



**University of  
Zurich**<sup>UZH</sup>

## **Institute of Computational Linguistics**

---

Andreasstrasse 15  
CH-8050 Zürich-Oerlikon  
Switzerland

**Prof. Dr. Martin Volk**  
Institute of  
Computational Linguistics  
Phone +41 44 635 43 17  
volk@cl.uzh.ch  
<https://www.cl.uzh.ch>

Andreas Schaufelbühl  
Binzmühlestrasse 14  
8050 Zürich

Matrikel-Nr. 12-918-843  
[andreas.schaufelbuehl@uzh.com](mailto:andreas.schaufelbuehl@uzh.com)

March 5, 2019

### **Master's Thesis Specification**

#### **Titel: ???**

#### **Introduction**

Natural machine translation [NMT] play a paramount role in the daily life of human beings. The usability of computational translated phrases is strictly bound to correctly and naturally translation close to these of professional human translator. Consequently, researcher put enormous effort in improving the state-of-the-art translation technologies, with the goal of producing natural accurate sentences. As it has the potential to support manual translation by increasing efficiency and reduce the amount of work drastic being done by humans in this area.

In recent years, the research of NMT have presented many different, novel models using Deep learning technologies. Using state of the art machine learning technologies, machine translation are able to produce more precise and natural translation and outperform existing stale models. Hokamp and Liu [2] introduces the model of Grid Beam Search, allowing lexically constrained decoding of text, which improves the translation quality in interactive scenarios. Using constraints models are able to generate more context near output or generate translation, which must include the given lexically constraints within its output translation

Post and Vilar [4] extends the Grid Beam Search model and introduces the Dynamic Beam Allocation, which improves the efficiency for lexically constrained decoding with beam search. Further Vaswani et al. [5] present their simple neural network architecture, called Transformer, based solely on attention mechanism. The transformer also achieves remarkable performance in NMT in terms of scoring, but also time efficiency.

These models show already very good performance in lexically constrained translation. However, non of the current research in this area is concerning about adaptive word endings in the process of sequence-to-sequence translation. As translation of the corresponding base word is done often correctly, we examine a lack of translating the correct word ending.



## The goals of this master's thesis

We argue that using state of the art technologies in NMT and extend existing models (e.g. [3]) with additional processes focusing on word endings can improve the quality of translated text. Focusing on ending of single tokens, we hope to detect patterns for generalizing endings of words better with respect to their context. Thus the high level goal of this thesis is to extend existing Decoder with a lemmatizing-approach and investigate their benefit in terms of performance improvement. The research questions that guiding this thesis are:

- RQ1: *To what extend is it possible to improve existing state of the art NMT by adding lemmatizer for token translation in a sequence-to-sequence model?*
- RQ2: *Is it possible to design an approach, which can use these improvements in practice?*

## Task description

The main tasks of this thesis are:

1. Read up on the current state of the art in industry and research in the areas relevant to the thesis, including Attention-based models, Natural machine translator and Lexically Constrained Decoding.
2. Get familiar with currently existing open-source solutions in the area of Multi-Source Transformer Decoding.
3. Setup up an environment using one of these solution and integrate an score testing approach.
4. Develop and implement the model extension in code.
5. Perform empirical test in order to identify research results.
6. Writing an academic report explaining and summarizing the results from the work on items 1 to 5.

## Deliverables

The major milestones of the project are as follows:

| When                                     | What  |
|--|---|
| 1 <sup>st</sup> and 2 <sup>st</sup> week | State of the Art review is finished. Open Source tools are examined in-depth, so the student understands them.                                      |
| 3 <sup>st</sup> and 4 <sup>st</sup> week | Basis tool for development is extended with the testing environment and ready for implementation of the model. The model is planned more in detail. |
| 2 <sup>st</sup> month                    | First approach of proof-of-concept is implemented.  |
| 3 <sup>rd</sup> month                    | Evaluation and adjustment of first version model is done.   |
| 4 <sup>rd</sup> month                    | Second approach with experiments is implemented.  |
| 5 <sup>th</sup> month                    | The proof-of-concept implementation is finished.  |
| 6 <sup>th</sup> month                    | Thesis is written, proof-of-concept is fully functional, documented and delivered.  |



## **Provided resources**

There is no need of special provided resources to realise this thesis

## **General thesis guidelines**

The typical rules of academic work must be followed. In [1] Bernstein describes a number of guidelines which must be followed. At the end of the thesis, a final report has to be written. The report should clearly be organized, follow the usual academic report structure, and has to be written in English. As implementing software is also part of this thesis, state-of-the-art design, coding, and documentation standards for the software have to be obeyed.

## **Advisors:**

### **Professor:**

Prof. Dr. Martin Volk

### **Responsible assistant:**

Samuel Läubli

## **Signatures:**

Andreas Schaufelbühl

Prof. Dr. Martin Volk



## References

- [1] A. Bernstein. So what is a (diploma) thesis? a few thoughts for first-timers. Technical report, Dynamic and Distribution Information System Group, University of Zurich, 2005.
- [2] C. Hokamp and Q. Liu. Lexically constrained decoding for sequence generation using grid beam search. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1535–1546. Association for Computational Linguistics, 2017".
- [3] J. Libovický, J. Helcl, and D. Marecek. Input combination strategies for multi-source transformer decoder. *CoRR*, abs/1811.04716, 2018.
- [4] M. Post and D. Vilar. Fast lexically constrained decoding with dynamic beam allocation for neural machine translation. *CoRR*, abs/1804.06609, 2018.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.