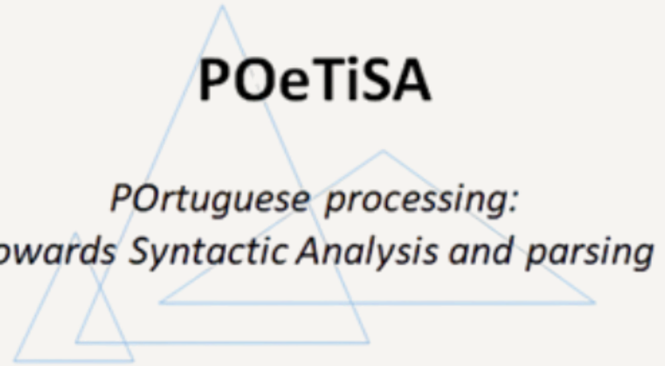
The revision of linguistic annotation in the Universal Dependencies framework:

a look at the annotators' behavior

Magali Duran Lucelene Lopes Thiago Pardo

Universidade de São Paulo - Brazil

The 19th Linguistic Annotation Workshop - Vienna, Austria, July 2025



POeTiSA: POrtuguese processing – Towards Syntactic Analysis and parsing

https://sites.google.com/icmc.usp.br/poetisa

THE **CHALLENGE:**

To produce a gold standard multigenre corpus annotated under Universal Dependencies framework aiming to train state-of-the-art models.

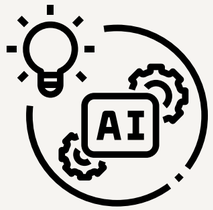
OUR SCENERY:

- No expertise
- No UD guidelines for Portuguese
- UD Portuguese parsers available, reporting 90% precision

Project Decisions

How to address this challenge?

Tackle **one genre at a time,** so that the next corpora would benefit from this first experience.



Pre-annotate using an available parser and **revise the entire corpus**.

Divide the revision task into **4 phases**

Write **guidelines for each phase**, enhancing them during the process.

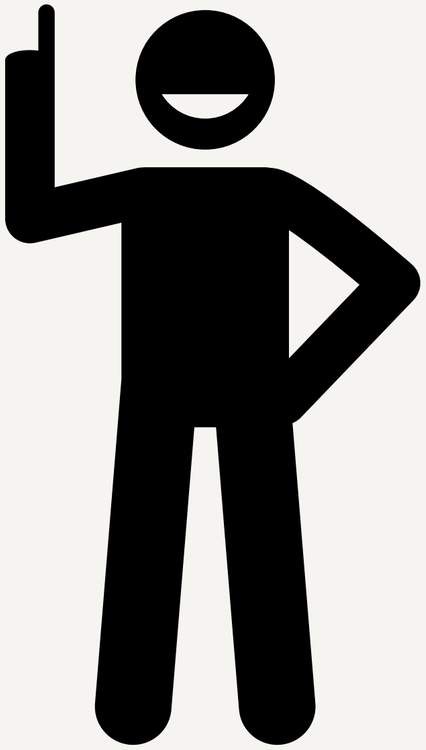
What means to annotate under UD framework?

It means to annotate in **CoNLL-U** format, using **UD tagsets** and employing general **UD guidelines**

CoNLL-U is a file containing **10 columns** and **one line for each token** in the sentence.

So, **each sentence** has its own CoNLL-U file.

What we first realized about CoNLL-U



Only 5 of the 10 columns require revision

There is an **ideal sequence** to perform the revision.

Dividing the revision task of CONLL-U into 4 steps:

1	2	3	4	5	6	7	8	9	10
ID	**FORM**	**LEMMA**	**UPOS**	**XPOS**	**FEATS**	**HEAD**	**DEPREL**	**DEPS**	**MISC**
Token identifier (numeric)	Token form (word or symbol)	Lemma of the token form	PoS tag in the UD tag set	Optional extended (language-specific) PoS tag	List of morphological features associated to the token	ID of the token's head for the dependency tree	Dependency relation tag of the token towards the token's head	HEAD-DEPREL pairs for the enhanced dependency graph	Any additional annotation

