

Towards a Universal Grammar for Natural Language Processing

Joakim Nivre

Uppsala University,
Department of Linguistics and Philology
`joakim.nivre@lingfil.uu.se`

Abstract. Universal Dependencies is a recent initiative to develop 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. In this paper, I outline the motivation behind the initiative and explain how the basic design principles follow from these requirements. I then discuss the different components of the annotation standard, including principles for word segmentation, morphological annotation, and syntactic annotation. I conclude with some thoughts on the challenges that lie ahead.

1 Introduction

The notion of a universal grammar in linguistics and philosophy goes back at least to Roger Bacon’s observation that “[i]n its substance, grammar is one and the same in all languages, even if it accidentally varies” [1, p. 27]. It can be traced forward through the speculative grammars of the Middle Ages and the Port-Royal grammar of Arnauld and Lancelot [2], all the way to the theories of Noam Chomsky [3,4]. What these theories have in common is the assumption that all human languages are species of a common genus because they have all been shaped by a factor that is common to all human beings. For the speculative grammarians working in the Aristotelian tradition of scholastic philosophy, this factor was simply thought to be the world itself. Arnauld and Lancelot replaced the external world by the human mind, which in Chomskyan linguistics has been further specified to an innate language faculty. Regardless of these differences, the main idea is that we can bring order into the chaos of linguistic variation by referring to a common underlying structure.

How is this relevant for natural language processing? Traditionally, research in our community has not paid much attention to language typology or linguistic universals. At one end of the scale, we find systems based on language-specific resources that cannot easily be ported or generalized even to closely related languages. At the other end, we find general statistical models that can be applied to any language and therefore are not attuned to the special characteristics of any individual language. The first approach eschews linguistic variation altogether; the second embraces it with ignorance; but neither manages to bring any order into the chaos.

There are definitely signs that this is about to change. Research on statistical parsing of morphologically rich languages has highlighted the interplay between language typology and parsing technology [5,6]. Studies on cross-lingual learning have shown that typological information can be exploited to improve learning and adaptation across languages [7,8]. However, a major obstacle to progress in these areas has been the fact that annotation standards vary across languages almost as much as the languages themselves. Hence, it is often very difficult to say exactly which differences in performance are due to real structural differences between languages, as opposed to more or less arbitrary differences in annotation conventions.

In this paper, I will present a recent initiative to overcome these difficulties by creating guidelines for cross-linguistically consistent grammatical annotation, called Universal Dependencies (UD). The UD project builds on and subsumes several earlier initiatives, including Intersect [9], Google Universal Part-of-Speech Tags [10], HamleDT [11], Universal Dependency Treebanks [12], and Universal Stanford Dependencies [13]. We may think of UD as a universal grammar for natural language processing, but as such it is fundamentally different from the notion of universal grammar found in linguistics and philosophy. Our goal is not to give an explanatory account of the structural variation found in the world's languages, but to represent it in a way that is revealing and useful for the purpose of natural language processing. Hence, as long as the representation is found to be practically useful, it is immaterial whether it captures the true universal grammar, or indeed whether such a grammar even exists. Needless to say, it would be very rewarding if our efforts turned out to have significance also for the more theoretical discussion of a universal grammar, but our current ambitions are more modest than that.

2 Motivation

Syntactic parsing is rarely an end in itself, so its main role in natural language processing is to extract information about grammatical structure from sentences for the benefit of applications like machine translation, information extraction, and question answering. It is an open question to what extent these applications benefit from grammatical information, and in what form it should be provided, but we currently see a trend towards an increased use of parsing, also in applications like information retrieval that traditionally have not considered grammatical information. The recent parsing trend has also clearly favored dependency-based representations, which provide a simple and transparent encoding of predicate-argument structure and which are also amenable to very efficient processing.

To develop efficient and accurate parsers, we currently need access to grammatically annotated corpora, or treebanks. Although unsupervised parsing is an interesting alternative in theory, the accuracy is still much too low for practical purposes. This is a bottleneck because grammatical annotation is expensive. Thus, treebanks are only available for a small fraction of the world's languages,

and the amount of data for each language is often quite limited. Moreover, the annotation schemes vary considerably across languages, which makes it hard to use data from rich-resource languages to bootstrap parsers for low-resource languages.

