Assignment

SRL

Teaching assistants:

Stefan Bejgu,

Abelardo Carlos Martinez Lorenzo,

Riccardo Orlando,

Alessandro Scirè,

Simone Tedeschi

SAPIENZA NLP

Semantic Role Labeling (SRL)

A recap

SRL is the task of addressing

"**Who** **did** **What** to **Whom**, How, Where and When?"

The **cat** (Who?) **ate** (Did what?) the **fish** (To whom?)

Semantic Role Labeling (SRL)

A recap: predicates, arguments, semantic roles

The cat ate the fish

Semantic Role Labeling (SRL)

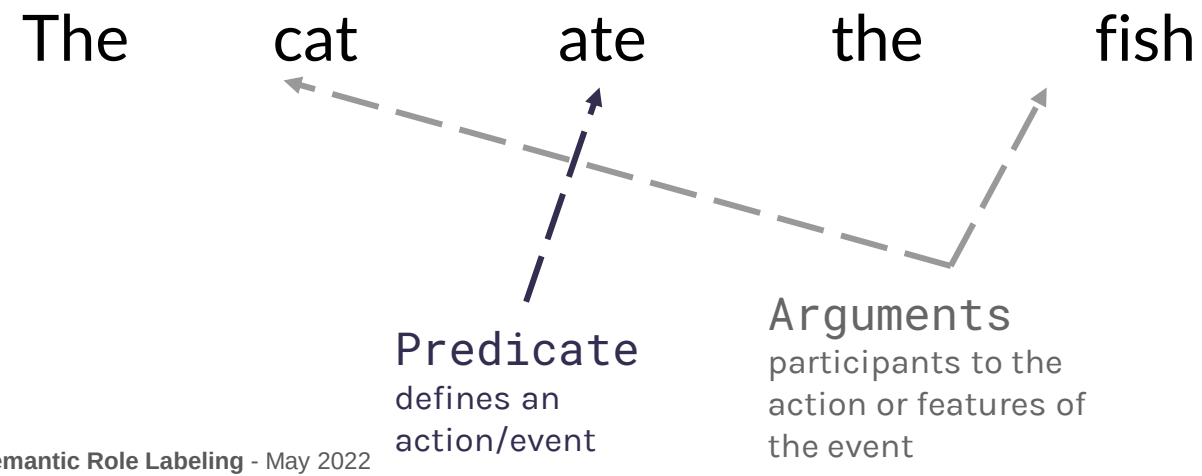
A recap: predicates, arguments, semantic roles

The cat ate the fish

Predicate
defines an action/event

Semantic Role Labeling (SRL)

A recap: predicates, arguments, semantic roles

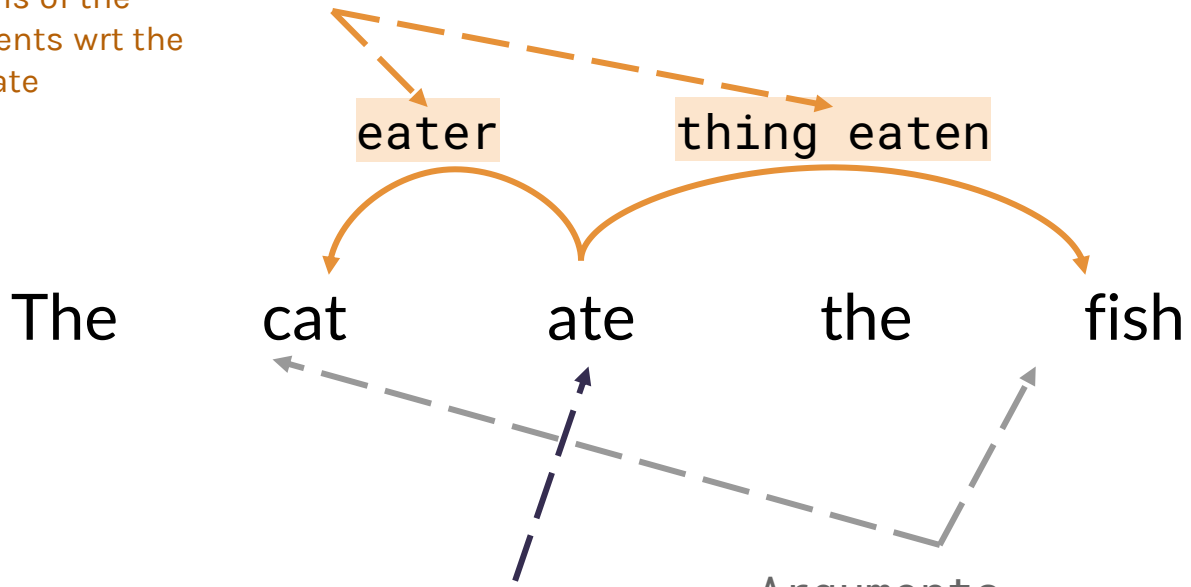


SAPIENZA NLP

Semantic Role Labeling (SRL)

A recap: predicates, arguments, semantic roles

Semantic Roles
relations of the arguments wrt the predicate



Predicate
defines an action/event

Arguments
participants to the action or features of the event

SAPIENZA NLP

Semantic Role Labeling (SRL)

SRL can be divided into 4 subtasks

Identify the predicate(s) in the target sentence

The cat **ate** the fish.

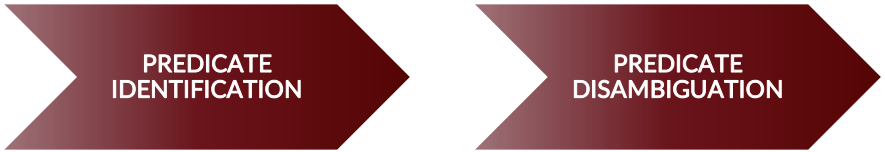
SAPIENZA NLP

Semantic Role Labeling (SRL)

SRL can be divided into 4 subtasks

Identify the predicate(s) in the target sentence

Associate the predicate with its sense



The cat **ate** the fish.

