
Chenlu Yu

A GERMAN-CHINESE E-DICTIONARY OF MANUFACTURING TECHNOLOGY IN AUTOMOTIVE INDUSTRY

Entry Design

Abstract We intend to create an online dictionary for manufacturing technology in the automotive industry, which will be available in German and Chinese. The dictionary is designed to improve specialized communication, and its target users include Chinese engineers, students, and interpreters in this sector. The development of entry structures using Frame-Based Terminology (FBT) as a dynamic process-oriented approach (cf. Faber et al., 2005; Faber, 2011) will be the main topic of this paper. Specifically, we will present a corpus-based methodology for integrating contextual information, especially syntactic-semantic features, into the entries. At this point, we start with the description of predicative terms. It might be conceivable to provide a semantic-based approach that can serve as a practical solution for future bilingual dictionaries in various engineering disciplines.

Keywords specialized lexicography; bilingual dictionaries; frame-based terminology; dictionary entries

1. Introduction

In recent decades, German original equipment manufacturers and automotive suppliers have set up overseas offices and production bases in China. Statistics show that Germany is rapidly becoming a top choice for Chinese students majoring in technology and engineering. Furthermore, as China and Germany collaborate in various of industrial sectors, skilled translators with in-depth knowledge are becoming increasingly important.

In this context, a high-quality German-Chinese automotive manufacturing dictionary is extremely useful. In contrast to the traditional approach used in most bilingual specialized dictionaries for the Chinese market, which typically focuses on collecting nouns and offers a static micro-structure, our dictionary incorporates exclusive predicative terms, like specialized verbs with their participants and their contextual information, into the dictionary entries. This approach recognizes the need for effective daily communication in the automotive manufacturing industry, especially when it comes to understanding and refining German technical documentation and manuals and facilitating the acquisition of subject-specific knowledge. According to ten Hacken (2016, p. 73), there are three key questions to consider when applying theories to word meanings in bilingual dictionaries:

- a) How can a dictionary entry be interpreted?
- b) How can we determine how good a dictionary entry is?
- c) How can we compile a dictionary entry?

These three questions are fundamental to our research project. In this paper, we introduce the current lexicographic stage of the project and offer a preliminary prototype of an entry. Section 2 summarizes the project, including the target user group, the dictionary's intended function, the constitution of the corpora, and the types of headwords included. Section 3 outlines the methodology for compiling a sample entry, including constructing definition patterns, collecting contextual information, and identifying equivalents within the entry. Section 4 presents a sample entry based on the outcomes of the preceding methodological steps. Finally, Section 5 concludes with a summary and offers recommendations for future work.

2. The Outline of the Project

2.1 The User Group

Schierholz (2003; 2014) and Gouws (2020) argue that specialized dictionaries are designed as reference tools for experts, laypeople, or learners in specific fields. By including specialized words within a certain subject field, these dictionaries assist in bridging language gaps and prompting effective communication. Given the widespread accessibility of online dictionaries, the target users include three groups: (1) Chinese engineers in the automotive and related industries; (2) students majoring in automotive engineering, mechanical engineering, and related subjects; and (3) translators working in the mechanical, automotive, and other technical sectors.

2.2 Functions of the Dictionary

Most specialized dictionaries and terminology databases prefer to emphasize definitions and concept relationships rather than starting from the user's perspective. According to Bergenholtz & Tarp (2010), the primary goal of any lexicographic or terminological product should be to address the users' needs. Considering the diverse characteristics of users, our dictionary should be multi-functional, accommodating the needs of three user groups involved in text comprehension, production, and translation. For instance, engineers frequently encounter specialized communicative contexts in which they must interpret German technical drawings and documents and communicate with their German counterparts via email. In such scenarios, having access to equivalent terms and contextual information is essential for accurate comprehension and effective communication.

2.3 Corpora

We are developing two comparable corpora consisting of handbooks and manuals pertinent to the automotive sector, academic literature and books used as teaching materials, and papers written by professors and experts.

Currently, three *Handbooks for Automotive Management* published by International Automotive Groups, along with three academic books, are included in each of the German and Chinese corpora. In the future, we plan to add papers and articles to each corpus covering topics related to automotive manufacturing, such as manufacturing technology, electric automobile production, vehicle repair, etc.

To sum up, the German corpus contains 750,867 tokens, equivalent to 565,143 words, while the Chinese corpus consists of 883,141 tokens, corresponding to 734,127 words. Our goal is to reach the size of 1 million words for each corpus.

