

Corpus-based Translation of Scientific and Technical Terms: A Case Study of "人工智能"

Jidong Gao¹, Xianghong Zeng²

(College of Foreign Languages/ University of Shanghai for Science and Technology, China)

ABSTRACT : With the rapid development of contemporary society, technology plays an increasingly important role in people's daily lives. Today, as technological globalization continues to intensify, scientific and technological cooperation among nations has grown closer. Consequently, the accurate translation of scientific and technological terms has become crucial for interdisciplinary communication. Using the term "Artificial Intelligence" as a case study, this paper employs methods such as frequency statistics and collocational analysis based on the COCA corpus to illustrate the varying levels of acceptance of different English translations of the term in the English-speaking world. The aim is to provide a reference for future scholars in developing translation strategies for scientific and technological terminology.

KEYWORDS - Corpus; Scientific and Technical Terminology; Artificial Intelligence

I. INTRODUCTION

In the current era of technological revolution and industrial transformation, artificial intelligence (AI) is demonstrating increasingly extensive applications and tremendous potential for development. Various AI technologies are not only serving as core driving forces in cutting-edge fields such as intelligent manufacturing and autonomous driving, but also as fundamental tools penetrating into all aspects of social life, including medical diagnosis, financial services, education, and entertainment. This has led to significant global attention on the research and development investment and application exploration of AI technologies. AI technology is highly complex and strategic, with extremely high research and development costs. It plays an indispensable and profound role in a country's industrial upgrading, economic growth, and even social transformation. Against the backdrop of the in-depth development of globalized science and technology, the English of artificial intelligence and its translation work are increasingly demonstrating their critical value. As global exchanges and cooperation in the AI industry continue to expand, China's technology enterprises and researchers must be more proactive in participating in the formulation of international technical standards, academic discussions, and commercial competition if they want to gain greater discourse power and development space on the international stage. Under such circumstances, the translation quality of AI professional terms directly affects the introduction, absorption, and re-creation of science and technology. Therefore, this article selects the well-known and frequently used term

¹ Jidong Gao, Graduate Student of Foreign Language School, University of Shanghai for Science and Technology; Research field: translation and technology.

² Xianghong Zeng, Corresponding author, D.A., Associate Professor of Foreign Language School, University of Shanghai for Science and Technology; Research field: translation and culture.

"人工智能" in the field of science and technology, and analyzes the usage frequency and context of its English translations "AI" and "Artificial Intelligence" to explore the different choices and the stylistic preferences and communication strategies they reflect.

II. BRIEF CONCEPT DESCRIPTION

The development laws of language can be gradually reflected through language research, thereby deepening our understanding of the language system (Huang & Ji, 2026). The development of the corpus originated in the 1950s. With the rise of computer network technology, the corpus was effectively combined with computer network technology. In the 1990s, the application of corpora and corpus methods saw a leapfrog development and was widely used in research in linguistics, foreign language teaching, dictionary compilation, translation, etc. The corpus is a language research method. Its research foundation is a corpus composed of real language data, and it adopts a research method combining qualitative and quantitative approaches. This major breakthrough in linguistic research has profoundly influenced translation research. Domestic research on translation based on corpora mainly consists of theoretical introduction articles, most of which are at the stage of introducing theories and making evaluations, and there is relatively little analysis and research based on examples. Furthermore, there are numerous synonyms in English, and distinguishing these synonyms has always been a major challenge in English learning. The involvement of the corpus provides a new approach for identifying synonyms (Lin & Xu, 2023). Therefore, empirical research on the translation of terms in translations using corpora and corpus methods can effectively improve the quantity, speed and quality of translation, avoid subjective speculation, and make up for the shortcomings of traditional qualitative research lacking data support. This is a new paradigm in translation research and is worth trying.

