The Application of Mandarin Industry Vocabulary Corpus Pages to Improve Translators Competence in the Industrial World

Ruth Kiana Nuratri[1*], Dani Putri Septi Kusumaningtyas[2], Kristina Indah Setyo Rahayu[3]
Clarosa Amanda Hasel[4], Juairiah Nastiti Sandyaningrum[5]
[1*] Universitas Sebelas Maret, Surakarta, Indonesia (ruthkiananuratri@staff.uns.ac.id)
[2] Universitas Sebelas Maret, Surakarta, Indonesia (putry_fang89@staff.uns.ac.id)
[3] Universitas Sebelas Maret, Surakarta, Indonesia (clarosahasel@staff.uns.ac.id)
[4] Universitas Sebelas Maret, Surakarta, Indonesia (kristinaisr@staff.uns.ac.id)
[5] Universitas Sebelas Maret, Surakarta, Indonesia (jnsnastiti6@staff.uns.ac.id)

Keywords:
Mandarin Language Corpus,
Industrial Vocabulary,
Translation,
Technical Terminology,
Digital Corpus Technology

ABSTRACT

The Mandarin language corpus has become an essential resource for enhancing translator competence, particularly amid the growing industrial cooperation between Indonesia and China. The availability of technical vocabulary sources remains limited, creating a gap between academic proficiency and the demands of professional translation in industry. This community service program aims to develop an online corpus of Mandarin industrial vocabulary as an effort to provide standardized terminology resources based on authentic data. The implementation methods include mapping user needs, collecting vocabulary from 30 undergraduate theses and student internship reports from 2021–2024, developing the digital platform, and conducting training on its use. A total of 4,221 vocabulary items were compiled and integrated into the corpus, equipped with search features based on hanzi, pinyin, and definitions. The results show that the corpus platform improves the accuracy of technical document translation, accelerates terminology retrieval, and strengthens both thematic and instrumental competencies of translators. This digital corpus has proven effective in supporting translation needs in academic and industrial settings.

INTRODUCTION

The ever-increasing multilateral partnership between Indonesia & China over the past decade has driven a growing demand for professional translators with a strong grasp of cross-sector industrial Mandarin vocabulary. Various studies have shown that translator competence is strongly influenced by the availability of relevant terminology sources based on authentic data (Liu & Wang, 2020). However, references related to industrial Mandarin vocabulary remain limited, thus it creates a concerning gap between students' academic abilities and the demands of a more advanced or professional translation in the workplace (Zhang, 2019).

Numerous efforts to improve translator competency through conventional learning, the use of dictionaries, or automated translation applications have not yet fully met the industry's needs for contextually accurate and consistent terminology (Chen & Huang, 2021). A situational analysis regulated along with the industry partners revealed that the low accuracy of technical document translation is largely due to a limited access to standardized terminology and the lack of a corpus system that records terminology utilization in the field (Sun, 2022). This circumstance not only slows down translators' work processes but also increases the risk of terminological errors in technical documents.

Industrial documents are an authentic source with great potential to produce an accurate terminology corpus. Therefore, this community service collected vocabulary data from 30 final projects as well as student internship reports throughout 2021–2024 period, focusing on the development of an industrial glossary. In its implementation, the

program also involved industry partners through data collection and interviews with eight companies: PT. Donlim, PT. Makuku Care, PT. Jiale Indonesia Textile, PT. Formosa Bag, PT. Young Tree Industries, PT. Matahari Tire, PT. LBM Energi Baru, and PT. Borine Technology. This authentic document-based approach has proven to be effective, as demonstrated in previous research on corpus development in the energy and manufacturing sectors (Huang & Li, 2020).

Further discussions done with industry partners indicate that a digital platform which allows for a fast and precise term search engine is essential to support daily technical translation work (Wang, 2021). Intrigued by this topic, the Diploma degree of Mandarin Language Study Program at UNS Vocational School implemented a Community Service Program through a Research Group Grant (PKM HGR UNS) for the 2025 Fiscal Year with Contract Number: 370/UN27.22/PT.01.03/2025 to develop a Mandarin industrial vocabulary corpus page yinnizhong-hyk.com, which is based on authentic academic and industrial data.

Digital corpus technology is known to increase term search efficiency up to four times faster than manual/traditional search in printed dictionaries (Li & Xu, 2018), making its implementation highly relevant to learning needs and professional practices. In addition to developing the corpus page, this activity also includes training and mentoring for students along with novice translators so they can optimally utilize the corpus features. This training approach has been proven to improve instrumental and thematic competencies in translation activities (Gao, 2022). Therefore, developing a Mandarin industry corpus website is a promising strategic step to bridge the gap between theory and practice, whilst also supporting & complementing the availability of skilled translators ready to meet industry needs.

METHOD

The community service team's method is divided into six stages: preparation, initial corpus, verification, website development, training, simulation, and dissemination.
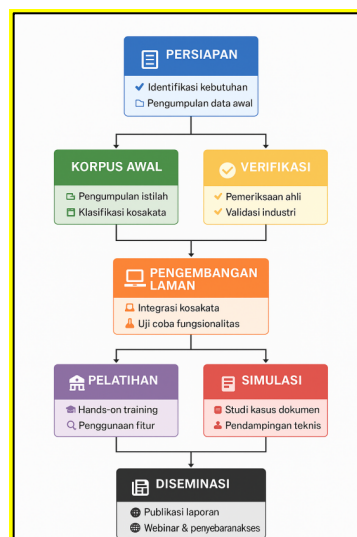

Figure 1. Community Service Activity Implementation Scheme

The followings describe each of the method:

1) **Preparation Phase**

A cooperation was conducted with eight industry partners to ensure the availability of technical documents and conduct interviews as sources for the corpus data. The team also developed envenomization guidelines and corpus material utilization in translation.

2) **Initial Corpus Stage**

The vocabulary data are collected from academic documents and authentic documents from industry partners, then categorized and compiled into an initial corpus draft adjusted to the needs of the Indonesian–Chinese industry.

3) **Verification Stage**

The lists of vocabularies & technical terms collected from the earlier stage are then verified by lecturers, translation practitioners, and industry professionals to ensure its accuracy, contextual appropriateness, and consistency of terminology. Revisions will be conducted based on partner feedbacks.

4) **Page Development Stage**

The verified corpus will be integrated into the yinnizhong-hyk.com website through database structuring, search feature development, as well as functional and user experience testing.

5) **Training Phase**

The training includes an explanation of the concept of digital corpus, its benefits in technical translation, and a demonstration of how to use the website, focusing on translation issues faced by partners.

6) **Simulation Phase**

Participants then translate technical documents from partners using the corpus page as a form of usability evaluation and basis for feature improvements.

7) **Dissemination Stage**

Activities are then documented through videos and mass media publications, complemented by comprehensive evaluations and efforts to build sustainable collaboration with industry partners.

RESULTS AND DISCUSSION

This Community Service Program focuses on developing a corpus of Mandarin industrial vocabulary and its application through the digital corpus website yinnizhong-hyk.com. This effort aims to improve the competence of Mandarin translators in the industrial world and helps partners in addressing the challenges of technical communication in Mandarin. This activity is conducted properly throughout the six-stage of methodology.

Preparation Phase

The practice began with intensive coordination with eight industrial partners: PT. Donlim, PT. Makuku Care, PT. Jiale Indonesia Textile, PT. Formosa Bag, PT. Young Tree Industries, PT. Matahari Tire, PT. LBM Energi Baru, and PT. Borine Technology. This coordination aimed to identify Mandarin industrial vocabulary needs, the availability of company documents, and interview opportunities with operational staffs who routinely interact with specialized Mandarin terms in various industrial fields. The community service team also developed a booklet introducing digital corpus technology, guidelines for cataloging specialized terms in various industrial fields, and training materials.

Figure 2. Preparation stage activities, (a) A visit to PT. LBM, (b) A visit to PT. Young Tree Industries, (c) A visit to PT. Matahari Tire, (d) A visit to PT. Jiale Indonesia Textile, (e) A visit to PT. Borine Technology, (f) A visit to PT. Formosa Bag, (g) A visit to PT. Makuku Care, (h) A visit to PT. Donlim

Initial Corpus Stage

Once all partners were ready to collaborate, data collection was initiated throughout two primary sources: academic documents, such as internship reports and student final projects related to Mandarin vocabulary in various industrial fields; and authentic company documents, such as machine SOPs, work safety guidelines, product specifications, quality control reports, and daily worker communication documents. The term data was then classified by industry sectors, such as textile manufacturing, household electronics, chemicals, etc., as well as general industrial communication terms. This stage resulted in an initial draft corpus which contains terms that reflect the industry's actual needs at the operational level.
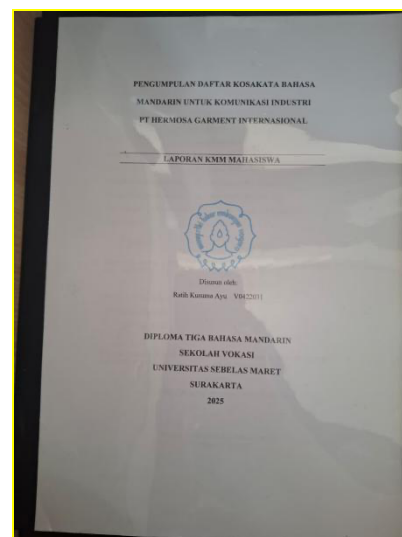

Figure 3. Examples Of Student Internship Reports & Final Assignments Used as Vocabulary Data

Verification Stage

Verification was conducted by expert Mandarin Chinese lecturers, professional translators, and technical representatives from partner companies. Each term was checked for accuracy of Mandarin-Indonesian terminology, consistency of term usage in industrial contexts, relevance of meaning to machine functions along with production stages, and appropriateness of *pinyin* and Chinese characters.

This verification process resulted in several revisions to terms previously used only in general academic settings. With input from eight industry partners, the team identified industry-specific terms that are not found in standard dictionaries to ensure that the corpus truly applied on field practices. Once the correction process had been completed, the corpus then was ready to be integrated into the digital platform.
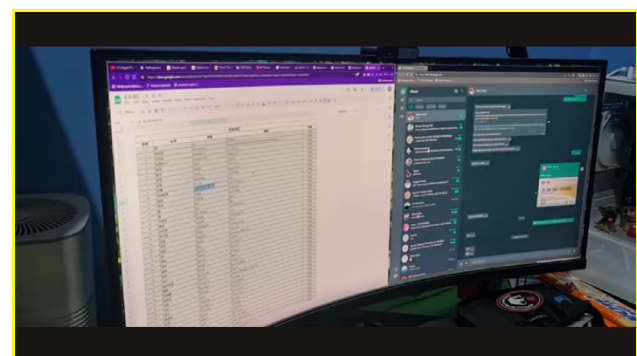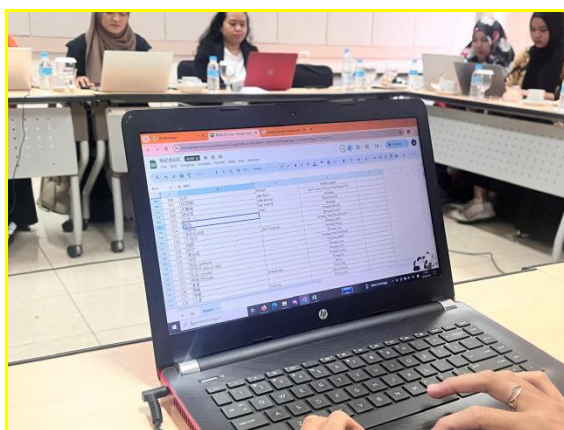

Figure 4. The Community Service Team (Consist of Lecturers and Students) Created A Draft of Mandarin Industrial Vocabulary

Page Development Phase

On this phase we focused on integrating the corpus into the yinnizhong-hyk.com (digital website). Activities included are structuring the database, developing a three-way search feature (Chinese, *pinyin*, and meaning), and testing the page's functionality to ensure a user-friendly interface.

The team also conducted user experience trials involving students and industry representatives to determine whether the search feature was fast, responsive, and met their technical translation needs. The trial results demonstrated that vocabulary search could ease or even speed up translation assignment completion and technical discussions between Indonesian and Chinese workers.
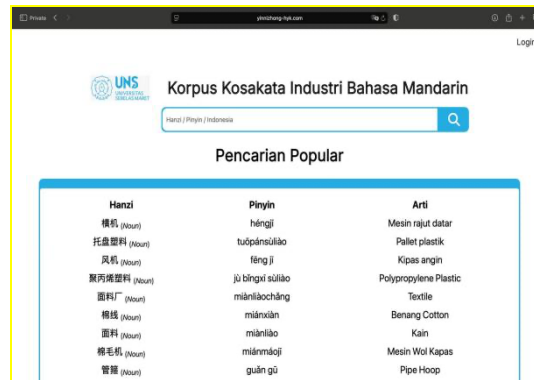
Figure 5. Team Inputted Data on The Corpus Page (Mandarin Industrial Vocabulary Corpus Page)

Training Phase

The training was conducted on campus and was attended by students, lecturers, and staffs from industry partners. The session began with an explanation of the basic concepts of digital corpus, how terminology databases work, and their role in improving industrial Mandarin translation competency.

The next session included a demonstration of how to use the website to search for terminology equivalents, meaning evaluations, and the vocabulary utilization in industrial communications. Participants were also given opportunities to ask questions about translation issues they frequently encounter in their daily work. The training demonstrated how digital tools such as corpus can reduce terminology misinterpretation and improve work efficiency.

Figure 6. Training stage for corpus utilization attended by students, lecturers, and staffs from industrial partners

Simulation Phase

In order to properly test the effectiveness of the website, a translation simulation was conducted using technical documents from partner companies. Participants were asked to complete translation tasks using the corpus website as the primary reference. This simulation served as a direct evaluation of: the terminology accuracy, the completeness of technical vocabularies, the consistency of terminology used in the document, and the participants' speed in understanding Mandarin terminology.

The simulation results showed improved translation accuracy and speed, especially for technical terms that were previously considered difficult to find in general dictionaries. Furthermore, several new terms emerged during the simulation, and these new materials were added as input for further development of the corpus.

During the simulation phase, the corpus proved to be effective in addressing several crucial translation challenges in the industrial world. For example, previously translators often relied on general dictionaries or literal translations, which could potentially lead to errors and ambiguity. The advantage of this corpus lies in its ability to provide context-tested equivalents. For example, a search for the term "注塑机 (zhùsù jī)" in the corpus not only yields the general equivalent "injection molding machine" but also reveals variations in the term used across different companies. PT. Donlim refers this term specifically to "injection machine," while PT. Young Tree Industries refers it to "molding machine." This nuance is crucial for ensuring the translation to align with the company terms cataloging and avoid technical misunderstandings.

Another example is the term "刀片 (dāopiàn)." A general dictionary would translate it to the equivalent "knife" or "blade," which is too general whilst on the yinnizhong-hyk.com corpus, this term refers it to "razor" or "cutting knife." This term specification is only possible by documenting the vocabularies into a databank of authentic sentences collected from each industry sector. This corpus reduces the search time which is usually spent on multiple sources. Previously, translators had to consult multiple online dictionaries, forums, and glossaries to verify a single term. With yinnizhong-hyk.com, this process is centralized in one platform. Search features based on pinyin, Chinese characters, or even Indonesian vocabularies allow translators to instantly search for relevant vocabulary, finding precise equivalents like "胎坯 (Tāi pī)" for "*ban rawa*" in just seconds, thus significantly improves translator efficiency.

Figure 7. The Community Service Team Held Simulations with Several Industrial Partners

Dissemination Phase

Documentation of the community service activities (including the training and simulation processes) was recorded and compiled into a publication video that was uploaded into the YouTube channel. Furthermore, the results of the activities were published through the electronic media portal Soerakarta.id to announce the breakthrough to the public.

At the end of the activities, a comprehensive evaluation throughout of all stages (from preparation to dissemination) was conducted. Feedbacks from participants and partners indicated that the development and application of a digital corpus of Mandarin industrial vocabulary to improve translator competency in the industrial world had a tangible impact on reducing technical communication barriers in the workplaces. The community service team also has established ongoing collaboration with industry partners to ensure continuous updates to the corpus and the integration of new terms and vocabulary from industrial technology developments.
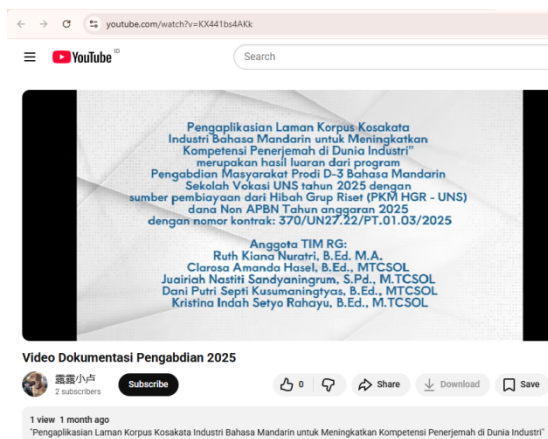

