Article

Improving Norwegian Translation of Bicycle Terminology Using Custom Named-Entity Recognition and Neural Machine Translation

Daniel Hellebust [ID] and Isah A. Lawal *[ID]

Department of Applied Data Science, Noroff University College, 4612 Kristiansand, Norway; danhel53935@stud.noroff.no
* Correspondence: isah.lawal@noroff.no

Abstract: The Norwegian business-to-business (B2B) market for bicycles consists mainly of international brands, such as Shimano, Trek, Cannondale, and Specialized. The product descriptions for these brands are usually in English and need local translation. However, these product descriptions include bicycle-specific terminologies that are challenging for online translators, such as Google. For this reason, local companies outsource translation or translate product descriptions manually, which is cumbersome. In light of the Norwegian B2B bicycle industry, this paper explores transfer learning to improve the machine translation of bicycle-specific terminology from English to Norwegian, including generic text. Firstly, we trained a custom Named-Entity Recognition (NER) model to identify cycling-specific terminology and then adapted a MarianMT neural machine translation model for the translation process. Due to the lack of publicly available bicycle-terminology-related datasets to train the proposed models, we created our dataset by collecting a corpus of cycling-related texts. We evaluated the performance of our proposed model and compared its performance with that of Google Translate. Our model outperformed Google Translate on the test set, with a SacreBleu score of 45.099 against 36.615 for Google Translate on average. We also created a web application where the user can input English text with related bicycle terminologies, and it will return the detected cycling-specific words in addition to a Norwegian translation.

Keywords: cycling; machine translation; named-entity recognition; domain-specific; transfer learning

1. Introduction

Value creation in the business-to-business (B2B) context has recently changed, becoming more digital [1,2]. The advent of COVID-19 changed B2B sales and helped jump-start the digitization process. B2B companies consider digital interactions more crucial to their customers than traditional sales interactions [3]. In addition, the demand for self-service in B2B has increased, elucidating the importance of the digital presence of vendors [3]. Sales communication has shifted from oral to written, which needs to be delivered online clearly and persuasively. Therefore, digitization increases the demand for product descriptions in the local language. The Norwegian B2B market within retail comprises more than one hundred vendors, where at least twenty vendors operate within the bicycle industry [4]. Most available brands are international, where the original product descriptions are available in English. There is a need for local translations, and these translations are either outsourced or performed manually by online tools such as Google. However, these product descriptions include bicycle-specific terminologies, which are challenging for online translators. Thus, local companies need an accurate and automatic translation tool.

Machine translation (MT) is a sub-field of Natural Language Processing (NLP) [5]. Recent developments in the field of NLP have led to the creation of several language models that are excellent at translating generic text, thus increasing the performance of MT to a level comparable to that of professional human translators in news translation [6].

However, its performance on domain-specific terminology needs some improvement [7]. A domain-specific term relates to the rare-word problem in MT, where out-of-vocabulary words are tokenized as unk [8]. This problem is resolved by identifying out-of-vocabulary words and translating them in a post-processing step [9]. Motivated by this approach, this study focused on creating a tool to identify bicycle-related words and then translate these words separately from the rest of the sentences. However, our method differs because several words in bicycle terminology are not out-of-vocabulary words, but they just happen to have different meanings. There are relatively few unique words in bicycle terminology [10]. Nonetheless, incorrect translations may produce a completely different context. For example, as shown in Figure 1, a bicycle wheel consists of a rim, hub, spokes, and nipples. Machine translation algorithms may classify spoke as a verb and not a noun or nipple as a body part instead of a nut connecting the spoke to the rim. Thus, we investigated methods to increase the accuracy of the machine translation of product descriptions in the domain of the bicycle industry by adapting neural machine translation (NMT) approaches from other sectors with similar word adoption problems. Due to a lack of publicly available datasets for training the NMT model to translate product descriptions related to the bicycle industry from English to Norwegian, we created one for our problem. Finally, we evaluated the model performance and compared it with that of Google Translate.

Figure 1. Bicycle parts and their naming conventions [10].

2. Related Work

NLP is how a computer understands natural language. The language text analysis is performed by splitting the text into tokens (tokenization), such as words, phrases, or sentences [11]. The task of domain-specific machine translation is composed of several sub-tasks, including a syntactic process that deals with the structure of sentences and a semantic process that deals with the meaning of sentences [11]. The semantic process is the most challenging due to the complexity of understanding context. A word-by-word translation is problematic without context knowledge. In the bicycle industry, a spoke is a rod connecting the hub and the rim, but spoke could also refer to a verb. In this section, we review the literature to understand the state of the art. The task of creating a domain-specific translation tool requires a domain-specific parallel corpus. However, creating a domain-specific parallel corpus is complex. Bago et al. [12] presented an overview of a two-year-long completed project of providing resources in Irish, Norwegian, Croatian, and Icelandic for purposes of language engineering (PRINCIPLE). The project helped develop a machine translation system to translate EU law into Norwegian. The recent development of machine translation (MT) has led to the creation of several language models based on the transformer architecture, which is excellent at translating generic

text. However, its performance for domain-specific terminologies is lower due to the lack of labeled data for the specific domain [7]. One approach to fine-tuning the machine translation model is customization through data [13]. In order to train the model for a specific domain, it will need annotated data from the target domain to learn from the relationships between words and thereby learn the domain-specific context. The language model architecture allows for fine-tuning, which makes it possible to apply transfer learning to the language model [14]. Castilho et al. [15] presented a study that leverages machine translation for product descriptions in e-commerce. The authors fine-tuned a model with a dataset of product listings in both English and German, which increased the quality of the MT models [15].

