

RESEARCH

The status of function words in dependency grammar: A critique of Universal Dependencies (UD)

Timothy Osborne¹ and Kim Gerdes²¹ Department of International Studies, Zijinggang Campus, Building East 5, Zhejiang University, CN² l'Institute de Linguistique et Phonétique Générales et Appliquées (IPGA) de la Sorbonne Nouvelle, Paris 3, FRCorresponding author: Timothy Osborne (tjo3ya@yahoo.com)

The article examines the Universal Dependencies (UD) annotation scheme. The UD project is an international initiative to produce treebanks of the world's languages, whereby the treebanks have been annotated in a cross-linguistically consistent manner. A central aspect of the UD annotation scheme is its analysis of function words. The scheme advocates subordinating function words to content words. This article discusses linguistic and practical motivations behind the UD decision to subordinate function words to content words. It demonstrates that UD choices in this area are not supported linguistically. At the same time, the near convertibility of the UD treebanks to a more linguistically motivated annotation format means that the UD initiative remains of great value to linguistics in general.

Keywords: annotation scheme; dependency grammar; function words; Universal Dependencies; treebank corpus

1 Introduction

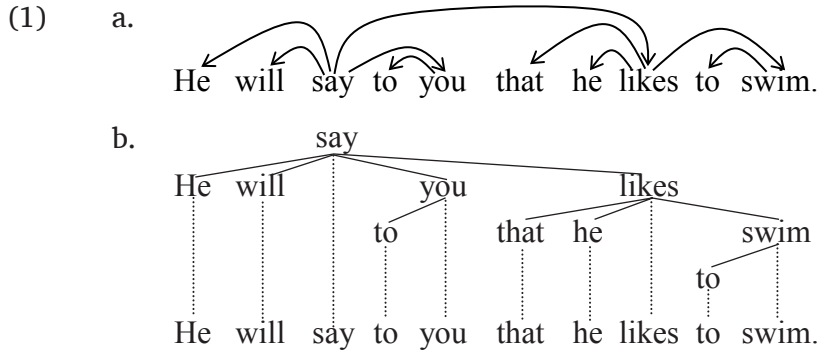
The *Universal Dependencies* (UD) project is a large-scale effort involving many dozens of researchers internationally to produce consistently annotated treebanks of the world's languages (UD webpage: <http://universaldependencies.org/>).¹ The consistency of annotation occurs in the form of adherence to the same one annotation scheme. One of the stated goals of the UD project is to promote the typological study of natural language syntax and grammar. The idea is in part that given the newly created, consistently annotated, and easily accessible treebanks, it should be possible to investigate natural language syntax and grammar with ease on an unprecedented scale. While the goal of promoting the typological study of natural language syntax is a worthy one, there is a problem with the UD annotation scheme. This article gives a critical account of the sentence structures being created by the UD annotation scheme and suggests solutions to the problems. At the same time, it emphasizes that the potential for automated conversion of the UD corpora to an annotation format that is linguistically well-motivated means that the UD project is of great value to linguistics in general.

The UD scheme is a type of dependency syntax (also called *dependency grammar*, DG) – DG in general is an approach to the syntax of natural languages that is prominent in corpus linguistics and favored in the field of natural language processing (NLP). UD is hence a set of guidelines for dependency annotation. It posits approximately a dozen parts of speech (POS) and two dozen grammatical functions. The aspect of the UD scheme addressed here is, however, not the inventories of the POS and grammatical functions

¹ The statements produced here about the nature of the UD project and its annotation scheme are based mainly on the information provided in the UD webpage: <http://universaldependencies.org/>. When citing the webpage, we give the URL of the relevant page and at times we point to a specific example, e.g. ex. 26.

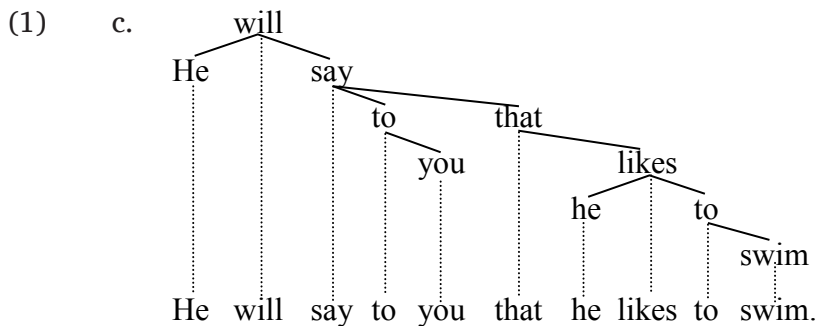
it posits, but rather its analysis of function words. UD advocates subordinating function words to content words, contrary to most other DGs and most other frameworks of syntax in general.²

The hierarchical analysis of function words that UD advocates is illustrated with the following sentence:



The arced arrows employed in (1a) are the convention to show dependencies preferred by UD researchers and by corpus linguists in general. The analysis in (1b) is equivalent to (1a), the only difference being that (1b) employs a tree-drawing convention to show dependencies. This article employs trees like the one in (1b) throughout because trees are transparent, showing the hierarchy of words clearly. Observe that contrary to much work in the DG tradition, examples (1a–b) do not provide information about the grammatical functions involved. This information could be included in the diagrams, but since it is not directly relevant to most of the points about the UD approach we make below, it can be ignored for the most part.

The noteworthy aspect of the hierarchical analysis in (1) concerns the function words. The modal auxiliary verb *will* is subordinate to the content verb *say*; the preposition *to* is subordinate to the content pronoun *you*; the subordinator *that* is subordinate to the content verb *likes*; and the particle *to* is subordinate to the content verb *swim*. The decision to subordinate function words to content words is, as stated above, contrary to the DG tradition. Most DGs would advocate a surface hierarchical analysis of the sentence more along the following lines (e.g. Kunze 1975; Hudson 1987; Schubert 1987; Starosta 1988; Engel 1994; Jung 1995; Heringer 1996; Bröker 1999; Groß 1999; Eroms 2000; Gerdes and Kahane 2007; Mel’čuk 2009; Osborne et al. 2012; Mille et al. 2013; Lacheret et al. 2014; Osborne and Groß 2017):³



² Determiners are the one major exception to this statement. UD subordinates determiners to nouns, which is consistent with most works in the DG tradition.

³ There are significant differences among the DGs listed. Concerning the main issues discussed in this article, however, they all position the finite verb as the root of the sentence, subordinate nouns to adpositions, and subordinate verbs to subordinators.

This analysis subordinates the content verb *say* to the auxiliary *will*, the content pronoun *you* to the preposition *to*, the content verb *likes* to the subordinator *that*, and the content verb *swim* to the particle *to*. This more traditional DG hierarchy is motivated by numerous facts of syntax. The UD analysis in (1a–b) is, in contrast, a mixture of semantic and syntactic motivations and as such, it is not well-motivated by linguistic reasoning.

This article argues extensively that when researching natural language syntax, DG analyses like the one in (1c) should be preferred over UD analyses like the one in (1a–b). This message does not, however, preclude the possibility that UD annotation might be more appropriate for certain downstream applications, such as relation extraction, reading comprehension, and machine translation (Desideratum 6 of the UD goals, see below). More importantly, the fact that one can, with little human intervention, automatically convert the current UD corpora to an annotation format that is in line with more traditional assumptions about sentence structure means that the UD treebanks are of great value – we have performed this conversion ourselves, as discussed below in Section 4.4.

2 Background on UD

The next two sections provide some background information on UD goals and guidelines and the consequences of these for the analysis of function words.

2.1 UD desiderata

UD is, as stated above, a project that proposes a “universal” annotation scheme and as such, this scheme should be applicable to all languages. To date the scheme has served as guidance for the creation of treebanks for more than 70 languages (<http://universaldependencies.org/>). The stated goal of the overall project is expressed as follows:

“Universal Dependencies (UD) is a project that is developing cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective. The annotation scheme is based on an evolution of (universal) Stanford dependencies (de Marneffe et al. 2006; 2008; 2014), Google universal part-of-speech tags (Petrov et al. 2012), and the Intersect interlingua for morphosyntactic tagsets (Zeman 2008). The general philosophy is to provide a universal inventory of categories and guidelines to facilitate consistent annotation of similar constructions across languages, while allowing language-specific extensions when necessary.”

(<http://universaldependencies.org/introduction.html>)

The desiderata of the current version of UD (version 2.0) were first presented in December 2016. These desiderata are listed next:

1. UD needs to be satisfactory on linguistic analysis grounds for individual languages. (Desideratum 1)
2. UD needs to be good for linguistic typology, i.e. providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families. (Desideratum 2)
3. UD must be suitable for rapid, consistent annotation by a human annotator. (Desideratum 3)
4. UD must be suitable for computer parsing with high accuracy. (Desideratum 4)
5. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. We

refer to this as seeking a *habitable* design, and it leads us to favor traditional grammar notions and terminology. (Desideratum 5)

6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...). (Desideratum 6)

(<http://universaldependencies.org/introduction.html>)

These goals can lead to contradictory conclusions, as shown by Gerdes and Kahane (2016) and as discussed at length below.⁴ To refer to these goals efficiently, the designations in parentheses are used (e.g. Desideratum 1, Desideratum 2, etc.).

The organization of this article follows to an extent the outline provided by these desiderata. Section 3 examines linguistic aspects of UD annotation (Desideratum 1). Section 4 considers the value of UD annotation for language typology (Desideratum 2). Section 5 considers further areas: the extent to which UD is easy for the human annotator (Section 5.1, Desideratum 3), the impact of UD annotation on parser accuracy (Section 5.2, Desideratum 4), the extent to which UD annotation is learner friendly (Section 5.3, Desideratum 5), and the extent to which UD annotation is good for downstream applications (Section 5.4, Desideratum 6). Section 6 concludes the article.

When all is said and done, only Desideratum 6 remains as a potential source of support for the current UD annotation scheme. We leave the door open in this area that UD annotation may be more suitable for the areas mentioned in Desideratum 6 (relation extraction, reading comprehension, machine translation, etc.).

2.2 Function words

The/a controversial aspect of the current UD annotation scheme concerns its analysis of function words, as stated above. UD annotation subordinates many (not all!) function words to content words, as illustrated above with examples (1a–b). The motivation for doing this is given in the following passage:

“Preferring content words as heads maximizes parallelism between languages because content words vary less than function words between languages. In particular, one commonly finds the same grammatical relation being expressed by morphology in some languages or constructions and by function words in other languages or constructions, while some languages may not mark the information at all (such as not marking tense or definiteness).”

(<http://universaldependencies.org/u/overview/syntax.html>)

The desire to subordinate function words to content words imposes a binary classification on all words; a given word is classified either as a function word or a content word. This is problematic, since the distinction between function and content word is not black and white. The distinction is, rather, more accurately captured in terms of a continuum, whereby prototypical function words and content words appear at opposite ends of the continuum, non-prototypical cases appearing somewhere on the continuum in-between.

