I've got a construction looks funny – representing and recovering non-standard constructions in UD

Josef Ruppenhofer
Leibniz-Institut für Deutsche Sprache
R5, 6-13
68161 Mannheim
ruppenhofer@ids-mannheim.de

Ines Rehbein
Data and Web Science Group
Universität Mannheim
68159 Mannheim
ines@informatik.uni-mannheim.de

Abstract

The UD framework defines guidelines for a crosslingual syntactic analysis in the framework of dependency grammar, with the aim of providing a consistent treatment across languages that not only supports multilingual NLP applications but also facilitates typological studies. Until now, the UD framework has mostly focussed on bilexical grammatical relations. In the paper, we propose to add a constructional perspective and discuss several examples of spoken-language constructions that occur in multiple languages and challenge the current use of basic and enhanced UD relations. The examples include cases where the surface relations are deceptive, and syntactic amalgams that either involve unconnected subtrees or structures with multiply-headed dependents.We argue that a unified treatment of constructions across languages will increase the consistency of the UD annotations and thus the quality of the treebanks for linguistic analysis.

1 Introduction

The Universal Dependencies (UD) initiative is a project that aims for crosslinguistically consistent annotation of morphosyntax (de Marneffe and Nivre, 2019). The sharing of representations across languages is intended to support multilingual NLP applications on the one hand and to facilitate the linguistic study of similarities and differences from a typological perspective.

While the UD framework also covers the annotation of parts of speech and of morphological features, the annotation of syntactic dependencies is at its core. UD currently uses 37 labels for broadly attested grammatical relations. In addition to these basic UD dependencies, UD allows for an enhanced representation that "aims to make implicit relations between content words more explicit by adding relations and augmenting relation names" (Schuster and Manning, 2016). The most prominent examples of added relations in enhanced representations are links that help propagate relations over conjunctions. While the basic dependency structure is assumed to form a (possibly non-projective) tree, enhanced UD representations often no longer are trees. The subtyping of relations serves to capture more fine-grained language-specific constructions (de Marneffe et al., 2014). For instance, in English, subjects of passive clauses bear the relation `nsubj:pass`.

In this paper, we discuss several constructions that are found in spoken language and data from social media, which are not trivial to analyze. The first set of constructions involves cases where a head and dependent are related in a hybrid way, exhibiting for instance properties of both coordination and subordination, or of coordination and predication. The second group of constructions involves syntactic amalgams where arguably the assumption that a basic UD analysis should form a tree is not met. We

Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020), pages 140–151

discuss possible trade-offs with respect to how basic and enhanced UD relations might be used in these cases to give linguistically adequate analyses. Note that, our unmarked examples come from Twitter. The others come from the ukWaC corpus (Baroni et al., 2009), UD treebanks, or the linguistic literature.

2 Constructions with hybrid properties

We first discuss constructions that superficially look as if a head and dependent are connected by a certain relation but where the construction also, or even mainly, exhibits properties of another type of relation.

2.1 Conditional coordinations

The first group of constructions involves superficial coordinations that are understood as conditionals, with the first conjunct being subordinated to the second conjunct (Culicover and Jackendoff, 1997; Culicover and Jackendoff, 2005). Some more common forms involve coordination of two declaratives (1), of an NP and a declarative (2), and of an imperative and a declarative (3). Constructions like this are found in German, French, Dutch and Russian (Fortuin and Boogaart, 2009), and probably in further languages.

(1) You say another word and I'll start hating you.

(2) One more word outta you and you're gettin kissed boy

(3) Say that again and imma fight you

In what follows, we focus on the simple declarative type in English, as illustrated in (1). As discussed by Culicover and Jackendoff (2005), while on the one hand the construction looks like syntactic coordination, it does not behave like it in every aspect but also has features of subordination. First, reversing the two conjuncts leads to a loss of the relevant meaning, which is somewhat unexpected for a simple coordination (4). Second, we cannot add coordinates with the intention of expanding the condition/protasis: (5) cannot be used to convey that 'If you drink another can of beer and Bill eats more pretzels, then I'm leaving'. At most, we can add coordinates if they are interpreted as part of the conclusion/apodosis (6).

(4) I'll start hating you and you say another word.

(5) # You say another word, Bill agrees with you and I'll start hating you.

(6) You say another word, I'll start hating you and Bill will stop talking to you.

Third, while regular coordination is compatible with gapping (7), conditional coordination is not (8). In this respect, it is similar to an explicit *if*-conditional, which also does not allow gapping (9).

