SCHOOL OF COMPUTATION,
INFORMATION AND TECHNOLOGY —
INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Morphological Inflection Generation

Murilo Escher Pagotto Ronchi

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY — INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Bachelor's Thesis in Informatics

Morphological Inflection Generation

Titel der Abschlussarbeit

Author:	Murilo Escher Pagotto Ronchi
Examiner:	Shu Okabe Ph.D.
Supervisor:	Advisor
Submission Date:	Submission date

I confirm that this bachelor's thesis is my own work and I have documented all sources and material used.

Munich, Submission date Murilo Escher Pagotto Ronchi

Acknowledgments

Abstract

Contents

1 Introduction

1.1 Section

Citation test [Lam94].

Acronyms must be added in `main.tex` and are referenced using macros. The first occurrence is automatically replaced with the long version of the acronym, while all subsequent usages use the abbreviation.

E.g. `\ac{TUM}, \ac{TUM}` $\Rightarrow$ Technical University of Munich (TUM), TUM

For more details, see the documentation of the `acronym` package[1].

1.1.1 Subsection

See Table 1.1, Figure 1.1, Figure 1.2, Figure 1.3.

Table 1.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

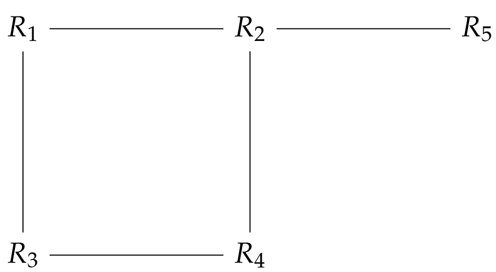

Figure 1.1: An example for a simple drawing.

[1] `https://ctan.org/pkg/acronym`

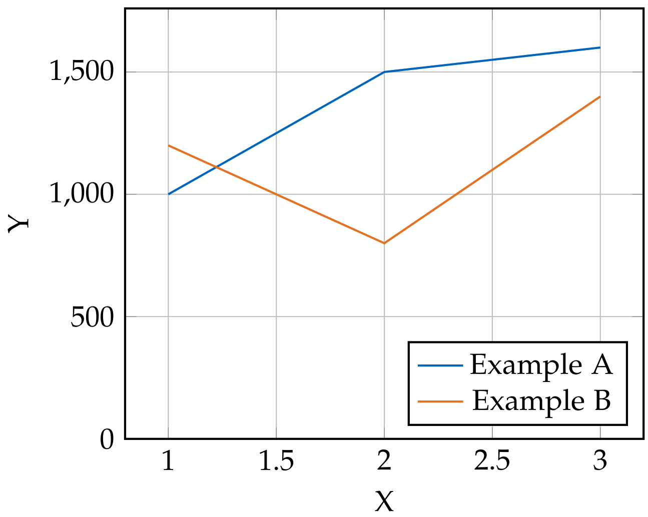

Figure 1.2: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 1.3: An example for a source code listing.

2 Background

Different models have been used in this thesis to tackle the problem of morphological inflection generation. This chapter presents the theoretical background of the models used, as well as some related concepts.

2.1 Non-Neural Baseline

The first model evaluated was the non-neural baseline provided in the Sigmorphon 2023 shared task [Gol+23] repository. This is a statistical model based on the work first presented in [Cot+17], which uses a set of edit rules extracted from the training data to perform morphological inflection generation.

This rule-based system learns a set of prefix and suffix transformation rules from the training data. For each morphological tag (MSD), the model extracts the possible prefix and suffix transformation rules that map the lemma to the inflected form. During inference, given a lemma and a target MSD, the model applies the most frequent rules associated with the specific transformation to generate the inflected form.

For example, given the Portuguese verb *andar* (to walk) and the target MSD `V;PRS;3;SG` (verb, present tense, third person singular), the model might learn the suffix transformation rule `AR -> A` from the training data. Applying this rule to the lemma *amar* would yield the inflected form *ama*.

Verb	Target MSD	Inflected Form	Transformation Rule
andar	`V;PRS;3;SG`	*anda*	`AR -> A`
amar	`V;PRS;3;SG`	*ama*	`AR -> A`

Table 2.1: Example of transformation rule learned by the non-neural baseline

How the model works, is that it first extracts all possible prefix and suffix transformation rules from the training data. For each lemma-inflected form pair, it identifies the longest common prefix and suffix, and derives the transformation rules by comparing the remaining parts of the lemma and inflected form. The model keeps track of the frequency of each rule for each MSD.

2.2 Neural Baseline

Placeholder text.

2.3 ByT5

Placeholder text.

2.4 Large Language Models

Placeholder text.

3 Datasets and Languages

3.1 Datasets

Explain dataset structure, with stem, lemma and inflection. Explain morphological features.

3.2 Languages

Placeholder text.

4 Methodology

4.1 Section

Placeholder text.

5 Results

5.1 Section

Placeholder text.

6 Conclusion

6.1 Section

Placeholder text.

Abbreviations

TUM Technical University of Munich

List of Figures

List of Tables

Bibliography

[Cot+17] R. Cotterell, C. Kirov, J. Sylak-Glassman, G. Walther, E. Vylomova, P. Xia, M. Faruqui, S. Kübler, D. Yarowsky, J. Eisner, and M. Hulden. *CoNLL-SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection in 52 Languages*. 2017. arXiv: 1706.09031 [cs.CL].

[Gol+23] O. Goldman, K. Batsuren, S. Khalifa, A. Arora, G. Nicolai, R. Tsarfaty, and E. Vylomova. "SIGMORPHON–UniMorph 2023 Shared Task 0: Typologically Diverse Morphological Inflection." In: *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*. Ed. by G. Nicolai, E. Chodroff, F. Mailhot, and Ç. Çöltekin. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 117–125. DOI: 10.18653/v1/2023.sigmorphon-1.13.

[Lam94] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.