WORDNET FOR DEFINITION AUGMENTATION WITH ENCODER-DECODER ARCHITECTURE

Konrad Wojtasik, Arkadiusz Janz, Bartłomiej Alberski, Maciej Piasecki
Email: konrad.wojtasik@pwr.edu.pl

Wrocław University of Science and Technology
Department of Artificial Intelligence

AGENDA

1. Introduction

2. Datasets

3. Model

4. Experiment Setting

5. Evaluation

6. Results

7. Future Work

INTRODUCTION

The main goals:

- Generate definitions for missing senses and augment plWordNet.

- Test models capacity for generalization and ability to use knowledge from pretraining.

- Incorporate the world knowledge into the model.

INTRODUCTION

Example:

Input: "**cat**: the **cat** jumped on the bed in the middle of the night"

Output: "cat: feline mammal usually having thick soft fur and no ability to roar, domestic cats"

DATASETS

- **The Corpus of Wrocław University of Science and Technology (KPWr)** – corpus with documents from various domains. 100 documents where manually annotated with plWordNet senses.

- **Verb's Valency Dictionary** – sense-annotated treebank (Składnica), benchmark for knowledge-based WSD

- **Sherlock Holmes: The Adventure of The Speckled Band** – translated to polish language and manually annotated with sense tags.

- **Monosemous Lemmas from KGR10** – single sense lemmas usage example were extracted from the KGR10 corpus (large corpus with 4 billion tokens)
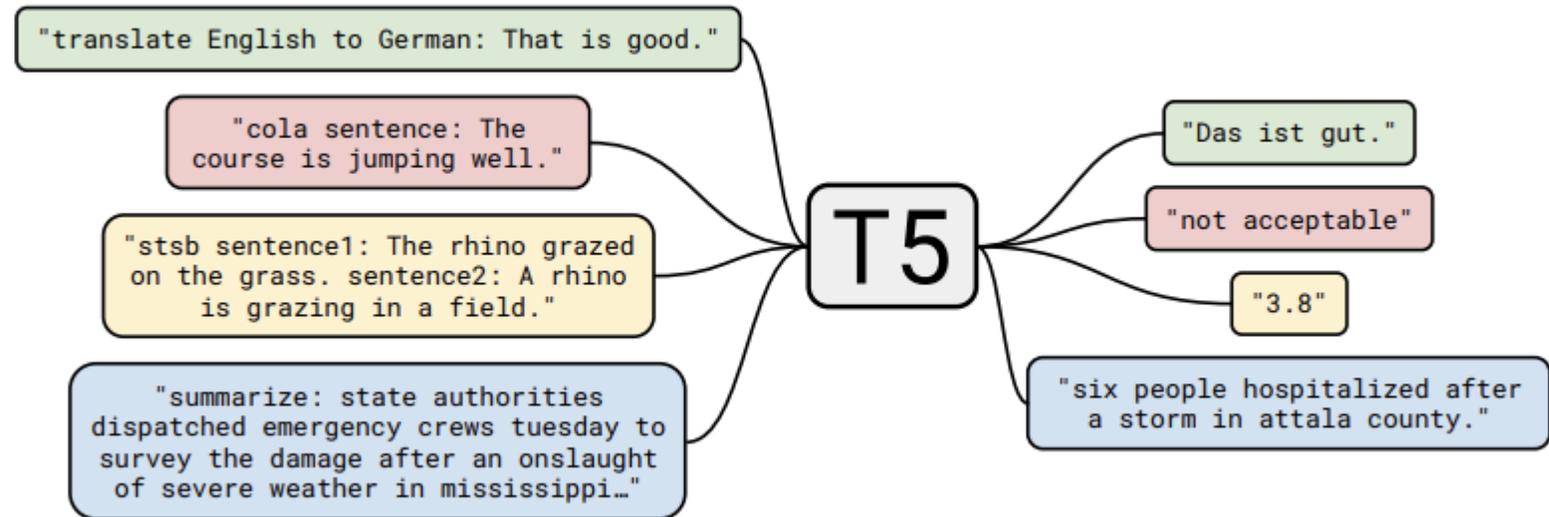
DATASETS

We acquired almost 237k examples: words, usage examples and definitions.