(7) Big Louie stole another car radio and Little Louie the hubcaps. (from Culicover and Jackendoff (2005))

(8) #Big Louie steals one more car radio and Little Louie the hubcaps.

(9) #If Big Louie steals one more car radio, then Little Louie the hubcaps.

However, a treatment in terms of subordination is not clear-cut either. While English subordinate clauses, including *if*-clauses, can precede or follow the main clause, the first conjunct of a subordinating conditional coordination cannot be moved (10–11), whether we assign *and* to the left or the right conjunct.

(10) # And I'll start hating you, you say another word.

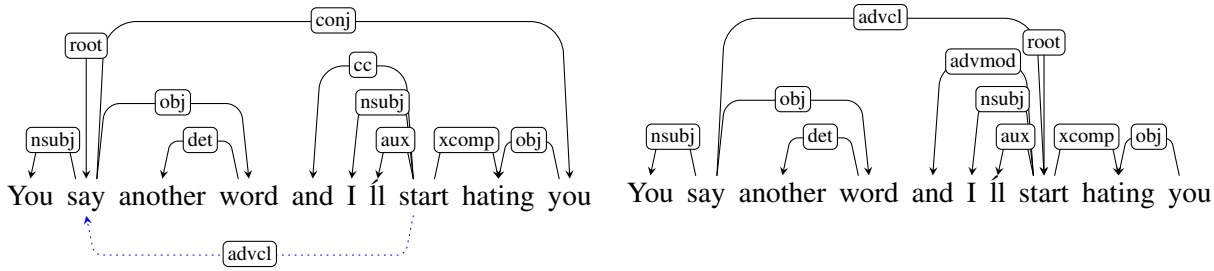
(11) # I'll start hating you, you say another word and.

Certainly, *and* would make a very odd subordinator: all other English subordinators are clause-initial. The fact that pauses in the conditional coordination construction precede, rather than follow *and* also suggests that it should be integrated with the right conjunct rather than the left, whatever the analysis. As a semantic peculiarity, we note that these conditionals are restricted to root conditional meanings. Speech act uses, for instances, are not possible unlike with *if*-conditionals (cf. 12–13).

(12) #You're hungry and there's pizza in the fridge. (constructed)

(13) If you're hungry, there's pizza in the fridge. (constructed)

Figure 1: Conditional coordination



(a) Using enhanced UD

(b) Basic UD only

What are the options for representing instances of this construction in UD? One option that preserves Culicover and Jackendoff's diagnosis that the construction involves a mismatch between syntax and semantics might be to proceed as follows. In terms of basic UD, we treat the structure as involving coordination while adding an `advcl` relation in the enhanced representation, as shown in Figure 1, to capture the subordination characteristics. However, as one of our reviewers argues, "the encoding of the fact that it's a different construction [from ordinary coordination] cannot be done at the enhanced level" and "[i]f you identify new constructions, you must introduce new labels to name them". We take this argument seriously but also want to point out that current UD annotations do not seem to differentiate fully all the constructions that one might want to recognize. For instance subject-auxiliary inversion constructions in English are not recoverable from specific relation labels. Likewise, instances of, for instance, the Group Identity NP construction such as *the rich/poor*, *the young/old*, etc. are not recoverable from specific labels. For instance, in example (14) below from UD-EWT (Silveira et al., 2014) *rich* is an adjective by POS and the NP-status of the subtree it heads is seen only from the fact that it is determined by *the* and governed as an `nmod` by *many*. But these relations also apply to NPs with regular nouns as heads.[1]

(14) . . . too many of **the rich** made their money not by succeeding in business, but . . .

Now, if we want to forego the use of enhanced relations but still point to the hybrid properties of the construction, we could introduce a new relation subtype `conj:advcl` for the relation between the conjuncts and still treat *and* as related to its head via `cc`. However, insofar as we understand current usage, subtype labels have not been used to indicate hybrids. E.g. while there is for instance a label `csubj:cop`, this label is used for a clause that acts as the subject of another, copular clause; it does not refer to something that is at the same time a `csubj` and a `cop`. To avoid the problem of how a subtyped relation name is likely to be interpreted, we could simply introduce a new relation name such as `conjcond`. But this relation may be sparse and hard to learn for statistical parsers. Yet another option is to treat this construction as involving subordination of the first conjunct as an `advcl` in the basic UD analysis, with *and* being tied to the main clause by a relation other than *cc*. In analogy to the treatment of *then* in *if-then* conditionals, *and* could be treated as an ADV that depends on the main clause head as an `advmod`. What is odd about this analysis is that it involves an obligatorily marked main clause and an obligatory unmarked subordinate clause. At least, *and* would not be the only conjunction that changes function in a pseudo-coordination construction. *But* similarly figures in constructions that are interpreted as (concessive) conditionals with readings such as 'despite P, Q' or 'even if P, Q', as shown by (15).

