

Enriching the Romanian Collection of Negative Polarity Items (CODII-NPI.ro) with Negative-polarity Multiword Expressions (NPMWEs)

Monica-Mihaela Rizea (Bucharest), Gianina Iordăchioaia (Stuttgart), Frank Richter (Frankfurt)

Goals

- ❖ Update the Romanian Collection of Negative Polarity Items (started in 2005) and enrich it with NPMWEs. This is part of a multilingual resource hosting German, English and Romanian collections of distributionally idiosyncratic items. www.english-linguistics.de
- ❖ Classification of Romanian NPMWEs into superstrong/strong/weak (following Zwarts 1997)
 - ✓ NPMWEs are items with idiosyncratic distributional patterns (van der Wouden 1997)
 - ✓ Corpus profiles in terms of licenser-NPMWEs collocations (distributional dependence on the licensing contexts documented with frequency data and real-use examples from large Romanian corpora - such as Romanian Web Corpus or OPUS2 parallel corpora - via the Sketch Engine online tool)
 - ✓ No corpus examples - use linguist intuition

How do NPMWEs differ from regular MWEs ?

- ❖ NPMWEs are distributionally restricted lexical units that occur in the scope of negation, but also in a variety of other contexts such as questions, the scope of conditional operators and of downward-entailing expressions, in the restrictor of universal quantifiers, the scope of imperatives, etc.
- ❖ They are a theoretically and practically challenging class of MWEs because their obligatory licensers are not simple lexemes but abstract grammatical and semantic categories.

STRONG NPMWE ‘urmă de’ (lit. ‘trace of’); idiomatic meaning ‘no...at all’

```

<licensors>
  <cmn given="yes" hits="urmade1"/>
  <ncmn given="yes" hits="urmade2"/>
  <nw given="yes" hits="urmade3"/>
  <nici given="yes" hits="urmade4"/>
  <dent given="no"/>
  <nv given="yes" hits="urmade5"/>
  <que given="yes" hits="urmade6"/>
  <if given="yes" hits="urmade7"/>
  <without given="yes" hits="urmade8"/>
  <only given="yes" hits="urmade9"/>
  <univ given="yes" hits="urmade10"/>
  <comp given="no"/>
  <sup given="no"/>
</licensors>

```

Romanian Web Corpus – n° of words = 44,729,032

Query: *urmă de* ‘trace of’ N=1918

negative(-like) (A+B+C+D+E+F)	N=1493 (78%)
----------------------------------	--------------

A:nu (n-) ‘not’ NM as the only licenser	N=62 (3%)
--	-----------

B:nicio ‘no’	N=313 (16%)
--------------	-------------

C:nici ‘not even’	N=720 (38%)
-------------------	-------------

D:fără ‘without’	N=200 (10%)
------------------	-------------

E: NV	N=190 (10%)
-------	-------------

F: dacă ‘if’	N=8 (1%)
--------------	----------

positive	N=409 (21%)
----------	-------------

other (irrelevant strings - not the intended structure)	N=16 (1%)
---	-----------

Table 1. Corpus Profile – Distribution of *urmă de* with respect to its licensers

Corpus examples:

- (1) *Nu mai rămăsese urmă de veselie pe fața lui.*
 NM still left trace of joy on face his
 No trace of joy remained on his face (→No joy **at all** remained on his face).
- (2) *Nu se simțea nici urmă de umor în vocea lui Conk.*
 NM CL feel not.even trace of humour in voice DET Conk
 No trace of humour could be felt in Conk’s voice. (→No humour **at all** could be felt in Conk’s voice).
- (3) *Totuși, episcopul nu simțea nicio urmă de bucurie în suflet.*
 However, bishop.the NM feel no trace of happiness in soul
 However, the bishop felt **no trace of** happiness in his soul. (→The bishop felt **no hapiness at all**).
- (4) *Namarti zâmbi fără urmă de veselie: Nu ai încredere? Nu putem pierde.*
 Namarti smiled without trace of joy [...]
 Namarti smiled without a trace of joy (→There was **no joy at all** in Namarti’s smile).