Machine translation has evolved from statistical-based to neural-network-based systems. Neural Machine Translation (NMT) [16] started to gain popularity in 2015 as a new approach to machine translation, with the goal of building an extensive neural network that reads a sentence as input and outputs the correct translation. NMT consists of two components: an encoder and a decoder. The encoder transforms the input sentence into a context vector representation using a Recurrent Neural Network (RNN). The decoder, on the other hand, predicts the next word given the context vector. However, the encoder–decoder approach has challenges with long sentences due to long context vectors, and the model must be trained using parallel corpora, which may not be available. A better solution was proposed by Bahdanau et al. [17], who extended the encoder–decoder by letting a model (soft-) search for a set of input words or their annotations computed by an encoder when generating each target word. Another revolution within machine translation was the use of an attention mechanism to make the models pay more attention to some words than others [18]. An important generalization of attention is the multi-head attention known as the transformer architecture, which was proposed by Vaswani et al. [18] and is shown in Figure 2. In 2018, Google launched a new language model called BERT based on the transformer architecture [19]. Due to its transformer architecture, it can be fine-tuned to specific tasks with high performance [19]. However, the BERT model is trained on an English corpus and performs best in English. Since BERT was released, several other versions have been developed for local languages, such as the Norwegian model NorBERT [20].

Fine-tuning a large language model requires a domain-specific adaptation [21,22]. Depending on the complexity of the domain, there are different approaches to tackling the domain adaption problem [23]. Recent research has combined tools and techniques in NLP to develop a custom Named-Entity Recognition (NER) tool to extract domain-specific entities from documents in the context of healthcare [24]. NER identifies named entities, such as a person, location, place, or thing, in a text [25]. Like transfer learning in computer vision [26], fine-tuning a language model for a specific task allows the NER model to be trained with an annotated corpus using a Python library called spaCy [24].

Inspired by the development in NER using spaCy, we exploited a similar approach to create a custom NER model for bicycle terminologies in this work. We explored transfer learning to improve the machine translation of bicycle-specific terms from English to Norwegian, including generic text. We note that the problem we are addressing might be related to a narrow topic; nevertheless, it covers all processes for the translation of domain-specific lingo to a low-resource language, i.e., creating the domain-specific dataset, building the language-translation model, and manually proofreading the sentences as a quality check, etc. For example, developing a language model for make-up/cosmetic products would involve gathering text from make-up magazines, identifying the domain-specific words, translating sentences to Norwegian/other languages, and then manually proofreading and editing to create a sentence-pair dataset. Then, the best-performing language model is fine-tuned for the specific language. We followed the same steps in our work.
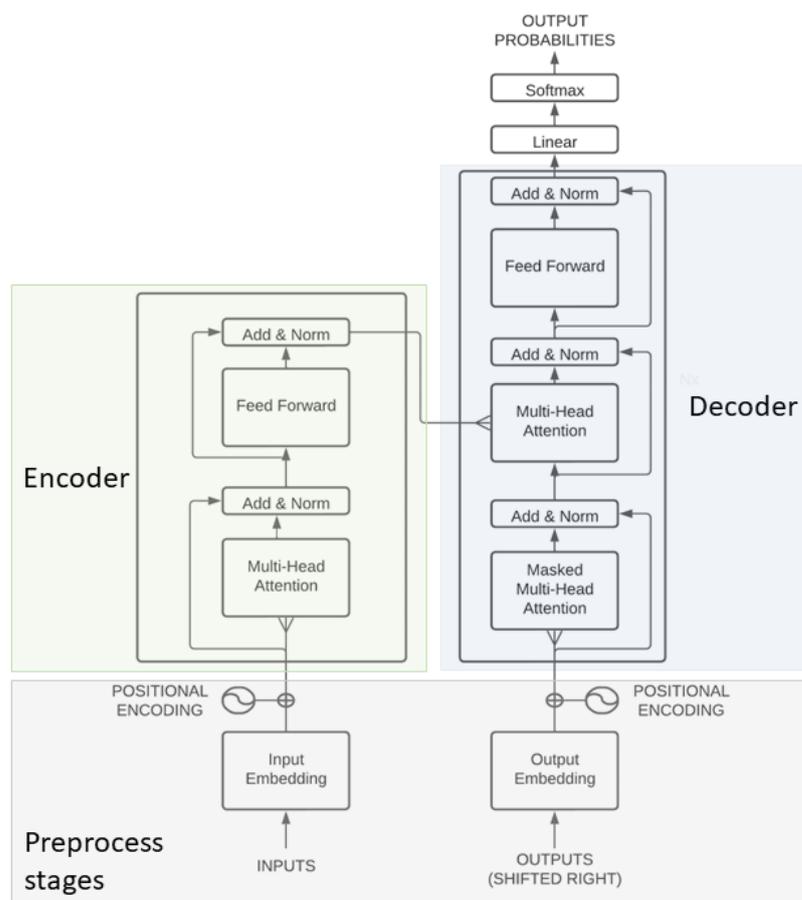
Figure 2. The transformer language model architecture [18].