(15) Say what you want about this label , but with Delight they really have picked a winner (ukWaC)

And further, whatever analysis we choose also applies to conditional coordinations involving *or* (cf. (16)), which correspond to explicit *unless*-conditionals.

(16) Give me the money or I'll bite.

[1]Unlike UD, construction grammar and HPSG can represent the fact that certain phrases/subtrees have unexpected external semantics or syntax through features on phrasal mother signs. For constructional treatments of the Group Identity NP construction see Fried (2015) and Fillmore et al. (2012), where the construction is called Adjective-as-nominal.

For consistency's sake we also suggest extending the same treatment to cases of paratactic conditionals such as (17)–(18).

(17) You do that again, I'm gonna smack you where the sun don't shine.

(18) One more word outta you I'll take the servers down all day

Finally, we want to note that conditional coordinations are not the only 'weird' constructions involving *and*. Other pseudo-coordination constructions with a hortative-mandative semantics have verbs like *try* and *remember* or *be*+Adjective combinations such as *be sure* in the left conjunct (Flach, 2017).

(19) If you 're a woman , then try and find the cheapest policy - whoever it 's marketed at . (ukWaC)

2.2 German non-finite predication construction (NFPK)

The example in (20) is an instance of a German construction that, while apparently a coordination, semantico-pragmatically serves to express the speakers' incredulity about a hypothetical state of affairs mentioned in prior discourse. The German and related constructions in other languages are known under various names that reflect either their form, function or distribution, among them: Oxymoron construction (Deppermann, 2007), Bare predication construction in polar echoes (Huddleston and Pullum, 2002), Echo exclamation (Quirk et al., 1985) Mad Magazine sentences (Akmajian, 1984) and Incredulity Reponse Construction (Lambrecht, 1990). We will use the descriptive name Non finite predication construction (NFPK) proposed by Bücker (2012). Alongside the variant featuring *und* (NFPK$_{und}$), there is a version of the construction where *und* is lacking (NFPK$_{bare}$) (21). Functionally similar constructions exist in English and Spanish. They are structurally different, though also grammatically special (Etxepare and Grohmann, 2005). For instance, the English construction is similar to the German NFPK$_{bare}$ variant but the English construction involves an object-form subject (cf. the gloss of (20).) whereas the German construction has a subject in the nominative case typical of subjects even though there is no finite verb.

(20) Der und ein Vorbild , - lachhaft ! . . .
that_one and a role_model , - ridiculous !
'Him (be) a role model, . ridiculous! . . . '

(21) Frau Merkel ehrlich zu sich selbst sein ???
Mrs. Merkel honest to 3.REFL self be ???
'Mrs Merkel be honest with herself???'

Example (20) is a case of the NFPK construction that is also identified as such by the presence of an optional coda (above: *lachhaft* 'ridiculous') that explicitly expresses the speaker's disbelief/incredulity. Many instances of the NPFK construction with *und* also involve coordination of unalikes (cf. 22–23).

(22) Der und gewonnen ?
that_one and won ?
'He is supposed to have won?'

(23) Der und im Himmel ? Wohl eher ein paar Stockwerke tiefer .
that_one and in heaven ? Likely rather a couple floors lower .
'Him in heaven? More likely a few floors lower down.'

Assuming a predication relation between the two conjuncts is supported by the fact that reflexives in the right conjunct can be bound by governors in the left (24), unlike in regular coordinations (25).

(24) A: Der Sänger postete ein Bild von sich mit stylischer Brille . . . B: Der und ein Bild
the singer posted a picture of 3.REFL with with stylish glasses . . . B: that_one and a picture
von sich/#ihm ?
of 3.REFL/#him ?
A: The singer posted a picture of himself with stylish glasses . . . B: Him take a picture of himself/#him?

(25) A: Was hast du gesehen ? B: Queen Elisabeth und ein Bild von ihr/#sich. (constructed)
a: What have you seen ? B: Queen Elizabeth and a picture of her/#3.REFL
'A: What did you see? B: Queen Elizabeth and a picture of her/#herself.'

As in English, the predicate, even when involving a verbal phrase as in (22), cannot normally bear tense or modal markings.
However, note that with nominal right conjuncts, the right conjunct need not always be a predicate nominal as in (20). In (26), the two nominals in the conjuncts are not related as subject and predicate. Rather, both are to be understood as dependents of a zero predicate that is inferable from prior discourse.

