



Transformer Translation for a European Language Near Extinction : A Halunder Case Study

Author: **Jakob Martens**

Supervisor: **Prof. Dr. Raad Bin Tareaf**

2nd Examiner: **Prof. Dr. Martin Hillebrand** (Chair of examination board)

Study Program

Digital Business
XU Exponential University of Applied Sciences

This thesis is submitted for the degree of Bachelor of Science (B.Sc)



En betjen fer Arkiaan, welk Halunder snakket.

Dear hid ik do reälik oalsni om toch, dat ik sallow no uun'e Loag keem skul, iaan fan'e Spreeken, wat blooat no fan men'n Lid' n snakket wart, ii faini Halunder, dat ik mi dearfer iinsat. Djüm kan djüm goorni feerstel, ho ik mi dearoawer froie – ofskoons ik de hiir Text med en Gefeel skriuw, wat ni soo gud es, wiil ik nä iaars market hoa, wat is al en heel Tid soo noa en noa ferlörsen gingen es en gungt...

Deät es machti stel würn om ii Spreek. Deät Önnerrrech fer Halunder, wat ik as Kin bi miin Skuulmaister Bettina Köhn gearn medmoaket hoa, djef' et dollung ni mear. Deät woort nä baal föfti Djooarn, dat' et med deät Halunder ni mear soo gud lapt, wiil de Spreek ümmer menner snakket wart. En as Mina Borchert, Nils Århammar en feln fan de ool Halunders, wat no Halunder snakket, stürrewen wear, es de Loag fer de Spreek no menner würn. Do nä es' et miin groot Heep, dat med miin Oarbooid oawer' t Halunder en nai Kapitel fer ii Spreek eepen moaket wart. Hellech kan wi' t ni aphool, dat' et med de Spreek ümmer menner wart, oawers med de hiir Oarbooid wel ik dearfer sürrige, dat ii Halunder, wat wi soo gearn snakken hear, uk no uun honnert Djooarn authentisch en rehti belewwet wü kan – dat' et en leäbendi Tschiiichnis fan is es en tu is heart.

Welk hid toch, dat' et meegelk es, sek en Spreek nä med en Computer en med „künstli Intelligenz“ tu oawersatten, en Spreek, wat fel uun Büln snakket en wearmed man soo gud ferhoaln kan en wat soo authentisch es as deät Halunder. Ii Spreek es machti fel mear as en poor frendli meent Würder, welk helech uf en tu nons mank' et Diitsk apdoage – dji heart tu ii Historii, es en Weerklang fan ii Feerooln. Halunder es fer mi laider ni miin Memmenspreek, oawers as Kin fan en Halunder Familiije lewwet dji uun mi, en kosboor Arrow, wat wi bewoore mut.

En nä es' et wichti fer mi, Danke tu sooin tu Mensken, wat mi hülpen hoa:

Miin Ooti Bertha, welk deät Lun uun her Hart hat; Bettina Köhn, welk mi ümmer weer önnertstettet hat med Korrekturen en Oawersattungen; Hanne Siemund-Dähn en Erika Dähn, welk mi ferkloort hoa, wat „Halunderi“ es en wat ni; Ritva en Nils Århammar fer herrem Teenken en Oarbooid om en fer deät Halunder, en Spreek, wat eentli goorni herrem Memmenspreek es; Jörg Singer, welk mi ümmer weer Mud moaket hat. En natiirli sen dear uk no miin Ooln, tuijaars miin halunder Foor. Hi hat djooarnloang alles fan' t Halunder en ii halunder Oort en Wiis fer mi apbewoort en, soo as Foor deät meäsens du, mi deät alles toop med deät Halunder fiiderden' n.

Signed Declaration of Authorship

I hereby declare that the work written and presented in this thesis is entirely my own. Where I have consulted the works of others, this is always clearly stated, with the consulted sources being individually named in the bibliography. The work at hand has not been presented for a degree at any other educational institution. It does not include material that to a substantial extent has been part of any other assignment for the duration of my studies at XU Exponential University and has not yet been published.

Jakob Martens

December 9, 2024

Disclosure Statement

While drafting this manuscript, the author utilized Claude 3.5 Sonnet for the purpose of quantifying experiment results, troubleshooting of code and alignment of a parallel corpus of 10K Halunder and German sentence pairs (text segmentation). The specific material produced by these AI tools within the manuscript is distinctly identified and elaborated within the paper. Before the final submission, the author has meticulously examined the AI-generated content and assumes complete accountability for the contents of the submitted thesis.

Jakob Martens
December 9, 2024

Abstract

This thesis addresses translation challenges extending beyond technical considerations, emphasizing linguistic, cultural, and historical factors that shape the complexity of the translation task. It focuses on “Deät Halunder”, the language of Heligoland—Germany’s most remote island, a tiny red rock in the North Sea. The Insular North Frisian language, highly idiomatic and context-dependent; Halunder uses no single word for “forest” yet offers at least three for “wind.” Its vocabulary shaped by its remote, rough maritime history. Once being the only language on the entire island, spoken by 2000 inhabitants around 1900, it now faces extinction, classified as 8b (Nearly Extinct) on the EGIDS scale by this research, with fewer than 100 speakers, about a dozen fully proficient, and just a handful capable of writing.

In response to this alarming decline, this thesis documents the development of the first fully functioning translator for Halunder, including the collection of 10K Halunder-German sentence pairs—the island’s only digital dataset suited for machine learning. This thesis likely represents the smallest and most endangered application of the Transformer model. Seven experiments and 31 configurations tested approaches ranging from training from scratch to fine-tuning pre-trained models and examining incremental data growth. Using NLLB-200-d-1.3B as a base, Halunder-to-German translations achieved about 60.5 BLEU and 75.6 Chrf++, and German-to-Halunder about 52.9 BLEU and 72.3 Chrf++. Manual evaluation of unseen data showed that 55.5% of outputs were rated good or flawless and 70.5% were at least acceptable.

By presenting a practical, community-centered approach that does not demand immediate linguistic expertise, this work is designed to empower other minority languages facing speaker decline to take resolute action before it is too late.

For Halunder, this may be the very last opportunity to rescue it from disappearing entirely.

Table of Contents

Abstract	v
List of figures	viii
List of tables	ix
Chapter 1 Introduction	10
1.1 Background and Motivation	11
1.2 Current State of the Halunder language	12
1.3 Research Objectives	13
1.4 Significance	14
1.5 Challenges and Contributions	15
1.6 Thesis Structure.....	16
Chapter 2 Theoretical Framework	17
2.1 The Need for a Translator.....	17
2.1.1 Critical Endangerment	17
2.1.2 Institutional and Legal Obligations	20
2.2 Translation Architecture	21
2.2.1 Beginnings of Machine Translation.....	21
2.2.2 Early Neural Architectures	22
2.2.3 Attention is All You Need – The Transformer Model	23
2.3 Transfer Learning	25
2.4 Historical Development and Similarities to other Languages	26
2.5 Related Work	28
2.6 Summary of the Theory.....	30
Chapter 3 Proposed Approach	31
3.1 Dataset Development	32
3.1.1 Data Availability	32
3.1.2 Data Collection and Preparation.....	33
3.1.3 Data Overview	34
3.1.4 Legal and Ethical Aspects.....	36
3.2 Dataset Analysis	37
3.2.1 Characteristics and Distribution	37
3.2.2 Limitations	41
Chapter 4 Translation Experiments	42
4.1 General Setup.....	44
4.2 Fine-Tuning existing Pre-Trained Models for Halunder to German Translation.....	45
4.2.1 Experiment Setup.....	45

4.2.2	Experiment Results	47
4.2.3	Hyperparameter Optimization for mBART50	49
4.2.4	Experiment Setup	50
4.2.5	Experiment Results	51
4.3	Hyperparameter Optimization OpusMT	52
4.3.1	Experiment Setup	52
4.3.2	Experiment Results	52
4.4	Manual Evaluation OpusMT vs. mBART	53
4.5	Impact of Dataset Size on Translation Performance	56
4.5.1	Experiment Setup	56
4.5.2	Experiment Results	57
4.6	German to Halunder Translation	60
4.6.1	Experiment Setup	60
4.6.2	Experiment Results	60
4.7	Feasibility of Training from Scratch	62
4.7.1	Experiment Evaluation	63
4.7.2	Experiment Results	64
Chapter 5	Discussion and Interpretation	67
5.1	Manual Review of Translation Quality	67
5.1.1	Results	68
5.1.2	Interpreting Translation Results	72
5.2	Future Research	72
5.3	Impact	73
Chapter 6	Conclusion	75
6.1	Summary of Key Contributions	75
6.2	Limitations and Future Recommendations	76
6.3	Stakeholder Recommendations	76
6.4	Closing Remarks	77
6.5	References	78
Appendix A	85
Appendix B	87

List of figures

Figure 1.1 Map highlighting Heligoland within the North Sea. Base map adapted from Ætoms (n.d.), CC BY-SA 4.0, with inset historical image of Heligoland (Library of Congress, 1890-1905).	12
Figure 2.1 A Conceptual Visualization of Self-Attention on a Halunder example. ...	24
Figure 3.1 Characters per unique word in Halunder and German Parallel Corpus..	37
Figure 3.2 Frequency of Top 70 Halunder and German Word Distribution	38
Figure 3.3 Words per Sentence for Halunder and German Parallel Corpus	39
Figure 3.4 The occurrences of each word in the dataset.	40
Figure 4.1 Comparison of fine-tuning mBART50 for Halunder-German	49
Figure 4.2 Hyperparameter Comparison for mbart50 using nl-de for Halunder-German Translation	51
Figure 4.3 Hyperparameter Comparison for C9, C8 and C2 between mBART50 and opusMTgem-gem.....	53
Figure 4.4 Comparison of Halunder to German translation with different amounts of sentences.....	57
Figure 4.5 Comparison for German to Halunder translation between Helsinki/opus-mt-gem-gem and 2 language proxies for facebook/nllb-200-distilled-1.3B	61
Figure 4.6 Chrf++ Result comparison for German to Halunder translation using two fine-tuned facebook/nllb-200-distilled-1.3B models.	62
Figure 4.7 Loss vs. Negative Log-Likelihood Loss of the Baseline Model	63
Figure 4.8 SacreBLEU Performance of a Baseline Non-Pretrained Model (Halunder to German)	64

List of tables

Table 2.1 EGIDS Levels (adapted from Lewis & Simons, 2010).....	19
Table 3.1 Distribution of dataset sources.	34
Table 4.1 Challenging Sentence used for Review	43
Table 4.2 Performance vs. Training Time between fine-tuned models for Halunder-German	48
Table 4.3 Distribution of SacreBLEU scores across different configurations	50
Table 4.4 Evaluated Sentences between models.	54
Table 4.5 Representative samples showcasing important differences between minimal automated metric differences	59
Table 4.6 Performance and training time of German to Halunder models.	62
Table 4.7 Representative Samples from the 250-Sentence Test Set Translated by a baseline model.....	65
Table 5.1 Distribution of manual evaluation scores by two evaluators.....	68
Table 5.2 Examples of translations containing major flaws.	69
Table 5.3 Representative examples of manually evaluated sentences	71

Chapter 1

Introduction

Deät Halunder, the Insular North Frisian language of Heligoland, (*Deät Lun*), Germany, is often referred to as one of Europe's smallest languages (Montigny, 1994b). Nearly seventy years ago, concerns were already voiced about its imminent extinction (Krogmann, 1952) – but the language persists, with continued further decline. By the early 1980s, less than 500, mostly elderly speakers remained, and more recent observations indicate that a much lower number, certainly less than a hundred speakers and less than a dozen proficient speakers live today (Köhn, 2020). With no younger generation stepping in, and disastrous hopelessness taking over (Wanke, 2008), the current situation likely represents the very last chance to rescue Halunder before it disappears entirely.

In contrast to many minority languages that retain thousands of speakers of every age, and at least some digitized materials (Flora, 2021), Halunder has been almost entirely absent from contemporary databases and digitization (Århammar, 1995). The extreme scarcity of both speakers and resources has relegated it to the far margins of machine translation efforts. Prior preservation attempts emphasized documentation and small-scale teaching, but these measures did not slow down the language's decline (Wanke, 2008).

This thesis employs Transformer-based neural machine translation to produce the first ever Halunder-German translator from severely limited data. By introducing

Halunder to state-of-the-art developments of the digital era, it seeks to enable meaningful preservation and promote the expansion of Halunder's future. Even if the decline cannot be stopped entirely, the translator aims to enable authentic and accurate translation of this culturally embedded island language, as a source of encouragement for language minorities facing a similar fate.

1.1 Background and Motivation

Heligoland, Germany's only high-sea island ("Hochseeinsel"), once lived by Halunder as the nearly exclusive local language (Siebs, 1910). Around 1905, the large majority of the island's roughly 2,300 inhabitants did not speak German (Siebs, 1910) and instead spoke Halunder, the most distinct Frisian language, far from mainland varieties (Århammar, 1995). Today, fewer than 100, and only a few dozen fluent speakers¹ remain (Pichler, 2024), showing a dramatic linguistic shift triggered by historical beatdowns—repeated changes in sovereignty that scattered people across Germany, WWII evacuations, the island's use as a bombing range culminating in the largest non-nuclear detonation, and later post-war rebuilding—followed by the rise of a tourism-centered development (Montigny, 1994a).

The remote island, only one square km large, approximately 70 kilometers off the German coast (

Figure 1.1), and primarily divided into "upper land" and "lower land" (Steensen, 1994), has seen its economy transition from maritime trades, fishing and to tourism (Montigny, 1994a). Visitors crowd the narrow streets in summer, drawn by duty-free shopping², rare wildlife and dramatic red cliffs, then disappear by evening ferries, leaving the community very quiet in the evenings, dreadfully empty throughout the winter months (Wanke, 2008). Cultural markers persist in local traditions, shanties, and historical narratives, but the structural forces threaten Halunder's survival, with a fear of the language becoming an artifact, rather than a means of communication

¹ Estimate provided by the most fluent native speakers of Halunder in 2024.

² Heligoland is tax-free since 1807 and neither part of the customs territory of the European Union nor of the German tax territory.

(Krüss, 1985).

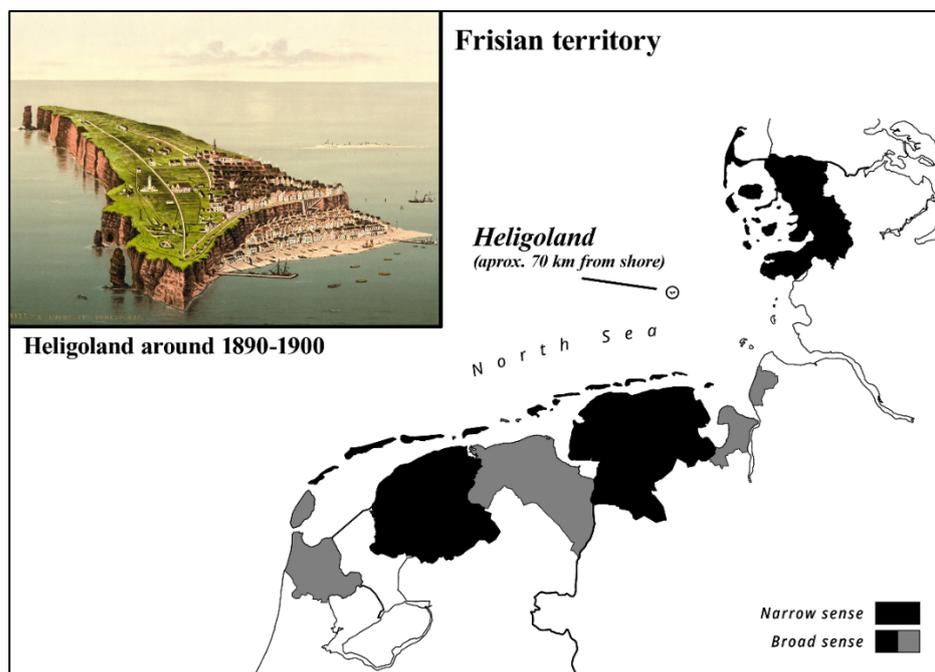


Figure 1.1 Map highlighting Heligoland within the North Sea. Base map adapted from Ætoms (n.d.), CC BY-SA 4.0, with inset historical image of Heligoland (Library of Congress, 1890-1905).

The discontinuation of Halunder as a school subject, limited institutional support, persistent hopelessness, and the absence of modern materials (Grützmacher, 2020) do not paint an optimistic picture. Although legal frameworks not only advocate preservation but also impose obligations to ensure promoting and enabling education, practical solutions remain scarce. Addressing this gap, this thesis investigates in what means Transformer-based neural machine translation can support Halunder’s preservation and revitalization.

1.2 Current State of the Halunder language

When visiting Heligoland, Halunder’s presence is visible in symbolic forms, such as the welcoming sign *Welkoam iip Lun* (“Welcome to Helgoland”) on its landing bridge, but this display masks a bitter reality. Bilingual public signage, street names like *Rekwai* or *Be de Spukkerbu*, and the occasional recycling of worn-out Halunder expressions in tourist brochures create the impression of vitality, yet only the oldest

residents speak it, only a small part of the island understands it. Halunder is not present in daily life (Köhn, 2020). James Krüss, a Halunder native and renowned children’s book author, observed already 40 years ago “his”³ language—once spoken fluently by nearly everyone in 1919—had declined so dramatically that not a single schoolchild said they spoke it when asked (Krüss, 1985).

In the distant past, maritime trades, fishing, and pilot services required Halunder, but a tourism-oriented island, with visitors mostly from Germany, favors German (Quast, 1994). Mixed marriages and economic migration have further accelerated the shift, while post-war evacuation fragmented the community (Wanke, 2008). Currently, there are no structured means for children to acquire Halunder (Grützmacher, 2020). Archival materials, monthly newspaper editions containing a section called *Halunder Spreek* (See Section 3.1.1), and partially completed dictionaries (Krogmann, 1964) persist, but remain underutilized. Without intergenerational transmission or formal teaching, Halunder’s domains of use have narrowed drastically (Wanke, 2008).

While legal frameworks call for minority language preservation (Council of Europe, 2009), neither formal recognition nor sporadic cultural events have led to sustained everyday usage (Nägler, 2010). The community’s linguistic identity is now maintained more as heritage than as a functioning communicative system. Researchers argue revitalization would require greater institutional support, updated teaching materials, and the practical integration of Halunder into daily life, none of which currently exist in sufficient measure (Århammar, 1995; Wanke, 2008).

1.3 Research Objectives

This study aims to develop a proof-of-concept Halunder-German translator using self-attention-transformer based methods (Vaswani et al., 2017). By digitizing existing texts and consulting remaining native speakers, it documents key steps and considerations in creating a first digital parallel corpus suitable for machine learning

³ James Krüss (1926-1997) born on Heligoland, was a German writer of children's and picture books, illustrator, poet, dramatist, scriptwriter, translator, and collector of children's poems and folk songs.

(See Section 3.2). Although not a final version, rather embodying a first documented checkpoint, the translator—once publicly available—may empower the local government, cultural institutions, and educators to reintroduce Halunder more authentically, more accurately and more credibly than it has been for many years. This thesis employs multiple approaches and translation models in applied experiments that demonstrate the training and evaluation of a natural language processing translator. The process, methods and experience should guide similar ultra-low-resource language communities, aiming to illustrate how incremental digital initiatives can support broader revitalization. And finally, it could spark the last initiative in preventing Halunder from becoming a relic of the past (Wanke, 2008).