3. Methodology

In this work, we aimed to investigate and develop solutions to improve the machine translation of bicycle parts and their terminologies. As mentioned in the previous sections, labeled data are the key to teaching a machine translation model a new domain. However, due to the lack of a public dataset of labeled data in the bicycle domain, we developed one in this work. We then used the dataset to train a language model for our specific domain translation. The following sections discuss the data generation process and the implementation of the translation model.

3.1. Dataset Generation

To our knowledge, there is no available dataset on cycling lingo. Thus, we decided to create a custom dataset and, by domain adaption, follow the process proposed by Tarcar et al. [24]. Figure 3 shows the workflow of the dataset creation process. First of all, we gathered a cycling-specific text corpus. Then, we created a cycling lingo dictionary from the text. The collected text corpus was annotated by identifying the cycling lingo location. The annotated dataset was then used to train a custom NER tool to identify cycling-specific words. The custom NER model was then used to tag cycling lingo as an entity, after which we translated these entities with a custom-made dictionary. This process allowed us to achieve the consistent translation of the cycling lingo. The dataset consists of sentences from online articles that are highly focused on specific bicycle parts, such as wheels, derailleurs, bottom brackets, brakes, etc. There are 1000 sentence pairs with 1298 identified cycling words, 227 of which are unique words that are difficult to translate by Google. Additionally, to allow the language model to translate generic text from English to Norwegian, we added the same 54,666-generic-word vocabulary employed in [27]. Table 1 highlights the components of the dataset and summarizes their sizes, while Figure 4 shows the most

frequent cycling words identified in the dataset. A dataset containing between 1000 and 3000 sentence pairs has been shown to be effective in fine-tuning language models [28,29]. Thus, in our case, because bicycle-related words are limited, using 1000 sentence pairs is enough to achieve our desired results. The sentence pairs were translated manually by a native Norwegian speaker and industry expert in the bicycle domain with ten years of professional experience.

Table 1. Summary of the dataset collected and used in our work.

Sentence Pairs	Generic-Word Vocabulary	Bicycle-Part Vocabulary
1000	54,666	1298
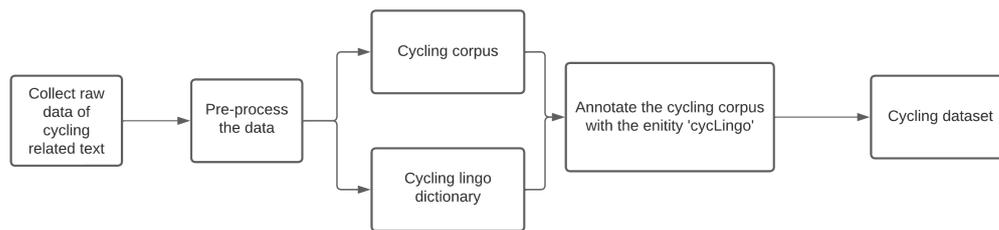
Figure 3. Illustration of the data generation processes.
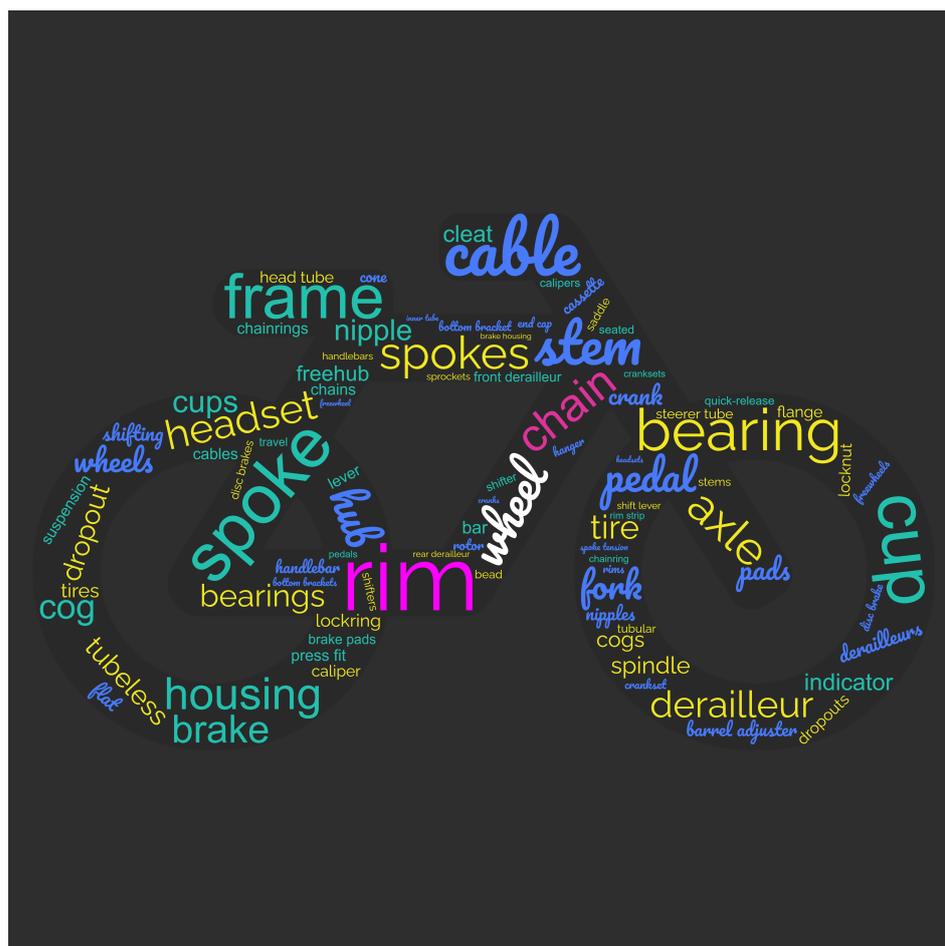
Figure 4. The identified cycling words in our collected dataset.

