Universal Dependencies for Irish

Teresa Lynn [1,2]   Jennifer Foster[1]

(1) ADAPT Centre, School of Computing, Dublin City University, Ireland
(2) Department of Computing, Macquarie University, Australia
`tlynn@computing.dcu.ie, jfoster@computing.dcu.ie`

RÉSUMÉ ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯
Dépendances universelles de l'irlandais

Les ressources linguistiques permettant aux études cross-langues de se développer sont très importantes pour les langues minoritaires telles que l'irlandais, car elles favorisent le partage des ressources pour palier au problème du manque de données. Le projet «Universal Dependencies » (UD) a pour but de faciliter les études cross-langues des arbres syntaxiques, des structures linguistiques et de l'analyse syntaxique. L'objectif principal de ce projet est de former un ensemble harmonieux d'arbres syntaxiques en utilisant un schéma d'annotations universelles. Dans cet article, nous présentons la transformation de l'arbre de dépendance syntaxique irlandais (IDT) (Lynn, 2016) au schéma d'annotations universelles du projet UD, suivie d'une description claire des changements structurels nécessaires à cette conversion. Le nouvel arbre est ainsi appelé « Irish Universal Dependency Treebank » (IUDT).

ABSTRACT ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯
Language resources that enable cross-lingual studies have become increasingly valuable for lesser-resourced languages such as Irish, as they allow for easier sharing of resources, thus overcoming the problem of data scarcity. The Universal Dependencies (UD) Project[1] is an initiative aimed at cross-lingual studies of treebanks, linguistic structures and parsing. Its goal is to create a set of multilingual harmonised treebanks that are designed according to a universal annotation scheme. In this paper, we report on the conversion of the Irish Dependency Treebank (IDT) (Lynn, 2016) to a UD version of the treebank which we term the Irish Universal Dependency Treebank (IUDT). We report on the mapping of the IDT labelling scheme to the UD scheme, along with a clear description of the structural changes required in this conversion.

MOTS-CLÉS : Analyse syntaxique, irlandais, langue irlandaise, arbre de dépendance syntaxique, dépendances syntaxiques universelles, conversion, étiquettes.

KEYWORDS: parsing, Irish, dependency treebank, universal dependencies, mapping, labels.

1 Introduction

Dependency treebanks exist for many languages (e.g. Turkish (Oflazer *et al.*, 2003), Czech (Hajič, 1998), Danish (Kromann, 2003), Slovene (Džeroski *et al.*, 2006) and Finnish (Haverinen *et al.*, 2010)). However, these treebanks vary significantly, with labelling notations and linguistic analyses that are usually specific to that language, and often influenced by linguistic theories to which the developers

---

[1]http://universaldependencies.org/

subscribe. As a result, cross-lingual research is often hampered by variations that exist across the annotation schemes of treebanks. From a statistical parsing perspective, if the labelled training data for both languages is based on different annotation schemes, parser output in one language cannot be easily compared or transferred to another (Søgaard, 2011; McDonald *et al.*, 2011). McDonald *et al.* (2013) reported improved results on cross-lingual transfer parsing using 10 uniformly annotated treebanks. Lynn *et al.* (2014) also reported on similar experiments using the same treebanks to bootstrap parsing for Irish.

In October 2014, the Universal Dependency (UD) Project released guidelines to assist with the creation of new UD treebanks, or mappings and conversions of existing treebanks to a *new* universal scheme. This new annotation scheme is based on (universal) Stanford dependencies (de Marneffe *et al.*, 2006; de Marneffe & Manning, 2008; de Marneffe *et al.*, 2014), Google universal part-of-speech tags (Petrov et al., 2012), and the Interset interlingua for morphosyntactic tagsets (Zeman, 2008). The UD scheme accounts for varying linguistic differences across languages by providing the option of defining language-specific label sub-types when the prescribed list of labels do not adequately cover all linguistic features of a given language. Nivre (2015) clearly explains the motivation behind the project. Ten treebanks were released in January 2015 including Czech, English, Finnish, French, German, Hungarian, *Irish*, Italian, Spanish and Swedish. Since then a large number of additional treebanks have been either (i) built from scratch or (ii) converted from existing treebanks to form new UD treebanks. To date[2], there are 54 treebanks representing 40 languages listed in the UD project.

We have mapped the Irish Dependency Treebank (IDT) (Lynn, 2016) to the UD scheme (v1) for purposes of cross-lingual studies and parser improvement. The IDT is a corpus[3] of Irish sentences that have been annotated with information on deep syntactic structure. This paper summarises the conversion and mapping of the IDT to the Irish Universal Dependency Treebank (IUDT), as part of the Universal Dependencies (UD) Project[4].

