

WG 1: Lexicon-Grammar Interface

4th General Parseme Meeting

Valletta, 19.&20.3.

Structure of the sessions

Session 10, 9.00-10.30

1. Opening
2. General remarks on the WG1 volume
3. Poster presentations for WG1 volume
4. Poster display

Structure of the session:

Session 11, 11.00-11.30

1. Report on surveys:
 1. Surveys on lexical resources
 2. Work on the templates
 3. WG-internal survey
2. MWEs, semantic relations, ontologies
 1. Input by Jelena Mitrović
 2. Discussion
3. Working plan for Year 3 (March '15 - March '16)

Session 10: 9:00-10:30

9:00-9:15 Welcome, sketch of the WG1 volume

9:15-9:45 Poster presentations for potential chapters

9:45-10:30 Poster session

Multiword Expressions: Insights from a Multilingual Perspective

- Initiated at the 3rd General Meeting
- Coordinators/editors: Stella Markantonatou, Manfred Sailer
- Publication plan: Edited volume with *Language Science Press*, series *Empirically Oriented Theoretical Morphology and Syntax*.
- Time line:
 - > Call: November 2014
 - > Outline submissions: January 2015 (received: 24, invited for submission: 16)
 - > Outline presentations: today!
 - > First versions due: August 1, 2015
 - > Reviews due: November 2015
 - > Final versions due: January/February, 2016

Purpose of the book

- MWEs are challenge for natural language applications and to linguistic theory.
- There is a rich body of descriptive work on MWEs for many European languages, but little comparative work
- The volume attempts to bring together MWE experts with individual languages as their background to explore the benefits of a multilingual perspective on MWEs, for both computational and theoretical linguistics.
- Topics: Classification, tests for classifications, special types of MWEs, cross-linguistic comparison of MWE-types

Purpose of today's presentations

- Inform about the content of the volume.
- Get feedback on the potential chapters.
- Get data from other languages for the chapters -> Strengthen the multi-lingual aspect

Problems:

- Not all potential chapters represented
- Videos, non-presented posters (Manfred), copies of non-presented outlines
 - > Send your feedback to the authors directly!

Posters with local presenters

- Sascha Bargmann & Manfred Sailer: Syntactic Flexibility of Non-decomposable Idioms
- Carla Parra Escartín et al.: Spanish Multword Expressions: typology and cross-lingual analysis from a traductological viewpoint
- Patrick Hanks et al.: Automatic extraction of MWEs for the Pattern Dictionary of English Verbs
- Cvetana Krstev et al.: Semantic and Syntactic Patterns of Multiword Names (a Cross-language Study)
- Petya Osenova et al.: Multiword Expressions: observations from a parallel BG-EN newsmedia corpus

Posters without local presenters

- Koenraad Kuiper: Parameters of variation in the cross-linguistic analysis of phraseology
- Éric Laporte: Choosing features for classifying multiword expressions
- Verginica Barbu Mititelu et al.: Derivation and multi-word expressions
- Jelena Parizoska: Idiom variation in English and Croatian

Additional papers

- Lars Borin: Is a cross-linguistic typology of multiword expressions useful or even possible?
- Natalia Filatkina: MWEs cross-linguistically: benefits and challenges from a historical perspective
- Aggeliki Fotopoulou & Voula Giouli: MWEs and the Emotion Lexicon: Typological and cross-lingual considerations
- Ho Jia Qian et al.: Losing one's mind over meaning: Analysing the decomposability of possessive idioms
- Vesna Pajic et al.: Multiword Term Extraction: The Multilingual Perspective
- Elisabeth Piirainen: MWEs from a multi-lingual perspective: New insights into syntactic, lexical and semantic parallels of multiword expressions in diverse languages

Session 11: 11:00 - 12:30

11:00-11:30 Overview over activities and discussion of survey results

11:30-12:00 Semantic relations, ontologies, ...

