FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

HABILITATION THESIS

Daniel Zeman

Cross-Language Harmonization of Linguistic Resources

Institute of Formal and Applied Linguistics (ÚFAL)

Prague 2023

Title: Cross-Language Harmonization of Linguistic Resources

Author: Daniel Zeman

Institute: Institute of Formal and Applied Linguistics (ÚFAL)

Abstract: The presented work consists of two parts. In the first part I summarize the main directions of my research since the defense of my PhD thesis in 2005. I start with cross-language transfer of parsing models to languages that have little or no annotated data. This section provides motivation for the subsequent sections, which revolve around designing a description of natural language systems that could be used for any language, leading to data resources that are interoperable and comparable cross-linguistically. The harmonization efforts culminate in the international project called Universal Dependencies (UD), to which I have contributed significantly. Finally, I discuss some more recent spin-offs from Universal Dependencies, showing the current and future directions of my research work.

The second part contains a selection of my publications from the same period. Each publication is accompanied with a comment that puts it in context and assesses its long-term impact. The publications in the second part are directly related to the individual topics in the first part and I highlight these connections using cross-references in both ways.

Keywords: annotated corpora; morphology; dependency syntax; delexicalized parsing; shared task

Contents

Introduction

This thesis focuses on linguistic resources for many different languages. More specifically, it focuses on corpora that are annotated both morphologically and syntactically; since syntactic structure is typically expressed as a rooted tree, such corpora are called *treebanks*. They are invaluable resources for the study of language systems and, more generally, for digital humanities. For several decades it was also assumed that morphosyntactic analysis is an essential first step towards any application that assumes computational understanding of natural language, including machine translation. This assumption has now been drastically reduced by the advances of deep learning models, which can be tuned for the end-user task and can capture morphology and syntax internally, without seeing corresponding human-made annotation; however, such models do not reveal how they arrived at the output they were asked for and, consequently, they do not bring much insight about the language itself. In contrast, some insight about the language system can be obtained if morphosyntactic analysis is taken as the target task and a model *(a parser)* is trained on a human-annotated treebank to predict the annotation for previously unseen data. (Note that deep learning still plays a role, now in solving the parsing task.) Furthermore, morphosyntactically parsed text is useful as input for heuristics solving downstream tasks whenever there is not enough training data in the given language annotated directly for those tasks.

Morphological annotation, as understood in the present thesis, consists of three main pieces of information: the lemma of a word, its part-of-speech (POS) category, and a set of morphological feature-value pairs that characterize the annotated word form within an inflectional (or derivational) paradigm. Not all treebanks separate the POS category and the features in the way we just did here; part of speech itself can be (and often is) viewed as another feature with a predefined set of possible values. Depending on the terminology used by individual authors, the lemma is then accompanied by a *POS tag* or a *morphological tag*, which is a more or less compact encoding of the feature-value pairs.

Tagsets come with different expectations about how much can and should be disambiguated by context. For example, the English word *can* is either a modal auxiliary (as in *I can give you a ride*), or a noun (as in *I have a can full of fruit*). We can also derive a verb from the noun (as in *How to can fruits*). The surface ambiguity between the first *can* and the other two is purely coincidental and we definitely want to disambiguate them in text. The second and third *can* are related, one is derived from the other, but we still want to distinguish them because the syntactic rules applying to nouns and verbs are not compatible [Zeman, 2018].

Many different standards have been proposed for morphological tagging. Some differences are differences between languages; but even within one language, tagsets vary substantially in their level of granularity and choice of phenomena to capture. Table 1 demonstrates this on the example of tags denoting adjectives.

The syntactic structure of a sentence can be annotated in various ways, depending on the underlying theory. Most frameworks represent the sentence hierarchically as a rooted directed tree. In the present thesis we focus on *dependency trees*, whose nodes correspond (mostly) to words, and edges connecting them are

Language	Tagset	Tag
English	Penn Treebank	`JJ`, `JJR`, `JJS`
Swedish	Mamba	`AJ`
Swedish	Stockholm-Umeå	`JJ|POS|UTR|SIN|IND|NOM`
		`JJ|POS|UTR|SIN|IND|GEN`
		`JJ|POS|UTR|SIN|DEF|NOM`
		…
Czech	Prague Dependency Treebank	`AAMS1----1A----`
		`AAMS2----1A----`
		`AAMS3----1A----`
		…

Table 1: Morphological / POS tag examples for various languages. The tags for adjectives as defined in the Penn Treebank [Marcus et al., 1993], Mamba [Teleman, 1974, Nilsson et al., 2005], Stockholm-Umeå Corpus [Gustafson-Capková and Hartmann, 2006, p. 20–21], and the Prague Dependency Treebank (PDT) [Hajič et al., 2000]. The three PDT tags represent only a fraction; as many as 378 feature combinations are possible in a regular adjective paradigm. Stockholm-Umeå is less rich, but still it has many more tags than the three displayed here.

typed dependencies. Usage of such structures in linguistics dates back to the seminal work of Tesnière [1959], and a number of dependency-syntactic theories evolved since then; therefore, narrowing syntactic annotation to dependency trees itself does not ensure that there is a single set of annotation rules that everyone uses. To illustrate this, we show two annotations of the English sentence *I saw the man who loves you* in Figure 1, one following the annotation guidelines of the Prague Dependency Treebank (henceforth Prague Dependencies, PD) [Hajič et al., 2000], and the other following Stanford Dependencies (henceforth SD) [de Marneffe et al., 2014]. Topologically, the sentence receives in both frameworks identical structure, but the labels of the dependency relations differ. Nevertheless, the tree shapes may differ, too, as we demonstrate on the sentence *Bell, based in Los Angeles, makes electronic and building products* (Figure 2). Note that in this case SD does not even treat all words as nodes (the function words *in* and *and* are reflected as parts of the dependency relation types `prep_in` and `conj_and`, respectively, but they are not nodes).

Structure of the Thesis

The thesis consists of two parts. In the first part, I summarize the main directions of my research from 2006 to the present. I start in Chapter 1 with cross-language transfer of parsing models to languages with little or no annotated resources. This provides motivation for cross-linguistic harmonization of data resources, the topic of Chapter 2 (morphological harmonization) and Chapter 3 (syntactic harmonization). Chapter 4 returns to parsing and discusses several shared tasks that took advantage of harmonized data. Finally, Chapter 5 discusses some recent projects and future directions based on the work described in the previous

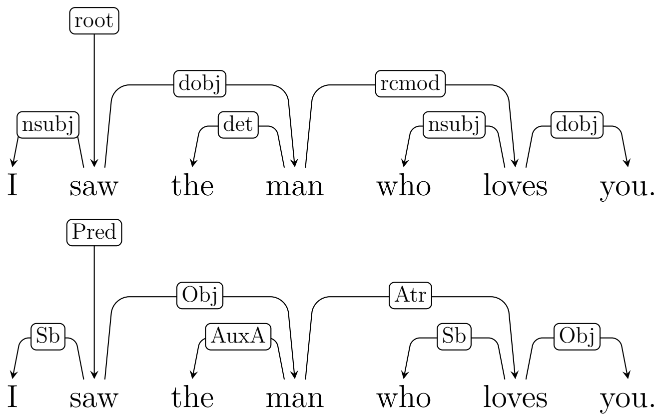


Figure 1: The sentence *"I saw the man who loves you"* in SD (up) and PD (down). Adapted from de Marneffe et al. [2006].

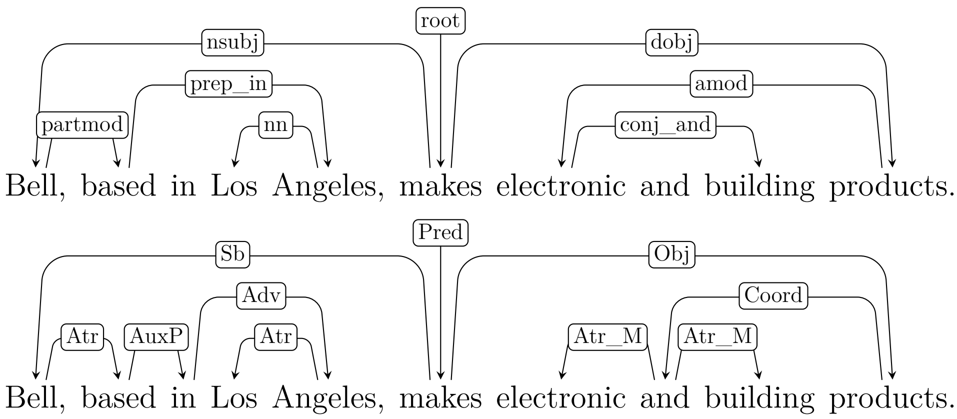


Figure 2: The sentence *"Bell, based in Los Angeles, makes electronic and building products"* in SD (up) and PD (down). Adapted from de Marneffe and Manning [2008].

chapters.

The second part (Chapter 6) is a selection of my publications directly related to the content of the first part. Most of the selected publications are joint work with other researchers, which is typical in the field. It goes without saying that I selected only papers where my contribution was essential.

1. Low-resource Languages

1.1 Dependency Parsing

The task of predicting the dependency tree structure for a previously unseen sentence is called **dependency parsing.** Nowadays it typically includes predicting the label (type) for each dependency relation, that is, for each word the **parser** must identify the word that should serve as its parent node, and the type of the relation between the two words. The parser is usually a model trained on manually annotated **('gold standard')** data. The performance of a parser is evaluated on test (evaluation) data, which is separate from training data. The parser is applied to unannotated **('blind')** version of the test data, and the parser's output is then compared to manually annotated version of the same data. The most widely used evaluation method is the Labeled Attachment Score **(LAS)** – we count a word as correct if both its parent and the dependency type have been predicted correctly, and we compute LAS as the percentage of correct words among all words[1] in the test data. In situations where prediction of the labels is considered uninteresting or too difficult, Unlabeled Attachment Score **(UAS)** is used instead. It counts a word as correct if its parent has been identified correctly, ignoring the dependency label.

In my PhD thesis [Zeman, 2004], I explored dependency parsing of Czech. My parser[2] was not only result of several years of my own work; it also rested on the shoulders of a large team of colleagues who had spent over five years designing annotation rules and annotating 70 thousand Czech sentences on multiple levels. It struck me that the Czech language was very lucky to have such rich computational resources, far exceeding most languages of the world (including languages with far more speakers). Regardless that I tried to keep my parsing algorithm as language-agnostic as possible, I could not apply it to most languages simply because there was no training data. The situation has improved since then, but the problem of **low-resource languages** has not disappeared and it is not going to disappear any soon. There are thousands of natural languages in the world [Dixon, 2010, p. xiii] and if we now have about 100 languages with decent treebanks, there are still thousands of languages that lack them. I became interested in language processing that could be applied to many languages, including those that possess little or no hand-annotated data. I started to explore techniques of parsing a low-resource language B, taking advantage of better-resourced, related language A. For instance, could we build a reasonably performing parser for Slovak, given that it is very close to Czech, and while Slovak did not have any treebank, there was so much data available for Czech?

[1] Most implementations of LAS work with all nodes, i.e., not only actual words, but also punctuation symbols and other tokens.

[2] With $UAS = 74.7\%$ on the d-test data of PDT 1.0 I fell significantly behind the state of the art (84.3%), but in combination with other parsers, my parser contributed to the new SotA $UAS = 85.5\%$.

1.2 Delexicalized Parsing

The technique I developed[3] [Zeman and Resnik, 2008] (Section 6.1) was based on four simple assumptions:

- It is easier and thus cheaper to obtain gold-standard data with morphological tags than with syntactic structures.
- Languages that are related are likely to have similar syntactic structures, even if their lexical forms differ.
- A model can predict the syntactic structure reasonably well with only morphological tags (but not the actual word forms) as input.
- The sets of morphological tags for the related languages are mutually compatible.

We did not attempt to quantitatively evaluate the first assumption but it seemed quite intuitive, and it was supported by the existence of tagged corpora for many languages for which no treebank was available.

As for the second assumption, there are varying levels of relatedness. An obvious candidate is the genealogic relationship, with Czech being most closely related to other West Slavic languages (Slovak, Upper Sorbian and Polish), then to other Slavic languages, then to Baltic languages, then to other Indo-European languages. Languages can be *typologically* related because of common ancestry, but also because of geographic proximity and mutual interaction; for example, Bulgarian and Macedonian are in some aspects closer to Greek or Romanian than to other Slavic languages. But even distant languages may share some common traits, such as nouns being typical subject dependents of verbs.