2   Mapping the Irish POS tagset to the Universal POS tagset

The UD part-of-speech (POS) tagset is an extension of the The Google Universal POS tagset (Petrov *et al.*, 2012) and contains 17 POS tags. The IDT was built upon a gold-standard POS-tagged corpus developed by Uí Dhonnchadha (2009), and is based on the PAROLE Morphosyntactic Tagset (ITÉ, 2002). The IDT's tagset contains both coarse- and fine-grained POS tags, both of which we map to the Universal POS tags (e.g. Prop Noun → NOUN). Note, however, that we only map to 16 of the UD tags as we do not identify auxiliary verbs in Irish to require the inclusion of AUX. We provide a mapping from the Irish POS tagset to the UD tagset in Table 1.

3   Universal Dependency Scheme

The IDT to UD treebank conversion required extensive work on dependency relation renaming, mapping and structural changes. We provide a mapping in Table 2 and describe the changes below.

---

[2]May 2016

[3]Current treebank size is 1020 trees with 23,684 tokens. See Appendix C of Lynn (2016) for additional statistics.

[4]http://universaldependencies.org

Part-of-speech (POS) mappings			
UD	IDT	UD	IDT
NOUN	Noun Noun, Pron Ref, Subst Subst, Verbal Noun,	ADP	Prep Deg, Prep Det, Prep Pron, Prep Simp, Prep Poss, Prep CmpdNoGen, Prep Cmpd, Prep Art, Pron Prep
PROPN	Prop Noun	ADV	Adv Temp, Adv Loc, Adv Dir, Adv Q, Adv Its, Adv Gn
PRON	Pron Pers, Pron Idf, Pron Q, Pron Dem	PART	Part Vb, Part Sup, Part Inf, Part Pat, Part Voc, Part Ad, Part Deg, Part Comp, Part Rel, Part Num, Part Cp,
VERB	Cop Cop, Verb PastInd, Verb PresInd, Verb PresImp, Verb VI, Verb VT, Verb VTI, Verb PastImp, Verb Cond, Verb FutInd, Verb VD, Verb Imper	NUM	Num Num
DET	Art Art, Det Det	X	Item Item, Abr Abr, CM CM, CU CU, CC CC, Unknown Unknown, Guess Abr, Foreign Foreign
ADJ	Prop Adj, Verbal Adj, Adj Adj	PUNCT	. . .. ? ? ! ! : : ? . Punct Punct
CONJ	Conj Coord	INTJ	Itj Itj
SCONJ	Conj Subord	SYM	(Abr)

Table 1: Mapping of the IDT's POS pairs (coarse fine) to the Universal Dependency POS tagset.

3.1 UD labels not used in the Irish UD Treebank

The following is a list of labels in the UD annotation scheme that do not apply to the Irish language:

- `aux`: This label is used for non-main verbs in a clause, i.e. auxiliary verbs. Examples in English are '*has* opened', '*will* be', '*should* say'. There are no equivalent auxiliary verbs in Irish.[5]

- `auxpass, nsubjpass, csubjpass`: These labels are used in passive constructions, respectively as: passive auxiliary verbs, passive nominal subjects and clausal passive subjects. There is no equivalent passive form in Irish (see The Christian Brothers (1988, p.120) and Stenson (1981, p.145)).

- `iobj`: In English, an example is 'Mary gave *John* the book'. There are no indirect objects in Irish, and constructions like these must follow the normal ditransitive verb structure using a preposition (i.e. 'Mary gave the book to John').

Some UD labels are not used in IUDT due to lack of instances observed in the data[6]:

- `reparandum`: This label is used to indicate disfluencies in text. The IDT data does not currently contain any disfluencies.

---

[5] Stenson (1981, p.86) notes that modal verbs such as *caithfidh* inflect as per regular verbs and are considered the main verb.
[6] This may be related to the well-structured, grammatical nature of the text in the IDT corpus (e.g. newswire, literature).

UD Dependency Label Mappings			
Universal	**Irish**	**Universal**	**Irish**
root	top	*foreign*	for
acl:relcl	relmod	*list*	quant †
advcl	comp †	*mark*	subadjunct, toinfinitive
advmod	adjunct †, advadjunct, advadjunct_q, quant †	*mark:prt*	advparticle, cleftparticle, particle, qparticle, vparticle
amod	adjadjunct	*name* ±	nparticle, nadjunct †
appos	app	*neg*	vparticle
case ±	padjunct †, obl_ag	*nmod*	aug, pobj †±, relparticle †
case:voc	vocparticle	*nmod:poss*	poss
cc ±	–	*nmod:prep*±	obl, obl2
ccomp	comp †	*nmod:tmod*	advadjunct, padjunct †, pobj †±, relparticle †
compound	nadjunct †	*nsubj*	relparticle †, subj, subj_q
compound:prt	particlehead	*nummod*	quant †
conj ±	coord	*parataxis*	comp †
cop ±	NEW	*punct*	punctuation
csubj:cop	csubj	*vocative*	addr
det	det, det2, dem	*xcomp*	xcomp
discourse	adjunct †	*xcomp:pred*	adjpred, advpred, npred, ppred ±
dobj	obj, vnobj, obj_q, relparticle †		

Table 2: Mapping of Irish Dependency Annotation Scheme to UD Annotation Scheme. † marks one-to-many mappings, and ± marks structural changes. The IUDT uses 26 of the 40 UD labels (and 9 Irish-specific sub-labels).