1.4 Significance

The work documented in this thesis creates what appears to be the first attention-based Transformer applied for the translation task of a language as small and endangered as Halunder (See Section 0). It also documents the creation of the first parallel corpus for Halunder and German sentence pairs suitable for machine learning (See Section 0), as well as the first Halunder machine translator of any kind.

By introducing a digital solution, the thesis is in line with the European Charter for Regional or Minority Languages (Council of Europe, 1992). Its approach addresses educational and cultural mandates that Heligoland has struggled to fulfill, leaving the Halunder minority vulnerable (Council of Europe, 2009; Grützmacher, 2020). Transforming scattered texts and historical materials into an accessible online resource may encourage more active usage among digitally native generations (Cassels, 2019). Similar minority languages, especially related North Frisian dialects, could adapt these methods to meet their own preservation needs with data sovereignty and equity (Galla, 2018). Rather than leaving Halunder’s legacy buried in archives, this translator could extend its presence into reliable everyday communication among the people of Heligoland but also enable the creation of further learning opportunities (Tan & Jehom, 2024).

1.5 Challenges and Contributions

The main research question: “Is Transformer technology the decisive technology for preserving Halunder and other endangered languages?” faces a set of challenges. The idea of developing a translation system that captures the authenticity of the Halunder language is to this day deemed impossible by many natives, who after being approached about the idea in this thesis, reacted with skepticism and caution, but also with hope. Halunder, being separated from the mainland, was seemingly cut off from the rest of the world for centuries, especially in terms of vocabulary development (Wanke, 2008). This separation is not only a geographical or political fact, but also visible in the core of the language as it frequently ignores direct German equivalents, requiring careful interpretation (Köhn, 2020). For example, a phrase that was provided by a natives, used for the word “baby” was *letj Putsch* (“little pit”, as in something wonderful, small, and cherished, much like a tiny fruit pit.⁴) A similar case exists for “kitten”, referred to as *letj Kat* (“small cat”). Terms like *tu beerigen keemen* (literally “to come to prayer”) mean attending confirmation lessons (Århammar, 1994) and expressions like *Ik mai di gearn* (“I like you with pleasure”), Halunder’s only construct for “I love you”, resist literal translation (Krogmann, 1964). Idioms such as *dear lait en Koks bi* (“there’s a sea snail in it”) convey suspicion and “something being fishy” through a locally known fishing reference (Krüss, 1985). Similarly, *Semm* describes a specific rope used in lobster baskets, a word not to be found in English or German vocabulary (Krogmann, 1952a). According to Krogmann, the initiator of the Halunder dictionary (1952), number expressions vary by context and the speaker’s occupation and compounds encode cultural narratives that cannot be translated literally (Wanke, 2008). Halunder, a language that does not contain a basic word like “forest” (*fel boamen* = “a lot of trees”) but yet offers at least three renditions of “wind” (Köhn, 2009). These are not just a isolated exceptions, as some might assume, but

⁴ This is how Halunder natives explained the meaning of the term.

the essence of the language in practice – there are hundreds of idiomatic examples (Siebs, 1910).

1.6 Thesis Structure

Subsequent chapters follow a chronologically similar progression to the development of the translator, describing each practical step needed to establish the first Halunder-German translation system according to scientific standards. Chapter 2 reviews other endangered-language translation attempts and explains why attention-based Transformer architectures uniquely meet Halunder’s contextual and idiomatic requirements. Chapter 3 details the dataset collection process, including selection criteria, digitization and preparation methods, and legal considerations. Chapter 4 presents a total of seven experiments in place to test three hypotheses—model comparisons, hyperparameter tuning, and dataset scaling. Chapter 5 manually evaluates the perceived translation quality, identifying successes and shortcomings, also suggesting recommendations for subsequent improvement. Finally, Chapter 6 evaluates the impact of this development, addresses constraints, and concludes the weight of the discoveries on the minority language preservation industry. In total, offering a comprehensive, step-by-step workflow for readers interested in replicating or adapting the methodology.

Chapter 2

Theoretical Framework

This chapter begins with a historical background (Section 2.1), outlining the critical endangerment of Halunder and its classification within EGIDS. It reviews literature on machine translation architectures (Section 2.2), detailing the evolution from rule-based to neural approaches, including transformer models. Section 2.3 focuses on transfer learning, emphasizing its potential for low-resource languages like Halunder. Section 2.4 provides a historical and linguistic overview of Halunder and its relationships with related languages. Section 2.5 highlights relevant endangered language projects and develops the conceptual framework for the study, positioning Halunder within the broader context of endangered language translation efforts.

2.1 The Need for a Translator

2.1.1 Critical Endangerment

The Expanded Graded Intergenerational Disruption Scale (Lewis & Simons, 2010) provides a standardized model for language vitality assessment and is grounded in sociolinguistic research on intergenerational transmission, speaker demographics, literacy, and institutional support (Fisherman, 1991; Lewis & Simons, 2010; UNESCO Language Vitality and Endangerment, 2003). It builds upon Fishman's original GIDS but offers more granular stages. Linguists recognize and use EGIDS to identify

degrees of endangerment and determine appropriate revitalization measures (Bird, 2020).

This thesis places Halunder near EGIDS Level 8b (Nearly Extinct), based on current speaker demographics, usage patterns, and lack of intergenerational transmission (Wanke, 2008). The Halunder speaker base consists of less than 100, mostly elderly individuals, with only a handful writing sporadically and no consistent child acquisition (Pichler, 2024; Wanke, 2008). Although a small number of individuals attempt to use the language, overall community engagement is severely limited, according to Wanke's empirical findings (2008). Halunder is almost never heard in daily conversation (Köhn, 2020). The restricted usage occurs in small, occasional gatherings—facilitated by circumstance rather than choice—due to Heligoland's tiny geography, where everything is reachable on foot within 5 minutes (Wanke, 2008).

The given conditions align closely with the criteria for EGIDS 8b, more than 8a indicating that Halunder has reached a critical level of endangerment (Lewis & Simons, 2010). The mostly elderly speakers are “so few [...] that they have little opportunity to use the language with each other” (Århammar, 1995; Lewis & Simons, 2010), reflecting the defining characteristics of 8b. Once a language has reached this stage, conventional revitalization—such as reintroducing it as a first language between family members (Fisherman, 1991) becomes nearly impossible without extreme measures (Bradley & Bradley, 2002). The immediate priority shifts to documentation, that linguistic features such as grammar, phonetics and lexicon are thoroughly recorded for potential future revival (Fisherman, 1991). As a byproduct of material collection for dataset curation (See Chapter 3), the development of a machine translator, as proposed in this thesis, itself becomes a means of preserving existing records, aligning with the recommendations for critically endangered languages (Kuhn et al., 2020). As Fishman (1991) underscores, once the natural transmission of a language ceases and remaining speakers dwindle to a negligible few, immediate intergenerational revitalization may be impractical. Instead, maintaining a well-organized linguistic and cultural record ensures that subsequent generations—should they choose—have the tools to attempt a revival.

Table 2.1 EGIDS Levels (adapted from Lewis & Simons, 2010)

Expanded Graded Intergenerational Disruption Scale (adapted from Fishman 1991)			
LEVEL	LABEL	DESCRIPTION	UNESCO
0	International	The language is used internationally for a broad range of functions.	Safe
1	National	The language is used in education, work, mass media, government at the nationwide level.	Safe
2	Regional	The language is used for local and regional mass media and governmental services.	Safe
3	Trade	The language is used for local and regional work by both insiders and outsiders.	Safe
4	Educational	Literacy in the language is being transmitted through a system of public education.	Safe
5	Written	The language is used orally by all generations and is effectively used in written form in parts of the community.	Safe
6a	Vigorous	The language is used orally by all generations and is being learned by children as their first language.	Safe
6b	Threatened	The language is used orally by all generations but only some of the child-bearing generation are transmitting it to their children.	Vulnerable
7	Shifting	The child-bearing generation knows the language well enough to use it among themselves but none are transmitting it to their children.	Definitely Endangered
8a	Moribund	The only remaining active speakers of the language are members of the grandparent generation	Severely Endangered
8b	Nearly Extinct	The only remaining speakers of the language are members of the grandparent generation or older who have little opportunity to use the language.	Critically Endangered
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community. No one has more than symbolic proficiency.	Extinct
10	Extinct	No one retains a sense of ethnic identity associated with the language, even for symbolic purposes	Extinct

2.1.2 Institutional and Legal Obligations

According to the Protection of National Minorities Framework Convention (Council of Europe, 2009), “the protection of national minorities is an integral part of the protection of human rights.” Consequently, the development of language technology tools, particularly translators, represents not just a practical measure but an obligation.

The European Charter for Regional or Minority Languages (1992) mandates that signatory states must “create and/or promote and finance translation and terminological research services, particularly with a view to maintaining and developing appropriate administrative, commercial, economic, social, technical or legal terminology in each regional or minority language.” Other obligations extend to public use, offering education, and using minority languages in administration, media, and all aspects of public life (Council of Europe, 1992). As the Charter emphasizes, minority languages can only survive “if they are used everywhere and not just at home” (Council of Europe, 1992). The development of digital language tools as explored by this thesis, directly supports this mandate by aiming to enable language use across all domains.

At the national level, Germany has incorporated these obligations into its legal framework. Schleswig-Holstein's implementation is most relevant, with its constitution explicitly guaranteeing that “The cultural independence and political participation of national minorities and ethnic groups are protected by the state”⁵ (Verfassung SH, Art. 6.2). In an official statement to the Parliament, Grützmacher (2020) emphasizes the urgent need to “create conditions for current teaching materials in both classical-analog and digital form”⁶, specifically highlighting digital resources as crucial for modern language preservation.

The already introduced EU Minority Preservation Framework declares the ability to

⁵ Original German: „die kulturelle Eigenständigkeit und die politische Mitwirkung nationaler Minderheiten und Volksgruppen stehen unter dem Schutz des Landes“ (Author's translation)

⁶ Original German: „in buchstäblich allen Bereichen des öffentlichen Lebens“ (Author's translation)

learn minority languages as one of the principal means by which these persons can maintain and develop their identity (Council of Europe, 2009). This requirement for states to provide opportunities for language learning necessitates modern technological solutions, especially considering the practical challenges further outlined by Grützmacher (2020) regarding the difficulties in sourcing traditional teaching materials, as even small publishers cannot meet minimum print runs. This has led to a situation where “almost all teachers must create their own materials”⁷ (Grützmacher, 2020)

2.2 Translation Architecture

The following two sections establish the theoretical foundations necessary for developing a modern translator. They introduce essential concepts and terminology from the technological field, ranging from Neural Machine Translation (NMT) and Natural Language Processing (NLP) to specific architectures like Statistical Machine Translation (SMT), Recurrent Neural Networks (RNN), Gated Recurrent Units (GRU), and Transformer models, as well as techniques such as machine learning, self-attention mechanisms, Transfer Learning, and fine-tuning.

2.2.1 Beginnings of Machine Translation

Early research on machine translation started with rule-based methods that attempted to encode linguistic structures, grammatical patterns, and bilingual dictionaries into explicit sets of instructions (Hutchins, 1995). While these approaches established a starting point, they did not handle language complexity well (Hutchins, 1995). For example, when words have multiple meanings (polysemy), and languages rely on subtle contextual hints, rule-based systems did not comprehend such nuances, resulting in unnatural or inaccurate outputs (T. Luong et al., 2015).

⁷ Original German: "fast alle Lehrer müssen ihr eigenes Material erstellen" (Author's translation)

Statistical machine translation offered a different approach by using large "parallel corpora" consisting of aligned text pairs where each entry or sentence in one language corresponds to its translation in another (Karakanta et al., 2018). Statistical methods estimated translation probabilities based on word frequency, alignment patterns, and how often certain words appeared together (Cho et al., 2014). In simpler terms, the translation often depended on what was statistically more common in the training data, not on what the context might subtly reveal (Cho et al., 2014). While this improved upon the rigidity of rule-based methods, statistical MT struggled in cases with limited data, especially for underrepresented languages (van der Wees et al., 2016).

2.2.2 Early Neural Architectures

A meaningful improvement came with neural machine translation (NMT) models, which relied on encoder-decoder⁸ neural networks (Sutskever et al., 2014). Early neural approaches employed recurrent neural networks (RNNs) or gated recurrent units (GRUs), which processed a sentence step by step (Cho et al., 2014). Compared to SMT, these models learned more complex patterns and were less rigid. They could capture some context by maintaining hidden states over time, allowing for generally better translations (Bahdanau et al., 2016).

At the same time, RNNs had limitations. When handling long sentences, they had difficulty remembering details from far earlier in the text (Vinyals et al., 2015). Adding layers or using bidirectional encoders (T. Luong et al., 2015) helped, but there was no direct mechanism to highlight crucial parts of the input. For languages similar to Halunder, where many words have multiple contextual meanings and distinct situational dependencies (Krogmann, 1952a) these model's translations lacked precision (T. Luong et al., 2015). They might, for example, fail to choose a situationally dependent term if the relevant contextual information was mentioned in a previous sentence, very subtly or far away in the text.

⁸ More on Encoder and Decoder in Section 2.2.3.

2.2.3 Attention is All You Need – The Transformer Model

The transformer architecture as introduced in the groundbreaking “Attention is All You Need” (Vaswani et al., 2017), provides a revolutionary strategy for machine translation and related tasks. Instead of processing words one by one like previous sequential models (Sutskever et al., 2014), the transformer examines all words at the same time, making it easier to find dependencies between distant parts of a sentence (Vaswani et al., 2017). At the core of the Transformer architecture is the use of “attention” and, defining it, “self-attention” (Luong et al., 2015; Vaswani et al., 2017). Attention enables the model to focus on relevant segments of a different sequence, while self-attention enables it to focus on all parts of the same input sequence. An example is provided in

Figure 2.1 A Conceptual Visualization of Self-Attention on a Halunder example.

Before a Transformer begins to translate, it processes input text through tokenization, where each word or sub-word (token) is converted into a high-dimensional numerical vector—typically 512 dimensions—forming rows in a matrix that represents the semantic features of the sequence (Vinyals et al., 2015). These dense vector representations encode complex linguistic properties and contextual relationships learned from large-scale training data. The transformer architecture then uses multi-head self-attention mechanisms to compute token interdependencies across the entire sequence in parallel (Vaswani et al., 2017).

The **encoder** uses self-attention to simultaneously evaluate relationships between all tokens, enabling the model to mathematically disambiguate meaning based on full contextual information. For instance, when processing polysemous⁹ words, the self-attention mechanism can determine appropriate interpretations by considering the entire surrounding context at once. Each layer refines these token representations through feed-forward neural networks, incorporating layer normalization and residual connections, maintaining training stability (Vaswani et al.,

⁹ Polysemy is the capacity for a sign (such as a word) to have multiple related meanings.

2017).

The **decoder** implements two critical modifications: masked self-attention to enforce sequential generation in the target language, and encoder-decoder attention (Cho et al., 2014) to maintain access to the source sequence's semantic representation. This architecture enables the transformer to process input sequences as a whole rather than sequentially, leading to increased efficiency in capturing complex linguistic patterns compared to traditional sequential models (Vaswani et al., 2017).

Application to Halunder

Applying theory to the case of Halunder, **Figure 2.1** presents a conceptual visualization of a transformer's self-attention mechanism that can find relationships between words. For instance, it can determine whether *Paik* or *Pakkloor* is the correct translation of “garb”¹⁰, a term that English or German treats neutrally but Halunder encodes with gender-specific detail (Krogmann, 1964).

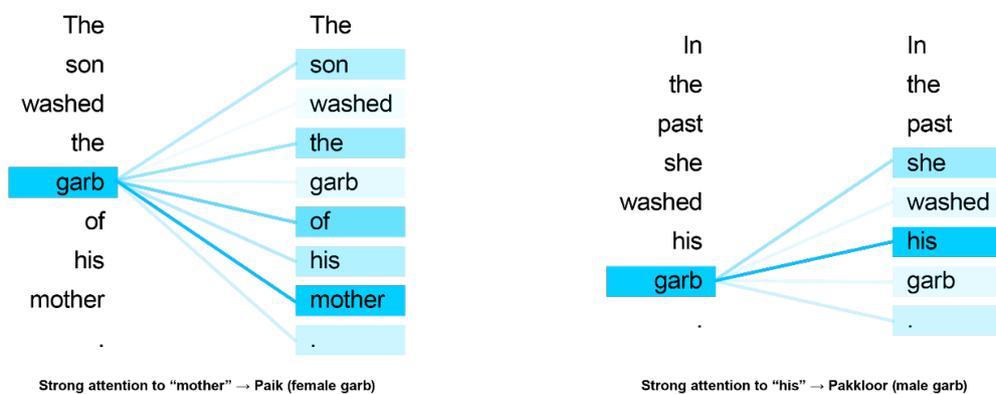


Figure 2.1 A Conceptual Visualization of Self-Attention on a Halunder example.

In a context where the source sentence describes a mother and her son's traditional attire, Halunder distinguishes *Paik* (a female-specific garment featuring a red skirt) from *Pakkloor* (the male-specific equivalent) based on the person wearing it (Krogmann, 1964). German uses the single term *Tracht* for both garments, while

¹⁰ garb in this context refers to traditional attire, often worn for special occasions or cultural representation.

Halunder distinguishes between gender:

German	The son washes the traditional garb of his mother. In the past his mother always washed his traditional garb .
English	De Seen wasket de Paik fan siin Mem. Feerdem hat ümmer siin Mem siin Pakkloor wusken.
Halunder	Der Sohn wäscht die Tracht seiner Mutter. Früher hat seine Mutter immer seine Tracht gewaschen.

Here, the transformer benefits from self-attention to locate contextual cues—whether they appear several sentences earlier or are only subtly implied—enabling it to select the correct term (Vaswani et al., 2017).

2.3 Transfer Learning

Transfer learning in NMT involves using knowledge gained from one language pair to improve translation performance on another (Zoph et al., 2016). This technique has shown consistent success for languages with small, limited datasets (Gu et al., 2018). The most common application is “fine-tuning” where a pre-trained model undergoes additional training on an additional parallel corpus (Kozhircbayev, 2024).

Studies show particularly good results between related languages (Nguyen & Chiang, 2017), suggesting a model trained on German-Dutch translation could provide valuable initial knowledge for German-Halunder¹¹ translation. Since Halunder belongs to the North Germanic language family, sharing roots and “borrowing” words from German, (Old-)English, and Danish (Montigny, 1994b), as explored in Section 2.4, the model can apply linguistic knowledge from these languages, for which large-scale training data exists (Galiano-Jiménez et al., 2023).

The fine-tuning process extends established neural network training methods (Sutskever et al., 2014). The model begins with parameters already optimized for

¹¹ and vice versa.

translation tasks rather than random initialization (Nguyen & Chiang, 2017). During training, it processes sentence pairs, evaluates a models translations against “expected” sentences provided in a separate validation set, and adjusts its internal parameters accordingly, restarting the process (Galiano-Jiménez et al., 2023). Transfer learning proves effective as the pre-trained model already understands many aspects of other languages and can recognize similar patterns despite vocabulary differences (Zoph et al., 2016).

While this approach has demonstrated success for many low-resource languages (Gu et al., 2018) one important consideration for Halunder exists. The author James Krüss (1985) lamented that younger generations had already started to adopt what he referred to as “German language-thinking” patterns, moving away from traditional Halunder thinking, which often uses distinct sentence constructions and relies heavily on circumlocution¹². This phenomenon observed in the Halunder (Krogmann, 1952a) might extend to machine learning systems (Galiano-Jiménez et al., 2023), potentially biasing translations away from Halunder's traditional patterns.

