

Tutti i diritti riservati.

© 1983 *Rivista di Studi Italiani*ISSN 1916-5412 *Rivista di Studi Italiani*(Toronto, Canada: in versione cartacea fino al 2004, online dal 2005)

LINGUISTICA

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY? AN EVALUATION OF AVAILABLE MACHINE TRANSLATION SERVICES ON A SAMPLE OF ‘ADJECTIVE + NOUN’ COMBINATIONS

BARBARA BERTI

Università degli Studi di Milano

Introduction

Recent technological development has had a deep and positive impact on automatic translation. The increase in terms of size and number of parallel corpora produced in many different languages, in conjunction with considerable improvement in the available algorithms, have seen machine translation enter a new stage, whose outcomes are right before our eyes. In particular, a shift in the paradigm from statistical to neural network-based machine translation (Brown *et al.* 1990, Brown *et al.* 1993) has successfully addressed many shortcomings of traditional machine translation systems (Wu *et al.* 2016), thus improving the quality of the proposed translations.

Indeed, for many decades statistical machine translation has been the framework of reference within which computer scientists and computational linguists have worked. Such a system is based on the translation of sequences of words (i.e. statistically identified ‘phrases’) of different lengths (Kohén, Och & Marcu 2003)¹. The machine translation system is trained on large bilingual corpora, which limits its versatility. In fact, some of the shortcomings of statistical machine translation concern translating material that is not similar to the content of the training corpus. Particularly problematic is the treatment of idiomatic and slang expressions which are more typical in informal language thus tend to be less represented in the training corpora. The delta in the

¹ The term ‘phrase’ refers to a sequence of words identified by the application of statistical models. A phrase in statistical machine translation might not coincide with a phrase in the syntactic sense.

performance of such systems can be rather impressive: when the text to be translated is consonant with the material included in the training corpus, the quality of the translation is generally very good. Conversely, when the text deals with topics that are not covered in the training corpus, the adequacy of the translation falls drastically.

The recent switch to neural network-based translation systems (Kalchbrenner & Blunsom 2013, Cho *et al.* 2014, Sutskever *et al.* 2014) has enabled computer scientists to overcome the phrase-based translation paradigm in which a sentence is divided up into sub-components that are dealt with separately. In neural network-based translation systems, each word is coded along a 500-layered vector that represents its unique characteristics. Based on the language pairs used for training, the neural network will self-define what these dimensions should be (e.g. part of speech, gender, level of politeness etc.). The translation is performed at the sentence level as each word is encoded in further vectors depending upon which other words it occurs with, while in statistical machine translation the maximum window comprises five words. Moreover, neural network models are based on continuous feedback and consequent fine-tuning.

So far, for the reasons outlined above collocations and idiomatic expressions have failed to be properly handled by machine translation services (henceforth MTSs). Moreover, collocations are undoubtedly one of the most complex and intricate concepts that pertain to the domain of lexis. Owing to their intrinsically fleeting nature, they have proven to be particularly difficult to pin down, especially from a theoretical point of view.

If looked at from the perspective of phraseology (Benson *et al.* 1986, Cowie 1998, Mel'čuk 1998, Moon 1998), although it has been possible to identify a series of distinctive features (i.e. fixedness, non-substitutability, non-compositionality, metaphoricity, opacity), such features could never be measured objectively, making any attempt to classify this type of word combination a matter of subjective judgment. As noted by many (Granger 1998: 147, Wei 1999: 4, Jackson & Amvela 2000: 114, Nesselhauf 2003: 225, Chang *et al.* 2008: 285), this has resulted in an intricate web of definitions as almost every scholar who has approached the topic has elaborated her/his own terminology.

Objectivity can be achieved by treating collocations as a statistical phenomenon (Sinclair 1991, Halliday 1966), whereby the 'attraction' between words can be quantified on the basis of the computation of their co-occurrence in a corpus. However, whilst figures provide a more impartial outlook, they do not resolve the issue of what lies at the core of 'words that go together' as different collocations can be extracted by means of a wealth of different statistical measures (mainly frequency, mutual information, T-score, Chi-square).