➢213k were acquired from plWordNet and KGR10

➢16k from Verb's Valency Dictionary

➢6.2k from The Corpus of Wroclaw University of Science and Technology (KPWr),

➢1.5k Sherlock Holmes.

TEST DATASETS

- 10k randomly sampled examples from the whole data collection

- 146 examples were used in human evaluation:
  - 102 senses were already in plWordNet
  - 44 had no definition in plWordNet.

MODEL

"translate English to German: That is good."

"cola sentence: The course is jumping well."

"stsb sentence1: The rhino grazed on the grass. sentence2: A rhino is grazing in a field."

"summarize: state authorities dispatched emergency crews tuesday to survey the damage after an onslaught of severe weather in mississippi…"

T5

"Das ist gut."

"not acceptable"

"3.8"

"six people hospitalized after a storm in attala county."

Source: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer, 2020

T5 is an encoder-decoder model pre-trained on multiple tasks, where each task is converted into a text-to-text format. Each task has its' own prompt, for example:

"translate English to German: "

EXPERIMENTAL SETTING

pIT5-base and pIT5-large models both available on huggingface.

Batch size:
- 16 (pIT5-base)
- 4 (pIT5-large)

Prompt: "generate definition:"

Trained on Nvidia RTX3090 GPU.

EVALUATION

To evaluate generated texts, we used **BLEU** and **ROUGE-L** metrics widely used in machine translation.

Both metrics are readily available in Evaluate library.

https://github.com/huggingface/evaluate

HUMAN EVALUATION

Mistakes classification:

- **Self-reference/definiendum** – word is defined by the word itself

HUMAN EVALUATION

przyświecać, verb	„W rzeczywistości zapewne **przyświecała** mi zupełnie bezsensowna myśl, żeby spróbować ułożyć z kamieni coś w rodzaju muru."	przyświecać, przyświecać komuś;	**definiendum**
(to guide)	(In fact, I was probably **guided** by a completely nonsensical thought to try to arrange something like a wall out of the stones.)	to guide, to guide someone;	

HUMAN EVALUATION

Mistakes classification:

- **Self-reference/definiendum** – word is defined by the word itself

- **Incoherent** – definition contains contradictions

HUMAN EVALUATION

| znieczulony, adj | „Nic nie czułem! — Był pan **znieczulony**, sir. Zresztą podobno to nie boli." | o człowieku: taki, który jest pod wpływem znieczulenia, ma kontakt z rzeczywistością; | **incoherent** |
| (anesthetized) | (I didn't feel anything! - You were **anesthetized**, sir. Anyway, supposedly it doesn't hurt.) | about a person: one who is under anesthesia, is in touch with reality; | |

HUMAN EVALUATION

Mistakes classification:

- **Self-reference/definiendum** – word is defined by the word itself

- **Incoherent** – definition contains contradictions

- **Completely-wrong** – word is defined as wrong sense

HUMAN EVALUATION

warkliwy, adj	„*Gdy spytałem go o tego człowieka, przyznał* **warkliwym** *tonem, że jest najlepszy z całej zgrai.*"	taki, który jest pełen warczenia, warczący;	**completely-wrong**
(snarky)	(*When I asked him about the man, he admitted in a* **snarky** *tone that he was the best of the bunch.*)	One that is full of growling, growling	

HUMAN EVALUATION

Mistakes classification:

- **Self-reference/definiendum** – word is defined by the word itself

- **Incoherent** – definition contains contradictions

- **Completely-wrong** – word is defined as wrong sense

- **Partially-wrong** – part of the definition is incorrect or refers to a different sense

HUMAN EVALUATION

| wikłać, verb	„Niestety, dziewczyna pada ofiarą zakazanej miłości i **wikła się** w niebezpieczny romans."	przenośnie: pakować się w coś, dawać się wciągnąć czemuś lub komuś;	**partially-wrong**
(get involved)	(*Unfortunately, the girl falls prey to forbidden love and **get involved** in a dangerous affair.*)	metaphorically: get sucked into something or someone;	

HUMAN EVALUATION

Evaluation settings:

- **Hard evaluation** – generated definition is correct without any errors

- **Soft evaluation** – generated definition is correct, even if self-reference or partially-wrong errors are spotted.

