

What is new in the annotation guidelines version 6.0 with respect to version 5.0?

PARSEME shared task on automatic detection of verbal MWEs

Agata Savary

7 August 2016

1. Decision tree

The [guidelines](#) now contain three additional decision trees (one for VMWE identification, one for VMWE categorisation and one LVC-specific) which lead the annotator through the categorisation process.

2. FAQ

The guidelines are now linked to a [Frequently Asked Questions](#) section.

3. Obsolete IPrepV category

This category (previously covering cases like *to rely on, to come across*) has disappeared since no reliable tests could be found to distinguish selected prepositions (belonging to the phenomenon of valency, and thus handled by a regular grammar) from inherent ones. Now, a preposition is part of a VMWE only if introduces a lexicalized complement (e.g. *to take somebody **by surprise***).

4. VPC category

It is now extended to one-token idiomatic particle+verb combinations, e.g. (DE) *vor**ber**eiten, auf**ge**macht*. The guidelines now contain not only tests discriminating prepositions from particles, but also VPCs from compositional verb+particle combinations.

5. IRefIV category

This category was previously called IPronV. The renaming is meant to signal that only combinations of verbs with reflexive pronouns, e.g. (FR) *se **suic**ider* 'to suicide', and not with other pronoun, e.g. *to **make** it*, belong to this category. Examples in German are now also added to the IRefIV guidelines.

6. LVC category

Linguistic tests for this category have been largely enhanced and simplified. They are now organized into a (shallow) decision tree.

7. OTH category

This category now covers only VMWEs whose lexicalized components do not include a head verb and at least one of its lexicalized arguments (*to **drink and drive**, to **pretty-print**, to **voice act***).

8. Color codes

Color codes were introduced for VMWE examples cited in the guidelines. Germanic languages use nuances of green, Romance languages - of blue, Balto-Slavic - of yellow/orange, other languages - of violet/rose. **Red** color is used for examples which resemble VMWEs but do not have their status. Lexicalized components in each example are highlighted in bold.

9. Obsolete hesitation labels

In this version of the guidelines we no longer admit "hesitation labels" (e.g. LVC/ID or OTH/_) used in the pilot annotation. Hesitation can, however, be expressed in a comment and a particular value of the annotator's confidence assigned to a particular VMWE occurrence.