To illustrate this, consider the sentence *My daughter tasted strawberry ice cream yesterday* in four Slavic languages (Figure 1.1). The Czech and Slovak versions are very close, even with half of the words identical. Ukrainian uses different words (and script) but the syntactic structure, as well as the sequence of part-of-speech tags is still the same. Polish slightly diverges from the other three languages in preferring the post-nominal position of the adjectival attribute; with that exception, its surface order mimics the other languages, and its dependency tree is still isomorphic with theirs.

A parsing model that relies on word forms can hardly be trained on one language and successfully applied to another – even the 50% of unknown words in Slovak could be devastating.[4] However, if the parser can obtain most of the required information from part-of-speech tags, its Czech model will work just as well for the Slovak and Ukrainian sentence, and probably almost as well for Polish (we cannot rule out that it will predict the dependency of the 'misplaced'

[3] This research was done during my stay at the University of Maryland in 2006. I am grateful for the interesting interactions with the colleagues there, in particular with Philip Resnik, under whose mentoring I did the work. I also acknowledge the funding provided jointly by the Fulbright-Masaryk Fellowship and by the Office of Naval Research.

[4] This motivational example should not be taken as a proof of anything. We have not provided evidence that the out-of-vocabulary rate will stay 50% on a larger data sample; we are just suggesting that the rate is not negligible.

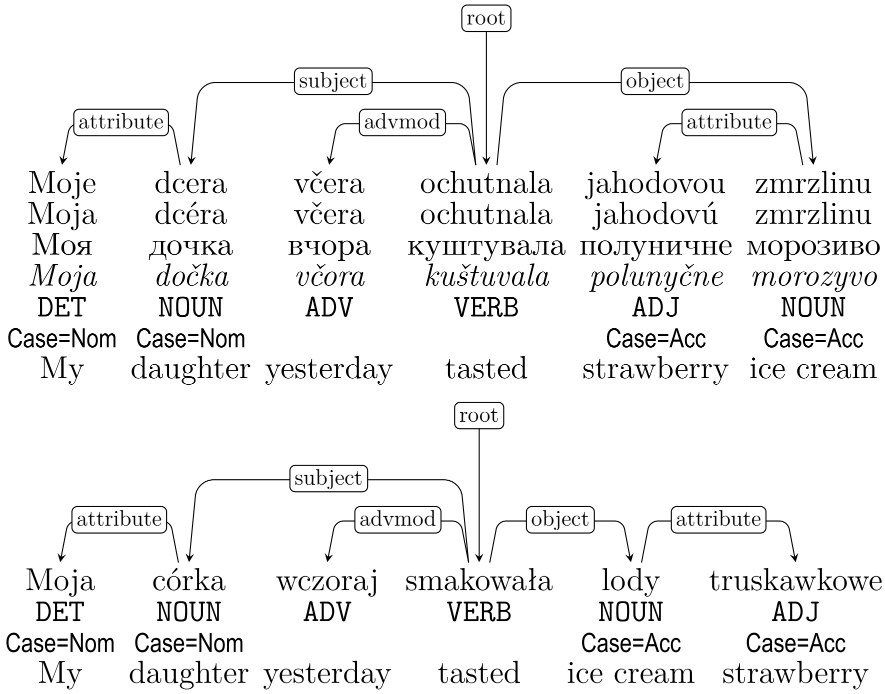


Figure 1.1: The sentence *"My daughter tasted strawberry ice cream yesterday"* in Czech, Slovak and Ukrainian (upper tree) and in Polish (lower tree).

adjective correctly). Even better if the tags are morphological, that is, if they reveal not only the part of speech but also the nominative case of words 1 and 2, and the accusative case of words 5 and 6.

The same part-of-speech sequence corresponds to many other sentences (or their parts) that have the same syntactic structure, for example

- `[cs]` *Tento sortiment také tvoří hlavní část [produkce společnosti]* "This assortment also forms the main part [of the company's production]"
- `[cs]` *[...] jejíž část zatím nemá hlasovací právo [...]* "[...] part of which does not yet have voting rights [...]"
- `[en]` *All offices also have free copies*
- `[pt]` *A direcção já mostrou boa vontade* "Management has already shown good will"
- `[zh]` 任何議員未曾作最後宣誓 *(Rènhé yìyuán wèicéng zuò zuìhòu xuānshì)* "No member has taken his final oath"

This leads us to the third assumption, namely that morphological information is a sufficient characterization of the input words for a parser. Of course, there may be other sentences with the same sequence of tags whose syntactic structure is different. It is also clear that there are cases that cannot be decided without understanding the lexical content, as in the Czech examples below, where *v Ústí* is a modifier of the university, while *v září* modifies the event, i.e., the verb.

- `[cs]` *Přestoupím na univerzitu v Ústí* "I will move to the university in Ústí"
- `[cs]` *Přestoupím na univerzitu v září* "I will move to the university in September"

The best way of testing the seriousness of this deficiency is to train a parser and evaluate it using the standard attachment score (see Section 1.4). We call a model that has been trained only on morphological tags, without any lexical information, a **delexicalized parser.**

Finally, there is the fourth assumption, which may not be obvious from the start, nevertheless it is very important: We need the tag sets for the languages in question to be compatible, that is, the same part of speech or morphological feature should be encoded the same way in every language. As demonstrated in Table 1, this is rarely the case; in fact, even within one language different corpora may use different tag sets. I will address this issue in Chapter 2.

Delexicalized parsing was later explored by many other authors. Most notably, McDonald et al. [2011] conducted large-scale experiments with delexicalized parser transfer among 9 Indo-European languages, and they also combined delexicalized parsing with part-of-speech tag projection across parallel data (see Section 1.3), removing the requirement that a tagged corpus be available in the target language. Aufrant et al. [2016] improved delexicalized parsing by adapting word order before training the model (cf. the word order difference between Polish and the other three languages in Figure 1.1).

More recently [Kondratyuk and Straka, 2019], parsers started using large multilingual neural language models to represent the words and their context. These models can also consider subwords (even individual characters), which allows them, e.g., to assess that the Czech adjective *jahodovou* and the Slovak *jahodovú* "strawberry" are equivalents. Such parsers can be viewed as occupying the middle ground between lexicalized and delexicalized. They have access to full lexical information, but they are also able to use it for an unknown word in a low-resource language if similarities can be observed on unannotated raw data.

1.3 Using Parallel Data

Other techniques that have been proposed for low-resource languages take advantage of **parallel texts,** that is, translations of the same text into multiple languages. They do not require that the text is annotated (specifically, morphological tags are not required). Again, the motivation is that unlabeled parallel texts are often available for pairs of languages where one language has rich annotated linguistic resources and the other does not. Indeed, there are many sources of such texts, ranging from multilingual legal documents (e.g., proceedings of the European Parliament) to open movie subtitles, to translations of the Bible.

Once a parallel corpus is available, unsupervised algorithms, well known from the machine translation field, can be used to first align sentences that are translations of each other, and then for each pair of parallel sentences compute the word alignment. The alignments provide links between elements of sentence structure, and these links can be used to project linguistic annotation from the resource-rich to the resource-poor language. As with delexicalized parsing, the techniques can be applied to any pair of languages, but better results are expected for languages that are closely related.

Training data projection. Run the source-language model on the source side of the parallel data, annotate it automatically. Project the annotation across word alignments to the target side of the parallel data. Train a target-language model on the now annotated target side of the parallel data.

Training data translation. Use the parallel data to obtain a simple word-to-word translation model. Apply it to the source-language annotated data. As a result, we have a 'translated' target-language corpus with exactly the same number of words, hence we can directly use the source-language annotation with the target-language word forms. Train a target-language model on the translated data. Of course, this technique makes sense only for closely related languages.

Test data translation. Use the parallel data to obtain a simple word-to-word translation model. Apply it to the target-language blind test data. Once 'translated' to the source language, we can apply the source-language model to annotate the data. Then the text can be 're-stuffed' with the original target words, and used for whatever purpose we needed the annotation. This resembles delexicalized parsing but instead of replacing the words with morphological tags, we replace the words with their equivalents in the other language.

Training data projection for part-of-speech tagging was first proposed by Yarowsky and Ngai [2001] and later refined by other authors. Das and Petrov [2011] used a word lattice in the target language to propagate tags to words that did not occur in the parallel data but were similar to words from the parallel data in that they preferred similar context. Agić et al. [2015] showed that part-of-speech projection is available for a large number of languages thanks to translations of the Bible. Mishra et al. [2017] experimented with "feature projection" for part-of-speech tagging of Indian languages. Their technique is similar to word-by-word translation of the training data.

Concerning dependency parsing, training data projection was proposed by Hwa et al. [2005]. In [Zeman and Resnik, 2008], we experimented with test data translation for dependency parsing and compared it to delexicalized parsing. The results we obtained spoke in favor of delexicalized parsing, but the translation approach fell not too far behind and it should not be ruled out for other datasets. Tiedemann [2014], Ramasamy [2014], Rosa [2018] compared the advantages and disadvantages of the projection and translation techniques. In 2017 our team won the shared task on similar language parsing [Rosa et al., 2017];[5] we used a variant of training data translation.

Annotation projection across parallel data has been applied even beyond surface syntax, for example to semantic roles that were projected from the English PropBank to several other languages [Jindal et al., 2022].

[5]The task consisted of parsing three target languages: Slovak (with Czech as the source language), Croatian (with Slovenian as the source), and Norwegian (with two source languages, Danish and Swedish). This shared task provided harmonized annotations for the languages in question.

1.4 Evaluation

The cross-lingual techniques outlined in the previous sections are useful if we do not have manually annotated data in the target language. However, in order to evaluate the performance of the techniques, we do need target gold-standard data. The evaluation is thus typically conducted on languages that possess annotated corpora, using those corpora only for evaluation, and hoping that the method would work similarly well when applied to a really resource-poor language. Once again, we need the annotation in the source and target languages to be compatible. If we are projecting parsing models, the compatibility requirement applies also to dependency trees – the rules for positing a dependency relation between two words, and the label (type) of the relation. None of that is granted (recall Figure 2); in fact, the opposite was the norm until about 2012.

The first CoNLL shared task in multilingual dependency parsing [Buchholz and Marsi, 2006] made available dependency treebanks of 13 languages.[6] The datasets were unified technically, using the same file format (later dubbed `CoNLL-X`), but their label sets were not harmonized, and neither were the linguistic decisions governing the dependency relations. On the other hand, the collection provided an opportunity to test cross-lingual transfer of parsers, as it included two closely related languages: Danish and Swedish.

The Danish data followed the annotation guidelines of the Danish Dependency Treebank [Kromann, 2002], while the Swedish data was taken from Talbanken [Nilsson et al., 2005]. These two treebanking schemes are very distant from each other. In [Zeman and Resnik, 2008], we employed various heuristics to make the annotations comparable; then we used Danish as the source language and Swedish as the target language. In contrast, McDonald et al. [2011] did not attempt to harmonize their data, and their results picture Danish as the worst possible source language for Swedish, among the eight European languages available.[7]

The actual attachment scores can be found in the respective papers cited here. They are not directly comparable, as they have been obtained on diverse datasets of various languages, and also with many different parsers (note that the delexicalization, projection and translation techniques can be used with any parser that can be trained on annotated data). Roughly speaking, one can expect around 65% UAS for closely related languages, meaning that two out of three words have the correct parent node. An interesting perspective to view this number is provided by a comparison with the learning curve of a fully supervised parser. The question we ask is: If manual annotations were available for the target language, how much of them would we need to train a parser that performs as well as our model transferred from the source language? Hwa et al. [2005] showed that their projection from English to Chinese corresponded to about 2000 Chinese gold-standard trees. The best Danish-based model from [Zeman and Resnik, 2008] ranked equal to a parser trained on 1546 Swedish sentences. I repeated the experiment in 2015 with more advanced parsers and better harmonized data. The UAS was still 66% but the learning curve was steeper, suggesting that the

[6] Not all the treebanks were available free of charge after the shared task.

[7] There were four other Germanic languages in the mix but none of them worked well, presumably also due to annotation divergences. The most helpful source, as evaluated on the Swedish data, turned out to be the Portuguese treebank.

same result can be obtained with just 75 Swedish sentences. Along the same lines, Ramasamy [2014, Table 6.6 on p. 100] found that with just 10 annotated training sentences, the UAS on his language set ranges from 57% (Bengali and Tamil) to 74% (Telugu) on in-domain target language data. Therefore, if a native speaker of the target language is available for a few days, the best technique might be to have the native speaker annotate a small sample of the target language. But this approach does not scale well to hundreds or thousands of target languages.

At any rate, we need annotations to be **harmonized** across languages in order to train and evaluate multilingual NLP tools, regardless of what particular approach we take. We will focus on harmonization in the following chapters.