- goeswith: This label links to parts of a word that has been split, due to poor editing. There are no instances of this in the Irish data.

- dep: This catch-all label is used for unknown relations. We do not require this in the Irish data.

In addition, there are some UD labels that we have not included in the first release version of this treebank, but which we expect will be included in future releases:

- expl: There is no existential 'there' in Irish. However, we have not yet fully researched uses of other types of expletives in the IDT data (e.g. *tá sé soiléir go..* 'it is clear that ..').

- mwe: Multiword expressions are not marked in the IDT. There is not sufficient linguistic literature on this topic for Irish on which we could base a complete analysis of idioms or multiword units in the treebank. This analysis therefore remains as a future enhancement to the treebanks when such resources are available.

- remnant: This label is used for remnants in ellipsis, where a predicate or verb is dropped (e.g. 'Marie went to Paris and Miriam [] to Prague'). Instances of remnants in Irish are not easily identified. Further study is required to identify cases, if any, including a possible analysis of crossing dependencies.

- dislocated: This label is used for fronted or postposed elements that are not core grammatical elements of a sentence. Example, 'he must not eat it, *the playdough*'. We have not yet identified such cases in the IDT data.

3.2   Manual label updates

Some of the treebank conversion was automated with straightforward mappings. However, there were a number of one-to-many label mappings that required manual mapping. These instances are marked with † in Table 2 and discussed here.

relative particles: In the IDT, the relative particle *a* is attached to a relative modifier verb with the label `relparticle`. In the UD scheme, this particle is labelled with the syntactic role it plays in the relative clause.[7] The *a* can therefore fulfil the role of `nsubj`, `dobj`, `nmod` or `nmod:tmod`[8]. For example, *an rud deireanach a chonaic sé* 'the last thing that he saw' is shown in Figure 1. In this case *a* refers to *rud* 'thing', and therefore is labelled as a `dobj` of *chonaic* 'saw'.

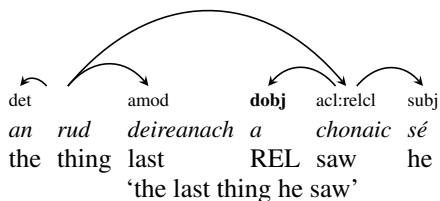| det	amod	**dobj**	acl:relcl	subj
an *rud*	*deireanach*	*a*	*chonaic*	*sé*
the thing	last	REL	saw	he

'the last thing he saw'

Figure 1: UD `dobj` relative particle analysis

quant → nummod, list, advmod Numerals and quantifiers are given more fine-grained descriptions in UD than the single IDT `quant` label. In addition, list numbering is represented by `list`.

comp → advcl, ccomp, parataxis The tokens labelled in the IDT with the closed complement label `comp` have been divided among three new labels. The UD labels are: `advcl` adverbial clause (normally connected with a subordinator such as *nuair* 'when', *má* 'if' etc); `ccomp` complement clauses that are normally introduced by the complementiser *go*, *nach*, *gur*, or quoting direct speech; `parataxis` labels two phrases or sentences set side-by-side without explicit linking through coordination or subordination, for example. Sometimes punctuation such as colons or semicolons connects the pairs. *Bhí an cál an-ghann; b'fheidir nach mbeadh i ngach baile ach aon gharraí amháin.* 'Kale was very scarce; maybe there would only be one garden in every town'.

nadjunct → compound, name The compound label is used for nominal modifiers. In Irish this could take the form of compounding (one noun modifying another) such as *deireadh seachtaine* 'weekend', or ownership *teach Mhichil* 'Michael's house'. Compounding can occur with a string of nouns as per the example in Figure 2.

The new label `name` is explained below in more detail in Section 3.3.

3.3   Structural Changes

Other labels required a manual annotation because they related to structural changes required in the treebank that were not easily automated. The following structural changes were made manually before the dependency labels were mapped to the universal scheme.

---

[7]This type of annotation that cannot be automated in the absence of additional data on the semantic properties of the element to which the relativiser refers.

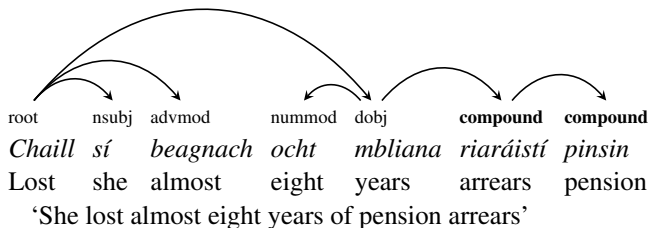[8]Irish language-specific label for temporal modifiers in nominal form.