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF ‘ADJECTIVE + NOUN’ COMBINATIONS

From a foreign language speaker’s point of view, collocations are a critical issue owing to anisomorphism between languages. Moreover, given that they tend to be poorly represented in English-Italian bilingual lexicography (Berti 2017), learners of English often have to resort to other resources, especially when they wish to translate Italian collocations into English. Thus this study sets out to investigate how such combinations are dealt with by automatic machine translation, a tool commonly utilised by learners.

Methodology

The first step in this work was the collection of a large enough sample of combinations to be translated automatically. In order to gain deeper understanding of the possible variables that shape the quality of automated translation, it is important that such combinations should display different strengths of association, ideally ranging from loose combinations that can be generated by rules of syntax and semantics (e.g. *great discomfort*) to fully-fledged collocations that are more idiomatic, therefore more fixed (e.g. *black anger*).

To this end, an initial list of 32 English nouns representing a wide range of different frequencies was selected:

English wordlist = [‘affection’, ‘amazement’, ‘anger’, ‘anxiety’, ‘astonishment’, ‘aversion’, ‘boredom’, ‘concern’, ‘contempt’, ‘curiosity’, ‘devotion’, ‘discomfort’, ‘disgust’, ‘dread’, ‘embarrassment’, ‘fear’, ‘fright’, ‘guilt’, ‘happiness’, ‘hate’, ‘hatred’, ‘interest’, ‘joy’, ‘love’, ‘nervousness’, ‘pride’, ‘rage’, ‘revulsion’, ‘sadness’, ‘scorn’, ‘shame’, ‘sorrow’].

The MTSs investigated were Bing Translator (henceforth BT) and Google Translate (henceforth GT), the former developed by Microsoft, the latter by Google.

A Python script was implemented by the author so that each word would be automatically looked up in the online Oxford Collocation Dictionary (available at <http://www.freecollocation.com/>), its adjectival collocates extracted, coupled with the noun so as to form an ‘adjective + noun’ combination that would finally be translated utilising the available BT and GT APIs.

As a result, a list of 282 English collocations was obtained, together with their Italian equivalents provided by both BT and GT. Table 1 (APPENDIX) shows the first 10 ‘adjective + noun’ combinations extracted and their Italian translation.

Given the complexity and multifaceted nature of the concept of collocation, it was felt that evaluation based on an ‘adequacy’ versus ‘inadequacy’ binary opposition would not capture the nuances connected with this phenomenon². For this reason, the idea of adequacy was explored at greater depth so that four defining criteria were eventually identified after carefully examining the equivalents proposed by the MTSs as well as the most common types of error that would occur. Adequacy was thus evaluated on the basis of syntactic plausibility, intelligibility, typicality, and semantic/pragmatic equivalence. Syntactic plausibility refers to whether the provided equivalent can be deemed syntactically correct in Italian, regardless of whether it appropriately translates the English collocation it stems from. In some cases a translated collocation would result in a syntactically nonsensical combination, such as the one of *cane-come devozione*, provided by BT.

Intelligibility has to do with the semantic aspect of the proposed equivalent. The score for intelligibility refers to whether the meaning of the provided equivalent can be understood by a speaker of Italian or whether it is a mere juxtaposition of words that do not make sense in the target language. The above example (i.e. *cane-come devozione*) is a case of non-intelligible combination (as well as a syntactically ill-formed one). Another example is *interessante interesse*, an expression that does abide by the rules of syntax but is not meaningful.

Typicality is seen as one of the truly defining features of collocations, connected as it is with the idea that usage determines which word combinations become established in the language. A word combination that can be deemed typical is one that is perceived as habitual, natural, and one that would be spontaneously produced by a native speaker of a language. Typicality is indeed

² The methods employed for assessing the performance of automatic translation can be either manual or automated (e.g. BLEU). Manual evaluation can be done on the basis of adequacy, ranking, or post-annotation. Adequacy refers to the correctness of a proposed translation and is computed by attributing a score to each translation. Adequacy can be expressed as a binary value or as a 5- or 7-point scale. Ranking is a method that bypasses scoring and only requires the translations to be ordered from the most appropriate to the least appropriate by the reviewer. Post-annotation is based on the count of the number of changes made by a reviewer in order to make the translation fully acceptable so that translations that require a higher number of changes rank lower. The manual evaluation based on adequacy was preferred in this case, especially because “metrics based on word-to-word matches are not really appropriate for collocation-oriented evaluation, as they underestimate the impact that the substitution of a single word (the collocate) has on the overall sentence quality” (Wehrli *et al.* 2009: 133).