Research shows transfer learning's strengths between languages with different vocabularies but similar structures (Nguyen & Chiang, 2017). Despite room for drawbacks, transfer learning remains central to translating languages with limited data (Kozhirbayev, 2024).

2.4 Historical Development and Similarities to other Languages

Krogmann (1952b) states, “Heligoland has changed hands more often than almost any other place on Earth.”¹³ With the island having shifted sovereignty at least thirteen times (Krogmann, 1952b), its language has similarly undergone many external influences (Århammar, 1985). *Deät Halunder*, also referred to as *Helgoländisch* in German, is officially classified as a North Frisian dialect, yet based on the

¹² circumlocution involves describing a concept indirectly due to the absence of a single, direct term.

¹³ Original German: "Wie kaum ein anderes Fleckchen Erde hat Helgoland seinen Besitzer gewechselt." (Author's translation)

distinctiveness of its linguistic form and system, one could equally justify calling it an independent language (Århammar, 1985). According to Montigny (1994b), Halunder is “without doubt” one of the smallest languages in Europe, and the island community has always treated it as such - an independent language (Wanke, 2008).

Alongside other Insular North Frisian dialects Feering, Öömrang, and Sölrang, Halunder diverged from mainland Frisian in the 8th or 9th century, developing significant differences that set island variants apart (Århammar, 1985). Within the North Frisian context, the Heligoland dialect is considered unique, both culturally and in terms of its vocabulary, tied to the island's identity and distinct from mainland Frisian traditions (Montigny, 1992). While Dutch and Low German adapted to inland influences during the Middle Ages, Frisian in Europe maintained its North Sea Germanic character (Århammar, 1985). It stands as the only language of an independent Germanic tribe to survive without a national state (Århammar, 1993).

The first written examples of Halunder appeared in the 18th century, initially in pilot examination handbooks and poetry (Århammar, 1985). Though traditionally spoken for hundreds of years (Nägler, 2010), its grammatical system and rules were not formally established until the 1970s (Montigny, 1992). Still, over centuries, a handful of proficient authors produced a substantial body of literature, including shanties, poems, and well over a thousand pages of printed narratives (Århammar, 1985).

The language's development reflects its complex history of external influences. Danish left a distinct imprint on Halunder (Århammar, 1985), as it did in English, visible in terms such as *Knüf* (“knife”) and *Skear* (“to cut”). Political and economic shifts led to strong German influences, reflected in words like *gearn* (gern), *Hemt* (Hemd), *Leepel* (Löffel), and *Skooster* (Schuster)¹⁴ (Århammar, 1985). The establishment of the Heligoland sea resort in 1826 and eventual inclusion into the German Empire in 1890, along with the expulsion and dispersal of its population during and after the World Wars, intensified the integration of High German

¹⁴ gern – with pleasure; Hemd – shirt; Löffel – spoon; Schuster – shoemaker

vocabulary (Montigny, 1985).

Despite these influences, Halunder's Ingvaemonic (North Sea Germanic) roots remain visible. Due to its common origin with Old English, Halunder is more closely related to English than German (Wanke, 2008). For those who speak the combination of German, Low German, and English, Halunder texts may be partially accessible, allowing for some independent navigation of written materials (Århammar, 1985).

As described in *Wi lear halunder*¹⁵ (p.84), it does not rely on four cases like German, making its noun and article usage more straightforward. Instead, it is limited to two articles: *de* for non-neuter nouns and much rarer *deät* for neuter nouns. Its pronoun forms align more closely with English than with German's complex, case-dependent system (Århammar, 1985). The characteristics, along with the historical English and Danish influences, reflects Halunder's preservation of an older and simpler grammatical system. Related Work (other similar projects)

2.5 Related Work

Machine translation for endangered languages has advanced significantly, focusing on addressing challenges such as small datasets, low speaker populations, and unique linguistic features. Research on languages like Kalamang, Erzya, and European Portuguese illustrates how varied conditions influence approaches and outcomes.

The Kalamang project (Tanzer et al., 2024) centered on a language with fewer than 200 speakers, primarily oral, and lacking any substantial web presence. This work introduced the Machine Translation from One Book (MTOB) framework, utilizing a single grammar book and a small parallel corpus of under 2,000 sentences. The goal was to teach models linguistic features through curated explanations, rather than traditional large-scale training corpora. Despite achieving chrF scores of 44.7 and

¹⁵ "Wi lear Halunder" (1987) is the island's only textbook, written by Mina Borchert, Ritva & Nils Århammar

45.8 for translation to and from English, this was notably lower than human translators working with the same materials. The study highlighted the difficulties of working with datasets of less than 250,000 tokens in such severely constrained settings (Tanzer et al., 2024). While Kalamang’s speaker base is extremely small, the availability of a comprehensive grammar book provided a foundation for this innovative translation approach.

For Erzya (Dale, 2022), spoken by around 300,000 people in Russia, a first neural machine translation system was developed using approximately 12,000 parallel sentences derived from religious texts, online bilingual dictionaries, and aligned books. An additional 21,000 sentence pairs were mined from monolingual sources using sentence alignment algorithms. These efforts yielded BLEU scores of 17 and 19 for translation to and from Russian, and over half of the outputs were rated acceptable by native speakers. The project relied heavily on fine-tuning multilingual pre-trained models, back-translation, and transfer learning (Dale, 2022). Unlike Kalamang, Erzya benefited from a larger speaker base and written tradition, enabling more substantial data collection efforts. However, its classification as “definitely endangered” by UNESCO underlines the urgency of technological intervention for its preservation.

European Portuguese (EP) presents a different challenge. Although not endangered, EP suffers from underrepresentation compared to its closely related variant, Brazilian Portuguese (BP). Researchers leveraged TED Talks and subtitles to fine-tune models such as mBART-50 and M2M100 for EP-BP translation tasks, alongside manual curation of evaluation datasets. These efforts addressed discrepancies in linguistic representation within the Portuguese-speaking world. The project highlighted the advantage of shared linguistic resources between EP and BP, which facilitated dataset creation and model training (Sanches et al., 2024). This contrasts with languages like Erzya or Kalamang, which lack such interlingual proximity to a larger, well-resourced language.

Halunder, with fewer than 100 speakers, is even more vulnerable than Kalamang or

Erzya. Its speaker population is highly scattered, and there is no significant written tradition or digital presence. Unlike Kalamang, which benefited from a grammar book, or Erzya, which had access to religious texts and bilingual dictionaries, Halunder has no substantial corpus to draw from. Its endangered status necessitates novel data generation strategies and collaboration with the remaining speaker community to construct even a minimal linguistic resource. The absence of such foundational materials places Halunder among the most severely underserved languages in terms of computational representation. This positions it as an extreme case within the broader "low-resource" language category, where resource availability and linguistic proximity to larger languages vary widely.

2.6 Summary of the Theory

Halunder's critical endangerment, classified at EGIDS Level 8b, reflects its negligible intergenerational transmission and limited speaker population of fewer than 1,000 individuals. Institutional frameworks, including the European Charter for Regional or Minority Languages, emphasize the need for documentation and technological solutions to preserve the language.

Modern machine translation technologies, such as transformer models and transfer learning, offer promising methods to overcome the data limitations faced by Halunder. Insights from other endangered language projects, such as Kalamang and Erzya, illustrate the potential of combining linguistic documentation and machine learning to address resource scarcity. However, Halunder's lack of foundational resources makes it one of the most critically underserved languages, necessitating urgent and innovative efforts for its preservation.

Chapter 3

Proposed Approach

The proposed approach of this thesis in its essence relies on the creation of a community-driven parallel corpus rather than depending on linguistic experts that are not available within the community (Wanke, 2008). By assembling and digitizing analog texts, and involving community involvement, this thesis collects accurate, context-rich parallel sentences in German and Halunder, enabling a Transformer model (Vaswani et al., 2017) to autonomously learn translation patterns and linguistic nuances. Regardless of the final training configuration and implementation (See Chapter 4), the development of a dataset consisting of parallel translations of Halunder and German sentences remains a strong foundation for the development of a machine translation system (Gu et al., 2018) and therefore plays a critical role. Subsequent sections detail the dataset development, including data requirements, availability, collection, preparation, as well as legal and ethical considerations. The second part of this chapter analyzes the data collected. Ultimately, the proposed method aims to empower minority communities to actively participate in the preservation and revitalization of their language through the provision of raw data for effective NMT training.

3.1 Dataset Development

3.1.1 Data Availability

Halunder faces considerable challenges in building digital corpora. Historically, it has remained primarily a spoken language, which limited digital written output (Wanke, 2008). Early preservation efforts note a scarcity of standardized materials, with the first grammatical frameworks emerging in the 1970s (Århammar, 1985). Despite the existence of teaching materials like *Wi lear Halunder*¹⁶—Halunder’s only textbook that is not available anymore—and a small selection of children’s books by James Krüss, there remains a shortage of large-scale, contemporary written content (Wanke, 2008). Apart from that, while a lot of material about the island exists, much less literature is written in its language (Århammar, 1985). As a result, digital Halunder datasets do not exist. Material is often scattered across analog publications spreading centuries and cultural narratives that have never been published (Montigny, 1994a). Even in modern times, more than 85% of Halunder speakers are not capable of reliably writing in Halunder, with the primary reason being the evolution of the grammar system that many did not accept because it did not match their spoken tradition (Wanke, 2008). Today the number is even smaller¹⁷. Fortunately, the “*Halunder Spreek*” series, initiated in 1974, published well over 360 texts written by the most proficient Halunder authors of the previous generation, covering personal stories, fictional short-stories, and comments on local developments—documenting culture, traditions and typical moments (Montigny, 1985). Favoring the effort of developing a digital translation system, all texts were adjusted for, or written in, the new grammar framework (Montigny, 1994c). While most of the published materials are monolingual, most publications contain footnotes for set expressions and contextual importance. Only some publications contain a parallel German version (Wanke, 2008).

¹⁶ “*Wi lear Halunder*” (“*We Learn Halunder*”) was the first and only textbook for learning Halunder, published in 1987 by Mina Borchert, Ritva & Nils Århammar. It is not produced nor sold anymore.

¹⁷ Natives estimate the number of proficient writers to only a handful in 2024.

3.1.2 Data Collection and Preparation

The foundation of the formation of a parallel corpus relied on digitizing materials using an on-desk book scanner with optical character recognition (OCR). The digitalized texts were processed using LLM Claude 3.5 Sonnet (Anthropic, 2024) to decompose book pages and texts into aligned sentence pairs. A model by Anthropic was selected for this task due to its data privacy guarantees¹⁸ and its ability to “process sensitive cultural materials without retaining or training on the content” (Anthropic, 2024). The model leveraged its ability to analyze parallel content in both languages, identifying matching sentence boundaries based on semantic and structural patterns while standardizing German spelling variants to modern conventions. Native speakers provided ongoing support throughout the data collection process, offering crucial guidance on language-specific questions and the verification of translations.

Each entry underwent verification for minimum length requirements¹⁹ and parallel text alignment, sentence length differences between languages of at least 15% were flagged for review. Punctuation consistency was checked, with particular attention to the frequency and placement of common marks across both languages. Special attention was given to Halunder-specific character combinations, such as the misinterpretation of *Hallem* (the neighboring, small island) as *Hallem* and confusion between *ü* and *ii*, or *lip* (*iip*, with a capitalized i) and *lip* (mistaken as a lowercase L). The process included screening for unexpected characters that fell outside the allowed set of linguistic and formatting symbols²⁰.

Additional quality control measures included the verification of consistent first-letter capitalization between source and target sentences, monitoring for matching numerical formatting (e.g. 20 vs. twenty), and the identification of abbreviations that might indicate incomplete sentences. To protect against potential copy-paste errors,

¹⁸ Anthropic's privacy policy specifies that Claude does not retain or train on user data processed during inference (Anthropic, 2024).

¹⁹ Minimum length required for a sentence was defined at 2 words.

²⁰ Halunder doesn't contain z (replaced by ts), x, v (replaced by f).

a similarity threshold check was implemented to flag a character overlap of 90%, marking almost identical texts between languages.

Native speakers reviewed a substantial portion of the translations, especially those flagged, providing verification and corrections where needed, with a particular focus on grammar, authenticity and linguistic quality. Metadata was documented for each sentence pair entry, enabling source verification, potential weighting and further analysis. For model training, the sentence rows were shuffled, and a 250-sentence pair subset was extracted for review and unbiased analysis purposes. It is ensured that this subset was never part of any training or validation process.

3.1.3 Data Overview

The establishment of the parallel corpus began with a thorough assessment of existing digital bilingual materials. Unsurprisingly, the only digital sources were four narrative texts published by Nils Århammar²¹. The texts provided foundational insights into parallel text structures between German and Halunder.

Table 3.1 Distribution of dataset sources.

Category	Subcategory	Entry Pairs
Teaching Materials	Language Textbook	840
	Native Examples	240
	Grammar Samples	230
	Study Materials	110
Dictionaries	Krogmann Dict.	6,260
	Århammar Dict.	260
Literature	Translations	550
	Island Life Stories	1,200
Cultural Sources	Historic Texts	260
	Personal Stories	150
Miscellaneous	Various Sentences	130
Total		10,250

²¹ Nils Århammar published these texts on his website www.helgolaendisch-halunder.de

The remaining sentences were sourced from printed materials. Stemming from multiple categories, a total of 10,250 Halunder and German sentence pairs have been digitized, reviewed, and documented. The sentence pairs consist of teaching materials, comprising 1,440 sentences (14% of the corpus), including content from the "*Wi lear Halunder*" textbook, native speaker contributions, sentences exemplifying grammar structures, and additional study materials. Dictionary-derived content represents the largest portion with 6,520 sentences (63.6%), primarily drawn from the Krogmann Dictionary²² examples and Århammar Dictionary²³ entries. Literary translations contribute 550 sentences (5.4%), featuring works such as the translations "*Wi Künner fan Bullerbü*"²⁴ and "*Heäwelman*"²⁵. Cultural materials add 1,610 sentences (15.7%), including the "*Snakketaffel*"²⁶ book, a collection of peer-reviewed, bilingual stories by natives, but also historical documents, and personal narratives. The remaining 130 sentences (1.3%) come from miscellaneous sources such as greetings and single scattered sentences. The corpus achieves substantial coverage through its diverse materials. Teaching materials provide fundamental language constructs and everyday vocabulary, while dictionary-derived content ensures thorough coverage of standard usage patterns, expanding vocabulary further, especially for single words in multiple contexts. Literary translations introduce more complex linguistic structures and creative or fictional language use. Cultural materials capture authentic sentence structures in historical, island-related, maritime, and traditional contexts. The miscellaneous sources add variety through informal communications, specialized vocabulary and fixed greeting clauses.

²² Helgoländer Wörterbuch. Lfg. 4. harti – Kortel (1964) by Willy Krogmann

²³ Deutsch-Helgoländisches Wörterbuch (1997-2021) by Ritva and Nils Århammar

²⁴ *Wi Künner fan Bullerbü iip Halunder* (1998) by Bettina Köhn, Berta Martens and other natives, a translation of Astrid Lindgren's "The Six Bullerby Children" (1947)

²⁵ *De letj Heäwelman iip Halunder* (1986) by Mina Borchert, a translation of Theodor Storm's "Der kleine Häwelmann" (1850)

²⁶ *Snakketaffel* published by Bettina Köhn in 2018, contains "short stories" as written by Halunder natives. Each story is published with a peer-reviewed Halunder text and German translation

3.1.4 Legal and Ethical Aspects

The Halunder-German dataset carries significant ethical weight, as it has the potential to directly influence the future of the language by establishing a de facto “standard” for translation. This responsibility demands accountable decisions and measures to preserve the cultural accuracy and integrity of Halunder. Translation outputs will likely shape perceptions of what constitutes “correct” or “authentic” usage (Galla, 2018). To prevent “vocabulary modernization” from eroding cultural identity—a concern frequently discussed by Halunder speakers (Montigny, 1994c)—the most fluent and experienced native speakers in both Halunder and German were directly involved in dataset creation and had ultimate authority over language-shaping decisions. These contributors validated entries, resolved uncertainties, and collaboratively determined “predominant” forms for terms, ensuring authenticity and cultural integrity.

From a legal standpoint, the project operates within a complex legal landscape. While text and data mining exceptions under §44b and §60d UrhG permit non-commercial scientific research and the automated evaluation of copyrighted material, very recent critiques argue that generative AI models often infringe on §16 UrhG by memorizing of syntactic and semantic structures from copyrighted works and inherently reproducing protected content (Dornis & Stober, 2024). However, the approach in this thesis differs fundamentally from generative AI models. The dataset consists of isolated, decontextualized sentences, not entire works in their original form, ensuring no direct reproduction or memorization of copyrighted material occurs. Personal materials were included only with explicit consent and without sensitive information, and no texts that explicitly prohibit its usage for data mining purposes were used. Further, the dataset remains private, avoiding public communication under §19a UrhG. The translator's technical process also complies with §44a UrhG, which permits temporary reproductions integral to lawful operations such as data mining.

3.2 Dataset Analysis

The following section explores the substance of the established dataset of over 10K sentence pairs established before. It offers an overview of domain coverage, sentence length distribution, total word counts, most frequently occurring words, and typical word lengths.

3.2.1 Characteristics and Distribution

The initial analysis of the collected dataset reveals distinctive patterns in word usage and structure between Halunder and German. While this dataset does not represent the entire vocabulary of either language, both contain approximately 92,000 words and around 15,000 unique words each.

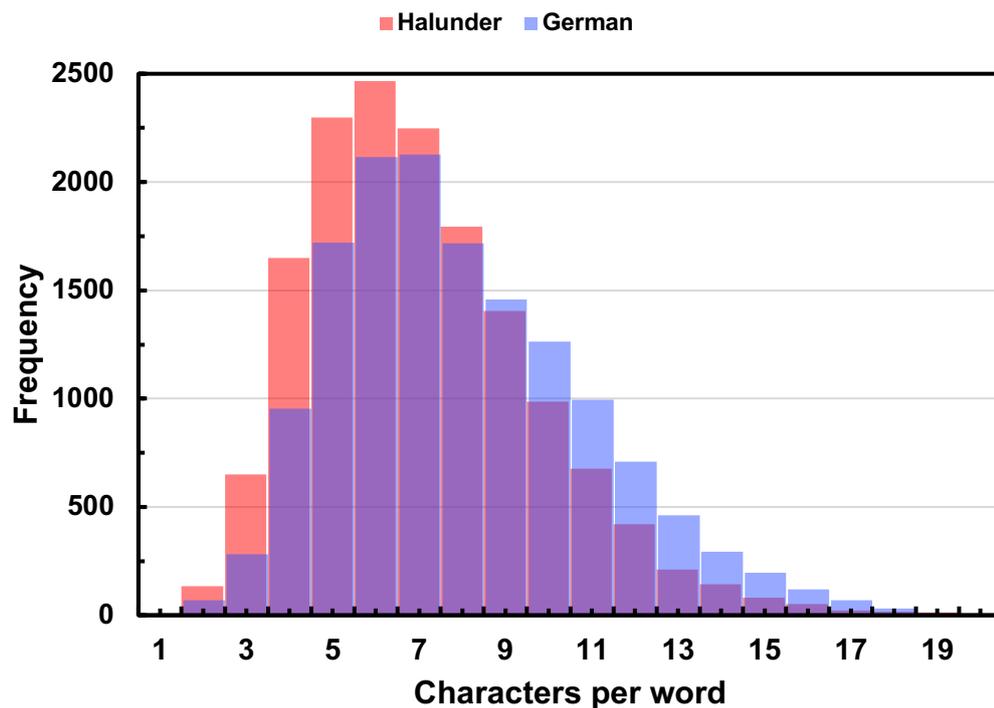


Figure 3.1 Characters per unique word in Halunder and German Parallel Corpus

A detailed examination of the word length distributions across unique words, as visualized in **Figure 3.1**, confirms that Halunder words are generally shorter than German words. In this dataset, the median length for Halunder is about 7 characters,

while German’s median is around 8. Halunder shows a pronounced concentration of words between 2 and 6 characters, with its most common length at 6 characters. In contrast, German words not only peak at 7 characters but also extend more frequently into lengths beyond 9 characters. While German words in the dataset reach up to 25 characters (e.g., *schreibmaschineschreiben*), Halunder’s longest entries are shorter, reaching a maximum of 23 characters (e.g., *blachoawerdibeerlainers*). Word frequency analysis further shows the fundamental differences between the two languages.