The cat **ate** the fish.

ate → **EAT/BITE**
→ CORRODE
→ CONSUME
→ ...

SAPIENZA NLP

Semantic Role Labeling (SRL)

SRL can be divided into 4 subtasks

Identify the predicate(s) in the target sentence

Associate the predicate with its sense

Identify the arguments of the predicate

The cat **ate** the fish.

The cat **ate** the fish.

ate → **EAT/BITE**
→ CORRODE
→ CONSUME
→ ...

The **cat** ate the **fish**.

SAPIENZA NLP

Semantic Role Labeling (SRL)

SRL can be divided into 4 subtasks

Identify the predicate(s) in the target sentence

PREDICATE IDENTIFICATION

The cat **ate** the fish.

Associate the predicate with its sense

PREDICATE DISAMBIGUATION

The cat **ate** the fish.

ate → **EAT/BITE**
→ CORRODE
→ CONSUME
→ ...

Identify the arguments of the predicate

ARGUMENT IDENTIFICATION

The **cat** ate the **fish**.

Associate each argument with its class

ARGUMENT CLASSIFICATION

The **cat** ate the **fish**.

cat → Agent
fish → Patient

The goal of this assignment

The core goal of this assignment is to develop a system that performs **argument identification** and **argument classification** (steps 3 and 4 in the previous slide).

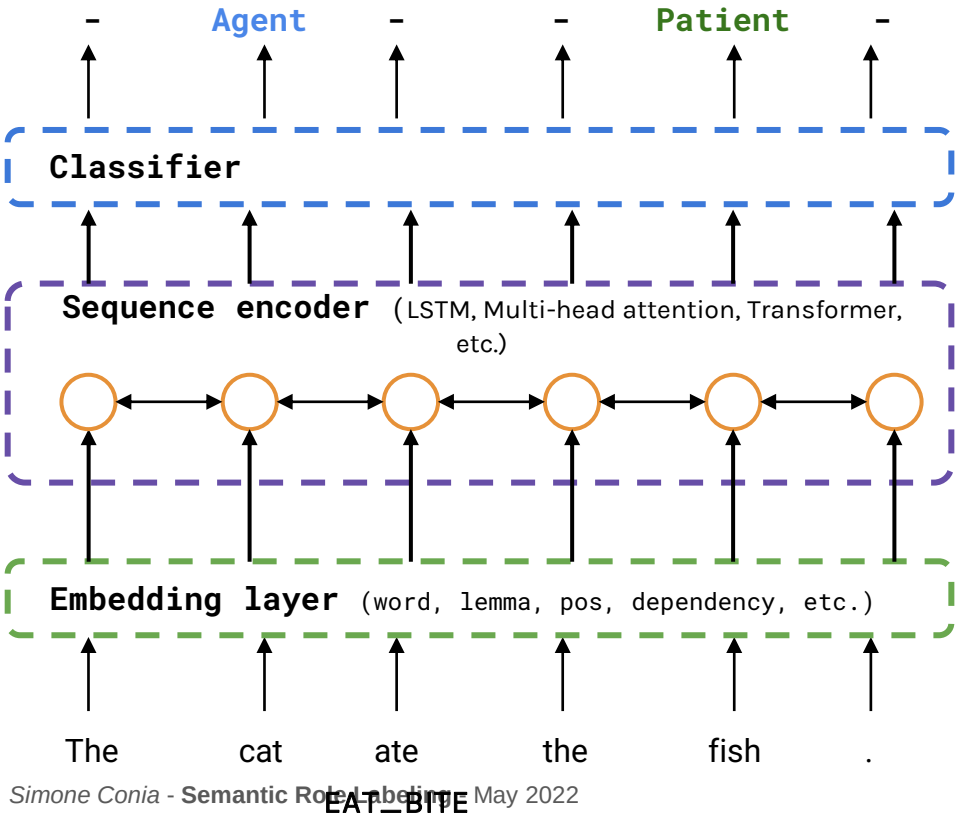
INPUT	The	cat	ate (EAT_BITE)	the	fish	.
OUTPUT	_	**Agent**	_	_	**Patient**	_

The predicate senses and semantic roles are defined according to a semantic resource. In this homework, we will use VerbAtlas[1], available at https://verbatlas.org

[1]: Andrea Di Fabio, Simone Conia, and Roberto Navigli. "VerbAtlas: a Novel Large-Scale Verbal Semantic Resource and Its Application to Semantic Role Labeling." *Proceedings of EMNLP-IJCNLP*. 2019.

