1. Introduction

Dependency treebanks are available for several dozen languages, and there is a growing interest in multilingual/cross-lingual syntactic parsing experiments. However, the individual treebanks use different annotation schemes. The differences can be found in the dependency-relation label sets (the names and definitions of the labels and the granularity of the set) as well as in the dependency structures capturing coordination, subordinate clauses, verb groups, prepositional phrases and other linguistic phenomena (Zeman et al., 2012). These divergences present a significant obstacle to the use of these resources in multilingual language technologies or for evaluation of cross-lingual syntactic parsers (McDonald et al., 2011). Therefore, a logical step to take is to convert the various treebanks into the same schema.

So far, the largest collection of harmonized treebanks has been HamleDT 1.0, a compilation of 29 existing dependency treebanks (or dependency conversions of other treebanks). The treebanks were harmonized into the Prague Dependencies (PRG) style of annotation (Zeman et al., 2012), which is a slight adaptation of the annotation style of the Prague Dependency Treebank (PDT) (Hajič et al., 2006). We list the source treebanks and describe PRG in Section 2.

Recently, the Stanford Dependencies (SD) representation (de Marneffe et al., 2006; de Marneffe et al., 2013) has gained in popularity and, although primarily defined for English, has been successfully used by researchers in various domains and for various languages. Moreover, Universal Stanford Dependencies (USD) (de Marneffe et al., 2014) have just been introduced, which focus on adapting the previous version to capture grammatical relations across languages. Until now, the largest collection of treebanks subscribing to the SD has been the Universal Dependency Treebank of McDonald et al. (2013), currently including 11 languages (7 of which have been annotated manually directly using SD; the others were converted automatically).

In this paper, we present HamleDT 2.0, a collection of 30 treebanks annotated in basic USD. The treebanks, already harmonized to PRG in HamleDT 1.0, were automatically converted from PRG to USD in a language-independent (and also source-treebank-independent) way, thereby creating the largest existing collection of treebanks annotated in USD. The resource is released on our website.1 We detail the stanfordization in Section 3. and discuss some encountered issues in Section 4.

2. Harmonization

The first step in creating the data resource presented was the collection of the source treebanks, and their automatic rule-based conversion (harmonization) to PRG, including conversion of part of speech tags and other annotated morphological information into Interset representation (Zeman, 2008). This has already been done in HamleDT 1.0 (Zeman et al., 2012) and further improved in HamleDT 1.5.1 In HamleDT 2.0, a new Slovak treebank was added.

In this section, we list the source treebanks in Table 1 and describe the current version of PRG used in HamleDT. For details about the harmonization itself, please refer to (Zeman et al., 2012).

As the current version of HamleDT contains just one treebank per language, we use the ISO language codes to refer to individual treebanks throughout this paper. Any claims are to be read as claims about the particular treebank, and not necessarily about the language in general.

2.1. Prague Dependencies

There are at least ten treebanks (both from Prague and from other places) that use the PRG label set natively. Eight of

---

1http://ufal.mff.cuni.cz/hamledt

The licenses allow us to distribute only a subset of the whole data set, consisting of 13 treebanks. For the rest of the treebanks, the user has to obtain the source treebank first; we then provide the conversion tools.
Table 1: List of treebanks included in HamleDT 2.0