2. Harmonization of Morphological Annotation

2.1 Interset

In Chapter 1, I stressed the necessity of working with corpora that have mutually compatible annotation. Specifically, for delexicalized parsing I needed a morphological tagset that could be applied to both the source and the target language. Since each of the available corpora used its own tagset, I had to either convert tags from tagset A to tagset B, or to define a hybrid tagset C covering features that are common to both corpora, and then convert A and B to C. While we described experiments with Danish and Swedish in [Zeman and Resnik, 2008], I conducted similar experiments with other language pairs, which means many different conversions had to be done. A typical conversion procedure is based on a large table or on a long sequence of `if-else` statements, and preparing it is tedious work. Therefore I was looking for ways how to reuse parts of the code written previously. Each conversion from tagset A to tagset B can be viewed as two steps done at once: understanding the information in tag A **(decoding)** and producing tag B that contains same or similar information **(encoding)**. If I separate the steps, I will be able to reuse them in the future when I encounter a new tagset C and need conversion from A to C, or from C to B. I will only have to implement the decoder and encoder for tagset C; then I can immediately convert tags between C and any previously covered tagset. I implemented this mechanism in Perl, and the Perl modules with encoders and decoders for individual tagsets were called **tagset drivers** [Zeman, 2008, 2018] (Section 6.2).

A crucial part of the conversion system is the intermediate feature structure where the information is stored between decoding from tagset A and encoding to tagset B. It functions as an Interlingua for morphological tagsets and I named it **Interset**.[1] Information from a morphological tag was decomposed and stored as a set of pre-defined morphological features (such as pos (part of speech), gender, number, tense) and one of their pre-defined values (such as pos=noun or tense=past). Interset turned out to be a useful framework for describing morphosyntax independently of individual corpora; as such, its significance grew beyond the engineering problem of preparing data for an experiment.

Conversion of a tag to a different tagset is often an information-losing process because the tag may make distinctions that the target tagset does not make. Nevertheless, we do not want to lose information during round-trip 'conversion' from a tagset to itself (i.e., when Interset is used as an internal data structure to easily access information about words, without the need to actually convert the tag). It may not be possible to capture all distinctions in a tagset because some of them may be too peculiar to deserve an Interset feature. Therefore, a decoder can always store additional data to a feature called other. The data is not expected to be understood by any other driver, hence Interset also remembers the identifier of the source tagset in the feature tagset. The encoder will consult

[1] By extension, 'Interset' also refers to the conversion software built around the data structure (`https://ufal.mff.cuni.cz/interset`).

the value of other only if it originates in the same tagset.

Interset was built bottom-up and new features or values were occasionally added when they were needed for newly added tagsets. If the existing feature-value pairs could not capture something in a new tagset, I had to assess whether it was worth adding a new feature (or value). If not, then it would be stored in other. In some cases, a feature was first stored in other but later revisited and made a regular Interset feature, when it was attested in another tagset.

In the current version, Interset covers 64 tagsets of 40 languages. It defines 63 features with 390 values in total. Some of the features are lexical, that is they pertain to the whole lexeme with all its morphological forms; they can be viewed as a finer partition of the part-of-speech space. Other features are inflectional, they describe the position of an inflected word form in the lexeme's inflectional paradigm. This classification is only approximate, for example, gender is lexical feature of Czech nouns but inflectional feature of Czech adjectives. However, the lexical-inflectional distinction serves only for orientation purposes and has no practical impact on work with Interset. Similarly, one could classify features as typically nominal (e.g., case) or typically verbal (e.g., tense), but many features would combine with multiple parts of speech, and plausible combinations would vary across languages (for example, Czech verbs do not inflect for case but some forms of Finnish verbs do).

Table 2.1 gives an overview of features and values in the current version of Interset together with a brief explanation of each feature.

Table 2.1: Interset features and their values.

pos	noun, adj, num, verb, adv, adp, conj, part, int, punc, sym	main part of speech
nountype	com, prop, class	special type of noun if applicable
nametype	geo, prs, giv, sur, nat, com, pro, oth, col, sci, che, med, tec, cel, gov, jus, fin, env, cul, spo, hob	named entity type
adjtype	pdt	special type of adjective: predeterminer
prontype	prn, prs, rcp, art, int, rel, exc, dem, emp, neg, ind, tot	pronominality and its type for nouns (pronouns), adjectives (determiners), numerals, adverbs
numtype	card, ord, mult, frac, sets, dist, range	numeral types; the main pos may be numeral, adjective, adverb
numform	word, digit, roman, combi	presentation form of numerals
numvalue	1, 2, 3	class of numeric values for numerals with special behavior
verbtype	aux, cop, mod, light, verbconj	special type of verb if applicable
advtype	man, loc, tim, sta, deg, cau, mod, adadj, ex	semantic type of adverb
adpostype	prep, post, circ, voc, preppron, comprep	special type of adposition if applicable
conjtype	coor, sub, comp, oper	conjunction type
parttype	mod, emp, res, inf, vbp	particle type

Continuation of Table 2.1

punctype	peri, qest, excl, quot, brck, comm, colo, semi, dash, root	punctuation type
puncside	ini, fin	distinction between opening and closing brackets and other paired punctuation
morphpos	noun, adj, pron, num, adv, mix, def	morphological part of speech – inflectional paradigm may behave like different pos than the word is assigned to
poss	yes	possessive word
reflex	yes	reflexive word
foreign	yes	foreign word
abbr	yes	abbreviation
hyph	yes	part of a hyphenated compound
typo	yes	incorrect form
echo	rdp, ech	reduplicated or echo word
polarity	pos, neg	polarity: affirmative or negative
definite	ind, spec, def, cons, com	definiteness and/or construct state
gender	masc, fem, com, neut	gender
animacy	anim, hum, nhum, inan	animacy
number	sing, dual, tri, pauc, grpa, plur, grpl, inv, ptan, coll, count	grammatical number
case	nom, gen, dat, acc, voc, loc, ins, abl, del, par, dis, ess, tra, com, abe, ine, ela, ill, ade, all, sub, sup, lat, per, add, tem, ter, abs, erg, cau, ben, cns, equ, cmp	grammatical case
prepcase	npr, pre	special case form after a preposition
degree	pos, cmp, sup, abs, equ, dim, aug	degree of comparison; also diminutives and augmentatives
person	0, 1, 2, 3, 4	person
clusivity	in, ex	inclusive vs. exclusive pronoun *we*
polite	infm, form, elev, humb	politeness, formal vs. informal word forms
possgender	masc, fem, com, neut	possessor's gender
possperson	1, 2, 3	possessor's person
possnumber	sing, dual, plur	possessor's number
possednumber	sing, dual, plur	possession's number; in Hungarian distinguished from main number and possessor's number
absperson	1, 2, 3	person of the absolutive argument of the verb (polypersonal agreement in Basque)
ergperson	1, 2, 3	person of the ergative argument of the verb (polypersonal agreement in Basque)

Continuation of Table 2.1

datperson	1, 2, 3	person of the dative argument of the verb (polypersonal agreement in Basque)
absnumber	sing, dual, plur	number of the absolutive argument of the verb (polypersonal agreement in Basque)
ergnumber	sing, dual, plur	number of the ergative argument of the verb (polypersonal agreement in Basque)
datnumber	sing, dual, plur	number of the dative argument of the verb (polypersonal agreement in Basque)
abspolite	infm, form, elev, humb	politeness of the absolutive argument of the verb (polypersonal agreement in Basque)
ergpolite	infm, form, elev, humb	politeness of the ergative argument of the verb (polypersonal agreement in Basque)
datpolite	infm, form, elev, humb	politeness of the dative argument of the verb (polypersonal agreement in Basque)
erggender	masc, fem, com, neut	gender of the ergative argument of the verb (polypersonal agreement in Basque)
datgender	masc, fem, com, neut	gender of the dative argument of the verb (polypersonal agreement in Basque)
position	prenom, postnom, nom, free	position / usage of adjectives, determiners, participles etc.
subcat	intr, tran	subcategorization (transitive vs. intransitive)
verbform	fin, inf, sup, part, conv, vnoun, ger, gdv	finite verb vs. infinitive, supine, participle, converb, verbal noun, gerund, gerundive
mood	ind, imp, cnd, pot, sub, jus, prp, opt, des, nec, qot, adm	mood
tense	pres, fut, past, aor, imp, pqp	tense
voice	act, mid, pass, rcp, cau, int, antip, dir, inv	voice
evident	fh, nfh	evidentiality
aspect	imp, perf, prosp, prog, hab, iter	aspect (lexical or grammatical)
strength	weak, strong	strong vs. weak forms of adjectives or pronouns
variant	short, long, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c	variant form of the same lemma and paradigm slot

Continuation of Table 2.1

style	arch, rare, form, poet, norm, coll, vrnc, slng, expr, derg, vulg	style (either of the lemma, or standard vs. colloquial suffix of the same lemma)
tagset	e.g. cs:pdt	source tagset identifier (determines relevance of other)
other	any value, possibly structured	tagset-specific information that does not fit elsewhere

2.2 'Google' Universal POS Tags

A few years after the first version of Interset, a team from Google and Carnegie-Mellon University proposed a set of 12 universally applicable and universally needed, coarse-grained part-of-speech tags for use in NLP applications [Petrov et al., 2012]; this tagset became informally known as the 'Google' universal tagset. Their goal was to harmonize the encoding of the main categories of words, ignoring finer morphological distinctions. In Interset, they would approximately correspond to the eleven non-empty values of the pos feature.

The authors also offered mappings from 25 existing tagsets of 22 languages to the universal tagset. An important shortcoming of their approach in comparison to Interset was that their mappings often relied exclusively on the top-level part of the source tag. So, for example, they defined a tag for numerals (`NUM`), but the source tagset for Danish did not have numerals as a top-level category. Instead, they were treated as a subclass of adjectives and consequently, they would end up as `ADJ` in the universal tagset, although by looking at other parts of the Danish tag, one could actually tell apart numerals from adjectives. Some of these issues were fixed in later versions of the mapping tables.[2]

2.3 Universal Dependencies

Having one annotation standard that fits all languages and applications is obviously beneficial for natural language processing. Also obviously, having more than one standard reduces the benefit. On the morphological level, there were universal POS tags, Interset, and some older standardization attempts which I survey in [Zeman, 2008]. There were at least two harmonization efforts also on the syntactic level (more on that in Chapter 3). In 2014, we joined forces with colleagues from Uppsala University, Stanford University, Google, University of Turku, Bar-Ilan University and the Open University of Israel. Our goal was to take the best from the previous harmonization efforts and try to build one standard that would supersede them. The team included authors of the competing harmonization projects, which was one important ingredient for success. The name of the new framework, Universal Dependencies[3] [de Marneffe et al., 2021] (Section 6.4), refers to syntactic annotation, but the framework defines cross-linguistic annotation both for syntax and morphology.

[2] `https://github.com/slavpetrov/universal-pos-tags`
[3] `https://universaldependencies.org/`

Universal Dependencies (UD) uses an extended version of the Universal POS tagset, now also abbreviated UPOS, with 17 tags instead of the original 12 (the additions included `PROPN` for proper nouns, `AUX` for auxiliaries, `SCONJ` for subordinating conjunctions, `INTJ` for interjections, and `SYM` for symbols other than punctuation. Besides UPOS, the UD standard has morphological features. The core set of features and values, documented as "universal features", are taken from Interset.[4] UD corpora can extend that set with their own features if needed, and some of the remaining Interset features have been used this way. I continue to maintain the feature set within the UD project and occasionally propose language-specific features or values, when they are attested in multiple corpora, to be promoted to the universal features. This ensures that people working on new languages for UD will use those features if they apply to their language, following the objective that same things be annotated same way in all languages. Interset proper still exists as a tagset conversion tool and I keep it compatible with UD.

2.3.1 Layered Features

In some languages, some features are marked more than once on the same word. For example, possessive pronouns (also called possessive determiners or adjectives in various terminological systems) may have two independent values of gender and two independent values of number. One of the values characterizes the possessor, the other characterizes the possessee. The possessor's gender and number is something that we observe also with normal personal pronouns: for instance, the English 3rd-person pronouns distinguish singular and plural, and they also distinguish three genders in the singular *(he, she, it)* but not in the plural *(they)*. Likewise, the corresponding possessive pronouns have three genders in singular *(his, her, its)* but only one form in plural *(their)*. English does not mark the possessee's features morphologically, but other languages do.