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

a particularly crucial aspect of the translation of collocations as the English combinations in the sample are expected to be rendered with equally typical combinations in Italian. This is clearly one of the most challenging issues to resolve, even for professional translators, owing to the very nature of the English and Italian languages and their dissimilarities (this issue will be discussed in the final section of this paper).

Finally, semantic/pragmatic equivalence examines the possibility of juxtaposing the semantic/pragmatic meaning of the English collocation to that of the given equivalent in Italian. In fact, there were cases in which, despite being semantically appropriate, a provided equivalent would not correspond at all to the English counterpart (e.g. *amorevole sorella* stemming from the English *sisterly love*).

The four criteria can be reduced to a set of four yes/no questions:

1. Is the proposed equivalent syntactically plausible in Italian?
2. Is the combination semantically intelligible?
3. Is the combination typical in Italian?
4. Is the meaning of the Italian translation equivalent to that of the English collocation?

Depending upon whether the answer to the questions is a 'yes' or a 'no', each of the above-mentioned criteria was given a score of either 1 (i.e. yes) or 0 (i.e. no). The sum of the scores obtained for each criterion by each combination results in an adequacy index, ranging from a minimum value of 0, for a syntactically and semantically ill-formed juxtaposition of words, to a maximum of 4, for a well-formed and adequate equivalent. For example, *affetto reciproco*, which translates *mutual affection*, is syntactically well-formed, perfectly intelligible, typical in Italian, and equivalent to its English counterpart. For this reason, it was given an adequacy score of 4. Conversely, *inattività curiosità*, which is provided by GT as the equivalent of *idle curiosity*, was given a score of 0 in that it does not translate the English collocation, it is syntactically ill-formed, and overall it makes no sense.

Results

The results of the translation assessment for BT and GT are summarised in Table 2 (APPENDIX). As can be gauged, both MTSs seem to behave very similarly: BT has a 62% success rate, while GT reaches 61%. This means that more than half of the translated combinations resulted in a perfectly acceptable equivalent in Italian (i.e. corresponding to an adequacy index of 4), while less than half were deemed partially or totally inadequate. In particular, about 18%

of the combinations were only slightly inadequate in both MTSs (i.e. score = 3), while the remaining 20-21% were hardly acceptable.

In general, the vast majority of the English collocations resulted in the same equivalent being produced by both MTSs. In fact, only 16% of the translated collocations display some differences. In terms of adequacy, the distribution is rather homogeneous with 19 more successful translations provided by BT, 16 by GT, and 12 that were equally bad. For example, the collocation *heady joy* was correctly translated as *gioia inebriante* by BT while it was rendered as *gioia sconcertante* by GT, a combination showing a problem of semantic prosody, given by the clash between the positive connotation of the noun *gioia* and the negative one of the adjective *sconcertante*. *Growing fear* is another example in which BT performed better by providing the user with the equivalent *crescente paura*, while GT presented *crescendo la paura*, in which the term *increasing* failed to be recognised as an adjective. However, GT offered a better equivalent for the collocation *icy contempt*, translated as *gelido disprezzo*, as opposed to *disprezzo ghiacciato* given by BT, and also for *lasting happiness*, properly rendered as *felicità duratura*, as opposed to the mistranslation of *felicità durevole* by BT. Examples of collocations translated inadequately by both MTSs are *terminal boredom*, given as *noia terminale* by BT and *noia del terminale* by GT, and *dog-like devotion*, rendered as *cane-come devozione* by BT and as *devozione simile al cane* by GT.