The large variation in annotation schemes across languages can to some extent be explained by different theoretical preferences among treebank developers. More important, however, is the fact that broad-coverage linguistic annotation almost inevitably has to rely on descriptive grammatical traditions established over long time for specific languages. These traditions are often roughly similar across similar languages but nevertheless with more or less arbitrary (and often quite subtle) differences in terminology and notation. When these differences are inherited into treebank annotation schemes, they give rise to a number of problems for researchers and developers in natural language processing, including but not limited to the following [12]:

- In multilingual applications involving parsers, it is a major drawback if downstream components cannot assume uniform output representations from parsing, because we then require specialized interfaces for each language.
- In cross-lingual learning and parser evaluation, inconsistent annotation standards make it impossible to properly evaluate system performance, because we cannot separate parse errors from discrepancies in the standards.
- In statistical parsing generally, it is hard to make effective use of linguistic knowledge, because we cannot assume a consistent representation of linguistic categories and structures across languages.

The major goal of the UD project is to alleviate these problems by creating a standard for cross-linguistically consistent grammatical annotation, a standard that brings out cross-lingual similarities in a perspicuous way without forcing all languages into the same mold. In a way, UD tries to do for statistical parsing what initiatives like ParGram [14] and DELPH-IN [15] have done for the grammar-based parsing community, and we focus on the needs of multilingual natural language processing from a mainly practical point of view. This does not preclude that UD can be useful for other purposes as well, but a few disclaimers may be appropriate in order not to create false expectations:

- UD is not proposed as a linguistic theory. While we like to think that most of our design decisions are informed by linguistic theory, we are also well aware that we sometimes have to make compromises in the interest of practical utility. The representations are in general oriented towards surface syntax, but in the interest of a transparent encoding of predicate-argument structure we also encode aspects of deep syntax, and the representations probably do not correspond to any well-defined level in theoretical grammar frameworks. We nevertheless think that UD could be a useful resource also for more linguistically oriented studies of grammatical structure across languages.
- UD may not be the ideal treebank annotation scheme for all projects. The main goal is to provide a kind of *lingua franca* for grammatical annotation,

which can be used for data interchange and development of multilingual systems, but we do not have the ambition to capture all the information that is encoded in specific treebank annotation schemes. Hence, we do not expect all treebank developers to abandon their language-specific schemes in favor of UD, but we do hope that treebank developers will find UD useful enough to make the extra effort to ensure that their own scheme can be converted to UD in a noiseless fashion (though possibly with some loss of information). In addition, we think that UD could be a convenient choice for quick-starting new annotation projects given the availability of consistent guidelines for many languages.

- UD is not necessarily an optimal parsing representation. It is clear that the need for cross-linguistic consistency and perspicuity often runs counter to the requirements of optimal parsability for specific languages. We therefore envisage that parsers expected to output UD representations often will have to use different representations internally. In fact, we believe that research on finding optimal representations for parsers, which has been a dominant theme in constituency-based parsing for the last twenty years, is an under-exploited area in dependency parsing. With a touch of irony, we could even say that the obvious suboptimality of UD representations for parsing is our way of encouraging more research into these problems.

3 Design Principles

Given our ambition to support both parsing research and system development in a multilingual setting, we have opted for an annotation standard that is close to common usage and based on existing de facto standards. The basic structure of the annotation is that *sentences* are segmented into *words* and that words are described by *morphological properties* and linked by *syntactic relations*. A typical representation of this kind is shown in Figure 1.¹

The decision to treat words as the basic units of analysis constitutes a commitment to the *lexicalist hypothesis* in syntax [13], but it is also consistent with common practice in practical natural language processing. This means that we do not attempt any morphological segmentation of words but instead use a word-based morphology [17], where morphological categories are represented as properties of whole words. By words, however, we mean *syntactic* words (not orthographic or phonological words), so clitics are treated as separate words regardless of how they are represented orthographically, and contractions are split if they consist of syntactically independent words. The principles of word segmentation are described in more detail in Section 4.

The morphological description of a word consists of three parts: a lemma (or base form), a universal part-of-speech tag, and a set of morphological features

¹ The format used to encode these annotation is a revised version of the CoNLL-X format for dependency treebanks [16] called CoNLL-U. For more information about this format, see <http://universaldependencies.github.io/docs/format.html>.