The UD choice to subordinate function words to content words results in a number of concrete decisions about the hierarchical status of the parts of speech. In particular, the following choices guide UD analyses:

1. Auxiliary verbs, including the copula, are subordinated to content verbs (or in the absence of a content verb, to another contentful predicative expression).

⁴ Gerdes and Kahane (2016) propose three finer-grained sets of annotation choices: conception-oriented considerations (including Desiderata 1 and 2 of the UD goals), annotator-oriented considerations (Desideratum 3 above), and end-user-oriented considerations (Desiderata 4, 5, and 6).

2. Adpositions (prepositions, postpositions) are subordinated to the nouns with which they co-occur.
3. Subordinators (subordinate conjunctions) are subordinated to the verbs with which they co-occur.
4. Many other function words are subordinated to content words (e.g. in English the particle *to* introducing an infinitive, *than* and *as* of comparatives, the coordinators *and*, *or* and *but*, etc.)

These choices result in structures that stand in stark contrast to most work in theoretical syntax over the past few decades. Most phrase structure grammars – e.g. HPSG (Pollard and Sag 1994), Lexical Functional Grammar (Bresnan 2001), Government and Binding (Chomsky 1981; 1986), Minimalist Program (Chomsky 1995) – as well as most DGs – Lexicase (Starosta 1988), Word Grammar (Hudson 1984; 1990; 2007; 2010), Meaning Text Theory (Mel'čuk 1988; 2003; 2009), the German schools (Kunze 1975; Schubert 1987; Engel 1994; Heringer 1996; Bröker 1999; Groß 1999; Eroms 2000) – assume that most function words are heads over content words.

The fact that most works in the DG tradition and in theoretical syntax in general contradict UD choices concerning the analysis of function words is not in itself an argument against UD. The linguistic reasoning behind the developments in theoretical syntax do, however, bear directly on UD choices – since as just established in the previous section (Section 2.1), UD annotation strives for representations that are linguistically motivated (Desideratum 1).

The authors of UD are aware that the UD analysis of function words breaks with the DG tradition. They write:

“We are aware that the choice to treat function words formally as dependents of content words is at odds with **many** [our emphasis] versions of dependency grammar, which prefer the opposite relation for many syntactic constructions.”

(<http://universaldependencies.org/u/overview/syntax.html>)

Osborne and Maxwell (2015) address this issue directly. They survey the DG tradition concerning the hierarchical status of function words, and their survey reveals that most prominent DG works over the decades have positioned auxiliary verbs above content verbs,⁵ and with the exception of Kern (1882; 1883; 1884; 1886; 1888) and Tesnière (1959/2015), who positioned a given function word together with a content word in the same one node/nucleus, all DGs since Tesnière and before UD (that we are aware of) have acknowledged the status of adpositions as heads over their nouns and subordinators as heads over their verbs.

3 Linguistic analysis

UD's first desideratum appeals to linguistic validity for individual languages; it is repeated here from Section 2.1:

Desideratum 1

UD needs to be satisfactory on linguistic analysis grounds for individual languages.

⁵ Kern (1883; 1884) and Tesnière (1959) positioned auxiliary verbs and adpositions together with a content word in a single nucleus/node, and in this regard, their analyses neither support nor refute the UD decision to subordinate function words to content words. The only noteworthy DG linguists considered in Osborne and Maxwell's survey who offer analyses that support UD are Hays (1964: 521), who provided an example that subordinates an auxiliary to a content verb, and Matthews (1981: 63), who did the same more consistently. Neither of these two linguists advocated subordinating adpositions to their nouns and subordinators to their verbs, however.

The next sections establish that current UD annotation choices are not satisfactory on linguistic analysis grounds because they result from a mixture of semantic and syntactic criteria. This mixture produces structures that are neither semantically nor syntactically satisfactory. The alternative we advocate for is consistent with most of the DG tradition in elevating syntactic motivations over semantic ones.

3.1 *Semantics over syntax*

The UD desire to subordinate function words to content words is a semantic motivation, for the distinction between function and content word is semantic in nature. This emphasis on semantics renders UD parses incapable of serving as a basis for addressing phenomena of syntax. The next sections briefly consider some such phenomena that cannot be addressed in a coherent manner due to UD's decision to take a semantic criterion as its guiding principle for annotation.⁶

3.1.1 Subcategorization (auxiliary verbs)

UD annotation is contrary to the general understanding of subcategorization. Subcategorization is assumed to operate down the syntactic hierarchy, that is, heads subcategorize for their dependents. UD annotation would at times, however, have subcategorization pointing up the hierarchy, that is, some dependents would subcategorize for their heads. Our understanding of subcategorization is expressed concretely as follows:

Subcategorization

Given the co-occurrence of two words W1 and W2, W1 subcategorizes for W2 if W1 requires that W2 appear as a specific category or subcategory (or as a particular lexical item).⁷

This relationship is asymmetrical, i.e. the appearance of W1 requires that W2 appear as a specific category or subcategory, but not vice versa (cf. Järventausta 2003: 785). Expressing this idea as generally as possible, subcategorization is the notion that *the general requires the specific*. Hence if W1 subcategorizes for W2, the form of W1 is more flexible than that of W2.

The following data demonstrate that a function verb, in this case the auxiliary of perfect aspect, is more flexible in its form than the co-occurring content verb:

- (2)
- a. Sam has/had eaten.
 - b. They have/had eaten.
 - c. They will have eaten.
 - d. Having eaten, they were content.
 - e. *Sam has eats/ate/eat/eating.

When combined with the participle *eaten* to express perfect aspect, the form of the auxiliary *have* is flexible, for it can appear in its varying finite forms (*has/had/have*), in its infinitive form (*have*), or in its progressive form (*having*). In contrast, the form of the content verb *eaten* is fixed, as the failed attempts to vary its form in (2e) illustrate.

⁶ If space allowed, additional problems could be discussed, such as UD's inability to model constituent structure and its inability to produce a coherent account of coordinate structures.

⁷ Our characterization of subcategorization here unifies two notions for identifying heads discussed by Zwicky (1985), namely *subcategorization* and *government*. Many accounts of subcategorization, for instance the ones that one finds in dictionaries of linguistic terms, remain general, considering mainly just the combinatory potential of content verbs. Our definition here provides a basis for characterizing head-dependent combinations in terms of subcategorization more generally.

The point at issue concerns the direction in which subcategorization operates. Consider the competing structural analyses:

- (3) a. – UD analysis
 b. – Purely syntactic analysis
- The UD analysis tree (a) shows 'eaten' as the root node with children 'Sam', 'has', and 'eaten'. Dotted lines connect 'Sam' to 'Sam', 'has' to 'has', and 'eaten' to 'eaten' in the sentence 'Sam has eaten.'. The purely syntactic analysis tree (b) shows 'has' as the root node with children 'Sam' and 'eaten'. Dotted lines connect 'Sam' to 'Sam', 'has' to 'has', and 'eaten' to 'eaten' in the sentence 'Sam has eaten.'.

The UD analysis sees subcategorization pointing up the hierarchy; the dependent function verb *has* subcategorizes for its head content verb *eaten*. The purely syntactic analysis, in contrast, is more plausible because it has subcategorization pointing down the hierarchy, from the head function verb *has* to the dependent content verb *eaten*.

3.1.2 Subcategorization (adpositions and particles)

Subcategorization is also a criterion that helps reveal the hierarchical status of adpositions. Prepositional phrasal verbs provide good examples of this point (cf. Hudson 1987: 120). There are numerous idiosyncratic verb-adposition combinations in natural languages, whereby the given adposition that co-occurs with the verb at hand is fixed, that is, it cannot appear with just any verb, but rather the meaning that the two convey together is idiosyncratic and hence non-compositional to a greater or lesser degree (e.g. English: *abscond with*, *focus on*, *get into*, *laugh at*, *pass for*, *pick on*, *stare at*, *take after*, *rely on*, *wait for*; German: *arbeiten an*, *denken an*, *freuen auf*, *halten für*, *kämpfen um*, *rechnen mit*, *stehen auf*, *warten auf*; French: *compter sur*, *dépendre de*, *prendre pour*, *se décider entre*, *considérer comme*).

The UD analysis of prepositions results in a situation that has the preposition hierarchically once removed from the verb. The diagrams on the left are those of UD, and those on the right, of the purely syntactic analysis:

- (4) a. b.
- (5) a. b.
- (6) a. b.
- The UD analysis trees (a) show the verb as the root node. In (4a), 'laughed' is the root with children 'He' and 'it'. 'at' is a child of 'it'. Dotted lines connect 'He' to 'He', 'laughed' to 'laughed', 'at' to 'at', and 'it' to 'it' in the sentence 'He laughed at it.'. In (5a), 'picked' is the root with children 'She' and 'us'. 'on' is a child of 'us'. Dotted lines connect 'She' to 'She', 'picked' to 'picked', 'on' to 'on', and 'us' to 'us' in the sentence 'She picked on us.'. In (6a), 'takes' is the root with children 'He' and 'her'. 'after' is a child of 'her'. Dotted lines connect 'He' to 'He', 'takes' to 'takes', 'after' to 'after', and 'her' to 'her' in the sentence 'He takes after her.'. The purely syntactic analysis trees (b) show the preposition as a child of the verb. In (4b), 'laughed' is the root with children 'He' and 'at'. 'it' is a child of 'at'. Dotted lines connect 'He' to 'He', 'laughed' to 'laughed', 'at' to 'at', and 'it' to 'it' in the sentence 'He laughed at it.'. In (5b), 'picked' is the root with children 'She' and 'on'. 'us' is a child of 'on'. Dotted lines connect 'She' to 'She', 'picked' to 'picked', 'on' to 'on', and 'us' to 'us' in the sentence 'She picked on us.'. In (6b), 'takes' is the root with children 'He' and 'after'. 'her' is a child of 'after'. Dotted lines connect 'He' to 'He', 'takes' to 'takes', 'after' to 'after', and 'her' to 'her' in the sentence 'He takes after her.'.

The problem with the UD account of these verb-preposition combinations should be apparent, for the verb and preposition are not directly linked to each other in the a-trees. The UD analysis therefore implausibly implies that the verb can somehow subcategorize

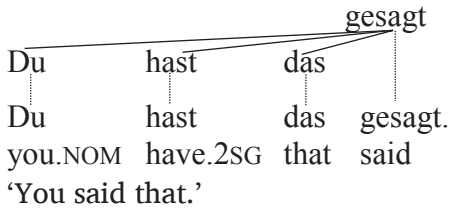
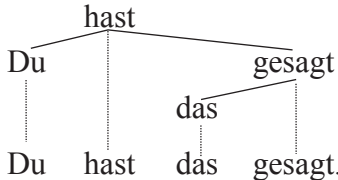
for a certain preposition that is its grandchild, rather than its child. The purely syntactic analysis is not faced with this difficulty, since it has the two words that constitute the idiosyncratic combination linked to each other directly.