<table>
<thead>
<tr>
<th>Language</th>
<th>Treebank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabic</td>
<td>Prague Arabic Dependency Treebank 1.0 / CoNLL 2007 (Smrž et al., 2008)</td>
</tr>
<tr>
<td>Basque</td>
<td>Basque Dependency Treebank, a larger version than the one included in CoNLL 2007, generously provided by IZA Group (Aduriz et al., 2003)</td>
</tr>
<tr>
<td>Bengali</td>
<td>BuTreeBank (Simov and Osenova, 2005)</td>
</tr>
<tr>
<td>Bulgarian</td>
<td>Bulgarian Dependency Treebank / ICON 2010 (Husain et al., 2010)</td>
</tr>
<tr>
<td>Catalan</td>
<td>Catalan Dependency Treebank / ICON 2010 (Husain et al., 2010)</td>
</tr>
<tr>
<td>Czech</td>
<td>Czech Dependency Treebank 3.0 / CoNLL 2007 (Bejček and Kopecký, 2011)</td>
</tr>
<tr>
<td>Danish</td>
<td>Danish Dependency Treebank / CoNLL 2006 (Kromann et al., 2004), now part of the Copenhagen Dependency Treebank</td>
</tr>
<tr>
<td>Dutch</td>
<td>Dutch Dependency Treebank / CoNLL 2006 (van der Beek et al., 2002)</td>
</tr>
<tr>
<td>English</td>
<td>English Dependency Treebank / CoNLL 2007 (Marcus et al., 1993)</td>
</tr>
<tr>
<td>Estonian</td>
<td>Estonian Dependency Treebank / CoNLL 2007 (Bick et al., 2004)</td>
</tr>
<tr>
<td>Finnish</td>
<td>Finnish Dependency Treebank / Haverinen et al., 2010)</td>
</tr>
<tr>
<td>German</td>
<td>German Dependency Treebank / CoNLL 2009 (Brants et al., 2004)</td>
</tr>
<tr>
<td>Greek</td>
<td>Greek Dependency Treebank (Prokopidis et al., 2005)</td>
</tr>
<tr>
<td>Hindi</td>
<td>Hindi Dependency Treebank / CoNLL 2007 (Srinivas et al., 2008)</td>
</tr>
<tr>
<td>Hungarian</td>
<td>Hungarian Dependency Treebank / CoNLL 2007 (Montemagni et al., 2003)</td>
</tr>
<tr>
<td>Italian</td>
<td>Italian Syntactic-Semantic Treebank / CoNLL 2007 (Montemagni et al., 2003)</td>
</tr>
<tr>
<td>Latin</td>
<td>Latin Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Persian</td>
<td>Persian Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Portuguese Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Romanian</td>
<td>Romanian Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Russian</td>
<td>Russian Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Slovak</td>
<td>Slovak Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Slovene</td>
<td>Slovene Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Spanish</td>
<td>Spanish Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Swedish</td>
<td>Swedish Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Tamil</td>
<td>Tamil Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Telugu</td>
<td>Telugu Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
<tr>
<td>Turkish</td>
<td>Turkish Dependency Treebank / CoNLL 2007 (Kataoka et al., 2000)</td>
</tr>
</tbody>
</table>

Note that the labels listed above are part of the PRG label set used in HamleDT. The list of PRG labels includes:

- **Pron** - Pronominal predicate. Typically a pronoun in nominative, though infinitives and other realizations are possible.
- **Adv** - Adverbial modifier of a verb, adjective, adverb or numeral ("approximately ten"). Typically realized as adverb, prepositional phrase or clause.
- **Atr** - Attribute of a noun, pronoun or numeral (one number attached to another, e.g. in dates). Realized as adjective, another noun, prepositional phrase or clause. A noun parent should never get an Atr child: even "stadium in London" will be analyzed as Atr.
- **Apposition** - A noun phrase attached to another noun phrase as a parenthetical explanation, e.g. "Elisabeth II, the Queen of England". This label is intentionally distinct from Apos, used in the original treebanks. The attachment of apposition in HamleDT is very similar to nominal attributes, while the original PRG annotation style is to treat it as a

2http://ufal.mff.cuni.cz/pdt2.0
paratactic structure, similar to coordination. This is the first major deviation from the style of PDT. Unlike the creators of PDT, we do not see apposition as an inherently paratactic construction; this is in accord with all the non-Prague treebanks in our collection.

Coord – Coordinating node. Usually a conjunction, sometimes comma or other punctuation. This label does not describe the relation of the node to its parent. Such relations are marked at the conjuncts. Children of the Coord node are classified as either conjuncts or shared dependents of the conjuncts. (Every conjunct may have its own private dependents in addition to the shared ones.)

AuxP – Primary preposition or part of secondary (compound) preposition. Preposition determines the case of the noun phrase in many Indo-European languages, therefore it is annotated as the parent of the noun in PRG. Still, it is considered an auxiliary node and the real lexical dependency goes from the parent of the preposition to its noun child (this is driven by the same lexicalist principle that led in USD to attaching prepositions as child nodes of their nouns). The function of the whole prepositional phrase is annotated at the noun. See Figure 1.