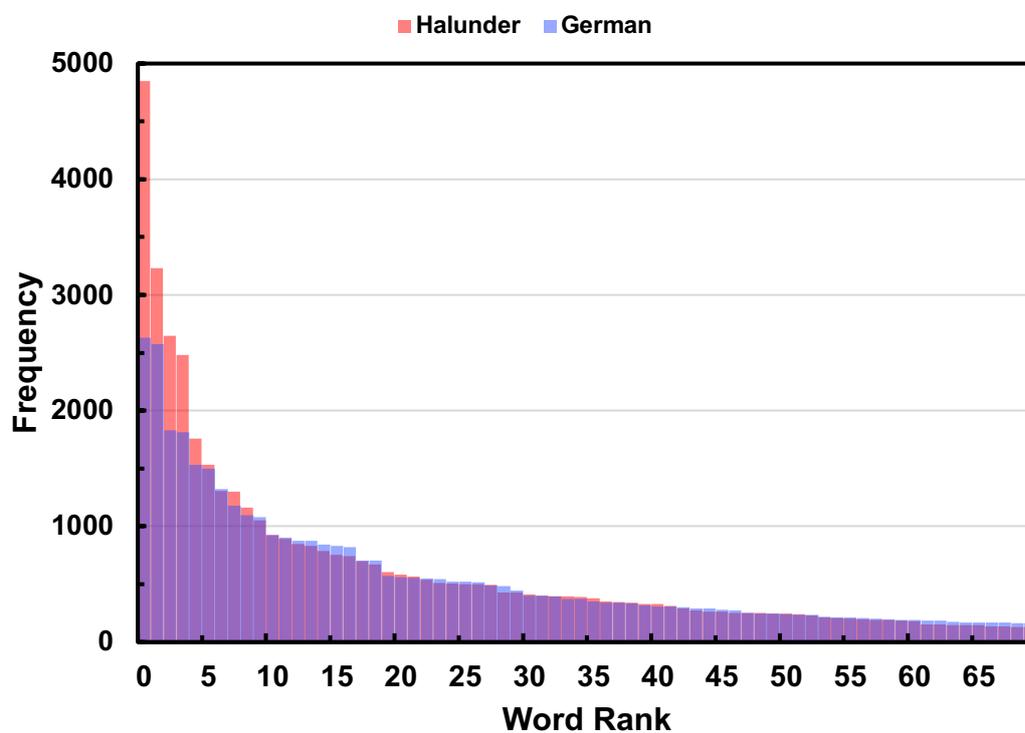


Figure 3.2 Frequency of Top 70 Halunder and German Word Distribution

The frequency analysis of the 70 most commonly appearing words reveals notable differences in usage concentration between the two languages. Halunder exhibits a heavier reliance on its top words, with *en* (“and”) appearing in 5.3% of the dataset. By contrast, German’s equivalent word *und* appears at only 2.9%. This difference continues throughout the most common words: Halunder’s top three words (*en*, *di*, *deät*) = “and”, “you” and “that”, each exceed 2.8% of all word usage, whereas only the top two German words (*und*, *die*) = “and” and “the” reach that level.

The common use of *en* suggests that Halunder prefers coordinating clauses directly with a conjunction rather than relying on punctuation rules. This observation aligns with its historical status as a primarily spoken language (Århammar, 1985). The further differences are likely attributable to the simpler grammar system that relies on only two articles, no cases and simple pronouns (Århammar, 1985).

Overall, the concentration demonstrated in

Figure 3.2 proves that Halunder relies more heavily on a core set of common words for sentence construction, which has important implications for translation model design.

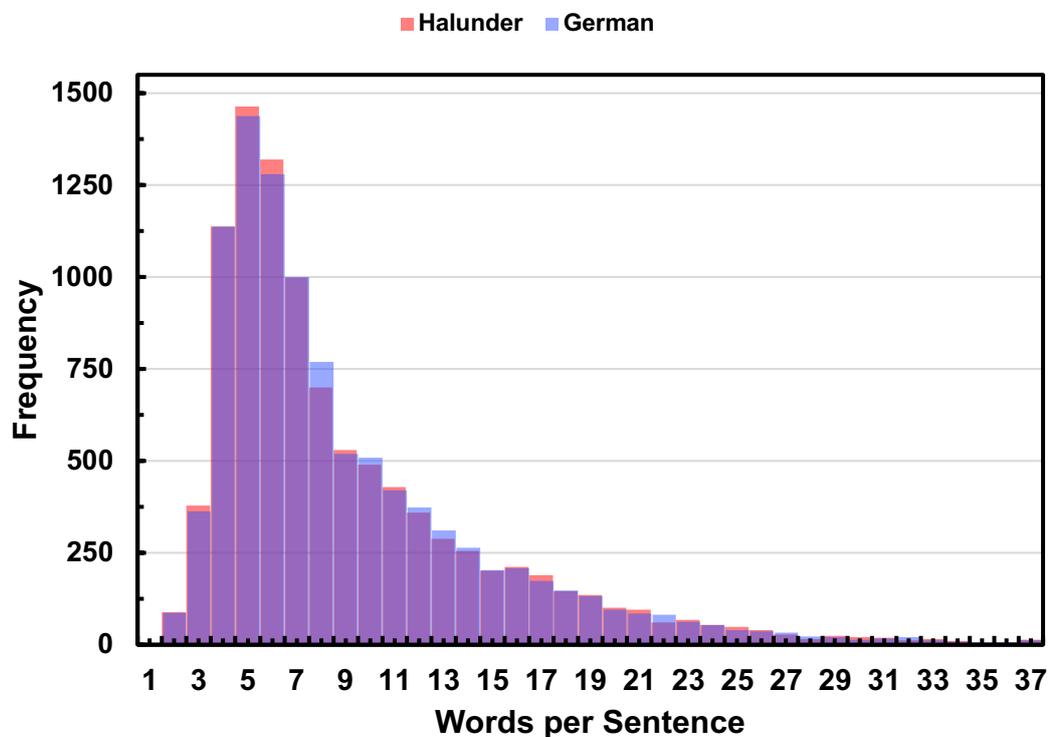


Figure 3.3 Words per Sentence for Halunder and German Parallel Corpus

The examination of sentence lengths reveals interesting similarities between the languages as visible in

Figure 3.3. Despite lexical differences, both languages show similar sentence lengths. In this right-skewed distribution that peaks at 5-7, Halunder averages 9.22 words per sentence and German 9.13 words. Such similarity suggests that while their word formation and usage differ, their basic sentence construction patterns,

averaging all sentences, align. Regarding likely the most important metric, the data shows a highly uneven distribution of word usage frequencies.

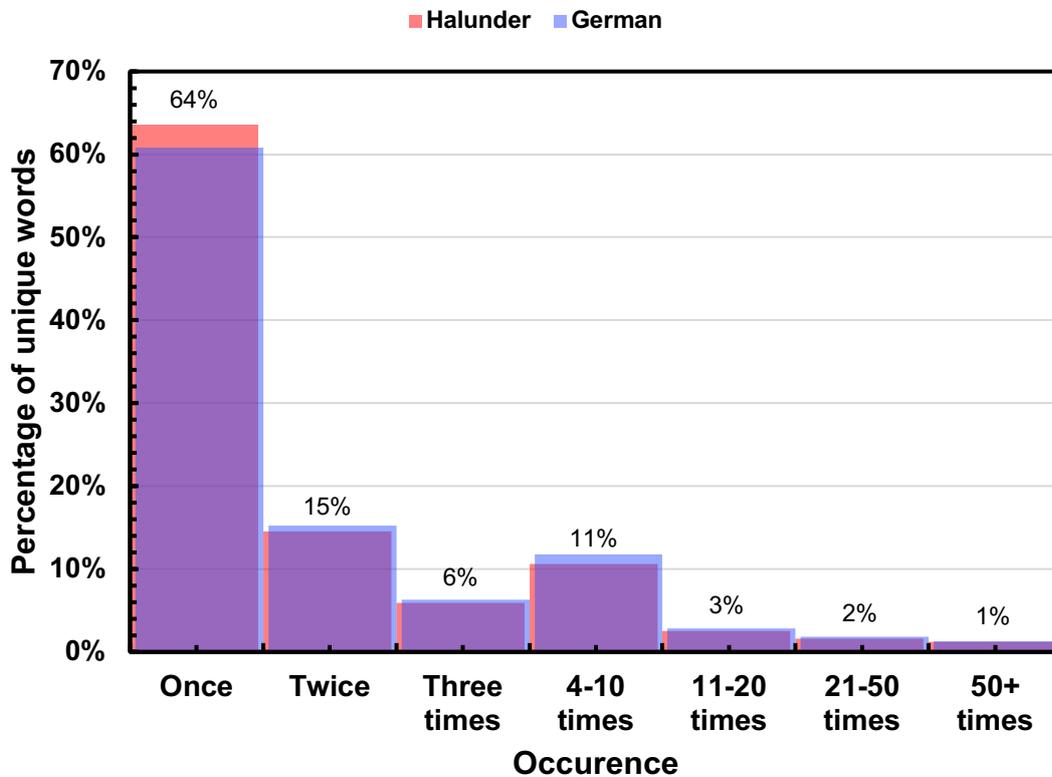


Figure 3.4 The occurrences of each word in the dataset.

While a few most used words in each language occur thousands of times—such as the Halunder word *en* at 4,846 occurrences and the German word *und* at 2,630 occurrences the large majority of words appear only once or twice. In fact, more than half of the Halunder words (9,785) and a similarly large proportion of German words (8,889) occur only once, and many others are found just a handful of times. Despite an overall average frequency of around six occurrences per word, this average is heavily influenced by a small number of extremely common words with a Median of 1 in both languages.

This pattern indicates that most of the vocabulary in the established dataset does not appear across multiple contexts, limiting the insight into how certain words might be used more broadly. It suggests that the current dataset may not be comprehensive enough to capture the full range of possible usage scenarios. Future data collection

efforts would be beneficial to ensure that the vocabulary is represented in a wider variety of examples, providing a more accurate picture of language use.

3.2.2 Limitations

The dataset presents several significant limitations that influence experimental design and results. The corpus shows a natural bias toward formal written communication, with limited representation of informal spoken language. Modern technical terminology receives restricted coverage due to the slow development of new terms (Krüss, 1985) reflecting Halunder's historical usage patterns rather than contemporary application. The absence of youth language limits the corpus's representation of linguistic diversity, although such language does not exist in Halunder (Wanke, 2008). The predominance of historical rather than contemporary usage patterns affect the model's ability to generate modern language constructions as seen in German.

These limitations directly impact the translation results (Zoph et al., 2016). While the corpus provides balanced coverage across documented domains, its size remains relatively small by modern machine translation standards (NLLB Team et al., 2022)

Chapter 4

Translation Experiments

This part of the thesis aims to produce the very first machine translation system for Halunder, primarily translating into German. The experiments have been designed to align with the research question defined in Section 1.5 and aim to identify factors that influence translation quality for a language as small as Halunder. A total of 31 configurations were trained and reviewed across seven experiments. The experiments rely on consistent evaluation protocols and build upon each other.

Three hypotheses were proposed.

(1) The first hypothesis focuses on incremental dataset increase. As the parallel corpus grows from 500 to 9,000 training sentences, the expectation is a measurable improvement in translation accuracy (Koehn & Knowles, 2017). Aiming to prove that further data collection efforts are worth exploring and that the translator can further be improved.

(2) The second hypothesis questions the effectiveness of fine-tuning existing pre-trained models against training from scratch as described in Section 2.3. Given Halunder's linguistic similarity to established North Sea Germanic languages (Montigny, 1994b), Halunder can benefit from adapting to various pre-trained models.

(3) The third hypothesis tests the capacity of transformer-based models to

handle a language as idiomatic and context-dependent, as Halunder, which are crucial for authentic Halunder translations (Montigny, 1992).

To systematically review the training progress and to test the evolution of translation models, a challenging sentence was selected:

Table 4.1 Challenging Sentence used for Review

Language	Sentence
Halunder (Original)	Miin Ooln wuune bedeel, bitu fan 'e Roathüs. Dja wuune toop med miin Bruur Jonas en miin Unkel Tomek. li Oolooti sooit, dat dear'n Koks bi lait
Ideal English Translation	My parents live in the lower land, next to the town hall. They live together with my brother Jonas and my uncle Tomek. My great grandma says that something is fishy / not right / that there is a catch (literally: "that there's a snail lying by")
Literal German Translation	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Jonas und meinem Onkel Tomek. Meine/Unsere Uroma sagt, dass da eine Wellhornschnecke bei liegt
Ideal German Translation	Meine Eltern wohnen im Unterland, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Jonas und meinem Onkel Tomek. Meine/Unsere Uroma sagt, dass „die Sache einen Haken hat“

It includes lexical ambiguity (*Ooln* means “parents” but stems from *Ool* = “old”, but also as in *Ooloot* = great (“old”) grandmother), compound nouns (*Oolooti* for great grandma), the difficulty of *bedeel* (meaning “downstairs” but only used in the context of location on Heligoland as “in the lower land”), the expression of locational relation *bitu fan 'e* (“next to the”), two family relations and gender-specific names, one of Germanic origin *Jonas* and one of foreign origin *Tomek*, and the prominent idiom *dat dear'n Koks bi lait* (indicating “that something is fishy/suspicious”). Another challenge with this expression is its overall token amount, as achieving good translation quality becomes increasingly difficult with longer inputs (Koehn & Knowles, 2017). This set of sentences, along with the 250-sentence test, help ensure that the assessment goes beyond simple literal reproduction and addresses nuanced contextual dependencies.

The chosen evaluation metrics: SacreBLEU (Post, 2018) and chrF++ (Popović, 2015) serve as automated methods to quantify performance by comparing the similarity of translations to expected translations provided in the parallel corpus, both on a scale of 0 to 100. SacreBLEU is a universally used metric measuring n-gram overlap. chrF++ adds another dimension by evaluating character-level n-gram overlaps rather than entire words, particularly relevant due to the irregular grammar evident in Halunder texts (Wanke, 2008). The maximum values are generally never achieved, as each input sentence can have multiple correct translations that are not accounted for (Callison-Burch et al., 2006).

4.1 General Setup

Training was conducted on an NVIDIA H100 NVL hosted via Runpod.io²⁷ using the Hugging Face Transformers library (Wolf et al., 2020) and PyTorch (Paszke et al., 2019), an open-source machine learning framework. A fixed random seed was chosen for reproducibility. The batch size was consistently set to 4 after initial testing.²⁸ The training proceeded for up to 40 epochs, with a minimum of 8 epochs before early stopping could be triggered. Early stopping used a patience of six evaluations to allow for unbiased stopping and overfitting prevention. The epoch limit was never hit; early stopping was triggered in every experiment.

Training parameters included a learning rate of 0.00002, a linear learning rate scheduler with a warmup ratio of 0.1, weight decay of 0.01, a label smoothing factor of 0.08, and a maximum token size of 256. Mixed precision (bf16) was used. Evaluation occurred every half epoch, to allow for early detection of plateaus and more frequent evaluation on the scarce data. The thesis follows recommendations for low-resource language translation (Kaplan et al., 2020).

All training and validation loss values, time, and automated metrics were logged and

²⁷ Runpod.io is an online GPU hosting service.

²⁸ Initial undocumented tests revealed that a batch size of 4 was superior to 8 and 16 in terms of training speed and translation quality.

saved to a CSV file at each evaluation interval. The same challenging Halunder sentence described in the introduction to this section was translated and logged after every adjustment to monitor progress. Following training, the 250-Sentence test set was translated and saved alongside experiment configuration values.

The controlled environment was designed to ensure that performance differences could be attributed to the tested factors rather than outside influences. The dataset was not adjusted throughout all experiments.

4.2 Fine-Tuning existing Pre-Trained Models for Halunder to German Translation

The first experiment aimed to establish a baseline for evaluating transfer learning effectiveness in Halunder-to-German translation. By applying equal parameters across multiple pre-trained models of different language coverage and comparing language proxies, it uncovers how well large models learn to adapt to a small language with many historical influences like Halunder (Wanke, 2008). The findings from this experiment will guide the selection of best performing candidates for subsequent experiments that focus on further optimizing transfer learning success.

4.2.1 Experiment Setup

All models were fine-tuned on the same dataset of 9,000 Halunder-German training sentences and 1,000 validation sentences. A diverse set of pre-trained models were tested, each with its own tokenizer. All models were integrated and fine-tuned using Hugging Face Transformers (Wolf et al., 2020) and PyTorch (Paszke et al., 2019):

- 1. NLLB-200-d-1.3B (Meta) with NllbTokenizer** (NLLB Team et al., 2022)

This model is part of the “No Language Left Behind” initiative by Meta, primarily intended for research in low-resource language translation. It allows for single sentence translation among 200 languages. Its large parameter count (1.3B distilled was chosen) and varied data exposure make it a strong

contender for Halunder, as it can utilize patterns learned across distant linguistic families. Two language proxies were tested: (nl-de) and (de-de), reflecting Dutch-to-German and German-to-German.

2. mBART-large-50 and mBART-large-50-m2m-mmt (Meta) with MBart50Tokenizer (Tang et al., 2020)

mBART-50 is a multilingual sequence-to-sequence model pre-trained for general text denoising tasks, while its many to many multilingual machine translation variant is specialized for direct translation between any pair of 50 languages. Proxies included (de-de), (en-de), and (nl-de), covering German-to-German, English-to-German, and Dutch-to-German.

3. OpusMT (Helsinki-NLP) with MarianTokenizer (Tiedemann & Thottingal, 2020)

OpusMT models, such as GEM to GEM (General Germanic) and GMW to GMW (West Germanic to West Germanic), are trained on subsets of OPUS multilingual corpora. These models are smaller and more domain-specific compared to NLLB or mBART. Their narrower focus on Germanic languages might help capture certain linguistic patterns in Halunder.

4. GermanToNorthFrisianV1 (Tritkoman)²⁹ with MarianTokenizer

This model, originally developed for North Frisian should be directly related to Halunder. It has a smaller and more focused pre-training corpus although no exact documentation is available. While North Frisian and Halunder share linguistic ancestry, the model's directionality and limited Halunder input data may make its adaptability difficult. It should provide a useful contrast to large-scale multilingual approaches.