(26) Der und ein klares ethisches Profil ?
that_one and a clear ethical profile ?
'Him have a clear ethical profile?'

The NFPK$_{und}$ variant presents a mismatch between form and semantics. An early analysis of this variant nevertheless treats it as a case of real coordination where the predicate (and sentence) that governs it is omitted (Behaghel, 1928). The two conjuncts are thus treated as metalinguistically mentioned rather than used, as in the English cases (27–28) where incongruity is explicitly stated or alluded to.

(27) The words Obama/Biden, and "scandal-free" should NEVER appear in the same sentence again. EVER.

(28) How often have you heard the terms ' entrepreneur ' and ' social conscience ' in the same sentence ? (ukWaC)

This analysis does not generalize well to the NFPK$_{bare}$ variant (21). A second type of ellipsis analysis assumes that NFPK$_{bare}$ features the elision of a finite verb form from what would otherwise be a regular finite sentence. This does not generalize to the NFPK$_{und}$ variant: adding a finite verb to instances of that variant does not produce grammatical sentences. Further, the NFPK$_{bare}$ version has constraints on the ordering of pronouns in particular that do not obtain in sentences where a finite verb has been inserted, which argues against NFPK being simply the result of finite verb elision (Bücker, 2012).

If we reject a simple ellipsis analysis, we have the same options as for the conditional coordination. Sticking close to the surface, we may propose annotating a coordination structure in the basic UD representation and making the subject-predicate relation, when present, explicit in the enhanced annotation, as shown in Figure 2a. Otherwise, we annotate an `orphan` relationship for cases like (26). A downside of this proposal is that we connect two words in the enhanced representation that are already related in the basic representation, with the roles of head and dependent being switched now. On the other hand, this would be in keeping with our enhanced UD analysis for the conditional coordination construction.

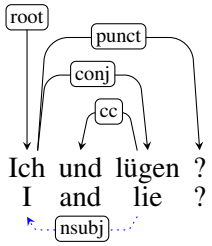
Avoiding the use of the enhanced representation, an alternative treatment directly uses the `nsubj` relation in the basic UD annotation for all cases where the second conjunct is understood as a predicate and the first conjunct as its subject (Fig. 2b). This would treat cases like (26) differently in the basic annotation since these cases would involve an `orphan` relationship. Further, the apparent conjunction, where present, would need to be changed both in terms of POS and syntactic dependency. We might, for instance, treat *und* somewhat arbitrarily as an ADV related by `advmod` to the head of the right conjunct.

2.3 Presentational relative clause construction (PRC)

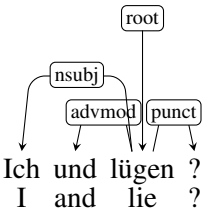
The presentational relative construction (PRC) involves the combination of a semantically weak main clause and a relative clause which, rather than the main clause, contains the assertion of the utterance (Lambrecht, 1988; Duffield et al., 2010). The PRC is, therefore, normally more aptly paraphrased by a single sentence (29a) than by a sequence of sentences (29b).

(29) You have some folks who deny she LOST when she suspended her campaign.
a. Some folks deny she LOST when she suspended her campaign.
b. #You have some folks. They deny she LOST when she suspended her campaign.

Figure 2: NFPK construction



(a) Subject and predicate, using enhanced UD



(b) Subject and predicate, basic UD only

Formally, PRCs mostly feature subject gaps in the relative clause.[2] By contrast, restrictive and non-restrictive relative clauses with a relative pronoun or a complementizer have no such restriction (30–31).

(30) OMG I want the car that shots from the headlights.

(31) My 68 month old has a toy truck that he rides up and down the street in.

(32) Whole family voting including my father who has never voted.

(33) In full survival mode, I forgot to water my lemon tree, which I love.

Restrictive relative clauses without a relative pronoun or complementizer do not allow subject gaps (34).

(34) #I chose the dress _ made me want to dance.

As argued by Lambrecht (1988), the PRC allows the speaker to avoid violating an information-packaging constraint that Lambrecht (1994) refers to as the Principle of Separation of Reference and Role (PSRR): "Do not introduce a referent and talk about it in the same clause". While the main clause serves to introduce the referent, the relative clause makes an assertion. In accord with the assertive pragmatic function of the relatives in PRCs, these relative clauses can be conjoined with assertive main clauses, unlike restrictive or appositive main clauses (cf. (35)).