RESULTS
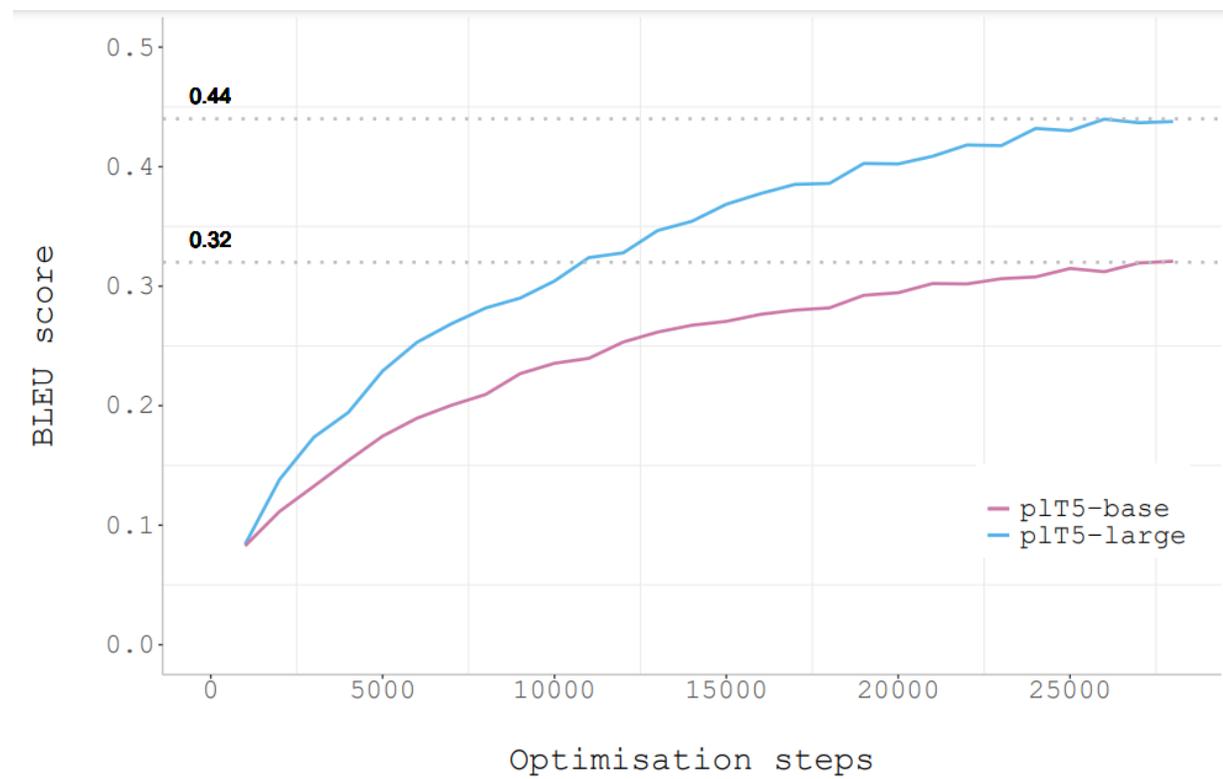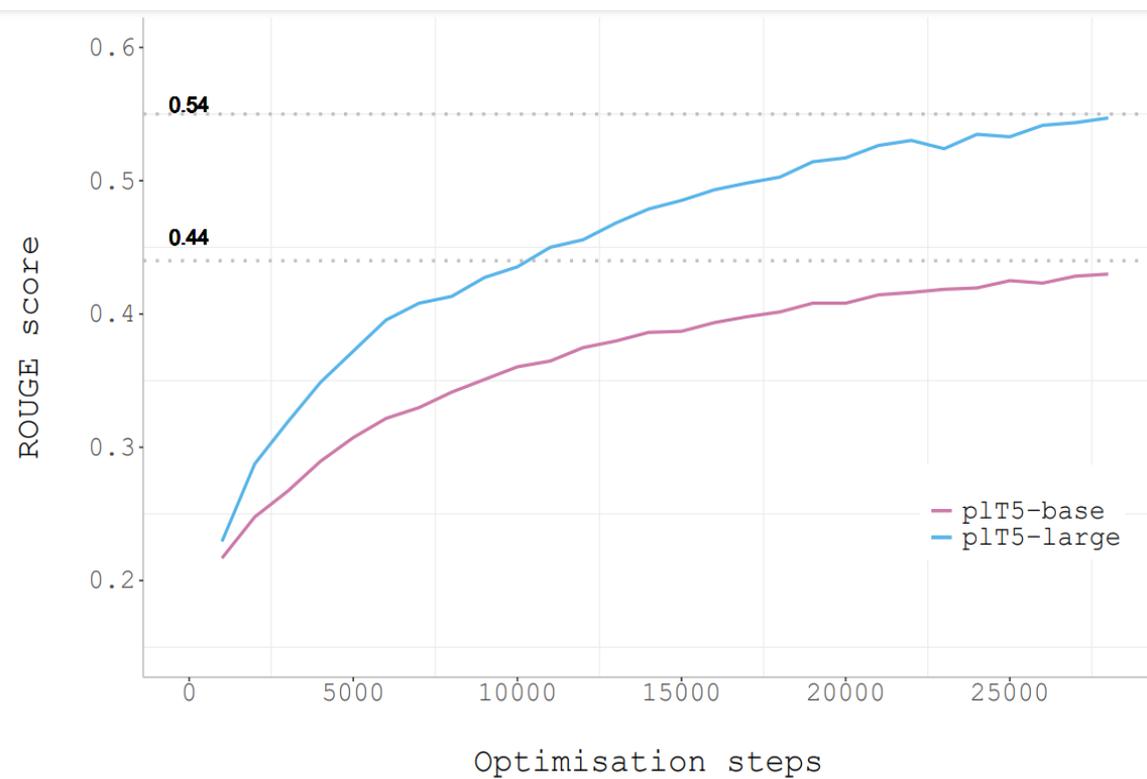
RESULTS