The data collection process consists of two parts, one for the cycling-related corpus and the other for our custom NER tool, which needs annotated data from the cycling lingo context to learn how to detect cycling lingo as an entity. We gathered the cycling-related

corpus by scraping several online sources, including https://en.wikipedia.org/wiki/List_of_bicycle_parts (accessed 8 January 2023), https://www.bicycling.com/ (accessed on 10 January 2023), https://www.parktool.com/ (accessed on 10 January 2023), and https://www.velominati.com/ (accessed on 10 January 2023). We decided to collect our data from these specific online sites because they have articles that are highly focused on bicycle parts, thus allowing us to obtain a highly relevant pool of information for our dataset. For example, Park Tool is an American company producing bicycle-specific tools for repair and maintenance. They have an extensive collection of publicly available articles for maintenance and repair. The overall raw text file contains 146,231 rows of various lengths, and to pre-process them, we loaded all raw data into a panda data frame. We counted each row for characters and excluded sentences with fewer than 200 characters, allowing us to have some context available in the sentences, as short sentences were not of value to us. After removing duplicates, we had 74,268 rows in our data frame. Then, we cleaned the data by removing all in-text citation numbers and reference symbols. Because we want to ensure that each row contains a single sentence for annotation purposes, we used the function sent in the spaCy library to split rows into sentences. Finally, the collected raw data of the cycling corpus were pre-processed and ready for annotation. Because of the large size of the corpus to be labeled, we created a Python script for annotation, which looped through every sentence in our sentence dataset and identified the words in our cycling lingo dictionary. However, we note that each word in our cycling lingo dictionary may appear in different casing, such as uppercase or lowercase, and could also have different suffixes due to the plural form. Thus, we accounted for each of those cases to make the annotation quality as good as possible.

Annotating single words from the lingo dictionary was straightforward; however, several words in cycling lingo consist of more than one word separated by space, such as bottom bracket, brake lever, and inner tube. To capture the different forms of a word, for example, the plural, we converted the word to its lemma before matching. We used the Python library "re" and its function "finditter" to find all matches of a word in the sentence to the lingo list. Furthermore, we identified the location of a word within a sentence by using the start function of "finditter". The position of the word was captured by adding the start index of the word in the sentence together with the length of the word, which together pin-points the end of the word. Figure 5 shows examples of the annotated sentences generated by the proposed data generation process. For each sentence, we identify the cycling lingo and its position in the sentence.

```
{'text': 'Chains should be cleaned before lubrication.', 'label': [[0, 6, 'cycLingo']]}
{'text': 'The same material can be used in handlebars.', 'label': [[33, 43, 'cycLingo']]}
{'text': 'Most cleats develop more float as they wear.', 'label': [[5, 11, 'cycLingo']]}
{'text': 'Air-sprung forks are generally lighter still.', 'label': [[11, 16, 'cycLingo']]}
{'text': 'Most adult mountain bikes use 26 inch wheels.', 'label': [[38, 44, 'cycLingo']]}
```

Figure 5. Samples of annotated sentences.

3.2. Implementation of the Language-Translation Tool

The implementation of the proposed translation tool consists of three parts. Firstly, we trained a custom NER model, and afterward, we learned an NMT model for the language-translation process. Lastly, we developed a web application of the developed translation tool to detect cycling-specific terminology and translate it accordingly using the spaCy, Dash, and Transformers libraries. The following sections discuss the implementation process of the various parts of the tool.

3.2.1. Custom NER Model

To train our custom NER model, we used the generated dataset discussed in Section 3.1. Because some cycling lingo words do not have ambiguous meanings and can be translated correctly by most language models, we used a dataset with sentences from bicycle repair articles collected from https://www.parktool.com/ (accessed on 10 January 2023). These

sentences contain mainly words with ambiguous meanings and chunks that include prepositions. For example, words such as an inner tube, bottom bracket, rear derailleur, front derailleur, and top tube include prepositions or positional words, which can be confused as two separate words by most NER models. Thus, we manually annotated these sentences before using the dataset to train the customized NER model. In total, 2090 rows of text and 8720 words were tagged as cycling lingo with the entity cycLingo. The annotated dataset was randomly shuffled and split into 80% training and 20% test sets. We employed the training set to learn the custom NER model using spaCy. SpaCy uses the transformer architecture shown in Figure 2 to build the NER model.