Thus in Croatian, the 3rd person pronouns distinguish three genders and two numbers in the nominative case, but in the other cases and in the possessives, the singular masculine is often identical to the singular neuter, and the plural forms are mostly common for all three genders. In most cases, there are three distinct forms (Table 2.2). There are also possessive pronouns for three different categories of possessors: masculine/neuter singular *(njegov)*, feminine singular *(njezin)*,[5] and plural *(njihov)*. However, in Croatian the possessive pronouns behave like adjectives and agree in gender, number and case with the possessed (modified) noun. If the possessee is masculine singular, such as *pas* "dog", the possessive pronoun will acquire a masculine suffix: *njegov pas* "his dog", *njezin pas* "her dog", *njihov pas* "their dog". If the possessee is feminine singular, the form of the possessive changes and takes the feminine suffix: *njegova mačka* "his cat", *njezina mačka* "her cat", *njihova mačka* "their cat". Similarly for singular neuter (*njegovo polje* "his field"), plural masculine (*njegovi psi* "his dogs") etc.

We thus need tags that distinguish the ordinary agreement suffixes (i.e., the possessee's gender, number and case) from the possessor's gender and number,

[4] Only capitalization is changed, e.g. the Interset feature gender=masc is Gender=Masc in UD.

[5] In fact, there are two feminine possessive variants: *njezin* and *njen*. We disregard the latter here.

	Case	Sing Masc/Neut	Sing Fem	Plur Masc/Fem/Neut
Prs	Nom	*on/ono*	*ona*	*oni/one/ona*
Prs	Gen	*njega*	*nje*	*njih*
	Number Gender Case			
Poss	Sing Masc Nom	*njegov*	*njezin*	*njihov*
Poss	Sing Fem Nom	*njegova*	*njezina*	*njihova*
Poss	Sing Neut Nom	*njegovo*	*njezino*	*njihovo*
Poss	Plur Masc Nom	*njegovi*	*njezini*	*njihovi*
Poss	Plur Fem Nom	*njegove*	*njezine*	*njihove*
Poss	Plur Neut Nom	*njegova*	*njezina*	*njihova*

Table 2.2: The nominative and genitive forms of Croatian 3rd person pronouns, and the nominative forms of the corresponding possessive pronouns. The rows represent various genders and numbers of the possessee, while the columns represent genders and numbers of the possessor.

which is encoded in the stem. Universal Dependencies call this *layered features:* there are two layers of gender, and two layers of number. There is also a specific notation: if a word is annotated more than once with a feature, the layers must be identified by a predefined string given in square brackets. For instance, a masculine possessor would be annotated as Gender[psor]=Masc. One layer can be treated as default and given without layer name; in our example, the agreement gender would be annotated simply as Gender=Masc. Note that Interset did not have such a flexible mechanism and had to define a separate feature for each layer. For instance, UD's Gender[psor] corresponds to possgender in Table 2.1. Another example where layered features help is polypersonal agreement in languages like Basque: when morphology of a ditransitive verb concurrently refers to three arguments distinguished by the absolutive, ergative and dative case, Interset would encode the verbal agreement as absperson, ergperson and datperson, while the layers in UD would lead to Person[abs], Person[erg] and Person[dat].

2.4 UniMorph

For completeness I also briefly mention another project that tries to capture morphology across languages: UniMorph. It started independently of UD, shortly after the first version of UD was released [Sylak-Glassman et al., 2015]. It took a top-down approach, trying to survey the known morphological categories from typological literature and project them all to the schema even before they were actually seen in corpora. Fortunately, UniMorph did not lead to a new competition between standards of morphological annotation. I took the proposal into account when designing the second version of the UD guidelines in 2016 and adopted some features that had been defined in UniMorph but not in UD. The two frameworks use similar level of granularity, and although they do not align perfectly, most UniMorph features can be represented in UD without loss of information. UniMorph and UD are now overlapping communities that take care to minimize potential incompatibilities between the two schemas.

3. Harmonization of Syntactic Annotation

3.1 HamleDT

I showed some examples of diverging approaches to syntactic annotation in Figures 1 and 2 in the Introduction, and in Section 1.4, I reported on experiments where the benefits of close relationship between Danish and Swedish were negated by the differences in the annotation of Danish and Swedish data. In [Zeman and Resnik, 2008] I used simple transformation heuristics to make the Danish and Swedish treebanks more comparable. However, this was an ad-hoc solution that did not consider datasets of other languages and did not lead to harmonized annotations that other researchers could reuse. In 2011, I and several my colleagues from Charles University decided to find a more principled and far-reaching solution.

We first inventoried the various dependency treebanks that were available at that time, and studied their annotation styles. To demonstrate the differences, in Figures 3.1–3.6 I show the coordination structure *apples, oranges and lemons* annotated according to 6 different treebanking styles.[1]

We implemented a technical conversion to a common file format – we used the `CoNLL-X` format defined by Buchholz and Marsi [2006], which had already become a de-facto standard used by various NLP tools. The morphological tags were converted to Interset features and stored in the file. Then we implemented transformations of the dependency structures.

It was almost a rule that each treebank had its own annotation style. An exception to this rule was a group of about ten treebanks inspired by the Prague Dependency Treebank [Hajič et al., 2000]; their annotation styles were not identical but they were reasonably similar. Since PDT was the home product of our institute, we naturally based our common annotation scheme on PDT. We named the collection HamleDT[2] (Harmonized Multi-Language Dependency Treebank) [Zeman et al., 2014] (Section 6.3). Its first version [Zeman et al., 2012] covered 29 languages but we later expanded it to 36 languages.

3.2 Stanford Dependencies

Another dataset with common annotation scheme was made available by a team of researchers from Google and Appen [McDonald et al., 2013]. Its first version contained six languages: English and Swedish were conversions of datasets that we also had in HamleDT; Spanish, French and Korean were newly annotated texts collected from the web, and German combined a pre-existing treebank with new data from the web. A year later the collection was expanded to 11 languages.[3] The authors called it 'Universal Dependency Treebank'; to distinguish it from

[1] See Popel et al. [2013] for more details on coordination styles in treebanks.

[2] `https://ufal.mff.cuni.cz/hamledt`

[3] `https://github.com/ryanmcd/uni-dep-tb`

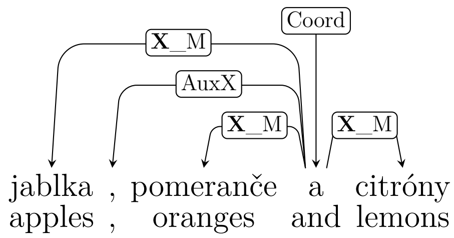


Figure 3.1: Coordination in the Prague style as seen in the Prague Dependency Treebank of Czech. **X** represents the relation between the coordination and its parent in the sentence.

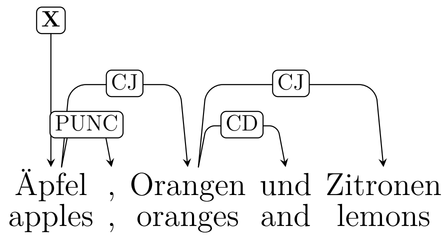


Figure 3.2: Coordination in the Mel'čukian style as seen in the Tiger treebank of German.

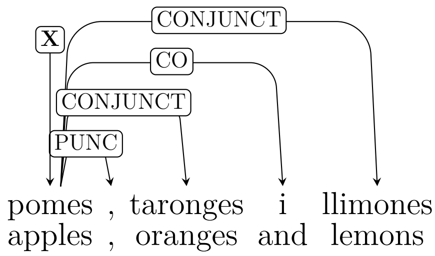


Figure 3.3: Coordination in the Stanford style as seen in the AnCora treebank of Catalan.

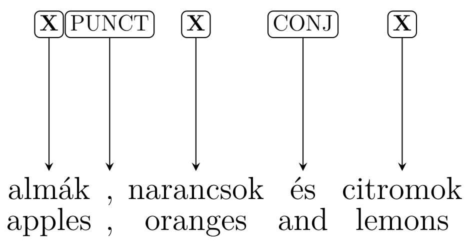


Figure 3.4: Coordination in the Tesnièrian style as seen in the Szeged Treebank of Hungarian. All participating nodes are attached directly to the parent of the coordination.

the Universal Dependencies project, it is sometimes informally dubbed 'Google' Universal Dependency Treebank. At the morphological level, it used the Google universal POS tags without additional features. At the syntactic level, they used a variant of Stanford Dependencies (SD) [de Marneffe et al., 2013]. As they said, the Stanford typed dependencies, partly inspired by the LFG framework, had emerged as a de-facto standard for dependency annotation in English and had

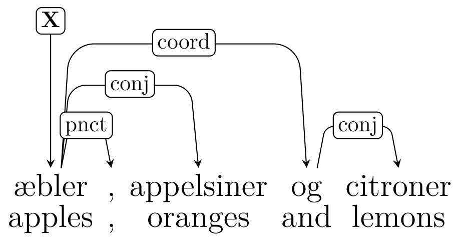


Figure 3.5: A mixture of Stanford and Mel'čukian coordination styles seen in the Danish Dependency Treebank.

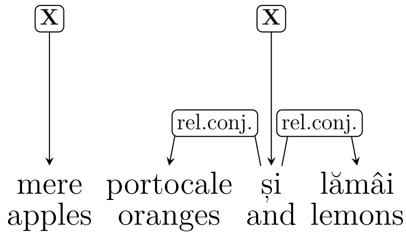


Figure 3.6: The Romanian treebank used Prague coordination style mixed with Tesnièrian because punctuation was missing from data.

then been adapted to several other languages; hence they decided to take SD as the point of departure for their representation.

The research group at Stanford University further developed their formalism to make it less biased towards English and more applicable to typologically diverse languages; the new proposal was called Universal Stanford Dependencies (USD) [de Marneffe et al., 2014]. In Prague, we noticed the growing popularity of Stanford-derived schemes and released HamleDT 2.0 with every treebank converted to two alternative schemes: Prague (based on PDT) and Stanford (based on USD) [Rosa et al., 2014].

3.3 Universal Dependencies

So in mid 2014 the problem of many diverging treebanks was replaced by the problem of several diverging standards, each of them hoping to solve the former problem. There were the Prague-style dependencies of HamleDT, and at least two flavors of the Stanford dependencies: the 'Google' flavor in the Google Universal Depenedncy Treebank, and the USD. In addition, there were Google UPOS and Interset on the morphological level. As I already outlined in Section 2.3, our ultimate answer to this muddle was Universal Dependencies [de Marneffe et al., 2021] (Section 6.4). In the present section I will focus on the syntactic aspects of UD. Unlike morphology, the syntactic part of the UD standard was not derived from my previous work. Nevertheless, as a founding member of the UD core group I contributed to its development, in particular to the formulation of the second version of the standard in 2016 [Nivre et al., 2020].

The syntactic structures in UD are based on a modification of the Universal Stanford Dependencies. Both USD and UD try to maximize parallelism in annotation of the same construction across languages. This naturally leads to