²⁹ Tritkoman. (n.d.). GermantoNorthFrisianV1 [Machine translation model]. Hugging Face. <https://huggingface.co/Tritkoman/GermantoNorthFrisianV1>. The model achieves a SacreBLEU score of 50.89 on the validation set and was trained using AutoTrain

4.2.2 Experiment Results

Overall, the compared models performed similarly well, with notable differences between NLLB200, OpusMT and mBART50 and finally GermanToNorthFrisianV1. The results demonstrate the significant impact of pre-training data diversity and volume on translation quality. A comparison of BLEU Scores and training time can be found in **Table 4.2**.

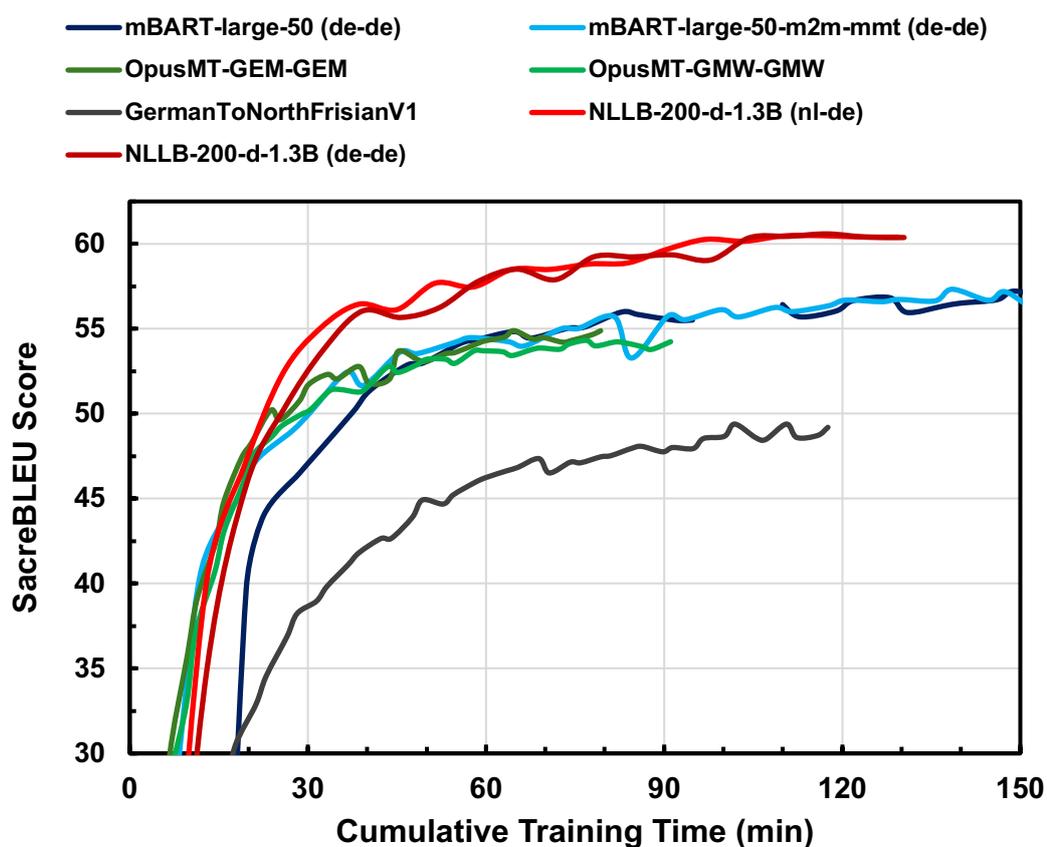


Figure 4.1 Comparison between fine-tuning pre-trained models for Halunder - German

The NLLB-200-d-1.3B model achieved the highest BLEU scores, reaching 60.48 (nl-de) and 60.38 (de-de) in approximately 156 minutes of training. The result highlights the strength of NLLB's large amount of multilingual pre-training, which spans hundreds of languages, enabling the model to generalize effectively to low-resource languages like Halunder. The minimal difference between the language proxies suggests that linguistic similarity to Halunder has a smaller influence compared to the sheer volume of pre-training data.

The mBART-large-50-m2m-mmt models, specifically designed for multilingual machine translation, performed slightly better than the general-purpose mBART-large-50 model. The m2m-mmt variants achieved BLEU scores ranging from 56.49 (nl-de proxy) to 57.33 (de-de and en-de proxies) with shorter training times (148–166 minutes) compared to mBART-large-50's score of 57.39 in 201 minutes. Although the score is comparable, the difference reflects increased time-efficiency of the task-specific model optimized for translation tasks.

The OpusMT models (GEM-GEM and GMW-GMW) demonstrated lower BLEU scores, with 54.89 and 54.32 achieved in 79 and 91 minutes. The general Germanic model reached a higher score and reached its best performing epoch significantly faster. While they are more efficient, their performance plateaued earlier than the larger models.

The GermanToNorthFrisianV1 model, despite being designed for North Frisian achieved only 49.39 BLEU after 118 minutes of training.

Table 4.2 Performance vs. Training Time between fine-tuned models for Halunder-German

Pre-trained model	Highest BLEU Score	Training time (min)
NLLB-200-d-1.3B (nl-de)	60.48	157
NLLB-200-d-1.3B (de-de)	60.38	156
mBART-large-50 (de-de)	57.39	201
mBART-large-50-m2m-mmt (de-de)	57.33	165
mBART-large-50-m2m-mmt (en-de)	57.33	166
mBART-large-50-m2m-mmt (nl-de)	56.49	148
OpusMT-GEM-GEM	54.89	79
OpusMT-GMW-GMW	54.32	91
GermanToNorthFrisianV1	49.39	118

The influence of language proxies is particularly evident in **Figure 4.2**, which focuses on mBART-50 variants. The de-de and en-de proxies performed nearly identically, with both achieving 57.33 BLEU scores. This suggests that Halunder aligns more closely with languages like German and English in structure and vocabulary. The nl-de proxy, while still competitive at 56.49, fell behind slightly, either due to a less fitting language similarity or due to a reduced amount of Dutch data. Showcasing the importance of selecting proxies that are not only linguistically related but also supported by large amounts of pre-training data.

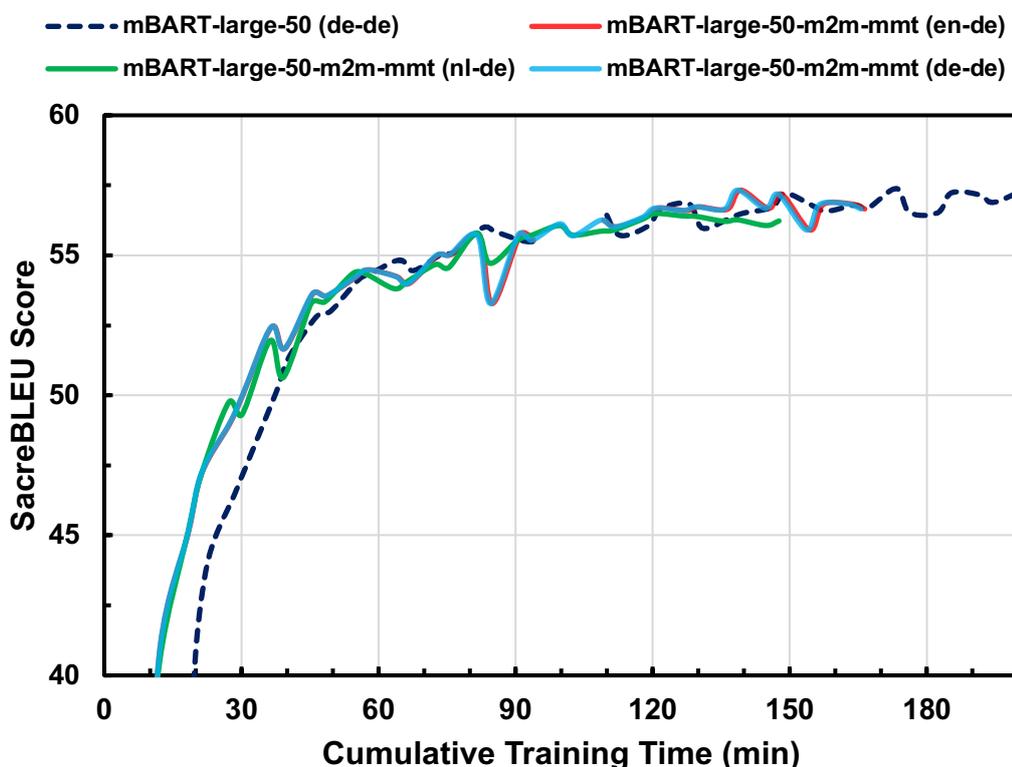


Figure 4.2 Comparison of fine-tuning mBART50 for Halunder-German

4.2.3 Hyperparameter Optimization for mBART50

The next experiment focused on fine-tuning the mBART-Large-50-many-to-many-mmt model using nl-de as a proxy for Halunder-German translation with various training parameters to explore room for improvement on this specific model. The model choice is based on closer evaluation, it is evident that that the Dutch proxy caught the idiomatic concepts better and provided a more precise translation.

4.2.4 Experiment Setup

Ten different parameter configurations (C1–C10) were established, adjusting parameters such as learning rate, dropout rates, weight decay, label smoothing, and learning rate schedulers.

Table 4.3 Distribution of SacreBLEU scores across different configurations

Config	Short Description	Test BLEU Score	Final BLEU Score	Total training time (min)
C9	Hybrid of top performing C2 and C8	58.24	58.06	122
C2	High dropout	58.13	57.84	211
C10	High regularization with pure cosine schedule	58.08	57.74	56
C8	Mixed regularization with cosine restarts	57.41	57.00	66
C6	Polynomial decay with medium-high regularization	57.02	56.60	138
C5	Conservative learning with heavy regularization	56.84	56.67	122
C7	Stepped learning rate with low dropout	56.70	55.76	56
C1	Baseline	56.31	56.00	120
C3	Cosine schedule with medium dropout	55.72	55.20	104
C4	Aggressive learning rate with minimal regularization	52.93	52.74	85

Configurations included variations like high dropout regularization, aggressive learning rates, different warmup ratios, and various learning rate schedulers (e.g., linear, cosine, polynomial decay) - comparing BLEU scores across cumulative training times. The goal was to identify the configurations that yield the best

translation performance in terms of BLEU score, idiom usage, training stability, and efficiency. The entire configuration setup can be found in Appendix A.

4.2.5 Experiment Results

The results highlight key differences in the performance of the hyperparameter configurations, particularly in terms of BLEU score progression and training efficiency. C9 delivered the best overall results, achieving the highest test-set BLEU score of 58.24 and maintaining a stable final BLEU of 58.06 within 122 minutes of training. Its use of a cosine learning rate schedule with moderate dropout levels provided balanced and consistent growth, making it a strong candidate for high-quality translation with efficient training.

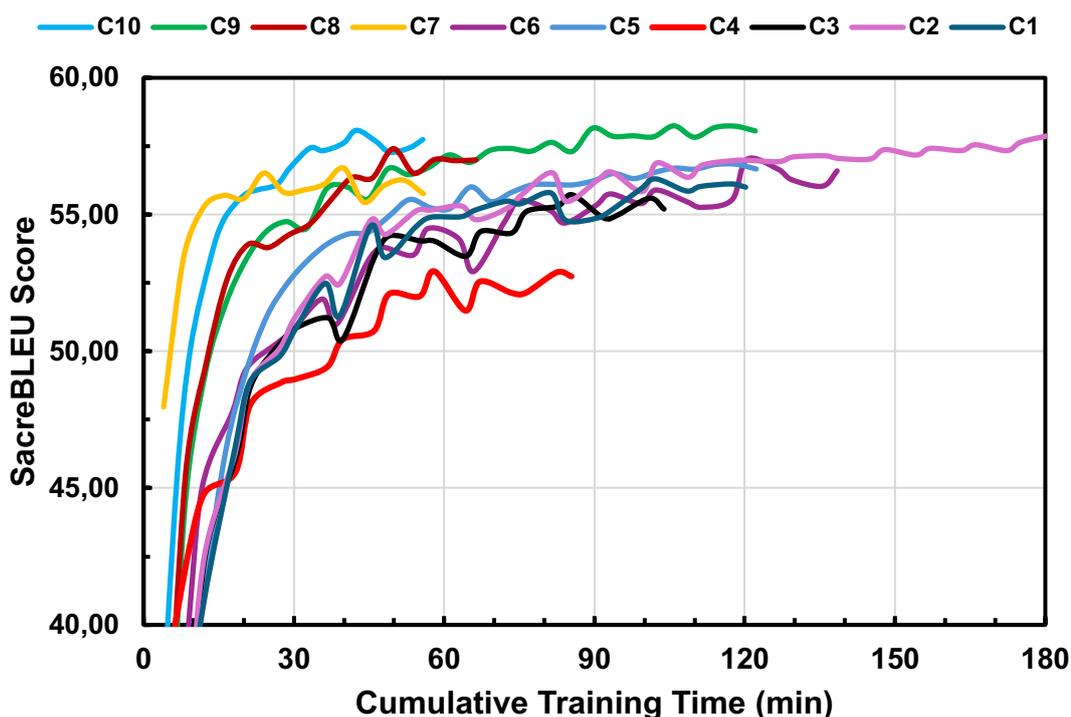


Figure 4.3 Hyperparameter Comparison for mbart50 using nl-de for Halunder-German Translation

C2 reached a comparable BLEU score of 58.13 but required significantly more training time (211 minutes). The high dropout regularization slowed its early performance, but it demonstrated consistent improvements throughout training, indicating strong long-term stability. This makes C2 more suitable for tasks where

training time is less of a constraint.

C10 stood out for its efficiency, achieving a BLEU score of 58.08 in just 56 minutes. The higher dropout rates and cosine learning rate schedule allowed rapid performance gains, making it ideal for scenarios requiring quick results. Meanwhile, C8 performed well with a BLEU score of 57.41 in 66 minutes, showing steady progress but triggering early stopping earlier than C9 or C2, which may limit its longer-term stability.

The weaker configurations, such as C6, C5, C3, and C1, demonstrated limited BLEU improvements or inconsistent growth patterns. C4 performed the worst, achieving only 52.93 BLEU, likely due to its aggressive learning rate and minimal regularization, which caused rapid stagnation. The findings demonstrate the importance of balancing regularization and learning rate strategies. Configurations like C9 and C10 are optimal for efficient training, while C2 offers stability for extended cycles.

4.3 Hyperparameter Optimization OpusMT

The results from the OpusMT Gem-Gem base model provide further insight into the applicability and transferability of the best-performing configurations across models.

4.3.1 Experiment Setup

The best-performing configurations under mBART50 (C2, C8 and C9) were applied to the previously introduced OpusMT gem-gem model under the same conditions and compared to the previously described results.

4.3.2 Experiment Results

C8 emerged as the top performer in this setup, achieving a slightly higher BLEU score (56.40) than C9 (56.00), while requiring the same amount of training time (63 minutes). While C9's consistency and efficiency make it a strong candidate across

models, the superior performance of C8 with OpusMT demonstrates the value of dynamic regularization in diverse contexts. The results also highlight that configurations like C2, although robust for mBART, may require adjustments to perform optimally on other models.

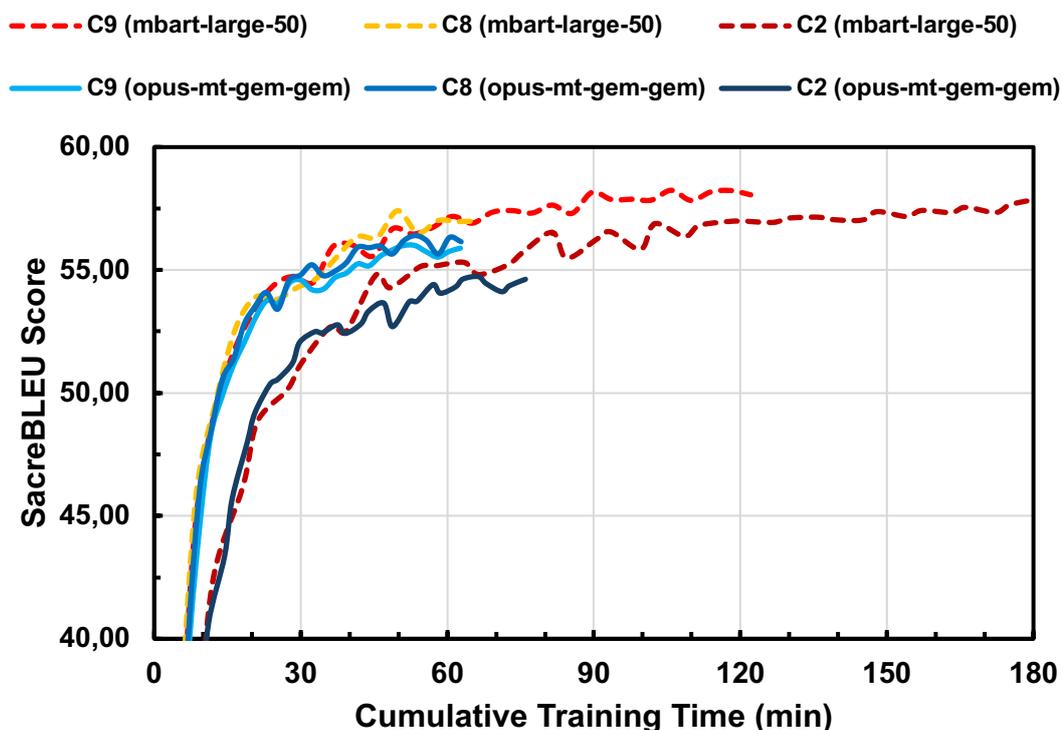


Figure 4.4 Hyperparameter Comparison for C9, C8 and C2 between mBART50 and opusMTgem-gem

This suggests that C8's combination of moderate regularization and a cosine learning rate schedule with restarts might generalize better to this model. C2, on the other hand, underperformed relative to its results with mBART, reaching a maximum BLEU score of 54.73 in 76 minutes. The longer training time and lower score indicate that C2's high dropout values and linear scheduler may be less suited to OpusMT's architecture.

4.4 Manual Evaluation OpusMT vs. mBART

To further compare these six configurations (C2, C8, and C9 across both mBART and OpusMT), the complex sentence previously introduced was translated using each

model. The resulting translations can be used to assess real-world performance beyond BLEU scores. The table below summarizes these results, providing insights into which configuration and model combination produced the most accurate and contextually appropriate translation and what the final model would generate.

Table 4.4 Evaluated Sentences between models.

Configuration		Sentence
Halunder (Original)		Miin Ooln wuune bedeeIn, bitu fan 'e Roathüs. Dja wuune toop med miin Bruur Jonas en miin Unkel Tomek. li Oolooti sooit, dat dear'n Koks bi lait.
English Translation		My parents live in the lower lands, next to the townhall. They live together with my brother Jonas and my uncle Tomek. My/Our great-grandma says that there is a catch (that a snail lays beside).
Expected German Translation		Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Jonas und meinem Onkel Tomek. Meine/Unsere Uroma sagt, dass da eine Wellhornschnecke bei liegt. Meine Eltern wohnen im Unterland, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Jonas und meinem Onkel Tomek. Meine/Unsere Uroma sagt, dass „die Sache einen Haken hat“.
C9 (mbart-large-50)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Johann und meinem Onkel Tomek. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt, das heißt, dass die Sache einen Haken hat.
	Final	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Johanna und meinem Onkel Tomek. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt.
C8 (mbart-large-50)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Johann und meinem Onkel Tomek. Meine Eltern sagen, dass eine Seeschnecke dabei liegt, das heißt, dass die Sache einen Haken hat.
	Final	Meine Eltern wohnen unten, neben dem Roadhaus. Sie wohnen zusammen mit meinem Bruder Johannes und meinem Onkel Tomek. Meine Eltern sagen, dass eine Wellhornschnecke dabei liegt.
C2 (mbart-large-50)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Jonas und meinem Onkel Tomek. Unsere Großmutter sagt, dass eine Wellhornschnecke dabei liegt.

