Learning verb phrases using a parallel corpus

Gulomjonova Nozigul Dilshodbek qizi

1st year Master's degree student at NUU
E-mail: gnusa44@gmail.com

ABSTRACT

Research on corpora and their potentials for special areas of language learning have been gaining extra attention in latest years. Albeit the utilization of corpora in language mastering and edifying is not incipient, what is novel in the present learn about is that it demonstrates for the first time the potential of a bilingual corpus in learning both grammar and translation. The paper endeavors first to define one of a kind types of corpora and concordancers, and then concentrate especially on the application of English-Uzbek parallel corpus in mastering the most possibly translations of complexified English verb phrases having no equipollent in Uzbek. It is intriguing to optically discern what options the parallel corpus would genuinely grant in such occasions. An experiment in this route reveals that the utilization of parallel corpus in translation will enhance the learners' overall performance at least up to forty-five percent. The remaining section of the paper discusses the more than a few purposes of such corpora in exceptional areas of language studies.

Keywords: Bilingual parallel corpus; English-Uzbek corpus; Translation; Verb phrase

We live in the world of technology in which data is unexpectedly and substantially accessible and networked. As a result, researchers choose to boost strategies for establishing required information exceptionally primarily based on this freely on hand statistics in order to be capable to make use of it properly.

Corpus linguistics if truth be told makes use of massive collections of written as well as spoken language manufacturing to study the language. It tries to describe different elements of language based totally on empirical evidences extracted from corpora. Today, corpus linguistics has been developing greater than earlier than due to widely availability of a vary of sorts of monolingual and bilingual corpora on one hand, and laptop computer software program enhancement on the exceptional hand. In fact, we can see many one of variety corpora in many one-of-a-kind languages and researchers and novices can get proper of entry to these sources as an alternative easily. The modern-day advances in corpus linguistics are now accountable for improvements in overseas language pedagogy. The current paper argues for the use of a bilingual parallel corpus in language instructing surroundings the region the beginners accumulate statistics about the way in which a grammatical sample is laid down in two comparing languages. Typical linguistic patterns are besides troubles and proper away on hand by concordances.

Corpora and Concordances A corpus is a "collection of linguistic data, both written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description or as a plausible of verifying

hypotheses about a language" (Crystal, 1991, p. 86).

Dealing with measurement corpora can be divided into reference and show corpora. Reference corpora have a consistent size; that is, they are no longer expandable (e.g., the British National Corpus), whereas show corpora are expandable; that is, texts are constantly being added (e.g., the Bank of English). Corpora can also be divided relying their shapes as the corpora which contain total texts and these which incorporate samples of a unique length. The latter lets in a greater variety of texts to be blanketed in a corpus of a given size.

In phrases of content, corpora can be both general, that is, attempt to reflect a specific language or range in all its contexts of use (e.g., the American National Corpus), or specialized, that is, reason to focal factor on distinctive contexts and users (e.g., Michigan Corpus of Academic Spoken English), and they can contain written or spoken language. Corpora can moreover represent the extraordinarily right varieties of a single language. For example, the International Corpus of English (ICE) consists of one-million-word corpora information of gorgeous varieties of English (British, Indian, Singaporean, etc.). As implied in the previous section, corpora can additionally in addition include language produced through native or non-native audio device (usually learners). Finally, corpora can be monolingual (i.e., contain samples of solely one language), or multilingual. Multilingual corpora are of two types: they can include the identical text-types in one of type languages, or they can encompass the equal texts translated into extraordinary languages, in which case they are moreover known as parallel corpora (Hunston, 2002; Kennedy, 1998; McEnery and Wilson, 2001; Meyer, 2002). Some corpora are enriched with linguistic archives such as part-of-speech annotation, parsing and prosodic transcription. In these instances it is a methods much less difficult and increased special to retrieve facts than in the case of unannotated corpora.