12:00-12:30 Working plan for year 3

Results of the surveys

Summary of ongoing and planned activities

- Wiki:
 - > Examples of lexical entries from existing MWE resources
 - > MWE template
- MWE resource survey
- Workshop on lexical encoding (September 2015, Iasi)
- SOA survey paper on lexical formalisms (planned)
- **SOA lexical encoding**
- Volume on "MWEs: Insights from a Multi-lingual Perspective" (in progress)

Examples of lexical entries on the wiki

Information sheets for resources at:

http://wiki.studiumdigitale.uni-frankfurt.de/FB10_Parseme/index.php/lexical_Resources

NorGram (Norwegian): with examples

SemLex (Czech): with examples

English DELA (English): with examples

SEJFEK (Polish): with examples

SAWA (Polish): with examples

Named entities in the National Corpus of Polish (Polish): one example

Greek verb MWEs (Greek): 1100+ MWEs with paraphrase

Greek MWEs (Greek): no examples yet

MWE Template

A lot of data: English, Greek, Macedonian, Norwegian, Serbian, Slovak,

Some data: German, Polish, Slovene,

Empty: Bulgarian, Dutch, Hebrew, Italian, Latin, Portuguese, Romanian, Turkish

No contributor: Croation, Czech, Danish, Estonian, French, Guadeloupe Creole, Swedish

Most popular dimension: syntactic structure

Partly controversial dimension: syntactic flexibility

Least discussed dimension: idiomaticity

Additional dimension(s)? semantic relations

Lexicon subgroups

Detailed overview with links to all subgroups:https://docs.google.com/document/d/1oOjdNYjbbddUsrLJpb7FPZpV9z-wl9sLjwnqpUG_guk/edit?pli=1



1. MWE resource survey
2. Lexical descriptions
3. Lexical encoding formalisms
4. Ontology
5. Lexicon-grammar interface

1. MWE resource survey

- The survey is still open: <http://goo.gl/iUTc06>
- When closed, the data must be organized into a database.
- Is it realistic to aim for a publication, e.g. a presentation at the MWE workshop 2016? Other suggestions?

Working document: https://docs.google.com/document/d/1JH_WFZSBmOzPOMsKbPre-B8JyFKRNofJvwhZqPFDJNA/edit?usp=sharing

2. Lexical descriptions

Done on the wiki: http://wiki.studiumdigitale.uni-frankfurt.de/FB10_Parseme/index.php/Lexical_Resources



- SOA article on lex. representation of MWEs: Agata, Adam, Agnieszka, Gyri, Shuly at JoLM
- Collecting examples on the wiki
- Hands-on workshop in Iași:
 - > Generating more examples from different languages
 - > How-to-videos for simple and complex examples

3. Lexical encoding formalisms

1. Hands-on workshop in Iași
2. Exchange within the subgroup
3. Aim: Publication if enough active interest
4. Contact: Gyri

4. Ontologies

-> in a few minutes

5. Lexicon-Grammar interface

Target these WG1 objectives:

- 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.
- 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.

No direct coordinated activities yet

Semantic relations, ontologies

MWEs and semantic relations

Today's discussion:

1. Input presentation by Jelena Mitrović + short discussion
2. Is this a topic we should integrate into WG1?
3. Do we want to formulate a goal/subtask for our year three working plan?
4. Do we want to include it as a dimension in the MWE template?

MWEs, semantic relations, ontologies

Input: Jelena Mitrović

Discussion:

mwe in normal lexicon

stella: framenet ontology

kiril: formal all lang, data as ontologies

-> Christian Chiarcos (GU)

MWEs, semantic relations, ontologies

- Internal survey: Interest in "semantic relations", classification:
 - > Ontologies
 - > Classification
 - > Translation, multilingual perspective
 - > Encoding of MWEs in conceptual lexical and linguistic ontologies
 - > Networks of related idiomatic MWEs (relations on a formal and/or semantic level) (*for German*)
 - > Classification, description models, method of identification and delimitation of MWEs
 - > Classifications of synonymous and antonymous MWEs
 - > Linking MWEs different on a morphosyntactic level into a unit of language (meaning level). E.g., in Polish there are several verbal MWEs that mean 'to die': how to link them on the semantic level of language description?

MWEs and semantic relations

A topic for WG1?

A goal/subtask for year 3?

Working Plan

- Working plan year 2
- Suggestions from the WG-internal survey
- Working plan year 3

Suggestions for topics/directions from the survey

- Focus more on the interface between lexicon and grammar than working in subgroups
 - > Mapping from lexical resource to lexica for computational grammars
 - > Topic 5 of the lexicon subgroup; interaction with WG 2
- Typology of MWEs:
 - > starting from syntactic structure (LMF/Duelme)
 - > starting from semantic properties (sem. relations)
 - > starting from syntactic flexibility (WG2?)
 - > Borin's and Kuiper's contribution to the WG1 volume

Comments: "fixedness"

Heuristics in our template: (Sag et al. 2002 and Baldwin/Kim 2010)

- *Fixed MWEs*: no flexibility allowed whatsoever.
- *Semi-fixed MWEs*: only meaningless, obligatory variation allowed.
- *Flexible MWEs*: variation allowed that indicates that part of the MWE should be assigned meaning.