Given that the list of English collocations is made up of combinations showing different degrees of associational strength (e.g. *great astonishment* is weaker than *blank astonishment*, while *unfounded fear* is stronger than *real fear*), the performance of weak versus strong collocations was assessed separately in order to ascertain whether the strength of association has an impact on the quality of the translation. The results are summarised in Figures 1 and 2 (APPENDIX), the first referring to BT, the second to GT. As can be appreciated from the figures, the majority of strong collocations was adequately translated by both MTSs. BT performed slightly better, with 52.5% of the combinations achieving a score of 4, while the same score was obtained by approximately 51% of the combinations translated by GT. As for the combinations that scored 0, 16% of them are made up by those translated by BT, whereas 21% belong to those translated by GT. Interestingly, very few combinations got a score of 1 (2.5% in both MTSs), while approximately the same amount scored 2 and 3 (around 15%). Compared to the overall performance of both MTSs, a focus upon strong collocations clearly reduced the quality of the proposed translations so much so that only one in two collocations could be considered an appropriate equivalent in Italian.

By taking a closer look at the kind of strong collocations there are in the sample, one can see that many happen to be literally translatable into Italian. Combinations such as *implacable hatred*, *avid interest*, or *irrational fear* can be translated *verbatim* into Italian as *odio implacabile*, *avidio interesse*, and

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

paura irrazionale. Such combinations were adequately translated by both MTSs. However, literality is no guarantee of adequacy. For instance, *burning interest* was translated by BT as *interesse burning*, *public concern* as *preoccupazione del pubblico* by GT.

As previously mentioned, some cases were treated differently by the two MTSs. For example, while *mild curiosity* was rightly rendered as *lieve curiosità* by GT, it was mistranslated as *curiosità mite* by BT. Conversely, while *mock amazement* was given as *finto stupore* by BT, it resulted in *stupore stupore* when translated through GT.

The translation of weak collocations (i.e. free combinations) was slightly less problematic, as can be gauged from Figures 3 and 4 (APPENDIX). In fact, as many as approximately 67.5% and 69% of the combinations were translated successfully through BT and GT respectively. It is particularly interesting to see how the distribution of the ill-translated collocations changed with respect to that of strong collocations: both bar charts show a decreasing trend that remains constant from score 4 through score 0. Combinations that have a score of 0 are very few. With strong collocations, instead, we observed the same decreasing trend from score 4 through score 1, but the trend rose again at score 0, thus showing a peak of completely unacceptable Italian equivalents. The number of English collocations that failed to be handled by both MTSs is rather high. Of course, whilst free combinations seemed to be dealt with well by both services, approximately 30% of them were still deemed unacceptable. This means that about one in three English collocations failed to be correctly translated into Italian.

After studying the variable 'strength of the collocation' (i.e. strong versus weak), we moved on to identify another factor that might play a role in the quality of machine translations: the impact of context. In fact, so far the collocations had been translated as mere juxtapositions of adjective and noun. Yet at this point it was hypothesised that additional information could help MTSs in the elaboration of more appropriate translations by narrowing the number of possible options by looking at the words that occur in proximity of a given collocation.

To this end, a group of 18 collocations that had been mistranslated by both BT and GT was further investigated. Each English collocation was looked up in the British National Corpus (BNC-BYU) and, among the various sentences in which the collocation appears, one was selected. Once collected, the 18 sentences containing the 18 collocations under study were translated through BT and GT. The process, applied to the collocation *mounting anxiety*, is exemplified in Table 3 (APPENDIX). The collocation was inadequately translated by both MTSs: BT rendered it as *montaggio ansia* while GT proposed *ansia di montaggio*. After the addition of more context, some changes were

detected. Although a radical improvement in the quality of the translation could not be appreciated, BT's Italian equivalent changed to *ansia di montaggio*. In terms of adequacy, both equivalents are clearly unusable. However, it must be noted that, while the former does not abide by either Italian rules of syntax or semantic, the latter is at least syntactically plausible. If on the one hand the changes brought about by BT were not sufficient to make the translation appropriate, in GT's new equivalent we can appreciate a major improvement. The new combination *crescente ansia* is indeed perfectly adequate both as a translation and as an Italian collocation *per se*.

For reasons of space, the complete table with all 18 collocations as translated by BT and GT before and after the addition of context cannot be attached. However, as for GT, in 11 cases out of 18 the translation of collocations embedded within a sentence resulted in different equivalents being produced. Some of the changes were successful (e.g. *redemptive love: amore di rendenzione* > *amore redentore*, *single-minded devotion: devozione singola* > *devozione assoluta*), others were not (e.g. *devozione simile al cane* > *devozione al cane* which are tentative translations of the English *dog-like devotion*).

