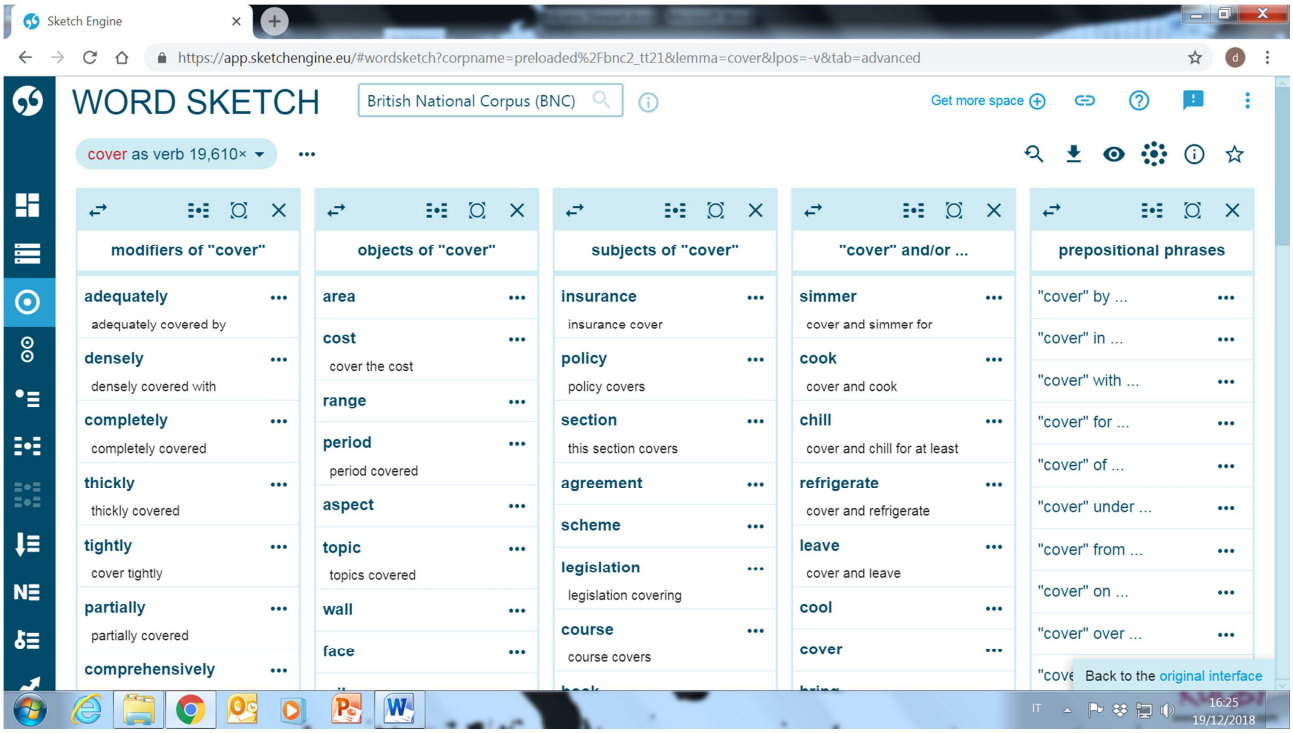

Figure 6. Extract of BNC Word Sketch query 'cover' as verb

Here the learner would scan the list of grammatical objects of *cover* (Figure 6 contains only an extract) to search for words semantically akin to *road*, but there are hardly any. The closest are *area* (619 hits), *mile* (106), *ground* (105), *acre* (60), *distance* (45), and *track* (43, though this occurs almost always within the idiom *cover one's tracks*), and a similar list results from the same search in the *ukwac*. The next step for the user is to click on these collocations to try to get an idea of their typical co-text, after which it becomes clear that the combinations *cover* + *area*, *cover* + *mile, cover* + *acre* habitually co-occur with exact distances or dimensions: "Covering an area of 20,000 sq km"; "the gardens and grounds covered eight acres"; "it had taken them four days to cover the hundred-odd miles". Some of these also include the time taken to complete the distance (as in the latter example), something which perhaps explains both the sequence generated by *DeepL* above and the one occurring in the text about Madeira. The user should at this point be in a position to hypothesise that the associated lexical item (see again Sinclair, 1996) of *cover* with this meaning could perhaps be expressed as follows:

"*cover* + *area/distance* + expression of dimensions and/or distance"

However one wishes to formulate the lexical item, *cover* + *road* is certainly unusual, so again the *DeepL* output would appear to be far from ideal.