Figure 2: UD compounding analysis

coordination   Significant changes were required to the analysis of coordination while mapping to IUDT. The IDT follows the Lexical Functional Grammar (LFG) (Bresnan, 2001) coordination analysis, where the coordinating conjunction (e.g. *agus* 'and') is the head, with each coordinate as its dependents, labelled as coord (see Figure 3). The UD annotation scheme, on the other hand, uses right-adjunction, where the first coordinate is the head of the coordination, and the rest of the phrase is adjoined to the right, labelling coordinating conjunctions as cc and subsequent coordinates as conj (Figure 4).
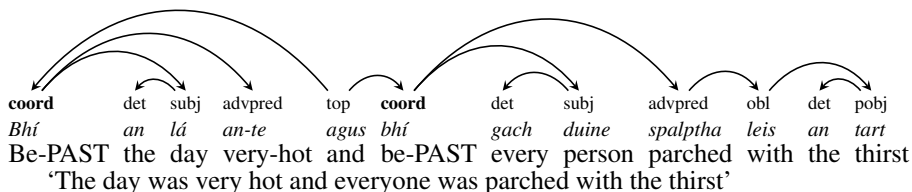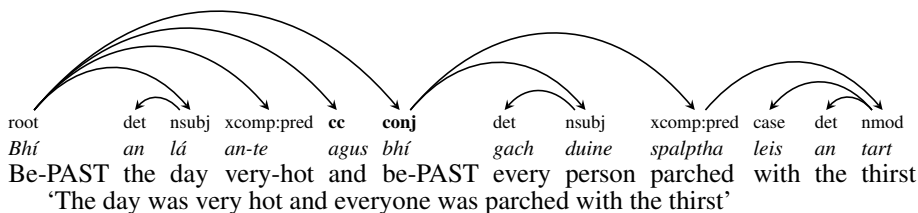

Figure 3: LFG-style coordination of the IDT

Figure 4: Coordination structure in the IUDT

subordinate clauses   In the IDT, the analysis of the relationship between the matrix clause and a subordinate clause is similar to that of LFG: the subordinating conjunction (e.g. *mar* 'because', *nuair* 'when') is a subadjunct dependent of the matrix verb, and the head of the subordinate clause is a comp dependent of the subordinating conjunction (Figure 5). In contrast, the UD scheme marks the head of the subordinate clause as a dependent of the matrix verb, and the subordinating conjunction is a dependent of the subordinate clause (Figure 6).
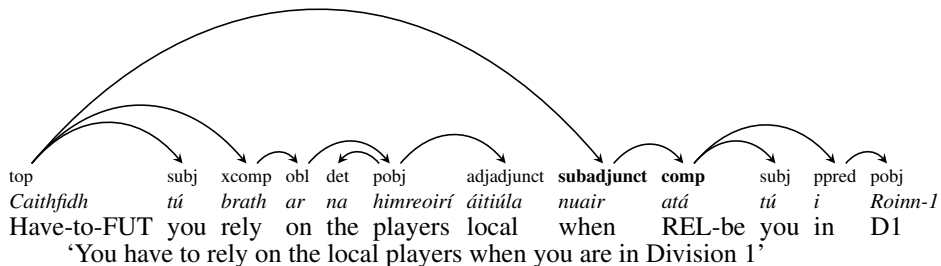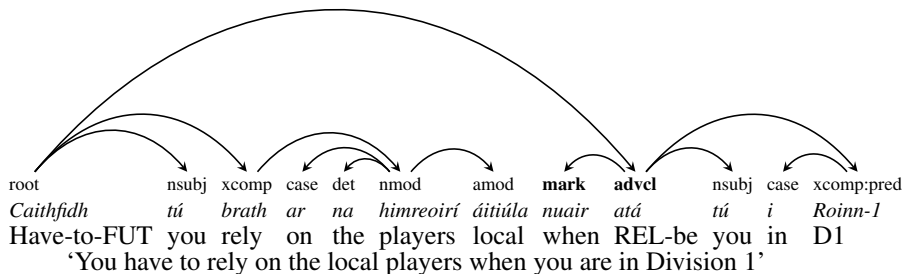
Figure 5: IDT subordinate clause analysis

top		subj	xcomp	obl	det	pobj	adjadjunct	**subadjunct**	**comp**	subj	ppred	pobj
Caithfidh		*tú*	*brath*	*ar*	*na*	*himreoirí*	*áitiúla*	*nuair*	*atá*	*tú*	*i*	*Roinn-1*
Have-to-FUT		you	rely	on	the	players	local	when	REL-be	you	in	D1

'You have to rely on the local players when you are in Division 1'

Figure 6: IUDT subordinate (adverbial) clause analysis

root		nsubj	xcomp	case	det	nmod	amod	**mark**	**advcl**	nsubj	case	xcomp:pred
Caithfidh		*tú*	*brath*	*ar*	*na*	*himreoirí*	*áitiúla*	*nuair*	*atá*	*tú*	*i*	*Roinn-1*
Have-to-FUT		you	rely	on	the	players	local	when	REL-be	you	in	D1