With regard to BT, the addition of some context brought some change to eight out of 18 collocations. Yet, only one combination really benefited from such change, namely *passing interest*, which was first translated as *passaggio di interesse* and then became *interesse di passaggio*. Some of the remaining combinations are worth mentioning, if only because they will bring a smile to the reader's face. These are: *single-minded devotion* given first as *singolare devozione* and then as *devozione di una singola mentalità*, and *dog-like devotion* given first as *cane-come devozione* and then *devozione cane-like*.

Discussion

This study shows that, with regard to collocations, the rate of successful automatic translations is at best two in three. A difference has emerged in the translation of weak and strong collocations. Overall, the former seem to be handled in a more satisfactory fashion by MTSs (more than two out of three English collocations were adequately translated into Italian) and pose fewer challenges. With the latter, the performance rather decreased by approximately 15% in both MTSs, a sign that the most metaphorical or typical senses taken on by the collocate are still problematic in automatic translation. This certainly has to do with the frequency with which the meaning of a word shifts towards a figurative level. Indeed, metaphorical extensions often emerge only when the collocate (i.e. the adjective in this case) establishes a lexical relation with a particular base, while in most other cases the word retains its literal sense. This may well account for the fact that more literal interpretations were favoured over more metaphorical ones.

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

The data confirm the decisive effect that lexical relations have in language, one which is greatly amplified where translation is concerned. An interesting example is that of the English collocation *mild curiosity*. The combination was rightly rendered as *curiosità lieve* by GT while mistranslated as *curiosità mite* by BT. Depending upon the nouns it collocates with, *mild* can be associated with different translations in Italian and *mite* is certainly one of them. However, when in collocation with the noun *curiosità*, *mite* is clearly not the right equivalent, referring mainly to the weather or to a person. This example shows that the same non-substitutability that characterises collocations in one language and accounts for the impossibility of replacing a collocate by one of its synonyms (e.g. *strong tea* versus **heavy tea*) holds across languages too. When *mild* collocates with nouns concerning weather, manners, or people, it can be translated as *mite*. Yet, when it accompanies nouns concerning illness, taste, feelings, or products (e.g. detergents), it corresponds to the Italian adjective *lieve*. Surprisingly, the collocation *mild discomfort* that is one of the 282 collocations under investigation was adequately translated as *lieve disagio* by BT.

Since GT provided the right translation of *mild* as *lieve* in our original example, it was worth looking into whether the MTS would adapt to a change of context and associate *mild* to *mite* when a new noun was juxtaposed with it. For this reason, the collocations *mild winter* and *mild man* were given to the machine. GT did indeed provide the correct translations *inverno mite* and *uomo mite*, proving to be flexible enough to distinguish between two different meanings in Italian.

Yet the study highlights that some of the issues concerning the automatic translation of collocations remain to be tackled. Some of these issues are of a technical nature. On a few occasions the handling of the part of speech resulted in the wrong tag being assigned to the components of a collocation. This was the case with *sisterly love*, translated by GT as *amorevole sorella* (A/N *loving sister*), in which the parts of speech were inverted: the adjective *sisterly* was turned into the Italian noun *sorella*, while the noun *love* was tagged as an adjective. BT was not immune to this kind of mistake either. For example, *withering contempt* was translated as *disprezzo appassimento*, showing that BT failed to recognise *withering* as an adjective and turned it into a noun in the Italian equivalent.

Occasionally, the MTSs came up with non-existent words. For example, GT translated *idle curiosity* as *curiosità idlea*, despite *idlea* not being a word in the Italian vocabulary. With respect to this, a major issue was the handling of the compound adjective *dog-like*, as seen in the previous section. In general, it seems that compound adjectives were not translated appropriately by either MTS.

On a few occasions BT showed mismatches in the agreement concerning the gender between adjective and noun. *Idle curiosity* was given as *inattivo curiosità*, *redemptive love* became *amore redentrice*, and *nagging anxiety* was rendered as *ansia fastidioso*.