3.2.2. Neural Machine Translation (NMT) Model

The Norwegian translation of bicycle terminologies can be accomplished in two ways. The first way is to translate the identified cycling lingo word by word and then employ an MT model to translate the rest of the sentences. This process allows complete control and the consistent translation of ambiguous words. In addition, this approach is fast, as it avoids translating every sentence from scratch. However, the word-by-word translation approach has its challenges. For example, by masking the words for translation, the model loses information and produces outputs that might be grammatically inaccurate. On the other hand, we obtain a more grammatically correct translation by training the language model with a domain-specific sentence-pair dataset. In this work, we employed the trained custom NER model discussed in Section 3.2.1 to create such a sentence-pair dataset. Figure 6 illustrates the steps to generate the sentence-pair dataset using our custom NER. We focused on ambiguous words in cycling terms, such as hub, spoke, nipple, housing, and bottom bracket. We translated one thousand sentences manually to create our sentence-pair dataset. We proofread and edited every sentence to obtain the best-quality translation. The dataset was then randomly shuffled and split into 80% training, 10% validation, and 10% test sets for learning the NMT model. There are very few models that are able to translate text from English to Norwegian that we could adapt for our work. These include the Marian translation model (MarianMT) [27] and Many-to-Many (M2M) multilingual translation model developed by Facebook researchers [30]. The MarianMT model is known to perform well in translating generic texts from English to Norwegian. It uses the transformer architecture shown in Figure 2, which has six decoder layers and six encoder layers in the encoder and decoder, respectively, with sixteen attention heads. The model is pretrained with the OPUS corpora, consisting of over 74 million English and Norwegian sentence pairs [31]. Other popular multi-language models, such as Multilingual Denoising Pre-training for Neural Machine Translation (mBART) [32], do not cover Norwegian. In addition, the Multilingual Text-to-Text Transfer Transformer (MT5) from Google [33] is not pre-trained on a Norwegian corpus, so it would require retraining from scratch for our translations. To develop our NMT model, we performed a preliminary selection test using M2M and MarianMT to translate some of the texts in our dataset. MarianMT outperformed the M2M model with an average SacreBleu score of 41.511 against the 24.497 obtained by the M2M model. The pre-trained MarianMT model performed well on generic text but needed domain-specific terminology improvement to apply to our purposes. Additionally, we chose to use the pre-trained model because it would reduce computation costs by allowing us to use a state-of-the-art model without having to train it from scratch [34]. Thus, we fine-tuned it using our custom sentence-pair dataset. To fine-tune the MarianMt model, we followed the same fine-tuning procedure in the Transformers library using the default training parameters accordingly (https://huggingface.co/docs/transformers/training) (accessed on 2 February 2023).
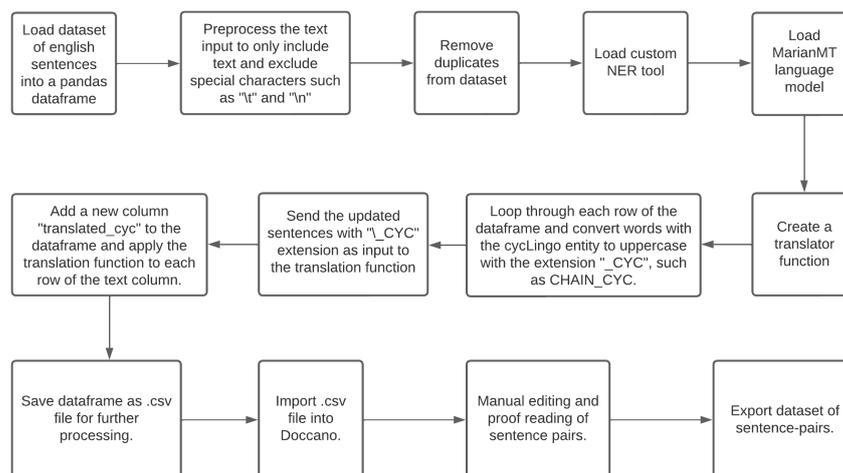
Figure 6. The steps for creating a sentence-pair dataset used for learning the NMT model.

3.2.3. Web Application for Language Translation

One of the main objectives of our research was to create a prototype software tool to translate text, including bicycle-specific terminology, from English to Norwegian. So far, we have developed our dataset to train a custom NER model, then used this model to create a sentence-pair dataset, and finally fine-tuned a language model on our custom dataset. To tie all parts together, we built a user-friendly web application of the language-translation tool using the Python Dash library, which is built on top of the Flask library. The web application has an input text field, a translation button, and an output text field. The application loads our trained NMT model and our custom NER model. The user can input English text, and the application will return the detected cycling lingo words, including a Norwegian translation. Figure 7 shows the user interface of the developed web application for our proposed machine translation of bicycle-specific terminology from English to Norwegian, including generic. The complete code for the developed language models is available at our GitHub repository (https://github.com/danielhellebust/cycLingo), while the web application is hosted on our Huggingface hub (https://huggingface.co/spaces/DanielHellebust/cycLingoTranslator).

Figure 7. The user interface of the proposed web application (cycLingo Translator) using custom NER detection and NMT language translation.

4. Evaluation and Discussion

4.1. Evaluation Setup

To facilitate the efficient development and evaluation of our proposed method, we used an open-source machine learning framework library produced by Meta AI known as PyTorch. We conducted our experiments on a consumer-graded personal computer with RTX3060ti 8 GB GPU and a Windows 11 operating system. We performed both qualitative and quantitative evaluations of the performance of our proposed models, i.e., the custom NER model and the NMT model. We evaluated the custom NER model quantitatively using the built-in evaluation metrics from the spaCy library, which calculates the F1 score of the model output on the test set. Similarly, we assessed the performance of the NMT model using the SacreBleu score, a machine translation evaluation method that is language-independent and correlates highly with human evaluation [35]. We compared the SacreBleu scores for the proposed NMT model with those obtained using the classic Google Translate app.