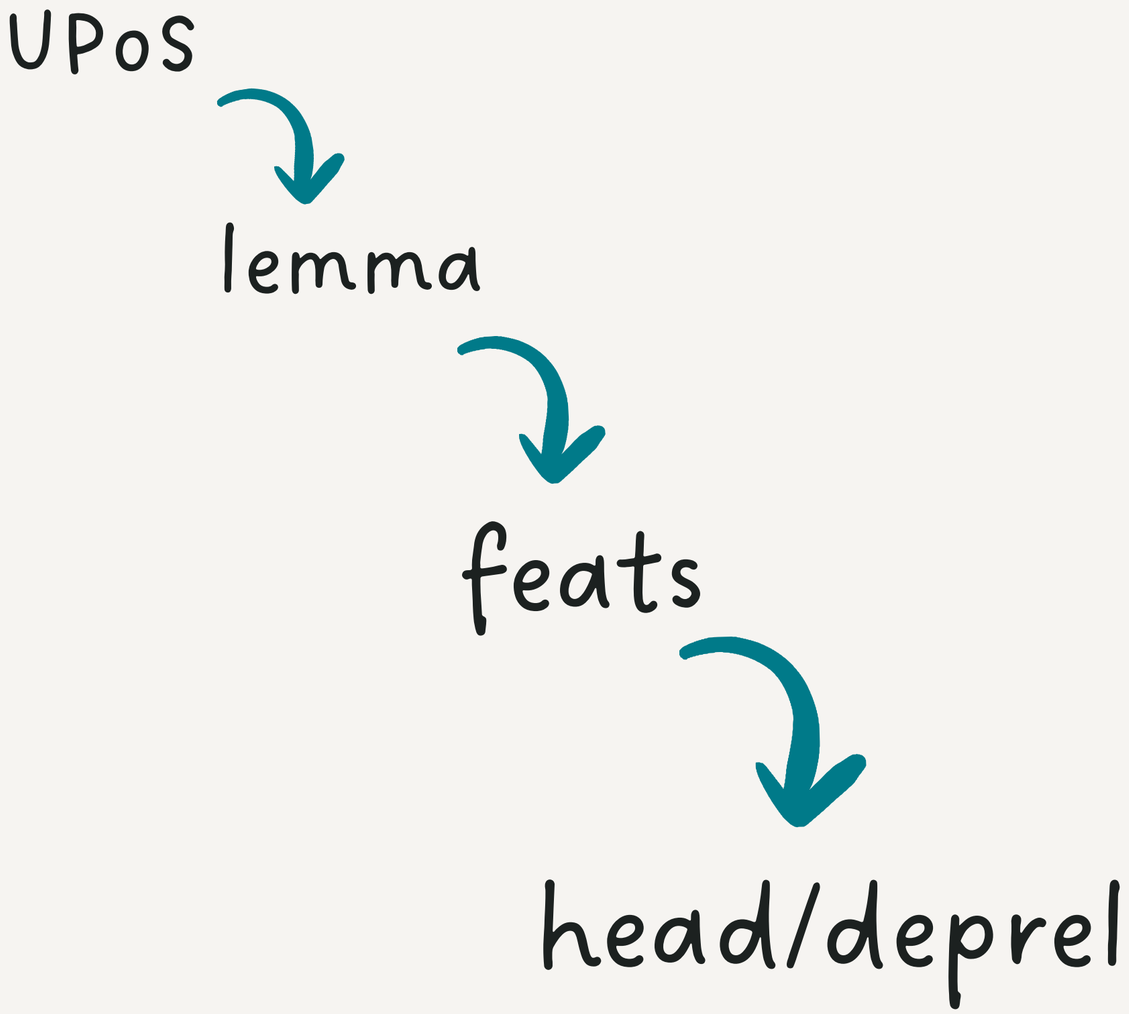
lexicon

human revision

lexicon

human revision

Cascade effect



U-PoS tagset contains 17 tags, most of them corresponding to well-known morphosyntatic functions. We felt that this was the least complex column in which to start training the annotators.

Given the right U-PoS, the vast majority of lemmas could be automatically revised using a lexicon. 1% required human decision.

Given the right lemma, most of the features could be automatically revised using our Portilexicon (a lexicon customized with upos and UD features). Less than 5% required human decision.

Deprel tagset contains 37 tags. The revision of dependency relations (head and deprel columns) is the most complex task and was performed last.