Conclusion

After analysing the Italian translation of a sample of 282 ‘adjective + noun’ English collocations, we can conclude that, although much progress has been made in recent years, collocations still pose a challenge to automatic translation. In truth, being so language-specific, collocations represent a hurdle in themselves when a contrastive perspective is embraced.

First, there is the issue of conceptual voids, i.e. collocations that exist in one language representing a conceptual unit but are not codified in another. For instance, how can combinations such as *agony aunt* or even *wishful thinking* be translated when they cannot be matched to a conceptual referent in Italian?

Second, even when the same concept is shared between two languages, it might take a different linguistic shape in each of them. For example, *split second* can be translated into Italian as *frazione di secondo*, while *secondo spaccato*, which is the literal counterpart and also keeps the same part of speech structure, would not be appropriate.

Third, collocations that do exist in both languages and refer to the same concept might be different in terms of style, register, and prosody. For instance, in Italian, *ansia crescente* and *ansia galoppante* have the exact same meaning but differ in register, the latter being very informal. Conversely, in English *mounting anxiety* and *rampant anxiety* do not stand in the same relation to one another in that the latter does not belong to an informal register.

Finally, issues of frequency dishomogeneity should also be considered, i.e. two collocations having the same structure, the same referent, and the same stylistic implications might differ in their popularity within a linguistic community. For example, the Italian collocation *estate rovente* is correctly translated into English as *scorching summer*. However, while the former is very common in the Italian language (especially in the media), the latter is not as common in English. Such disparity between the relative frequency of the two collocations influences the way in which speakers of Italian and speakers of English respond to them.

In conclusion, to answer the original question ‘Can collocations be translated automatically?’, we have to bring into play the aims and expectations of MTSs users. In reception tasks, MTSs can certainly help users obtain basic understanding of the collocation in the foreign language. Yet, if users are to translate collocations from their mother tongue into the foreign language, such services cannot be trusted blindly as more work is needed for the production of

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

well-formed, natural stretches of language, particularly with respect to the translation of strong collocations.

Last but not least, we have caught a glimpse of how the presence of context might affect the quality of the translation of collocations. Undoubtedly, a more systematic study conducted on a larger sample of collocations is needed in order to produce conclusive evidence regarding an improvement in the quality of the translation of contextualized collocations. Nevertheless, it is our hope that the present study marks a step in that direction.

REFERENCES

- Benson, M., E. Benson, and R. Ilson. *Lexicographic Description of English*, Philadelphia: John Benjamins Publishing, 1986.
- Berti B. *Lexical collocations in Bilingual Lexicography*, Milano: Mimesis International, 2017.
- Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., and P. S. Roossin. "A Statistical Approach to Machine Translation", *Computational linguistics*, 16, 2, 1990, pp. 79-85.
- Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and R. L. Mercer. "The Mathematics of Statistical Machine Translation: Parameter Estimation", *Computational Linguistics*, 19, 2, June 1993, pp. 263-311.
- Chang, Yu-Chia, J. S. Chang, H. J. Chen, and H. C. Liou. "An Automatic Collocation Writing Assistant for Taiwanese EFL Learners: A Case of Corpus-based NLP Technology", *Computer Assisted Language Learning*, 21, 3, 2008, pp. 283-299.
- Cho, K., van Merriënboer, B. Gulcehre, C. Bougares, F. Schwenk, H. and Y. Bengio. "Learning Phrase Representations Using RNN Encoder-decoder for Statistical Machine Translation", in *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014.
- Cowie, A. P. (ed.). *Phraseology: Theory, Analysis and Applications*, Oxford: Clarendon Press, 1998.
- Granger, S. "Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae", in A. P. Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press, 1998, pp. 145-60.
- Halliday, M. A. K. "Lexis as a Linguistic Level", in C. E. Bazell, J. G. Catford, M. A. K. Halliday, and R. H. Robbins (eds.), *In Memory of J. R. Firth*, London: Longmans, 1966, pp. 148-62;