(5) perched on the Pasubio with a daring path protected from enemy actions. […] you go through very bold tunnels […] pay attention not to lean over the daring road

Here the source text author shows a predilection for the Italian adjective *ardito* (= *daring, bold*), and more specifically in its superlative form *arditissimo* (*very daring, very bold*), which occurs three times in this brief description to underline the extraordinary and perilous work of engineering required to forge the path described through the mountain rock. As one might expect of a personification of this type, the searches 'daring path', 'daring road' and 'bold tunnel' retrieve no occurrences in the *BNC* or the *ukwac* aside from one instance where *path* refers to a course of action. The Word Sketches 'daring' and 'bold' disclose that although these adjectives are certainly not restricted to qualifying animate nouns – for example there are plenty of occurrences such as *bold plan, bold pattern, daring experiment, daring clothes* – there would appear to be scarcely any combinations of the type *bold road, daring track*, so the user would be persuaded to seek alternative solutions.

(6) as always crowded with walkers and cyclists is the refuge

Finally, a problem of a different order is the sequence [*the refuge is*] *crowded with hikers*. This is not so much an issue of combinatorial possibilities, inasmuch as *crowded with* collocates with all sorts of animate nouns (for instance *visitors, skiers, students*) in both the *BNC* and the *ukwac*. The point of interest here is that the adjective used twice in the source text is *affollato* (both occurrences are in its superlative form *affollatissimo*), which is the first equivalent of *crowded* provided in bilingual dictionaries, but which in actual fact has a more favourable meaning than *crowded*. Whereas the Italian adjective emphasises that the refuge is a very popular destination and a great place to visit (and thus would ideally be rendered by something like *busy* or *bustling*), the English adjective evokes fairly unpleasant scenarios of not being able to move or get served, and is therefore best avoided in the current context. It goes without saying that, at least for the present, such contrastive subtleties are beyond the reach of machine translation, but they may also lie beyond the reach of advanced learners of English, to whom it may not occur that *crowded* is a less positive word than *affollato*, and who may as a result accept it without question. This raises the thorny issue of to what extent the learner of English is able to interpret as problematic the outcomes of machine translation, an issue that will be addressed in the next section.

Non-native speaker intuitions

The above sections provide examples of searches that can be performed with Sketch Engine by learners of English in order to tidy up outcomes of MT applications. This modus operandi, however, rests upon a paradox, which is that learners of English may lack the necessary language skills to recognize as problematic the precarious outcomes that MT applications inevitably generate. This connects to the following scenario which is no doubt familiar to language teachers. The instructor marks as wrong a given structure, and asks the students why they did not check it in dictionaries or other language resources, upon which they reply that it had never crossed their minds that the structure in question was in any way problematic or that it might warrant investigation, and this could apply to several of the examples analysed. For instance, in example (1) above the sequence *of great interest* combines with all sorts of preceding nouns, so students might assume axiomatically that *of great satisfaction* is similarly eclectic. This conundrum, which has significant repercussions not only for translation training in tourism but for the teaching of English for tourism *tout court*, is not easily resolved. One might be tempted to suggest that learners should check any word combination they have not seen or heard before, or at least that they do not recall having seen or heard before, but this is simplistic. Learners may not have come across the combination *pink stapler* before, but they could intuitively be sure that the two words combine successfully, inasmuch as colours co-occur fairly freely with nouns, and inasmuch as there is no reason why a stapler should never be pink. In any case this co-occurrence is unlikely to turn up in dictionaries or corpora. There can be no tailor-made solutions to the question of precisely when learners should avail themselves of language resources, because personal intuitions about language are unique, but learners adopting MT applications should be particularly wary of renderings which sound rather too literal (*attend an excursion, educational sign, daring road*).

