This paper: opportunities & challenges for unifying these frameworks. We offer short-, medium-, and long-term recommendations.

**CHALLENGE #1**: Scope of what is identified as a multiword expression (MWE)
- In MWE community, defined in terms of morphosyntactic and/or semantic idiosyncrasy. PARSEME has developed rigorous crosslinguistic guidelines + corpora for categories of verbal MWEs: Inherently Reflexive Verbs, Verb-Particle Constructions, Multi-Verb Constructions, Light Verb Constructions, Verbal Idioms.
- In UD guidelines, used loosely as a cover term for fixed, flat, compound relations (+ in some languages, subtypes like compound:lv, expl:pv). But not all compounds are idiosyncratic.

**short term**: dispense with the usual use of “MWE” in the UD guidelines

**medium term**: extend PARSEME work beyond verbal MWEs to include nominal MWEs, multiword connectives, etc.; consider relationship to named entities

**long term**: UD: better guidelines for productive grammatical subsystems like templatic named entities, numbers, measurements, dates; PARSEME: partially productive constructions (as in Construction Grammar)

**CHALLENGE #2**: UD tokenization is sometimes too coarse to capture idiomatic combinations (e.g., synthetic compounds).
- German & Korean examples where the ideal MWE annotation is more granular than UD syntactic words

**short term**: indicate subword character spans in PARSEME annotation

**medium term**: disentangle things like :lvc and :pv, which are MWE classifications, from the syntax by moving them to an MWE layer; address inconsistencies in some of the other deprels

**long term**: implement a finer-grained notion of word in UD. Splitting synthetic compounds would also disambiguate cases like Swedish *bildrulle*: *bil*d*rulle ‘car maniac (bad driver)’ vs. *bild*rulle ‘picture roll (roll of film)’.

**CHALLENGE #3**: Idiosyncrasy at the lexical type level is not always reflected at the token (occurrence) level.
- MWEs can have regular syntax, even if the meaning is idiomatic and the variability of the type is restricted (fossilization).
- UD mostly targets token-level analysis, and is agnostic to type-level variability or meaning. But this is muddled by labels like fixed, compound:lv, expl:pv, compound:prt vs. advmod.

**short term**: merge fixed and flat under a new label, headless?

**medium term**: link token occurrences in corpora to entries in a lexicon

**long term**: link token occurrences in corpora to entries in a lexicon

---

**Parts of unsplit tokens (MWT) contains several morphological terms as units bearing morphological properties other. In UD, words are de-... customers. For instance, in Fig. 6: Complex names with mixed dependencies.

**Figure 6: Complex names with mixed dependencies.**