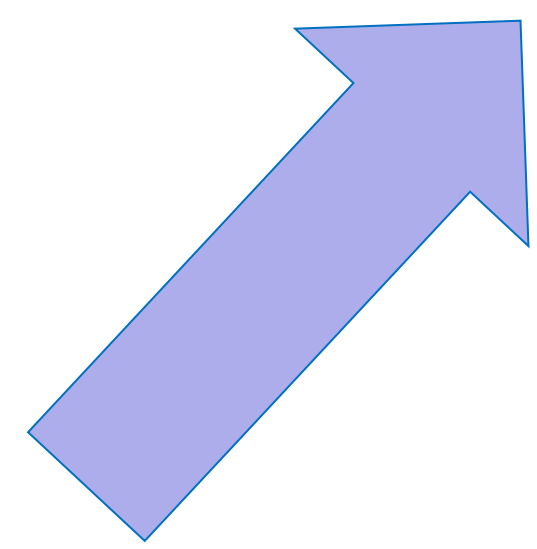




## UNIVERSAL DEPENDENCIES TREEBANKS

- Treebanks are used as training data by NLP tools such as spaCy and StanfordNLP for tagging and parsing text
- Universal Dependencies (UD): framework for consistent crosslinguistic treebank annotation

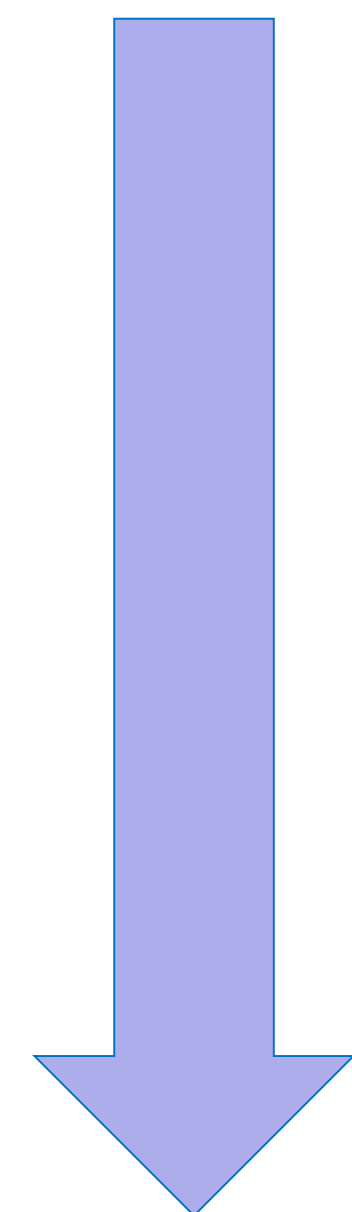
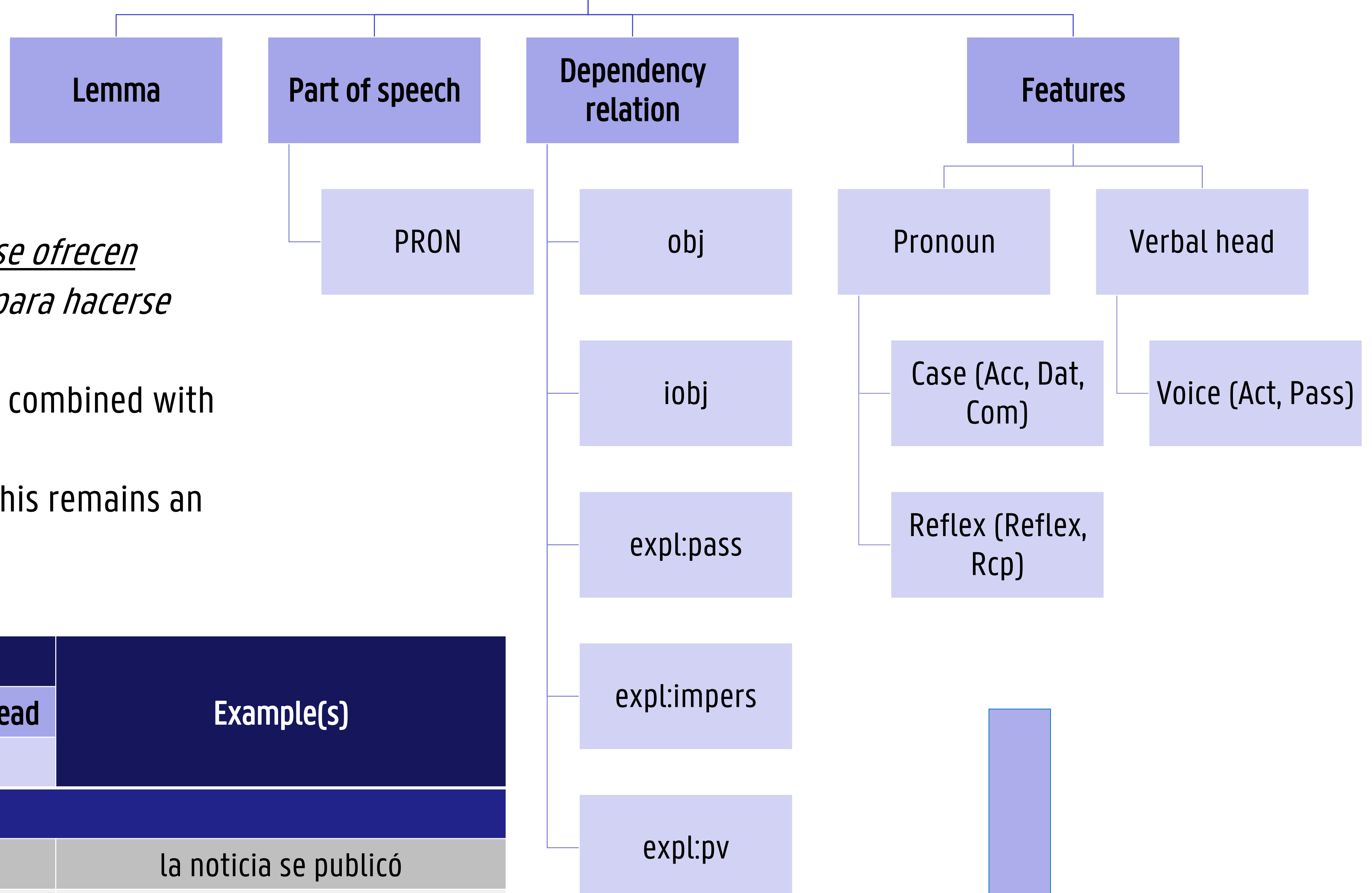


```
# sent_id = test-s18
# text = - Se acuerdan ustedes de eso que se llamaba la voz de la conciencia?
# orig_file_sentence 001#18
1 - PUNCT   FunctType=Dash 3 punct
2 se PRON   Case=Acc,Dat|Person=3|PrepCase=Npr|PronType=Prs|Reflex=Yes 3 obj
3 acuerdan VERB Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin 0 root
4 ustedes PRON Case=Acc,Nom|Number=Plur|Person=2|Polite=Form|PronType=Prs 3 nsubj
5 de ADP AdpType=Prep 6 case
6 eso PRON Number=Sing|PronType=Dem 3 obj
7 que PRON PronType=Int,Rel 9 nsubj
8 se PRON Case=Acc,Dat|Person=3|PrepCase=Npr|PronType=Prs|Reflex=Yes 9 obj
9 llamaba VERB Mood=Ind|Number=Sing|Person=3|Tense=Imp|VerbForm=Fin 6 acl
10 la DET Definite=Def|Gender=Fem|Number=Sing|PronType=Art 11 det
11 voz NOUN Gender=Fem|Number=Sing 9 obj
12 de ADP AdpType=Prep 14 case
13 la DET Definite=Def|Gender=Fem|Number=Sing|PronType=Art 14 det
14 conciencia NOUN Gender=Fem|Number=Sing 11 nmod SpaceAfter=No
15 ? PUNCT FunctSide=Fin|FunctType=Qest 3 punct
```