'You have to rely on the local players when you are in Division 1'

cop[9] In the IDT, the copula is treated similarly to a verb, and can function as the root of a sentence, or as the head of a dependency clause. However, the UD scheme analyses copula constructions differently. Instead, the predicate is regarded as the head of the phrase, and the copula is its dependent, as indicated by the `cop` label. This also applies to copula use in fronting or cleft structures. See Figure 7 and Figure 8 for comparison.[10]
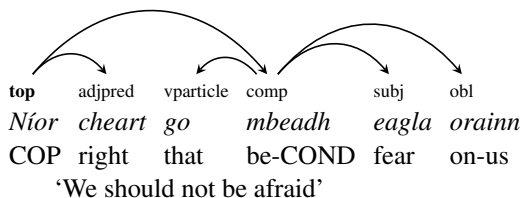

Figure 7: IDT copula analysis

name: The UD relation `name` is used with compounding proper nouns, typically for names of people,

---

[9]Note that Irish has two forms of the verb 'to be' – the copula and the substantive verb *bí*. Constructions using the substantive verb are not analysed using the UD `cop` label and are treated like regular verbs instead. For example, *tá sé fuar* 'it is cold'

[10]The labels have also been mapped between examples, but the structural change is of interest here.

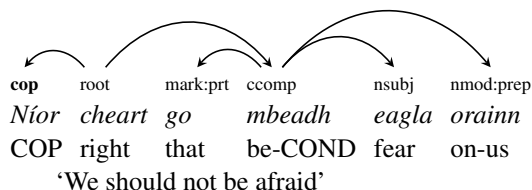| cop	root	mark:prt	ccomp	nsubj	nmod:prep
Níor	*cheart*	*go*	*mbeadh*	*eagla*	*orainn*
COP	right	that	be-COND	fear	on-us

'We should not be afraid'

Figure 8: UD copula analysis

places, organisations and so on. In Irish, this not only includes surnames, but also surname particles such as *Mac, Mc, Ó, de, Uí* and *Ní*. In the IDT, the surname is the head noun, and its dependents can either be first names (nadjunct) or nominal particles (nparticle). See Figures 9 for example. However in the UD analysis, the first word is the head, modified by the rest of the words as name. See Figure 10 for comparison.
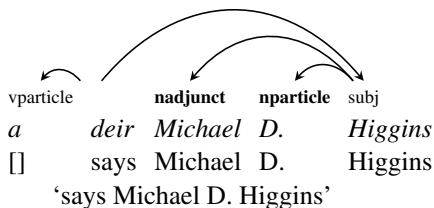

vparticle		nadjunct	nparticle	subj
a	*deir*	*Michael*	*D.*	*Higgins*
[]	says	Michael	D.	Higgins

'says Michael D. Higgins'

Figure 9: IDT name analysis

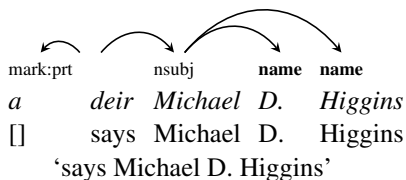| mark:prt		nsubj	name	name
a	*deir*	*Michael*	*D.*	*Higgins*
[]	says	Michael	D.	Higgins

'says Michael D. Higgins'

Figure 10: UD name analysis

nmod, case, xcomp:pred In the IDT, the preposition is the head of a prepositional phrase (PP). UD recognises the head noun of the object NP as the PP head. This affects the Irish treebank in a number of ways:

In the UD analysis, the head of regular preposition phrases (object of the preposition) is attached to the verb as nmod (formerly pobj in IDT). The preposition is a dependent of the object, and this relation is labelled as case. Compare Figures 11 and 12 to observe the difference in analyses.

Irish progressive aspectual phrases are constructed with the preposition *ag* followed by a verbal noun. The IDT regards *ag* as the head of the prepositional phrase, and thus the open complement label (xcomp) marks the relation between the matrix verb and the preposition. In the UD scheme however, the verbal noun is regarded as the head of the prepositional phrase. Compare Figures 13 and 14.
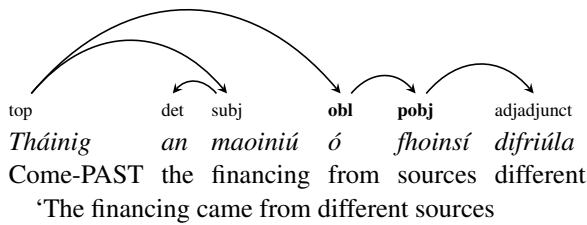
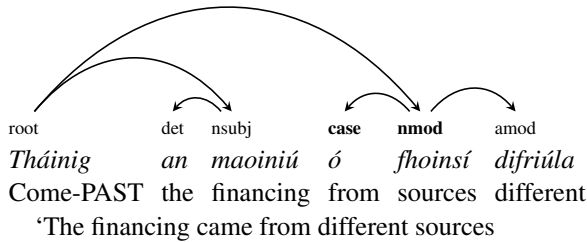Figure 11: IDT prepositional phrase analysis