4.2. Evaluation of Custom NER Model

Figure 8 shows a sample of the qualitative outputs of the custom NER model. The NER model correctly identified all the cycling lingo entities in the English texts, including capturing noun chunk entities such as "shifter cable" and "rear derailleur", which are challenging to detect because they consist of two tokens. Thanks to the quality of the generated dataset and the learned custom NER model, the accurate detection of cycling lingo entities in a sentence allows direct word-by-word translation using a custom dictionary. We also evaluated the performance of the custom NER model by computing the F1 score of its outputs using the spaCy library. Our custom NER model achieved, on average, an F1 score of 88.91% in detecting the cycling lingo words in the test set.

Figure 8. Sample qualitative results of the proposed custom NER model. It is was able to detect all cycLingo entities (see the purple color boxes) in the English texts, including capturing noun chunk entities such as "shifter cable" and "rear derailleur", which are challenging.

4.3. Evaluation of the NMT Model

To evaluate the trained NMT model, we used a test set that consists of 100 sentences manually translated by domain experts who are also Norwegian native speakers. The goal is to use our trained NMT model to translate the sentences as closely as possible to human translation. Figure 9 shows a screenshot of a sample of the qualitative output of our

trained NMT model. The model correctly translated all English texts presented to it. We also evaluated the trained NMT model quantitatively by computing the SacreBleu score of its translation. SacreBleu tokenizes the translated sentences and counts matching words against the manually translated test set to calculate the scores. Our trained NMT model achieved, on average, a SacreBleu score of 45.099 on the test set.

We compared the performance of the proposed method with the classic Google Translate. Figure 10 shows the SacreBleu scores for both approaches on some test samples. In general, our model outperformed Google Translate on the entire test set, with an average SacreBleu score of 45.099 against 36.615 for Google Translate. Figure 11 shows the outputs of Google Translate and our proposed method on the same bicycle-related article from the internet. Google Translate translated the text shown in Figure 11-top as Den brukes også når du installerer en ny eiker. Hver eike er festet til hjul felgen med en eike nippel, som kan snus for å enten stramme eller løsne eikens spenning. (see Figure 11-bottom left). In contrast, our proposed model translated the text as Det brukes også når du installerer en ny eike. Hver eike er festet til hjul felgen med en eike nippel, som kan dreies for enten å stramme eller løsne eikens spenning. (see Figure 11-bottom right). Notably, Norwegian grammar is better with our model than with the Google Translate app. The proposed NMT model also translates "turned" to "dreies", which is more appropriate in this context than Google's translation of "snus".

cycLingo Translator

Translate cycling specific text from English to Norwegian

> When replacing disc brake pads, the first thing to consider is what type of pads you are using. Always replace your pads with the SAME TYPE of pad material, unless you plan on replacing both the pad and the rotor

Translate

Detected cycLingo entities:

When replacing [disc brake pads **cycLingo**], the first thing to consider is what type of [pads **cycLingo**] you are using. Always replace your [pads **cycLingo**] with the SAME TYPE of [pad **cycLingo**] material, unless you plan on replacing both the [pad **cycLingo**] and the [rotor **cycLingo**]

Norwegian Translation:

> Når du erstatter skivebremser, er det første du bør vurdere hvilken type bremseklosser du bruker. Erstatt alltid bremseklosser med den samme typen bremsekloss materiale, med mindre du planlegger å erstatte både bremseklossen og bremseskiven.

Figure 9. Sample qualitative result of the proposed custom NMT model. The text in the purple color boxes are the cycLingo entities detected in the text. It was able to accurately translate English texts including domain-specific bicycle terms to Norwegian.

English Text	Goundtruth (Norwegian)	Proposed Method	Google Translator	SacreBleu - Proposed Method	SacreBleu - Google Translator
This is because the right side is displaced forward by the hanger bracket.	Dette skyldes at høyre side forskyves fremover av girøre braketten.	Dette skyldes at den høyre siden er forskjøvet fremover av girøret.	Dette er fordi høyre side er forskjøvet fremover av opphengsbraketten.	8.2	8.5
Check the dish reference on both sides, and repeat as necessary until the wheel shows a 1mm or less gap at the hub.	Kontroller sentreringsreferansen på begge sider, og gjenta om nødvendig til hjulet viser en 1 mm eller mindre gap på navet.	Sjekk sentreringsreferansen på begge sider, og gjenta om nødvendig til hjulet viser et gap på 1 mm eller mindre ved navet.	Sjekk tallerkenreferansen på begge sider, og gjenta etter behov til hjulet viser en åpning på 1 mm eller mindre ved navet.	57.2	38.8
It is also recommended to wear protective gloves such as the MG-2, and use eye protection.	Det anbefales også å bruke beskyttelseshansker som MG-2, og bruke øyebeskyttelse.	Det anbefales også å bruke beskyttelseshansker som MG-2, og bruke øyebeskyttelse.	Det anbefales også å bruke vernehansker som MG-2, og bruk øyevern.	72.2	31.5
The design is a variation of the cantilever brake.	Designet er en variant av cantilever brems.	Konstruksjonen er en variant av cantilever brems.	Designet er en variant av utkragerbremsen.	57.1	27.8
Pick up the front of the bike, and gently swing the handlebars back and forth from center.	Plukk opp forsiden av sykkelen, og forsiktig sving styret frem og tilbake fra sentrum.	Plukk opp forsiden av sykkelen, og forsiktig sving styret frem og tilbake fra sentrum.	Ta opp fronten av sykkelen, og sving styret forsiktig frem og tilbake fra midten.	77.0	29.7

Figure 10. Comparison of the SacreBleu scores for both Google Translate and our trained NMT models.