One's intuitions about language are crucial for the purposes not only of judging the correctness and appropriateness of texts but also of appraising the relevance and usefulness of outcomes generated by language resources, whether these be dictionaries, MT, or corpora. In the literature on corpus studies much has been made of a supposed opposition between personal intuition and empirical data, but in reality the two cannot plausibly be opposed within the realm of corpus linguistics, because without intuitive or introspective reactions we as users of corpora would be totally unable to interpret the data they contain. For further discussion see Stewart (2010, pp. 122-151).

Conclusions

Despite important progress in the area of machine translation in recent times, even the limited number of examples provided in this paper demonstrate that users cannot rely blindly on machine outputs.

While advanced students of English and trainee translators into English as L2 can reasonably be expected to weed out the majority of MT shortcomings without too much difficulty, the slippery area of collocation is in many cases beyond their intuitive reach. Since for obvious reasons the remit of dictionaries does not extend to an exhaustive treatment of collocation, it is at this point that language learners – in the present case learners of English for tourism – can make rewarding use of electronic corpora in order to make empirical investigations into the frequency and plausibility of given word combinations. Word Sketch, one of the battery of options provided by Sketch Engine, would appear to be the most useful in this regard, because it takes into account not only mere proximity but also the grammatical relations obtaining between the constituents of syntagms.

Naturally, raw frequency alone cannot suffice for meaningful conclusions. Learners need to be sensitive to both qualitative and quantitative data in order to prioritise relevant usage and discard irrelevant usage, but they also require awareness of when or when not to make the corpus search in the first place. The last example analysed (*affollatissimo* / *very crowded*) constitutes an interesting case precisely because although corpus searches attest that *very crowded* habitually co-occurs with people engaged in all sorts of diverse activities, they do not help to establish that *very crowded* describes a less favourable scenario than *affollatissimo*. Therefore the user's intuitions about language are crucial, particularly within the domain of tourist texts where persuasiveness is a defining feature of the discourse.

Appendix

La strada delle 52 gallerie

Opera straordinaria di ingegneria militare che conduce dalla Bocchetta Campiglia alle 'Porte del Pasubio' (m.1935) (rifugio gen. Achille Papa) e che consentiva l'approvigionamento delle truppe arroccate sul Pasubio con un arditissimo percorso al riparo dalle azioni nemiche.

E' una escursione di grandissima soddisfazione, un percorso storico che tutti gli appassionati di escursionismo, e a maggior ragione coloro che s'interessano di storia, dovrebbero almeno una volta frequentare. Lungo il percorso numerosi cartelli didattici interessanti illustrano la storia e i dettagli costruttivi della strada. Si tratta, probabilmente, della più bella ed entusiasmante tra le escursioni

nelle Prealpi Venete. Sempre affollatissima di escursionisti, come sempre affollatissimo di camminatori e ciclisti è il rifugio Papa del Cai di Schio, specie nei fine settimana estivi.
La strada si percorre in ore 3:00/3:30 partendo dal passo di Xomo o raggiungendo (da quest'ultimo) la bocchetta Campiglia. Il dislivello è di circa 800 metri e si percorrono gallerie molto ardite, anche con curve e biforcazioni, ed è indispensabile una affidabile torcia elettrica. Specie nelle gallerie, alcuni tratti a volte sono scivolosi causa lo stillicidio d'acqua. Il soffitto, in alcuni casi, è abbastanza basso, quindi le persone molto alte prendano in considerazione l'uso del caschetto. In particolare prestare attenzione a non sporgersi oltre l'arditissima stradina e a non inoltrarsi in gallerie secondarie, spesso semicrollate o pericolose. L'escursione non è da sottovalutare e non è una passeggiata turistica, pur non presentando particolari difficoltà o pericoli.