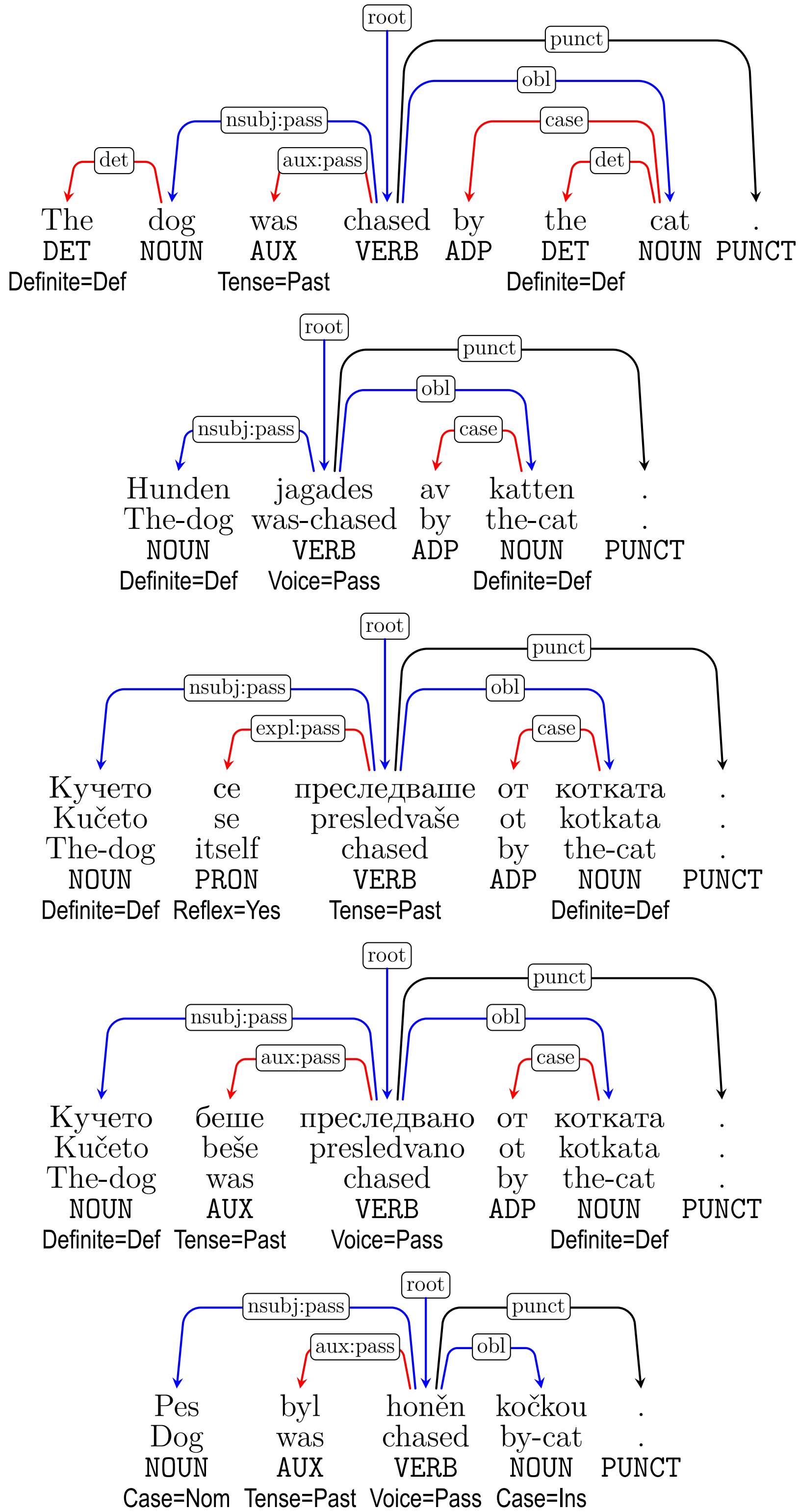


Figure 3.7: Parallel UD trees for the sentence *The dog was chased by the cat* in English, Swedish, Bulgarian (two versions) and Czech. Relations leading to content words are highlighted in blue, relations to function words in red and punctuation in black. Only selected features are shown.

preferring relations that place content words higher in the tree. Function words, which are more likely to vary across languages, are typically represented by leaf nodes. If we compare two languages where a function word in one language corresponds to a morphological feature in the other, the lexical backbones of the two trees stay parallel. This is demonstrated on the parallel sentences in Figure 3.7. The main meaning is expressed by the passive predicate *chase*, its subject *dog* and oblique agent *cat*; the relations between these three nodes are identical in all five trees. Relations attaching function words vary but they do not disrupt the main structure because their dependents are leaves. So in English there are separate nodes for the definite articles, while in Czech definiteness is not marked and in Swedish and Bulgarian it is marked directly on nouns. The oblique agent is marked by preposition in all languages but Czech, which uses the instrumental case (morphology). The passive voice is encoded with the help of an auxiliary in English, Czech and the second Bulgarian translation, by a reflexive pronoun in the first Bulgarian translation, and morphologically on the main verb in Swedish.

There were numerous typologically interesting constructions from many languages that we had to study when designing the UD guidelines. No doubt there are many others we will encounter as new languages and language families get covered by UD. I am not going to survey such constructions now because I have done so in [de Marneffe et al., 2021, § 4], which is incorporated in Section 6.4 of this thesis.

Universal Dependencies is a thriving project and community, which keeps growing and adding annotated resources for several new languages every year. In many cases UD literally helped to "put the language on the digital map." UD treebanks are used in natural language processing but also in various areas of digital humanities, in particular linguistics and linguistic typology. While UD treebanks are probably too small to study the language system, parsers trained on these treebanks can be used to process additional data, often with decent accuracy. UD includes quite a few classical languages such as Ancient Greek or Sanskrit, thus aiding historical studies. Diversity of the collection is further increased by fieldworkers who create treebanks while documenting endangered languages (for example, we have samples of 15 indigenous languages from South America). The success of UD may lay in various factors which are difficult to evaluate, but the crucial point is that we tried to balance different perspectives and needs, however conflicting they may be. We tried to make it linguistically adequate but still simple enough for non-linguists, we built it on de-facto standards, kept the guidelines relatively stable over time, and maintained a regular cycle of two releases per year. This, together with the supporting infrastructure, makes it easy for newcomers to start a treebank and see it become part of UD in relatively short time. And once UD became known in the NLP community, the snowball effect went off: People who did not see their language in UD decided to do something about it and started annotating data. That is why we now[4] cover 148 languages from 31 families and all parts of the world, the combined size of the treebanks exceeds 31 million words, it exists thanks to 577 contributors and it has cumulatively reached nearly 200 thousand downloads.

[4]UD release 2.13 from November 2023.

4. Multilingual Shared Tasks

It is a tradition in the field of natural language processing to organize evaluation campaigns – shared tasks – focused on concrete NLP problems. Such tasks serve multiple purposes. They help establish what is the current state of the art of solving the problem at hand on a given dataset; they typically also lead to advancing the state of the art by the best systems developed by task participants. In many cases, the evaluation data used in the task are also a new contribution, available to the research community after the task.

I have already mentioned (Section 1.4) the importance of the CoNLL 2006 and 2007 tasks for the area of multilingual dependency parsing. Now it is natural to ask how the parsing accuracy would change when parsers are evaluated on the annotation schema of Universal Dependencies. We thus decided to organize a new series of parsing shared tasks at CoNLL 2017 [Zeman et al., 2017] and 2018 [Zeman et al., 2018] (Section 6.5).

The algorithms of machine learning and dependency parsing had improved since 2007, so even a mere repetition of the 2007 task would have been interesting. However, our tasks were novel and brought new insights in a number of ways:

- Thanks to the uniform annotation scheme, it was now possible to compare parsing results across languages.
- It was now possible to combine training data from different languages to increase the robustness of parsing models. Participants were able to take advantage of data combination for well-resourced languages (e.g., a Swedish parser gave better results if it also saw Danish and Norwegian data besides Swedish), but it was especially useful for languages with little or no training data.
- To encourage multilingual and crosslingual parsing techniques, we included several low-resource languages, some of them without any training data. In the 2017 task we even introduced four 'surprise languages' (Buryat, Kurmanji, North Sámi, and Upper Sorbian) that had not been previously released in UD and the participants only got their names and a small data sample shortly before the test phase of the task. The default approach taken by the participants to such languages was a delexicalized parser (Section 1.2) trained on another language, but more successful were lexicalized models trained on multiple languages with weights for individual training datasets.
- An annotation effort was launched that yielded new parallel UD test sets (PUD), consisting of 1000 sentences from online news and Wikipedia, translated into 18 languages. Although this treebank collection was first used for parser evaluation in the shared task, it was later used in various contrastive studies, taking advantage of having the same contents with same annotation scheme in multiple languages.
- In addition to annotated treebanks, we also collected and made available large raw text corpora in 45 languages from Common Crawl to help the participants obtain word embeddings for their parsers.

- With a total of 82 test sets for 57 languages, the 2018 task became the largest and most multilingual evaluation campaign in dependency parsing to date. It set a new trend in NLP that tools and algorithms should be evaluated on large and typologically diverse sets of languages.

- Unlike the older parsing tasks, ours were designed as 'end-to-end' tasks, meaning that the submitted systems could not rely on gold-standard sentence segmentation, tokenization or part-of-speech tags in the input. We effectively redefined the standard setup of a parsing task. Before 2017 it would be common to assume that sentences and tokens are given;[1] since our shared tasks it is expected that a parser should be able to process raw text, which is more like a real-world scenario. Moreover, we also evaluated predicted POS tags and morphological features in the system output. These annotations, while interesting for human users, are typically not needed by modern parsers to predict the syntactic structure; by making them part of the evaluation we encouraged the participating systems to become full-fledged analyzers of natural language morphology and syntax.

With 32 participating teams in 2017 and 25 in 2018, the shared tasks can be considered a success. They also set the stage for a significant flow of follow-up research where multilingual parsing systems were evaluated using the same methodology and same type of data (the latest release of UD).

As cross-linguistic comparison of parsers was one of the goals of the shared tasks, we paid a lot of attention to comparability of the evaluation scores. The uniform annotation scheme was a necessary condition, but not a sufficient one. The standard labeled attachment score (LAS) is affected by various language-specific factors, such as the number of function words. The same grammatical meaning may be encoded by function words, by morphology, or not encoded at all; and while attachment of function words would be reflected in LAS, errors in morphological features would not. This is illustrated in Figure 4.1 with English and Finnish version of the same sentence. English uses a preposition to mark an oblique dependent while Finnish uses the elative case suffix instead. And the three definite articles in English have no counterpart on the Finnish side. Analytical languages like English use more words than synthetic languages like Finnish – in the example, the same meaning is expressed by 8 English words but only 4 Finnish words. If a parser makes one error in each language, its LAS will be 87.5% on English but only 75% on Finnish. One could object that more words also provide more opportunity to make an error; but it often seems to be the case that function words are easy to attach, making it easier for the parsers to reach higher scores on analytical languages. To be able to evaluate the impact of such language differences, we used additional evaluation metrics in the shared tasks. In 2017 the additional metric was CLAS [Nivre and Fang, 2017], which disregards attachment of function words in the total score. For the 2018 task I proposed MLAS [Zeman et al., 2018], which instead combines attachment of content words, attachment of function words and morphological features into

[1] In the 2006 and 2007 tasks one would even expect gold-standard POS tags on input, so the evaluation of the parsing algorithm is not 'biased' by possible tagging errors, but by 2017 it was generally acknowledged that it is important to also evaluate parsing with machine-predicted tags—if the parser needs to see the tags at all.

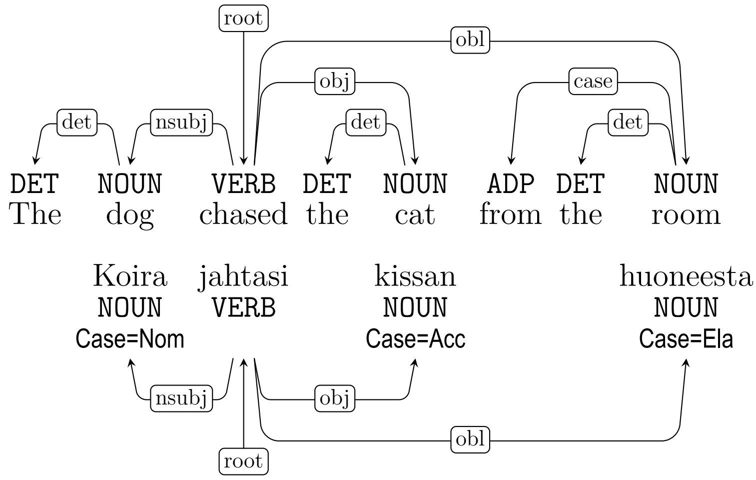


Figure 4.1: Impact of function words on parser evaluation. Adapted from Nivre and Fang [2017].

one score.[2] In the example in Figure 4.1, both English and Finnish have just 4 content words that can be correct or wrong, and to be correct the word must have its incoming dependency relation as well as all morphological features and all dependent function words analyzed correctly.

The shared task overview papers analyze the parsing results from many different angles. Here we just note that in the 2018 shared task, the best system's LAS, macro-averaged over 61 'bigger' datasets (those with large training data) reached 84%; the same figure for MLAS is 73%. The easiest dataset was one of the Polish treebanks (LAS 95%, MLAS 87%); the best result on Czech was LAS 92% and MLAS 85%; on Finnish it was LAS 90% and MLAS 84%; and on English LAS 88% and MLAS 76%. Low-resource languages obviously received much lower scores, especially under the stricter MLAS evaluation. Nine languages in the 2018 task were categorized as low-resource because they had either no labeled training data at all (Breton, Faroese, Naija, and Thai), or there was only a tiny sample of a few dozen sentences (Armenian, Buryat, Kazakh, Kurmanji, and Upper Sorbian). The average score on these languages achieved by the best system was 28% LAS but only 6% MLAS, showing that prediction of morphological features for an unknown language was still an extremely hard task. Nevertheless, there were significant differences among these languages. Some of them benefited from resource-rich siblings and ranked high above the low-resource average: Faroese (Germanic languages; LAS 49%, MLAS 1%), Upper Sorbian (Slavic languages; LAS 46%, MLAS 9%), Breton (Celtic languages; LAS 39%, MLAS 14%), and Armenian (Indo-European; LAS 37%, MLAS 13%).

[2] I also proposed a third metric, BLEX, which reflects syntax and lemmatization. All three metrics (LAS, MLAS, BLEX) were declared equally important – we wanted to encourage the participants to submit systems that predict all types of annotation.