Model	All samples		WordNet$^+$		WordNet$^-$	
	hard eval.	*soft eval.*	*hard eval.*	*soft eval.*	*hard eval.*	*soft eval.*
plT5-base	0.43	0.62	0.82	0.95	0.27	0.54
plT5-large	0.59	0.74	0.95	0.99	0.37	0.64

WordNet+ – already in plWordNet

WordNet- – not defined in plWordNet

RESULTS

Polysemic Word	Context	Generated Definition
zamek, noun	„Na górze widniał wielki, piękny **zamek** otoczony fosą.”	budowla mieszkalna, rezydencja pana, króla, księcia lub magnata.
(castle)	(On the mountain was a large, beautiful **castle** surrounded by a moat.)	A residential building, the residence of a lord, king, prince or magnate.
zamek, noun	„ Masz rozpięty **zamek** w spodniach.”	wyrób tekstylny, który ma zamek, jest skonstruowany w taki sposób, że można go zamknąć.
(zipper)	(You have the **zipper** of your pants undone.)	a textile product that has a lock is constructed in such a way that it can be closed.
zamek, noun	„Dorobił sobie klucz do **zamka**.”	urządzenie do zamykania np. drzwi, szuflad, walizek.
(lock)	(He made up a key for the **lock**.)	A device for locking, for example, doors, drawers, suitcases.
zamek, noun	„Po raz któryś z kolei odciągnął **zamek** i zajrzał do komory nabojowej swego kałasznikowa.”	mechanizm broni palnej, wyposażony w ruchomy zamek.
(bolt)	(For the umpteenth time, he pulled back the **bolt** and looked into the cartridge chamber of his kalashnikov.)	firearms mechanism, equipped with a movable bolt.

FUTURE WORK

➢Apply the definition generation task to Word Sense Disambiguation (WSD)

➢Improve on knowledge intensive tasks like Entity Linking or Question Answering, which require a lot of knowledge and language understanding.

THANK YOU FOR YOUR ATTENTION!

Email: konrad.wojtasik@pwr.edu.pl