Problems:

- Depending on the language and the synt. category, there may not be any processes to help classify. Which type should be chosen?
- Passive: For English used as test for "flexible", German/French: also "semi-fixed" MWEs
- Obligatory variation: German V2, fronting in root clauses: Not completely "meaningless", but very weak constraints
- Similar observations for French, Italian, Estonian

Solution?

"flexible"

Greek: flexibility criteria don't work for the entire expression

Slovenian: flexibility: not usable

Polish: flexibility not useful

"totally fixed - not totally fixed" might be useful

Comments: "fixedness"

Solutions?

- stick to the classification, because it is widely used in the literature
 - Eric Laport's volume contribution
 - but specify for each language, which operations are taken as diagnostics for the degrees of fixedness
 - reject fixedness altogether and rather stick to clearly observable properties such as syntactic structure
 - leave it open to each template to which of the dimensions is filled in (ideally with a remark if a particular dimension is not used for principled reasons or just for reasons of time/resources
 - ...
- diagnostics instead of 3 labels
- but: maybe: which diagnostics correspond to which label

Suggestions for further topics

- Classification of MWEs
- Comparison of different formalisms -> *Hands-on in Iași*
- Building bilingual lexicons of idioms
- Take some entries from existing open-source lexicons and compare them (at least for languages belonging to close families of language) -> *Hands-on in Iași*
- A synchronic study of phraseological borrowings and calques in European languages -> *Contact to Piirainen? (in WG1 volume)*
- Description of MWUs in lexicons, particularly multi-word verbs how could they be described for a language with rich inflection and free word order in the lexicon.
- Work towards a broader typological (as in WALS-style language typology) treatment of MWEs.

Send a message to our mailing list to ask for people to join in on your favorite topic! (Use the BSCW-folder or our wiki to sketch the topic/ coordinate the work)

What do we do with data collected in WG1?

- Examples of MWE encoding in various frameworks
- MWE examples with comments from the Prague training school
- MWE examples gathered in the WG1-internal survey

Results of the WG sessions

- Poster presentations on potential chapters of the volume "MWEs. Insights from a multi-lingual perspective"
 - > Send your comments to the chapter authors!!!
 - > <http://typo.uni-konstanz.de/parseme/index.php/2-general/117-book-project-multiword-expressions-insights-from-a-multi-lingual-perspective>
- Template:
 - > new dimensions: rhetorical figures; lexical semantic relations
 - > "fixedness": instead of fixed/semi-fixed/flexible, have language-specific diagnostic constructions and possibly notes on how they relate to Baldwin/Kim
- Lexical resources:
 - > Ongoing activities documented at: https://docs.google.com/document/d/1oOjdNYjbbddUsrLJpb7FPZpV9z-wl9sLjwnqpUG_guk/edit?pli=1

Year 2 (till March 2015)

Linguistic properties of MWEs:

- Complete MWE templates for considerable amount of languages. ✓
- Concept and first versions of contributions to an edited volume related to the MWE template. ✓

Lexical formalisms:

- Information sheets on existing lexical resources for MWEs ✓
- Concept and first versions of contributions to an edited volume related to standards and practices in lexical representations of MWEs ~
- Shared task for WG1: Example analysis to challenging cases of MWEs. ~

Interaction with other WGs:

- WG2: MWE templates can help to design test sequences for parsers. 2
- WG3: MWE templates provide useful data for multilingual considerations. 0
- WG4: Cooperation on classification and encoding standards 0

Year 3 activities

- Volume, SOA on lexical encoding
- Hands-on workshop on lexical encoding (Iași)
- WG sessions at the 5th General Meeting (Iași)
- Joint workshop with *European Network of e-Lexicography* (ENeL), perhaps on *Using Resources for Automatic Generation of Lexical Encoding?* (+WG3, WG4) (collocated with Parseme Meeting (Skopje) or ENeL meeting)
- 6th General Meeting (Struga)

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

Plan for Iași

WG1 contributions for poster session

- topics suggested in the WG1-internal survey
- explicit interface questions, i.e. mapping from lexical description to grammars

WG1 sessions

- 1 session on the results of the hands-on workshop
 - > Presentation of the results
 - > Discussion of additional examples (other languages, other formalism, other MWE types...)
- 1 session on the inapplicability of English classification to other languages
 - > 2-3 10' input presentations
 - > discussion