- https://www.issco.unige.ch/en/staff/seretan/publ/EAMT2009_EW_VS_LN_LR.pdf
- Kalchbrenner, N. and P. Blunsom. “Recurrent Continuous Translation Models”, in *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, 2013, pp. 1700-1709.
- Koehn, P., F. J. Och, and D. Marcu. “Statistical Phrase-based Translation”, in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics*, 2003.
<http://www.aclweb.org/anthology/N03-1017>
- Mel’čuk, I. A. “Collocations and Lexical Functions”, in A. P. Cowie (ed.), *Phraseology. Theory, Analysis and Applications*, Oxford: Clarendon Press, 1998, pp. 23-53.
- Moon, R. *Fixed Expressions and Idioms in English*, Oxford: Clarendon Press, 1998.
- Nesselhauf, N. “The Use of Collocations by Advanced Learners of English and Some Implications for Teaching”, *Applied Linguistics*, 24, 2003, pp. 2223-2242.
- Sinclair, J. M. *Corpus, Concordance, Collocation*, Oxford: Oxford University Press, 1991.
- Sutskever, I., O. Vinyals, and V. Q. Le. “Sequence to Sequence Learning With Neural Networks”, in *Advances in Neural Information Processing Systems (NIPS 2014)*, 2014, pp. 1-9.
- Wehrli, E., V. Seretan, L. Nerima, and L. Russo. “Collocations in a Rule-based MT System: A Case Study Evaluation of Their Translation Adequacy”, in *Proceedings of the 13th Annual Conference of the EAMT, Barcelona, May 2009*, pp. 128-135.
<https://pdfs.semanticscholar.org/15fe/d034a4ac3873a74c7d6702e53063193e965f.pdf>
- Wei, Y. “Teaching Collocations for Productive Vocabulary Development”, *Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (New York, 1999)*, 1999;
<http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED457690>.
- Wu, Y., M. Schuster, Z. Chen, Le V. Q., M. Norouzi, W. Macherey, W. Krikun, M. Cao, Y. Gao, Q. Macherey, K. Klingner, J. Shah, A. Johnson, M. Liu, X. Kaiser, Ł. Gouws, S. Kato, Y. Kudo, T. Kazawa, H. Stevens, K. Kurian, G. Patil, N. Wang, W. Young, C. Smith, J. Riesa, J. Rudnick, A. Vinyals, O. Corrado, G. Hughes, M. and J. Dean. “Google’s Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation”, *arXiv preprint arXiv:1609.08144*, 2016.

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

Dictionary

Oxford Collocations Dictionary for Students of English, Oxford: Oxford University Press, 2002.

Corpus

British National Corpus accessible at <https://corpus.byu.edu/bnc/>^[1]_[SEP]

APPENDIX

ENGLISH COLLOCATION	ITALIAN (BT)	ITALIAN (GT)
deep affection	profondo affetto	affetto profondo
great affection	grande affetto	grande affetto
real affection	vero affetto	vero affetto
special affection	affetto speciale	speciale affetto
strong affection	forte affetto	forte affetto
mutual affection	affetto reciproco	affetto reciproco
black anger	rabbia nera	rabbia nera
deep anger	rabbia profonda	rabbia profonda
great anger	grande rabbia	grande rabbia
sudden anger	rabbia improvvisa	improvvisa rabbia

Table 1. Sample of extracted collocations and their automated translation

	BT	GT
adequacy index 4	174	172
adequacy index 3	53	49
adequacy index 2	23	20
adequacy index 1	9	9
adequacy index 0	23	32

Table 2. Performance of BT and GT on 282 collocations

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
 AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
 SERVICES ON A SAMPLE OF ‘ADJECTIVE + NOUN’ COMBINATIONS

Collocation	mounting anxiety
Collocation in context (BNC)	“This is a common pattern, says Matthew, triggered more by mounting anxiety than by deep-seated dissatisfaction.”
Translation (BT)	“Questo è un modello comune, dice Matthew, innescato più da <u>ansia di montaggio</u> che da profonda insoddisfazione.”
Translation (GT)	“Questo è un modello comune, dice Matteo, innescato più dalla <u>crescente ansia</u> che dall'insoddisfazione profonda.”
Pre- vs. Post context (BT)	montaggio ansia > ansia di montaggio
Pre- vs. Post context (GT)	ansia di montaggio > crescente ansia