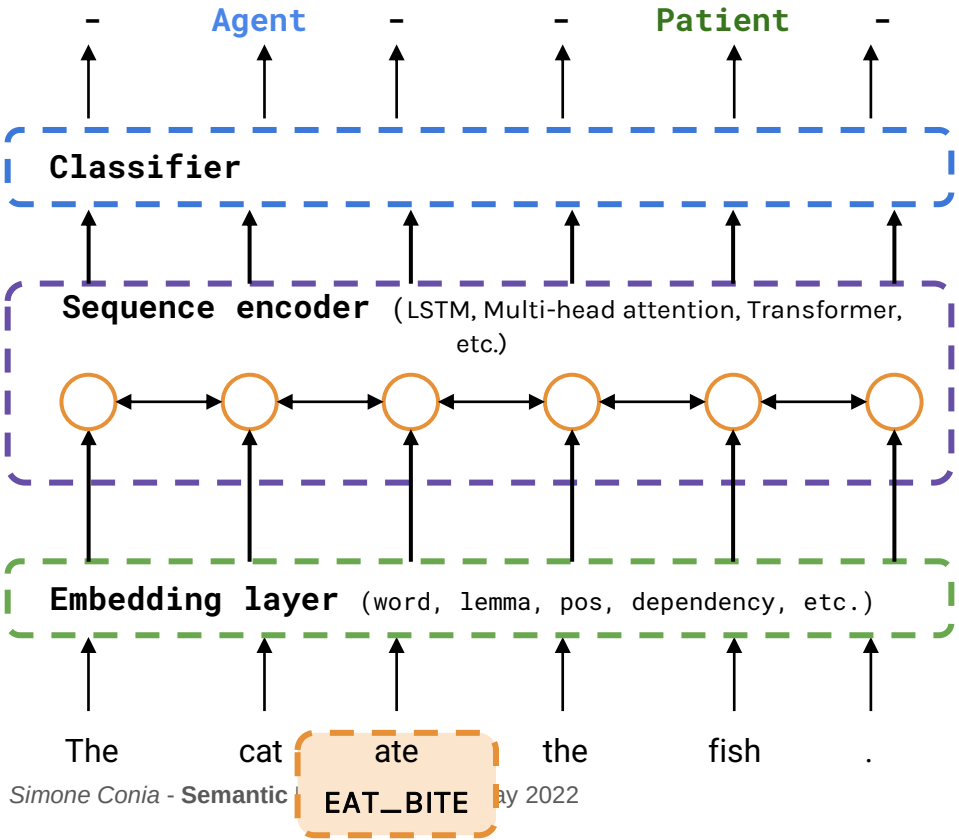
Model: possible approaches

SRL can be seen as a sequence labeling task, so you can start from your NER architecture 👀!



Model: possible approaches

SRL can be seen as a sequence labeling task, so you can start from your NER architecture 👀!

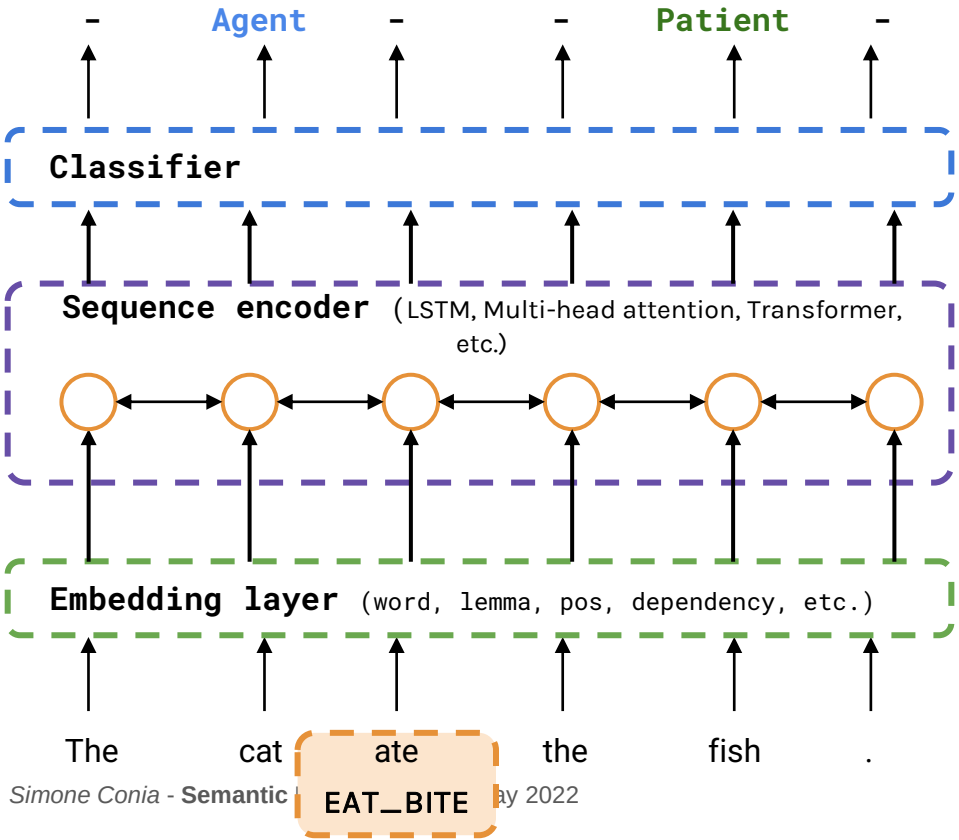


KEY DIFFERENCE:
You should indicate which and where the predicate is in the sentence.

There are multiple ways to do this: from a simple 0/1 flag to a trainable flag embedding. Be creative!

Model: possible approaches

SRL can be seen as a sequence labeling task, so you can start from your NER architecture 👀!



KEY DIFFERENCE:
You should indicate which and where the predicate is in the sentence.

There are multiple ways to do this: from a simple 0/1 flag to a trainable flag embedding. Be creative!

QUESTION:
What if there is more than one predicate in a sentence?

SAPIENZA NLP

What if there is more than one predicate in a sentence?

Simple solution: replicate the sentence for each predicate-argument pair

The Committee insists that all perpetrators of violations of the Convention be brought to justice.

What if there is more than one predicate in a sentence?

Simple solution: replicate the sentence for each predicate-argument pair

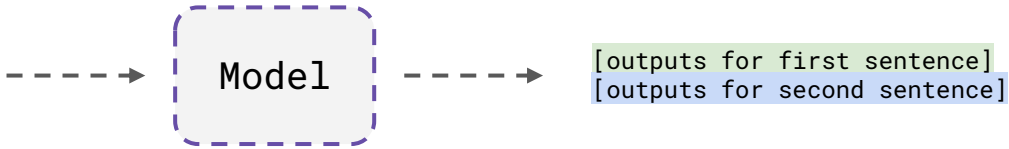
The Committee insists that all perpetrators of violations of the Convention be brought to justice.