	Final	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Johanna und meinem Onkel Tomek. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt.
C9 (opus-mt-gem-gem)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder und meinem Onkel Tomick. Unsere Urgroßmutter sagt, dass eine Seeschnecke dabei liegt.
	Final	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder und meinem Onkel Tome. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt.
C8 (opus-mt-gem-gem)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Andreas und meinem Onkel Tome. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt, das heißt, dass die Sache einen Haken hat.
	Final	Vati wohnt unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Andreas und meinem Onkel Tome. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt, das heißt, dass die Sache einen Haken hat.
C2 (opus-mt-gem-gem)	Best	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Johannis und meinem Onkel Tomik. Unsere Urgroßmutter sagte, dass eine Seeschnecke dabei liegt.
	Final	Meine Eltern wohnen unten, neben dem Rathaus. Sie wohnen zusammen mit meinem Bruder Andreas und meinem Onkel Tome. Unsere Großmutter sagt, dass eine Seeschnecke dabei liegt.

The evaluation of translations revealed several positive trends across all configurations, alongside notable differences in performance. A key strength shared by all configurations was the accurate recognition of familial relationships and proper nouns, correctly identifying "Jonas" and "Tomek" as names for the brother and uncle in most cases. Additionally, all models successfully maintained the sentence's overall structure and meaning, reflecting strong foundational translation capabilities.

C9 (mbart-large-50) delivered the most accurate results overall, correctly translating the idiom "die Sache einen Haken hat" and preserving key terms like *Eltern* and *Rathaus*. Despite occasional errors in handling "Urgroßmutter" (translated as "Großmutter") and name fluctuations (e.g., "Johanna" for "Jonas" in the final output),

it consistently provided contextually appropriate translations.

C8 (mbart-large-50) and C2 (mbart-large-50) performed well but exhibited more variability. C8 handled idiomatic expressions well in its best outputs but introduced lexical issues in final translations, such as “Roadhaus” for “Rathaus.” C2 demonstrated solid handling of familial relationships, retaining "Jonas" and "Wellhornschnecke" in its best outputs, but inconsistencies emerged in later stages, including gender mismatches in names.

The OpusMT Gem-Gem configurations also displayed strengths. Both C9 and C8 (OpusMT) effectively recognized familial relationships and maintained overall sentence structure. However, they showed more frequent lexical errors, such as "Ratshaus" for “Rathaus” and inconsistent handling of names ("Andreas" for "Jonas," "Tomick" for "Tomek"). Notably, C9 and C2 (OpusMT) occasionally captured "Urgroßmutter" correctly, a distinction not consistently observed in mbart-large-50 outputs.

In summary, all configurations excelled in preserving sentence structure, familial relationships, and identifying proper nouns. However, C9 (mbart-large-50) stood out for its balance of accuracy and idiomatic understanding, while OpusMT configurations, particularly C9, demonstrated potential but required further refinement to match the lexical precision of mbart-large-50.

4.5 Impact of Dataset Size on Translation Performance

This experiment investigated the hypothesis that increasing the dataset size would improve translation performance using a fine-tuned Helsinki/opus-mt-gem-gem model.

4.5.1 Experiment Setup

The datasets were divided into cumulative subsets of 500, 1,000, 3,000, 5,000, 7,000, and 9,000 sentences, where each larger subset included the smaller ones (e.g., the

3,000-sentence set included the 500- and 1,000-sentence subsets). Models trained on smaller datasets (500, 1,000, and 3,000 sentences) used 500 validation sentences, while larger datasets (5,000 sentences and above) used 1,000 validation sentences. Each experiment was trained on the OpusMT gem-gem model with C8 configuration. The goal was to evaluate how dataset size impacts BLEU scores and translation quality.

4.5.2 Experiment Results

As seen in **Figure 4.5**, increasing the dataset size results in higher BLEU scores, but the rate of improvement diminishes with larger datasets. The 3,000-sentence dataset shows a clear improvement, reaching a highest BLEU score of **30.79** and ChrF++ of **54.36**.

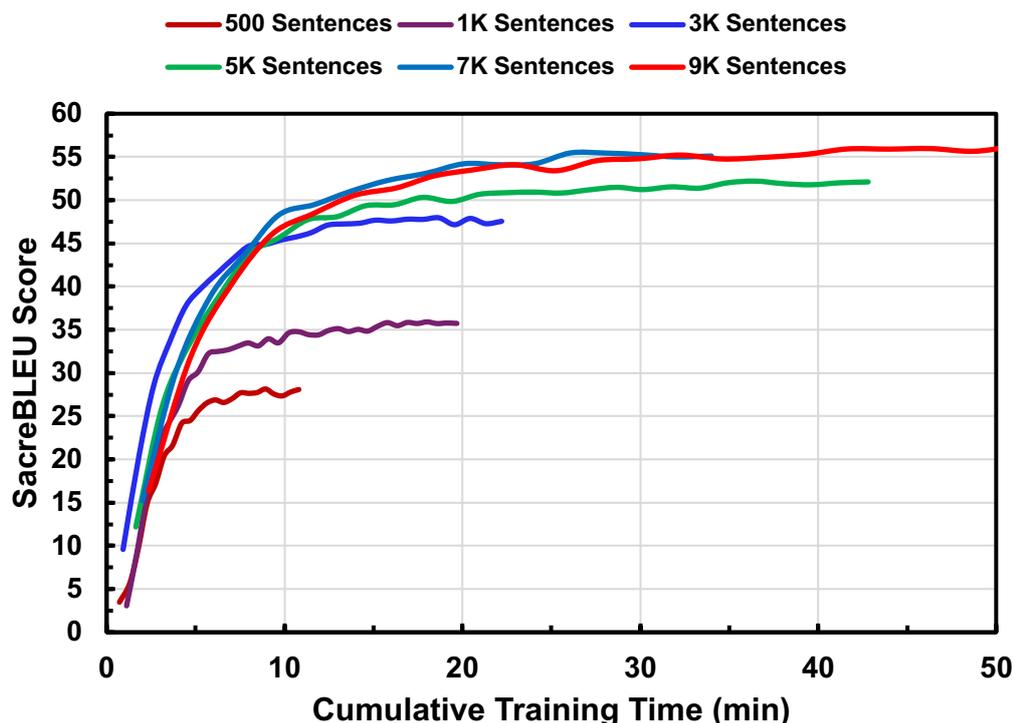


Figure 4.5 Comparison of Halunder to German translation with different amounts of sentences

This dataset introduces enough additional data to provide the model with better generalization capabilities and improve translation accuracy substantially.

While BLEU scores continue to improve with larger datasets, the gains become smaller. The 7,000-sentence model reaches a highest BLEU score of **60.13** and Chrf++ of **76.53** while the 9,000-sentence model achieves **63.13** and **78.54**. The minimal difference between these two models reflects diminishing returns in numerical performance as the model saturates its ability to learn from the available data.

The BLEU scores for the smaller datasets plateau early, with the 500-sentence model reaching a highest BLEU score of **30.79** and Chrf++ of **54.36** and the 1,000-sentence model slightly higher at **42.39** and **63.54**. The small performance gap indicates that the additional 500 sentences lack sufficient linguistic diversity to make a significant difference.

Although the BLEU score difference between the 7,000- and 9,000-sentence models is marginal, qualitative analysis reveals significant differences in translation accuracy and fluency. The 9,000-sentence model consistently produces translations that are closer to the expected output, capturing grammatical differences, general nuances, and resolving ambiguities missed by the 7,000-sentence model. **Table 4.5** highlights these differences with representative examples, showing that even a small numerical improvement in BLEU can correspond to meaningful real-world gains in translation quality.

Table 4.5 Representative samples showcasing important differences between minimal automated metric differences

Type	Sentence
Source (Halunder)	En groot Skolk es hoggelt würn.
Expected Translation	Eine große Scholle ist geangelt worden.
Translation 9K Dataset	Eine große Scholle ist geangelt worden.
Translation 7K Dataset	Eine große Scholle ist gepflastert worden.
Translation 1K Dataset	Eine große Scholle ist geholt worden.
Source (Halunder)	Noa de Krich wear iip Lun alles uuntau.
Expected Translation	Nach dem Krieg war auf Helgoland alles kaputt.
Translation 9K Dataset	Nach dem Krieg war auf Helgoland alles kaputt.
Translation 7K Dataset	Nach dem Krieg war auf Helgoland alles an.
Translation 1K Dataset	Nach dem Krieg war auf Helgoland alles zerdrückt.
Source (Halunder)	Ik reekene no gau it, wat ik itden'n hoa.
Expected Translation	Ich rechne noch schnell aus, was ich ausgegeben habe.
Translation 9K Dataset	Ich rechne noch schnell aus, was ich ausgegeben habe.
Translation 7K Dataset	Ich rechnen noch schnell aus, was ich getan habe.
Translation 1K Dataset	Ich rechne noch schnell aus, was ich gehabt habe.
Source (Halunder)	Wi siise deer 'e Goater en Stroaten.
Expected Translation	Wir sausen durch die Gassen und Straßen.
Translation 9K Dataset	Wir sausen durch die Gassen und Straßen.
Translation 7K Dataset	Wir sausten durch die Gassen und Straßen.
Translation 1K Dataset	Wir sehen durch die Gassen und Straßen.

4.6 German to Halunder Translation

Although the primary focus of this thesis is Halunder to German translation with the aim to preserve material and enabling the access to cultural texts to German speakers, translating German to Halunder was explored to assess the feasibility of creating a complete bidirectional translator. This task seems inherently more challenging due to the low availability of Halunder data. While Halunder texts are relatively well-documented, they are scarce in pre-trained model corpora, unlike German, which benefits from extensive pre-training coverage (Kozhirbayev, 2024). As a result, translating into Halunder poses additional challenges, particularly in terms of spelling, grammar, and overall fluency, likely as the models are less familiar with the target language.

4.6.1 Experiment Setup

To evaluate this, the **NLLB-200-d-1.3B** model with de-de and de-nl proxies was compared to the **OpusMT-GEM-GEM** model, using the **C8** configuration. The results are summarized in **Table 4.6**, with BLEU and ChrF++ score progressions shown in **Figure 4.5** and **Figure 4.6**.

4.6.2 Experiment Results

Despite the inherent challenges, all three models achieved competitive BLEU scores, with **NLLB-200-d-1.3B (DE2DE)** achieving the highest score of **52.94**. The DE2NL proxy closely followed with **52.72**, while **OpusMT-GEM-GEM** reached an impressive **52.33**, despite its smaller pre-training corpus. The **OpusMT-GEM-GEM model**, in particular, stands out for its rapid growth in BLEU scores, converging significantly faster than the NLLB models. This suggests that while its pre-training data is less extensive, it is well-suited for low-resource scenarios where efficient training is critical.

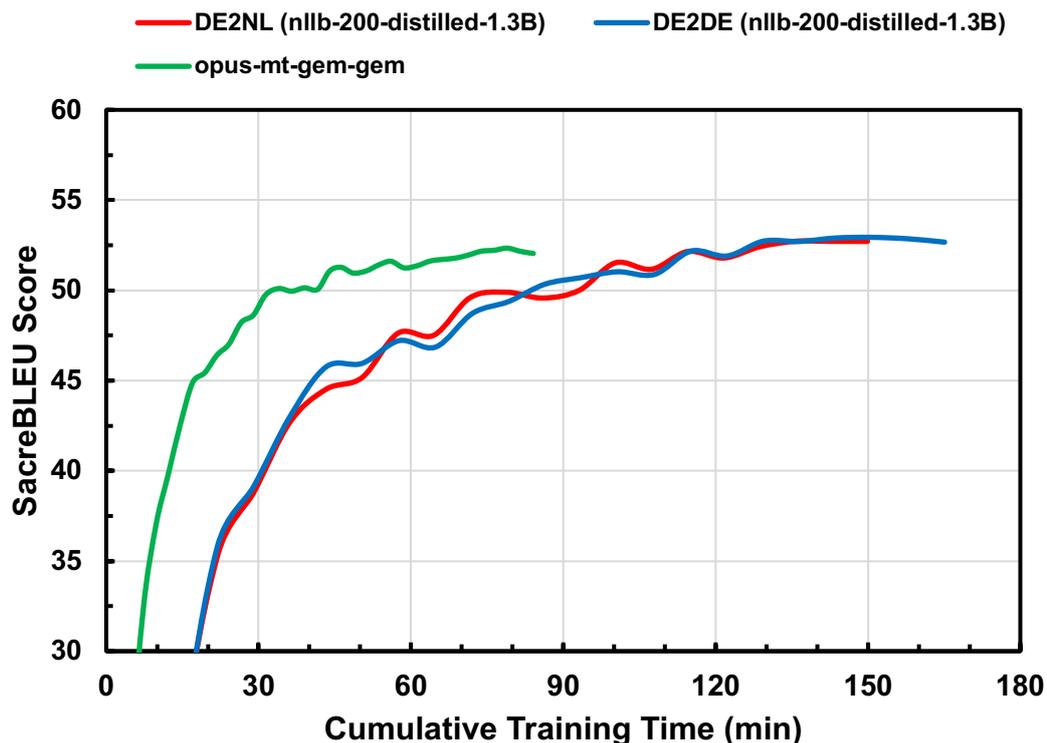


Figure 4.6 Comparison for German to Halunder translation between Helsinki/opus-mt-gem-gem and 2 language proxies for facebook/nllb-200-distilled-1.3B

The **ChrF++ scores**, which emphasize character-level accuracy, further highlight the strong performance of the NLLB models, with de-de reaching **72.34** and de-nl achieving **71.79**. The difference between the proxies is minimal, but DE2DE converges slightly slower before surpassing de-nl in final performance. Unfortunately, ChrF++ was not tested for OpusMT-GEM-GEM, but its BLEU results suggest that it could remain competitive in character-level evaluations as well.

The OpusMT-GEM-GEM model is the most efficient, requiring just **84 minutes** to reach its highest BLEU score. In contrast, NLLB-de-nl and NLLB-de-de took **151 minutes** and **166 minutes**, respectively. While the NLLB models ultimately achieved slightly higher BLEU and ChrF++ scores, OpusMT-GEM-GEM's ability to converge quickly and perform near the level of much larger models highlights its efficiency and practicality for rapid deployment in resource-constrained scenarios.

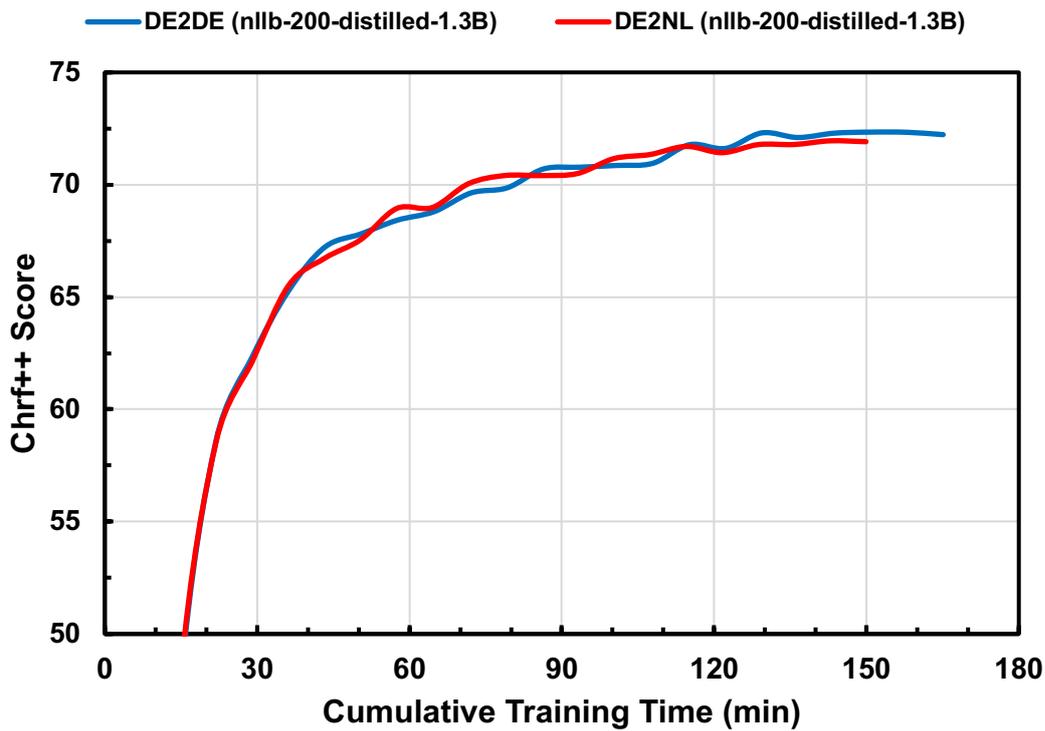


Figure 4.7 Chrf++ Result comparison for German to Halunder translation using two fine-tuned facebook/nllb-200-distilled-1.3B models.

Table 4.6 Performance and training time of German to Halunder models.

Pre-trained model	Highest BLEU Score	Chrf++ Score	Training time (min)
NLLB-200-d-1.3B (de-de)	52.94	72.34	166
NLLB-200-d-1.3B (de-nl)	52.72	71.79	151
OpusMT-GEM-GEM (C8)	52.33	-	84

4.7 Feasibility of Training from Scratch

To evaluate the feasibility of developing a Halunder-to-German translator without Transfer Learning and the reliance on pre-trained models, a model was trained from scratch using FAIRSEQ (Ott et al., 2019). The training process utilized the full dataset with Byte Pair Encoding (BPE) for tokenization (Sennrich et al., 2016) and spanned a

total of 220 epochs over 40.4 minutes. Early stopping was triggered after a patience of 15 epochs.

4.7.1 Experiment Evaluation

While the training reached a maximum **BLEU** score (Post, 2018) of **28.13**, the overall translation quality was poor, with most outputs failing to produce meaningful or accurate translations.