A concordancer is a variety of software application software for watching a corpus which suggests a listing of occurrences of a special word, area of a phrase or mixture of words, in its contexts. The search phrase is referred to as key word. The most common way of exhibiting a concordance is via the usage of a sequence of traces with the key-word in context (KWIC-format) in which each and every phrase is primarily based in a fixed-length area.

According to the Collins Cobuild English Dictionary, a "concordance" is: "An alphabetical record of the words in a e book or a set of books which verbalizes where each phrase can be situated and frequently how it is used"

A concordance demonstrates authentic patterns of a language higher explicitly and accurately in terms of language samples in their mundane context. In fact, the learners can get an awful lot better notion of the utilization of the phrase looking into the context than they would gain through utilizing merely probing it up in the dictionary.

Concordancers are in particular recommended when utilized in conjunction with a thesaurus. One cannot find information about unknown words and their authentic utilization in the context of the language, but with the avail of a concordancer s/he can.

There is immensely colossal and ever-growing physique of lookup on more preponderant than a few factors of language assimilating erudition of alongside with translation predicated on monolingual as nicely as multilingual corpora (Biber et al., 1998; Hunston, 2002; Kennedy, 1998; McEnery and Wilson, 2001; McEnery et al., 2005, in press; Meyer, 2002; Partington, 1996; Stubbs, 1996, 2001; Tognini-Bonelli, 2001). Many other studies have addressed the utilization of corpora, in general, and parallel corpora, in particular, as a plausible avail in language authoritatively mandating and reading (Johns 1991; Barlow 1995,1996).

Zanettin states a variety of methods in which homogeneous bilingual corpora or data derived from them can be utilized in the translation lecture room as follows:

• Utilizing the Corpus to Translate: an experiment carried out at the School for Translators and Interpreters at the University of Bologna, in which a crew of trainee

translators utilized the Olympics corpus as a predominant implement in translating one short Italian sports article into English.

• Utilizing the Corpus to Learn about Terminology and Content: In the classroom, commensurable corpora can be acclimated to substantiate translation hypotheses and to suggest doable solutions to genuine translation troubles associated to a categorical text. They can supplementally furnish an aptitude to check out kindred domains or sub-domains across languages. A specialized kindred corpus can offer files about terminology and concepts, and about the attestedness of expressions within a certain context.

• Utilizing the Corpus to Explore Texts: A 0.33 approach to cognate corpora involves the utilization of them to investigate a particular genre and/or theme area, as in pre- or post-translation activities. Commensurable corpora can be the source of a possibly illimitable "serendipity process" (Johns 1998) as one phrase or phrase leads to another, depending on the learner's instinct and man or woman proficiency, pastimes or needs. Commensurable corpora furnish newcomers with a plausible of checking out the relationship between contrivances of language which they perceive as preserving some marginally homogeneous attribute or equipollence. Learners can be authoritatively mandated to appear to be for homogeneous attributes between languages and to consider phrases and phrases by betokens of way of deducing edifications which have a sturdy formal resemblance, such as suited names and cognates (visually perceive Partington 1995), or which are proposed as translation equipollents in dictionaries (Zanettin, F. 1998). Bernhard Kettemann studied the utilization of concordances in English Language Edifying (ELT) in grammar, lexicon works as congruous as stylistics and literature edifying. In the grammar area, for instance, he seemed at if-clauses, reported verbalization, the distinction between subsisting best and precedent traumatic and some examples of feasible contrasts between in view that and for (Kettemann, B. 1995). Kennedy investigated the methods of expressing quantification and

