

# Proposal: Unsupervised Machine Translation with Word2Vec

II2202, Fall 2015

Bram Leenders  
KTH, Royal Institute of Technology  
Brinellvägen 8, 114 28 Stockholm  
Sweden  
b.c.leenders@gmail.com

Marc Romeyn  
KTH, Royal Institute of Technology  
Brinellvägen 8, 114 28 Stockholm  
Sweden  
marc.romeyn@gmail.com

## ABSTRACT

Add an abstract or comment this part out

## Keywords

Word2Vec, Machine Translation

## 1. INTRODUCTION

Since the introduction of word2vec [2, 4], blah blah

### 1.1 Problem Statement

The project will investigate how to...

### 1.2 Prior Work

Mention (at least) [1, 3, 5]

Describe both the broader spectrum of word2vec/NLP and machine translation. Position our research (i.e. we skip grammar), and be sure to limit our project to per-word translation (not full grammatically correct sentences).

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

## 2. RESEARCH GOAL

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

### 2.1 Research Purpose

Abstract description of goals

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

## 2.2 Goals

Pretty much what our deliverables are...

## 2.3 Tasks

What do we have to do to achieve this?

- Build (at least) two working setups for automated translation
- Make a verification mechanism
- Verify which setup works better

## 3. PLANNING

A global overview of the milestones we defined in our research is as follows;

### *Half September - October: Literary Study*

During this phase, a list of relevant papers (e.g. [1, 3, 5]) is collected, short-listed to a readable size and read. Each paper will get a brief (informal) summarization to capture the essence of the paper insofar as it is relevant to this research. These summaries help process the information and provide quick access during the practical research.

Not all papers relevant to the research will be read during this phase, so next phases will include a fair share of reading.

### *October: Preparing experiments*

Description

### *November: Running experiments*

Description

### *Before 12 December: Incorporating results in paper*

During this phase, we will incorporate the results obtained in the experiments into our final paper.

## 4. ALLOCATION OF RESPONSIBILITIES

Marc pledges to provide sufficient beer. Bram will take care of coffee and possibly some Club-Maté.

Nunc sed pede. Praesent vitae lectus. Praesent neque justo, vehicula eget, interdum id, facilisis et, nibh. Phasellus at

purus et libero lacinia dictum. Fusce aliquet. Nulla eu ante placerat leo semper dictum. Mauris metus. Curabitur lobortis. Curabitur sollicitudin hendrerit nunc. Donec ultrices lacus id ipsum.

## 5. REFERENCES

- [1] O. Levy, Y. Goldberg, and I. Ramat-Gan. Linguistic regularities in sparse and explicit word representations. *CoNLL-2014*, page 171, 2014.
- [2] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [3] T. Mikolov, Q. V. Le, and I. Sutskever. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*, 2013.
- [4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [5] L. Wolf, Y. Hanani, K. Bar, and N. Dershowitz. Joint word2vec networks for bilingual semantic representations. *International Journal of Computational Linguistics and Applications*, 5(1):27–44, 2014.