AuxC – Subordinating conjunction. Parallel to prepositions, subordinating conjunctions are auxiliary nodes on the path between the predicate of the subordinate clause and its governor. The real function of the clause is annotated at the predicate.

AuxR – Auxiliary verb attached to main verb.

Neg – Particle that negates the meaning of its parent. Language-specific, as some languages express negation using bound morphemes. This label is not used in PDT (it is not needed for Czech) but it is useful in quite a few languages and we recently introduced it in HamleDT. We still do not recognize it in all the treebanks where it would apply. In some of the treebanks, negative particles are labeled as adverbial modifiers.

AuxT – In PDT, this label is reserved for reflexive pronouns attached to inherently reflexive verbs (reflexiva tantum). In HamleDT, we also use it for particles that modify the meaning of verb, e.g. “make up”.

AuxX – Garbage can for adverbs, conjunctions, particles etc. that do not qualify for a better label. Example: actually. This label is also used for dependent parts of multiword conjunctions (such as such as), extra conjunctions in coordinations etc.

ExD – Externally dependent (see Figures 2 and 3). This label in general signals that the structural information is incomplete. If an item was elided but its dependents remain in the sentence, they get ExD. Example: John brought some apples and Mary oranges. Mary and oranges cannot be attached to the second brought because it is missing. They are therefore treated as conjuncts, together with the first brought. They cannot be labeled as predicates (which they are not), nor can they get Sb/Obj (the visible dependency link does not lead to the verb of which they are subject/object). So they get the technical label ExD.

3. Stanfordization

3.1. Universal Stanford Dependencies in HamleDT

In the structure of the dependency tree, we try to fully adhere to USD as defined by de Marneffe et al. (2014). In copular constructions, the nominal predicate is the head governing the copula as a complement. The adpositions (and any other case-marking functional words) and subordinating conjunctions are structurally treated the same; they are governed by the word they introduce (see Figure 1). In
coordination constructions, the conjuncts are siblings except for the first one, which is the head, and the shared modifiers and coordinating conjunctions are governed by the head conjunct – see Figure 2 for an example.

We base our label set on USD; however, we add one new label, and a few other labels remain unused. The reason is not differing linguistic views, but rather a technical one – the relations marked by the unused labels cannot be reliably distinguished from others in PRG in a language-independent way, and some of them are indistinguishable even in the source treebanks. Figure 4 shows the hierarchy of USD with our modifications; additional labels are marked by + and the labels not used by us are marked by ∗. The reasoning behind these differences is given in Section 4.

USD assume the extension of the label set with language-particular labels as subtypes of existing labels. We leave this for future work; in HamleDT 2.0, we are only trying to convert the treebanks from PRG to the general label set, which should be a basis for future extensions thereof.

3.2. The conversion

Stanfordization of the harmonized treebanks consists of rehanging some of the nodes, and mapping the PRG labels to USD labels. The pipeline is implemented in the Treex framework3 and is language-independent, as all the language-specific and treebank-specific conversions take place already in the harmonization phase (Section 2.). Structural changes were performed for constructions that involve adpositions, copula verbs, coordinations, subordinations, and punctuation – in all of these cases, USD differ from PRG in their principle of nodes considered as auxiliary being leaf nodes.

A USD label for a node is devised based mainly on its PRG label and its part of speech, often also taking into account the label and part of speech of its parent, and, with verbs, the finiteness and passiveness marked in the Interset morphological representation. A basic overview of the mapping is provided in Table 2.

4. Discussion

In this section, we discuss the differences between USD and PRG, the effect of these differences on the process of stanfordization of HamleDT, and specify resulting implications for future work.

Both PRG and USD have a common goal of providing a linguistic formalism that can capture the syntax of sentences across diverse languages. Both of them were originally designed primarily for one language only (SD for English, PRG for Czech) and had to be adapted for the multilingual setting; in both cases most of the adaptation happens in the label set, while the tree structure remains largely unchanged.