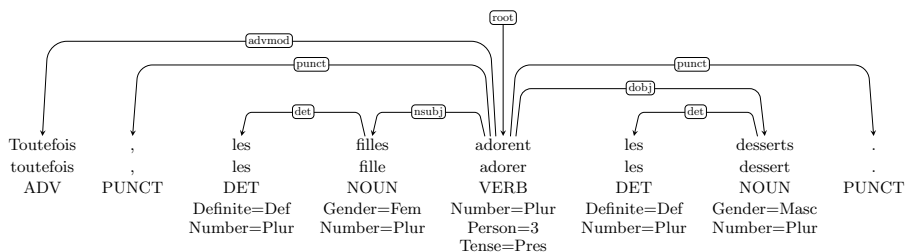


Fig. 1. UD representation of a French sentence

(attribute-value pairs). The UD tagset is a revised and extended version of the Google Universal Part-of-Speech Tagset [10], and the inventory of morphological attributes and values is based on Intersect [9]. The morphological annotation is further described in Section 5.

The adoption of lexicalism and word-based morphology fits very well with a dependency-based view of syntax [18], which is also the most widely used form of syntactic annotation in available treebanks. In addition, as noted earlier, it is currently favored over other representations by system developers. UD adopts a version of the Universal Stanford Dependencies [13], containing 40 universal dependency relations, further described in Section 6.

The overriding principle of the guidelines for morphological and syntactic annotation is to bring out similarities and differences between languages by maximizing parallelism in annotations. This can be summed up in two slogans:

- Don’t annotate the same thing in different ways!
- Don’t make different things look the same!

However, it is important to apply these principles with reason, to avoid turning the annotation scheme into a Procrustean bed. Hence, we also apply the slogan:

- Don’t annotate things that are not there!

For instance, we do not introduce empty subjects in pro-drop languages just because other languages have obligatory overt subjects. We also allow language-specific extensions of the annotation scheme in two places. In the morphological features, each language selects a subset of the universal features and can in addition add language-specific features (cf. Section 5). In the syntactic relations, each language may define language-specific subtypes of the universal relations (cf. Section 6). The subtyping is important because it gives us a mechanism for backing off to completely homogeneous representations in contexts where this is important. The language-specific documentation for each treebank should specify what language-specific extensions are used (if any), so that users of the treebanks can make informed choices about how to use the resource.

Finally, it must be emphasized that the UD scheme is still evolving. The first version of the guidelines, released in October 2014, will remain stable for at least

a year (and probably longer) to give the community a chance to explore it and apply it in different projects. But it is very unlikely that the first version is also the final version. Therefore, we are very eager to get feedback on the first version from treebank developers, parsing researchers and system developers alike. Universal Dependencies is an open project and anyone is invited to contribute by developing guidelines for a new language, contributing treebank data, or just providing feedback on existing guidelines and treebanks.

4 Word Segmentation

It is hard to give exact criteria for distinguishing syntactic words in all languages, but the basic idea is that a syntactic word can be assigned a single consistent morphological description with a unique lemma, part-of-speech tag and morphological feature set, as well as a single syntactic function in relation to other words of the sentence. A consequence of this characterization is that clitics normally need to be separated from their hosts. For example, in the Spanish orthographic word *dámelo* (give it to me), there are three different parts of speech (verb, pronoun, pronoun) and three different syntactic functions (predicate, indirect object, direct object). Hence, it should be split into three separate words: *da*, *me*, *lo*. Similarly, for a contraction like the French *au*, we need to postulate two words *à* and *le* with different parts of speech (adposition, determiner) and syntactic functions (case marker, determiner). In principle, the word-based view could also be taken to imply that certain fixed multiword annotations should be treated as single words in the annotation. So far, however, multiword expressions are annotated using special dependency relations, instead of collapsing multiple tokens into one, which has the additional advantage that we can accommodate discontinuous multiword expressions.