frequency in ESL (English as a 2nd language) textbooks (Kennedy, 1987a, 1987b). Holmes moreover examined methods of expressing doubt and stroll in the park in ESL textbooks (Holmes, 1988), whilst Mindt appeared at future time expressions in German textbooks of English (Mindt, 1992). The remaining three studies have homogeneous methodologies, that is, evaluating the applicable constructions or lexica in the pattern textbooks to these in Standard English corpora. Most of these lookup found that there have been cyclopean distinctions between what textbooks are edifying and how native audio contrivance in authenticity use language as evidenced in the corpora. Costas Gabrielatos in an astronomically immense discover out about first defined corpora and their sorts and then mentioned their contribution to language getting acquainted with and injuctively authorizing even as provided examples of their utilization in class. His learn about outlined the transmutations in erudition, faculties and postures that are needed for abecedarians and edifiers to take gain of the opportunities offered by utilizing the availability of corpus resources. Determinately, the paper referred to the obstacles of the utilization of corpora in language edifying, and the potential pitfalls arising from their uncritical use (Costas Gabrielatos. 2005). An implement for work in parallel corpora (including Skandinavian languages and English) has supplementally been developed by Danielsson and Ridings and their efforts to cumulate it into an academic application for coaching translators established (Danielsson and Ridings, 1996). From the above literature it can be considered that parallel corpora have already been efficiently employed in a range of studies. However, what is novel in the present day research about is that we utilized a bilingual parallel corpus for gaining expertise of each grammar and translation. In unique words, a contrivance incorporating these two elements of language reading has been discussed. Here, a parallel corpus comes to avail for positive patterns investigation in language studying as well as evaluating the patterns in 2nd language with these in the first language and deducing some

generalizations from it. As a long way as the engenderer kens, there is no current work in this connection for the dyad of English and Persian languages.

As the assimilating cognizance of manner consists of processing information predicated absolutely on obtained information, the acquisition of incipient cognizance is conventionally got here about due to ever growing the incipient information. Language studying is now now not an exception to this proposition. It is optically discerned as a dynamic and interactive method at some stage in which incipient understanding is often received when some assets are denuded to the abecedarians in nice situations. As a consequence, they will be capable to draw incipient conclusions from such form of expertise processing. A corpus affords the learners with the kinds of sentences that they will encounter when the utilization of the language in genuine esse situations. Corpus use contributes to language edifying in a vary of methods (Aston, 2000; Leech, 1997; Nesselhauf, 2004). The two most referred ones are translation and grammar to which this locate out about is concerned. Today Translators (professional as well as pupil translators) are in a position to investigate and manipulate the statistics contained in corpora utilizing corpus evaluation gear in one-of-a-kind ways (Mosavi Miangah, 2006). Researches on comparison between unique texts and their translations in the structure of parallel corpus have proved that translators can make translation techniques by referring to the installed equipollents below positive situations made by different translators. If the parallel corpus is amply giant and comprehensive, the translators can safely pick out some broadly familiar patterns encountering recurrent translation equivalencies in such circumstances, and consequently generalize some rules. The undetachable section of a parallel corpus is a parallel concordancer. A parallel concordancer is in a role to stumble on the equal sentences in the translated texts. It can provide the researchers with imperative information about the evaluation of more than a few linguistic patterns occurring in the

bilingual corpus. The evaluation of these habitual patterns exhibits the demeanor of certain linguistic contrivances in context, and in some cases challenges or supplement the information provided by denotes of distinct mundane reference implements such as dictionaries and textbooks.

Parallel corpora are viewed as valuable resources for fixing not fully different troubles of linguistics alternatively moreover herbal language processing as nicely as computational linguistics issues. What follows are amongst the variety applications of such corpora in one of form areas.

Translation Memory Systems

Translation reminiscence can be regarded as a very specialized form of parallel corpus in which parallel texts in a database are saved and then an expression (a word, phrase or sentence) can be retrieved with its previous translation equivalents when it desires to be translated again. Translation memory systems are quintessential gear for translators in cutting-edge years. Compiling large equal corpora is the first step towards making such advisable systems. Sentences and their translations can be besides difficulty extracted from the corpus and put into the translation memory. Of course, parallel corpora in particular compiled for effective domains are of more aid to be used as translation memory. But it is now not handy to create such domain-specific bilingual corpora of very massive sizes. The future intention is to assemble general translation reminiscence software using the present English-Uzbek parallel corpus. In fact, it is a sort of parallel concordance from which translators can mechanically retrieve fragments of previous translations that totally or partly healthy a brand new phase to be translated.