Table 3. Example of a contextualised collocation

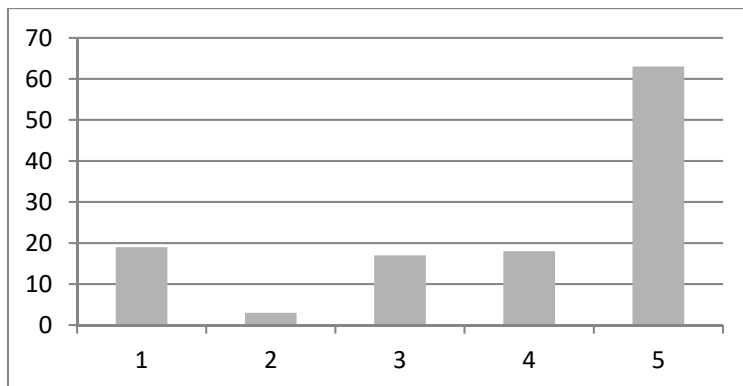


Figure 1. Performance of translation of strong collocations by BT

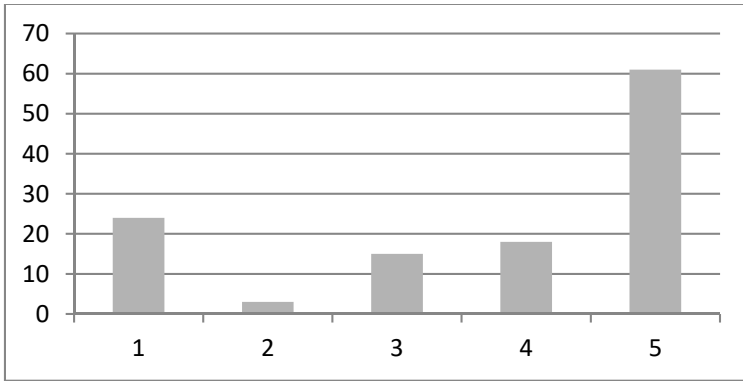


Figure 2. Performance of translation of strong collocations by GT.

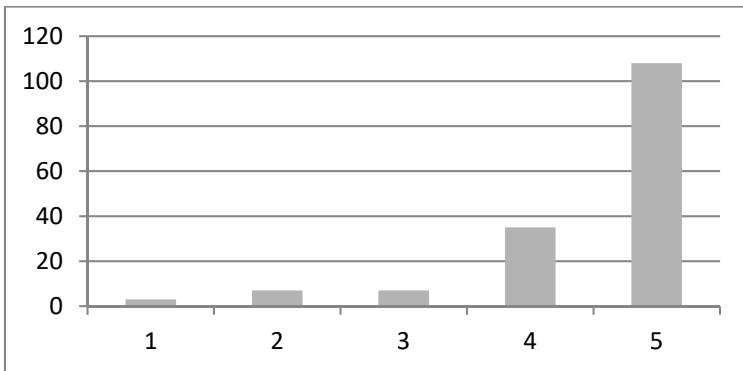


Figure 3. Performance of translation of weak collocations by BT.

CAN COLLOCATIONS BE TRANSLATED AUTOMATICALLY?
AN EVALUATION OF AVAILABLE MACHINE TRANSLATION
SERVICES ON A SAMPLE OF 'ADJECTIVE + NOUN' COMBINATIONS

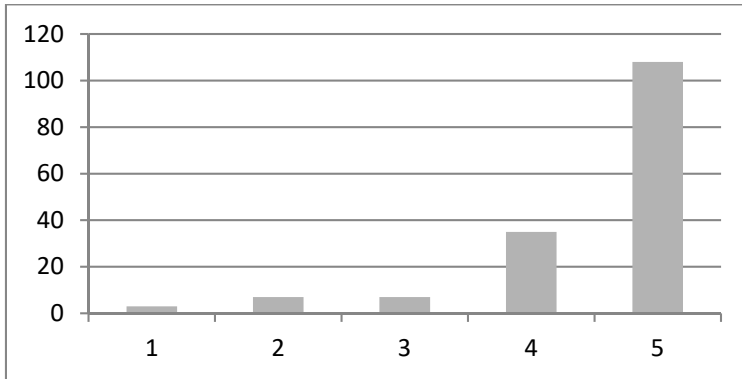


Figure 4. Performance of translation of weak collocations by GT.