(35) Once upon a time there was an old cockroach who lived in a paper bag and he was very poor. (from Lambrecht (1988))

The assertive status of the relative clause is underscored by the fact that its proposition can be challenged by the lie-test (36), which is not the case for restrictive or appositive relative clauses (37).

(36) A: have you tried simultaneous Twitter comments during lecture? I know colleagues who have done this.
B: That's a lie!
   a. Nobody's doing this.
   b. #You don't know colleagues.

(37) A: I'm looking for a gift for my son. He wants a car that he can build.
B: That's a lie.
   a. He doesn't want a car.
   b. # He can't build it.

Presentational relative clauses are also not the same as the relative clauses in cleft-sentences. Contrast the effect of the lie-test in response to a cleft (38) with that in response to the PRC (36)

(38) A: It's colleagues of mine who have done this.
B: That's a lie!

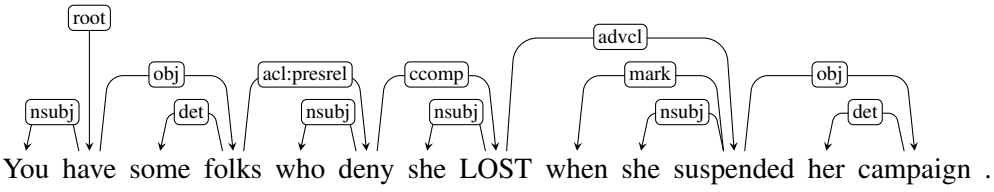
[2]Lambrecht (1988) assumes PRCs are restricted to involve subject gaps in the relative clause but Duffield et al. (2010) report instances with object gaps in the relative clause.

a. It's not them (but Pat who did it).

b. #Nobody has done this.

The Italian PoSTWITA treebank (Sanguinetti et al., 2018) contains an instance of what we call PRC that is simply annotated as a relative clause.[3] The treebank for spoken French (Lacheret et al., 2014) handles PRC instances the same way. But given their special properties, presentational relative clauses should be treated differently from restrictive and appositive relative clauses. A simple option involving only basic UD would be to only subtype the `acl`-relation further and introduce `acl:presrel`.

Figure 3: Presentational Relative Construction



The annotation strategy for the English PRC can also be carried over to Italian, French and German. Note that German presentational relative clauses, unlike restrictive and appositive relative clauses, exhibit the verb-second word order that is characteristic of matrix clauses (39).

(39) Ich hatte mal nen Freund, der hatte was im Auge und ging damit in die Notaufnahme.
I had some_time a friend, who had something in eye and went therewith in the ER.
'I had a friend once who got something into his eye and went to the ER about it.'

There likely are quite a few more similar constructions in other languages. We know of (i) complement clauses in Danish, Icelandic, Swedish, Norwegian and German as well as (ii) marked subordinate clauses in German that have unexpected main clause (verb-second) word order and which are said to have main clause properties (Günthner, 1996; Gärtner and Michaelis, 2010; Antomo and Steinbach, 2010; Reis, 2013; Wiklund, 2009; Bentzen, 2014). For French, there is discussion of constructions involving *reverse subordination* (Benzitoun, 2013).

3 Syntactic amalgams

We now turn to two constructions where arguably a basic UD analysis mirroring the analyses proposed in the theoretical literature from the construction grammar tradition would not result in simple trees and which for that reason need some special treatment.

3.1 Presentational amalgam construction (PAC)

Examples (40)–(41) exemplify what Lambrecht (1988) calls the presentational amalgam construction (PAC), which is common in spoken language but also found in social media. The construction consists of a sequence [[NP1 V NP2] [VP]], which apparently combines an existential *there*-clause or a clause with *have (got)* with a final VP. The postverbal NP in the initial presentational clause is usually indefinite.

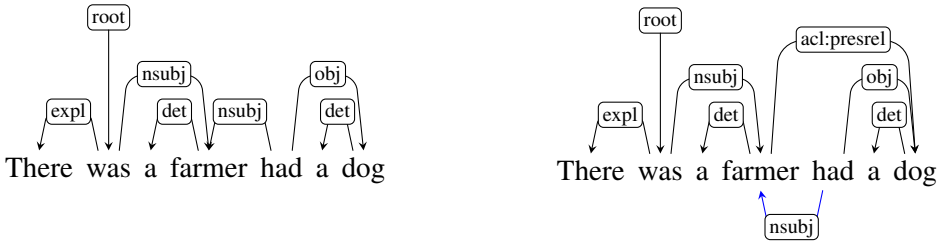
(40) There's a guy says he's gonna run a marathon every time Wigan win this season!

(41) I've got a friend of mine, does this all the time, especially during PE.

The instances of PAC thus are like instances of PRC (cf. section 2.3) but without a relative pronoun or complementizer. Like PRC, PAC allows only subject gaps. Also, instances of PAC behave like instances of PRC with respect to the lie-test (cf. §2.3): challenges do affect the final VP.

[3] Cruschina (2018) calls the instances of the construction presentational ci-sentences (PCS).

Figure 4: Presentational amalgam construction



(a) non-tree (b) tree

(42) A: there was a guy stole a yacht down south.
B: That's a lie. He only borrowed it.

On the analysis of Lambrecht (1988), PAC features an NP, namely NP2, that is simultaneously a dependent in the existential clause and the subject of a clause it forms with the final VP (cf. Figure 4a). Part of the motivation for this treatment is semantic. The presentational part of some instances such as (41), repeated below in (43), is by itself an odd proposition because of the redundancy.

(43) ??I've got a friend of mine.

But, as Lambrecht argues, NP2 is specifically produced with an eye towards its role as subject of the final VP. From that vantage point, an NP2 like that in (41) is not odd when considered within the PAC construction. This is so because the purpose of PAC is to solve the same communicative problem that the presentational relative construction addresses. PAC sentences are alternatives to simple sentences such as (44), which is a modified version of (40). Sentences like (44) are dispreferred in spoken language because they involve predicating something about a newly introduced referent in the same clause, while speakers normally like to first separately introduce a referent before predicating about it.

(44) A guy says he's gonna run a marathon every time Wigan win this season!

However, unlike with PRC, in the case of PAC there is full fusion between the presentational clause and the clause containing the main assertion. If we accept Lambrecht's analysis, then we have the undesirable situation that the same content word is a dependent of two different heads (Fig. (4a)). We could avoid this, by shifting the subject relation to the enhanced representation. The question is then how to connect the final VP to the initial existential clause. Our suggestion is to use a subtype of `acl` that we call `acl:presrelbare` to connect the VP to the head of NP2 (4b) . Because this subtype of `acl` always lacks a relativizer or pronoun, we cannot reuse the `acl:presrel` subtype introduced for PRC.

3.2 Double *is* construction (ISIS)

Consider examples (45–47). All three feature sequences where two instances of *is* immediately or very closely follow each other.

(45) He's certainly not as dominant in that role as he is, uh, is his, in his normal role.

(46) @RetroAperture @BeardedGenius what he is is stupid.

(47) But the thing is is that I'm naturally thin... (ISIS)

Example (45) involves a disfluent repetition where UD's `reparandum` relation would be used. Sentence (46) involves a specifying pseudo-cleft construction where a so-called fused relative (*what he is*) (Huddleston and Pullum, 2002) serves as the subject of the second instance of *is*, which takes the final phrase as `ccomp`. Sentence (47) is an example of the double *is* or ISIS construction.[4] As argued by Coppock et al. (2006), ISIS sentences do not involve disfluent repetition, although they might appear to do so. With ISIS, the copula *be* is typically repeated after certain nouns, such as *issue* and *point*, indicating that ISIS is a subtype of the specificational construction (Mikkelsen, 2005). Further, ISIS is often

[4] Another designation of the construction is copula doubling.

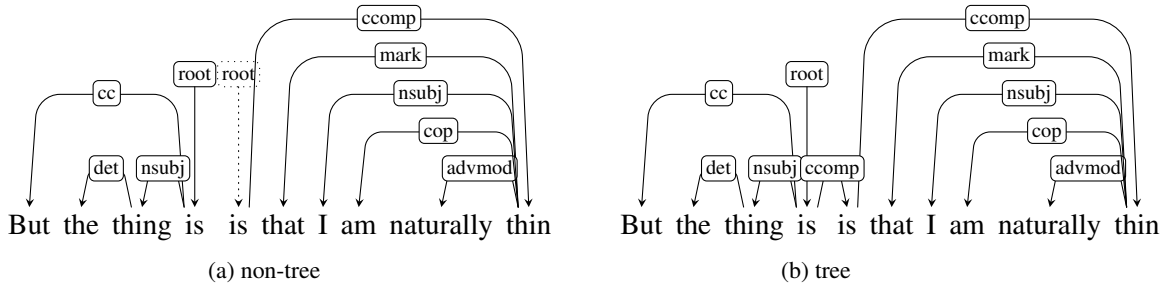
found before short, easy-to-process clauses where disfluencies are unexpected. And in fact human raters in Coppock et al. (2006)'s study found the ISIS instances involving declarative clauses highly fluent.