Results:

(in number of tokens)

	CONLL-U COLUMN	HUMAN		AUTOMATIC		TOKENS CHANGED	
STEP 1	UPOS	168,080	100%	-	-	6,440	3.83%
STEP 2	LEMMA	1,825	1.09%	166,255	98.91%	3,649	2.17%
STEP 3	FEATS	8,050	4.79%	160,030	95.21%	19,274	17.42%
STEP 4	HEAD	168,080	100%	-	-	15,358	9.14%
	DEPREL	168,080	100%	-	-	13,816	8.22%

What we have learned about annotators behaviour

What we have learned about annotators behaviour

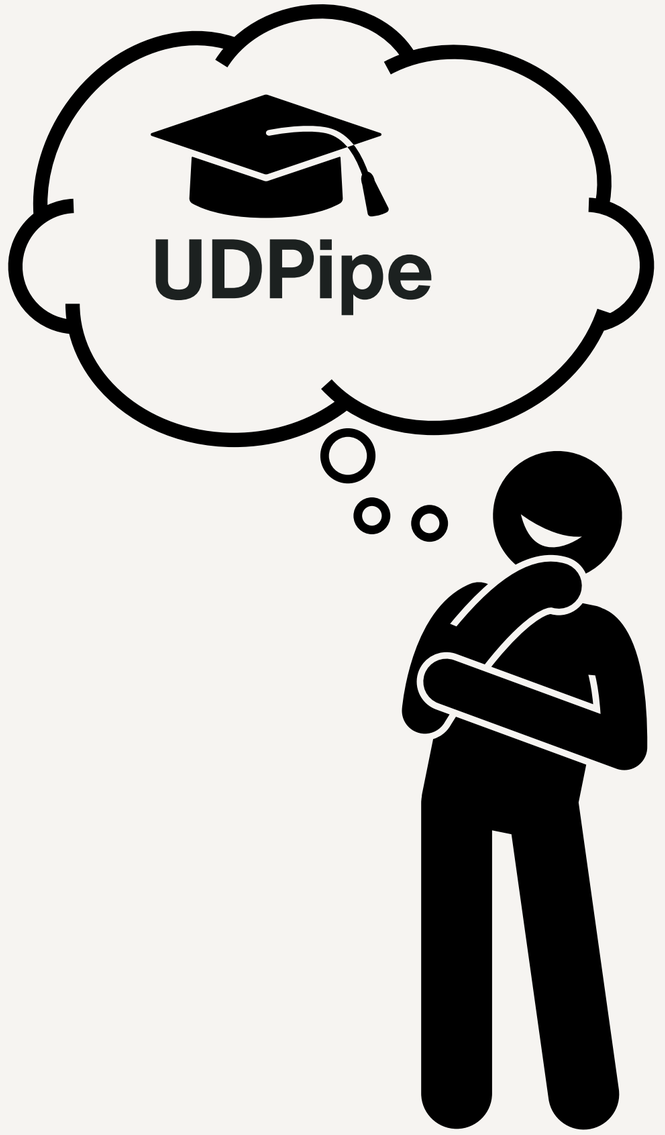
When annotators realize that the **parser makes few mistakes**, they begin to “**trust**” the parser and start to question the annotation less, missing the errors.

UDPipe

What we have learned about annotators behaviour

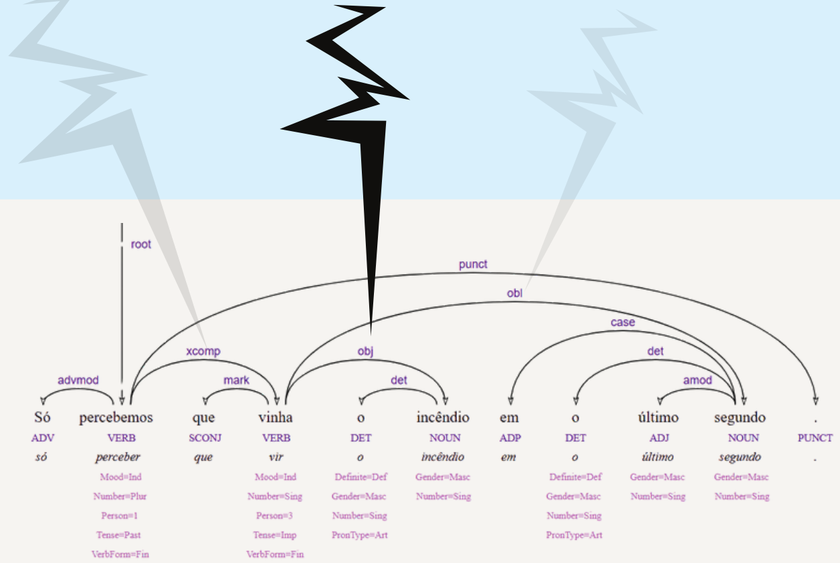
Annotators believe that, if the parser gets difficult things right, **it will not get easy things wrong**.