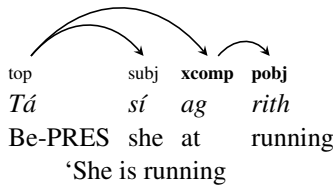Figure 12: UD prepositional phrase analysis

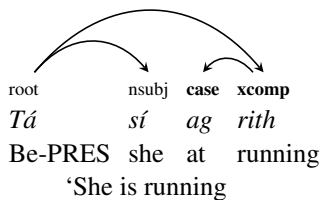Figure 13: IDT progressive aspectual phrase analysis

Figure 14: UD progressive aspectual phrase analysis

Prepositional predicates are labelled as `ppred` in the Irish Dependency Treebank. In keeping with the other PP analyses, the preposition is the head of the prepositional phrase. The IDT label `ppred` maps to `xcomp:pred` in the UD scheme.[11] In addition, the object of the preposition is now regarded as the head of the phrase. See Figures 15 and 16 for comparison of prepositional predicate analyses.

---

[11] The label `xcomp:pred` is an Irish-specific label, these language specific labels are discussed in Section 3.4.
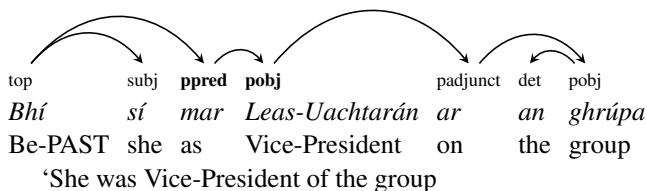
top	subj	**ppred**	**pobj**		padjunct	det	pobj
Bhí	*sí*	*mar*	*Leas-Uachtarán*		*ar*	*an*	*ghrúpa*
Be-PAST	she	as	Vice-President		on	the	group

'She was Vice-President of the group

Figure 15: IDT prepositional predicate analysis

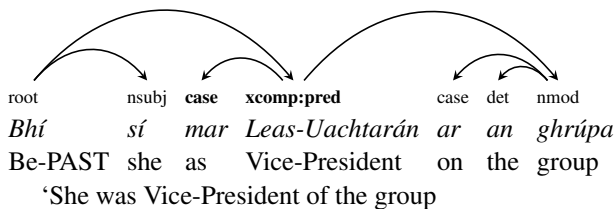|						
root	nsubj	**case**	**xcomp:pred**	case	det	nmod
Bhí	*sí*	*mar*	*Leas-Uachtarán*	*ar*	*an*	*ghrúpa*
Be-PAST	she	as	Vice-President	on	the	group

'She was Vice-President of the group

Figure 16: UD prepositional predicate analysis

3.4 Irish-specific relations

The UD scheme provides scope to include language-specific subtype labels. The label naming format is *universal:extension*, which ensures that the core UD relation remains identifiable, making it possible to revert to this coarse label for cross-lingual analysis. During the conversion of the IDT, we defined some labels required to represent Irish syntax more concisely. These labels are discussed below.

acl:relcl: This label is used for relative clause modifiers. We use this subtype label `acl:relcl` in cases where the head of the relative clause is a predicate (usually a verb), and is dependent on a noun in a preceding clause. It is also used in the English, Finnish and Swedish schemes. An example of this subtype used in the converted IUDT is in Figure 17.
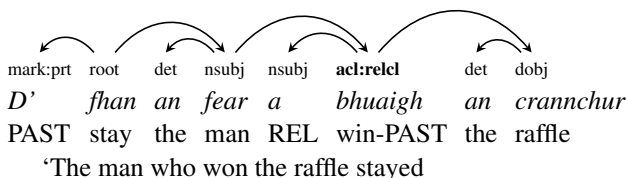

mark:prt	root	det	nsubj	nsubj	**acl:relcl**	det	dobj
D'	*fhan*	*an*	*fear*	*a*	*bhuaigh*	*an*	*crannchur*
PAST	stay	the	man	REL	win-PAST	the	raffle

'The man who won the raffle stayed

Figure 17: UD relative clause analysis

case:voc: The vocative particle *a* is a case marker in Irish and precedes an addressee. We therefore use the `case:voc` label for vocative particles. For example, *Slán **a** chara* 'Goodbye, friend'.

compound:prt We use `compound:prt` for verbal particle-heads, in order to distinguish them as particles as opposed to nominal compounds (e.g. *leagtha amach* 'laid out').