2.4 The Type of Headwords

Scholars have recently recognized that employing predicative units to represent processes, events, and attributes can improve the utility of specialized dictionaries and terminological resources. Unlike nouns that refer to entities, verbs can be analyzed in terms of their argument structures and the semantic roles of each participant. This approach allows for the construction of a conceptual paradigm that is based on both syntax and semantics (cf. L'Homme, 2012; 2015), and it has been proven that concentrating on specialized verbs can be broadly implemented in specialized dictionary projects and knowledge bases, such as those in the medical field (cf. Wandji et al., 2013) and the legal domain (cf. Pimentel, 2012). Specialized verbs, morphological nominalizations and compounds will be considered for headwords inclusion in our project. In addition to single terms, many multi-word terms paired with certain verbs through concordance and clustering could also be added into headwords.

3. Methodology

We employ a bottom-up approach, starting with the selection of specialized verbs from corpora (Section 3.1). Then we define frame categories and definition patterns to describe specialized verbs (Section 3.2), annotate syntactic-semantic information using concordance analysis (Section 3.3), and finally assign the equivalent (Section 3.4). The data collection and pre-processing are completed with Sketch Engine.

3.1 Term Selection

Before term selection, we divide the term candidates into three stages of the automotive production process: pre-production, in-production, and post-production. Equipment installation and setup, as well as material preparation, are all part of the pre-production stage, which can also occur throughout production. Therefore, sub-categories such as *manipulation* and *preparation* are essential at these stages. According to the German Industry Standard DIN 8580 (cf. Förster & Förster, 2018), the manufacturing process within the production stage is organized into six sub-categories: primary shaping, material forming, dividing, joining,

modifying material property, and coating. The post-production stage includes sub-categories such as testing and shipment, among others.

Using frequency indicators suggested by L'Homme (2015), we examine the frequency of verb occurrences in their context based on this classification to find potential term candidates. To be specific, we first go through verbs that have an absolute frequency value greater than 100 and work with industry experts to perform the preliminary filtering, producing a list of verbs for each language. Following that, each of these potential verbs is examined separately to find multi-word term candidates using clustering and concordance. This process is used for both corpora in the same way.

After completing these procedures, we chose the German predicative term *einstellen* 'to regulate' as a candidate term. As seen in the corpora, this candidate term has a frequency value of 127 in the German corpus and a frequency value of 313 in the Chinese corpus.

3.2 Frame-Based Definition Pattern

Frame Semantics (cf. Fillmore, 1982; 1985) proposes that to understand the meaning of a word, one must activate the associated frame and the comprehensive knowledge involved in the frame. Expanding upon Fillmore's notion, Faber (2011) defined frames in a broader sense to describe knowledge structures. In addition to enriching the dynamism and coherence of the terminological knowledge base, this active, process-oriented approach extends the conceptual description beyond the hierarchical structure.

Based on concordance analysis (see Figure 1), the participants of the predicative term *einstellen* 'to regulate' are classified into two types:

- (1) Concrete tools or devices: the spring, machine, gear, and valve can be adjusted, as shown in items 2, 5, 7, and 9.
- (2) Abstract units, like parameters: items 1, 3, 4, 8, 10, and 14 indicate that the frequency, limit, and time can be regulated.

A human being performs either (1) or (2) of these tasks at the initial production stage or while the machine is operating. As a result, we designate the sub-category *manipulation* as the frame for characterizing the predicative term *einstellen*. Three types of conceptual relations are included in the definitional pattern for the frame *manipulation*: (1) *is_a*; (2) *has_patient*; and (3) *has_instrument*. Within this frame, other related actions at the pre-production stage, such as *festlegen*, *überprüfen*, and *kalibrieren*, share similar features and argument templates.