Turning now to particles of comparison, the same reasoning from subcategorization also extends to these particles. Numerous verbs subcategorize for a particle of comparison (e.g. English: *serve as*, *take as*, *view as*, *feel like*, *taste like*; German: *ansehen als*, *arbeiten als*, *dienen als*, *schmecken wie*; French: *servir de*, *paraître comme*, *utiliser comme*, *employer comme*, *sentir comme*, *prendre pour*). On a purely syntactic analysis of these data, the two words that need each other, i.e. the verb and the particle, are linked directly to each other, whereas there is no such direct link between them on the UD analysis.

3.1.3 Nominative case assignment

In many languages, the finite verb enjoys a special relationship with the subject (cf. Järventausta 2003: 784, 790–1). The salient property is the correlation of the nominative case (marking the subject) with the presence of the finite verb. This relationship is the/a main motivation for the assumption of IP/TP (inflection phrase/tense phrase) in Chomskyan grammars, and it is also the/a main trait of finite verbs and subjects that has motivated many DGs to position the finite verb as the root of the sentence and the subject as an immediate dependent of the finite verb (e.g. Kern 1883: 1; 1884: Prologue; Hudson 1984: 91; Schubert 1987: 90–96; Starosta 1988: 239–240; Engel 1994: 107–109; Jung 1995: 62–63; Heringer 1996: 82–84; Eroms 2000: 129–133; Mel’čuk 2009: 44–45, 79–80).

Consider the following sentence of German with respect to the assignment of nominative case:

- (7) a.  – UD analysis
 Du hast das gesagt.
 you.NOM have.2SG that said
 ‘You said that.’
- b.  – Purely syntactic analysis
 Du hast das gesagt.
 Du hast das gesagt.

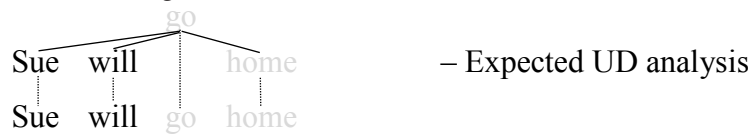
The appearance of the nominative-marked pronoun *Du* is reliant on the presence of the finite auxiliary verb *hast*. In the absence of the finite verb, the nominative-marked subject cannot appear, e.g. **Du das gesagt*, lit. ‘You.NOM that said’. The UD structure in (7a) does not accommodate this reliance, whereas the purely syntactic analysis in (7b) does. The analysis in (7b) expresses the relationship by subordinating the subject directly to the finite verb. This direct dependency between the two accommodates two aspects of nominative case assignment: the fact that it is the finite verb *hast* that is assigning nominative case and the fact that the nonfinite verb *gesagt* is NOT assigning nominative case.

3.1.4 VP-ellipsis

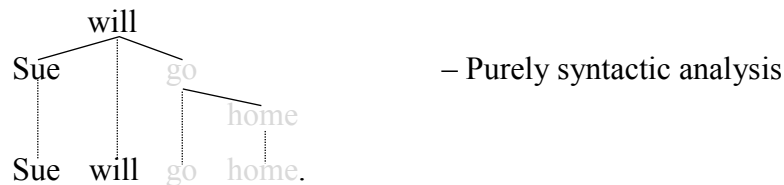
The nature of VP-ellipsis in those languages that have it, like English, is easily accommodated if the auxiliary verb is head over the content verb. If the opposite is assumed as with the UD scheme, the words that survive VP-ellipsis end up disconnected (cf. Hudson 1987: 118; Groß and Osborne 2015: 115), e.g.

(8) cf. <http://universaldependencies.org/u/overview/specific-syntax.html>, ex. 7–8

a. Fred won't go home, but



b.



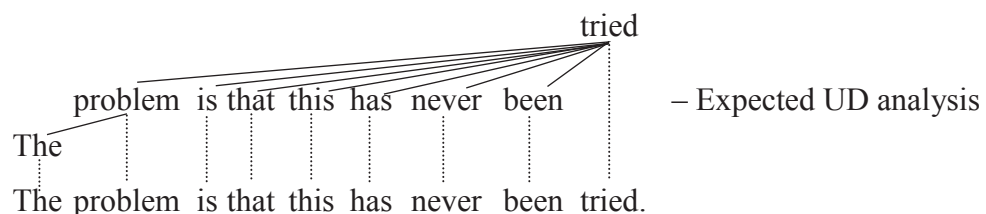
The elided string *go home* is a complete subtree on the purely syntactic analysis, whereas on the analysis that one would expect from UD, the elided words *go home* do not form a complete subtree. For the UD account, the situation therefore seems to result in disconnected words, since the surviving words, here *Sue* and *will*, are not linked together by a dependency. The expectation, then, might be that the UD scheme posits the presence of empty nodes for the elided material in order to maintain a tree-based analysis of sentence structure.

Interestingly, the basic dependencies of UD do not assume the presence of empty nodes to accommodate VP-ellipsis (or any other phenomenon of syntax). UD opts instead to promote the auxiliary to root status; it calls this promotion “promotion by head elision”. In other words, UD in fact assumes the analysis shown in (8b) (but with no empty material present), not the one in (8a) (<http://universaldependencies.org/u/overview/specific-syntax.html>, ex. 7–8). The solution is *ad hoc*; it reveals the difficulties that the UD scheme has in accommodating VP-ellipsis, a frequent occurrence in English.⁸

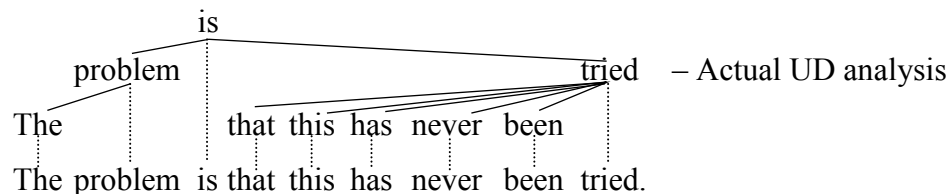
3.1.5 Predicative clauses

UD's principle of positioning content words over function words can result in a content word of a subordinate clause serving as the root of the entire sentence. This occurs with predicative clauses, e.g.

(9) a. (cf. <http://universaldependencies.org/u/overview/complex-syntax.html>, ex. 21)



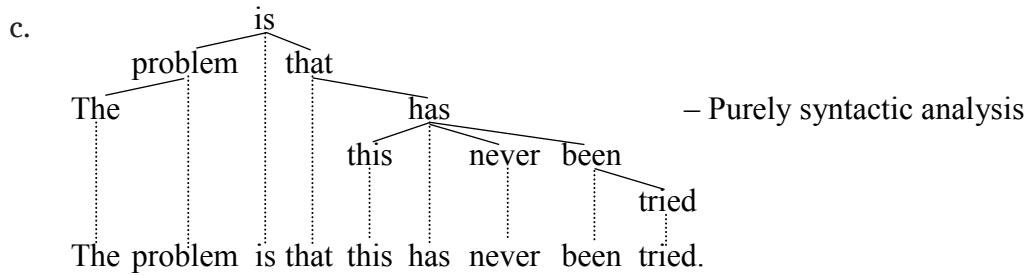
b.



⁸ Note that a similar type of *ad hoc* solution is also assumed with certain cases of coordination: For *He gave a book to Mary and a disk to Ann*, an *orphan* link is proposed between *disk* and *Ann* – an analysis that could be applied to (8), too:

(i) *Sue* –*orphan*→ *will*

This shows nonetheless that UD acknowledges the pivotal status of auxiliaries in these rare cases of ellipsis.



UD reasoning dictates that the copula *is* must be subordinated to the predicative expression that follows it. Similarly, the subordinator *that* must be subordinated to the content verb that follows it. The result of these choices is a situation in which the content verb in the subordinate clause becomes the root of the entire sentence, as shown with *tried* in (9a).

This result is quite implausible as there are now two competing subjects, *problem* and *this*, a fact that the authors of UD have realized, since they actually reject the analysis in (9a) and adopt the one in (9b). Ironically, they thus elevate the copula *is* to root status to overcome the problem, which means that their analysis overlaps to an extent with the purely syntactic analysis in (9c). The solution they adopt is again *ad hoc*.

3.2 Syntax over semantics

In response to the points just discussed (and to other phenomena of syntax that are problematic for UD annotation choices – see footnote 6), the proponents of current UD annotation might counter that the ability to address phenomena of syntax is not one of UD's stated goals. What is important, rather, is that UD annotation result in sentence parses that are consistent insofar as semantically loaded words (content words) appear over semantically impoverished words (function words). The problem in this area is that the actual annotation choices are expressed in terms of syntactic category (see the UD choices 1–4 in Section 2.2). There is hence a basic contradiction in UD annotation: positioning content words over function words is a semantic criterion, but the actual annotation choices are expressed in terms of syntactic category, a syntactic criterion. The next sections consider two areas where the contradiction is apparent.

3.2.1 Semi-auxiliaries

When semi-auxiliaries (e.g. *is going to*, *have to*, *ought to*, *used to*) appear, UD annotation positions semantically impoverished words as heads rather than as dependents. From a semantic perspective, semi-auxiliaries are more like pure auxiliary verbs than content verbs because they have little semantic content, but from a syntactic perspective, they are more like content verbs than auxiliary verbs because their distribution is that of content verbs. To illustrate this point, take the semi-auxiliary verb *used to* as an example. Unlike true auxiliary verbs such as *do*, the semi-auxiliary *used to* does not license subject-auxiliary inversion: **Used he to smoke? vs. Did he used to smoke?*

The UD binarity of categorization forces a decision in this regard. It elevates syntax over semantics because it positions semi-auxiliaries as heads over content verbs, and in so doing, it contradicts its own principle of categorization in terms of semantic content. To provide an example, the UD scheme takes the English modal auxiliary *will* to be a function verb and thus subordinates it to the full verb with which it co-occurs. In contrast, the analysis of the semantically similar near-future semi-auxiliary *is going to*, is such that *going* is positioned as the head, with the auxiliary *is* and the *to*-infinitive as its dependents. Despite the semantic near-equivalence of *will* and *is going to*, their UD structures are drastically distinct.