SD were adapted to USD by simplifying and generalizing the original label set to capture important phenomena attested in a significant proportion of languages, intricately building upon the hierarchical structure of the labels to allow and encourage language-specific extensions of the label

---


---

Figure 4: USD as used in HamleDT 2.0.

set by adding subtypes of the universal types, e.g. poss for possessiveness marking ("’s" in English) as a subtype of case. This allows great flexibility for accurately representing the sentences of various languages while always
having a common backbone of the universal types. (We have not yet resorted to introducing such subtypes in stan-
fordized HamleDT, as we decided to first focus on having at least the common backbone.)
The adaptation of PRG labels in HamleDT has been con-
siderably less systematic, trying to map the phenomena in the various languages onto the PRG labels by choosing one closest in its definition, introducing a new label only very rarely. This is one of the reasons that harmonization to PRG labels leads to loss of information in some cases, such as in case of the direct/indirect object distinction which PRG does not capture even if the source treebank does.
The two formalisms themselves differ in many aspects, especially because of the following reasons:

**The underlying theory.** PRG build upon the Function-
Generative description by Sgall (1967), which under-
stands the shallow syntax tree as only one of the an-
notation layers, built on morphological layer and lead-
ing to the tectogrammatical (deep-syntactic) layer. USD build upon Lexical-Functional Grammar of Bresnan (2001).

**The main goal.** USD try to make the representation easy
to use for the user in applications such as Information Retrieval, while PRG aim at being as linguistically ac-
curate as possible, even providing quite complex an-
notation rules rather than resorting to simplifications.

Thus, converting PRG annotations to USD annotations is not a trivial task. While there are some cases of 1:1 map-
ing of labels with no rehanging of nodes, this is usually not the case, and one has to be creative when trying to map the labels; the transformations of tree structure are less common and usually easier to perform. We therefore broadened the definitions of some of the labels to allow us to use them for many less-common phenomena, and we currently have one extra label and 13 unused labels from the perspective of USD. We illustrate some non-trivial conversion issues in the following paragraphs; a graphical representation of frequencies of the individual labels in the treebanks after stanfordization is shown in Figure 5.

In future, we plan to investigate whether the source tree-
banks contain information that is lost in harmonization and is then missing for accurate stanfordization, which will probably lead to enrichment of PRG so that no informa-
tion deemed important is lost and the stanfordization can remain language-independent. In this paper, however, we limit ourselves to exploring the extent to which USD repre-
sentation can be devised from the current PRG.

### 4.1. Coordinations

Coordinating structures are known to be difficult to cap-
ture by dependency trees, as the coordination itself is not a dependency relation. The properties of coordination rep-
resentation in PRG and SD and conversions thereof have already been studied by Popel et al. (2013), showing PRG to have more expressive power than SD; thus, conversion of coordinations structures from PRG to USD is easy, although lossy.

The main advantage of PRG over (basic) SD is the ability
to distinguish private and shared modifiers. It is arguable whether this distinction should be captured in syntactic an-
notation, because the construction is often truly ambiguous, as in “green tables and chairs”. In other cases, disambigua-
tion can be done based on semantics, as in “juice from black currant and bananas” versus “we serve fish steaks and fin-
gers”; it is still questionable whether the disambiguation is to be made on syntactic level, but the same can be said about e.g. PP attachment. However, there are still other cases, where the construction is clearly non-ambiguous, as in “Peter plays and sings” vs “Peter plays and Mary sings”.

Therefore, we believe that an annotation style that can capture the private/shared modifier distinction is superior to one that can not.

On the other hand, many of the source treebanks do not use PRG, the aforementioned distinction is often not present in the source treebank in the first place, and heuristics had to be used to convert these into PRG. From that point of view, SD are a better fit, as no heuristics are neces-
ary to convert most of the coordinations into them.

In future, we would like to explore the extended SD as well (de Marneffe and Manning, 2008), as these are able to make the private/shared modifier distinction, among other benef-
fits they bring.