## CURRENT ANNOTATION IN SPANISH UD ANCORA TREEBANK

- Faulty, inconsistent and coarse-grained annotation of potentially reflexive pronouns (*me, te, nos, os* and especially *se*) → faulty, inconsistent and coarse-grained tagging and parsing
- 30% of *se* annotated as passive marker are false positives
- 60% of reflexive pronouns annotated as indirect object are false positives
- No distinction between *se* as passive marker (*en este volumen se ofrecen textos económicos*) and *se* as reflexive object (*Pedro se ofrece para hacerse cargo del bebé*)
- Non-coreferential indirect objects where *se* substitutes *le* when combined with accusative *lo/a(s)* as in *se lo pago* annotated as reflexive
- Current proposals for improving annotations<sup>[1, 2]</sup> point out that this remains an unsolved question

### Reflexive pronouns in Spanish UD treebanks (new proposal)



## REANNOTATION OF UD ANCORA

- Use existing properties in UD annotation fields DEPREL (dependency relation to the head) and FEATS (additional morphological features) to distinguish between different uses of reflexive pronouns in annotations
- Distinction between reflexive (*Reflex*) and reciprocal use (*Rcp*) in annotation of *Reflex* feature
- Fine-grained disambiguation of reflexives that are part of inherently pronominal verbs
- Annotation of the *Voice* feature of the reflexive's verbal head

Dependency relation	Features			Example(s)
	Pronoun		Verbal head	
	Case	Reflex	Voice	
<b>Reflexive uses</b>				
<i>expl:pass</i> (expletive expressing passive use)	Acc	Reflex	Pass	la noticia se publicó
<i>obj</i> (direct object)	Acc	Reflex	Act	Pedro se ve en el espejo
	Acc	Rcp	Act	Pedro y María se vieron en la calle
<i>iobj</i> (indirect object)	Dat	Reflex	Act	Pedro se quita la ropa
	Dat	Rcp	Act	Pedro y María se dieron la mano
<i>expl:impers</i> (expletive expressing impersonal use)	/	Reflex	Act	se trabaja mucho
<i>expl:pv</i> (expletive expressing inherently pronominal use)	With corresponding transitive non-reflexive verb			
	Acc	Reflex	Pass	el fenómeno se manifiesta
	Acc	Reflex	Act	la gente se manifiesta
	Acc	Rcp	Act	Pedro y María se ponen de acuerdo
	Dat	Reflex	Act	Pedro se da cuenta de algo
	Dat	Rcp	Act	?
	Com	Reflex	Act	Pedro se llevó el regalo
	With corresponding intransitive non-reflexive verb			
	/	Reflex	Act	Pedro se muere; Pedro se va
	Without corresponding non-reflexive verb			
Acc	Reflex	Act	Pedro se atreve a hacer algo	
<b>Non-reflexive uses</b>				
<i>obj</i>	Acc	/	<unknown>	me/te/nos/os ven
<i>iobj</i>	Dat	/	<unknown>	me/te/nos/os/se lo dijeron