CAUSE-MENTAL-STATE

BRING

```
The Committee insists that all
perpetrators of violations of the
Convention be brought to justice.

The Committee insists that all
perpetrators of violations of the
Convention be brought to justice.
```



SAPIENZA NLP

Model performance evaluation

Argument identification

The performance of a SRL system is usually measured in terms of F1 score, excluding the NULL ("_") labels.

INPUT	The	cat	ate (EAT_BITE)	the	fish	.
GOLD	_	Agent	_	_	Patient	_
PRED	_	ARG	_	ARG	ARG	_

ARGUMENT IDENTIFICATION

Precision = (CORRECTLY identified arguments) / (PREDICTED arguments) = 2/3 = 0.66

Recall = (CORRECTLY identified arguments) / (GOLD arguments) = 2/2 = 1.0

F1 = 2 * (P * R) / (P + R) ~= 0.80

Model performance evaluation

Argument classification

The performance of a SRL system is usually measured in terms of F1 score, excluding the NULL ("_") labels.

INPUT	The	cat	ate (EAT_BITE)	the	fish	.
GOLD	_	Agent	_	_	Patient	_
PRED	_	Agent	_	Patient	Theme	_

ARGUMENT CLASSIFICATION

Precision = (CORRECTLY classified arguments) / (PREDICTED arguments) = 1/3 = 0.33

Recall = (CORRECTLY classified arguments) / (GOLD arguments) = 1/2 = 0.5

F1 = 2 * (P * R) / (P + R) ~= 0.40

What you will receive

- We will provide you with a folder organized as follows (some files are omitted):

```
- nlp2022-hw2/
  - data/
  - hw2/
    - model.py
    - stud/
  - model/
  - requirements.txt
  - test.sh
```

- You are allowed to edit **only** the items in bold!

What you will receive

- We will evaluate your work using Docker
  - You should be fine even if you don't know anything about it
- If **`test.sh`** runs on your side, it will run on ours as well
  - Just keep in mind: do not change any file but those we marked in bold as editable in the previous slide
- Additionally, we wrote a **`README.md`** to get you everything up and running
- You can find the code repository here!

What we expect from you

- The zip folder we gave you (but populated :))
- Put your training code (if you used Colab, download the notebook `.ipynb` and place it) in **`hw2/stud/`**
- If you use any additional library, modify the **`requirements.txt`** file as needed (click here for info)
- Use the data (train and dev) in the data folder
  - use each file as defined in the **standard ML conventions** (*train for training*, *dev for model selection*)

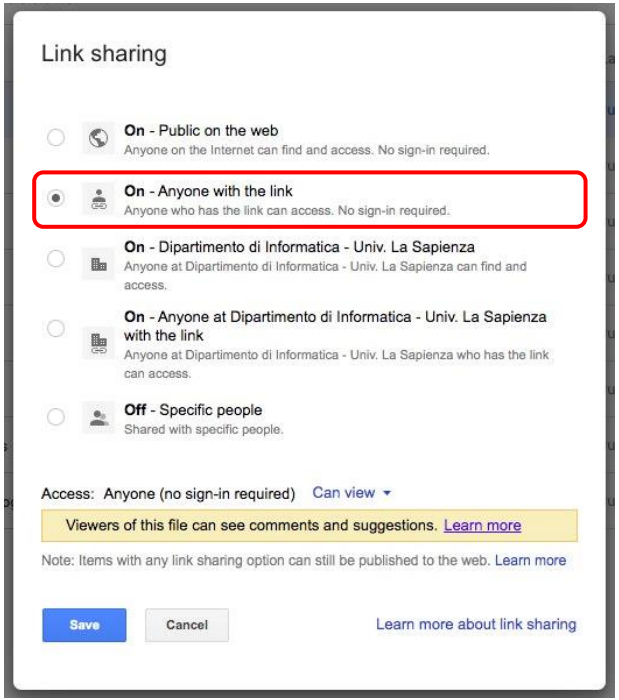
What we expect from you

- Put **everything** your model needs (vocabulary, weights, ...) inside the **`model/`** folder, and **be sure to properly load them** in your model
- In **`hw2/stud/implementation.py`** implement the **`StudentModel`** class
  - Load your model and use it in the **`predict`** method
  - You **must respect** the signature of the predict method!
  - You can add other methods (i.e. the constructor)
- In **`hw2/stud/implementation.py`** implement the **`build_model`** function
  - It should initialize your **`StudentModel`** class.

What we expect from you

- Use **`test.sh`** to check that everything works
- Add your **`report.pdf`** to the folder (yes, export it in PDF even if you are using Word!)
- Name the zip folder **`lastname_studentid_hw2.zip`**:
  - Ex: Luigi D'Andrea will submit a file named **`dandrea_1234567_hw2.zip`**
  - If you are unsure which name to put, use the one in your institutional email account

Submission Instructions



- Upload the zip on your **institutional** Drive and make it **link-shareable** and **public** to anyone (an automatic script will download it).
- Make sure it is accessible via an incognito page of your browser!
- Do **NOT modify** the folder structure
- You have to submit the homework through the submission form on Google Classroom. You will be asked to fill a form with the requested information and the **link** to the zip you uploaded on Drive.

SAPIENZA NLP

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

WARNING!

The dataset shall be used **EXCLUSIVELY** for this homework:

- **DO NOT REDISTRIBUTE** this dataset in any way!
- **DO NOT UPLOAD** this dataset online (GitHub, GitLab, your website, whatever).
- **DO NOT SHARE** the dataset with people outside of the NLP course.

