

# Analyzing the Potential of Source Sentence Reordering in Statistical Machine Translation for Chinese

Master Thesis of

Ge Wu

At the Department of Informatics  
Institute for Anthropomatics and Robotics (IAR)

Advisor: Alex Waibel  
Second Advisor: Yuqi Zhang

Duration: 1st February 2014 – 15th June 2014



## Abstract

fsaffafa asf fsa sa fa

---

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

**15th June 2014**

# Contents

<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Objective and Contribution . . . . .	1
1.3. Structure . . . . .	2
<b>2. Related Work</b>	<b>3</b>
2.1. Section 1 . . . . .	3
2.2. Section 2 . . . . .	3
<b>3. Foundations</b>	<b>5</b>
<b>4. Reordering Approach</b>	<b>7</b>
<b>5. Evaluation</b>	<b>9</b>
5.1. Section 1 . . . . .	9
5.2. Section 2 . . . . .	9
5.3. Section 3 . . . . .	9
<b>6. Conclusion</b>	<b>11</b>
<b>Appendix</b>	<b>13</b>
A. First Appendix Section . . . . .	13
<b>Bibliography</b>	<b>15</b>



# 1. Introduction

## 1.1. Motivation

Word reordering is a general issue when we want to translate text from one language to the other. Different languages normally have different word reordering and the difference could be huge, when two languages are isolated from each other. Depend on the language itself, the word reordering could have very distinguish features. For example, 45% of the languages in the world has a subject-object-verb(SOV) order. Unlike in English, verbs are put after object in these languages. Japanese is a popular language among them. Instead of saying “The black cat climbed to the tree top.”, people would say “The black cat the tree top to climbed.” in Japanese. Another example is Spanish, in which people often put the adjective after the modified nouns. An example from the paper [LP13] shows how people would order the words differently:

English	The black cat climbed to the tree top.
Japanese	The black cat the tree top to climbed.
Spanish	The cat black climbed to the top tree.

Figure 1.1.: Word orders of three different languages

Since different word orders are a common issue among languages, we propose several pre-reordering methods and evaluate them in this thesis. Before translation, the words in source language are rearranged into a similar word order as the target language’s through these methods. With the appropriate word order, better translation quality will be achieved.

## 1.2. Objective and Contribution

The ground of this thesis are three papers about data driven, rule based pre-reordering: [RV], [NK], [HWNW].

asset is data driven

original (mltilayer)

try to extend to other language

hiarchical

conclusion goal is

### 1.3. Structure



## 2. Related Work

a The content chapters of your thesis should of course be renamed. How many chapters you need to write depends on your thesis and cannot be said in general.

Check our the examples theses in the Wiki.

Of course, you can split this .tex file into several files if you prefer.

### 2.1. Section 1

...

### 2.2. Section 2

...

Add additional content chapters if required.



### 3. Foundations



## 4. Reordering Approach



## 5. Evaluation

...

### 5.1. Section 1

...

### 5.2. Section 2

...

### 5.3. Section 3

...





## 6. Conclusion

...



# Appendix

## A. First Appendix Section

ein Bild

Figure A.1.: A figure

...



# Bibliography

- [HWNW] T. Hermann, J. Weiner, J. Niehues, and A. Waibel, “Analyzing the potential of source sentence reordering in statistical machine translation.”
- [LP13] U. Lerner and S. Petrov, “Source-side classifier preordering for machine translation,” in *Proc. of EMNLP ’13*, 2013.
- [NK] J. Niehues and M. Kolss, “A pos-based model for long-range reorderings in smt.”
- [RV] K. Rottmann and S. Vogel, “Word reordering in statistical machine translation with a pos-based distortion model.”