### 4.2. Objects (adding a label)

USD distinguish direct and indirect objects, which PRG do not. De Marneffe et al. (2014) make a point by attesting that this distinction can be traced in many languages, while in PRG, such distinctions are not made, as the notion of direct and indirect object is not usual in Czech linguistics. We want to address this shortcoming in future by expl-
oring the source treebanks and extending PRG appropriately to capture object type distinctions that are common across languages. Currently, we simply introduce an additional **obj** label, which we use instead of **dobj** and **iobj**.
4.3. Currently unused labels

There are thirteen USD labels we do not use. They can be divided into the following categories:

- 
  - `dobj`, `iobj`, `name`, `foreign`, and `vocative` are unused because we cannot make the necessary distinctions based on the current PRG harmonization. We are planning to revisit the harmonization process to see whether we could get the necessary information from the source treebanks, as we find these labels useful.
  - `discourse`, `goeswith`, `reparandum`, and `list` also cannot be distinguished in PRG; however, we are still considering whether to try to include these labels in future or not, as we find them to be of little importance. We believe that using such labels is useful when annotating a new treebank, as the respective phenomena are typically hard to label consistently, but if the source treebank did not choose to use them, it will be nearly impossible to detect and label them in the harmonization; moreover, the added value brought by such an effort would be rather limited in our opinion.
  - `ncmod`, `expl`, `dislocated`, and `parataxis` are currently unused, as we are unsure about their exact definition in the multilingual setting. We will reconsider their inclusion once this is clearer.

4.4. Broadening the definitions of labels

During the conversion, we found that the definitions of labels by de Marneffe et al. (2014) are rather narrow, for instance frequently distinguishing clausal and nominal elements but disregarding other possibilities. This is insufficient for many languages, including English – consider e.g. the sentence “Quickly does not always mean efficiently.”, where both the subject and object are adverbs and should probably be classified neither as nominal nor clausal. We therefore broaden the definition of `nsubj` and `[di]obj` from “nominal” to “non-clausal” subject and object, as we believe the clausal/non-clausal distinction to be of a higher importance than the nominal/non-nominal distinction. Similar generalizations are made for other labels where necessary, which we prefer to defining new labels that would be only slightly different from the existing ones and at the same time very rare in the data. In future, such distinctions may be made by defining new subtypes of existing labels if the current label set is found to be too coarse.

4.5. Clauses

It is a non-trivial task to correctly identify and classify clauses. Following the definitions of de Marneffe et al. (2014), we treat every non-auxiliary verb as a clause head. PRG do not help us in classifying a clause as finite or non-finite; therefore, we resort to inspecting the Interset morphological representation of the clause head, and annotate a clause as finite if it is headed by a verb form marked as finite. This is clearly an approximation, as even finite clauses can be headed by infinite verb forms in many languages, their finiteness being marked by the auxiliary verbs; in other languages, the infiniteness is marked by a particle, similar to English “to”. Consequently, the morphological annotation in some source treebanks does not always dis-
tistinguish finite and infinite verb forms; currently, we treat
the verb as finite in such cases, as we have no language-

dependent way of making the distinction. It is also not
clear how to treat verb forms that are marked as neither
finite nor infinite, such as participles or transgressives; fol-
lowing de Marneffe et al. (2014), we currently treat them
as infinite, but this might not be appropriate for all languages.
Thus, the necessary distinction is currently hard to make in
a language-independent way, and it is clear that the so far
rather neglected harmonization of verbs in HamleDT will
have to be improved.

5. Conclusion

In this paper, we presented HamleDT 2.0, a collection of
30 pre-existing treebanks, which we automatically harmo-
nized into PRG, and subsequently converted to USD. We
encountered several issues during the conversion process,
leading us to make slight modifications to the USD defini-
tion and leave some future work for us in improving both
the harmonization step and the conversion step.

We release 13 of the harmonized and converted treebanks
at http://ufal.mff.cuni.cz/hamledt. The li-
censes for the rest of the original treebanks do not allow us
to redistribute them, but our harmonization and conversion
pipeline is available freely.

6. Acknowledgements

This research was supported by the grants FP7-ICT-2013-
10-610516 (QTLB), SVV 260 104, Czech Science Founda-
tion grant no. P406/14/06548P, and GAUK 1572314.
This work has been using language resources developed,
stored and distributed by the LINDAT/CLARIN project of
the Ministry of Education, Youth and Sports of the Czech
Republic (project LM2010013).
We would also like to thank de Marneffe et al. (2014) for
providing us with a draft of their paper, and apologize for
any inaccuracies in referring to it due to its possible later
changes that we were not aware of.