Bicycling.com article

Google translation

Our translation

Figure 11. Comparison of the translation of the https://www.bicycling.com/repair/a35937378/how-to-use-a-spoke-wrench/ (accessed 8 April 2023) article using Google Translate app and our proposed method. The text in the purple color boxes are the cycLingo entities detected in the text by our proposed model before translating the entire text.

5. Conclusions

In this study, we developed a tool to translate domain-specific text for the bicycle industry from English to Norwegian. Firstly, with the help of domain experts in the bicycle industry and Norwegian native speakers, we created a custom dataset for our work. To the best of our knowledge, this is the first domain-specific dataset for product descriptions related to the bicycle industry in Norwegian. The dataset also contains sentence-pair sets translated manually by an industry expert. Then, we created a custom NER model that identifies bicycle-specific words and labels them as a cycling lingo entity. We trained an NMT model using a sentence-pair dataset. The learned NMT model was then used to translate bicycle-related English texts. Our model outperformed Google Translate on our test set, with an average SacreBleu score of 45.099 against 36.615 for Google Translate. Our translation tool can translate bicycle-specific text from English to Norwegian with higher performance than Google Translate. The source code, generated dataset, and web application have been made available to the public for research purposes.

Notably, our work is limited to bicycle terminology of fixed, repair, and maintenance parts. This is because we focused mainly on challenging bicycling words that are difficult for the classic Google Translate app. Thus, due to the specific domain vocabulary used during the development of our language model, the model might not be directly applicable for translating text containing product or service terminologies in other domains different from the one it was trained for. However, the process of how the models were built can be applied to other domains as well. Future work could include the creation of a more comprehensive sentence-pair dataset to increase the performance and applicability of the model to other areas of human endeavors. A possible scenario is to liaise with native Norwegian speakers and industry experts in some sectors to help manually translate product and service texts from English to Norwegian, thus creating a comprehensive sentence-pair database to further fine-tune the NMT model. Many companies do already have all sentence pairs from previous translations, but they are not used; instead, they repeat the translations over and over again, which is cumbersome and costly.

Author Contributions: D.H. carried out this research as part of his bachelor's degree project. I.A.L. was involved in planning and supervising the work and drafting and reviewing the manuscript. The methodology, analysis, and discussion were conducted by all participants. All authors have read and agreed to the published version of the manuscript.

References

1. Corsaro, D.; Anzivino, A. Understanding value creation in digital context: An empirical investigation of B2B. *Mark. Theory* **2021**, *21*, 317–349. [CrossRef]
2. Rusthollkarhu, S.; Hautamaki, P.; Aarikka-Stenroos, L. Value (co-)creation in B2B sales ecosystems. *J. Bus. Ind. Mark.* **2021**, *36*, 590–598. [CrossRef]
3. Gavin, R.; Harrison, L.; Plotkin, C.L.; Spillecke, D.; Stanley, J. The B2B Digital Inflection Point: How Sales Have Changed during COVID-19. McKinsey & Company. 2020. Available online: https://www.mckinsey.com/ (accessed on 5 March 2023).
4. Norsk Sportsbransjeforening. Bransjeregisteret. 2022. Available online: https://sportsbransjen.no/bransjeregister (accessed on 20 February 2023).
5. Khurana, D.; Koli, A.; Khatter, K.; Singh, S. Natural language processing: State of the art, current trends and challenges. *Multimed. Tools Appl.* **2022**, *82*, 3713–3744. [CrossRef] [PubMed]
6. Popel, M.; Tomkova, M.; Tomek, J.; Kaiser, Ł.; Uszkoreit, J.; Bojar, O.; Žabokrtský, Z. Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals. *Nat. Commun.* **2020**, *11*, 4381. [CrossRef] [PubMed]

7.   Yan, R.; Li, J.; Su, X.; Wang, X.; Gao, G. Boosting the Transformer with the BERT Supervision in Low-Resource Machine Translation. *Appl. Sci.* **2022**, *12*, 7195. [CrossRef]

8.   Sennrich, R.; Haddow, B.; Birch, A. Neural Machine Translation of Rare Words with Subword Units. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, 7–12 August 2016; pp. 1715–1725. [CrossRef]

9.   Luong, T.; Sutskever, I.; Le, Q.; Vinyals, O.; Zaremba, W. Addressing the Rare Word Problem in Neural Machine Translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 27–31 July 2015; pp. 11–19. [CrossRef]

10.   Wikipedia. List of Bicycle Parts. 2022. Available online: https://en.wikipedia.org/wiki/ (accessed on 1 March 2023).

