

## Relevant UniDive working group: WG1

### Objectives

- To define and refine the class of Inherently Adpositional Verbs (IAVs) in Romanian language;
- To offer a classification of the Ro IAVs, based on the PARSEME diagnostic test;
- To offer solid ground for improving the PARSEME guidelines wrt IAVs.

### State of the art

- PARSEME guidelines: IAVs consist of “a verb or a VMWE and an idiomatic selected preposition or a postposition that is always required or, if absent, changes the meaning of the verb or of VMWE significantly”; optional and experimental category;
- So far, IAVs have been annotated for English, Italian and Irish VMWEs corpora;
- Challenges raised by borderline cases between IAVs and other classes, i.e. verb-particle constructions (VPCs), light verb constructions (LVCs).

### PARSEME-Ro annotated with IAVs

- In a second, recent step, after the manual annotation of the universal and quasi-universal types of VMWEs (LVCs, full, LVCs.cause, VIDs and IRVs), during the 3 Shared Tasks;
- IAVs annotation was automatic, starting from a manually compiled list of 1,725 prepositional verbs (Geană 2013):
  - ❖ all their occurrences in the corpus were identified and annotated as IAVs, using a Python script;
  - ❖ span window – to capture situations where other tokens were interleaved with the actual IAV in the corpus text;
- Manual validation and correction, in two phases:
  - ❖ 2 students in linguistics, using an annotation platform – they modified the automatic annotations, by deleting expressions or adjusting their size (i.e. adding or removing parts), using the BRAT tool (Stenetorp *et al.* 2012), integrated in the RELATE platform (Păiș, *et al.* 2020) → several sources of errors could be identified in the automatic annotation of IAVs;
  - ❖ consistency of annotation:
    - ✓ all cases of agreement between the two student annotators were considered correct decisions (2,981);
    - ✓ all cases of disagreement between them were further checked by two linguists experienced in the PARSEME annotation;
    - ✓ two thirds of the automatically annotated IAV are actually correct IAVs; they represent 94.47% of the IAVs that should have been annotated = “gold IAVs”.

### IAVs annotation

- PARSEME Test IAV.1 – [CIRCUM-QUEST] Circumstantial question with no adposition: *in response to a declarative sentence with the verb-preposition combination, is there a natural way to query the circumstances of the verbal event using the verb, but not the adposition?*
  - **YES** → it is not an IAV
    - *I care about the environment. (Why do you care?)*
  - **NO** → annotate as an IAV
    - *I came across a nice restaurant downtown. (#When did you come?)*
- IAVs in the PARSEME-Ro corpus:
  - for some IAVs, the PARSEME test holds:
    - *Ne bazăm pe tine. 'We count on you.' (#Când ne bazăm? 'when do we count?')*
  - for other IAVs, the test does not hold:
    - *Am aplicat la facultate. 'I applied to college.' (Când ai aplicat? 'when did you apply?')*

## Proposal for a classification of Romanian IAVs (based on the PARSEME test)

### I. IAV. functional (test IAV.1 holds)

- verbs which never occur without the required preposition;
- 1. The verb is polysemous, but it always requires the same preposition, irrespective of its meaning: **a se baza pe** ‘rely on’, **a depinde de** ‘depend on’, **a consta în/din** ‘consist in/of’; => IAV for all its senses
- 2. The verb is polysemous, but it requires a certain preposition only when used with a certain meaning: **a da (peste)** ‘run into’, **a aparține (de)** ‘belong to’, **a conta (pe)** ‘count on’; => IAV only for the meaning in which the preposition is mandatory.
- 3. The verb is symmetrical/reciprocal and when used with a prepositional object (and not with a plural noun as subject, for example), it always requires the preposition *cu* corresponding to the English with: **a merge (cu)** ‘match with’, **a coincide (cu)** ‘coincide with’, **a colabora (cu)** ‘collaborate with’.

### II. IAV.semi-lexical (test IAV.1 does not hold)

- verbs that normally require a specific preposition;
- they are polysemous and when used with certain meanings, they are always followed by a specific preposition: **a aplica (la/pentru)** ‘apply for’, **a adera (la)** ‘adhere to’, **a-i păsa (de)** ‘care about’.

		correctly annotated	
#	Total IAVs	#	%
AUTO annot.	4,686	3,128	66.75
annot. 1	3,462	3,085	89.11
annot. 2	3,519	3,185	90.50
both annots	-	2,981	
gold IAVs	-	3,311	

General statistics of the IAV annotation process

Type	Number
IRV	3,826
LVC.cause	182
LVC.full	516
VID	1,644
IAV	3,311
TOTAL	9479

Number of VMWEs of each type