In addition utility of parallel corpora is to act as main database for statistical laptop translation systems. In fact, the teaching corpus of all statistical computer translation systems is a bilingual sentence-aligned corpus (Brown at al., 1993). Brown et al have produced a probabilistic computing device translation device trained on an aligned French-English corpus. Their computer tries to locate the most

in all likelihood translation sentence in the goal language given a sentence in the source language the use of a trigram language mannequin particularly based totally on three-word sequences and a translation mannequin derived from the word-level alignment of their English and French parallel corpora. That is, their system depends on ample volume of parallel corpora to educate the translation model (Brown at al., 1991). The corpus can be used for discovering statistics on terminology, phrasing and collocations. Parallel corpora can have a sizeable contribution to the field of lexicography by means of extracting and inspecting empirical facts from corpora. This consists of improving, updating and classifying phrases or expressions as properly as gathering extra data such as well-known collocations, sub-categorization requirements, or definitions. The value and advantages of corpus studies in dictionary compiling can be located in the considerable wide variety of dictionary publishers investing in corpus science and it is reachable to find out many corpus-based monolingual as properly as bilingual dictionaries. Bilingual lexicographers can also discover the practicable equivalent(s) of the expressions (e.g. slang, neologisms, etc.) that are now not available in the existing bilingual dictionaries. There are a large range of studies discussing the approaches in which lexicographical problems may also additionally be except subject solved the use of bilingual parallel corpora. The primary attempt is to show the effectiveness of a bilingual parallel corpus in finding the most special and complete fluctuate of equivalents for dictionary entries. In most cases, parallel corpora can furnish archives that bilingual dictionaries do not usually include (Mosavi Miangah, 2007).

A parallel aligned corpus can be retrieved for multiword devices making less difficult to build a multilingual dictionary or to construct terminology databases the use of a specialized corpus. Now, availability of unique corpora has eased the task of lexicographers to a high degree in that they can get entry to to all the examples of the utilization of a phrase or phrase out of corpora in a few seconds. Consequently, compiling and revising

dictionaries would be greater specific and faster.

Cross-language information retrieval can be defined as supplying a question in one language and searching file collections in one or higher special languages. Traditional tactics for cross-language records retrieval were making use of bilingual dictionaries (dictionary-based methods). But the experiments show off that the use of bilingual corpora for computerized extraction of term equivalences would result in lots higher overall overall performance (Savoy, J. 2003). Constructing a bilingual lexicon from a bilingual parallel corpus which can be directly applied to cross-language records retrieval duties can be considered as some other software program of such corpora in this admire (Mosavi Miangah, 2008). There are more than a few situations in which the customers desire to have a dependable retrieval system. Some customers would possibly also grant a query in one language and desire their retrieved documents to be written in different languages. Some might additionally want information written in a quantity of languages. In this case they typically can examine documents in distinctive languages however cannot formulate a query in these languages. (Oard and Resnik, 1999). So, there arises a serious want to have a confident parallel corpus to be used for prompting a dependable cross-language information retrieval.

Teachers also can reap from the parallel corpora in many ways. Using the examples taken from a vary of corpora, a trainer can make excellent exercises which can be used in the classroom. These exercises can be in the form of cloze tests, regular collocations, a issue of grammar, or ideal vocabulary utilization or greater integral discovering the most brilliant equal for certain phrases or collocations. Language instructors can moreover make use of concordancers to help their college students recognize phrase associations.

Conclusion

In recent years technological advances have facilitated setting up and exploiting

extraordinary kinds of monolingual and bilingual corpora. Now, language researchers, teachers, students, translators as good as lexicographers have the satisfactory risk to make use of such corpora and related software in order to extract the recommendations of grammar or utilization and question some of the rules, criticize certain dictionary entries, consider texts produced with the useful resource of native and non-native audio gadget of a language in terms of appropriate role of sure lexical objects in the context, discover the most probable and top notch equivalent for collocations and the like. These corpora act as the pleasant complementary assets to textbooks, dictionaries and different materials. In this paper we tried to show the usefulness of parallel corpora and concordance software application in language learning, in particular translational problem related to positive verb phrases. However, there are nonetheless a lot of grounds to be covered via more than a few sorts of corpora for lookup to enhance the integration of corpora in extraordinary areas like lexicon, stylistics, cultural studies and the like. The presence and use of such corpora in language learning environments appears a sine qua non and desires to come to be ever greater. In the new technology of data and dialog technology (ICT) we prefer some sorts of modifications in strategies to teaching and gaining knowledge of to furnish the future generations with larger opportunities for exploiting technology in learning. Unfortunately, many language researchers which includes translators have little or no interest of corpus potentials in language learning, partly due to their limited get admission to corpora and corpus tools. Hopefully, however, many corpora can be reached freely or at decrease priced price. Teachers favor to be knowledgeable not only of linguistic corpora and their contents, but also of extra than a few corpus analysis tools (software) to grow to be expert users and be capable to statistics their students in dealing with corpus data. From what has been mentioned so far, it is concluded that it is necessary to rethink some of the methods and requirements of teaching and reading in order to enrich studying environments of the future.

References

1. Aston, G. (2000). I corpora come risorse per la traduzione e l'apprendimento. In Silvia Bernardini and Federico Zanettin (eds.) I corpora nella didattica della traduzione. Bologna: CLUEB, 21-29.
2. Barlow, M., 1995, ParaConc: a concordancer for parallel texts, Computers and texts, 10.
3. Barlow, M. (1996b). Parallel texts in language teaching. In S. Botley, J. Glass, T. McEnery, and A. Wilson (Eds.), Proceedings of teaching and language corpora 1996, Lancaster, UK: UCREL Technical Papers, 9, 45-56
4. Brown, P., Della Pietra, S., Della Pietra, V., and Mercer, R. (1993), "The mathematics of machine translation: parameter estimation," Computational Linguistics, 263-312.
5. Collins Cobuild English Language Dictionary (1987). ed. John Sinclair, London and Glasgow: Collins.
6. Crystal, D. (1991). A Dictionary of Linguistics and Phonetics. (third ed.), Blackwell, London.
7. Gabrielatos, C. (2005). Corpora and Language Teaching: Just a fling or wedding bells? TESL-EJ, 8(4), A1.
8. Holmes, J. (1988). Doubt and certainty in ESL textbooks. Applied Linguistics, 9, 21-44.
9. Hunston, S. (2002). Corpora in applied linguistics, Cambridge: Cambridge University Press.
10. Johns, T. (1988): Whence and Whither Classroom Concordancing?, E. van Els et al. (Eds), Computer Applications in Language Learning (Foris).
11. Johns, T., (1991). Should you be persuaded: two examples of data-driven learning, in T. Johns and P. King (eds.), Classroom concordancing, (ELR journal, 4), Centre for English language studies, Birmingham,1- 16.

12. Kennedy, G. (1987a). Expressing temporal frequency in academic English, TESOL Quarterly, 21, 69-86.
13. Stubbs, M. (2001) Text, corpora, and problems of interpretation: A response to Widdowson Applied Linguistics 22 (2), 149-172
14. Tognini-Bonelli, E. (2001). Corpus linguistics at work. Amsterdam: John Benjamins. [31]
15. Zanettin, F. (1998). Bilingual comparable corpora and the training of translators [1], Meta, XLIII, 4