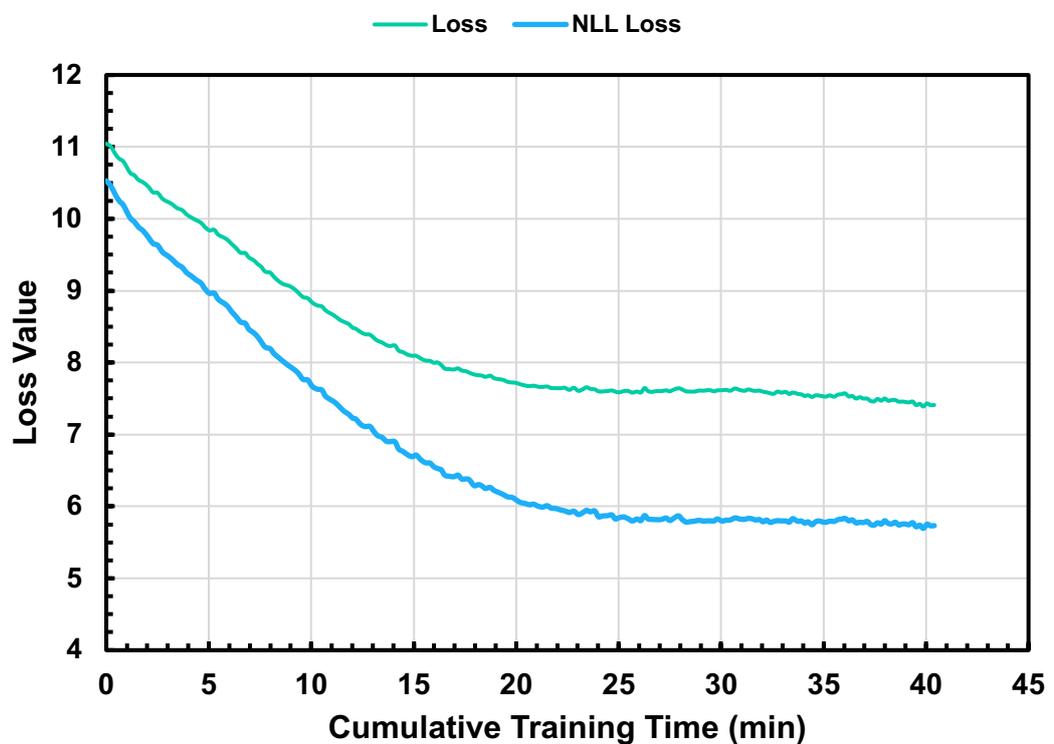


Figure 4.8 Loss vs. Negative Log-Likelihood Loss of the Baseline Model

The training process was monitored using two primary metrics: overall loss and the negative log-likelihood loss (NLL loss). Negative log-likelihood measures how well the model predicts the correct token during training by calculating the logarithm of the predicted probability for the correct output token. It penalizes incorrect predictions, making it a fundamental metric in language modeling and translation tasks (Ziyin et al., 2020)

The overall loss, which includes additional regularization terms such as label

smoothing, was consistently higher than the NLL loss. The graph shows that both metrics converged steadily, with the overall loss starting at **11.0** and NLL loss at **10.5**, decreasing to **7.4** and **5.7**, respectively, by 40 minutes of training. This indicates that the model optimized effectively, but the relatively high final values reflect the challenge of achieving fine-grained accuracy in this low-resource setup.

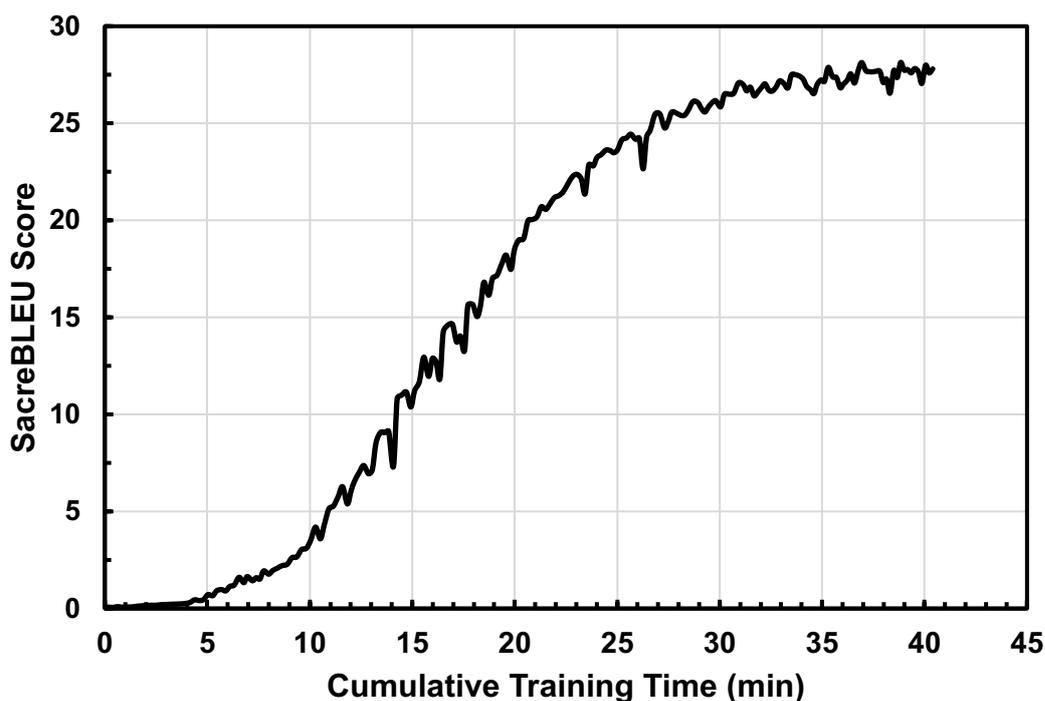


Figure 4.9 SacreBLEU Performance of a Baseline Non-Pretrained Model (Halunder to German)

The **BLEU** score steadily increased throughout the training process, peaking at **28.13**. The score reflects modest success in capturing basic sentence structure and simple word translations. However, the generated translations were largely nonsensical for complex or nuanced sentences, indicating that the score alone does not fully capture semantic fidelity.

4.7.2 Experiment Results

Examining the results of a test set sample however reveals the limitations of using BLEU score as an indicator of translation quality. While the improving BLEU score reflects that some parts of the sentences are translated correctly, the overall

translations are mostly incoherent, inaccurate and to a large part non-sensical. The table provides representative examples that demonstrate how partial correctness can inflate the score, even when the translations are far from usable.

Table 4.7 Representative Samples from the 250-Sentence Test Set Translated by a baseline model.

Type	Sentence
Source (Halunder)	Halunder Hart - Halunder Spreek!
Expected Translation	Helgoländisches Herz - Helgoländische Sprache!
Model Translation	Helgoländisches Herz - Helgoländische Sprache!
Source (Halunder)	De Roodbreschet songt faini Leeten.
Expected Translation	Das Rotkehlchen singt schöne Lieder.
Model Translation	Die Spiechen singt es für schöne Grabe gegessen.
Source (Halunder)	Nä lai de Rudders weer uun 'e Hoawen.
Expected Translation	Nun liegen die Börteboote wieder im Hafen.
Model Translation	Nun liegen die Börteboote wieder in der Haken.
Source (Halunder)	Wi hoa Kathalsen, Miisken en Letj Buurn.
Expected Translation	Wir haben Silbermöwen, Dreizehenmöwen und Sturmmöwen.
Model Translation	Wir haben Katzenhaie vorhorn.

Chapter 5

Discussion and Interpretation

The experimental results provide quantitative insights into the performance of Transformer-based technology applied to Halunder. However, the main research question—“Is Transformer technology the decisive technology for preserving Halunder and other endangered languages?”—requires further analysis to interpret these findings in a broader context. Beyond numerical performance, this discussion examines the practical usefulness of manual translation reviews, evaluates the impact of the proposed technology on language preservation, and explores future recommendations for advancing the role of machine translation in addressing the unique challenges of endangered languages. By connecting experimental outcomes with real-world applications, this section aims to assess whether Transformers truly offer a sustainable solution for Halunder and similar languages.

5.1 Manual Review of Translation Quality

In addition to the automated metrics (See Section 4.1), a manual evaluation was conducted on the best-performing model for Halunder to German translation (NLLB with Dutch to German proxy). While automated metrics for this fine-tuned model, specifically the BLEU score (60.48) and chrF++ score (75.62) provide quantitative measures of translation quality, manual evaluation is essential to evaluating the actual and perceived quality of translations

The evaluation scale followed a protocol similar to XSTS as used by NLLB Team et al., 2022. It ranged from 1 (unusable translation) to 6 (a perfect translation surpassing the provided reference). Within this scale, a score of 3 indicated an acceptable translation without serious errors. Although the scale extended to 6, the effective maximum for standard scoring was set to 5 points per translation, resulting in a total of 500 possible points for 100 samples. Assigning a 6 did not add extra points beyond this total, but rather indicated that the translation went beyond what was expected and compared to. The score 6 is designated for machine translations delivering exceptionally natural German expressions and flow. The complete evaluation protocol can be found in Appendix B.

5.1.1 Results

As with any human judgment, some subjectivity is expected despite the use of a standardized protocol (Dale, 2022). The score distributions for both evaluators were as follows:

Table 5.1 Distribution of manual evaluation scores by two evaluators.

Rating	Evaluator 1	Evaluator 2	Average
1 (Unusable)	14	16	15.0
2 (Major mistake)	18	11	14.5
3 (Acceptable)	12	18	15.0
4 (Good)	9	14	11.5
5 (Flawless)	28	23	25.5
6 (Perfect)	19	18	18.5

Evaluator 1 assigned a total of 376 points out of the 500-point maximum, and Evaluator 2 assigned 371 out of 500. This close alignment (within about 1%) suggests a reasonable level of agreement. Notably, 64% of the translations received identical

scores from both reviewers, indicating moderate inter-annotator consistency.

Table 5.2 Examples of translations containing major flaws.

Type	Sentence
Street name confusion (rated 2)	
Source	“Es heult zwischen den Gassen und saust um die Ecke beim Bakenweg .“
Source (Expected)	(It howls through the alleys and whizzes around the corner at Beacon Way.)
Translation	„Es heult zwischen den Gassen und saust um die Ecke bei der Buckelstraße .“
Translation (English)	(It howls through the alleys and whizzes around the corner at Hunchback Street.)
Semantic shift (rated 2)	
Source	"Ich rechne noch schnell aus, was ich ausgegeben habe."
Source (Expected)	(I'll quickly calculate what I have spent.)
Translation	"Ich rechne noch schnell aus, was ich ausgedacht habe."
Translation (English)	(I'll quickly calculate what I have made up.)
Critical meaning alteration (rated 1)	
Source	"Einige hatten auch noch einen offenen Herd in der Küche mit einer Stange darüber zum Fisch räuchern."
Source (Expected)	(Some also had an open fireplace in the kitchen with a pole above it for smoking fish.)
Translation	"Einige hatten auch noch einen offenen Herd in der Küche mit einem Netz darüber zum Fischen geschmuggelt ."
Translation (English)	(Some had also smuggled an open fireplace in the kitchen with a net above it for fishing.)

Analysis of the results reveals that 55.5% of the translations were rated as good, flawless, or perfect (scores 4-6), while 70.5% of the translations were considered at least acceptable. The remaining 29.5% contained major mistakes that significantly

impacted meaning. These errors typically involved incorrect word choices that completely altered the sentence meaning.

However, the model also produced translations that surpassed expectations (rated 6), demonstrating sophisticated handling of both linguistic and cultural elements. In the following example, the Halunder word *blinket* could be translated to *blinken* (to blink/flash), *funkeln* (to sparkle/twinkle), *glänzen* (to shine/gleam) or *glitzern* (to glitter). While the reference translation in the parallel corpus provides *glitzert*, the evaluators coincided that *funkelt* matches a more natural German expression.

Halunder	"Deät Weeter blinket as Sellewer."
Reference Translation	"Das Wasser glitzert wie Silber." (The water glitters like silver.)
Model Translation	"Das Wasser funkelt wie Silber." (The water sparkles like silver.)

Out of the 100 reviewed samples, the most remarkable translation demonstrated the model's advanced preservation of recognizing context:

Halunder	Eewifelns wear de Trap no med en Klap, de Ferrislap uuder Ferrislach namt, ufskottet."
English Translation	Often the stairs were additionally separated/locked by a hatch, called Ferrislap or Ferrislach.
Reference Translation	"Oft war die Treppe noch mit einer Klappe, Falltür oder Bodenluke genannt, abgetrennt
Model Translation	"Häufig war die Treppe noch mit einer Klappe, die Ferrislap oder Ferrislach genannt wurde, abgeschlossen."

This translation retains the original Halunder terms *Ferrislap* and *Ferrislach*, which refer to a specific attic hatch people had inside their house on Heligoland, rather than replacing them with presumed German equivalents. By employing a relative clause ("die *Ferrislap* oder *Ferrislach* genannt wurde"), it appears the model recognized that the sentence is presenting the name of the hatch rather than requiring a literal translation, thereby producing a more natural German sentence structure.

The following table (**Table 5.3**) demonstrates a sample of translation results reviewed and compares the result of the fine-tuned model to the expected translation.

Table 5.3 Representative examples of manually evaluated sentences

Rating	Halunder -- Expected -- Translation Result	Explanation
6	De Witkläow wear iiwensoo hoog as 'e Kläowkant fan ii Lun, oawers fel smaaler.	The model recognized “Witkläow” as an original name and further excelled in terms of fluency, word choice and flow.
	Die Weiße Klippe war ebenso hoch wie die Kliffkante von unserem Helgoland, aber viel schmaler.	
	Die Weiße Klippe war ebenso hoch wie die Felskante von Helgoland, aber viel schmaler.	
6	li leäs halunder Kudder lait dear uk, de "Heimat".	The model excelled over the expected translation by choosing a more precise term.
	Unser letzter helgoländischer Kutter liegt dort auch, die "Heimat".	
	Unser letzter Helgoländer Kutter liegt dort auch, die "Heimat".	
4	Allerhand Boaders stun iip 'e Miir tu hoggeln noa Gölken en Makkereal en uk Skolken.	Even though the model did not use the proper term for “wall”, it has generated a much more fluent translation that differs from the word order in Halunder.
	Allerhand Badegäste stehen auf der Mauer zum Angeln nach Dorschen, Makrelen und auch Schollen.	
	Allerlei Badegäste stehen auf der Elbe um nach Dorschen und Makrelen und auch Schollen zu angeln.	
2	Wi hoa Kathalsen, Miisken en Letj Buurn.	While the general sentence is correct, the model has mistaken a bird’s name with “little bees”.
	Wir haben Silbermöwen, Dreizehenmöwen und Sturmmöwen.	
	Wir haben Silbermöwen, Möwen und kleine Bienen.	
1	Ho feels 'e di bürigen dren iip Lun uun 'e Wonter!	This translation is unacceptable and introduces a major flaw that can lead to embarrassment if used.
	Wie fühlst du dich geborgen zuhause auf Helgoland im Winter!	
	Wie fühlt man die Geburten im Winter auf Helgoland!	

5.1.2 Interpreting Translation Results

The manual evaluation reveals a strong polarity in the model's Halunder-to-German translations: while more than half (56%) are rated as "really good" (i.e., good, flawless, or perfect) and could be mistaken for human-produced text, a third (32%) are "really bad", containing major errors that make them practically unusable. Such a pattern reveals the model's ambivalence—on the one hand demonstrating authentic and accurate translations that could have substantial impact (see Section 5.2), on the other creating doubt in the overall trust one could have regarding the translator. The majority of these severe issues, such as confusing *Möwen* (seagulls) with *Bienen* (bees) (See Table 5.3) though at least retaining that a flying animal is used are rooted in incorrect vocabulary choices rather than deep grammatical or structural flaws. But shortcomings must be considered in the context of the evolution of Halunder itself: even many natives consistently produce errors, primarily due to the late introduction of grammar rules (Århammar, 1995). The available material that was used for training likely contains inconsistent vocabulary usage itself (See Section 3.2.2). Thus, while the model's results are encouraging, it will require further refinement, additional data, and targeted error prevention strategies proposed in Section 5.3 to achieve a performance that users and institutions can trust in and rely on.

5.2 Future Research

Building on the results of this thesis, several options for future research remain viable, unexplored and likely beneficial. Future improvements could involve translating texts such as the large amount of Halunder Spreek (See Section 3.1) texts using the existing translator and systematically reviewing confidence levels of the output. Translations falling below a defined confidence threshold would be transferred to native speakers for approval or correction.

Expanding the dataset through the systematic construction of sentences would not expand the sentence structure variety but could further enrich the vocabulary known

to the transformer. Standardized formulas could be employed to generate new sentences based on dictionary entries or other linguistic materials, ensuring consistency and extending the translator's lexical coverage. This effort should be coordinated with locals to only perform data augmentation - in this case adding artificially generated sentences – that conforms to contextual correctness. For instance, the creation of standardized sentence structures for a) Names, b) Animals, c) places, d) family relationships, e) occupations, and more seem feasible.

One example could be the substitution of “Today I caught three cods” to “Yesterday my father caught two trouts”.

Incorporating conventional data augmentation approaches, such as back-translation, synonym replacement, and paraphrasing, could help increase the diversity and volume of training data in relationship to the German pre-trained data (Oh et al., 2023). These techniques would enhance the model's robustness and generalizability, particularly for low-frequency words, grammatical issues at the input level, and uncommon syntactic structures (Oh et al., 2023).

5.3 Impact

The development of the Halunder-German translation system demonstrates that Transformer-based technology can serve as a crucial element in preserving an endangered language facing extinction. Drawing from as few as 10,000 parallel sentence pairs and relying on a very small speaker community, this work achieves a level of translation quality that enables the creation of learning materials, general communication, and the creation of documents suitable for educational and cultural initiatives. In this way, it directly aligns with the European Charter for Regional or Minority Languages, which supports the presence of minority languages in education, media, and public life (Council of Europe, 1992).

By successfully producing translations and confirming that Transformer models function well even with extremely limited data, this research encourages other communities to undertake similar projects. Related languages facing similar issues,

yet at a slower pace (Steensen, 1994)—such as Fering, Öömrang, Sylter, and Saterfrisian—can adapt these methods to produce their own digital resources. The same approach might also be applied beyond North Frisian languages, wherever urgent conservation measures are needed to maintain linguistic diversity. Previous attempts to preserve Halunder through printed guides or occasional classes struggled to engage a younger generation (Wanke, 2008), but this new translator can help provide accessible and modern tools that resonate with contemporary community members.

In addition, this system lays the groundwork for future technological advancements. Expanded datasets, AI-driven conversational tools, and mobile applications could bring Halunder closer to everyday life. Enhanced resources may include interactive language learning platforms that introduce the language to those with little or no prior exposure. As more materials become available, accuracy and cultural specificity could improve over time. By increasing the presence of Halunder across various formats, there is a stronger possibility of long-term maintenance, allowing future generations to value their linguistic heritage (Tan & Jehom, 2024).

This project supports the idea that languages, even those at severe risk, can gain renewed traction if supported by innovative tools combined with local participation (Galla, 2018). Digital translation efforts can provide a model for addressing language endangerment caused by socioeconomic pressures, limited prestige, and reduced intergenerational transmission (Århammar, 1993; Council of Europe, 2009)

Finally, this achievement answers the main research question affirmatively: the work shows that Transformer technology can be considered decisive for preserving Halunder and similarly at-risk languages. It proves that technical solutions, backed by communities and designed for practical use, may hold the key to ensuring that languages once nearing extinction find new life and relevance in a modern, connected world.

Chapter 6

Conclusion

6.1 Summary of Key Contributions

This thesis presents the first machine translation system for Halunder, one of Europe's smallest and most endangered languages (Krüss, 1985; Montigny, 1994c). The development of www.halunder.ai, a freely accessible translation platform, creates the foundation for future improvements through community engagement and continuous development. The creation of a 10,000-sentence parallel corpus represents a significant contribution, as no comparable digital resource previously existed for Halunder.

The translation system serves several critical functions beyond direct language conversion. It enables the development of modern educational materials, preserves currently untranslated texts, and creates a digital archive of linguistic knowledge that will persist regardless of future speaker population trends. This work demonstrates that meaningful language technology development remains possible even with extremely limited resources, supporting Grützmacher's (2020) call for urgent action in Frisian minority language preservation.