The problem is now illustrated with the sentences *Frank will stay* and *Frank is going to stay*:

- (10) a. – UD analysis
 b. – Purely syntactic analysis
- (11) a. – UD analysis
 b. – Purely syntactic analysis

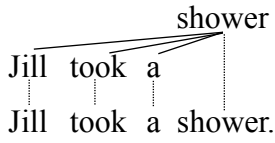
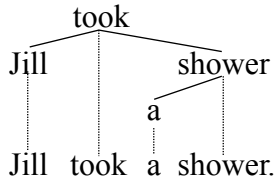
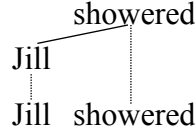
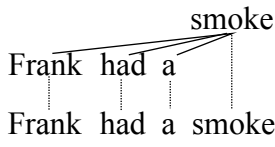
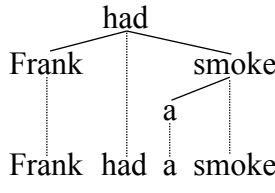
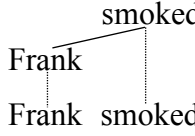
The purely syntactic analysis is consistent insofar as verb chains in English are layered and head-initial; hence sentences that are closely similar in meaning receive hierarchical analyses that are accordingly closely similar. In contrast, the UD analysis results in quite different hierarchical analyses of sentences that are closely similar semantically.⁹ This difficulty for the UD analysis repeats itself in any language that can distinguish between pure auxiliaries and semi-auxiliaries, e.g. in French where *aller* ‘be going to’ (near future) and *venir de* ‘has come to’ (recent past) are analyzed as roots of the sentence, contrary to the analysis of the pure auxiliaries *avoir* ‘have’ and *être* ‘be’.

3.2.2 Light verbs

The problem just discussed with respect to semi-auxiliaries also occurs with light verbs. Given a sentence such as *Jill took a shower*, the light verb *took* is poor on semantic content, whereas the noun *shower* expresses the main meaning of the predicate. This situation suggests that in order to be consistent about its analysis of function words, the UD analysis should subordinate *took* to *shower*, since *took* is semantically more like a function word than a content word. However, this is not what UD advocates; it instead positions the light verb *took* as the sentence root as though it were a semantically loaded full verb (cf. <http://universaldependencies.org/u/overview/specific-syntax.html>, ex. 1). In so doing, UD again elevates syntax over semantics to arrive at its analysis.

The point at issue is best illustrated by considering pairs of sentences that are almost synonymous, e.g.

⁹ Gerdes and Kahane (2016) characterize this aspect of the UD annotation scheme as resulting in a *catastrophe*, as understood in the mathematical sense of *catastrophe theory*. A seemingly minor difference in the analyses results in drastic differences in the resulting structures.

- (12) a.  – Expected UD analysis
 Jill took a shower.
- b.  – Actual UD analysis
 Jill took a shower.
- c.  – Corresponding sentence with full verb
 Jill showered.
- (13) a.  – Expected UD analysis
 Frank had a smoke.
- b.  – Actual UD analysis
 Frank had a smoke.
- c.  – Corresponding sentence with full verb
 Frank smoked.

As discussed in Gerdes and Kahane (2016), the UD decision to subordinate function words to content words predicts the UD approach to assume the a-analyses. However, UD actually advocates the b-analyses instead. The c-examples involving the corresponding full verb are included to draw attention to the manner in which the b-analyses do not follow UD assumptions about the distribution of content words in the hierarchy of structure. The a-analyses would in fact be more congruent with the distribution of content words present in the c-trees, where the semantically loaded words *Jill* and *showered* as well as *Frank* and *smoked* are directly linked to each other by dependencies.

4 Relevance for language typology

Desideratum 2 appeals to the value of the consistently annotated treebanks for typological studies. It is repeated here from Section 2.1:

Desideratum 2

UD needs to be good for linguistic typology, i.e. providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.

This section considers this goal, that is, the desire for cross-linguistic parallelism across diverse languages and language families. It also examines the value of UD corpora for two areas of syntactic typology, head-dependent ordering and the concept of dependency distance. The message is that these areas are negatively impacted by current UD annotation choices.

4.1 Structural parallelism

Concerning the desire to increase syntactic parallelism across typologically diverse languages – see the first paragraph in Section 2.2 – two points are relevant. The first is that from a linguistic perspective, leveling structural differences across typologically diverse languages is contrary to the nature of typological studies. These studies investigate and classify languages in part based on syntactic differences. Leveling these differences would hence seem counterproductive, since the differences are precisely what are of interest to the typologist. This is particularly true in the area of head-marking and dependent-marking languages, as discussed and investigated in the World Atlas of Language Structures (WALS: <http://wals.info/chapter/23>). Classifying a given language as head- or dependent-marking obviously relies directly on the hierarchical analysis of the structures in that language.

The second point is that if one nevertheless wishes to establish hierarchical parallelism across typologically diverse languages, doing so is in fact possible on purely syntactic annotation. It involves loosening strict lexicalism and acknowledging hierarchical organization among the morphological segments that constitute words, as done by Groß (2011; 2014a; b) and Groß and Osborne (2013; 2015). One allows what is a free morpheme in the one language to correspond to a bound morpheme in the other. In so doing, the typological differences across diverse languages are still present in the distinction between free and bound morphemes, but the hierarchical differences are leveled insofar as the hierarchy of functional and content elements becomes consistent across diverse languages.

This point is now illustrated briefly by considering examples across English and Japanese, two typologically quite distinct languages. On either analysis, the desired parallelism in hierarchical structure is present:

- (14) a. b.
 will eat tabe -yoo
 will eat tabe -yoo
 will eat eat -will
 – UD analyses
- c. d.
 will eat tabe -yoo
 will eat tabe -yoo
 – Purely (morpho)syntactic analyses
- (15) a. b.
 to Kyoto Kyooto -e
 to Kyoto Kyooto -e
 Kyoto to Kyoto to
 – UD analyses
- c. d.
 to Kyoto Kyooto -e
 to Kyoto Kyooto -e
 – Purely (morpho)syntactic analyses

The dotted edges mark dependencies between the segments that constitute words. These examples demonstrate that regardless of the approach, auxiliary/adposition as head or content verb/noun as head, the desired parallelism in hierarchy of structure (i.e. in the vertical dimension) holds across English and Japanese in these cases.

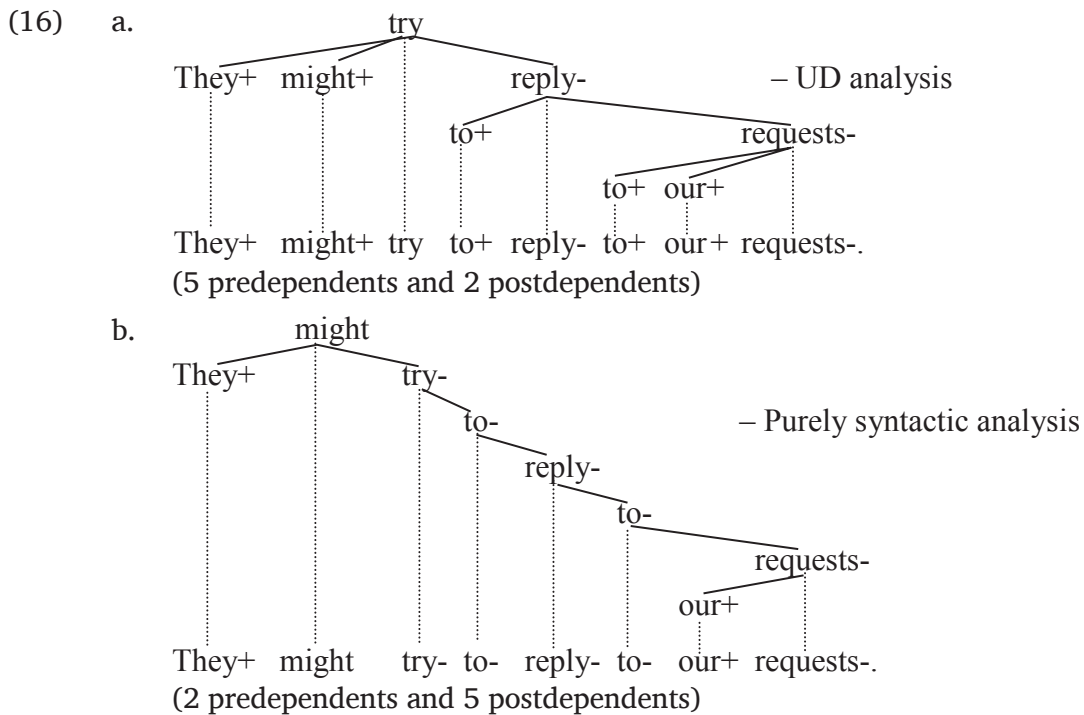
This one brief set of examples merely suggests how a measure of the desired hierarchical parallelism can in fact be achieved in a principled manner. There are many

outstanding issues of such an approach that cannot be addressed here. We instead point to the literature cited that explore the issues in some detail.

4.2 Head-dependent ordering

Head-dependent ordering has been a mainstay of language typology since Tesnière’s (1959/2015: Chapter 14) pioneering efforts to classify the world’s languages in terms of predependents, postdependents, and combinations of the two (see also Greenberg 1963; Hawkins 1983; Dryer 1992). Languages such as Turkish, Japanese, and Mongolian are viewed as syntactically similar insofar as heads appear in phrase-final position. Other languages, such as Arabic, Irish, and Welsh, are positioned at the opposite end of the continuum, since they have a large majority of heads appearing phrase-initially. Languages such as English and French are viewed as mixed, but mixed with more head-initial than head-final structures. This state of affairs is evident in the frequent phrase structure trees of English and French sentences (and the sentences of many other related languages) that one encounters in syntactic studies. These tree structures grow primarily down to the right (as opposed to down to the left).

The current UD annotation scheme undoes these insights about the positions of heads in phrases. Current UD annotation renders languages like English and French – which are, again, widely taken to be more head-initial than head-final – more head-final than head-initial. The following trees illustrate how this happens. Predependents are marked with a +, and postdependents with a –:



The frequent occurrence of auxiliary verbs, prepositions, and the particle *to* in English means that the number of predependents and postdependents present for the competing analyses varies drastically. These two structures illustrate that the numbers of predependents and postdependents on the one analysis are the opposite of what they are on the competing analysis. Thus, given that the UD annotation scheme is not supported by linguistic insight – the message established at length above in Section 3 – the UD numbers for the classification of head-initial and head-final structures are misleading.