Since word segmentation in general is a non-trivial task in many languages, and since the usefulness of tools trained on treebank data ultimately depends on how well the word segmentation can be reproduced for new data, it is important to document the principles of word segmentation for each language. The nature of this documentation will vary from one language to the next, depending on properties of the language and the writing system. For languages where word segmentation can be performed by a simple script given white-space and punctuation, only the words need to be represented in the treebank. For languages not using white-space at all, such as Chinese and Japanese, a complex word segmentation algorithm has to be employed, but there is no need to represent the basic character sequence in the treebank since it is completely recoverable from the word representation. By contrast, in languages where the mapping between white-space delimited tokens and syntactic words is highly ambiguous, such as Arabic and Hebrew, we provide the option of including both tokens and words in the treebank using a two-level indexing scheme. The morphological and syntactic annotation is only defined at the word level, but a heuristic mapping to the token level can usually be provided. The language-specific documentation for each treebank must describe how word segmentation has been performed,

Table 1. Morphological annotation: universal part-of-speech tags and features

Part-of-Speech Tags		Features	
ADJ	adjective	Animacy	animacy
ADP	adposition	Aspect	aspect
ADV	adverb	Case	case
AUX	auxiliary verb	Definite	definiteness or state
CONJ	coordinating conjunction	Degree	degree of comparison
DET	determiner	Gender	gender
INTJ	interjection	Mood	mood
NOUN	noun	Negative	if a word is negated
NUM	numeral	NumType	numeral type
PART	particle	Number	number
PRON	pronoun	Person	person
PROPN	proper noun	Poss	possessive
PUNCT	punctuation	PronType	pronominal type
SCONJ	subordinating conjunction	Reflex	reflexive
SYM	symbol	Tense	tense
VERB	verb	VerbForm	form of verb
X	other	Voice	voice

whether the treebank includes (multiword) tokens as well as words, and what types of white-space separated tokens are split into multiple words (if any).

5 Morphological Annotation

The morphological description of a word consists of three components:

- A lemma representing the semantic content of the word.
- A part-of-speech tag representing the abstract lexical category of the word.
- A set of features representing lexical and grammatical properties associated with the lemma or the particular word form.

Lemmas are typically determined by language-specific dictionaries. By contrast, the part-of-speech tags and features are taken from two universal inventories. The list of universal part-of-speech tags is a fixed list containing 17 tags shown in Table 1 (left). It is based on the Google Universal Part-of-Speech Tagset [10], which in turn is based on a generalization over tagsets in the CoNLL-X shared task on multilingual dependency parsing [19]. In the new version, the category VERB has been split into AUX and VERB, NOUN into NOUN and PROPN, and CONJ into CONJ and SCONJ; the two new categories INTJ and SYM have been added; and the category PRT has been renamed PART to dissociate it from the label commonly used for verb particles because the universal tag covers a larger class of grammatical particles. The universal tags must be used in all UD treebanks. Some tags may not be used in all treebanks, but the list

cannot be extended with language-specific categories. Instead, more fine-grained classifications can be achieved via the use of features.²

Features give additional information about the word, its part of speech and morphosyntactic properties. Every feature has the form Name=Value and a word can have any number of features.³ Table 1 lists our current set of universal features, which are all attested in multiple corpora and need to be standardized. The list is certainly not exhaustive and later versions of the standard may include new features or values found in new languages, corpora or tagsets. Users can extend the set of universal features with language-specific features as needed. Such features must be described in the language-specific documentation and follow the general format principles. In addition to simple features, we also provide a mechanism for *layered* features in cases where the same feature is marked more than once on the same word. This happens, for instance, in the case of possessives that agree with both the possessor and the possessed.

6 Syntactic Annotation

The syntactic annotation consists of typed dependency relations between words, with a special relation *root* for words that do not depend on any other word. Every sentence is associated with a set of *basic* dependencies that form a rooted tree representing the backbone of the syntactic structure. Many grammatical constructions introduce additional dependencies, which can be represented in an *enhanced* dependency structure, which is a general directed graph. Examples of such constructions are secondary predication, control structures and dependencies that need to be propagated over coordination structures. The guidelines for the enhanced structure are still under development and will not be discussed further in this paper.

The universal dependency relations are meant to capture a set of broadly observed grammatical functions that work across languages. More precisely, we want to maximize parallelism by allowing the same grammatical relation to be annotated in the same way across languages, while making enough crucial distinctions to differentiate constructions that are not the same. As a fundamental principle we assume that dependency relations hold primarily between content words, rather than being indirect relations mediated by function words. This principle is illustrated in Figure 2, where the solid arcs represent direct dependencies between content words.

