

# WG1: Lexicon-Grammar Interface

6th General Parseme Meeting

Struga, 7.&8.4.2016

## Structure of the sessions

Session 9, Friday, 10:35-11:50: Multilingual MWE resources (Johanna)

Session 11, Friday, 13:50-15:05: WG1 activities

- WG1 overview
- ENeL/Parseme workshop Skopje (Simon, Carla)
- Publications (Carla, Manfred)
- Future of the template (Manfred)
- Working plan for Year 4

## Session 9: Multilingual MWE resources

<slides for this session in an extra file, see:

<http://typo.uni-konstanz.de/parseme/images/WorkingGroups/WG1/Monti-Multilingual%20MWE%20resources.pdf>

## Session 11

- WG1 member development
- Working plan year 3
- ENeL/Parseme workshop (Carla)
- Publications:
- Encoding paper, LREC (Carla)
- WG1 volume (Manfred)
- Future of the template
- Working plan year 4
- Preparing for Dubrovnik

## WG 1 members

- Members: 120
- ESR: 71

## WG1-related STSM in year 3

- Bargmann (DE->UK): Idiom Licensing in Non-restrictive Relative Clauses
- Jelena Mitrovic (RS->EL): Greek and Serbian MWEs in Ontology-Based Lexical Resources
- Sailer (DE->EL): MWEs: Insights from a Multilingual Perspective
- Savary (FR->PL): Polish Verbal MWEs and PARSEME

## Working Plan for Year 3: (till March 2016)

### Linguistic properties:

- Submit WG1 volume to LSP ~
- More languages in the template; changes: flexibility, dimensions ✓

### Lexical encoding:

- Adapt resource survey to ENeL-members ✓
- Examples from various languages and of various MWE types ✓
- ~~How-to-do material~~ (to be produced in the hands-on workshop)
- Publication of the lexical resources ✓ and lexical encoding surveys ~

### Others:

- Lexicon-grammar interface? ~
- Initiate decentral cooperations on other topics (lexical encoding subgroups etc) ?

### Interaction with other WGs:

- WG2: input for ~~abstract "test suite"~~
- WG3: Parseme+ENeL workshop; input for ~~abstract "test suite"?~~
- WG4. Parseme+~~ENeL~~ workshop

## ENeL/Parseme Workshop

<extra set of slides by Simon, see:

[http://typo.uni-konstanz.de/parseme/images/WorkingGroups/WG1/Krek-Report\\_ENeLParseme\\_final.pdf](http://typo.uni-konstanz.de/parseme/images/WorkingGroups/WG1/Krek-Report_ENeLParseme_final.pdf)

>

Publications: LREC, Encoding paper

## WG1 volume

- Received papers: 10
- Currently: reread revised papers
- Last round of revision: May
- Submission to LSP: June?
- If possible: In new LSP series

## MWE Templates

- Wiki: [http://wiki.studiumdigitale.uni-frankfurt.de/FB10\\_Parseme/index.php/Main\\_Page](http://wiki.studiumdigitale.uni-frankfurt.de/FB10_Parseme/index.php/Main_Page)
- Wiki only visible to registered contributors

## MWE Template

**A lot of data:** Croatian, English, German, Greek, Latin, Macedonian, Polish, Serbian, Slovak,

**Some data:** Romanian, Norwegian, Slovene, Spanish

**Empty/no contributor:** Bulgarian, Dutch, Czech, Danish, Dutch, Estonian, French, Guadeloupe Creole, Hebrew, Italian, Latin, Portuguese, Romanian, Swedish, Turkish

Most popular dimension: syntactic structure

Least discussed dimension: idiomaticity

Not filled in yet? semantic relations, rhetoric relations

## What to do with the material from the wiki?

- Turn into a publicly accessible resource
- Publications based on the material

## What to do with the material from the wiki?

- Create a publicly accessible resource (wiki?)
  - > old, password-protected wiki remains
  - > copy advanced MWE templates into the new wiki
  - > can be expanded in the future
  - > hosting should be guaranteed for the next 10 years
- Should it be a wiki?
- Change the structure of the templates?