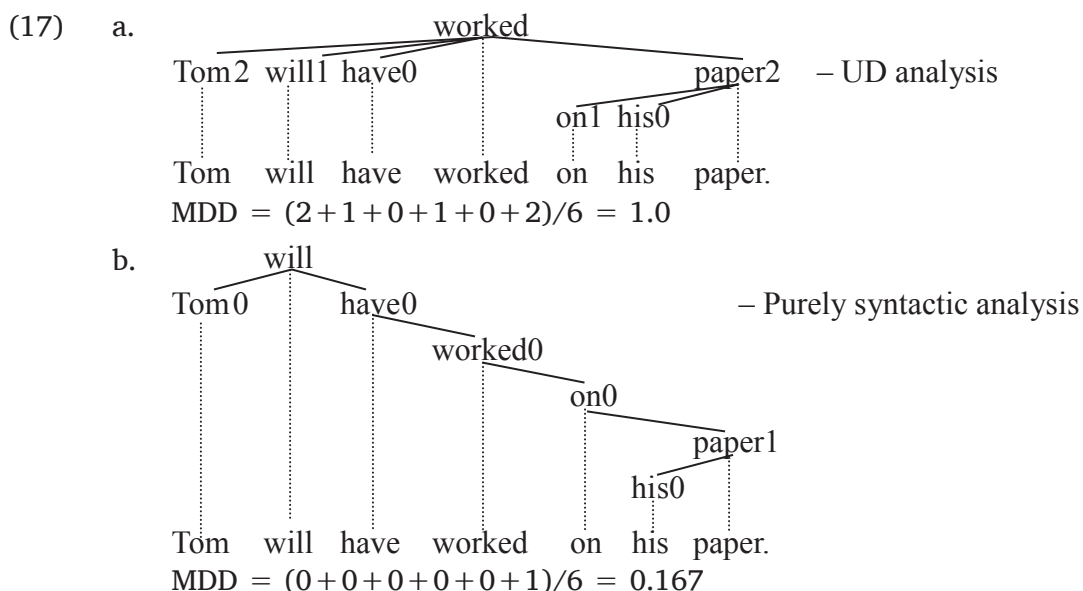
Faced with this discrepancy in numbers across the two annotation schemes, the choice is clear. Since the purely syntactic DG analysis is supported by linguistic insight, its account

of head-initial and head-final structures is solid. In contrast, current UD analyses sew confusion in our understanding of head-dependent ordering in natural language syntax. The discussion returns to this point below in Section 4.4, where the same message is delivered, but based on large quantities of corpus data.

4.3 Dependency distance (DD)

Dependency distance (DD) is a well-established metric for assessing syntactic complexity. It has been widely used to measure the complexity of syntactic structures within a given language and across languages in general (Temperley 2007; 2008; Liu 2008; 2010; Futrell et al. 2015; Liu et al. 2017). From a typological perspective, DD is a helpful measure for assessing variation in word order across languages. The hypothesis is that all or most languages adopt syntactic structures that despite great variation in word order, tend to keep the *mean dependency distance* (MDD) of their syntactic structures manageably low.

The UD annotation scheme results in sentence structures that have much higher MDDs compared to more traditional structures. More traditional structures are more layered (i.e. taller) than UD structures. The flatter UD structures can dramatically increase the MDD. This point is illustrated here first with a simple example:

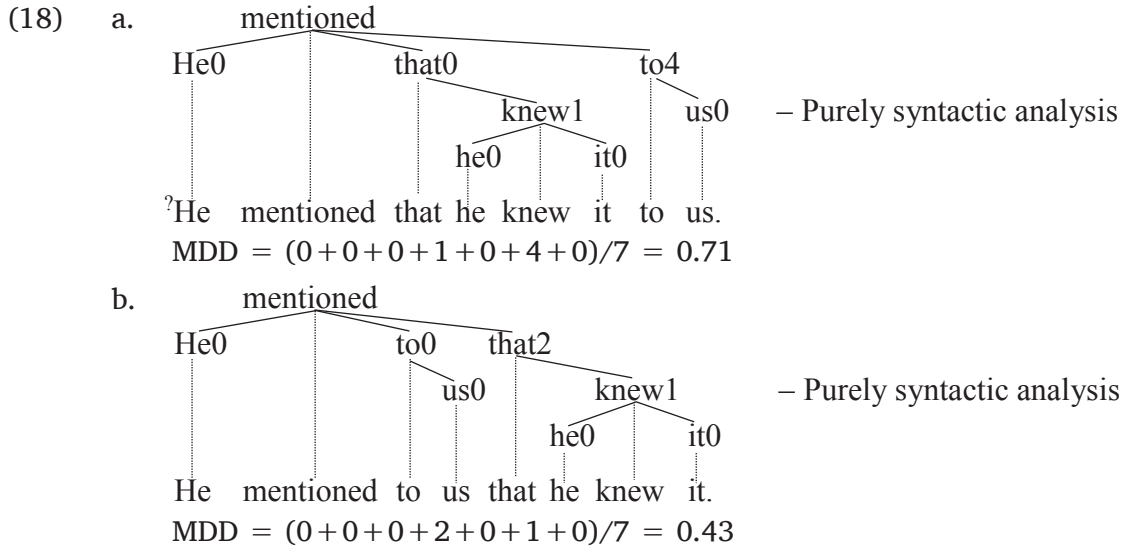


The number given immediately following each word is the DD of that word from its head word. For instance, the DD of the subject *Tom* in (17a) from its head *worked* is 2, because there are two words that separate *Tom* from *worked*.¹⁰

Observe the flatness of the UD analysis compared to the more layered traditional analysis (3 layers vs. 6 layers). The flatness of structure in (17a) results in a much higher MDD than for the more layered structure in (17b) (1.0 vs. 0.167). While this example has been constructed to emphasize the difference in MDD across UD and traditional structures, the UD annotation scheme also produces significantly higher MDD values when actually occurring sentences are involved, a point established in the next section with large quantities of corpus data.

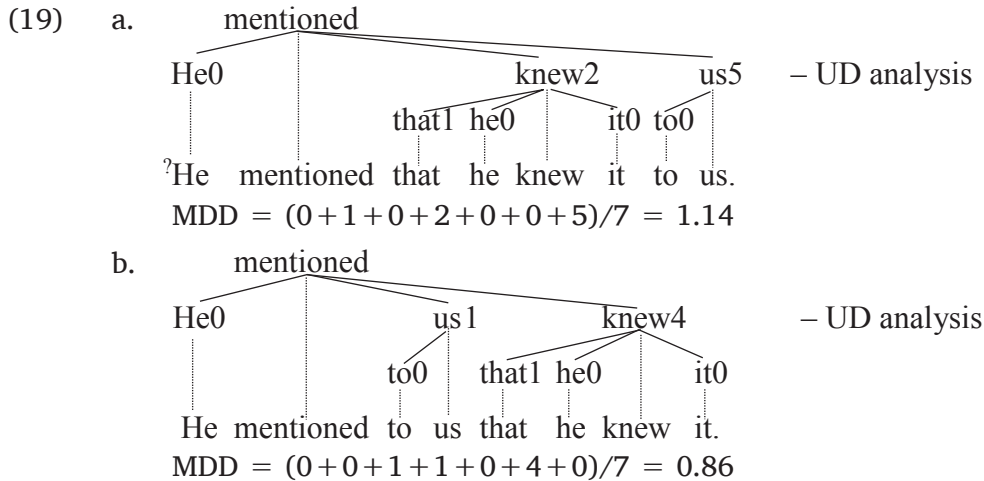
¹⁰ The method of calculating DD we use here is the original method, which was first employed by Hudson (1995). There is a competing method (used by Liu, his collaborators, and others), though, one which subtracts the linear index value of the dependent word from that of the head word and then takes the absolute value of the result. The method we employ here, which simply counts the number of intervening words between dependent and head, results in DD values that are lower by 1.

A brief illustration is now given that shows the extent to which a reduction in dependency distance sheds light on the nature of a mechanism of syntax, namely shifting. Two sentences are considered, whereby both are analyzed according to both annotation schemes. The purely syntactic analyses are given first:



The preference for sentence (18b) is apparent; the lower MDD number of 0.43 (vs. 0.71 in 18a) allows the sentence to be produced and processed more efficiently. The value of 0.71 is 65% greater than 0.43.

Turning to the UD analyses of the same sentence, they also predict the b-sentence to be preferred:



The percentage discrepancy between the two values is now not as great: 1.14 is just 33% greater than 0.86. Hence while both analyses deliver numbers that correctly predict the preference for the order in the b-sentences, the discrepancies in the MDD values is greater on the purely syntactic analyses, thus providing a more compelling basis for the account of shifting in general.

The discussion of examples (18–19) has suggested that the lower MDD values of purely syntactic annotation serve generally as a more solid basis for the investigation of the shifting mechanism. If space allowed, a similar point could be established for the extraposition mechanism. Further, the general tendency of natural language structures to avoid center embedding receives a straightforward account in terms of dependency distance, whereby

the purely syntactic analyses more effectively identify center embedding than the current UD analyses due to the more layered structures.

In the crosslinguistic big picture, an annotation scheme that results in lower MDD numbers is linguistically more plausible – other things being equal – since it is more consistent with the human tendency to reduce linguistic complexity in the interest of easing the burden on working memory.

4.4 Converting to purely syntactic annotation

The examples in the previous two sections illustrating head-dependent ordering and dependency distance are anecdotal. To establish the validity of the message broadly, we have converted the UD treebanks to purely syntactic annotation, and we then calculated the changes in head-dependent ordering and dependency distance across the two sets of treebanks (UD annotation vs. purely syntactic annotation). The results of this exercise validate our message from the previous two sections. We recorded significant changes in the numbers of head-initial and head-final dependencies across the two sets of treebanks as well as a large drop in mean dependency distance moving from the UD treebanks to the converted treebanks.

The rule-based conversion process involved a number of steps.¹¹ Auxiliary verbs were promoted to heads over content verbs, and the subject was positioned as an immediate dependent of the auxiliary. When more than one auxiliary verb were present, the finite/first auxiliary was made the root.¹² Copular verbs were positioned as heads over predicative expressions. Adpositions and particles of comparison were promoted to heads over their nouns. The *to* of *to*-infinitives was positioned as head over the infinitive. Coordinators were positioned as dependents of the immediately preceding conjunct and as head over the following conjunct. All other relations, in particular concerning multi-word expressions, remain intact. These changes resulted in structures that were mostly in line with the purely syntactic analyses of the sort given frequently above. All of the languages currently present in the UD inventory of treebanks (UD version 2.2) were converted to the corresponding structures according to these guidelines, 71 languages in total.¹³

The UD treebanks include relations of different types, some of which are not meaningful for typological computations of dependency direction and dependency distance because the structure and distance is fixed by the annotation scheme itself: this includes multi-word expressions that are represented in a bouquet structure (all words depend on the first word of the expression), coordination (also in bouquet structure), punctuation (e.g. head-initial languages having much longer *punct* relations to the final punctuation mark than head-final languages), and the *root* relation that has neither direction nor distance. To make the measures more meaningful, it was useful to restrict our measures to the 23 *syntagmatic* (parent-child) relations between words (*nsubj*, *csubj*, *obj*, *iobj*, *ccomp*, *xcomp*,

¹¹ Transformations of the sort we exacted on UD treebanks had already been proposed with the goal of measuring the impact of the annotation scheme on parser performance (see Section 4.4 for details), but the results have not been made available, so we thus had to redo the transformations.

¹² Note that this can lead to transformation errors in strict verb-final languages when two or more auxiliaries are present. Such stacking of auxiliaries occurs rarely, though, so its impact on overall numbers is minimal. A more precise conversion would require access to typological data about headedness and possibly even idiosyncratic lexical information for each language, for example indications on how to transform cases of multiple auxiliaries, adpositions, and markers on the same UD-governor. These cases are also the reason that our syntactic UD conversion is not perfectly isomorphic to the original UD scheme.

