category of "expl:pv", both with inanimate (b) and with animate subjects (d). These two examples are far from being clear-cut passives and reflexives, respectively, and thus would better be labeled as "expl:pv", but they also clearly differ from each other. Intuitively speaking, the first example seems more passive than the second, and the second more reflexive than the first. We believe that these intuitions can be captured by combining the different annotation layers: (b) and (d) receive the same dependency label "expl:pv", but their underlying features allow distinguishing between them. Concretely, with transitive verbs "Case" has to be disambiguated between "Acc" and "Dat" (for simplicity we leave "Com" out of the discussion), "Reflex" between "Reflex" and "Rcp", and "Voice" between "Act" and "Pass". With *manifestar*, "Case" would be "Acc" in the four cases, since this is the role that the "reflected argument" would play in a non-expletive construction (namely an active transitive construction for (b) or a true reflexive for (d)). "Reflex" would also be "Reflex" in the four cases, but "Voice" would be "Pass" in (b) and "Act" in (d), reflecting the intuition that the core semantic role of the subject is to undergo the process in (b), and to control it in (d).

		Dependency relation	Features reflexive pronoun		Features verbal head
			Case	Reflex	Voice
a	*como se manifestó en el periódico*	expl:pass	Acc	Reflex	Pass
b	*los problemas se manifestaron desde el primer día*	expl:pv			
c	*Dios se manifestó a sí mismo en Cristo*	obj	Acc	Reflex	Act
d	*la gente se manifiesta por tercer día consecutivo; el presidente se manifestó de acuerdo con … (*a sí mismo)*	expl:pv			

Table 1: Feature annotation on passives, reflexives and their corresponding "expl:pv". Translations: (a) 'as was said in the newspaper', (b) 'the problems became clear from the first day', (c) 'God materialized himself in Christ', (d) 'people demonstrated for the third consecutive day'; 'the president said he agreed with the proposal' (*him/herself)

Crucially, the feature sets link (b) with (a), and (d) with (c), respectively. This means that the expletive reflexive in (b) modifies an inherently passive construction (converting it, prototypically, into a spontaneous process, see the middle voice above), and that in (d), the expletive modifies an inherently reflexive construction, evoking event structures in which it is not relevant to distinguish two separate thematic roles for the reflected argument.

Similarly, the "Case=Acc/Dat" and "Reflex=Reflex/Rcp" properties also enable us to distinguish different underlying structures within the broad category of "expl:pv" examples. As is illustrated in Table 2, the difference between accusative (f) and dative reflexive (h) shows similarities with (e) and (g), respectively, and the difference between accusative reflexive (f) and reciprocal (j) is similar to the difference between (e) and (i).

		Dependency relation	Features reflexive pronoun		Features verbal head
			Case	Reflex	Voice
e	*se ve en el espejo*; *se mete en líos*	obj	Acc	Reflex	Act
f	*se ve amenazado de*; *se mete a hacer algo*	expl: pv			
g	*se quita la ropa*; *se da un baño*	iobj	Dat	Reflex	Act
h	*se da cuenta*	expl:pv			
i	*se saludan*; *se quieren mucho* (*el uno al otro*)	obj	Acc	Rcp	Act
j	*se llevan bien*; *se ponen de acuerdo* (**el uno al otro*)	expl:pv			

Table 2: Feature annotation on accusative/dative reflexives, accusative reflexives/reciprocals and their corresponding "expl:pv". Translations: (e) 'he sees himself in the mirror'; 'he gets himself in trouble', (f) 'he is threatened by; 'he starts doing something', (g) 'he takes off his clothes'; 'he takes a bath', (h) 'he realizes something', (i) 'they greet each other'; 'they love each other', (j) 'they get along well'; 'they agree' (*each other)

2.4 Reannotating Spanish UD AnCora

In Table 3 we present a comprehensive view on the proposed encodings. First, the pronouns were disambiguated according to their general reflexive character, distinguishing between *me veo* ('I see myself') and *me ven* ('they see me'). In the latter group, a distinction is made between "obj" and "iobj" (*me dieron algo*, 'they gave me something') at the level of the dependency relation.

Secondly, the reflexive uses were assigned one of the dependency labels "expl:pass", "obj", "iobj", "expl:impers" and "expl:pv". This means that reflexive and non-reflexive "obj" and "iobj" have the same dependency label but are distinguished by the feature "Reflex", which is absent in the case of non-reflexives. Reflexive "obj" and "iobj" are further subdivided according to their genuine reflexive versus reciprocal use.

Thirdly, the umbrella category "expl:pv" consists of three subgroups, namely constructions with corresponding transitive verbs, constructions which show an alternation with intransitive verbs, and constructions without corresponding (in)transitive verbs. The first group of "transitivity-based" reflexive constructions is then further subdivided by assigning different combinations of feature sets, as was explained in Section 2.3.4. These feature sets overlap with other "non-expl:pv" constructions, showing their shared characteristics. The proposal also foresees an "expl:pv" category with "Case=Dat", "Reflex=Rcp" and "Voice=Act", although this use does not seem to occur in Spanish.

Based on this annotation scheme, we then manually reannotated the AnCora treebank (both the test set and the training set). Table 4 includes a quantitative overview of the original dependency relation labels of all potentially reflexive pronouns (note that "expl:impers" and "expl:pv" do not occur in the original treebank), compared to their new labels after manual reannotation. Apart from the (very numerous) changes in dependency label, it is also worth noting that our reannotation removed the "Reflex" feature from 26 non-coreferential instances of *se*, that adding "Voice=Pass" to the feature set of the verbal head now allows identifying the passive reading of 2715 verb forms, and that, finally, the reciprocal character of 105 pronouns is now reflected in the feature set thanks to the introduction of the "Reflex=Rcp" property.

