

# Supplemental Material

## A Typology of Syntactic Constructions

We summarise the names of the languages appearing in the paper and their corresponding ISO 639-1 codes in Table 1. Subsequently, we provide examples of different strategies in syntactic constructions to illustrate how pervasive their impact may be on UD.

Arabic	AR	French	FR	Kazakh	KK
Basque	EU	Greek	EL	Korean	KO
Bulgarian	BG	Hebrew	HE	Latvian	LV
Chinese	ZH	Hindi	HI	Portuguese	PO
Croatian	HR	Hungarian	HU	Slovenian	SL
Danish	DA	Indonesian	ID	Russian	RU
Estonian	ET	Irish	GA	Turkish	TR
Finnish	FI	Italian	IT	Vietnamese	VI

Table 1: Languages used throughout this work and their corresponding ISO 639-1 codes.

**Polar questions** are signaled by a particle at the beginning of the sentence in Arabic (*hal* or *a*), by word order in Dutch (verb in first position), by a suffix in Indonesian (*-han*), and by prosodic intonation in Portuguese.

**Predicate possession** consists in an intransitive verb with existential meaning (*ada*) in Indonesian and a non-verbal predication in Arabic, the subject of both being the owned item. In the former, the possessor is prefixed by some locative adpositions coalescing with the noun (such as *'ind-*, *li-*, *ma'a-*), in the latter it is a topic (positioned before the verb and unmarked).<sup>1</sup> In Dutch and Portuguese, a specific transitive verb expresses possession (*hebben* and *ter*): its subject is the possessor and its object is the owned item.

**Relative clauses** in Arabic depend on the definiteness of the relativised noun: if definite, a rela-

<sup>1</sup>In Indonesian, the verb does not count as copula *cop* because equational predications in Indonesian use different copulas (such as *adalah*) or different constructions (null copula with optional focalized predicate).

tive pronoun (*allabī*) with the same function of the antecedent is used, if indefinite nothing links the clauses (hence the name gap strategy). In the second case, indefinite pronouns resume the antecedent if it has a role different from subject. In Indonesian, the same optional particle (*yang*) connects a noun with relative clauses and adjectives as *mark*. In Dutch and Portuguese, a relative pronoun (*dat/die/wat/wie* and *quel/quem/lo qual*, respectively) is used instead, with a function corresponding to the one it carries within the subordinate clause.

## B Morphology and Parsing

In the paper, we showed that compatibility of morphological inventories is crucial for cross-lingual parsing with DeSR. In Figure 2 we show that the same holds true for SyntaxNet. In fact, the degree to which languages convey information with morphology varies: to shed further light on this variation, we also perform an ablation study of morphology on monolingual parsers (DeSR and SyntaxNet). In particular, we compare their performances with two different feature sets. The first set includes forms, lemmas, POS tags, morphological features, and dependency relations of items on the stack. The second set leaves out all the morphological features.

The ablation study yields results (Figure 1) that are almost identical for the two parsers and demonstrate that the degree to which languages bestow information to morphology is uneven. Languages where LAS drops more dramatically are agglutinative (Basque, Estonian, Finnish, Hungarian, and Turkish) or fusional with a rich inventory of grammatical categories and values (Croatian, Latvian and Russian). Interestingly, UAS scores are affected in proportion but more mildly. This implies that morphology contributes especially to classi-