The project establishes documented procedures for other endangered language communities, particularly within the North Frisian language family. The methodologies developed here indicate that effective preservation efforts can begin

immediately, without waiting for perfect conditions or extensive resources - a critical consideration given the small number of remaining speakers.

6.2 Limitations and Future Recommendations

Several limitations require acknowledgment. The rapidly evolving nature of machine learning technology means current approaches may quickly become outdated (Bird, 2020). Translation performance varies significantly based on architectural choices and hyperparameter settings, making it challenging to establish definitively optimal approaches for languages facing the same fate as Halunder.

Dataset quality and composition significantly impact translation performance (Koehn & Knowles, 2017). The current corpus, while substantial given the language's situation, may not fully represent all linguistic aspects of Halunder. The scope focuses solely on text-to-text translation, leaving other important applications unexplored. Technical improvements could include collecting additional parallel texts, incorporating monolingual data, and developing specific handling for idioms and cultural expressions. The real-world impact of the Halunder translator cannot be evaluated at this stage, it requires longer-term observation and community feedback (Mager et al., 2018).

6.3 Stakeholder Recommendations

The preservation of Halunder requires coordinated action across multiple domains. Educational institutions must prioritize the reintegration of Halunder into formal curricula, addressing what Grützmacher (2020) identifies as critical gaps in current language education. The European Charter for Regional or Minority Languages mandates active support for minority languages in public life (Council of Europe, 1992), requiring local governments to increase visible Halunder usage through bilingual documentation, signage, and institutional communications.

Community engagement remains essential for sustainable language preservation.

This includes addressing what Wanke (2008) terms *Sprachselbstmord* (“language suicide”) - where minority languages gradually fade through the exoticization of speakers and their reduction to cultural performances rather than daily communication. Cultural institutions must support natural language use through increased Halunder-language publications and programming, while also facilitating exchanges with other minority language communities facing similar challenges.

The translator developed in this thesis serves as a foundational tool supporting these broader preservation efforts. Its implementation helps fulfill governmental obligations under both European and regional frameworks, while providing practical support for educational activities and communication.

6.4 Closing Remarks

While technological solutions alone cannot save a language from extinction, they can document it for future generations who might wish to reclaim their heritage. Ultimately, a language lives through its speakers - through parents teaching their children first words, *through people daring to use it in shops and streets despite pressures to switch to German, through poets and writers expressing their thoughts in Halunder verses*, through the courage of people choosing to learn and use their heritage language in a world that pushes toward linguistic uniformity. *The future of Halunder depends not just on tools and technology, but on its community finding the strength to carry their language forward.*

6.5 References

- Anthropic. (2024). *Claude 3.5 Sonnet* [Large Language Model]. <https://anthropic.com>
- Århammar, N. (1985, May). Leitfaden des Helgoländischen. *Der Helgoländer*, Nr. 250, 19.
- Århammar, N. (1993, May). Helgoländische Sprache ist bedroht. *Der Helgoländer*, Nr. 347, 24.
- Århammar, N. (1994, April). Beerigers und tu beerigen keem—Ein Beitrag zum Helgoländer Brauchtum um die Konfirmation. *Der Helgoländer*, Nr. 358, 18.
- Århammar, N. (1995, December). Überlebenschancen des Helgoländischen. *Der Helgoländer*, Nr. 378, 26–30.
- Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate* (No. arXiv:1409.0473). arXiv. <http://arxiv.org/abs/1409.0473>
- Bird, S. (2020). Decolonising Speech and Language Technology. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 3504–3519). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.313>
- Bradley, D., & Bradley, M. (2002). *Language Endangerment and Language Maintenance: An Active Approach* (1st ed.). <https://doi.org/10.4324/9781315028811>
- Callison-Burch, C., Osborne, M., & Koehn, P. (2006). Re-evaluating the Role of Bleu in Machine Translation Research. In D. McCarthy & S. Wintner (Eds.), *11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 249–256). Association for Computational Linguistics. <https://aclanthology.org/E06-1032>
- Cassels, M. (2019). *Indigenous languages in new media: Opportunities and challenges for language revitalization*.
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder–Decoder Approaches. In D. Wu, M. Carpuat, X. Carreras, & E. M. Vecchi (Eds.), *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation* (pp.

103–111). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-4012>

Council of Europe. (1992, November 5). *European Charter for Regional or Minority Languages (ETS No. 148)*. <https://www.coe.int/en/web/conventions/full-list?module=treaty-detail&treatynum=148>

Council of Europe. (2009, January 8). *The Council of Europe's Framework Convention for the Protection of National Minorities*. <https://doi.org/10.1017/CBO9780511575372.006>

Dale, D. (2022). The first neural machine translation system for the Erzya language. In O. Serikov, E. Voloshina, A. Postnikova, E. Klyachko, E. Neminova, E. Vylomova, T. Shavrina, E. L. Ferrand, V. Malykh, F. Tyers, T. Arkhangel'skiy, V. Mikhailov, & A. Fenogenova (Eds.), *Proceedings of the first workshop on NLP applications to field linguistics* (pp. 45–53). International Conference on Computational Linguistics. <https://aclanthology.org/2022.fieldmatters-1.6>

Dornis, T. W., & Stober, S. (2024). *Copyright Law and Generative AI Training—Technological and Legal Foundations (Urheberrecht und Training generativer KI-Modelle—Technologische und juristische Grundlagen)* (SSRN Scholarly Paper No. 4946214). Social Science Research Network. <https://papers.ssrn.com/abstract=4946214>

Fisherman, J. (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages*.

Flora, C. (2021, September). *Preserving and Revitalizing Endangered Minority Languages*. https://entan.org/wp-content/uploads/2021/10/Entan-Policy-Paper-No.-2_final-1.pdf?utm_source=chatgpt.com

Galiano-Jiménez, A., Sánchez-Martínez, F., Sánchez-Cartagena, V. M., & Pérez-Ortiz, J. A. (2023). Exploiting large pre-trained models for low-resource neural machine translation. In M. Nurminen, J. Brenner, M. Koponen, S. Latomaa, M. Mikhailov, F. Schierl, T. Ranasinghe, E. Vanmassenhove, S. A. Vidal, N. Aranberri, M. Nunziatini, C. P. Escartín, M. Forcada, M. Popovic, C. Scarton, & H. Moniz (Eds.), *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 59–68). European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.7>

Galla, C. (2018). Digital Realities of Indigenous Language Revitalization: A Look at Hawaiian Language Technology in the Modern World. *Language and Literacy*, 20(3), 100–120. <https://doi.org/10.20360/langandlit29412>

- Grützmacher, L. (2020). *Schriftliche Stellungnahme zum Thema „Maßnahmen für mehr Friesischunterricht“*.
<https://www.landtag.ltsh.de/infothek/wahl19/umdrucke/03700/umdruck-19-03786.pdf>
- Gu, J., Hassan, H., Devlin, J., & Li, V. O. K. (2018). Universal Neural Machine Translation for Extremely Low Resource Languages. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 344–354). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-1032>
- Hutchins, W. J. (1995). Machine Translation: A Brief History. In *Concise History of the Language Sciences* (pp. 431–445). Elsevier. <https://doi.org/10.1016/B978-0-08-042580-1.50066-0>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). *Scaling Laws for Neural Language Models* (No. arXiv:2001.08361). arXiv. <http://arxiv.org/abs/2001.08361>
- Karakanta, A., Dehdari, J., & Van Genabith, J. (2018). Neural machine translation for low-resource languages without parallel corpora. *Machine Translation*, 32(1–2), 167–189. <https://doi.org/10.1007/s10590-017-9203-5>
- Koehn, P., & Knowles, R. (2017). Six Challenges for Neural Machine Translation. In T. Luong, A. Birch, G. Neubig, & A. Finch (Eds.), *Proceedings of the First Workshop on Neural Machine Translation* (pp. 28–39). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W17-3204>
- Köhn, B. (2020, February 19). Große kleine Sprache Helgoländisch. *TraLaLit*. <https://www.tralalit.de/2020/02/19/grosse-kleine-sprache-helgolaendisch/>
- Kozhircbayev, Z. (2024). Enhancing Neural Machine Translation with Fine-Tuned mBART50 Pre-Trained Model: An Examination with Low-Resource Translation Pairs. *Ingénierie Des Systèmes d'Information*, 29(3), 831–838. <https://doi.org/10.18280/isi.290304>
- Krogmann, W. (1952a). Aus der Werkstatt des Helgoländischen Wörterbuchs. In *Helgoland ruft*.
- Krogmann, W. (1952b). Helgoland und die Helgoländer. In *Helgoland ruft*.
- Krogmann, W. (1964). *Helgoländer Wörterbuch. Lfg. 4. Harti – Kortel*. Verlag der Akademie der Wissenschaften und der Literatur.

-
- Krüss, J. (1985, April). Helgoland: Eine Sprache verstummt. *GEO*, Nr. 4.
- Kuhn, R., Davis, F., Désilets, A., Joanis, E., Kazantseva, A., Knowles, R., Littell, P., Lothian, D., Pine, A., Running Wolf, C., Santos, E., Stewart, D., Boulianne, G., Gupta, V., Maracle Owennatékhá, B., Martin, A., Cox, C., Junker, M.-O., Sammons, O., ... Souter, H. (2020). The Indigenous Languages Technology project at NRC Canada: An empowerment-oriented approach to developing language software. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 5866–5878). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.516>
- Lewis, M., & Simons, G. F. (2010). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique*.
- Luong, M.-T., Pham, H., & Manning, C. D. (2015). *Effective Approaches to Attention-based Neural Machine Translation* (No. arXiv:1508.04025). arXiv. <http://arxiv.org/abs/1508.04025>
- Luong, T., Sutskever, I., Le, Q., Vinyals, O., & Zaremba, W. (2015). Addressing the Rare Word Problem in Neural Machine Translation. In C. Zong & M. Strube (Eds.), *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)* (pp. 11–19). Association for Computational Linguistics. <https://doi.org/10.3115/v1/P15-1002>
- Mager, M., Gutierrez-Vasques, X., Sierra, G., & Meza, I. (2018). *Challenges of language technologies for the indigenous languages of the Americas* (No. arXiv:1806.04291). arXiv. <http://arxiv.org/abs/1806.04291>
- Montigny, J. C. (1985, May). Wichtig ist immer noch das gedruckte Wort. *Der Helgoländer*, Nr. 250.
- Montigny, J. C. (1992, May). Spracharbeit auf Helgoland und das Helgoländer Wörterbuch von Prof. Nils Arhammar. *Der Helgoländer*, Nr. 335.
- Montigny, J. C. (1994a). *Ist die helgoländische Sprache zum Untergang verurteilt? Konzeption, Anfertigung und Analyse eines journalistischen Produkts über Rettungsversuche einer Kleinsprache in Deutschland. Abschlussarbeit im weiterbildenden Studium Journalisten-Weiterbildung am Fachbereich Kommunikationswissenschaften der Freien Universität Berlin.*
- Montigny, J. C. (1994b, February). Halunder Spreek ist zahlenmäßig ganz weit hinten

anzusiedeln. *Der Helgoländer*, Nr. 356.

Montigny, J. C. (1994c, February). Lernt und Sprech mehr Helgoländisch. *Der Helgoländer*, Nr. 356.

Nägler, C. (2010). *Zukunftsperspektiven des Helgoländischen. Eine Studie zur Sprachbedrohung und Sprecherhaltung* [Bachelor Thesis].

Nguyen, T. Q., & Chiang, D. (2017). *Transfer Learning across Low-Resource, Related Languages for Neural Machine Translation* (No. arXiv:1708.09803). arXiv. <https://doi.org/10.48550/arXiv.1708.09803>

NLLB Team, Costa-jussà, M. R., Cross, J., Çelebi, O., Elbayad, M., Heafield, K., Heffernan, K., Kalbassi, E., Lam, J., Licht, D., Maillard, J., Sun, A., Wang, S., Wenzek, G., Youngblood, A., Akula, B., Barrault, L., Gonzalez, G. M., Hansanti, P., ... Wang, J. (2022). *No Language Left Behind: Scaling Human-Centered Machine Translation* (No. arXiv:2207.04672). arXiv. <https://doi.org/10.48550/arXiv.2207.04672>

Oh, S., Lee, S. A., & Jung, W. (2023). *Data Augmentation for Neural Machine Translation using Generative Language Model* (No. arXiv:2307.16833). arXiv. <https://doi.org/10.48550/arXiv.2307.16833>

Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., & Auli, M. (2019). *fairseq: A Fast, Extensible Toolkit for Sequence Modeling* (No. arXiv:1904.01038). arXiv. <https://doi.org/10.48550/arXiv.1904.01038>

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). *PyTorch: An Imperative Style, High-Performance Deep Learning Library* (No. arXiv:1912.01703). arXiv. <https://doi.org/10.48550/arXiv.1912.01703>

Pichler, W. (2024, June 27). *Helgoländisch ist kein deutscher Dialekt: Halunder in Not*. General-Anzeiger Bonn. https://ga.de/meinung/auf-helgoland-kaempft-eine-kleine-sprache-ums-ueberleben_aid-115126673

Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. In O. Bojar, R. Chatterjee, C. Federmann, B. Haddow, C. Hokamp, M. Huck, V. Logacheva, & P. Pecina (Eds.), *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 392–395). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W15-3049>

Post, M. (2018). *A Call for Clarity in Reporting BLEU Scores* (No. arXiv:1804.08771).

arXiv. <https://doi.org/10.48550/arXiv.1804.08771>

- Quast, G. (1994, April). Bedroht Sprachen, sterbende Völker. *Der Helgoländer*, Nr. 358.
- Repplinger, R. (2009, January 7). Helgoland: Halunder für Anfänger. *Die Tageszeitung: taz*. <https://taz.de/Helgoland/!5169954/>
- Sanches, J., Ribeiro, R., & Coheur, L. (2024). *From Brazilian Portuguese to European Portuguese* (No. arXiv:2408.07457). arXiv. <http://arxiv.org/abs/2408.07457>
- Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1715–1725). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P16-1162>
- Siebs, T. (1910). *Helgoland und seine Sprache. Beiträge zur Volks- und Sprachkunde* (2. Auflage).
- Steensen, T. (1994). *The Frisians in Schleswig-Holstein*. Nordfriisk Instituut.
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to Sequence Learning with Neural Networks* (No. arXiv:1409.3215). arXiv. <https://doi.org/10.48550/arXiv.1409.3215>
- Tan, Y., & Jehom, W. J. (2024). The Function of Digital Technology in Minority Language Preservation: The Case of the Gyalrong Tibetan Language. *Preservation, Digital Technology & Culture*, 53(3), 165–177. <https://doi.org/10.1515/pdte-2024-0021>
- Tang, Y., Tran, C., Li, X., Chen, P.-J., Goyal, N., Chaudhary, V., Gu, J., & Fan, A. (2020). *Multilingual Translation with Extensible Multilingual Pretraining and Finetuning* (No. arXiv:2008.00401). arXiv. <http://arxiv.org/abs/2008.00401>
- Tanzer, G., Suzgun, M., Visser, E., Jurafsky, D., & Melas-Kyriazi, L. (2024). *A Benchmark for Learning to Translate a New Language from One Grammar Book* (No. arXiv:2309.16575). arXiv. <https://doi.org/10.48550/arXiv.2309.16575>
- Tiedemann, J., & Thottingal, S. (2020). OPUS-MT – Building open translation services for the World. In A. Martins, H. Moniz, S. Fumega, B. Martins, F. Batista, L. Coheur, C. Parra, I. Trancoso, M. Turchi, A. Bisazza, J. Moorkens, A. Guerberof, M. Nurminen, L. Marg, & M. L. Forcada (Eds.), *Proceedings of the 22nd Annual*

Conference of the European Association for Machine Translation (pp. 479–480). European Association for Machine Translation. <https://aclanthology.org/2020.eamt-1.61>

UNESCO Language Vitality and Endangerment. (2003). *Language Vitality and Endangerment. International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages.*

van der Wees, M., Bisazza, A., & Monz, C. (2016). Measuring the Effect of Conversational Aspects on Machine Translation Quality. In Y. Matsumoto & R. Prasad (Eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 2571–2581). The COLING 2016 Organizing Committee. <https://aclanthology.org/C16-1242>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (No. arXiv:1706.03762). arXiv. <http://arxiv.org/abs/1706.03762>

Vinyals, O., Kaiser, L., Koo, T., Petrov, S., Sutskever, I., & Hinton, G. (2015). *Grammar as a Foreign Language* (No. arXiv:1412.7449). arXiv. <http://arxiv.org/abs/1412.7449>

Wanke, P. (2008). *Das Helgoländische*. Kulturwissenschaftliche Fakultät der Europa-Universität Viadrina in Frankfurt (Oder).

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. In Q. Liu & D. Schlangen (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>

Ziyin, L., Chen, B., Wang, R., Liang, P. P., Salakhutdinov, R., Morency, L.-P., & Ueda, M. (2020). *Learning Not to Learn in the Presence of Noisy Labels* (No. arXiv:2002.06541). arXiv. <https://doi.org/10.48550/arXiv.2002.06541>

Zoph, B., Yuret, D., May, J., & Knight, K. (2016). Transfer Learning for Low-Resource Neural Machine Translation. In J. Su, K. Duh, & X. Carreras (Eds.), *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1568–1575). Association for Computational Linguistics. <https://doi.org/10.18653/v1/D16-1163>

Appendix A

Config	Learning Rate	Warmup Ratio	Weight Decay	Label Smoothing	Dropout	Attention Dropout	LR Scheduler	Gradient Accum. Steps	Description
C1	2e-5	0.1	0.01	0.08	0.1	0.1	linear	1	Baseline configuration
C2	2e-5	0.1	0.01	0.1	0.3	0.3	linear	1	High dropout configuration
C3	3e-5	0.15	0.02	0.1	0.2	0.2	cosine	1	Cosine schedule with medium dropout
C4	8e-5	0.05	0.001	0.05	0.05	0.05	linear	1	Aggressive learning with minimal regularization
C5	1e-5	0.2	0.04	0.15	0.4	0.4	linear	2	Conservative learning with heavy regularization
C6	4e-5	0.12	0.02	0.12	0.25	0.25	polynomial	1	Polynomial decay with medium-high regularization
C7	5e-5	0.1	0.015	0.1	0.15	0.15	constant_with_warmup	4	Stepped learning rate with low dropout
C8	3e-5	0.15	0.03	0.2	0.35	0.15	cosine_with_restarts	2	Mixed regularization with cosine restarts
C9	2.5e-5	0.15	0.025	0.15	0.32	0.25	cosine_with_restarts	2	Hybrid of top performing C2 and C8 configurations
C10	3.5e-5	0.13	0.015	0.13	0.28	0.28	polynomial	2	Enhanced polynomial decay with balanced regularization
C11	2e-5	0.18	0.035	0.18	0.38	0.28	cosine	3	High regularization with pure cosine schedule

Appendix B

Points	Description
6	A translation that surpasses the reference translation and is superior in terms of natural flow in the target language.
5	A perfect translation. The meaning and style are fully conveyed, grammar and word choice are correct, and the text reads naturally.
4	A good translation. The meaning is fully or almost fully conveyed, and the style and word choice feel natural to the target language.
3	An acceptable translation. The overall meaning is conveyed; errors in word choice and grammar do not prevent understanding, and most of the text is grammatically correct and in the target language.
2	A poor translation. The text is mostly understandable and largely in the target language, but there are significant errors in meaning, grammar, or word choice.
1	An unusable translation. A large portion of the text is in the wrong language, unintelligible, or bears little relation to the original.