5. Future Directions

After nine years of existence, the UD project is still growing and getting more diverse. New languages are added in every release,[1] new treebanks and genres are added to existing languages, annotated data is added to existing treebanks. Also growing is the community of researchers that contribute to UD and those that use it for their research. I am happy to be part of this endeavor and I hope it will keep growing for many years, as there are still hundreds of languages that lack digital resources. Nevertheless, morphosyntax is not the only area of language processing where annotated data are needed.

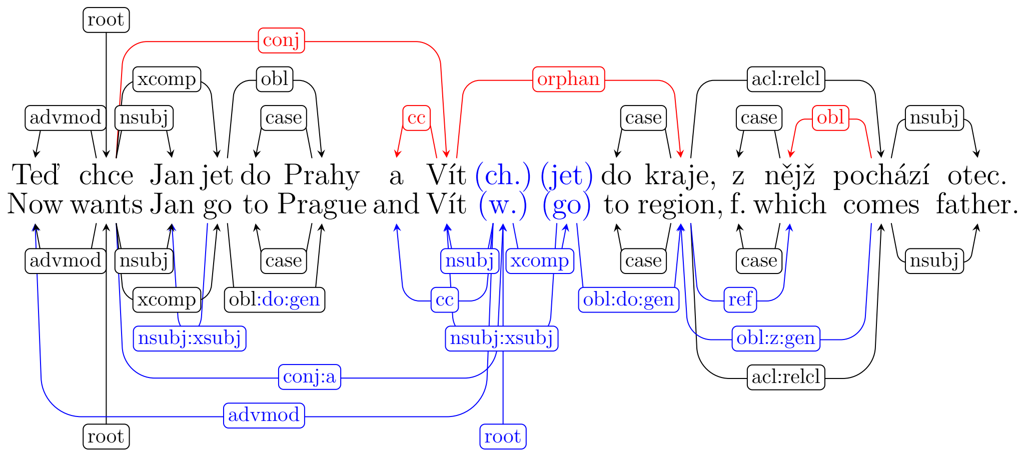
There are multiple proposals to either enhance the UD collection with new annotation layers, or to build other multilingual resources that are separate from UD but strive to follow a similar model of "universal" guidelines that would be applied to all languages. I will now discuss some of these new projects that I am involved in. Most of them revolve around getting closer to the semantics of natural language [Žabokrtský et al., 2020].

UD itself has always foreseen an optional second layer of annotation, called **enhanced representation** or **Enhanced Universal Dependencies (EUD).** A similar layer existed already in Stanford Dependencies and the corresponding UD proposal was first presented by Schuster and Manning [2016]. EUD is a moderate attempt to make explicit some of the relations that are implicitly contained in the syntactic representation and that may be useful for language understanding applications. It is a deep syntactic layer but it does not aspire to provide a complete account of deep syntax (as opposed to other multi-layered syntactic frameworks, most notably the tectogrammatical layer of the Prague Dependency Treebank [Hajič et al., 2000]). Figure 5.1 exemplifies all major enhancements in EUD: 1. abstract nodes for predicates in gapping constructions (the verbs *chce* "wants" and *jít* "go"); 2. parent propagation across coordination (the second root relation to the abstract *chce*); 3. shared dependent of coordination (the second `advmod` relation to the adverb *teď* "now"); 4. grammatical coreference between the subjects of the control verb *chce* and the controlled infinitive *jít*; 5. grammatical coreference between the relative pronoun *nějž* "which" and its antecedent *kraje* "region"; and 6. relation labels enriched by case markers (`obl:do:gen`) and conjunction lemmas (`conj:a`). Note that the enhanced structure is a directed graph but it is no longer a tree.

Some of the enhancements can be derived almost deterministically from the basic dependency structure, others can be estimated with reasonable accuracy using language-specific heuristics. This has been suggested already by Nivre et al. [2018] and confirmed during two shared tasks in Enhanced UD parsing that I co-organized [Bouma et al., 2020, 2021]. In spite of it, only a fraction of the present UD treebanks[2] have the enhanced annotation layer. Ensuring that the other UD treebanks contain at least this minimal deep-syntactic representation is one research direction worth pursuing. However, I believe that we can also go deeper. The rather arbitrary selection of six enhancements can be extended in

[1] UD releases occur regularly twice a year, in May and November.

[2] There are 32 treebanks of 17 different languages that have at least one of the six officially defined enhancements.



‘Now Jan wants to go to Prague while Vít wants to go to the region (his) father comes from.’

Figure 5.1: Example of basic UD tree (above the sentence) and corresponding enhanced UD graph (below). Colors highlight differences between the two structures.

the same spirit to constructions that are similar to those already covered by the guidelines, yet the guidelines do not mention them – sometimes perhaps because the constructions look different from English. For example, in languages such as Tamil the only way of creating a relative clause is a participle. Not only could the relative clause enhancement be extended to relative participles, it could also be extended to participial modifiers in English. Semi-automatic methods could be applied to normalize syntactic alternations [Candito et al., 2017] such as passives, antipassives, reflexives or causatives.

I outlined these ideas together with my PhD student Kira Droganova in Droganova and Zeman [2019] (Section 6.6); to distinguish the new extensions from the already defined Enhanced UD layer, we call it **Deep Universal Dependencies.** We envision a two-speed scenario. On one hand, we want to have cross-linguistically applicable guidelines for many different phenomena that exist between surface syntax and semantics and can be captured in annotated corpora. Conversion procedures could be defined to translate corresponding language resources to the ‘universal’ framework for languages for which such resources already exist. On the other hand, we are well aware that annotations of this kind are difficult and expensive to obtain, so we cannot hope for a growth rate comparable to Universal Dependencies. That is why semi-automatic approaches and heuristics are important, as we can use them to obtain less detailed and less accurate, but still useful annotation for a much larger set of languages. Kira is currently looking into a unified taxonomy of deep syntactic relations that would identify the common ground between several influential frameworks such as the tectogrammatical functors from PDT [Hajič et al., 2000], the PropBank roles [Palmer et al., 2005], or the MTT-inspired annotation of AnCora [Taulé et al., 2008].

Another large area is annotation of entities and **coreference** between them. Not just grammatical coreference, which is conditioned by syntax and which is at least partially covered by Enhanced UD, but all other **mentions** (by name, common noun, pronoun…) that can be said, based on context, to be representing the same entity. Delimitation of mentioning expressions is based on syntactic units, which provides a potential link between Universal Dependencies and coreference annotation. There are coreference-annotated datasets for multiple languages, some of them with and others without syntax, but each following its own annotation scheme. My colleagues and I have thus launched a project called **CorefUD** [Nedoluzhko et al., 2022] where we collect such resources, convert them to a common format and combine them with UD-style morphosyntactic annotation. It currently contains 17 datasets of 12 languages. These datasets have been harmonized at the level of file format and a bit beyond, e.g. with regard to the set of entity types used. However, the common linguistic guidelines are yet to be defined: for example, how exactly should we delimit a mention given its syntactic environment? How do we capture 'zero' mentions that are reflected solely by agreement on the verb? Another PhD student supervised by me, Dima Taji, is just starting research along these lines.

The third multilingual project I want to mention is **Uniform Meaning Representation (UMR)** [Van Gysel et al., 2021]. This one really belongs to the level of semantics, rather than deep syntax. There is no effort at present to map it to syntactic frameworks such as UD, yet the meeting point is that both UD and UMR's objective is to design structured annotation of sentences that would use the same set of concepts across all human languages. Pilot annotations already exist for six languages from six different families. With my colleagues from ÚFAL I am now investigating how UMR can be applied to other languages, primarily to Czech, and (together with my third PhD student Federica Gamba) to Latin.

The last two directions I want to mention here are back in the realm of morphosyntax. Both of them are potential extensions of UD annotation and both of them attempt to overcome problems that stem from taking the word as the basic unit of annotation. The two research directions try to loosen the impact of word boundaries and are complementary: One looks at small phrases, i.e., above the word level, the other looks at morphemes and other sub-word units, i.e, below the word level [Zeman, 2023].

The morphological features in Interset and in UD are defined for individual words; but in many languages, grammatical meanings such as tense and aspect are expressed analytically, using a content word in combination with one or more function words. For example, past perfect (pluperfect) in English is constructed using a finite past tense of the auxiliary *have* and the past participle of a content verb, as in *We had spoken.* None of the words involved is specific to pluperfect, and none of them will get the feature `Tense=Pqp` that encodes pluperfect. Therefore the annotation does not reveal that it is the same construction as Portuguese *Nós faláramos* – here the verb will be annotated as pluperfect, which is expressed purely morphologically. To facilitate such comparisons, we can define a new annotation layer in which UD-like features will be attributed to phrases, possible discontinuous. So in Czech *Nejsem a nikdy jsem nebyl vázán touto smlouvou* "I am not and never have been bound by this contract", we could say that the phrase *nejsem vázán* is finite indicative present tense passive, while *jsem nebyl vázán* is finite indicative past tense passive; note that both of them share the word *vázán*, which itself is only a passive participle (non-finite, with no tense feature).

On the other hand, dependency relations in UD are defined between words but not between smaller units. This is not ideal in certain use cases and certain languages. One cannot see parallel structure between compounds in English,

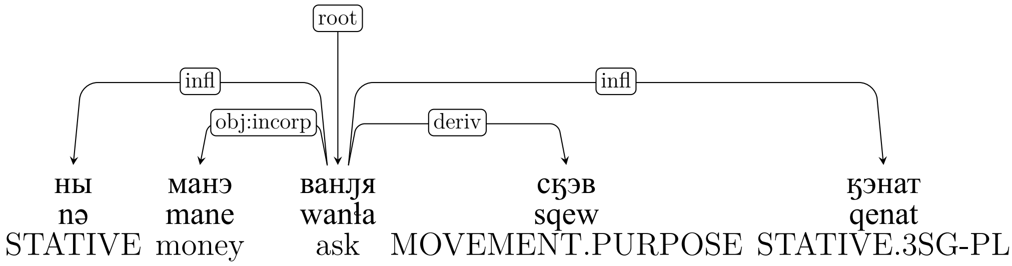


Figure 5.2: A dependency tree over the morphemes of the Chukchi word *ныманэванԓясӄэвӄэнат (nəmanewanɬasqewqenat)* "they constantly asked for money".

where they are usually written as multiple words *(life insurance company)* and in German, where the same compound is typically written as one word *(Lebensversicherungsgesellschaft)*. In other languages there are other reasons why a word may cover an entire sentence: agglutinating languages such as Turkish support long derivation chains (*çöplüklerimizdekilerdenmiydi* "was it from those that were in our garbage cans?"), polysynthetic languages like Chukchi may incorporate object of a verb inside the verb (*ныманэванԓясӄэвӄэнат (nəmanewanɬasqewqenat)* "they constantly asked for money" incorporates the object *манэ* "money" in the verb). A syntactic tree of a sentence with one or two words will not reveal the structure and relations that exist inside the word. One can thus ask whether we can define a similar dependency structure over morphemes rather than words, or at least over sub-word units that have their own lexical content and may correspond to words in other languages. Such extensions have been proposed in the UD community [Tyers and Mishchenkova, 2020] (Figure 5.2) and similar ideas are also pursued by my colleagues at ÚFAL [Žabokrtský et al., 2022].

To summarize, Universal Dependencies and its predecessors have shown that there is a need for linguistically annotated data that cover many human languages and apply a unified annotation framework to all these languages. Almost 150 languages now have such resources at the level of segmentation, morphology and surface syntax, and these resources are widely used in natural language processing, linguistics and digital humanities in general. This effort can and should be extended to other languages, but also to other areas of natural language understanding, such as deep syntax and semantics.

6. Selected Publications

6.1 Cross-Language Parser Adaptation between Related Languages

Full reference: Daniel Zeman and Philip Resnik. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*, pages 35–42, Hyderabad, India, January 2008. Asian Federation of Natural Language Processing. URL `https://aclanthology.org/I08-3008.pdf`. [Zeman and Resnik, 2008]

Comments: The term *delexicalized parsing* was coined in this paper. We presented experiments with transfer of parsing models from Danish to Swedish, where Swedish served as a surrogate for a low-resource language. Besides delexicalized parsing (Section 1.2), we also evaluated test data translation (Section 1.3), and found the former to perform better on our dataset. Our proposals were further developed and evaluated on multiple languages by McDonald et al. [2011], which sparked more interest by a number of other researchers. Nowadays, delexicalized parsing is still occasionally used as a cheap and quick first step for resourceless languages; however, lexicalized parsers using large multilingual language models typically perform better (even on languages not contained in their training data). My contribution: about 70%. Number of citations according to Google Scholar (retrieved 2023-07-21): **240**.

6.2 Reusable Tagset Conversion Using Tagset Drivers

Full reference: Daniel Zeman. Reusable tagset conversion using tagset drivers. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 213–218, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2008/pdf/66_paper.pdf`. [Zeman, 2008]

Comments: This is the first and main reference for Interset (Chapter 2). A preliminary version of the tagset conversion system was used already in [Zeman and Resnik, 2008]. Besides being used to convert tags between existing tagsets, Interset gradually became a framework that could be used to describe and access word features in any language. It became part of the language-processing framework Treex [Popel and Žabokrtský, 2010],[1] it was extensively used in the HamleDT project (Section 6.3), and finally, selected features from Interset provided the morphological annotation layer in Universal Dependencies (Section 6.4). I continue to oversee and maintain the set of features documented in UD, as I did previously for Interset; I also keep the conversion libraries in sync with

[1] `https://ufal.mff.cuni.cz/treex`

newly added features. Furthermore, my experience with morphosyntactic harmonization has projected into my monograph on the topic [Zeman, 2018]. My contribution: 100%. Number of citations according to Google Scholar (retrieved 2023-07-21): **209**.

6.3 HamleDT: Harmonized Multi-language Dependency Treebank

Full reference: Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. HamleDT: Harmonized multi-language dependency treebank. *Language Resources and Evaluation*, 48:601–637, 2014. URL `https://link.springer.com/content/pdf/10.1007/s10579-014-9275-2.pdf`. [Zeman et al., 2014]

Comments: The first paper about HamleDT (Section 3.1) was Zeman et al. [2012], presented at LREC in İstanbul. This is an extended version of that paper, which we were invited to submit to the LRE journal. HamleDT was a pioneering project, which provided the first collection of harmonized treebanks; it was also the largest one. Later at LREC in Reykjavík we presented a new version of HamleDT, which provided an alternative conversion of the treebanks to Stanford Dependencies [Rosa et al., 2014]. When the Universal Dependencies initiative started in 2014, the consensus was reached that the syntactic annotation in UD will be derived from Stanford (rather than Prague) dependencies. During 2015, we made all HamleDT treebanks compatible with the new UD standard. We made one final release, HamleDT 3.0. All HamleDT treebanks with permissive licenses were then incorporated in UD, which became a successor of HamleDT. My contribution: about 25%. Number of citations according to Google Scholar (retrieved 2023-07-21): **84**, together with the other two papers: **215**.

6.4 Universal Dependencies

Full reference: Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, 2021. DOI `10.1162/COLI_a_00402`. URL `https://aclanthology.org/2021.cl-2.11.pdf`. [de Marneffe et al., 2021]

Comments: The story of Universal Dependencies (Section 3.3) is atypical. Many projects are first publicized in a paper, then the impact of the publication is observed and eventually new work and new papers emerge. In the case of UD, the impact of the project was already well observable when the first descriptive paper appeared at LREC 2016 [Nivre et al., 2016]. The paper described version 1 of the annotation guidelines but later that year we projected the initial experience to version 2, which is still in use today. A paper describing version 2 was published at LREC 2020 [Nivre et al., 2020]. However, here I wish to emphasize and include the article we published a year later in Computational Linguistics. In comparison to the LREC papers it puts less weight on the growth and coverage of the data collection and focuses more on the linguistic theory behind the UD framework, which it lays out in much finer detail, with numerous examples from typologically diverse languages. Besides, I can claim significantly larger share of authorship of the latter article. My contribution: 25%. Number of citations according to Google Scholar (retrieved 2023-07-21): **278**, together with the other two papers: **2113**.

Besides working on the UD annotation scheme, I have also converted, annotated or contributed to dozens of UD treebanks. A few of these contributions were described in separate papers:

- Catalan and Spanish [Martínez Alonso and Zeman, 2016]
- Russian [Lyashevskaya et al., 2016, Droganova et al., 2018]
- Arabic [Taji et al., 2017]
- Slovak [Zeman, 2017]
- Latin [Cecchini et al., 2018, Gamba and Zeman, 2023]
- Sanskrit [Dwivedi and Zeman, 2018]
- Bhojpuri [Ojha and Zeman, 2020]
- Yoruba [Ishola and Zeman, 2020]
- Albanian [Toska et al., 2020]
- Indonesian [Alfina et al., 2020]
- Malayalam [Stephen and Zeman, 2023]

6.5 CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies

Full reference: Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. CoNLL 2018 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–21, Bruxelles, Belgium, October 2018. Association for Computational Linguistics. DOI `10.18653/v1/K18-2001`. URL `https://aclanthology.org/K18-2001v1.pdf`. [Zeman et al., 2018]

Comments: The overview paper of the second UD shared task in 2018 is presented here as a culmination of the two-year long evaluation campaign (Chapter 4); the first task was described in Zeman et al. [2017]. Seven years later this paper remains an important reference for multilingual end-to-end parsing, although new and better parsing models have emerged since then, especially with the advent of transformer-based multilingual large language models. We also organized two more shared tasks collocated with the IWPT conference [Bouma et al., 2020, 2021], which were focused on Enhanced UD parsing (Chapter 5) but all the previous annotation levels were evaluated as well. Unlike the pre-UD parsing tasks, new parsers are usually not evaluated on the shared task data except for comparison purposes; instead, they are evaluated on the most recent release of UD, which includes new languages and potentially also fixes of annotation errors in the older datasets. End-to-end parsing evaluation has become standard, and the shared task evaluation script is freely available among UD tools so that everyone can evaluate their parser following the same methodology. As for the newly proposed evaluation metrics, they cannot compete in popularity with the well-established LAS, yet they are occasionally used by other authors (e.g., Dary and Nasr [2021]). My contribution: about 45%. Number of citations according to Google Scholar (retrieved 2023-07-21): **569**.[2]

6.6 Towards Deep Universal Dependencies

Full reference: Kira Droganova and Daniel Zeman. Towards Deep Universal Dependencies. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 144–152, Paris, France, August 2019. Association for Computational Linguistics. DOI `10.18653/v1/W19-7717`. URL `https://aclanthology.org/W19-7717.pdf`. [Droganova and Zeman, 2019]

Comments: This paper is the first step on the journey from Universal Dependencies to a similarly broad and multilingual approach to deep syntax and semantics. As such, it is a representative of the possible future directions I outline in Chapter 5. We have already released several automatic enhancements of Universal Dependencies with deep-syntactic annotations and received some feedback from other researchers. Nevertheless, Deep UD has to be considered work in progress: it will be really useful when it can incorporate existing manually curated resources such as Prague tectogrammar or PropBank. My contribution: 50%. Number of citations according to Google Scholar (retrieved 2023-07-21): **16**.

[2]Google Scholar has merged the two papers about the two shared task years. This is the aggregate number of citations for both [Zeman et al., 2017] and [Zeman et al., 2018].

Bibliography

Željko Agić, Dirk Hovy, and Anders Søgaard. If all you have is a bit of the Bible: Learning POS taggers for truly low-resource languages. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 268–272, Beijing, China, July 2015. Association for Computational Linguistics. URL `https://aclanthology.org/P15-2044.pdf`.

Ika Alfina, Daniel Zeman, Arawinda Dinakaramani, Indra Budi, and Heru Suhartanto. Selecting the UD v2 morphological features for Indonesian dependency treebank. In *Proceedings of the International Conference on Asian Language Processing (IALP 2020)*, pages 104–109, Kuala Lumpur, Malaysia, 2020. Chinese and Oriental Languages Information Processing Society. ISBN 978-1-7281-7689-5. URL `https://colips.org/conferences/ialp2020/proceedings/papers/IALP2020_P87.pdf`.

Lauriane Aufrant, Guillaume Wisniewski, and François Yvon. Zero-resource dependency parsing: Boosting delexicalized cross-lingual transfer with linguistic knowledge. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 119–130, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee. URL `https://aclanthology.org/C16-1012.pdf`.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. Overview of the IWPT 2020 shared task on parsing into enhanced Universal Dependencies. In Gosse Bouma, Yuji Matsumoto, Stephan Oepen, Kenji Sagae, Djamé Seddah, Weiwei Sun, Anders Søgaard, Reut Tsarfaty, and Daniel Zeman, editors, *Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies*, pages 151–161, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.iwpt-1.16. URL `https://aclanthology.org/2020.iwpt-1.16`.

Gosse Bouma, Djamé Seddah, and Daniel Zeman. From raw text to enhanced Universal Dependencies: The parsing shared task at IWPT 2021. In Stephan Oepen, Kenji Sagae, Reut Tsarfaty, Gosse Bouma, Djamé Seddah, and Daniel Zeman, editors, *Proceedings of the 17th International Conference on Parsing Technologies and the IWPT 2021 Shared Task on Parsing into Enhanced Universal Dependencies (IWPT 2021)*, pages 146–157, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.iwpt-1.15. URL `https://aclanthology.org/2021.iwpt-1.15`.

Sabine Buchholz and Erwin Marsi. CoNLL-X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 149–164, New York, NY, USA, June 2006. Association for Computational Linguistics. URL `https://aclanthology.org/W06-2920.pdf`.

Marie Candito, Bruno Guillaume, Guy Perrier, and Djamé Seddah. Enhanced UD dependencies with neutralized diathesis alternation. In *Proceedings of the*

Fourth International Conference on Dependency Linguistics (Depling 2017), pages 42–53, Pisa, Italy, September 2017. Linköping University Electronic Press. URL https://aclanthology.org/W17-6507.pdf.

Flavio Massimiliano Cecchini, Marco Passarotti, Paola Marongiu, and Daniel Zeman. Challenges in converting the Index Thomisticus treebank into Universal Dependencies. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 27–36, Bruxelles, Belgium, November 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6004. URL https://aclanthology.org/W18-6004.pdf.

Franck Dary and Alexis Nasr. The Reading Machine: a Versatile Framework for Studying Incremental Parsing Strategies. In *The 17th International Conference on Parsing Technologies*, Bangkok (virtual), Thailand, August 2021. URL https://hal.science/hal-03328439.

Dipanjan Das and Slav Petrov. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 600–609, Portland, OR, USA, June 2011. Association for Computational Linguistics. URL https://aclanthology.org/P11-1061.pdf.

Marie-Catherine de Marneffe and Christopher D. Manning. Stanford typed dependencies manual (technical report), September 2008. URL https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. Generating typed dependency parses from phrase structure parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genova, Italy, May 2006. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2006/pdf/440_pdf.pdf.

Marie-Catherine de Marneffe, Miriam Connor, Natalia Silveira, Samuel R. Bowman, Timothy Dozat, and Christopher D. Manning. More constructions, more genres: Extending Stanford dependencies. In *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*, Praha, Czechia, August 2013. Charles University in Prague, Matfyzpress. URL https://aclanthology.org/W13-3721.pdf.

Marie-Catherine de Marneffe, Natalia Silveira, Timothy Dozat, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavík, Iceland, May 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4. URL http://www.lrec-conf.org/proceedings/lrec2014/pdf/1062_Paper.pdf.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, 2021. doi: 10.1162/COLI_a_00402. URL https://aclanthology.org/2021.cl-2.11.pdf.

R. M. W. Dixon. *Basic Linguistic Theory. Volume 1.* Oxford University Press, Oxford, UK, 2010. ISBN 978-0-19-957106-2.

Kira Droganova and Daniel Zeman. Towards Deep Universal Dependencies. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 144–152, Paris, France, August 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-7717. URL `https://aclanthology.org/W19-7717.pdf`.

Kira Droganova, Olga Lyashevskaya, and Daniel Zeman. Data conversion and consistency of monolingual corpora: Russian UD treebanks. In *Proceedings of the 17th International Workshop on Treebanks and Linguistic Theories (TLT 2018)*, pages 53–66, Linköping, Sweden, 2018. Linköping University Electronic Press. ISBN 978-91-7685-137-1. URL `https://ep.liu.se/ecp/155/ecp18155.pdf#page=61`.

Puneet Dwivedi and Daniel Zeman. The forest lion and the bull: Morphosyntactic annotation of the Panchatantra. *Computación y Sistemas*, 22(4):1377–1384, 2018. ISSN 1405-5546. URL `https://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/3076/2576`.

Federica Gamba and Daniel Zeman. Universalising Latin Universal Dependencies: a harmonisation of Latin treebanks in UD. In *Proceedings of the Sixth Workshop on Universal Dependencies (UDW, GURT/SyntaxFest 2023)*, pages 7–16, Washington, DC, USA, March 2023. Association for Computational Linguistics. URL `https://aclanthology.org/2023.udw-1.2.pdf`.