7. References

Aduriz, Itzair, Aranzabe, María Jesús, Arriola, Jose Mari,
Atutxa, Aitziber, Díaz de Iiarraza, Arantxa, Garmendia,
Basque dependency treebank. In Proc. of the 2nd Work-
shop on Treebanks and Linguistic Theories.

Afonso, Susana, Bick, Eckhard, Haber, Renato, and San-

Atalay, Nart B., Ofllater, Kemal, Say, Bilge, and Inst, Infor-
matics. (2003). The annotation process in the Turkish
treebank. In Proc. of the 4th Intern. Workshop on Lin-
guistically Interpreted Corpora (LINC).

Bamman, David and Crane, Gregory. (2011). The Ancient
Greek and Latin dependency treebanks. In Sporleder,
Caroline, Bosch, Antal, and Zervanou, Kalliopi, editors,
Language Technology for Cultural Heritage, Theory and
Applications of Natural Language Processing, pages 79–
98. Springer Berlin Heidelberg.

Bejček, Eduard, Hajičová, Eva, Hajič, Jan, Jínová, Pavlína,
Kettnerová, Václava, Kolářová, Veronika, Mikulová,
Marie, Mírovský, Jiří, Nedoluzhko, Anna, Panevová,
Jarmila, Poláková, Lucie, Ševčíková, Magda, Štěpánek,
Jan, and Zíkánová, Šárka. (2013). Prague dependency

treebank 3.0.

Croatian dependency treebank: Recent development and
initial experiments. In Proc. of LREC’12, Istanbul,
Turkey. European Language Resources Association
(ELRA).

Arborest – a VISL-style treebank derived from an Eston-
ian constraint grammar corpus. In Proc. of Treebanks
and Linguistic Theories.

Boguslavsky, Igor, Grigorieva, Svetlana, Grigoriev, Niko-
loi, Kreidlin, Leonid, and Frid, Nadezhd. (2000). De-
pendency treebank for Russian: Concept, tools, types of
information. In Proc. of the 18th conference on Com-
putational linguistics-Volume 2, pages 987–991. Associa-
tion for Computational Linguistics Morristown, NJ,
USA.

Brants, Sabine, Dipper, Stefanie, Eisenberg, Peter, Hansen,
Silvia, König, Esther, Lezius, Wolfgang, Rohrer, Chris-

TIGER: Linguistic Interpretation of a German Corpus.
Special Issue.

Bresnan, Joan. (2001). Lexical-functional syntax. Wiley-
Blackwell.

Csendes, Dóra, Csirik, János, Gyimóthy, Tibor, and Koc-
sor, András. (2005). The Szeged treebank. In Matoušek,
Václav, Mautner, Pavel, and Pavelka, Tomáš, editors,
TSD, volume 3658 of Lecture Notes in Computer Sci-
ence, pages 123–131. Springer.

Călancean, Mihaela. (2008). Data-driven dependency pars-

de Marneffe, Marie-Catherine and Manning, Christopher D. (2008). The Stanford typed dependencies rep-
resentation. In Coling 2008: Proc. of the workshop on
Cross-Framework and Cross-Domain Parser Evalu-
ation, pages 1–8. ACL.

de Marneffe, Marie-Catherine, MacCartney, Bill, and Man-
ning, Christopher D. (2006). Generating typed depend-
cy (TSD) from phrase structure parses. In Proc. of
LREC, volume 6, pages 449–454.

de Marneffe, Marie-Catherine, Connor, Miriam, Silveira,
Natalia, Bowman, Samuel R., Dozat, Timothy, and Man-
ning, Christopher D. (2013). More constructions, more
Second International Conference on Dependency
Linguistics (DepLing 2013), pages 187–196. Charles
University in Prague, Matfyzpress, Prague, Czech Re-
public.

de Marneffe, Marie-Catherine, Silveira, Natalia, Dozat,
Timothy, Haverinen, Katri, Ginter, Filip, Nivre, Joakim,
and Manning, Christopher D. (2014). Universal Stan-
of LREC’14, Reykjavik, Iceland. European Language