The advantages of using the corpus in translation practice are mainly reflected in the following aspects: First, over the past few decades, the corpus has continuously developed and has become a large-scale comprehensive database with a large volume of domain texts and extensive language information. Second, by integrating with computer networks, the corpus enables rapid, flexible, accurate and complex analysis work to be carried out through computers, improving the speed and quality of translation. Third, by using the corpus and combining quantitative and qualitative analysis methods, it becomes more scientific and objective, changing the traditional shortcomings of lacking data support, making the research results of translation more persuasive.

The infrastructure and scale of the Corpus of Contemporary American English have significant advantages. It was developed by the team led by Mark Davies at Brigham Young University and has now accumulated over 1.2-billion-word tokens of dynamic text data, covering five major registers and sampled evenly at a 20% ratio, ensuring the representativeness of the samples and cross-genre comparability (Lin & Yuan, 2025). The COCA corpus, officially known as the Corpus of Contemporary American English, consists of 1 billion words of text. The text includes various types of materials such as spoken dialogues, literary works, news reports, and academic papers, and is evenly and uniformly distributed across these fields (Chen & Lan, 2025). Generally speaking, people often analyze the translation of words by referring to dictionaries. Although dictionaries are authoritative, identifying synonyms solely through dictionaries has limitations (Zhou, 2025). A large number of articles on corpus research have proved that the emergence of the corpus provides new options for the analysis of different versions of translation, which greatly improves the accurate use of near-synonyms in English (Qin & Zhang, 2023). In addition to its powerful text search function, COCA is also finely divided according to the year of the corpus. All these features make COCA an ideal corpus for researchers, teachers, and students.

III. STUDY DESIGN

3.1 Research corpus

The corpus selected for this study is the term "人工智能". Its abbreviation is AI, which refers to a new technological science that studies, develops, and applies theories, methods, technologies, and application systems for simulating, extending, and expanding human intelligence. Artificial intelligence is a branch of computer science, aiming to understand the essence of intelligence and produce a new intelligent machine that can respond in a manner like human intelligence. This field of research includes robotics, language recognition, image

recognition, natural language processing, and expert systems, etc. In the COCA corpus, the term "人工智能" is mainly translated as "AI" and "Artificial Intelligence". By inputting "AI" into the COCA corpus, it was found that this English translation appeared 72,872 times (Figure 1). When inputting "Artificial Intelligence" into the COCA corpus, it was found that this English translation appeared 2,237 times (Figure 2).

HELP	①	★	ALL FORMS (SAMPLE): 100 200 500	FREQ
1	①	★	AI	72240
2	①	★	A.I.	577
3	①	★	A.I	55
TOTAL				72872

0.281 seconds

Figure 1: Frequency of "AI" appearance

HELP	①	★	ALL FORMS (SAMPLE): 100 200 500	FREQ
1	①	★	ARTIFICIAL INTELLIGENCE	2237

0.219 seconds

Figure 2: Frequency of "Artificial Intelligence" appearance

3.2 Contextual Analysis

By entering the word "AI" into the COCA corpus and clicking on the "CONTEXT" page, based on the analysis results of this corpus, the usage characteristics of the word "AI" can be interpreted from two aspects: in terms of context distribution and time trend. It is found that the usage scenarios of "AI" vary greatly. The high-frequency scenarios shown in the table are TV/M and FIC. Among them, in the TV/M scenario, the frequency of using "AI" is as high as 41,054, which is the highest among all scenarios; PER MIL (per million-word occurrences) is 320.55. In the FIC context, its PER MIL reaches 104.59, second only to TV/M. However, the scenarios where the usage frequency of "AI" is low are ACAD and NEWS: PER MIL is only 12.93 and 13.85, far below the overall average (73.32). In addition, the PER MIL of "AI" in MAG is 26.58, also at a relatively low level.