csubj:cop: The supertype label `csubj` indicates a clausal subject (a clause whose role is the subject of another). In English '[what she `said`] makes sense'. However, Finnish uses an additional specific

subtype label `csubj:cop` to indicate clausal subjects that act as a subject of a copular clause. We observed in the IDT data that clausal subjects in Irish are only ever subjects of copula clauses. For this reason we use only the subtype label `csubj:cop` for clausal subjects (see Figure 18).
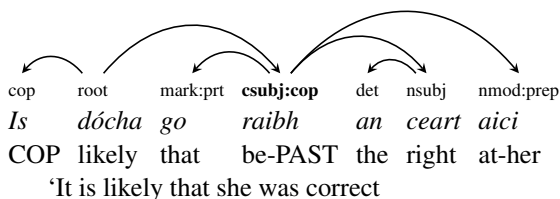
cop	root	mark:prt	**csubj:cop**	det	nsubj	nmod:prep
Is	*dócha*	*go*	*raibh*	*an*	*ceart*	*aici*
COP	likely	that	be-PAST	the	right	at-her

'It is likely that she was correct

Figure 18: UD copular clausal subject analysis

mark:prt: We introduce a new subtype label `mark:prt` for adverbial particles, cleft particles, quantifier particles, comparative/ superlative particles, verb particles and days of the week particles.

nmod:poss: In Irish, possession is denoted by possessive pronouns (*mo, do, a, ár, bhur*). English, Finnish and Swedish use the subtype label `nmod:poss` to indicate possession, and we also adopt it for Irish. The pronoun is a dependent of the noun to which it denotes ownership. For example, *Chuir mé ceist ar **mo mhúinteoir** 'I asked **my** teacher a question'.

nmod:prep: 16 of the most common Irish simple prepositions can be inflected to mark pronominal objects (e.g. *le* 'with' inflects as *liom* 'with-me') and are referred to as pronominal prepositions or prepositional pronouns.[12] In the UD scheme, where the object is the head of a PP, these inflected prepositions play nominal roles instead of prepositional roles.[13] We introduce the language-specific label `nmod:prep`, thus retaining information on the presence of the preposition within this synthetic form. An example is given in Figure 19.[14]
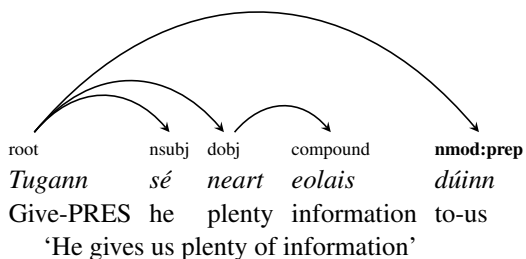
root	nsubj	dobj	compound	**nmod:prep**
Tugann	*sé*	*neart*	*eolais*	*dúinn*
Give-PRES	he	plenty	information	to-us

'He gives us plenty of information'

Figure 19: UD prepositional pronoun analysis

nmod:tmod: Temporal modifiers specifying time, in nominal form, are labelled as `nmod`. English also uses this subtype label. An example in Irish is *daoine a mhair na milliúin **bliain** ó shin* 'people who lived a million **years** ago'.

---

[12]Inflected prepositions were most frequently marked as either `obl` or `obl2` in the IDT.

[13]Their POS-tag remains `ADP`, however.

[14]Note that in some cases, prepositional pronouns behave like nominal modifiers of noun phrases. E.g. *an bheirt **acu*** 'the two **of them**'. These cases take the label `compound`.

xcomp:pred: The IDT uses the following fine-grained labels for predicates: `npred` (nominal), `adjpred` (adjectival), `advpred` (adverbial) and `ppred` (prepositional). These were typically used in copular constructions but are now no longer relevant in the UD, where the predicate heads the copular phrase. However, adjective, adverbial and prepositional predicates can also be arguments of the substantive verb *bí*. Therefore, we extend the open complement label to include the subtype `xcomp:pred`.[15] See Figure 20 for an example of an adjectival predicate.

root	nsubj	**xcomp:pred**	mark:prt	ccomp	nsubj	xcomp:pred
Bhí	*sé*	*dochreidte*	*go*	*raibh*	*sé*	*beo*
Be-PAST	it	unbelievable	COMP	be-PAST	he	alive

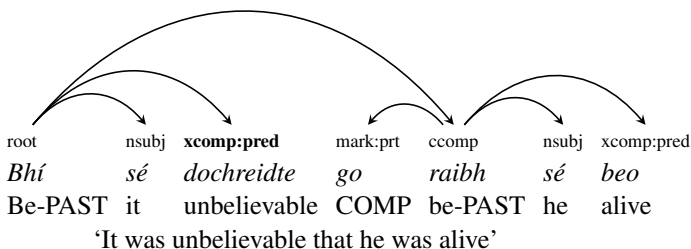'It was unbelievable that he was alive'

Figure 20: UD adjectival predicate analysis

4   Summary and Future Work

In this paper, we have summarised the conversion of the Irish Dependency Treebank (IDT) to a UD format (IUDT). We have described in detail the mapping and conversion process, including structural changes required, for the release of the IUDT as part of the Universal Dependencies project. We have also discussed linguistic analyses and motivations for choice of Irish language-specific label types. The Irish UD treebank (IUDT) is available to download under an open-source licence from The Universal Dependencies Project repository[16].

We have not discussed here the inclusion of morphological information in the IUDT as this still requires extensive documentation within the UD project. We plan to report on this at a later stage. In addition, as the IDT grows in size (a work in progress), we plan to extend the IUDT in parallel.

Acknowledgements

---

[15]This follows the LFG use of xcomp (open complement) to represent predicates.

[16]v1.3 `https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1699`

References

BRESNAN J. (2001). *Lexical Functional Syntax*. Oxford: Blackwell.

DE MARNEFFE M.-C., DOZAT T., SILVEIRA N., HAVERINEN K., GINTER F., NIVRE J. & D. MANNING C. (2014). Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC-2014)*, p. 4585–4592, Reykjavik, Iceland.

DE MARNEFFE M.-C., MACCARTNEY B. & MANNING C. D. (2006). Generating typed dependency parses from phrase structure trees. In *Proccedings of the 5th international conference on Language Resources and Evaluation (LREC 2006)*, p. 449–454, Genoa, Italy.

DE MARNEFFE M.-C. & MANNING C. D. (2008). The Stanford typed dependencies representation. In *Workshop on Crossframework and Cross-domain Parser Evaluation (COLING2008)*, Manchester, U.K.

DŽEROSKI S., ERJAVEC T., LEDINEK N., PAJAS P., ŽABOKRTSKY Z. & ŽELE A. (2006). Towards a Slovene dependency treebank. In *Proceedings of the Fifth Conference on Language Resources and Evaluation (LREC2006)*, p. 1388–1391, Genoa, Italy.

HAJIČ J. (1998). Building a syntactically annotated corpus: The Prague Dependency Treebank. In E. HAJIČOVÁ, Ed., *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, p. 12–19. Prague Karolinum, Charles University Press.

HAVERINEN K., VILJANEN T., LAIPPALA V., KOHONEN S., GINTER F. & SALAKOSKI T. (2010). Treebanking Finnish. In *Proceedings of The Ninth International Workshop on Treebanks and Linguistic Theories (TLT-9)*, p. 79–90, Tartu, Estonia.

ITÉ (2002). PAROLE Morphosyntactic Tagset for Irish. Institiúid Teangeolaíochta Éireann.

KROMANN M. (2003). The Danish Dependency Treebank and the DTAG Treebank Tool. In *Proceedings of the Second Workshop on Treebanks and Linguistic Theories (TLT2003)*, p. 217–220, Växjö, Sweden.

LYNN T. (2016). *Irish Dependency Treebanking and Parsing*. PhD thesis, Dublin City University.

LYNN T., FOSTER J., DRAS M. & TOUNSI L. (2014). Cross-lingual transfer parsing for low-resourced languages: An Irish case study. In *Proceedings of the First Celtic Language Technology Workshop*, p. 41–49, Dublin, Ireland.

MCDONALD R., NIVRE J., QUIRMBACH-BRUNDAGE Y., GOLDBERG Y., DAS D., GANCHEV K., HALL K., PETROV S., ZHANG H., TÄCKSTRÖM O., BEDINI C., BERTOMEU N. & LEE C. J. (2013). Universal dependency annotation for multilingual parsing. In *Proceedings of ACL '13*, p. 92–97, Sofia, Bulgaria.

MCDONALD R., PETROV S. & HALL K. (2011). Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, p. 62–72, Stroudsburg, PA, USA.

NIVRE J. (2015). Towards a Universal Grammar for Natural Language Processing. In A. GELBUKH, Ed., *Computational Linguistics and Intelligent Text Processing*, volume 9041 of *Lecture Notes in Computer Science*, p. 3–16. Springer International Publishing.

OFLAZER K., SAY B., HAKKANI-TÜR D. Z. & TÜR G. (2003). Building a Turkish treebank. In A. ABEILLE, Ed., *Building and Exploiting Syntactically-annotated Corpora*. Kluwer Academic Publishers.

PETROV S., DAS D. & MCDONALD R. (2012). A Universal Part-of-Speech Tagset. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, p. 2089–2096.

SØGAARD A. (2011). *Data point selection for cross-language adaptation of dependency parsers*, In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*. Association for Computational Linguistics.

STENSON N. (1981). *Studies in Irish Syntax*. Tübingen: Gunter Narr Verlag.

THE CHRISTIAN BROTHERS (1988). *New Irish Grammar*. Dublin: C J Fallon.

Uí DHONNCHADHA E. (2009). *Part-of-Speech Tagging and Partial Parsing for Irish using Finite-State Transducers and Constraint Grammar*. PhD thesis, Dublin City University.

ZEMAN D. (2008). Reusable tagset conversion using tagset drivers. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08).*, p. 213–218, Marrakech, Morocco.