	Features			
	Pronoun		Verb	
	Case	Reflex	Voice	
Reflexive uses				
expl:pass	Acc	Reflex	Pass	*la noticia se publicó*
obj	Acc	Reflex	Act	*Pedro se ve en el espejo*
	Acc	Rcp	Act	*Pedro y Juan se vieron en la calle*
iobj	Dat	Reflex	Act	*Pedro se quita la ropa*
	Dat	Rcp	Act	*Pedro y Juan se dieron la mano*
expl:impers	-	Reflex	Act	*se trabaja mucho*
expl:pv	***with corresponding non-reflexive transitive verb***			
	Acc	Reflex	Pass	*el fenómeno se manifiesta*
	Acc	Reflex	Act	*la gente se manifiesta*
	Acc	Rcp	Act	*Pedro y Juan se ponen de acuerdo*
	Dat	Reflex	Act	*Pedro se da cuenta*
	Dat	Rcp	Act	*?*
	Com	Reflex	Act	*Pedro se llevó el regalo*
	with corresponding non-reflexive intransitive verb			
	-	Reflex	Act	*Pedro se muere*
	without corresponding non-reflexive verb			
	Acc	Reflex	Act	*Pedro se atreve a …*
Non-reflexive uses				
obj	Acc		-	*me/te/nos/os ven*
iobj	Dat		-	*me/te/nos/os/se lo dijeron*

Table 3: Overview of the annotation scheme for potentially reflexive pronouns in Spanish

reannotated / original	expl:impers	expl:pass	expl:pv	iobj	obj	**Total**
expl:pass	285	139	28	1	5	**458**
iobj	1	15	217	142	38	**413**
obj	52	1880	2603	573	618	**5726**
Total	**338**	**2034**	**2848**	**716**	**661**	**6597**

Table 4: Overview of the dependency relation changes in Spanish UD AnCora (test + train)

3 Conclusion

We have argued that in current Spanish Universal Dependencies treebanks, the annotation of reflexives is an unsolved problem. Given the frequency of this construction, occurring for example in more than 20% of the sentences in written texts, this has considerable consequences for parser accuracy and/or granularity. Reflexives, and particularly so-called *se* constructions, have been heavily debated in the tradition of Spanish linguistics. Although it cannot be the aim of morpho-syntactic and dependency parsing to reflect all possible semantic nuances, we have shown that a layered annotation strategy, which combines a relatively limited number of UD dependency relations and feature set properties, can capture both constructional similarities and diversity. We applied this proposal to the v2.5 Spanish UD AnCora treebank and provide categorized conversion tables that can be run as a Python script (see Appendix A and B).

Bibliography

Bouma, G., Hajic, J., Haug, D., Nivre, J., Solberg, P. E., and Øvrelid, L. 2018. Expletives in Universal Dependency Treebanks. In *UDW 2018*, 18-26.

Croft, W., Nordquist, D., Looney, K., and Regan, M. 2017. Linguistic Typology meets Universal Dependencies. In *TLT* 2017: 63-75.

Goethals, P. (2018). Customizing vocabulary learning for advanced learners of Spanish. In T. Read, B. Sedano Cuevas, and S. Montaner-Villalba (Eds.), *Technological innovation for specialized linguistic domains* (pp. 229-240). Berlin: Éditions Universitaires Européennes.

Maldonado, R. 2008. Spanish middle syntax: A usage-based proposal for gramar teaching. In S. De Knop and T. De Rycker (eds.) *Cognitive Approaches to Pedagogical Grammar*, 155-196. Berlin: Mouton De Gruyter.

Marković, S., and Zeman, D. 2018. Reflexives in Universal Dependencies. In TLT 2018.

Martínez Alonso, H. and Zeman D. 2016. Universal Dependencies for the AnCora treebanks. In *Procesamiento de Lenguaje Natural*, 57, 91-98.

Mendikoetxea, A. 1999. Construcciones inacusativas y pasivas. In *Gramática descriptiva de la lengua española*, 2, 1575-1629. Espasa Calpe.

Nivre, J., M.-C. de Marneffe, F. Ginter, Y. Goldberg, J. Hajič, C. Manning, R. McDonald, S. Petrov, S. Pyysalo, N. Silveira, R. Tsarfaty, and D. Zeman. 2016. Universal dependencies v1: A multilingual treebank collection. *LREC* 2016.

Peregrín Otero, C. 1999. Pronombres reflexivos y recíprocos. In *Gramática descriptiva de la lengua española*, 1, 1427-1518. Espasa Calpe.

Silveira, N. 2016. *Designing syntactic representations for NLP: An empirical investigation.* PhD Thesis. Stanford University.

Taulé, M., M. A. Martí, and M. Recasens. 2008. AnCora: Multilevel annotated corpora for Catalan and Spanish. In LREC 2008.

Appendix A: Conversion table

The conversion table includes all occurrences of *me*, *te*, *nos*, *os* and *se*. Other users can modify or customize the annotation decisions.

Appendix B: Python script

The Python script reads in the original CoNLL-U AnCora files, and applies all the changes to the corresponding dependency relations and feature sets. The appended files are available upon request (by email, to Jasper.Degraeuwe@UGent.be).