Given the dependency relations between content words, function words attach as direct dependents of the most closely related content word (dashed arcs), while punctuation marks attach to the head of the clause or phrase to which they belong (dotted arc). Preferring content words as heads maximizes parallelism between languages because content words vary less than function words

² In addition, the CoNLL-U format allows the inclusion of language-specific tags on top of the universal ones.

³ For readability, Figure 1 displays multiple features on top of each other, whereas the CoNLL-U format uses a vertical bar to separate them: Gender=Fem|Number=Plur.

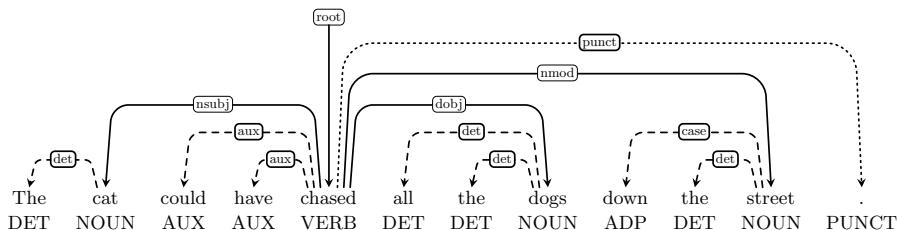


Fig. 2. Dependency tree for an English sentence (lemmas and features omitted)

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). Therefore, we treat adpositions as dependents of nouns, rather than the other way round, because they often correspond to case markers or nothing at all in other languages. We also treat auxiliary verbs as dependents of main predicates, with the copula as a dependent of a nominal or adjectival predicate as a special case.

We assume the taxonomy of the Universal Stanford Dependencies [13], which posits the 40 syntactic relations listed in Table 2.⁴ The main organizing principle of this taxonomy is the distinction between three types of syntactic structures:

- Nominal phrases, primarily denoting entities but also used for other things.
- Clauses headed by a predicate, usually a verb but sometimes an adjective, an adverb, or a predicate nominal.
- Miscellaneous other kinds of modifier words, which may allow modifiers but which do not expand into rich structures like nominal phrases and clauses.

This distinction is reflected in two dimensions in the upper part of Table 2, where the three columns represent *dependents* of all three types, while rows represent different types of constructions where the *head* is either a clausal predicate or a nominal. The taxonomy also distinguishes between core arguments (subjects, objects, clausal complements) and other dependents, but it makes no attempt to distinguish adjuncts from oblique arguments. The latter distinction has proven notoriously difficult to draw in practice and is often omitted in treebank annotation schemes, notably that of the Penn Treebank.

The first row in Table 2 shows relations for core arguments of predicates, with one series for nominal arguments (*nsubj*, *nsubjpass*, *dobj*, *iobj*) and one for clausal arguments (*csubj*, *csubjpass*, *ccomp*, *xcomp*). For subjects, we differentiate canonical voice (*nsubj*, *csubj*), where the proto-agent argument is the subject, from non-canonical voice (*nsubjpass*, *csubjpass*), where another argument is the

⁴ The current UD inventory of syntactic relations differs from that described in [13] by omitting the relations *nfincl*, *relcl* and *nmod* and adding the relation *acl*.

Table 2. Universal dependency relations. Dependents of predicates are divided into core (top), non-core (middle), and special (bottom) dependents. Dependents marked * are auxiliary verbs rather than real predicates. Note that some relations occur in more than one place.

	Nominal Dep	Predicate Dep	Other Dep
Predicate Head	nsubj	csubj	
	nsubjpass	csubjpass	
	dobj	ccomp	
	iobj	xcomp	
	nmod	advcl	advmod neg
	vocative	aux*	mark
	discourse	auxpass*	punct
	expl	cop*	
	dislocated		
Nominal Head	nummod	acl	amod
	appos		det
	nmod		neg
			case
No Head	root		
	Compounding	Coordination	Other
	compound	conj	list
	name	cc	parataxis
	mwe	punct	remnant
	goeswith		reparandum foreign dep

subject. For clausal complements, we differentiate clauses with obligatory control (*xcomp*) from clauses with other types of subject licensing (*ccomp*), but we do not differentiate finite from nonfinite clauses.

