

Learning to Understand Languages with Neural Networks



Jörg Tiedemann
Department of Digital Humanities
University of Helsinki
jorg.tiedemann@helsinki.fi

HELSINGIN YLIOPISTO
HELSINKIUNIVERSITET
UNIVERSITY OF HELSINKI



Learning to understand

Dad - What is “gin” ?

From explanations

- “It’s something to drink”
- “... for adults, you know ...”



Grounding

- this is a bottle of gin ...
- this is someone who drank gin ...



Reinforcement

- “Give me that bottle!”
- “This is alcohol! ... nothing for you, man!

HELSINGIN YLIOPISTO
HELSINKIUNIVERSITET
UNIVERSITY OF HELSINKI



Learning from multilingual data

Translations as explanations



- each translation gives a different view
- translations are like naturally occurring reformulations

Grounding through multimodality



- add visual features
- add sound

Reinforce semantics

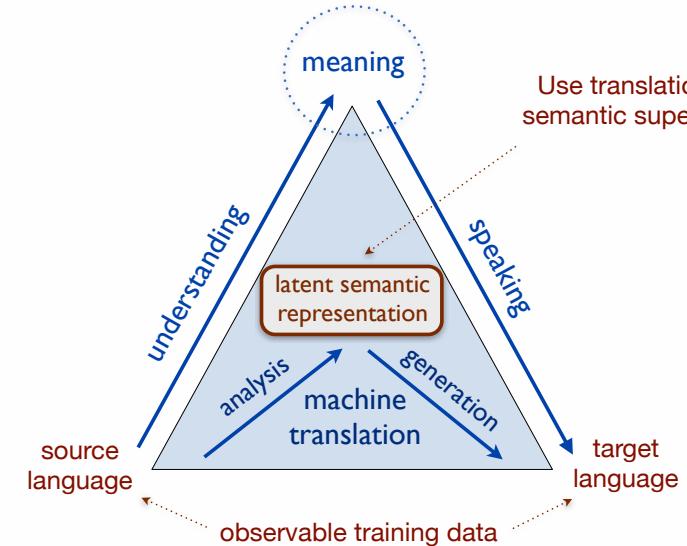
- supervised training objectives that require understanding
- machine translation and other tasks

HELSINGIN YLIOPISTO
HELSINKIUNIVERSITET
UNIVERSITY OF HELSINKI



Machine translation and meaning

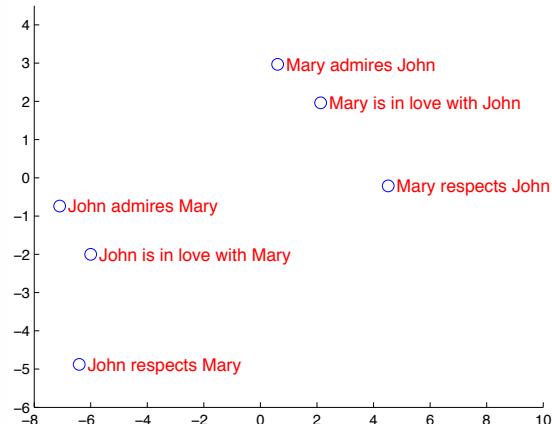
Use translations as semantic supervision!



HELSINGIN YLIOPISTO
HELSINKIUNIVERSITET
UNIVERSITY OF HELSINKI



Semantic representation learning

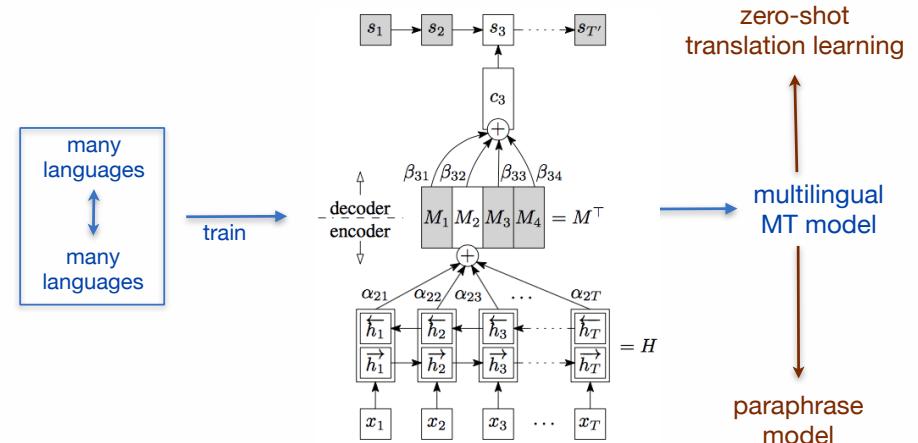


(From Sutskever et. al. et. al: "Sequence to Sequence Learning with Neural Networks")

HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



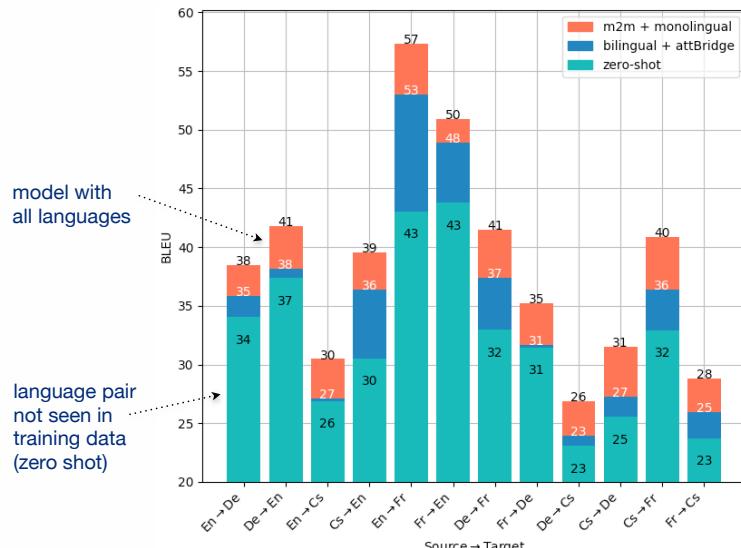
Multilingual Machine Translation (MT)



HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



Multilingual MT & Zero-Shot Translation



HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



Multilingual MT & Paraphrasing

register login

Fix your language!

detect language Afrikaans Danish Dutch Catalan Estonian Finnish Faroese French Frisian Galician German Hungarian Icelandic Italian Norwegian Occitan Portuguese Spanish Swedish

Valsch geschreibt is nich gut!
Das Pferd hat gelaufen.
Ich haben fertig.
Wir sein kommen.
wat morkelst du denn da rum?
Icke geb dir dann och noch wat zu trinken.
Dat is nix für meinereiner!
Mein Fuß ist brechen! Ich muss nach die dokter.

Falsch geschrieben ist nicht gut! Das Pferd ist gelaufen.
Ich bin fertig. Wir kommen. Was hast du denn da zu suchen? Dann gebe ich dir noch etwas zu trinken. Das ist nichts für mich! Mein Fuß ist gebrochen! Ich muss zum Arzt.

https://translate.ling.helsinki.fi/fix_language

HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



Multilingual MT & Paraphrasing



Fix your language!

detect language	Afrikaans	Danish	Dutch	Catalan	Estonian	Finnish	Faroese	French
Frisian	Galician	German	Hungarian	Icelandic	Italian	Norwegian	Occitan	Portuguese
Spanish	Swedish							

Huset är mögligt.
Framgång är möglig

moldy / musty

Huset är mögligt. Framgång är möglig

The Haus is
musty.

Success is
possible

https://translate.ling.helsinki.fi/fix_language

HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HELSINKIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI



What does the system attend to?

we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.

size of
attention-
bridge

(a) k = 1

we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.
we cannot afford to lose more of the momentum that existed at the beginning of the N_ineties.

(b) k = 10



Translation-related projects

Data collection for MT
> 200 languages



<http://opus.nlpl.eu>

Data collection & MT
in 2 languages



<https://blogs.helsinki.fi/fiskmo-project/>

<http://blogs.helsinki.fi/language-technology/>

audiovisual data & MT
in 6 languages



MeMAD

Methods for Managing
Audiovisual Data

<https://memad.eu>

semantics & MT > 1,000 languages