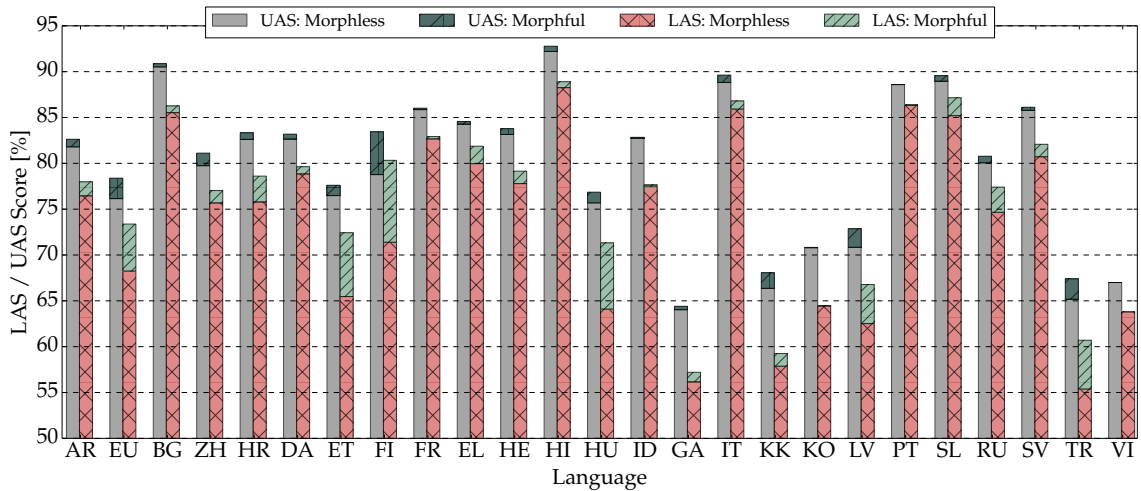


Figure 1: UAS scores (left bars) and LAS scores (right bars) without (*morphless*) and with universal morphological features (*morphful*). Results are presented for the DeSR parser.

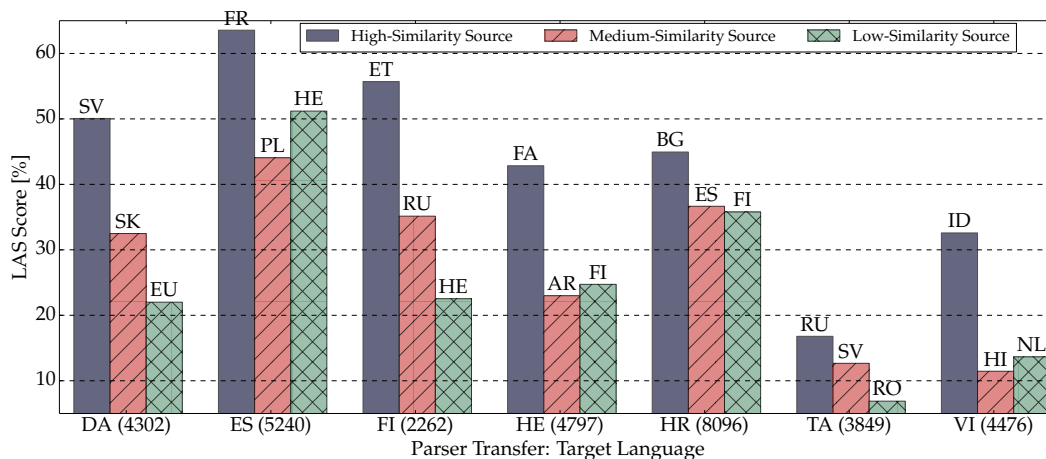


Figure 2: Results of monolingual SyntaxNet.

ifying dependency relations correctly, and is somewhat superfluous for the correct unlabeled attachments.

These results extend the conclusions reached in the shared tasks SPMRL 2013 and 2014 (Seddah et al., 2013, 2014) to a larger and more diverse set of languages, and based on a language-independent annotation scheme. This additional experiment clarifies the cross-lingual parsing task in the paper, as it shows how one of the sources of anisomorphism (morphological richness) affects performances even in mono-lingual settings, by making some languages easier to parse without morphological features than others.

## References

- Djamé Seddah, Sandra Kübler, and Reut Tsarfaty. 2014. Introducing the spmrl 2014 shared task on parsing morphologically-rich languages. In *Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages*, pages 103–109.
- Djamé Seddah, Reut Tsarfaty, Sandra Kübler, Marie Candito, Jinho Choi, Richárd Farkas, Jennifer Foster, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, et al. 2013. Overview of the spmrl 2013 shared task: cross-framework evaluation of parsing morphologically rich languages.