The second row in Table 2 shows the relations for non-core dependents of predicates, which differentiates nominal, clausal and other dependents:

John talked [in the movie theatre] (*nmod*)
 John talked [while we were watching the movie] (*advcl*)
 John talked [very quickly] (*advmod*)

The third row contains special dependents of the predicate, including function words like auxiliary verbs (*aux*, *auxpass*, *cop*), complementizers (*mark*), and punctuation (*punct*), as well as dependents that are more loosely connected to the predicate, such as vocatives and discourse particles. The fourth row contains dependents of nominal heads, again divided into three structural classes. Finally, we have the *root* relation, which is used for independent words, usually the predicate of a main clause.

The lower part of Table 2 displays relations that can occur with (almost) any type of head and that do not necessarily correspond to traditional grammatical relations. The first class covers lexical relations like compounding (*compound*), which we take to be fundamentally different from phrasal modification, fixed multiword expressions (*mwe*), and complex names that lack compositional structure (*name*). The second class is concerned with coordination, which is analyzed as an asymmetrical relation, where the first conjunct is the head on which all other conjuncts depend via the *conj* relation. Coordinating conjunctions and punctuation delimiting the conjuncts are attached using the *cc* and *punct* relations, respectively.

The primacy of content words implies that function words normally do not have dependents of their own. In particular, it means that multiple function words related to the same content word always appear as siblings, never in a nested structure, regardless of their interpretation. Typical cases in Figure 2 are auxiliary verbs (*could have*) and multiple determiners (*all the*). One possible interpretation of these flat structures is that they constitute dissociate nuclei in the sense of Tesnière [20], rather than regular dependency structures, and that the function words modify the syntactic category of their head, rather than performing a grammatical function in relation to a nominal or predicate. Nevertheless, there are a few exceptions to the rule that function words do not take dependents:

- Multiword function words: The word forms that make up a fixed multiword expressions are connected into a head-initial structure using the special dependency relation *mwe*. When the multiword expression is a functional element, the initial word form will then superficially look like a function word with dependents. Typical examples are *in spite of*, *because of*, *by* and *large*.
- Coordinated function words: Head coordination is a syntactic process that can apply to almost any word category, including function words like conjunctions and prepositions. In line with the general analysis of coordination, the first conjunct will then be the head of both the conjunction and the second conjunct, regardless of whether it is a function word or a content word. Examples: *to and from*, *if and when*.
- Promotion by head elision: When the natural head of a function word is elided, we “promote” the function word to the function normally assumed by the content word head (instead of introducing a null node representing the head). Typical examples are:

Bill could not answer, but Ann could. [*conj*(answer, could)]
 She forgot which address she wrote to. [*nmod*(wrote, to)]
 I know how. [*ccomp*(know, how)]

In addition, certain types of function words can take a restricted class of modifiers, mainly negation and light adverbials. Typical cases are modified determiners like *not every* and *exactly two* as well as modifiers of subordinating conjunctions *right when*.

In addition to the basic universal dependencies, it is always possible to add language-specific subtypes for constructions that are of special significance in a given language. These relations have the form *uni:spec*, where *uni* is one of the 40 universal relations, and *spec* is a descriptive label. Commonly used subtypes in the first release are *acl:relcl*, for relative clauses as a subtype of clauses modifying nouns, and *compound:prt* for verb particles. Language-specific subtypes of the universal relations must be described in the language-specific documentation.

7 Conclusion

I have presented Universal Dependencies, a recent initiative to create guidelines for cross-linguistically consistent grammatical annotation. So far, a first version of the guidelines has been released, as well as a first batch of treebanks for ten languages: Czech, English, Finnish, French, German, Hungarian, Irish, Italian, Spanish and Swedish.

In order to increase the usefulness of these resources, there are two important challenges for the future. The first is to improve the quality of the annotated treebanks with respect to real cross-linguistic consistency, as opposed to merely notational consistency. The second is to expand the coverage of languages and increase the typological diversity. This will require contributions from the entire treebank and parsing community, so we invite anyone who is interested to take part in this development.