Figure 8. Publication Of Community Service Activities on Youtube Channels And Mass Media

CONCLUSION

Based on the entire COMMUNITY SERVICE activity program, it can be concluded that: The development of the yinnizhong-hyk.com corpus has been successfully carried out by utilizing authentic data from the academic world (final assignments, internship reports, industrial lecture notes) and the professional world (8 industrial partners). A total of 4,221 vocabulary words have been integrated into a functional corpus page, with key search and filter features. The yinnizhong-hyk.com corpus page has made a significant impact on the development of industrial translator competencies. The yinnizhong-hyk.com corpus page can enhance two competencies of translators, which are: thematic competency and instrumental competency.

Thematic competency improvement is achieved through a structured industrial classification system within the corpus. Translators can easily identify equivalent terms specific to each industrial sector. For example, the term "成型" (chéngxíng), which has the general meaning of "forming," can be accurately translated as "printing" in the context of the manufacturing industry. "裁断机" (cáiduàn jī) which generally means "cutting machine" can now be translated alternatively as "Rotary cutting tables". This classification allows translators to develop a deep understanding of the terminology characteristics for each industrial field, thus speeding up the translation process to any specific industries.

Strengthening Instrumental Competence reflects/relies on translators' ability to optimally utilize the corpus's technical features. The Corpus responsive search system allows quick access to 4,221 curated vocabulary words. The industry sector filter facilitates the navigation for specific industry vocabulary searches. This shift from dependency on conventional/traditional dictionaries to the use of structured digital databases shows that utilizing technological resources can be done sensibly to raise the bar of professionalism into the next level.

The yinnizhong-hyk.com corpus has evolved into more than just a digital glossary, but also a competency development platform that empowers translators to meet the demands of increasingly complex industry dynamics.

ACKNOWLEDGEMENTS

REFERENCES

Avenida, Q., dkk. Buku Saku Kosakata Mekanik Garment Bahasa Indonesia-Mandarin. 2023. Surakarta: UNS Press.

Baker, M. (1995). Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. Target, 7(2), 223-243.

Göpferich, S. (2009). Developing Translation Competence. In Behind the Mind: Methods, Models and Results in Translation Process Research (pp. 5-32). Samfundslitteratur.

Laviosa, S. (2002). Corpus-based Translation Studies: Theory, Findings, Applications. Rodopi.

Liu, Y. (2021). Application of Monolingual Corpus in Technical Translation Teaching. Journal of Language Teaching and Research, 12(5), 789-797.

McEnery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.

PACTE Group. (2018). Researching Translation Competence by PACTE Group. John Benjamins Publishing Company.

Sandyaningrum, J.N., dkk. (2024). Kosakata Bahasa Mandarin Industri Teknologi Energi. Surakarta: UNS Press.

Sandyaningrum, J.N., dkk. (2024). Kosakata Bahasa Mandarin Industri. Surakarta: UNS Press.

Sinclair, J. (2005). Corpus and Text: Basic Principles. In Developing Linguistic Corpora: A Guide to Good Practice (pp. 1-16). Oxbow Books.

Xun, E., Rao, G., Xiao, X., & Zang, J. (2016). The BCC Corpus: Its Design and Methodology. Lexicographical Studies, (4), 1-17.

Yang, Y., & Wang, X. (2020). Machine Translation in Chinese-English Business Communication. Translation and Interpreting Studies, 15(2), 245-263.

Zhong, Q., & Zhang, W. (2019). Corpus-based Study on the Translation of Chinese Economic Terms into English. International Journal of English Language Teaching, 6(2), 1-10.

https://www.bps.go.id/id/statistics-table/2/MTg0MyMy/realisasi-investasi-penanaman-modal-luar-negeri-menurut-negara.html (accessed on: Tuesday, July 15, 2025 at 22:52)