Lemma	Frequency	Relative [?]
1 <input type="checkbox"/> Grenze einstellen	11	14.65 ...
2 <input type="checkbox"/> Ventil einstellen	3	4.00 ...
3 <input type="checkbox"/> Betriebsgrenze einstellen	3	4.00 ...
4 <input type="checkbox"/> Drehzahl einstellen	2	2.66 ...
5 <input type="checkbox"/> Maschine einstellen	2	2.66 ...
6 <input type="checkbox"/> Mann einstellen	2	2.66 ...
7 <input type="checkbox"/> Wechselzahnrad einstellen	2	2.66 ...
8 <input type="checkbox"/> Werkstückhöhe einstellen	2	2.66 ...
9 <input type="checkbox"/> Anpressfeder einstellen	1	1.33 ...
10 <input type="checkbox"/> Zeit einstellen	1	1.33 ...
11 <input type="checkbox"/> Viskosität einstellen	1	1.33 ...
12 <input type="checkbox"/> Drahtvorschub einstellen	1	1.33 ...
13 <input type="checkbox"/> Gegenkoppelnetzwerk einstellen	1	1.33 ...
14 <input type="checkbox"/> Stromfrequenz einstellen	1	1.33 ...
15 <input type="checkbox"/> Iuftstromes einstellen	1	1.33 ...
16 <input type="checkbox"/> Schweißstromstärke einstellen	1	1.33 ...
17 <input type="checkbox"/> Stückzahl einstellen	1	1.33 ...

Fig. 1: The participants of the predicative term *einstellen*

3.3 Syntactic-Semantic Contextual Information

Contextual information plays a crucial role in language comprehension and production. Domínguez Vázquez & Gouws (2023) highlight the importance of contextual information, noting that speakers combine words with contextual cues in various communicative situations. In the Frame-based approach, contextual information can be categorized by its scope (local and global) and the type of information conveyed (syntactic, semantic, and pragmatic variables) (cf. Faber et al., 2016). We focus on extracting local contextual data and generating syntactic-semantic patterns using CQL queries in Sketch Engine. This process involves the following steps:

- (1) **Select predicative units:** Currently, we focus on specialized verbs.
- (2) **Define participants and the valence structure:** Identify the participants and the valence structure syntactically linked to the verb, including prepositional phrases, to generate prototypical patterns.
- (3) **Extract data using CQL Queries:** Employ CQL queries in Sketch Engine to gather data for each prototypical pattern and identify the relevant slots.
- (4) **Annotate part of speech and morphological variation:** Annotate the part of speech for the extracted data and document any morphological variations of the verb.
- (5) **Establish semantic roles and add examples:** Define the semantic roles for each valence slot and include example sentences.

3.4 Equivalents Assignment

After grouping the German verbs into frames and completing their descriptions, we plan to link these verbs to their Chinese equivalents using the frame labels. The criteria for selecting equivalents as outlined by Pimentel (2012, p. 1793) include (1) the way the verbs evoke the frame; (2) the actantial structures of the verbs; and (3) the syntactic patterns of the verbs.

4. A Sample Entry

Let's address each of the three questions mentioned in Section 1:

- Our dictionary entries contain at least eight item categories: headwords, grammatical information, equivalents, frames, definitions, semantic roles and relations, syntactic-semantic structures, and examples extracted from corpora (see Figure 2).
- A comprehensive dictionary entry must address users' needs. When encountering technical texts in German, Chinese equivalents should facilitate comprehension promptly. The provided syntactic-semantic patterns aid users in completing writing tasks efficiently. Additionally, the description of verbs, especially in the highlighted section (yellow and grey sections) of the entry, serves as a concise knowledge structure to help users understand and recognize concepts.
- The Frame-based approach, which integrates semantic and syntactic behaviors, has proven to be an effective method for developing bilingual specialized dictionary entries, especially for predicative terms such as verbs.

einstellen (Eng: to regulate) vt.

Definition 定义: etwas mit Hilfe von etwas zu regeln

Chinese equivalent 中文: 调节, 设置

[IS_A] Handlung

Frame Category 类别: Manipulation

Semantic network 语义组合

[Object] Zubehörteil, Parameter

[Instrument] Ersatzteil, Tool

Structure and example 结构和例子

Object [Gerät, Zubehörteil, Parameter] (in Akkusativ) + verb

	German	Chinese
Example	Grenze; Frequenz	范围; 频率
	Maschinen; Zylinder	机器; 气缸

Preposition + Object [Werkzeug, Ersatzteil] (in Dativ) + verb

	German	Chinese
Example	mit Ventil	阀

Sentences 例句

- Damit das Motormanagement ein definiertes Luft-Kraftstoff-Gemisch einstellen kann, muss diese Luftmenge genau bekannt sein. 必须精确地知道这个空气量, 以便发动机管理系统能够调节一个确定的空气/燃油混合气量.
- Bei der Steuerung mit Proportionalventil werden alle Geschwindigkeiten mit diesem Ventil eingestellt. 在带比例阀的控制系统中, 所有速度均由该阀设置.