It still remains to be seen whether we will ever manage to construct something that deserves to be called a universal grammar for natural language processing. After all, the quest for a real universal grammar is still on almost 800 years after Bacon's initial observation. So even if our goals are more modest, we may need another decade or two to figure it out.

Acknowledgments. Sections 4–6 of this paper are based on the UD guidelines, which is collaborative work with Jinho Choi, Marie-Catherine de Marneffe, Tim Dozat, Filip Ginter, Yoav Goldberg, Jan Hajič, Chris Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Dan Zeman. I also want to thank Masayuki Asahara, Cristina Bosco, Binyam Ephrem, Richárd Farkas, Jennifer Foster, Koldo Gojenola, Hiroshi Kanayama, Jenna Kanerva, Veronika Laippala, Alessandro Lenci, Teresa Lynn, Anna Mäsilä, Verginica Mititelu, Yusuke Miyao, Simonetta Montemagni, Shinsuke Mori, Petya Osenova, Prokopis Prokopidis, Mojgan Seraji, Maria Simi, Kiril Simov, Aaron Smith, Takaaki Tanaka, Sumire Uematsu, and Veronika Vincze, for contributions to language-specific guidelines and/or treebanks. The views expressed in this paper on the status and motivation of UD are primarily my own and are not necessarily shared by all contributors to the project.

References

1. Nolan, E., Hirsch, S. (eds.): *The Greek Grammar of Roger Bacon and a Fragment of his Hebrew Grammar*. Cambridge University Press (1902)
2. Brekle, H.E., Lancelot, C., Arnauld, A.: *Grammaire générale et raisonnée, ou La Grammaire de Port-Royal*. Friedrich Frommann Verlag (1966)
3. Chomsky, N.: *Aspects of the Theory of Syntax*. MIT Press (1965)
4. Chomsky, N.: *Cartesian Linguistics*. Harper and Row (1965)
5. Tsarfaty, R., Seddah, D., Goldberg, Y., Kuebler, S., Versley, Y., Candito, M., Foster, J., Rehbein, I., Tounsi, L.: Statistical parsing of morphologically rich languages (spmrl) what, how and whither. In: *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, pp. 1–12 (2010)
6. Tsarfaty, R.: A unified morpho-syntactic scheme of Stanford dependencies. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 578–584 (2013)
7. Naseem, T., Barzilay, R., Globerson, A.: Selective sharing for multilingual dependency parsing. In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 629–637 (2012)
8. Täckström, O., McDonald, R., Nivre, J.: Target language adaptation of discriminative transfer parsers. In: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, pp. 1061–1071 (2013)
9. Zeman, D.: Reusable tagset conversion using tagset drivers. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC)*, pp. 213–218 (2008)
10. Petrov, S., Das, D., McDonald, R.: A universal part-of-speech tagset. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)* (2012)
11. Zeman, D., Mareček, D., Popel, M., Ramasamy, L., Štěpánek, J., Žabokrtský, Z., Hajič, J.: HamleDT: To parse or not to parse? In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pp. 2735–2741 (2012)
12. McDonald, R., Nivre, J., Quirmbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Bertomeu Castelló, N., Lee, J.: Universal dependency annotation for multilingual parsing. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 92–97 (2013)
13. de Marneffe, M.C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., Manning, C.D.: Universal Stanford Dependencies: A cross-linguistic typology. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*, pp. 4585–4592 (2014)
14. Butt, M., Dyvik, H., Holloway King, T., Masuichi, H., Rohrer, C.: The parallel grammar project. In: *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pp. 1–7 (2002)
15. Bender, E.M., Flickinger, D., Oepen, S.: The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In: *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pp. 8–14 (2002)

16. Buchholz, S., Marsi, E.: CoNLL-X shared task on multilingual dependency parsing. In: Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL), pp. 149–164 (2006)
17. Blevins, J.P.: Word-based morphology. *Journal of Linguistics* 42, 531–573 (2006)
18. Mel'čuk, I.: *Dependency Syntax: Theory and Practice*. State University of New York Press (1988)
19. McDonald, R., Nivre, J.: Characterizing the errors of data-driven dependency parsing models. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 122–131 (2007)
20. Tesnière, L.: *Éléments de syntaxe structurale*. Editions Klincksieck (1959)