¹³ An annotation scheme similar to the one we used to convert UD treebanks is the so-called *Surface-syntactic Universal Dependencies* (SUD) annotation scheme, as presented and discussed in Gerdes et al. (2018). The main difference between our annotation scheme and that of SUD is the position of coordinators. In SUD, a coordinator is positioned as a dependent of the following conjunct rather than as head over it.

aux, cop, case, mark, cc, advmod, advcl, obl, dislocated, vocative, expl, nummod, nmod, amod, discourse, acl, det), leaving 14 other UD relations aside.¹⁴

Table 1 gives the numbers for head-initial dependencies. We list the numbers we calculated for just ten of the 71 languages – the numbers for all the languages are given in the Appendix. The number of head-initial dependencies in the UD treebank for each language is then followed by the same number for the converted treebank.

For most of the 71 languages, there was a major shift in the percentage of head-initial dependencies across the two annotation styles, the average being a shift of 14.8 percentage points. In most cases of typologically head-initial languages, the conversion to purely syntactic annotation resulted in a greater percentage of head-initial dependencies.

Having converted all the treebanks, it was also possible to calculate the mean distance of syntagmatic dependencies for each language. With few exceptions, purely syntactic annotation resulted in lower mean dependency distance, as shown in Table 2.

The mean distance of syntagmatic dependencies was reduced significantly in many cases. In fact, the mean DD was reduced for all but six of the 71 languages, and the change in the other direction for the six exceptions was negligible. The mean DD averaged across all the languages on UD annotation was 1.51, and 1.31 on purely syntactic annotation. The decrease of dependency distance moving from UD structures to purely syntactic structures is highly significant and a paired T-test on the average syntagmatic dependency relations of each language gives a p-value below 10^{-10} . See the Appendix for all of the numbers.

A word of caution is warranted when interpreting these numbers. We checked the results of the conversion process for accuracy for the languages that we know (English, German, French). For the languages we do not know, however, we could not be sure that problems did not arise from the conversion process.

Table 1: Percentage of head-initial syntagmatic dependencies on UD annotation and on purely syntactic annotation.

Language	% head-initial UD	% head-initial purely syntactic
Chinese	26.4	32.5
Czech	35.5	59.9
English	29.5	59.3
French	33.7	62.0
German	23.3	47.1
Hindi	41.2	11.8
Italian	33.9	62.0
Japanese	50.7	13.3
Russian	36.3	58.5
Spanish	35.1	64.4

¹⁴ Since the remaining relations were not altered by the transformation, including them would not have changed the general trend of the transformation, but rather the numbers across the two annotation schemes would be harder to compare and interpret. In terms of Surface-syntactic Universal Dependencies (SUDs – see footnote 13), we included the SUD relations *dep, comp, mod, and subj* and the original UD relations *cc, det, discourse, dislocated, expl, and vocative* (which SUD leaves untouched) in our measures.

Table 2: Mean distance of syntagmatic dependencies on UD annotation and on purely syntactic annotation.

Language	Mean DD UD	Mean DD purely syntactic
Chinese	2.17	2.11
Czech	1.41	1.18
English	1.55	1.09
French	1.48	1.07
German	2.13	1.91
Hindi	2.48	2.13
Italian	1.43	1.08
Japanese	2.25	1.89
Russian	1.30	1.15
Spanish	1.58	1.20

5 Desiderata 3–6

The next sections consider UD annotation with respect to human annotation (Desideratum 3), parser accuracy (Desideratum 4), learner friendliness (Desideratum 5), and downstream applications (Desideratum 5).

5.1 Human annotation

Desideratum 3 addresses the extent to which UD annotation taxes human annotators. It is repeated here from Section 2.1:

Desideratum 3

UD must be suitable for rapid, consistent annotation by a human annotator.

In view of this guideline, questions arise concerning traditional annotation choices as compared to current UD choices. Is it, for instance, more or less difficult for a human annotator to identify adpositions and subordinators and position them as roots of phrases/clauses as opposed to as dependents. In most cases, we do not see that it is any easier or harder for human annotators to do this either way. If the human annotator can identify the categorial status of such words, then positioning them as heads or dependents is a mechanical issue that cannot be construed as increasing or decreasing difficulty of annotation.

If, however, the human annotator hesitates concerning the category status of a given word, making it necessary to consult a dictionary, the purely syntactic analysis will be easier to annotate. To provide examples of this point, imagine a human annotator encountering the following sentences:

- (20) a. The analysis is difficult regarding determiners.
b. The analysis ought to hold up to scrutiny.

UD annotation requires of the human annotator that he or she decide about the category status of *regarding* in (20a), i.e. preposition or gerund. If *regarding* is deemed a preposition, it must be subordinated to *determiners*, whereas if it is deemed a gerund, it should be

positioned as head over *determiners*. A similar difficulty arises for the human annotator in sentence (20b). If *ought* is deemed an auxiliary, then it should be subordinated to *hold*, but if it is deemed a semi-auxiliary or non-auxiliary, then it should be positioned as head over *hold*. The purely syntactic analysis is not confronted with such difficulties, because it positions *regarding* as head over *determiners* regardless of whether it is deemed a preposition or a gerund. Similarly, the purely syntactic analysis positions *ought to* over *hold* regardless of whether *ought* is deemed an auxiliary or semi-auxiliary (or full content verb). Hence the purely syntactic annotation results in fewer decisions about the correct analysis and there is thus less room for incorrect choices.

The message, then, is that Desideratum 3 motivating UD annotation choices actually mildly favors the purely syntactic analysis of most function words in dependency syntax.

5.2 Parser accuracy

Desideratum 4 is repeated here from Section 2.1:

Desideratum 4

UD must be suitable for computer parsing with high accuracy.

High accuracy for statistical parsers is achieved by both a large number of coherently annotated similar structures and an easy-to-learn annotation scheme. The previous section has shown how the coherence of human annotation is negatively impacted by the UD scheme; this negative impact then has a direct effect on parser accuracy. Further, independent of treebank quality, the so-called *learnability* of the grammatical rules by a statistical parser has attracted considerable attention in the literature.

This section looks in particular into results that allow for conclusions about the UD annotation choices concerning parsing. It must be kept in mind, however, that comparison of parser performance on differently annotated treebanks is never straightforward, as factors like the complexity of two different annotation schemes are hard to compare formally. The addition of one difficult distinction in one scheme can make all the difference, not to mention factors such as language, genre, treebank size, parser type, and coherence of the annotation. The first comparison of treebank schemes in view of parser performance was conducted on German on functionally augmented phrase structures (Kübler 2005); this comparison showed just how sensitive the results are to slight differences in the annotation scheme.

Schwartz et al. (2012) show, using five parsers and two learnability measures, that training on different dependency transformations of the Penn Treebank gives a clear advantage to functional heads concerning prepositional phrases and verb groups (including auxiliaries), and the results were mixed concerning coordination and the particle *to* of infinitives. Looking only at a transition-based parser on English data, Silveira and Manning (2015) report consistent improvements for the parser when transforming to prepositional heads, noting that positioning prepositions as heads “shortens dependency lengths, which benefits the transition-based parser”. They remark, however, that the transformation itself is error-prone and considerably reduces the accuracy gains of parsing functional-heads if the goal is to obtain a UD representation.

Kirilin and Versley (2015) show a general tendency, independent of treebank size, across a set of five different state-of-the-art parsers and five languages (four Indo-European languages and Finnish) that the “content-head schema seems to make things significantly worse”. Most recently, Rehbein et al. (2017) experiment on 15 UD treebanks from various typological groups by means of three algorithmically different parsers (graph-based,

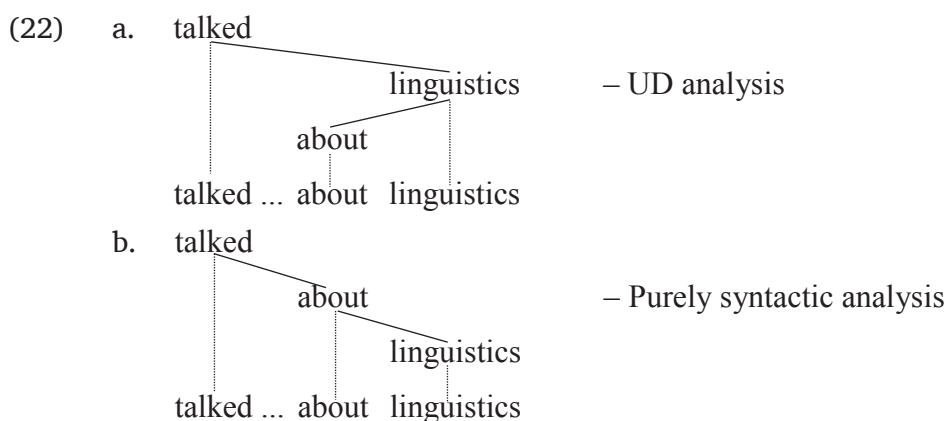
transition based, and head-selection parsers). They conclude that functional heads are beneficial for the parser accuracy for all languages where the treebanks are sufficiently large and coherently annotated.

These results on statistical parsing correspond to linguistic intuition and experience with annotation. One of the biggest problems facing accurate parsing is the attachment of prepositional phrases,¹⁵ e.g.

- (21) a. I talked to the students **about linguistics**.
 b. I talked to the students **of linguistics**.

It is hard for the parser to choose the attachment points of the PPs *about linguistics* in (21a) and *of students* in (21b). Concerning the former, the parser has to have encountered more cases of *talking about* than of *students about* to correctly attach *about* to *talked*, and not to *students*. Or if the POS tagging has already been accomplished as a first independent step, the parser has to have encountered more verbs than nouns having an *about*-PP as a dependent to produce the correct analysis of (21a) and more nouns than verbs having an *of*-PP as a dependent to produce the correct analysis of (21b).

At this point, the difficulty facing UD annotation concerning prepositions is again evident (see Section 3.1.2):



To produce the correct attachment of *about linguistics* to *talked* (and not to *students*), the UD analysis requires the parser to somehow access the preposition *about* from *talked* through the noun *linguistics*. On the purely syntactic analysis, the parser is not confronted with this difficulty because of the direct dependency that can be established between the two words that belong together, i.e. *talked* and *about*. The rule-based transformability of UD into functional head schemes and back implies that this transformation could be picked up implicitly by a sufficiently complex learning algorithm, i.e. some parsers may be able to recognize the structural fact about the UD analysis illustrated with (22a) and still recognize that *about* is in the subcategorization frame of *talked*, but doing so is undoubtedly more difficult than if they are operating on the purely syntactic analysis of PPs illustrated with (22b).

5.3 Learner friendliness

The UD project foresees the UD annotation scheme as learner friendly, a fact that is present in Desideratum 5, which is repeated here from Section 2.1:

¹⁵ PP attachment remains the central difficulty of all types of parsers (cf. Kummerfield et al. 2012), and even small advances in this area are important for parser improvement.