Sofia Gustafson-Capková and Britt Hartmann. Manual of the Stockholm Umeå Corpus version 2.0, December 2006. URL `https://spraakbanken.gu.se/parole/Docs/SUC2.0-manual.pdf`. [online; accessed 2018-07-26].

Jan Hajič, Alena Böhmová, Eva Hajičová, and Barbora Vidová Hladká. The Prague Dependency Treebank: A Three-Level Annotation Scenario. In Anne Abeillé, editor, *Treebanks: Building and Using Parsed Corpora*, pages 103–127. Amsterdam:Kluwer, 2000.

Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. Bootstrapping parsers via syntactic projection across parallel texts. *Natural Language Engineering*, 11(3):311–325, September 2005. doi: 10.1017/S1351324905003840. URL `https://doi.org/10.1017/S1351324905003840`.

Ọlájídé Ishola and Daniel Zeman. Yorùbá dependency treebank (YTB). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5178–5186, Marseille, France, May 2020. European Language Resources Association (ELRA). URL `https://aclanthology.org/2020.lrec-1.637.pdf`.

Ishan Jindal, Alexandre Rademaker, Michał Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu, and Yunyao Li. Universal proposition bank 2.0. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1700–1711, Marseille, France, June 2022. European Language Resources Association (ELRA). URL `https://aclanthology.org/2022.lrec-1.181.pdf`.

Dan Kondratyuk and Milan Straka. 75 languages, 1 model: Parsing universal dependencies universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1279. URL https://aclanthology.org/D19-1279.pdf.

Matthias T. Kromann. The danish dependency treebank: Linguistic principles and semi-automatic tagging tools, August 2002. URL https://www.researchgate.net/publication/228824624_T_The_Danish_Dependency_Treebank_Linguistic_Principles_and_Semi-automatic_Tagging_Tools.

Olga Lyashevskaya, Kira Droganova, Daniel Zeman, Maria Alexeeva, Tatiana Gavrilova, Nina Mustafina, and Elena Shakurova. Universal Dependencies for Russian: A new syntactic dependencies tagset, 2016. URL http://olesar.narod.ru/papers/44LNG2016.pdf.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, 19(2):313–330, 1993. URL https://aclanthology.org/J93-2004.pdf.

Héctor Martínez Alonso and Daniel Zeman. Universal Dependencies for the AnCora treebanks. *Procesamiento del Lenguaje Natural*, 57:91–98, September 2016. URL http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5341.

Ryan McDonald, Slav Petrov, and Keith Hall. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK, July 2011. Association for Computational Linguistics. URL https://aclanthology.org/D11-1006.pdf.

Ryan McDonald, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. Universal dependency annotation for multilingual parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL https://aclanthology.org/P13-2017.pdf.

Pruthwik Mishra, Vandan Mujadia, and Dipti Misra Sharma. POS tagging for resource poor languages through feature projection. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)*, pages 50–55, Kolkata, India, December 2017. NLP Association of India. URL https://aclanthology.org/W17-7507.pdf.

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Amir Zeldes, and Daniel Zeman. CorefUD 1.0: Coreference meets Universal Dependencies. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid

Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4859–4872, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.520.pdf.

Jens Nilsson, Johan Hall, and Joakim Nivre. MAMBA meets TIGER: Reconstructing a Swedish treebank from antiquity. In *Proceedings of the NODALIDA Special Session on Treebanks*, 2005. URL http://www.msi.vxu.se/users/nivre/research/Talbanken05.html.

Joakim Nivre and Chiao-Ting Fang. Universal Dependency evaluation. In Marie-Catherine de Marneffe, Joakim Nivre, and Sebastian Schuster, editors, *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 86–95, Göteborg, Sweden, May 2017. Association for Computational Linguistics. URL https://aclanthology.org/W17-0411v2.pdf.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL https://aclanthology.org/L16-1262.pdf.

Joakim Nivre, Paola Marongiu, Filip Ginter, Jenna Kanerva, Simonetta Montemagni, Sebastian Schuster, and Maria Simi. Enhancing universal dependency treebanks: A case study. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 102–107, Bruxelles, Belgium, November 2018. Association for Computational Linguistics. URL https://aclanthology.org/W18-6012.pdf.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France, May 2020. European Language Resources Association (ELRA). URL https://aclanthology.org/2020.lrec-1.497.pdf.

Atul Kr. Ojha and Daniel Zeman. Universal Dependency treebanks for low-resource Indian languages: The case of Bhojpuri. In *Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation*, pages 33–38, Marseille, France, May 2020. European Language Resources Association (ELRA). URL https://aclanthology.org/2020.wildre-1.7.pdf.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106,

2005. doi: 10.1162/0891201053630264. URL https://aclanthology.org/J05-1004.

Slav Petrov, Dipanjan Das, and Ryan McDonald. A universal part-of-speech tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2089–2096, İstanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/274_Paper.pdf.

Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, and Zdeněk Žabokrtský. Coordination structures in dependency treebanks. In Hinrich Schuetze, Pascale Fung, and Massimo Poesio, editors, *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 517–527, Sofia, Bulgaria, August 2013. Association for Computational Linguistics. URL https://aclanthology.org/P13-1051.pdf.

Martin Popel and Zdeněk Žabokrtský. TectoMT: Modular NLP framework. In *Advances in Natural Language Processing. 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16-18, 2010, Proceedings*, pages 293–304, Reykjavík, Iceland, August 2010. Springer. doi: 10.1007/978-3-642-14770-8_33. URL https://ufal.mff.cuni.cz/~popel/papers/2010_icetal.pdf.

Loganathan Ramasamy. *Parsing under-resourced languages: Cross-lingual transfer strategies for Indian languages*. PhD thesis, Univerzita Karlova v Praze, Praha, Czechia, 2014. URL http://ufal.mff.cuni.cz/biblio/attachments/2014-ramasamy-m251064576937367469.pdf.

Rudolf Rosa. *Discovering the structure of natural language sentences by semi-supervised methods*. PhD thesis, Univerzita Karlova v Praze, Praha, Czechia, 2018. URL http://ufal.mff.cuni.cz/biblio/attachments/2018-rosa-p4772924917445474076.pdf.

Rudolf Rosa, Jan Mašek, David Mareček, Martin Popel, Daniel Zeman, and Zdeněk Žabokrtský. HamleDT 2.0: Thirty dependency treebanks stanfordized. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2334–2341, Reykjavík, Iceland, May 2014. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2014/pdf/915_Paper.pdf.

Rudolf Rosa, Daniel Zeman, David Mareček, and Zdeněk Žabokrtský. Slavic forest, norwegian wood. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 210–219, Valencia, Spain, April 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-1226. URL https://aclanthology.org/W17-1226.pdf.

Sebastian Schuster and Christopher D. Manning. Enhanced English Universal Dependencies: An improved representation for natural language understanding tasks. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth*

International Conference on Language Resources and Evaluation (LREC'16), pages 2371–2378, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL `https://aclanthology.org/L16-1376.pdf`.

Abishek Stephen and Daniel Zeman. Universal Dependencies for Malayalam. *The Prague Bulletin of Mathematical Linguistics*, (120):31–46, 2023. ISSN 0032-6585. URL `https://ufal.mff.cuni.cz/pbml/120/art-stephen-zeman.pdf`.

John Sylak-Glassman, Christo Kirov, David Yarowsky, and Roger Que. A language-independent feature schema for inflectional morphology. In Chengqing Zong and Michael Strube, editors, *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 674–680, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-2111. URL `https://aclanthology.org/P15-2111.pdf`.

Dima Taji, Nizar Habash, and Daniel Zeman. Universal Dependencies for Arabic. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 166–176, Valencia, Spain, April 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-1320. URL `https://aclanthology.org/W17-1320.pdf`.

Mariona Taulé, M. Antònia Martí, and Marta Recasens. AnCora: Multilevel annotated corpora for Catalan and Spanish. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2008/pdf/35_paper.pdf`.

Ulf Teleman. Manual för grammatisk beskrivning av talad och skriven svenska (Mamba), 1974.

Lucien Tesnière. *Eléments de syntaxe structurale*. Klincksieck, Paris, France, 1959.

Jörg Tiedemann. Rediscovering annotation projection for cross-lingual parser induction. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1854–1864, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics. URL `https://aclanthology.org/C14-1175.pdf`.

Marsida Toska, Joakim Nivre, and Daniel Zeman. Universal Dependencies for Albanian. In *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*, pages 178–188, Barcelona, Spain (Online), December 2020. Association for Computational Linguistics. URL `https://aclanthology.org/2020.udw-1.20.pdf`.

Francis Tyers and Karina Mishchenkova. Dependency annotation of noun incorporation in polysynthetic languages. In *Proceedings of the Fourth Workshop on Universal Dependencies (UDW 2020)*, pages 195–204, Barcelona, Spain (Online), December 2020. Association for Computational Linguistics. URL `https://aclanthology.org/2020.udw-1.22.pdf`.

Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. Designing a Uniform Meaning Representation for natural language processing. *Künstliche Intelligenz*, 35(0):343–360, October 2021. ISSN 1610-1987. doi: 10.1007/s13218-021-00722-w. URL `https://par.nsf.gov/servlets/purl/10288899`.

David Yarowsky and Grace Ngai. Inducing multilingual pos taggers and np bracketers via robust projection across aligned corpora. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA, USA, June 2001. Association for Computational Linguistics. URL `https://aclanthology.org/N01-1026.pdf`.

Zdeněk Žabokrtský, Daniel Zeman, and Magda Ševčíková. Sentence meaning representations across languages: What can we learn from existing frameworks? *Computational Linguistics*, 46(3):605–665, September 2020. doi: 10.1162/coli_a_00385. URL `https://aclanthology.org/2020.cl-3.3.pdf`.

Zdeněk Žabokrtský, Niyati Bafna, Jan Bodnár, Lukáš Kyjánek, Emil Svoboda, Magda Ševčíková, and Jonáš Vidra. Towards universal segmentations: UniSegments 1.0. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1137–1149, Marseille, France, June 2022. European Language Resources Association. URL `https://aclanthology.org/2022.lrec-1.122.pdf`.

Daniel Zeman. *Parsing with a Statistical Dependency Model.* PhD thesis, Univerzita Karlova v Praze, Praha, Czechia, 2004. URL `http://ufal.mff.cuni.cz/biblio/attachments/2004-zeman-m5440617933930313730.pdf`.

Daniel Zeman. Reusable tagset conversion using tagset drivers. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 213–218, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA). URL `http://www.lrec-conf.org/proceedings/lrec2008/pdf/66_paper.pdf`.

Daniel Zeman. Slovak dependency treebank in Universal Dependencies. *Journal of Linguistics / Jazykovedný časopis*, 68(2):385–395, December 2017. doi: 10.1515/jazcas-2017-0048. URL `https://sciendo.com/article/10.1515/jazcas-2017-0048`.

Daniel Zeman. *The World of Tokens, Tags and Trees*. ÚFAL MFF UK, Praha, Czechia, 2018. ISBN 978-80-88132-09-7. URL https://ufal.mff.cuni.cz/books/preview/2018-zeman_full.pdf.

Daniel Zeman. Subword relations, superword features. In *UniDive General Meeting at Paris-Saclay posters*, Orsay, France, February 2023. URL https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:2023-saclay:abstracts:39_zeman_subword_relations_superword_features.pdf. https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:2023-saclay:abstracts:wg1-2-zeman-poster.pdf.

Daniel Zeman and Philip Resnik. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*, pages 35–42, Hyderabad, India, January 2008. Asian Federation of Natural Language Processing. URL https://aclanthology.org/I08-3008.pdf.

Daniel Zeman, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. HamleDT: To parse or not to parse? In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2735–2741, İstanbul, Turkey, May 2012. European Language Resources Association (ELRA). URL http://www.lrec-conf.org/proceedings/lrec2012/pdf/429_Paper.pdf.

Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, and Jan Hajič. HamleDT: Harmonized multi-language dependency treebank. *Language Resources and Evaluation*, 48:601–637, 2014. doi: 10.1007/s10579-014-9275-2. URL https://link.springer.com/content/pdf/10.1007/s10579-014-9275-2.pdf.

Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gokirmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Urešová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher D. Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj, and Josie Li. CoNLL 2017 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–19, Vancouver, Canada, August 2017. Association for Computational Linguistics. doi: 10.18653/v1/K17-3001. URL https://aclanthology.org/K17-3001v1.pdf.

Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, and Slav Petrov. CoNLL 2018 shared task: Multilingual parsing from raw text to Universal Dependencies. In *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 1–21, Bruxelles, Belgium, October 2018. Association for Computational Linguistics. doi: 10.18653/v1/K18-2001. URL `https://aclanthology.org/K18-2001v1.pdf`.

List of Figures

List of Tables