The formal construction grammar analysis of ISIS by Brenier and Michaelis (2005) treats this construction as a syntactic amalgam, a phrasal construction with two daughters: (i) the so-called setup clause including the subject NP and the first form of *be* and (ii) the second *is* and the following clause. This phrase-based analysis does not translate into a connected dependency tree (Figure 5a). To produce such a tree, we need to employ some sort of dependency relation to connect the two instances of the copular verb. This is not easy to decide, however, since common headedness criteria are inconclusive. Copulas do not normally depend on copulas. Both parts of the construction are non-optional. However, since semantically the second phrase seems to complement the first, we treat the head of the first as `root`.

(48) The real question was is are we getting a reasonable return on our investment. (from Brenier & Michaelis 2005)

The subject NP would be connected to it as an `nsubj`; the second copula would be connected to the first copular instance by a `ccomp` relation; and the finite clause in turn would be connected to the second copular instance by another `ccomp` (Figure 5b). This treatment would preserve the analysis of Brenier and Michaelis (2005) with the exception that the setup clause and the final *is*+clause are here connected by a dependency rather than both simply being daughters of the same phrase.

Figure 5: Double *is* construction



(a) non-tree

(b) tree

4 Conclusion

We have presented several constructions that are associated with social media or spoken language and that have special properties that do not readily fit the current use of basic and enhanced UD relations. The first set of constructions involved pseudo-coordinations whose surface structure is mismatched with their semantics. The second set of constructions featured cases of syntactic amalgams that either involved unconnected subtrees (double *is*) or structures with multiply-headed dependents (PAC).

While one can disagree both with the analyses of these constructions in the theoretical literature and with our ideas for dealing with them in UD, we are confident that as UD treebanks expand to more languages, domains, genres and registers, further constructions with similar challenges will be encountered. To allow for consistent and expressive analyses, we think the UD community would benefit from discussing which mechanisms to use for which kinds of constructions. We have explored the use of enhanced UD annotations, keeping the relation inventory the same, and as an alternative introducing new dependency relations and relation subtypes, which may lead to sparsely attested relation types. One option we have not explored but which the GUM corpus uses is a kind of constructional annotation (specifically, of sentence types) in the metadata (Zeldes and Simonson, 2016). This would, however, not localize which words are part of the construction. Other ways to track constructions may be conceivable.

Whatever the annotation mechanisms used, we think that adding linguistic analyses for such constructions to the UD guidelines might help to improve annotation consistency across languages, and thus the quality of the treebanks. Existing UD treebanks already feature some well-represented constructions that are treated inconsistently between treebanks and/or languages along the lines we discussed. For instance, the two clauses of the paratactic correlative construction (*the X-er, the Y-er*) are related by `conj` in German treebanks (HDT (Borges Völker et al., 2019), TüBa (Çöltekin et al., 2017), GSD (McDonald et al., 2013)), while English treebanks (EWT, GUM (Zeldes, 2017)) mostly use `advcl`.

References

Adrian Akmajian. 1984. Sentence types and the form-function fit. *Natural Language & Linguistic Theory*, 2(1):1–23.

Mailin Antomo and Markus Steinbach. 2010. Desintegration und Interpretation: Weil-V2-Sätze an der Schnittstelle zwischen Syntax, Semantik und Pragmatik. *Zeitschrift für Sprachwissenschaft*, 29(1):1–37.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226, September.

Otto Behaghel. 1928. *Deutsche Syntax: eine geschichtliche Darstellung*, volume 3. Carl Winters's Universitätsbuchhandlung, Heidelberg.

Kristine Bentzen. 2014. Embedded verb second (v2). *Nordic Atlas of Language Structures Journal*, 1(1).

Christophe Benzitoun. 2013. Faut-il remettre les pendules de la subordination temporelle à l'heure ? Description de deux fonctionnements de quand et avant que/de. *Cahiers Chronos*, 26:419–435.

Emanuel Borges Völker, Maximilian Wendt, Felix Hennig, and Arne Köhn. 2019. HDT-UD: A very large universal dependencies treebank for German. In *Proceedings of the Third Workshop on Universal Dependencies (UDW, SyntaxFest 2019)*, pages 46–57, Paris, France, August. Association for Computational Linguistics.

Jason M Brenier and Laura A Michaelis. 2005. Optimization via syntactic amalgam: Syntax-prosody mismatch and copula doubling. *Corpus Linguistics and Linguistic Theory*, 1(1):45–88.

Jörg Bücker. 2012. *Sprachhandeln und Sprachwissen [Linguistic Action and Linguistic Knowledge]*. De Gruyter, Berlin, Boston.