You **MUST** conform to these rules **even after** your homework submission! If you wish to share your code online (e.g., GitHub), please do after the beginning of next year's NLP course and without the dataset!

Any breach - even those of accidental nature - of the above-mentioned rules will result in an immediate **AUTOMATIC FAIL**. Please, be careful!

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

- UniteD-SRL[1]
- 3 languages: **English***, French and Spanish
- **English** is **mandatory** while French and Spanish are considered as extras
- Parallel sentences

	Train	Dev	Test
English	5501	1026	1027
French	464	1026	1027
Spanish	464	1026	1027

- Number of sentences for each split.
- Dev and Test are parallel for each language.
- FR and ES train sets only contain a subset of EN.

[1]: Rocco Tripodi, Simone Conia and Roberto Navigli: "UniteD-SRL: A unified dataset for span- and dependency-based multilingual and cross-lingual Semantic Role Labeling." *Proceedings of EMNLP* 2021.

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

The dataset is a JSON file where each entry contains the following fields:

```
sentence_id: {
  “words”: [“The”, “cat”, “ate”, “the”, “fish”, “and”, “drank”, “the”, “milk”, “.”],
  “lemmas”: [“the”, “cat”, “eat”, “the”, “fish”, “and”, “drink”, “the”, “milk”, “.”],
  “pos_tags”: [“DET”, ..., “PUNCT”],
  “dependency_relations”: [“NMOD”, ..., “ROOT”, ..., “P”],
  “dependency_heads”: [1, 2, 0, ...],
  “predicates”: [“_”, “_”, “EAT_BITE”, “_”, “_”, “_”, “DRINK”, “_”, “_”, “_”],
  “roles”: {“2”: [“_”, “Agent”, “_”, “_”, “Patient”, “_”, “_”, “_”, “_”, “_”],
                “6”: [“_”, “Agent”, “_”, “_”, “_”, “_”, “_”, “_”, “Patient”, “_”]}
}
```

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

The dataset is a JSON file where each entry contains the following fields:

```
sentence_id: {
  "words": ["The", "cat", "ate", "the", "fish", "and", "drank", "the", "milk", "."],
  "lemmas": ["the", "cat", "eat", "the", "fish", "and", "drink", "the", "milk", "."],
  "pos_tags": ["DET", ..., "PUNCT"],
  "dependency_relations": ["NMOD", ..., "ROOT", ..., "P"],
  "dependency_heads": [1, 2, 0, ...],
  "predicates": ["_", "_", "EAT_BITE", "_", "_", "_", "DRINK", "_", "_", "_"],
  "roles": {"2": ["_", "Agent", "_", "_", "Patient", "_", "_", "_", "_", "_"],
            "6": ["_", "Agent", "_", "_", "_", "_", "_", "_", "Patient", "_"]}
}
```

Syntactic Information

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

The dataset is a JSON file where each entry contains the following fields:

```
sentence_id: {
  “words”: [“The”, “cat”, “ate”, “the”, “fish”, “and”, “drank”, “the”, “milk”, “.”],
  “lemmas”: [“the”, “cat”, “eat”, “the”, “fish”, “and”, “drink”, “the”, “milk”, “.”],
  “pos_tags”: [“DET”, ..., “PUNCT”],
  “dependency_relations”: [“NMOD”, ..., “ROOT”, ..., “P”],
  “dependency_heads”: [1, 2, 0, ...],
  “predicates”: [“_”, “_”, “EAT_BITE”, “_”, “_”, “_”, “DRINK”, “_”, “_”, “_”],
  “roles”: {“2”: [“_”, “Agent”, “_”, “_”, “Patient”, “_”, “_”, “_”, “_”, “_”],
                “6”: [“_”, “Agent”, “_”, “_”, “_”, “_”, “_”, “_”, “Patient”, “_”]}
}
```

VerbAtlas frame of a predicate. If the token is not a predicate, then __.

Dataset

UniteD-SRL (Tripodi et al., EMNLP 2021)

The dataset is a JSON file where each entry contains the following fields:

```
sentence_id: {
  "words": ["The", "cat", "ate", "the", "fish", "and", "drank", "the", "milk", "."],
  "lemmas": ["the", "cat", "eat", "the", "fish", "and", "drink", "the", "milk", "."],
  "pos_tags": ["DET", ..., "PUNCT"],
  "dependency_relations": ["NMOD", ..., "ROOT", ..., "P"],
  "dependency_heads": [1, 2, 0, ...],
  "predicates": ["_", "_", "EAT_BITE", "_", "_", "_", "DRINK", "_", "_", "_"],
  "roles": {"2": ["_", "Agent", "_", "_", "Patient", "_", "_", "_", "_", "_"],
            "6": ["_", "Agent", "_", "_", "_", "_", "_", "_", "Patient", "_"]}
}
```

Roles of predicate `i`. E.g. 2 lists the roles for predicate at index 2 in `predicates`, i.e., `EAT_BITE`.

Extras

Get up to 6 extra points!

You can get up to 6 extra points by doing extra work. What can you do?

Extras

Get up to 6 extra points!

You can get up to 6 extra points by doing extra work. What can you do?

Whatever you want!

Extras

Get up to 6 extra points!

You can get up to 6 extra points by doing extra work. What can you do?

Whatever you want!

However...

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Cross-lingual performance: can you improve your results in ES and FR?

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Cross-lingual performance: can you improve your results in ES and FR?

- **Deep transfer learning:** train your model in English and transfer what it learns to Spanish and/or French.

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Cross-lingual performance: can you improve your results in ES and FR?

- **Deep transfer learning:** train your model in English and transfer what it learns to Spanish and/or French.
- **Project the annotations:** develop a system that "projects" the labels from an English sentence to a parallel Spanish and/or French sentence.

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Cross-lingual performance: can you improve your results in ES and FR?

- **Deep transfer learning:** train your model in English and transfer what it learns to Spanish and/or French.
- **Project the annotations:** develop a system that "projects" the labels from an English sentence to a parallel Spanish and/or French sentence.
- **Silver training corpora:** create high-quality automatic annotations on top of an unannotated corpus (e.g., part of Wikipedia). You can start from InVeRo-XL!

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Explicit Semantics: can you improve your results by exploiting BabelNet and/or VerbAtlas?

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Explicit Semantics: can you improve your results by exploiting BabelNet and/or VerbAtlas?

- **Take advantage of WSD:** try to exploit the knowledge provided by synsets to improve your results. You can annotate a sentence using AMuSE-WSD.

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

Explicit Semantics: can you improve your results by exploiting BabelNet and/or VerbAtlas?

- **Take advantage of WSD:** try to exploit the knowledge provided by synsets to improve your results. You can annotate a sentence using AMuSE-WSD.

- **Combine BabelNet and VerbAtlas:** VerbAtlas's frames are clusters of BabelNet's synsets. Can you find a way to combine the complementary knowledge provided by frames and synsets?

Extras

Get up to 6 extra points!

We would be extremely interested in (= award with more points):

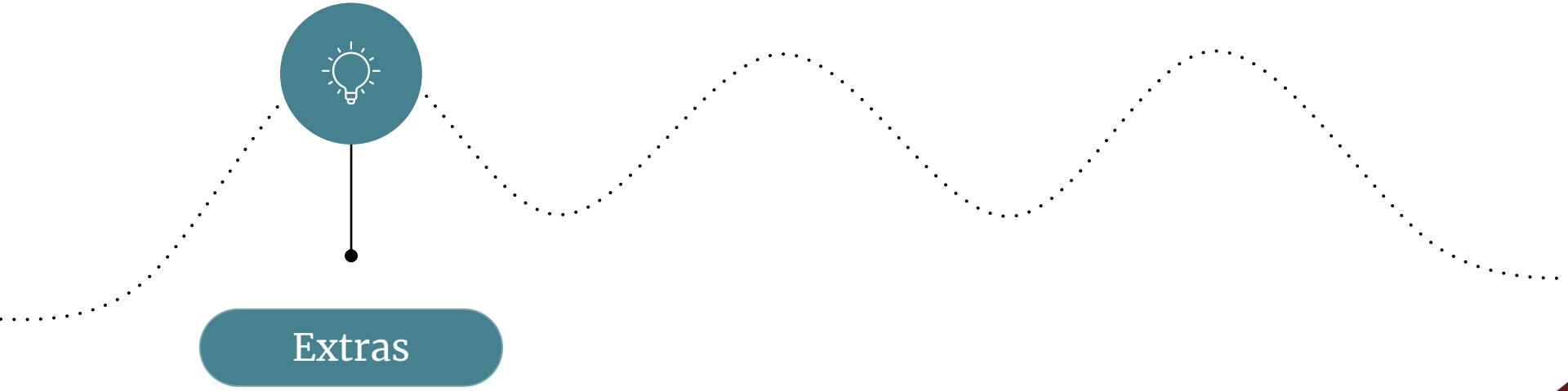
Explicit Semantics: can you improve your results by exploiting BabelNet and/or VerbAtlas?

- **Take advantage of WSD:** try to exploit the knowledge provided by synsets to improve your results. You can annotate a sentence using AMuSE-WSD.
- **Combine BabelNet and VerbAtlas:** VerbAtlas's frames are clusters of BabelNet's synsets. Can you find a way to combine the complementary knowledge provided by frames and synsets?
- **Exploit VerbAtlas:** can you find clever ways of encoding the predicate-argument structures of VerbAtlas within your model? VerbAtlas provides a lot of information that is still under-explored.

Extras

Get up to 6 extra points!

Why are we interested in your extras?



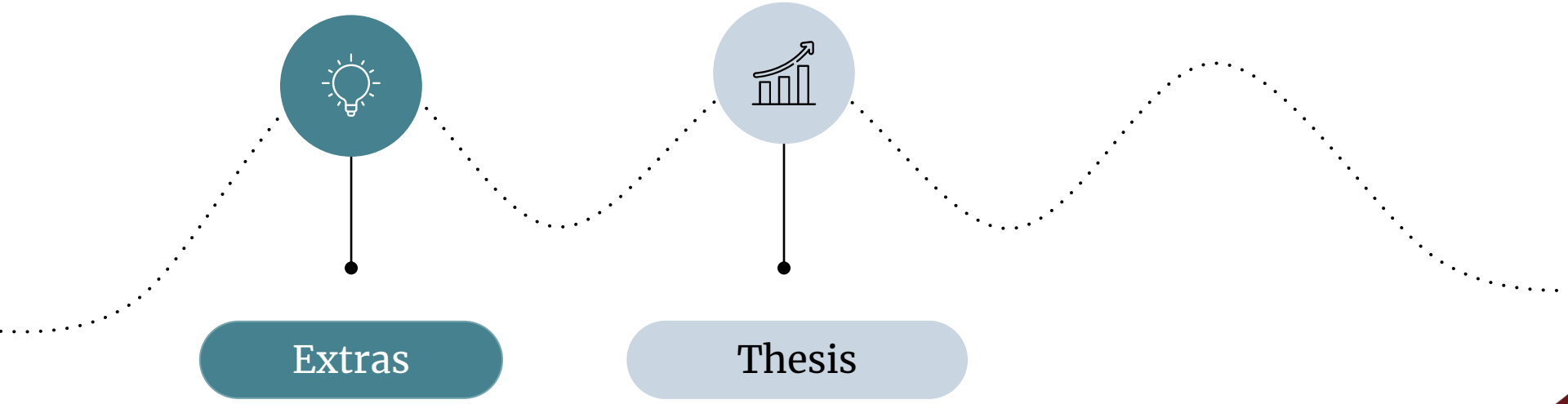
The work you do for the extra can be an **excellent starting point for a thesis!**

SAPIENZA NLP

Extras

Get up to 6 extra points!

Why are we interested in your extras?



The work you do for the extra can be an **excellent starting point for a thesis!**

You can better study the problem and your approach during a thesis.

SAPIENZA NLP

Extras

Get up to 6 extra points!

Why are we interested in your extras?

Extras

The work you do for the extra can be an **excellent starting point for a thesis!**

Thesis

You can better study the problem and your approach during a thesis.

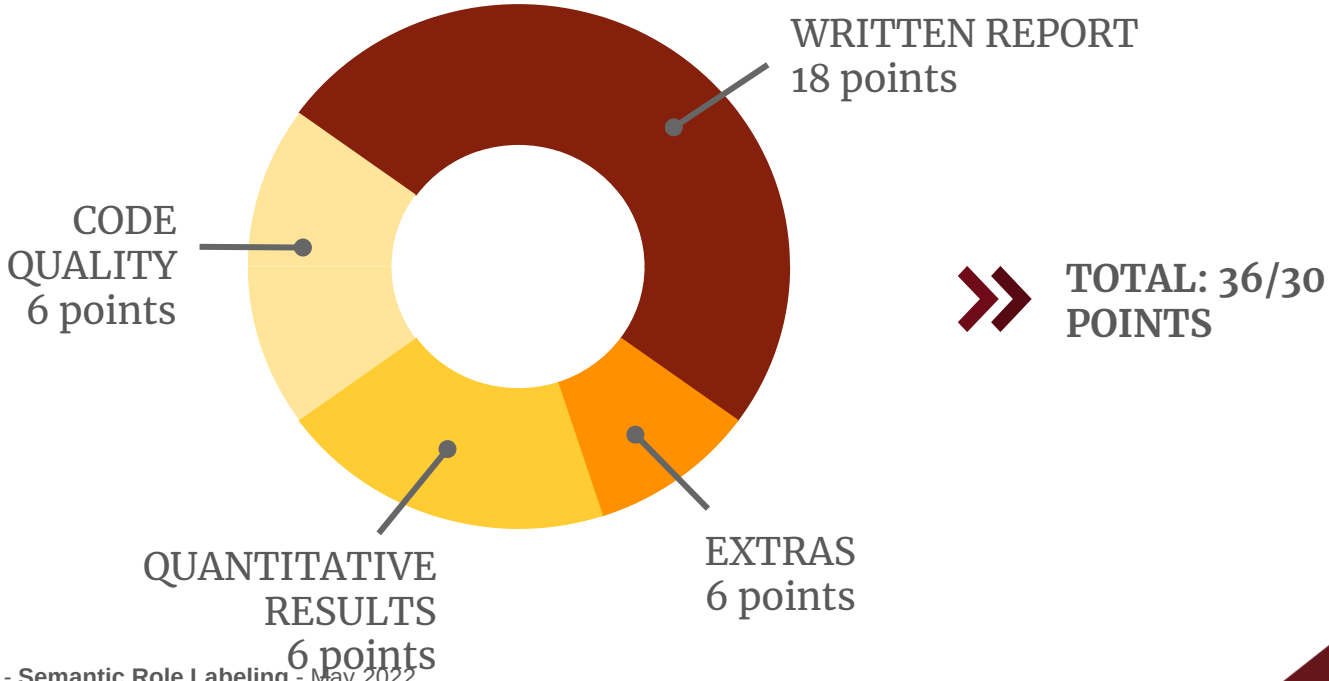
To the moon?

Your work could potentially be published to a **top-tier AI conference!**

SAPIENZA NLP

Evaluation

We will take into account the following criteria:



TOTAL: 36/30 POINTS

SAPIENZA NLP

Report: dos and don’ts

- **ACL 2021 paper template**
  - Available here (Word and LaTeX direct download) or here (Overleaf LaTeX template)
  - You can use either the LaTeX or the Word template, your choice
  - **DO NOT MODIFY** the template (margins, spacing, font size)
  - Use the non-anonymous flag, so you can enter your name
- **Max 3 pages**
  - For the report, including title, subtitles, etc.
  - This is a **STRICT RULE!**
- **Unlimited extra pages for images, tables and references**
  - Be sure to **include** and properly **comment** a confusion matrix, visualized as heat map
  - Every image and table must have a caption (don’t abuse them please :))
  - Tables and images must be referenced in the report

Report: what you are expected to do

We expect a good report to be:

- **Readable** and **understandable**
  - We will not give penalties for English errors, but we expect the report to follow a clear flow. We don't want to read just a sequence of statements on what you did without showing the reasoning behind your choices
- **Well-structured** and **organized**
  - Take inspiration from the many papers available online and organize your report in well-defined sections (e.g. method, setup, experiments, results...)

Report: what you are not expected to do

We expect a good report **NOT** to include:

- Unnecessary **task** or **dataset descriptions**
  - just focus on your solution to the problem.
- **Code** copy-paste
  - Your code should be self-explanatory, so no need to show it in the report. You can add **pseudocode** to show some particular algorithm, but **no code or screenshots,** please!

Report: what you are not expected to do

We expect a good report **NOT** to include:

- Unnecessary **low-level implementation details**
  - Avoid any **low-level implementation/technical details** like “I used a dictionary to store these values”, “I had to use configuration X to solve this exception”, “I could not use Y because there was a dependency issue with Z”, etc.
  - Instead, **we are interested in high-level abstractions/strategies** you decide to use to tackle the homework, as well as the **intuitions behind your choices**.