Therefore, things that are considered “**easy**” are taken **out of the focus** of the revision and “**silly**” mistakes are **no longer corrected**



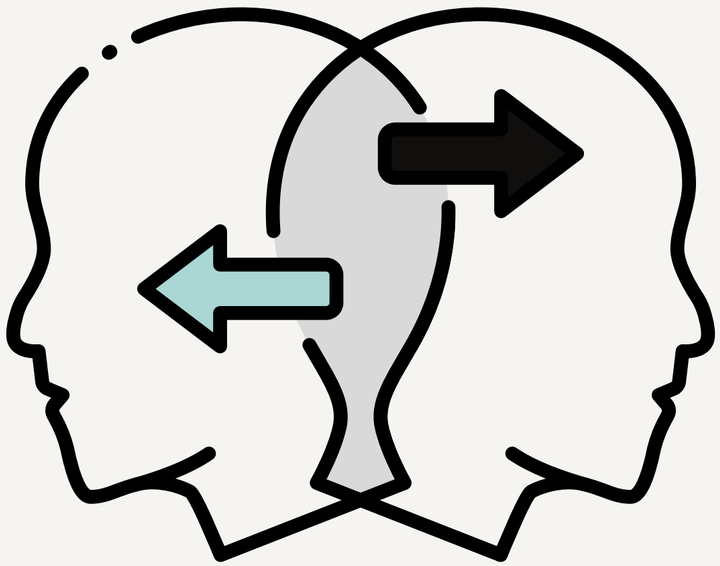
What we have learned about annotators behaviour

Annotators believe that the
"**lightning does not strike the same tree twice**"
and, when they find an error in a sentence, they sometimes are blind to other errors in the same sentence.



What we have learned about annotators behaviour

Annotators often do **not recognize patterns** in less frequent constructions, separated by a **long time interval** (3 days or more)

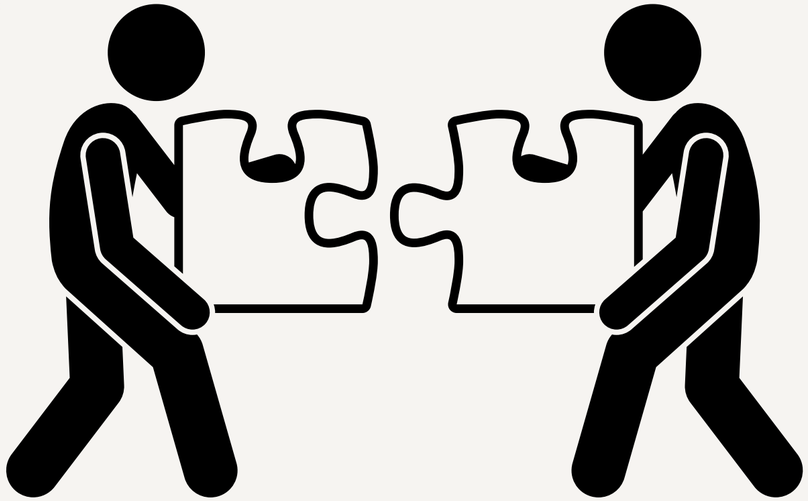


This leads them to annotate similar constructions in different ways. We call this **intra-annotator disagreement** (probably a problem of **memory decay**).

What we have learned about annotators behaviour

Annotators miss most **frequently errors regarding functional words,** as **they naturally tend** to **engage** in a **“skimming** and **scanning” reading process, focusing more** on **content words.**

These and other lessons learned about annotators' behaviour made us to adopt the **double non-blind revision** in the last step of the revision task: the annotators checked each other's work and were allowed to communicate to discuss disagreements.



Combining their revision capacities, they generated **synergy.**

Moreover, we noticed **greater motivation** when their task was no longer totally solitary.

The cases in which they were **unable to reach a consensus** were **revised by an experienced linguist.**

Resources developed during the revision process are contributing to the annotation of other genres:

Annotation Manuals:
https://drive.google.com/file/d/1ile8Wfxu1qdrZOmLGqkvVuQ4fXvHgVMo/view
https://drive.google.com/file/d/1BddPswn-_Ioo-A5GsIdA1cO1kqbcCahb/view

Portilexicon: https://portilexicon.icmc.usp.br/
an entry for each combination of form, lemma, UD PoS tag and UD features

Verifica-UD: http://verificaud.icmc.usp.br/
several rules to identify errors or possible errors in annotation (combining features, PoS tags and deprel restrictions)

Porttinari-base, our first annotated corpus, was launched in 2023 (Duran et al., 2023) and has been used to train a state-of-the-art **parser** (Lopes and Pardo, 2024), reaching **over 96% of accuracy**. We have been using this parser to preannotate corpora of new genres within the larger multi-genre project Porttinari.

Corpora and other resources are freely available on the POeTiSA project website:

https://sites.google.com/icmc.usp.br/poetisa

Thank you!