By analyzing the usage frequency of the term "AI" from the perspective of its temporal distribution trend, it was found that the overall trend shows a fluctuation pattern of "initially high, then decreasing, and then rising again". During the period from 1990 to 2004, the usage frequency of "AI" was at a relatively high level, with PER MIL ranging from 85 to 101, and FREQ remaining above 10,000. From 2005 to 2014, the usage frequency of "AI" continued to decline, with PER MIL dropping from 79.62 to 69.10, and FREQ also decreased simultaneously. From 2015 to 2019, the usage frequency of "AI" slightly increased, with PER MIL at 92.10 and FREQ at 1,130.

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	72810	4146	4934	41054	3714	12375	3352	1686	1549	12234	11210	10661	9797	8523	11305
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	121.1	125.2	124.6	123.1	123.3	122.8
PER MIL	73.32	32.24	39.71	320.55	29.44	104.59	26.58	13.85	12.93	101.02	89.53	85.54	79.62	69.10	92.10

SEE ALL SUB-SECTIONS AT ONCE

Figure 3: The Usage of "AI" in Different Contexts

NO.	YEAR	CONTEXT	TEXT
1	2012	BLOG	...tionalgeographic.com
2	2012	BLOG	scifi4me.com
3	2012	BLOG	...esternjournalism.com
4	2012	BLOG	blog.hemmings.com
5	2012	BLOG	blog-ah.typepad.com
6	2012	BLOG	100rsns.blogspot.com
7	2012	BLOG	kateharding.net
8	2012	BLOG	kow626.com
9	2012	BLOG	dailynos.com
10	2012	BLOG	dailynos.com
11	2012	BLOG	...indcitygridiron.com
12	2012	BLOG	thepewterplank.com
13	2012	BLOG	thepewterplank.com
14	2012	BLOG	blog.chasejarvis.com
15	2012	BLOG	...swolfe.wordpress.com
16	2012	BLOG	...swolfe.wordpress.com
17	2012	BLOG	vg247.com
18	2012	BLOG	lesswrong.com
19	2012	BLOG	sarahbessey.com
20	2012	BLOG	dailynos.com
21	2012	BLOG	dailynos.com
22	2012	BLOG	dailynos.com
23	2012	BLOG	dailynos.com
24	2012	BLOG	dailynos.com

Figure 4: Examples of "AI" in Different Contexts

By entering the word "Artificial Intelligence" into the COCA corpus and clicking on the "CONTEXT" page, based on the analysis results of this corpus, the usage characteristics of "artificial intelligence" can be interpreted from two aspects: context distribution and time trend:

In terms of context distribution, the usage of the word "Artificial Intelligence" varies significantly in different scenarios. The most frequently used scenario is in MAG. Its PER MIL (per million-word occurrences) is as high as 5.68, the highest among all contexts; at the same time, its FREQ (frequency) is 716, the most among all sub-scenarios - indicating that magazines are the most common carriers of the word "Artificial Intelligence". Meanwhile, in ACAD, its PER MIL is 3.71, second only to MAG; combined with the professionalism of academic texts, it also conforms to the high-frequency discussion of "artificial intelligence" in the academic field. However, the less frequently used scenario is in BLOG, where its PER MIL is only 1.10, the lowest among all contexts; its FREQ is only 142. Other scenarios, such as TV/M, SPOK, etc., have PER MILs below 1.5, indicating that the usage frequency of "artificial intelligence" is relatively low in non-written and non-professional scenarios.

In terms of time distribution, the usage of the word "Artificial Intelligence" also varies significantly in different time periods. Through the analysis of COCA charts, it is known that the usage frequency of the word "Artificial Intelligence" increased explosively from 2015 to 2019. Before 2015: the FREQ (frequency) in each time period was at a relatively low level (1990-94 was 150, 2010-14 was 221), and PER MIL fluctuated basically between 1.2-1.8, indicating that "artificial intelligence" was not a high-frequency word before this period. From 2015 to 2019, the usage frequency of the word "Artificial Intelligence" directly jumped to 1006 (about 4.5 times that of 2010-14); its PER MIL reached 8.20 (about 4.6 times the previous highest value).

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIG	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	2237	142	259	158	168	161	716	189	444	150	147	177	135	221	1006
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	121.1	125.2	124.6	123.1	123.3	122.8
PER MIL	2.25	1.10	2.08	1.23	1.33	1.36	5.68	1.55	3.71	1.24	1.17	1.42	1.10	1.79	8.20

Figure 5: The Usage of "Artificial Intelligence" in Different Contexts

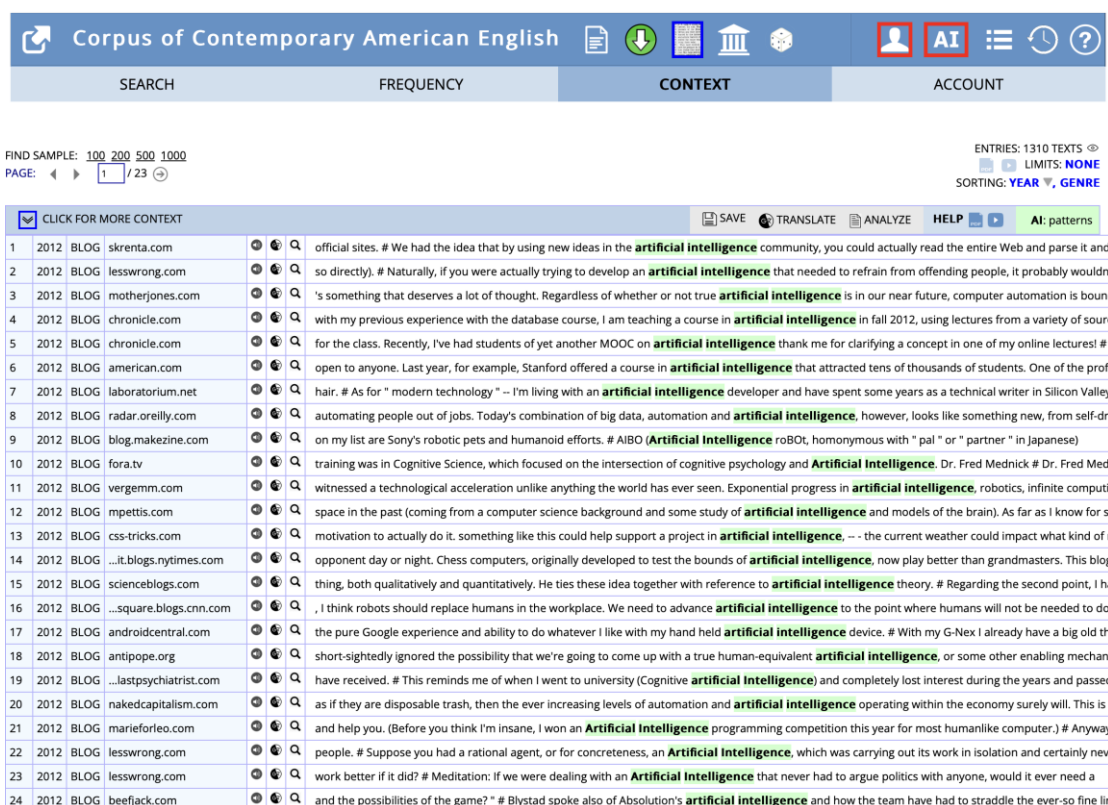


Figure 6: Examples of "Artificial Intelligence" in Different Contexts

3.3 Research Finding

Through the above analysis, it was found that the usage frequency of "AI" and "Artificial Intelligence" varies significantly in different scenarios and at different time periods. The data analysis shows that the usage of "AI" and "Artificial Intelligence" presents a "dual track" phenomenon: one leans towards the general public, while the other focuses on professional and academic discussions. In different contexts, "AI" exhibits a distinct tendency towards popularization. The word appears most frequently in TV/M programs (320.55 times per million words) and reaches 104.59 times per million words in fiction (FIC) scenes. This indicates that "AI" is mainly used by the public. In contrast, the usage frequency of "AI" in serious contexts such as academic (ACAD) and news (NEWS) is significantly lower, at 12.93 and 13.85 times per million words respectively, far below the overall average of 73.32. In sharp contrast, "Artificial Intelligence" shows a clear professional characteristic. The word has the highest usage frequency in magazines (MAG), at 5.68 times per million words, and also reaches 3.71 times per million words in academic texts (ACAD). This indicates that it is more academic-oriented. In non-formal scenarios such as blogs (BLOG), TV/M programs (TV/M), and spoken language (SPOK), its usage frequency is at a relatively low level.

From the perspective of historical development, the usage frequencies of the two words have shown different trajectories. The usage of the word "AI" has undergone a distinct fluctuating process: it remained at a relatively high level (85-101 times per million words) from 1990 to 2004; it continued to decline to 69.10 times per million words from 2005 to 2014; and it rebounded to 92.10 times per million words from 2015 to 2019. This fluctuation reflects the ups and downs in public attention to artificial intelligence technology. While "Artificial Intelligence" shows a completely different development trend. Before 2015, its usage frequency remained at a relatively low level (1.2-1.8 times per million words); but from 2015 to 2019, it experienced an explosive growth, with a usage frequency of 1006 times, reaching 8.20 times per million words, which was 4.5 times and 4.6 times higher than the previous periods.