Source: Magico Veneto: https://www.magicoveneto.it/index.html

References

Atkins, S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Campbell, S. (1998). *Translation into the Second Language*. Harlow: Longman.

Davies, A. (2003). *The Native Speaker: Myth and Reality*. Clevedon: Multilingual Matters.

Grosman, M., Kadric, M., Kovaĉiĉ, I., Snell-Hornby, M. (Eds). (2000). *Translation into Non-Mother Tongues in Professional Practice and Training*. Tübingen: Stauffenburg Verlag.

Heiss, C., & Soffritti, M. (2018). DeepL Traduttore e Didattica della Traduzione dall'Italiano in Tedesco: Alcune Valutazioni Preliminari. *Intralinea*. Retrieved from http://www.intralinea.org/specials/article/deepl_traduttore_e_didattica_della_traduzione_dallitaliano_in_tedesco

Kearns, J. (2007). Translate and be damned! Inverse translation and professional recognition. In Sokoloski, R., Duda, H., Klimkowski, K., Klimek J. (Eds). *Warsztaty Translatorskie / Workshop on Translation, 4*, 171–191. Lublin/Ottawa: Towarzystwo Naukowe.

Kelly, D., Martín, A., Nobs, M-L., Sánchez, D., Way, C. (Eds). (2003). *La Direccionalidad en Traducción y Interpretación: Perspectivas Teóricas, Profesionales y Didácticas*. Granada: Editorial Atrio.

Kilgarriff A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: Ten Years On. *Lexicography ASIALEX, 1*, 7-36.

Kilgarriff, A., & Kosem, I. (2012). Corpus Tools for Lexicographers. In Granger, S., and Paquot M. (Eds). *Electronic Lexicography* (pp. 31–55). Oxford: Oxford University Press.

Kilgarriff, A., Kovář, V., Krek S., Srdanovic, I., Tiberius, C. (2010). A Quantitative Evaluation of

Word Sketches. In Dykstra, A., & Schoonheim, T. (Eds). *Proceedings of the 14th EURALEX International Congress 2010* (pp. 372–379). Leeuwarden: Fryske Akademy.

Kilgarriff, A., Rychlý, P., Smrz, P., & Tugwell, D. (2008). The Sketch Engine. In Fontenelle, T. (Ed). *Practical Lexicography: A Reader* (pp. 297-306). Oxford: Oxford University Press.

Newmark, P. (1988). *A Textbook of Translation*. London: Prentice Hall.

Pokorn, N. K. (2005). *Challenging the Traditional Axioms: Translation into a Non-mother Tongue*. Amsterdam: John Benjamins.

Sinclair, J. (1996). The Search for Units of Meaning. *Textus*, *9*, 75-106.

Stewart, D. (2012). *Translating Tourist Texts from Italian to English as a Foreign Language*. Naples: Liguori.

Stewart, D. (2011). Translation Textbooks: Translation into English as a Foreign Language. *Intralinea* ('Review Articles'). Retrieved from http://www.intralinea.it/reviewarticles/eng_more.php?id=959_0_51_0_M%

Stewart, D. (2010). *Semantic Prosody: A Critical Evaluation*. London and New York: Routledge.

Stewart, D. (2000). Poor Relations and Black Sheep in Translation Studies. *Target: International Journal of Translation Studies*, *12:2*, 205-228.

Thomas, J. (2015). Stealing a March on Collocation: Deriving Extended Collocations from Full-text for Student Analysis and Synthesis. In Leńko-Szymańzka, A., & Boulton, A. (Eds), *Multiple Affordances of Language Corpora for Data-driven Learning* (pp. 85-108). Amsterdam and New York: John Benjamins.

Widdowson, H. (1994). The Ownership of English. *TESOL Quarterly 28:2*, 377-389.

Zanettin, F. (2018). Electronic Tools and Resources for Translating and Writing in the Digital Age. *Intralinea*. Retrieved from http://www.intralinea.org/specials/article/electronic_tools_and_resources_for_translating_and_writing_in_the_digital