11.   Chowdhary, K. Natural language processing. In *Fundamentals of Artificial Intelligence*; Springer: New Delhi, India, 2020; pp. 603–649.

12.   Bago, P.; Castilho, S.; Dunne, J.; Gaspari, F.; Andre, K.; Kristmannsson, G.; Olsen, J.A.; Resende, N.; Gíslason, N.R.; Sheridan, D.D.; et al. Achievements of the PRINCIPLE Project: Promoting MT for Croatian, Icelandic, Irish and Norwegian. In Proceedings of the 23rd Annual Conference of the European Association for Machine Translation, Ghent, Belgium, 1–3 June 2022; pp. 347–348.

13.   Ramirez-Sanchez, G. Custom machine translation. *Mach. Transl. Everyone Empower. Users Age Artif. Intell.* **2022**, *18*, 165.

14.   Tunstall, L.; von Werra, L.; Wolf, T. *Natural Language Processing with Transformers*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022.

15.   Castilho, S.; Moorkens, J.; Gaspari, F.; Calixto, I.; Tinsley, J.; Way, A. Is Neural Machine Translation the New State of the Art? *Prague Bull. Math. Linguist.* **2017**, *108*, 109–120. [CrossRef]

16.   Sennrich, R.; Firat, O.; Cho, K.; Birch, A.; Haddow, B.; Hitschler, J.; Junczys-Dowmunt, M.; Läubli, S.; Miceli Barone, A.V.; Mokry, J.; et al. Nematus: A Toolkit for Neural Machine Translation. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Valencia, Spain, 3–7 April 2017; pp. 65–68.

17.   Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.

18.   Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 6000–6010.

19.   Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.

20.   Kutuzov, A.; Barnes, J.; Velldal, E.; Øvrelid, L.; Oepen, S. Large-Scale Contextualised Language Modelling for Norwegian. In Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), Reykjavik, Iceland, 31 May–2 June 2021.

21.   Liu, X.; Yoo, C.; Xing, F.; Oh, H.; Fakhri, G.E.; Kang, J.W.; Woo, J. Deep Unsupervised Domain Adaptation: A Review of Recent Advances and Perspectives. *APSIPA Trans. Signal Inf. Process.* **2022**, *11*, e25. [CrossRef]

22.   Jiang, J.; Zhai, C. Instance weighting for domain adaptation in NLP. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, 25–26 June 2007; pp. 264–271.

23.   Hu, J.; Xia, M.; Neubig, G.; Carbonell, J. Domain Adaptation of Neural Machine Translation by Lexicon Induction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 2989–3001. [CrossRef]

24.   Tarcar, A.K.; Tiwari, A.; Dhaimodker, V.N.; Rebelo, P.; Desai, R.; Rao, D. Healthcare NER models using language model pretraining. *arXiv* **2019**, arXiv:1910.11241.

25.   Yadav, V.; Bethard, S. A survey on recent advances in named entity recognition from deep learning models. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 2145–2158.

26.   Brodzicki, A.; Piekarski, M.; Kucharski, D.; Jaworek-Korjakowska, J.; Gorgon, M. Transfer Learning Methods as a New Approach in Computer Vision Tasks with Small Datasets. *Found. Comput. Decis. Sci.* **2020**, *45*, 179–193. [CrossRef]

27.   Junczys-Dowmunt, M.; Grundkiewicz, R.; Dwojak, T.; Hoang, H.; Heafield, K.; Neckermann, T.; Seide, F.; Germann, U.; Aji, A.F.; Bogoychev, N.; et al. Marian: Fast Neural Machine Translation in C++. *arXiv* **2018**, arXiv:1804.00344.

28.   Adelani, D.; Alabi, J.; Fan, A.; Kreutzer, J.; Shen, X.; Reid, M.; Ruiter, D.; Klakow, D.; Nabende, P.; Chang, E.; et al. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 10–15 July 2022; pp. 3053–3070. [CrossRef]

29.   Eisenschlos, J.; Ruder, S.; Czapla, P.; Kadras, M.; Gugger, S.; Howard, J. MultiFiT: Efficient Multi-lingual Language Model Fine-tuning. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 5702–5707. [CrossRef]

30.   Fan, A.; Bhosale, S.; Schwenk, H.; Ma, Z.; El-Kishky, A.; Goyal, S.; Baines, M.; Celebi, O.; Wenzek, G.; Chaudhary, V.; et al. Beyond English-Centric Multilingual Machine Translation. *J. Mach. Learn. Res.* **2021**, *22*, 4839–4886.

31.   Tiedemann, J.; Thottingal, S. OPUS-MT–Building open translation services for the World. In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, Online, 3–5 November 2020.

32.   Liu, Y.; Gu, J.; Goyal, N.; Li, X.; Edunov, S.; Ghazvininejad, M.; Lewis, M.; Zettlemoyer, L. Multilingual Denoising Pre-training for Neural Machine Translation. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 726–742. [CrossRef]

33.   Xue, L.; Constant, N.; Roberts, A.; Kale, M.; Al-Rfou, R.; Siddhant, A.; Barua, A.; Raffel, C. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 July 2021; pp. 483–498. [CrossRef]

34. Alt, C.; Hübner, M.; Hennig, L. Fine-tuning Pre-Trained Transformer Language Models to Distantly Supervised Relation Extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1388–1398. [CrossRef]
35. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318. [CrossRef]