

$\Sigma$

# Paraphrasing In Context Of Computer Aided Translation

*Turan Rustamli*



Master of Science  
Artificial Intelligence  
School of Informatics  
University of Edinburgh  
2013

# **Abstract**

This doctoral thesis will present the results of my work into the reanimation of lifeless human tissues.

# Acknowledgements

Many thanks to my mummy for the numerous packed lunches; and of course to Igor, my faithful lab assistant.

# Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

*(Turan Rustamli)*

To my mummy.

# Table of Contents

# Chapter 1

## Introduction

### 1.1 Motivation

Today, taking advantage of advancements in machine translation, we can read and understand text in languages unknown to us. However, producing a high quality translation is a task that is still solved by humans. Computer Aided Translation (CAT) systems aim to support and assist translators, providing them with tools that partially automate and simplify their workflow. Demand for these systems is huge, each year billions are spent on translation by international organisations such as European Union.

Several assistance ways, that could be offered by a CAT system, were introduced since investigation of the question began. The most popular ones are sentence-completion and post-editing. While the first approach is providing suggestions while user is typing translation, the second approach prefills text input area with a machine translation.

In this report we present a novel assistance way that aims to help translator to fix an initial machine generated translation, by providing paraphrases for any user-specified part of the sentence.

Paraphrasing is the process of representing the same idea with different words. Some recent publications suggest that paraphrasing and translation tasks are closely related and information used to perform one task, can be useful for another (ccb). Considering this we are introducing paraphrasing approach that efficiently reuses the information available after machine translation process which output is used to provide post-editing assistance within a CAT system.

Moreover, taking advantage of interactivity of a CAT process, we investigated paraphrasing as an interactive task, implementing a system that accepts and responds to user's feedback on the paraphrases that were initially produced.



We will also describe a novel automatic evaluation approach that we used to test feasibility and usefulness of paraphrasing results. Outcomes of a set of user studies will be provided to back up the results of automatic evaluation.

## 1.2 Overview

In this section we would like to provide an outline of the report.

We will start the following chapter by introducing background material in the next chapter. Recent advancements in CAT will be presented and references to previous publications about paraphrasing will be provided. We will also discuss some basis concepts of machine translation, particularly Search Graph representation, which will be used by us to solve paraphrasing task.

In the next chapter we will focus on providing description of work that was done in scope of our project. First, we will start by discussing design of our approach and explaining it's possible ways of integration with existing CAT systems. We will continue by providing details of implementation of our design. Description of challenging problems, that were solved during design and implementation stages, and their solutions will be given as well.

We will present our evaluation methodology, stating results of testing various versions of our paraphrasing algorithm in the Chapter 4. This includes description of an automatic evaluation framework which was implemented by us during the project, it's advantages and disadvantages. Next, we will provide results of a set of user studies carried out by us in order to support results of automatic evaluation.

Following chapter will introduce novel experiments by making paraphrasing within a CAT system more interactive. Specifically, we will present a design and implementation of a feature that lets a user to instantly interact with the system and provide improved results. In this chapter we will discuss "Show More Like This" button that was added by us to the paraphrases list displayed to a user, which is similar to ideas of relevance feedback () in Information Retrieval in order to provide user with more relevant paraphrases. A modification that was made to our automatic evaluation framework in order to support this new feature will be also introduced, as well as results of both automatic and manual evaluation.

Next, we will look into another way of adding more interactivity to the user interface of the paraphrasing functionality, by implementing and adaptive paraphrase button that will let user to initially suggest us the reason of the paraphrasing request. We will

discuss situations when users might find paraphrasing functionality useful interacting with a CAT system, providing results of user studies.

In our final chapter we will conclude our report and suggest interesting directions for the future work.