Çağrı Çöltekin, Ben Campbell, Erhard Hinrichs, and Heike Telljohann. 2017. Converting the TüBa-d/z treebank of German to universal dependencies. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 27–37, Gothenburg, Sweden, May. Association for Computational Linguistics.

Elizabeth Coppock, Jason Brenier, Laura Staum, and Laura Michaelis. 2006. Isis: It's not disfluent, but how do we know that? In *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society. Berkeley, CA*.

Silvio Cruschina. 2018. Setting the boundaries: Presentational ci-sentences in Italian. *Belgian Journal of Linguistics*, 32(1):53–85.

Peter Culicover and Ray Jackendoff. 1997. Semantic Subordination Despite Syntactic Coordination. *Linguistic Inquiry*, 28:195–217, 06.

Peter Culicover and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford linguistics. Oxford University Press.

Marie-Catherine de Marneffe and Joakim Nivre. 2019. Dependency grammar. *Annual Review of Linguistics*, 5(1):197–218.

Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4585–4592, Reykjavik, Iceland, May. European Language Resources Association (ELRA).

Arnulf Deppermann. 2007. *Grammatik und Semantik aus gesprächsanalytischer Sicht [Grammar and Semantics from a Conversation Analytic Perspective]*. De Gruyter, Berlin, Boston.

Cecily Jill Duffield, Jena D Hwang, and Laura A Michaelis. 2010. Identifying assertions in text and discourse: The presentational relative clause construction. In *Proceedings of the NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pages 17–24.

Ricardo Etxepare and Kleanthes K Grohmann. 2005. Towards a grammar of adult root infinitives. In *Proceedings of the 24th West Coast Conference on Formal Linguistics*, pages 129–137.

Charles Fillmore, Russell Lee-Goldman, and Russell Rhomieux. 2012. The framenet constructicon. In *Sign-Based Construction Grammar*.

Susanne Flach. 2017. Idiomatic singleton or prototype? a productivity analysis of be-adj-and-v. *Yearbook of the German Cognitive Linguistics Association*, 5(1):129 – 142.

Egbert Fortuin and Ronny Boogaart. 2009. Imperative as conditional: From constructional to compositional semantics. *Cognitive Linguistics*, 20, 01.

Mirjam Fried. 2015. Construction grammar. In T. Kiss and A. Alexiadou, editors, *Syntax - Theory and Analysis. Volume 2*, Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK), pages 974–1003. De Gruyter.

Hans-Martin Gärtner and Jens Michaelis. 2010. On modeling the distribution of declarative V2-clauses: The case of disjunction. *Judgements and propositions*, pages 11–25.

Susanne Günthner. 1996. From subordination to coordination? Verb-second position in German causal and concessive constructions. *Pragmatics*, 6(3):323–356.

Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press.

Anne Lacheret, Sylvain Kahane, Julie Beliao, Anne Dister, Kim Gerdes, Jean-Philippe Goldman, Nicolas Obin, Paola Pietrandrea, and Atanas Tchobanov. 2014. Rhapsodie: a prosodic-syntactic treebank for spoken french. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*, pages 295–301. European Language Resources Association (ELRA).

Knud Lambrecht. 1988. There was a farmer had a dog: Syntactic amalgams revisited. In *Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 319–339.

Knud Lambrecht. 1990. "What, me worry?"–'Mad Magazine Sentences' Revisited. In *Annual Meeting of the Berkeley Linguistics Society*, volume 16, pages 215–228.

Knud Lambrecht. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge Studies in Linguistics. Cambridge University Press.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria, August. Association for Computational Linguistics.

Line Mikkelsen. 2005. *Copular clauses: Specification, predication and equation*, volume 85. John Benjamins Publishing.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman, London.

Marga Reis. 2013. „Weil-V2 "-Sätze und (k) ein Ende? Anmerkungen zur Analyse von Antomo & Steinbach (2010). *Zeitschrift für Sprachwissenschaft*, 32(2):221–262.

Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, Oronzo Antonelli, and Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. In Nicoletta Calzolari (Conference chair), Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In *Proceedings of the 10th Language Resource and Evalutation Conference (LREC 2016)*, pages 2371–2378. ELRA.

Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.

Anna-Lena Wiklund. 2009. In search of the force of dependent V2: A note on Swedish. *Working papers in Scandinavian syntax*, 83:27–36.

Amir Zeldes and Dan Simonson. 2016. Different flavors of GUM: Evaluating genre and sentence type effects on multilayer corpus annotation quality. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*, pages 68–78, Berlin, Germany, August. Association for Computational Linguistics.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.