



IMPACT  
**OLKi**



Utrecht University



# CoDWoE

## SemEval 2022 Shared Task 1

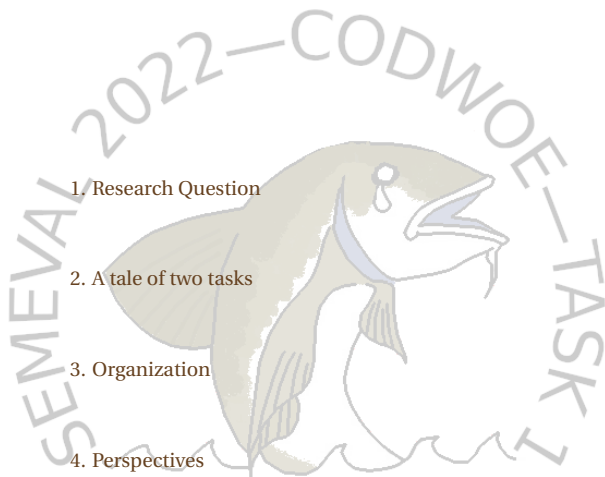
Timothee Mickus



Utrecht University



# Outline



1. Research Question

2. A tale of two tasks

3. Organization

4. Perspectives



Utrecht University



ANALYSE ET TRAITEMENT  
INFORMATIQUE  
DE LA LANGUE FRANÇAISE



UNIVERSITÉ  
DE LORRAINE

# Outline



Utrecht University



ANALYSE ET TRAITEMENT  
INFORMATIQUE  
DE LA LANGUE FRANÇAISE



UNIVERSITÉ  
DE LORRAINE

# Research Question

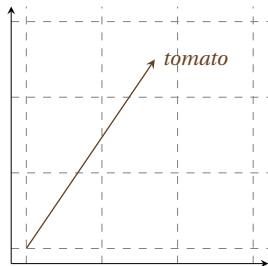
- We can show whether  $AMR = SDRT$  or whether  $HPSG = LFG$

# Research Question

- ▶ We can show whether  $\text{AMR} = \text{SDRT}$  or whether  $\text{HPSG} = \text{LFG}$
- ▶ What about non-formal semantics?

# Research Question

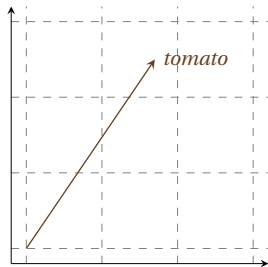
- We can show whether AMR = SDRT or whether HPSG = LFG
- What about non-formal semantics?



*A widely cultivated plant,  
Solanum lycopersicum,  
having edible fruit.*

# Research Question

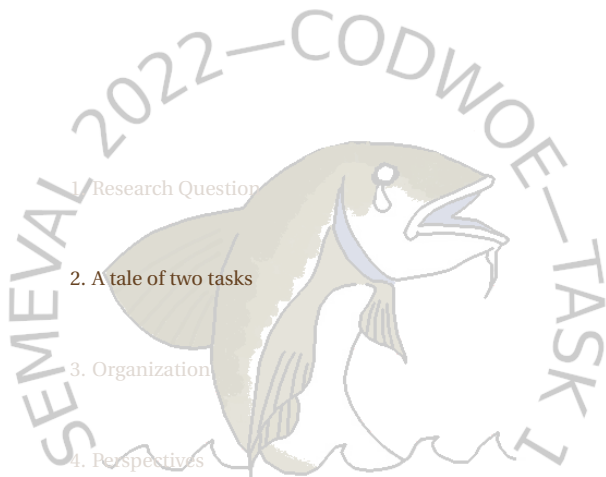
- We can show whether AMR = SDRT or whether HPSG = LFG
- What about non-formal semantics?



*A widely cultivated plant,  
Solanum lycopersicum,  
having edible fruit.*

- **Research topic:** the equivalence of dictionaries and distributional semantics models as semantic theories

# Outline



1. Research Question

2. A tale of two tasks

3. Organization

4. Perspectives



Utrecht University

atilf

ANALYSE ET TRAITEMENT  
INFORMATIQUE  
DE LA LANGUE FRANÇAISE



UNIVERSITÉ  
DE LORRAINE



# A tale of two tasks

## Definition Modeling

- ▶ Noraset et al. (2017): generate a definition from the embedding of the word being defined

# A tale of two tasks

## Definition Modeling

- ▶ Noraset et al. (2017): generate a definition from the embedding of the word being defined
- ▶ Gadetsky, Yakubovskiy, and Vetrov (2018), Mickus, Paperno, and Constant (2019), Zhang et al. (2019), and Bevilacqua, Maru, and Navigli (2020): ambiguous definienda need context

# A tale of two tasks

## Definition Modeling

- ▶ Noraset et al. (2017): generate a definition from the embedding of the word being defined
- ▶ Gadetsky, Yakubovskiy, and Vetrov (2018), Mickus, Paperno, and Constant (2019), Zhang et al. (2019), and Bevilacqua, Maru, and Navigli (2020): ambiguous definienda need context
- ▶ Yang et al. (2019), Mickus, Constant, and Paperno (2020), and Kabiri and Cook (2020): more than English

# A tale of two tasks

Reverse dictionary

- Reverse dictionary: retrieve a word from a description of its meaning

# A tale of two tasks

Reverse dictionary

- ▶ Reverse dictionary: retrieve a word from a description of its meaning
- ▶ Approaches:

# A tale of two tasks

Reverse dictionary

- ▶ Reverse dictionary: retrieve a word from a description of its meaning
- ▶ Approaches:
  1. Crawford and Crawford (1997), Bila et al. (2004), El Khalout and Oflazer (2004), and Shaw et al. (2013): IR-based

# A tale of two tasks

Reverse dictionary

- ▶ Reverse dictionary: retrieve a word from a description of its meaning
  
- ▶ Approaches:
  1. Crawford and Crawford (1997), Bila et al. (2004), El Khalout and Oflazer (2004), and Shaw et al. (2013): IR-based
  2. Dutoit and Nugues (2002) and Thorat and Choudhari (2016): graph-based

# A tale of two tasks

Reverse dictionary

- ▶ Reverse dictionary: retrieve a word from a description of its meaning
  
- ▶ Approaches:
  1. Crawford and Crawford (1997), Bila et al. (2004), El Khalout and Oflazer (2004), and Shaw et al. (2013): IR-based
  2. Dutoit and Nugues (2002) and Thorat and Choudhari (2016): graph-based
  3. Méndez, Calvo, and Moreno-Armendáriz (2013), Calvo, Méndez, and Moreno-Armendáriz (2016), and Zanzotto et al. (2010): vector-based



# A tale of two tasks

## Reverse dictionary

- ▶ Reverse dictionary: retrieve a word from a description of its meaning
  
- ▶ Approaches:
  1. Crawford and Crawford (1997), Bila et al. (2004), El Khalout and Oflazer (2004), and Shaw et al. (2013): IR-based
  2. Dutoit and Nugues (2002) and Thorat and Choudhari (2016): graph-based
  3. Méndez, Calvo, and Moreno-Armendáriz (2013), Calvo, Méndez, and Moreno-Armendáriz (2016), and Zanzotto et al. (2010): vector-based
  4. Hill et al. (2016), Zhang et al. (2020), Yan et al. (2020), Qi et al. (2020), and Malekzadeh, Gheibi, and Mohades (2021): LM-based
  
- ▶ Zhang et al. (2020), Yan et al. (2020), Qi et al. (2020), Malekzadeh, Gheibi, and Mohades (2021), Bila et al. (2004), and El Khalout and Oflazer (2004): more than English

# A tale of two tasks

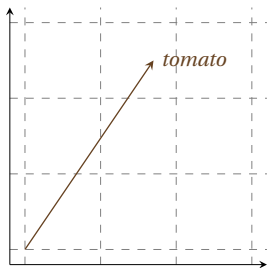
Inverse functions

- Two inverse functions?

# A tale of two tasks

Inverse functions

## ► Two inverse functions?



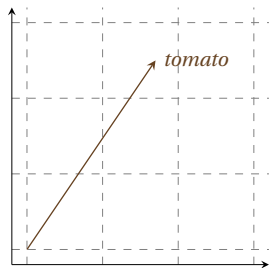
Definition Modeling

*A widely cultivated plant,  
Solanum lycopersicum,  
having edible fruit.*

# A tale of two tasks

Inverse functions

## ► Two inverse functions?



Definition Modeling

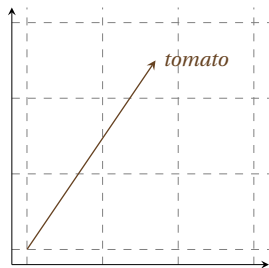
Reverse Dictionary

*A widely cultivated plant,  
Solanum lycopersicum,  
having edible fruit.*

# A tale of two tasks

Inverse functions

- Two inverse functions?



Definition Modeling

Reverse Dictionary

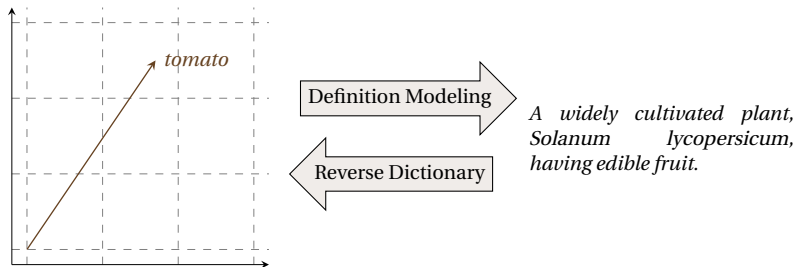
*A widely cultivated plant,  
Solanum lycopersicum,  
having edible fruit.*

- We need to reframe the two tasks:

# A tale of two tasks

Inverse functions

## ► Two inverse functions?



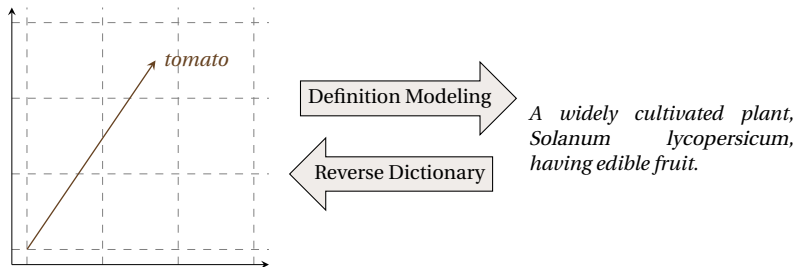
## ► We need to reframe the two tasks:

- Reverse Dictionary as the task of reconstructing an embedding given its definition gloss

# A tale of two tasks

Inverse functions

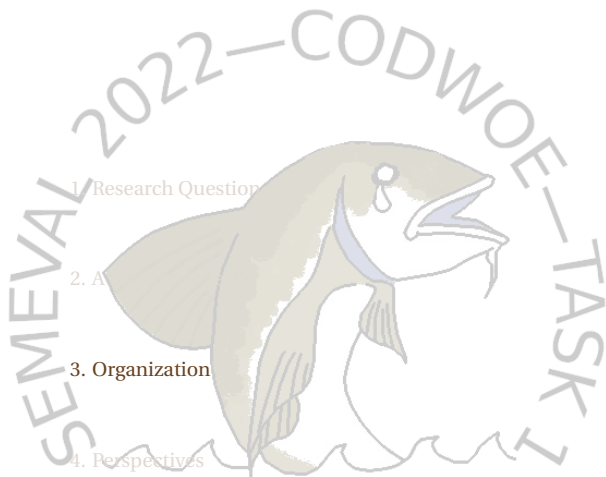
## ► Two inverse functions?



## ► We need to reframe the two tasks:

- Reverse Dictionary as the task of reconstructing an embedding given its definition gloss
- Definition Modeling as context-less Definition Modeling

# Outline



1. Research Question

2. A

3. Organization

4. Perspectives



Utrecht University

atilf

ANALYSE ET TRAITEMENT  
INFORMATIQUE  
DE LA LANGUE FRANÇAISE



UNIVERSITÉ  
DE LORRAINE



# Organization

## Datasets

- We need diverse data:

# Organization

## Datasets

- ▶ We need diverse data:
  - ▶ Diverse embedding architectures: `w2v`, `electra`, character-based embeddings (for comparison)

# Organization

## Datasets

- ▶ We need diverse data:
  - ▶ Diverse embedding architectures: `w2v`, `electra`, character-based embeddings (for comparison)
  - ▶ Diverse languages: `en`, `es`, `fr`, `it`, `ru`

# Organization

## Datasets

- ▶ We need diverse data:
  - ▶ Diverse embedding architectures: `w2v`, `electra`, character-based embeddings (for comparison)
  - ▶ Diverse languages: `en`, `es`, `fr`, `it`, `ru`
- ▶ We need comparable data:

# Organization

## Datasets

- ▶ We need diverse data:
  - ▶ Diverse embedding architectures: `w2v`, `electra`, character-based embeddings (for comparison)
  - ▶ Diverse languages: `en`, `es`, `fr`, `it`, `ru`
- ▶ We need comparable data:
  - ▶ Data in comparable amounts

# Organization

## Datasets

- ▶ We need diverse data:
  - ▶ Diverse embedding architectures: `w2v`, `electra`, character-based embeddings (for comparison)
  - ▶ Diverse languages: `en`, `es`, `fr`, `it`, `ru`
- ▶ We need comparable data:
  - ▶ Data in comparable amounts
  - ▶ Data from comparable sources (`dbnary` & Wikipedia, OPUS, Gutenberg)

# Organization

Formalization and metrics

- How to best test models?

# Organization

Formalization and metrics

- ▶ How to best test models?
  - ▶ Reverse dictionary: MSE, cos, closest match




# Organization

Formalization and metrics

- ▶ How to best test models?
  - ▶ Reverse dictionary: MSE, cos, closest match
  - ▶ Definition Modeling: sense-level BLEU, lemma-level BLEU, MoverScore

# Organization

Formalization and metrics

- ▶ How to best test models?
  - ▶ Reverse dictionary: MSE, cos, closest match
  - ▶ Definition Modeling: sense-level BLEU, lemma-level BLEU, MoverScore
- ▶  **MoverScore is based on embeddings, viz. precisely what we try to assess**

# Organization

Formalization and metrics

- ▶ How to best test models?
  - ▶ Reverse dictionary: MSE, cos, closest match
  - ▶ Definition Modeling: sense-level BLEU, lemma-level BLEU, MoverScore
- ▶ ⚠ MoverScore is based on embeddings, viz. precisely what we try to assess
- ▶ ⚠ We do not measure factual accuracy

# Organization

## Timeline

- ▶ September 3, 2021: Training data & development data made available
- ▶ January 10, 2022: Evaluation data made available & evaluation start
- ▶ January 31, 2022: Evaluation end
- ▶ February 23, 2022: Paper submission due
- ▶ March 31, 2022: Notification to authors

# Organization

Starter pack

- ▶ A codalab instance: `competitions.codalab.org/competitions/34022`

# Organization

Starter pack

- ▶ A codalab instance: `competitions.codalab.org/competitions/34022`
- ▶ A git with a fully running transformer baseline for both tracks:  
`git.atilf.fr/tmickus/codwoe`

# Organization

Starter pack

- ▶ A codalab instance: `competitions.codalab.org/competitions/34022`
- ▶ A git with a fully running transformer baseline for both tracks:  
`git.atilf.fr/tmickus/codwoe`
- ▶ A google group:  
`groups.google.com/g/semEval2022-dictionaries-and-word-embeddings`
- ▶ A discord server: `discord.gg/y8g6qXakNs`

# Organization

Starter pack

- ▶ A codalab instance: `competitions.codalab.org/competitions/34022`
- ▶ A git with a fully running transformer baseline for both tracks:  
`git.atilf.fr/tmickus/codwoe`
- ▶ A google group:  
`groups.google.com/g/semEval2022-dictionaries-and-word-embeddings`
- ▶ A discord server: `discord.gg/y8g6qXakNs`
- ▶ Data for 5 languages, 3 embedding architectures + richly annotated trial data



# Outline



Utrecht University

atilf

ANALYSE ET TRAITEMENT  
INFORMATIQUE  
DE LA LANGUE FRANÇAISE



UNIVERSITÉ  
DE LORRAINE

# Perspectives

- ▶ What do we hope to learn?

# Perspectives

- ▶ What do we hope to learn?

1. Do all architectures yield comparable results?

# Perspectives

## ► What do we hope to learn?

1. Do all architectures yield comparable results?
2. What are the effects of combining different inputs?

# Perspectives

## ► What do we hope to learn?

1. Do all architectures yield comparable results?
2. What are the effects of combining different inputs?
3. Do contextual embeddings help to define polysemous words?

# Perspectives

► What do we hope to learn?

1. Do all architectures yield comparable results?
2. What are the effects of combining different inputs?
3. Do contextual embeddings help to define polysemous words?

**Thanks for your attention! Any questions?**