Fig. 2: The Sample entry for the verb *einstellen*

5. Conclusion and Future Work

Compiling a bilingual dictionary for scientific and technical terms is not a new topic in specialized lexicography. Our contribution is to use a Frame-based approach to improve the dynamism and coherence of entries in German-Chinese bilingual dictionaries in the field of the automotive industry.

Regarding the limitations, the next step involves broadening experts' articles and papers into the comparable corpora and further refining the microstructure of the dictionary entries. The future dictionary will contain 50 preliminary predicative terms and the comprehensive entries as shown in Figure 2. A number of multi-word terms will be taken into account in headwords as well.

References

- Bergenholtz, H., & Tarp, S. (2010). CHAPTER 1. LSP Lexicography or Terminography? The lexicographer's point of view. In P. A. Fuertes-Olivera (Ed.), *Specialised Dictionaries for Learners* (pp. 27–38). De Gruyter.
- Domínguez Vázquez, M. J., & Gouws, R. H. (2023). The Definition, Presentation and Automatic Generation of Contextual Data in Lexicography. *International Journal of Lexicography*, 36(3), 233–259. <https://doi.org/10.1093/ijl/ecac020>
- Faber, P., Márquez Linares, C., & Vega Expósito, M. (2005). Framing Terminology: A Process-Oriented Approach. *Meta*, 50(4). <https://doi.org/10.7202/019916ar>
- Faber, P. (2011). The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. *Terminology*, 17(1), 9–29.
- Faber, P., Reimerink, A., & Leon-Araúz, P. (2016). Encoding context in bilingual specialized dictionaries. In T. Margalitadze, & G. Meladze (Eds.), *Lexicography and Linguistic Diversity: Proceedings of the XVII EURALEX International Congress* (pp. 491–498). Ivane Javakhishvili Tbilisi State University.
- Fillmore, C. J. (1982). Frame Semantics. In *Linguistics in the Morning Calm: Selected Papers from SICOL-1981* (pp. 111–137). Hanshin Publishing Co.
- Fillmore, C. J. (1985): Frames and the Semantics of Understanding. *Quaderni di Semantica*, 6, 222–253.
- Förster, R., & Förster, A. (2018). Einteilung der Fertigungsverfahren nach DIN 8580. In R. Förster, & A. Förster (Eds.), *Einführung in die Fertigungstechnik* (pp. 23–135). Springer. https://doi.org/10.1007/978-3-662-54702-1_2
- Gouws, R. H. (2020). Special Field and Subject Field Lexicography Contributing to Lexicography. *Lexikos*, 30(1), 143–170. <https://doi.org/10.5788/30-1-1568>
- L'Homme, M.-C. (2012). Adding syntactico-semantic information to specialized dictionaries: an application of the FrameNet methodology. *Lexicographica*, 28, 233–252. <https://doi.org/10.1515/lexi.2012-0012>

L'Homme, M.-C. (2015). Predicative Lexical Units in Terminology. In N. Gala, R. Rapp, & G. Bel-Enguix, (Eds.), *Language Production, Cognition, and the Lexicon. Text, Speech and Language Technology*, vol. 48 (pp. 75–93). Springer, Cham.

Pimentel, J. (2012). Identifying equivalents of specialized verbs in a bilingual comparable corpus of judgments: A frame-based methodology. In N. Calzolari et al. (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 1791–1798). European Language Resources Association (ELRA).

Schierholz, S. J. (2003). Fachlexikographie und Terminographie. *Zeitschrift für Angewandte Linguistik*, 39, 5–28.

Schierholz, S. J. (2014). 17. New developments in lexicography for special purposes I: An overview of linguistic dictionaries. In R. Gouws, U. Heid, W. Schweickard, & H. Wiegand (Eds.), *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography* (pp. 431–442). De Gruyter Mouton.

ten Hacken, P. (2016). Bilingual Dictionaries and Theories of Word Meaning. In T. Margalitadze, & G. Meladze (Eds.), *Lexicography and Linguistic Diversity: Proceedings of the XVII EURALEX International Congress* (pp. 61–76). Ivane Javakhishvili Tbilisi State University.

Wandji, O., L'Homme, M.-C., & Grabar, N. (2013). Discovering Semantic Frames for a Contrastive Study of Verbs in Medical Corpora. In *Actes de la 9e conférence internationale Terminology and Artificial Intelligence (TIA'13)*. Paris, France.

Contact information

Chenlu Yu

Universitat Pompeu Fabra
chenlu.yu@upf.edu