## Possible publications

Type A: *A computationally oriented overview of the phraseology of XXX*  
(useful for languages with limited (computational) MWE resources)  
*Overview of the phraseology of XXX*

Type B: *Cross-linguistic issues in phraseology*

Comparing various aspects of the data collected in the templates

## Computationally oriented overviews

- Relatively analogue structure (following the template?)
- Combined with a small resource (for example: 150 MWEs)
- Candidate languages?
- Publication venue? Individual papers or volume?



## MWE overviews:

- Example structure:
  - > Introduction: Language XX
  - > State of the art:
    - Computational resources of XX
    - Phraseological research on XX
  - > MWEs in XX
    - Syntactic categories
    - Tests for syntactic flexibility
    - Idiomaticity
    - Semantic relations
    - Rhetoric relations
  - > Conclusion
  - > References
  - > Appendix: 150 selected MWEs

Proposed structure available on google docs, write to  
sailer@em.uni-frankfurt.de to get the link!

## Cross-linguistic considerations

- Start with one overview paper?
- Various papers on individual aspects?

## Possible topics of the cross-linguistic overview

- Based on advanced templates
- Based on topics discussed in parseme contributions

## Possible structure

- Introduction
- Cross-linguistic differences in the domain of MWEs: compounds, particle verbs, one-word MWEs; computational vs. grammar tradition
  - A: Lexical variation in MWEs: none/restricted/open slots; predicative MWEs, *give-get-have* alternation, etc.
  - B: Structural peculiarities: encoding of possessive relations; MWEs in non-canonical syntactic structures
  - C: Syntactic flexibility: useful from a cross-linguistic or computational perspective?
  - D: Idiomaticity
  - E: Semantic relations
  - F: Rhetoric relations
- Conclusion

## Working plan for year 4

### Memorandum of understanding:

#### objectives:

- a better understanding of linguistic properties of MWEs, in particular at the lexical and syntactic level,
- enhancing the usability of MWE lexicons and valence dictionaries in parsing,
- paving the way towards interoperability of lexicons and the reduction of their production cost.

#### aspects to be studied:

- How to account for the fixed character of MWEs with respect to some linguistic phenomena on the one hand, and their similarities to regular syntactic structures on the other hand?
- In particular, how to represent, at the lexical level, phenomena most relevant to parsing, i.e., agreement, discontinuity and free word order?
- How should MWE lexicons be structured to be easily convertible and maximally reusable in different frameworks?

## Memorandum of understanding:

expected outcomes:

- reports on the contrastive analysis of lexical and syntactic properties of MWEs indifferent European languages,
- already existing lexicons and valence dictionaries enhanced and enriched with MWEs, for several European languages,
- design proposals for cost-saving abstract models of MWEs' properties, such as meta-grammars that could be automatically mapped to different lexicon and grammar formalisms; these models would apply to different languages in question.

## Working Plan for Year 3: (till March 2016)

Linguistic properties:

- Submit WG1 volume to LSP
- More languages in the template; changes: flexibility, dimensions

Lexical encoding:

- Adapt resource survey to ENeL-members
- Examples from various languages and of various MWE types
- How-to-do material (to be produced in the hands-on workshop)
- Publication of the lexical resources and lexical encoding surveys

Others:

- Lexicon-grammar interface?
- Initiate decentral cooperations on other topics (lexical encoding subgroups etc)

Interaction with other WGs:

- WG2: input for abstract "test suite"
- WG3: Parseme+ENeL workshop; input for abstract "test suite"?
- WG4. Parseme+ENeL workshop

## Working plan year 4

Linguistic properties:

- submit WG1 volume to LSP
- Transform the templates; work on papers.

Lexical encoding/resources:

- Finalize papers
- Multilingual MWE survey
- Interaction with ENeL

Others:

Interaction with other WGs:

WG2: Training school, Przepiorkowski/Patejuk: Verbal phraseology:  
from a valency dictionary to a parser

WG3: multilingual resources survey

WG4: annotation guidelines for verbal MWEs (shared task)

## Preparing for Dubrovnik

- WG3/WG1: Multilingual resources of MWEs
- WG session Lexicon: Overview publications, encoding publications
- WG session Grammar: Template-Wiki, Papers