    E.g. use and description of a particular model, explanation of how and why an architecture works, etc.

Application: what you are expected to do

Your project should conform to the following rules:

- You **MUST** use PyTorch.
  - TensorFlow and other deep learning frameworks are **NOT** allowed.
  - PyTorch Lightning **is allowed**
- **Frameworks** that use PyTorch (e.g. AllenNLP, torchtext…) are **NOT** allowed.
- Libraries (such as tqdm, sklearn, NLTK) are fine, but since the line between a framework and a library is sometimes blurred, please ask in the Google Classroom group before using any external library: **any other** library **MUST be agreed with the TAs**.

Application: what you are not expected to do

Your project should conform to the following rules:

- **You are now allowed** to use any architectures that have been explained in the course, in particular:
  - word embeddings (Word2Vec, GloVe, etc.) **are allowed**,
  - contextualized word embeddings (ELMo, etc.) **are allowed**,
  - Transformer-based models (BERT, BART, RoBERTa, etc.) **are allowed**.
- For any doubt, please ask the TAs on Google Classroom.
- **Comment** your code, please!

Quantitative Results

We will evaluate the **performance of your model** on a SECRET test set.

Also in this homework, your model must pass a **baseline score** for each (optional and mandatory) subtask for us to consider it:

- (Optional) Predicate identification: F1 > 70%
- (Optional) Predicate disambiguation: F1 > 60%
- (Mandatory) Argument identification: F1 > 25%
- (Mandatory) Argument classification: F1 > 25%

The homeworks that surpass the baseline will be graded according to different thresholds, that will be defined based on an internal reference model and the normalized distribution of YOUR scores.

Extras

You can achieve **up to 6 points with some extra work!**

See Homework Extras section for some suggestions about what we consider an extra.

Don’t forget to **explain your choices** in the report! Extras that are not explained in the report will not be considered for evaluation.

Please note that you are not allowed to use tools that have not been explained **yet** in the course. For any doubt, please contact the TAs.

Evaluation

- `test.sh` is identical to what we will be using
- **If it does not run on your side, we will not correct your homework**
- Note that, if you use **any kind of hard-coded paths**, this script **won't work**
- Use **paths relative to the project root folder**, e.g.:
  - **NO:** `/home/pincopallino/my_folder/model/weights.pt`
  - **OK:** `model/weights.pt`

Please be aware that

This is an **individual exercise!** Collaboration among the students is **not** allowed.

We will check for **plagiarism** both manually and automatically.

It is **not allowed** to:

- Copy from other students
- Share your code with other students
- Copy from online resources (StackOverflow, GitHub, Medium, Kaggle and so on).

While we release the homework on GitHub, **DO NOT FORK THE PROJECT.**

However, you are allowed to use material from **external sources** as long as it is **not central** to the homework.

You are also allowed to use the **SOME** parts of the presented class notebooks. However, you **MUST** explicitly specify these parts in your code comments.

Use of external data

- For your experiments, **use the provided data** (train and dev) in the data folder; use each file as defined in the standard ML conventions (train for training, dev for model selection).
- If you train it using external data, include it in the data folder.
- If you train it on dev set, it will be a **FAIL**.

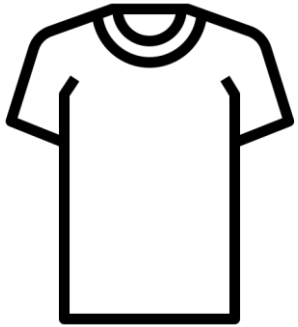
A few tips to organize your work:

- Test the testing infrastructure as soon as possible so if you have any issue you can contact us and try to solve the issue
- Start **as soon as possible**!
  - Training a neural network requires time, possibly hours, depending on your hardware
- Start **small**!
  - If you don't get decent results with a very simple neural network, there is a good chance that adding other things won't make your model perform better
- Leave some time for **hyperparameter** tuning!
  - Sometimes good hyperparameter combinations can do wonders for your neural network
- Use **Google** Colab (free GPUs!)

Win a Sapienza NLP t-shirt!

We will hand out amazing Sapienza NLP t-shirts to the **overall top-5** students!

The final ranking will be computed according to the scores on our **secret** test set.



Deadlines

The students **who passed the first homework** may deliver the second one in one of the four available deadlines (2022):

1. Early submission: June 23rd (23:59 CEST) → only this date allows **late submission!**
   Late submission: June 26th (23:59 CEST)
   Presentation: 30th June

2. Submission: July 14th (23:59 CEST)
   Presentation: July 21st

3. Submission: July 19th (23:59 CEST)
   Presentation: July 26th

4. Submission: September 12th(23:59 CEST)
   Presentation: September 19th

Late submission policy

- For the first deadline **ONLY**, we will allow **late submissions** until the 26th of June.
- Also in this case, we will apply 1 point of penalty for each day of delay.
- Example: A student delivers their homework on June 26rd (17:35) → max possible grade 36 - 3 = 33.

Final Exam

Presentation

To complete the exam, you will have an **oral presentation** of all the homeworks.

Presentation will be done in classroom/Google meet **with slides.**

Homework presentation:

- 12 minutes presentation. This is a **very strict requirement.**
  A split of 5 minutes for the first homework and 7 for the second should work.
- 10 minutes for questions

Questions?

If you have a question that may interest your colleagues, **please ask it on Google Classroom.**

Otherwise, for personal or other questions, email **ALL** of us (but please, only reach for things that can't be asked on the Google Classroom).

Our emails are:

{bejgu, martinez, orlando, scire, tedeschi}@diag.uniroma1.it

Good Luck!!