IV. CONCLUSION AND SIGNIFICANCE

This article first introduces the development, changes and significance of technical terms, as well as the COCA corpus. Taking the well-known term "人工智能" as an example, it conducts a study on the translation of related words and contextual collocations using the COCA corpus, analyzes the usage of the two English translations "AI" and "Artificial Intelligence" in different contexts and at different time periods, and briefly discusses the underlying reasons. These findings provide important references for translation practice. When choosing translation words, translators should fully consider the text type and target readers: for a general audience or involving cultural and entertainment content, it is advisable to use the more popular "AI"; while in academic papers, professional reports, etc., which require a sense of rigor, the complete "Artificial Intelligence" should be used. This distinction in usage not only conforms to the language habits in the English-speaking world but also enables a more accurate conveyance of the stylistic features and professional level of the original text.

In conclusion, the translation of technical terms is of great significance. To ensure the accuracy and comprehensiveness of term translation in professional fields, translators should master professional terms, while also preparing theoretical knowledge, cultivating good habits. They should also have context awareness and reader awareness. As translators, they should constantly learn various translation knowledge and translation tools and explore various fields to improve their translation skills.

REFERENCES

- [1] Y.T. Lin and Y. Sen. Comparative Study of Synonyms Based on the COCA Corpus: A Case Study of "Outcome" and "Consequence". *China Terminology*, 28(1),2026,180-183.
- [2] W. Huang and T.H. Ji. An Investigation of the Evolution of the Syntactic Functions of Chinese Adjectives Based on a Diachronic Corpus and Word Vector. *Contemporary Linguistics*, 28(1),2026,80-97.
- [3] Y.F. Chen and C.S. Lan. A Contrastive Study of Obvious Near-synonyms Based on the COCA Corpus: Take Apparent and Evident for Example. *Journal of Quanzhou Normal University*, 43(4),2025,58-63.
- [4] M.Y. Zhou. A Study of Synonym Discrimination Based on COCA Corpus: A Case Study of Award and Reward. *Journal of International Social Science*, 2(4),2025,133-139.
- [5] M.Y. Qin and X. Zhang. Comparative analysis of English synonyms based on COCA corpus: Taking "value, worth, merit" as an example. *International Journal of New Developments in Education*, 5(22),2023,36-40.
- [6] H. Lin and S.J. Xu. A Study of Synonyms Based on COCA Corpus Road and Street as Examples. *Lecture Notes on Language and Literature*, 6(3),2023,27-32.