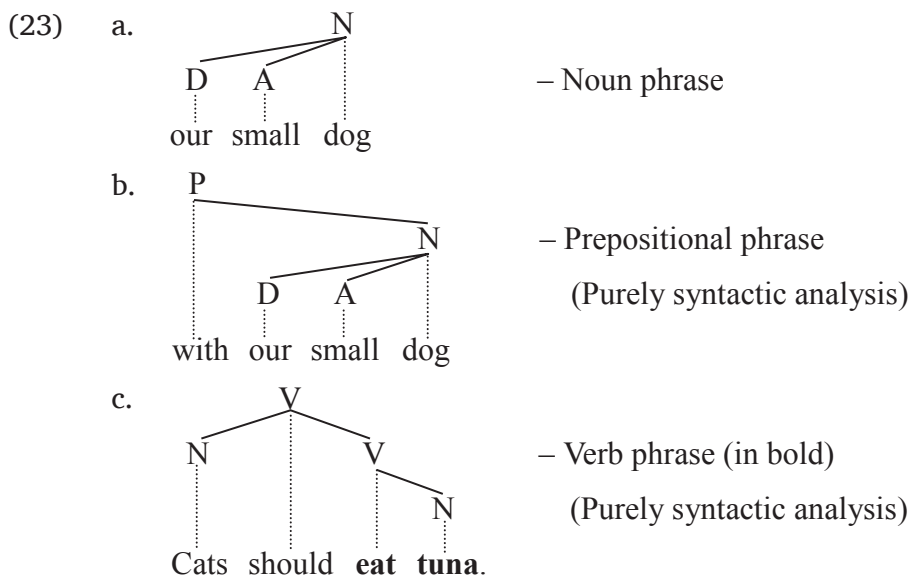
Desideratum 5

UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. We refer to this as seeking a *habitable* design, and it leads us to favor traditional grammar notions and terminology.

The DG tradition has been associated with pedagogical applications from the very start. This fact is particularly evident in the works of Franz Kern (1882; 1883; 1884; 1886; 1888), whose primary interest was to reform the manner in which grammar was taught in Prussian schools, and in the works of Lucien Tesnière (1953; 1959/2015: Book F), who was striving to reform the way in which grammar was taught in French schools. Both of these grammarians advocated an approach to teaching school grammar that emphasized the role of the verb in establishing sentence structure, and both advocated the use of sentence diagrams in the classroom, which Tesnière called *stemmas*, to illustrate sentence structure.

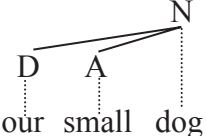
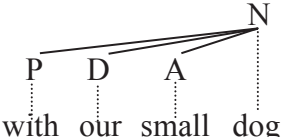
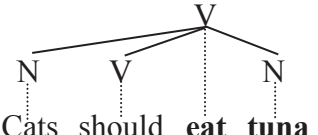
While Desideratum 5 itself is consistent with the DG tradition regarding pedagogical applications, we disagree with the notion that UD sentence structures can aid grammar learning or be consistent with traditional grammar notions. UD analyses of sentence structure are in fact contrary to the standard use of grammar terminology, a fact that is visible with the analyses of traditional phrases (e.g. prepositional phrase, noun phrase, verb phrase, etc.). From a diagrammatic standpoint, UD analyses actually deny the existence of prepositional phrases and of some verb phrases as well.

An intuitive and pedagogically helpful explanation of phrases is that the root of a given phrase determines what type of phrase it is. The root of a noun phrase is a noun, the root of a verb phrase is a verb, etc. Traditional dependency structures illustrate this state of affairs clearly:



The root of each of these phrases matches the traditional designation employed to denote the phrase in category status in traditional dependency and phrase structure grammars. The root of the noun phrase *our small dog* is the noun *dog*; the root of the prepositional phrase *with our small dog* is the preposition *with*; and the root of the verb phrase *eat tuna* is the verb *eat*.

In contrast, the analysis of these phrases becomes confused on current UD assumptions:

- (24) a.  – Noun phrase
- b.  – Prepositional phrase or noun phrase?
(UD analysis)
- c.  – Verb phrase or clause?
(UD analysis)

The prepositional phrase *with our small dog* can no longer be construed as a prepositional phrase, but rather it has become a noun phrase with the root noun *dog*. Similarly, the verb phrase *eat tuna* can no longer be construed as a verb phrase – because it is not a complete subtree – but rather it has become a non-phrasal part of the clause that includes *Cats should*.

The message delivered with these examples is that the intended utility of UD structures for pedagogical applications is reduced by the mismatch in root status and the traditional terms *prepositional phrase* and *verb phrase*. The UD account has to explain the fact that the root of a prepositional phrase is in fact not a preposition but rather a noun and that some verb phrases are in fact not phrases but rather non-phrasal parts of other phrases or clauses. The purely syntactic analysis is not faced with these difficulties, since its analysis of phrases is ideally consistent with traditional terminology.

5.4 Downstream applications

The final issue addressed in this article concerns Desideratum 6, which is repeated here from Section 2.1:

Desideratum 6

UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...).

Sentence condensation and semantic similarity queries are additional areas that can be added to this list of downstream language understanding tasks, as mentioned by an anonymous reviewer. The reviewer points to Riezler et al. (2003) and Crouch et al. (2004) concerning sentence condensation and Riezler and Maxwell (2006) and Owczarzak et al. (2008) in the area of machine translation. These studies all assume LFG f-structures as a basis for analysis, whereby the hierarchical organization of the f-structures is similar to dependency structures.

The extent to which UD annotation is more suited to areas mentioned in Desideratum 6 remains to be shown. The studies just named all pre-date the UD project and more importantly, the f-structures they assume are in line with the purely syntactic analyses above and thus contrary to current UD annotation. These f-structures have auxiliaries as heads over content verbs (Crouch et al. 2004: 176), the copula as head over predicative elements (Riezler 2003: 120), and prepositions as heads over nouns (Riezler 2003: 120; Crouch et al. 2004: 170; Riezler and Maxwell 2006: 250). Further, general accounts of f-structures tend to view auxiliaries as features (e.g. Bresnan 2001: 116–117; Falk 2001:

82–84), which means they are not granted node status and so cannot be interpreted as heads or dependents. Semantically loaded prepositions, however, are viewed as predicates and are hence positioned in f-structures as heads over their complement NPs (e.g. Bresnan 2001: 50).

Nevertheless, if other studies that we are unaware of and/or future investigations support the current UD annotation scheme by demonstrating that UD structures promote the goals targeted in the areas mentioned, then that is not problematic for our message here because of the ability to convert the one annotation format to the other as we have done in the course of our critique of current UD annotation.

Part of the message we wish to deliver is that the UD treebanks already in existence and the alternative treebanks that have been converted from the current UD format to the purely syntactic format should be made available in parallel. Which of the two sets of treebanks one uses can then depend on the particular goals of the researcher at hand. Those linguists who are investigating the nature of syntactic structures or interested in typological comparisons in areas such as head-dependent ordering and/or dependency distance can then choose to use the alternative set of treebanks that have been converted from the existing treebanks. Otherwise, the current set of UD treebanks can be chosen – assuming that they do prove to be more suited for work in relation extraction, reading comprehension, machine translation, sentence condensation, and/or semantic similarity queries.

6 Concluding comments

This article has critiqued the current annotation scheme of the UD project. In doing so, the six desiderata employed to motivate the scheme have been examined, whereby the majority of the discussion focused on Desideratum 1 (linguistic analysis) and Desideratum 2 (linguistic typology). Linguistic considerations have revealed that the current UD annotation scheme results in structures that are a mixture of semantic and syntactic motivations. These structures are hence not well-motivated from the linguistic point of view. As an alternative, we have advocated for a more traditional annotation scheme, one that consistently elevates syntactic criteria for determining headhood over semantic criteria. This alternative annotation scheme positions auxiliary verbs as heads over content verbs, the copula as head over predicative elements, and adpositions and subordinators as heads over nouns and verbs.

The discussion also considered Desideratum 3 (human annotation), Desideratum 4 (parser accuracy), and Desideratum 5 (learner friendliness). The discussion of these three desiderata was less extensive and the conclusion reached about them less robust. Nevertheless, the issues we have raised in this area are also more congruent with the purely syntactic annotation format.

Of the six desiderata, only Desideratum 6 (relation extraction, reading comprehension, machine translation) remains as a potential source of support for current UD annotation choices. This potential support need not be viewed as a problem for the message delivered above, though. The ability to automatically convert treebanks annotated according to the one annotation scheme into treebanks annotated according to the other means that both sets of treebanks can be made available, allowing the individual researcher to choose which of the two sets of treebanks best matches his or her goals.

Abbreviations

DD = dependency distance, DG = dependency grammar, MDD = mean dependency distance, NOM = NOMINATIVE, IP/TP = inflection phrase/tense phrase, 2SG = second person singular, UD = Universal Dependencies, VP-ellipsis = verb phrase ellipsis, WALS = World Atlas of Language Structures

Additional File

The additional file for this article can be found as follows:

- **Appendix.** DOI: <https://doi.org/10.5334/gjgl.537.s1>

Acknowledgements

The authors would like to thank Thomas Groß for his contributions to early drafts of this article.

Competing Interests

The authors have no competing interests to declare.

References

- Bresnan, Joan. 2001. *Lexical-functional syntax*. Malden, MA: Blackwell Publishers.
- Bröker, Norbert. 1999. *Eine Dependenzgrammatik zur Kopplung heterogener Wissensquellen*. Tübingen: Niemeyer. DOI: <https://doi.org/10.1515/9783110915952>
- Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris publications.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Crouch, Richard, Tracy Holloway King, John Maxwell III, Stefan Riezler & Annie Zaenen. 2004. Exploiting f-structure input for sentence condensation. In *Proceedings of the LFG04 Conference*, 167–187. Christchurch: University of Canterbury.
- de Marneffe, Marie-Catherine, Bill MacCartney & Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *The International Conference on Language Resources and Evaluation (LREC) 2006*, 449–454. Genoa.
- de Marneffe, Marie-Catherine & Christopher D. Manning. 2008. The Stanford typed dependency representation. In *Proceedings of the COLING Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, 92–97. Sofia. DOI: <https://doi.org/10.3115/1608858.1608859>
- de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silvaire, Katrin Haverinen, Filip Ginter, Joakim Nivre & Christopher D. Manning. 2014. Universal Stanford Dependencies: A cross-linguistic typology. In *The International Conference on Language Resources and Evaluation (LREC) 2014*, 4585–4592. Reykjavik.
- Dryer, Mathew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138. DOI: <https://doi.org/10.1353/lan.1992.0028>
- Engel, Ulrich. 1994. *Syntax der deutschen Gegenwartssprache*. 3rd edition; Berlin: Erich Schmidt Verlag.
- Eroms, Hans-Werner. 2000. *Syntax der deutschen Sprache*. Berlin: de Gruyter.
- Falk, Yehuda. 2001. *Lexical-functional grammar: An introduction to parallel constraint-based syntax*. Stanford, CA: CSLI Publications.
- Futrell, Richard, Kyle Mahowald & Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *PNAS* 112. 10336–10341. DOI: <https://doi.org/10.1073/pnas.1502134112>
- Gerdes, Kim, Bruno Guillaume, Sylvain Kahane & Guy Perrier. 2018. SUD or Surface-Syntactic Universal Dependencies: An annotation scheme near-isomorphic to UD. In *Proceedings of Universal Dependencies Workshop 2018*, 66–74. Brussels.
- Gerdes, Kim & Sylvain Kahane. 2007. Phrasing it differently. In *Selected lexical and grammatical issues in the Meaning Text Theory*, 297–336. Amsterdam/Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/slcs.84.10ger>
- Gerdes, Kim & Sylvain Kahane. 2016. Dependency annotation choices: Assessing theoretical and practical issues of Universal Dependencies. In *LAW X (2016), The 10th linguistic annotation workshop 131*, 131–140. Berlin.

- Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 58–60. Cambridge, MA: MIT Press.
- Groß, Thomas. 1999. *Theoretical foundations of dependency syntax*. Munich: Iudicium.
- Groß, Thomas. 2011. Catenae in morphology. In Kim Gerdes, Eva Hajičová & Leo Wanner (eds.), *Proceedings of the First International Conference on Dependency Linguistics (Depling 2011)*, 47–57. Barcelona.
- Groß, Thomas. 2014a. Clitics in dependency morphology. In Kim Gerdes, Eva Hajičová & Leo Wanner (eds.), *Dependency linguistics: Recent advances in linguistic theory using dependency Structures*, 229–252. Amsterdam/New York: John Benjamins. DOI: <https://doi.org/10.1075/la.215.11gro>
- Groß, Thomas. 2014b. Some observations on the Hebrew desiderative construction: A dependency-based account in terms of catenae. *SKY Journal of Linguistics* 27. 7–41.
- Groß, Thomas & Timothy Osborne. 2013. Katena und Konstruktion: Ein Vorschlag zu einer dependenziellen Konstruktionsgrammatik. *Zeitschrift für Sprachwissenschaft* 32(1). 41–73. DOI: <https://doi.org/10.1515/zfs-2013-0002>
- Groß, Thomas & Timothy Osborne. 2015. The dependency status of function words: Auxiliaries. *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 111–120. Uppsala.
- Hawkins, John A. 1983. *Word order universals*. New York: Academic Press.
- Hays, David. 1964. Dependency theory: A formalism and some observations. *Language* 40. 511–525. DOI: <https://doi.org/10.2307/411934>
- Heringer, Hans. 1996. *Deutsche Syntax: Dependentiell*. Tübingen: Stauffenburg.
- Hudson, Richard. 1984. *Word Grammar*. Oxford, UK: Basil Blackwell.
- Hudson, Richard. 1987. Zwicky on heads. *Journal of Linguistics* 23(1). 109–132. DOI: <https://doi.org/10.1017/S0022226700011051>
- Hudson, Richard. 1990. *An English Word Grammar*. Oxford: Basil Blackwell.
- Hudson, Richard. 1995. Measuring syntactic difficulty. Draft of manuscript. Available at: <http://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf>.
- Hudson, Richard. 2007. *Language networks: The new Word Grammar*. Oxford: Oxford University Press.
- Hudson, Richard. 2010. *An introduction to Word Grammar*. Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511781964>
- Järventausta, Marja. 2003. Das Subjektpromblem in der Valenzforschung. In Vilmos Ágel, Ludwig M. Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research* 1. 781–794. Berlin: Walter de Gruyter.
- Jung, Wha-Young. 1995. *Syntaktische Relationen im Rahmen der Dependenz-grammatik*. Hamburg: Helmut Buske Verlag.
- Kern, Franz. 1882. *Die deutsche Satzlehre: Eine Untersuchung ihrer Grundlagen*. Berlin: Nicolaische Verlags-Buchhandlung.
- Kern, Franz. 1883. *Zur Methodik des deutschen Unterrichts*. Berlin: Nicolaische Verlags-Buchhandlung.
- Kern, Franz. 1884. *Grundriss der deutschen Satzlehre*. Berlin: Nicolaische Verlags-Buchhandlung.
- Kern, Franz. 1886. *Betrachtungen über den Anfangsunterricht in der deutschen Satzlehre*. Berlin: Nicolaische Verlags-Buchhandlung.
- Kern, Franz. 1888. *Leitfaden für den Anfangsunterricht in der Deutschen Grammatik*. Berlin: Nicolaische Verlags-Buchhandlung.
- Kirilin, Angelika & Yannick Versley. 2015. What is hard in Universal Dependency Parsing. In *Proceedings of the 6th Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL 2015)*, 31–38. Bilbao.

- Kübler, Sandra. 2005. How do treebank annotation schemes influence parsing results? Or how not to compare apples and oranges. In *Proceedings of Recent Advances in Natural Language Processing (RANLP) 2005*. Available at: <https://www.researchgate.net/publication/228621065>.
- Kummerfield, Jonathan K., David Hall, James R. Curran & Dan Klein. 2012. Parser showdown at the Wall Street corral: An empirical investigation of error types in parser output. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1048–1059. Jeju Island, Korea.
- Kunze, Jürgen. 1975. *Abhängigkeitsgrammatik (Studia Grammatika XII)*. Berlin: Akademie-Verlag.
- Lacheret, Anne, Sylvain Kahane, Julie Beliao, Anne Dister, Kim Gerdes, Jean-Philippe Goldman, Nicolas Obin, Paola Pietrandrea & Atanas Tchobanov. 2014. Rhapsodie: A prosodic-syntactic treebank for spoken French. In *Language Resources and Evaluation Conference*, 295–301. Reykjavik.
- Liu, Haitao. 2008. Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science* 9(2). 159–191. DOI: <https://doi.org/10.17791/jcs.2008.9.2.159>
- Liu, Haitao. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120(6). 1567–1578. DOI: <https://doi.org/10.1016/j.lingua.2009.10.001>
- Liu, Haitao, Chunshan Xu & Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews* 21. 171–193. DOI: <https://doi.org/10.1016/j.plrev.2017.03.002>
- Matthews, Peter. 1981. *Syntax*. Cambridge, UK: Cambridge University Press.
- Mel'čuk, Igor. 1988. *Dependency syntax: Theory and practice*. Albany, NY: State University of New York Press.
- Mel'čuk, Igor. 2003. Levels of dependency description: Concepts and problems. In Vilmos Ágel, Ludwig M. Eichinger, Hans-Werner Eroms, Peter Hellwig, Hans Jürgen Heringer & Henning Lobin (eds.), *Dependency and valency: An international handbook of contemporary research* 1. 188–229. Berlin: Walter de Gruyter.
- Mel'čuk, Igor. 2009. Dependency in Natural language. In Igor Mel'čuk & Alain Polguère (eds.), *Dependency in linguistic description*, 1–110. Amsterdam, Philadelphia: John Benjamins. DOI: <https://doi.org/10.1075/slcs.111.03mel>
- Mille, Simon, Alicia Burga & Leo Wanner. 2013. AnCora-UPF: A multi-level annotation of Spanish. In *Proceedings of the Second International Conference on Dependency Linguistics (Depling 2013)*, 217–226. Prague.
- Osborne, Timothy & Daniel Maxwell. 2015. A historical overview of the status of function words in dependency grammar. *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 241–250. Uppsala.
- Osborne, Timothy, Michael Putnam & Thomas Groß. 2012. Catenae: Introducing a novel unit of syntactic analysis. *Syntax* 15(4). 354–396. DOI: <https://doi.org/10.1111/j.1467-9612.2012.00172.x>
- Osborne, Timothy & Thomas Groß. 2017. Left node blocking. *Journal of Linguistics* 53. 641–688. DOI: <https://doi.org/10.1017/S0022226717000111>
- Owczarzak, Karolina, Josef van Genabith & Andy Way. 2008. *Machine Translation* 21. 95–119. DOI: <https://doi.org/10.1007/s10590-008-9038-1>
- Petrov, Slav, Dipon Das & Ryan McDonald. 2012. A universal part-of-speech tagset. In *The International Conference on Language Resources and Evaluation (LREC) 2012*, 2089–2096. Istanbul.
- Pollard, Carl & Ivan Sag. 1994. *Head-driven phrase structure grammar*. Chicago, IL: The University of Chicago Press.

- Rehbein, Ines, Julius Steen & Bich-Ngoc Do. 2017. Universal Dependencies are hard to parse—or are they? In *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling 2017)*, 218–228, Pisa.
- Riezler, Stefan & John Maxwell III. 2006. Grammatical machine translation. In *Proceedings of the human language technology conference of the North American chapter of the ACL*, 248–255. New York: Association of Computational Linguistics.
- Riezler, Stefan, Tracy King, Richard Crouch & Annie Zaenen. 2003. Statistical sentence condensation using ambiguity packing and stochastic disambiguation methods for Lexical-Functional Grammar. In *Human Language Technology Conference – North American Chapter of the Association for Computational Linguistics (HLT-NAACL) 2003*, main papers, 118–125. Edmonton.
- Schubert, Klaus. 1987. *Metataxis: Contrastive dependency syntax for machine translation*. Dordrecht: Foris Publications.
- Schwartz, Roy, Omri Abend & Ari Rappoport. 2012. Learnability-based syntactic annotation design. In *COLING 24*. 2405–2422.
- Silveira, Natalia & Christopher Manning. 2015. Does Universal Dependencies need a parsing representation? An investigation of English. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, 310–319. Uppsala.
- Starosta, Stanley. 1988. *The case for Lexicase: An outline of Lexicase grammatical theory*. London: Pinter Publishers.
- Temperley, David. 2007. Minimization of dependency length in written English. *Cognition* 105. 300–333. DOI: <https://doi.org/10.1016/j.cognition.2006.09.011>
- Temperley David. 2008. Dependency-length minimization in natural and artificial languages. *Journal of Quantitative Linguistics* 15. 256–282. DOI: <https://doi.org/10.1080/09296170802159512>
- Tesnière, Lucien. 1953. *Esquisse d'une syntaxe structurale*. Paris: Klincksieck.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Tesnière, Lucien. 2015 (1959). *Elements of structural syntax*. Translated by Timothy Osborne and Sylvain Kahane. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/z.185>
- Zeman, Daniel. 2008. Reusable tagset conversion using tagset drivers. In *The International Conference on Language Resources and Evaluation (LREC) 2008*, 213–218. Marrakech.
- Zwicky, Arnold. 1985. Heads. *Journal of Linguistics* 21. 1–29. DOI: <https://doi.org/10.1017/S0022226700010008>

How to cite this article: Osborne, Timothy and Kim Gerdes. 2019. The status of function words in dependency grammar: A critique of Universal Dependencies (UD). *Glossa: a journal of general linguistics* 4(1): 17.1–28, DOI: <https://doi.org/10.5334/gjgl.537>

Submitted: 03 October 2017

Accepted: 07 November 2018

Published: 30 January 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS