

Corpus Linguistics: A Panacea for Just in Time Learning

OPEN ACCESS

Volume: 12

Special Issue: 1

Month: December

Year: 2023

P-ISSN: 2320-2645

E-ISSN: 2582-3531

Received: 24.10.2023

Accepted: 05.12.2023

Published: 14.12.2023

Citation:

Bhavana, AR. "Corpus Linguistics: A Panacea for Just in Time Learning." *Shanlax International Journal of English*, vol. 12, no. S1, 2023, pp. 266–67.

DOI:

<https://doi.org/10.34293/rtdh.v12iS1-Dec.53>

Dr. A.R. Bhavana

*Professor & Head, Department of Educational Technology
Bharathiar University, Coimbatore, India*

“One language sets you in a corridor for life. Two languages open every door along the way.” Frank Smith

In the present scenario there is a necessity for youth to hone their language skills both at the time of graduation or on the job training. Due to the exponential growth in digital technology, and Artificial Intelligence corpuses can be designed for just in time learning as per the demand. In the domain of scientific study of languages “corpus is a collection of text available / produced by users of the language to study how words, phrases are used in a particular language.

So Linguists, lexicographers, social scientists, anthropologists, and experts in natural language processing use corpus tailored for several research domains. In the domain of language learning for a specific purpose within a short time this is inevitable due to globalization, the COVID pandemic and exponential growth of AI applications. There is a necessity for people to develop certain skills /languages within a short timespan or ‘just in time’. The aftermath of the pandemic has left the work at home corridors free and the job market is open to those who can go along with the tide. This necessitates the development of a corpus which is either monolingual or bilingual and can be used for learning certain skills that require language learning for a specific purpose. Tools like Smart CAT (Smartcat Translation Tool – Sophisticated Language Tool are used for translation and these Computer Assisted Translation (CAT) tools have unique features like AI and Generative Pre-Trained Transformers (GPT) for accurate and efficient localization. These tools are corpus based to pick up marketing skills, conversation skill for a specific purpose like call centers, answering across online clients etc. Electronically processed corpora provide fast search options. In addition, text processing procedures such as word-sense disambiguation enrich corpus texts with detailed linguistic information. This enables to narrow the search to a particular parts of speech, word sequences or a specific part of the corpus. The pandemic has catalysed the entire populace and has percolated the social prism leading to exponential growth of online tools, apps, interfaces, using technologies like AI, Augmented Realty, and Virtual Reality. Entertainment applications like Netflix are creating content tailored for different regions to improve their business in the media domain. However there are some instances

where there is miss match of the lexical item and semantic meaning. So the corpora has to be designed with a team of subject experts and linguists.

The learner populace / employees are in need of a perfect corpus based solution to manage translation projects and collaborate with teams across organizations by providing an advanced translation management system for localization, and learning with a simple intuitive corpus based interface that anyone can use for just in time learning. A Data Driven Corpus can do the work of localization with various specialized software tools that help software developers, project managers, and translators to localize software, applications, documents, target markets and other content. With a localization tool, the translation quality can be increased and the consistency is maintained.

A customized corpus can offer just in time learning for translation of a text from one language into another language, and transcribing option to produce a copy of the source text with correct grammar and punctuation. This type of corpus can benefit language learners around the globe to communicate in the target language.

While training learners/ trainees/ employees, a parallel corpus comprising of two or more monolingual corpora translations can be aligned with their segments, like sentences or paragraphs, that are to be matched. The user can then search for examples of a word or phrase in one language and the results will be displayed together with the corresponding sentences in the other language. The user can then observe how the search word or phrase is translated. This kind of corpora is a boon for people who have to produce reports, presentations, narratives and communicate verbally for educational or economic purposes.

A specialized corpus can be built to study the diachronic changes in society, historical , anthropological, epigraphic content documentations etc. This kind of diachronic corpus can be used for recreating the cultural heritage of the universe. This kind of corpora can be used to create a knowledge bank /repositories according to regional , world history literary texts, news, legal aspects, etc. E-schemas are semantic memory structures that help people organize new information they encounter.

Conclusion

Thus corporuses based upon a Data Driven Learning(DDL) Approach is essential because there is a steady growing demand for people with ‘e-skills’ for business of all sizes and sectors. In addition they may help a person reconstruct bits and pieces of knowledge to overcome on the job challenges. Therefore, a strong knowledge based corpora is essential to bridge the knowledge gap using just in time learning. Therefore multiple corpora can be developed for use in the classroom, and on the job training because in the present scenario, education and training is a lifelong process.

References

1. Biber, D., & Conrad, S. *Lexical Bundles in Conversation and Academic Prose. Out of Corpora*, John Benjamins, 1999.
2. Nattinger, J. R., & DeCarrico, J. *Lexical Phrases and Language Teaching*. OUP, 1992.
3. Hunston, S., & Francis, G. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. *John Benjamins Publishing*, 2000.
4. Danielsson, P., & Utka, A. *Academic research and standards: A discussion on standards for multi-lingual language resources*.
5. Yunus, K., & Awab, S. The effects of the use of module-based concordance materials and Data-driven learning (DDL) approach in enhancing the knowledge of collocations of prepositions among Malaysian undergraduate law students. *International Journal of Learning*, 18(9), 2012.
6. Danielsson & Mahlberg. *Corpus based Approach to ELT*, 2003.