Minutes from the eleventh WG 13 meeting (online 2025-03-12 12:00 CET)

Agenda

Subtask reports:

- Task 3.2: Shared task on morphosyntactic parsing

(Omer Goldman, Leonie Weissweiler, Reut Tsarfaty) Google group Training data and future evaluation code UniDive webpage

- Task 3.4: Evaluation campaign PARSEME 2.0

(Manon Scholivet, Agata Savary)

- Taks 3.5: Evaluation campaign AdMIRe

(Thomas Pickard, Aline Villavicencio) General discussion

Next meeting: May 6, 12.00 CEST (on line)

List of Participants

- Gülşen Eryiğit (chair)
- Joakim Nivre (co-chair)
- Roberto Antonio Díaz Hernández
- Ali Basirat
- Csilla Horváth
- Manon Scholivet
- Rob van der Goot
- Agata Savary
- Ranka Stanković
- Thomas Pickard
- Aline Villavicencio
- Tanja Samardzic
- Dawit J
- Alina Wróblewska
- Luka Terčon
- Olha Kanishcheva
- Dan Zeman
- Takuya Nakamura
- Federica Gamba
- Carlos Ramisch
- Flavio Massimiliano Cecchini
- Gosse Bouma

- Rusudan Makhachashvili
- Voula Giouli
- Ebru Çavuşoğlu
- Omer Goldman
- Reut Tsarfaty
- Chaya Liebeskind
- Faruk Mardan
- Adriana Pagano
- Ilan Kernerman
- Kutay acar
- Ludmila Malahov
- Teresa Lynn
- Lucía Amorós-Poveda

PARSEME shared task (Manon, Agata)

subtask 1 (PARSEME 2.0) quite established framework novelty: non-verbal MWEs, diversity measures subtask 2 (MWE generation) given a context with eliminated MWEs, restore this MWE Problems: how to evaluate the system [ALINE] Consider taking into account the level of difficulty of the items? For example, some items will be more ambiguous and more difficult to determine [JOAKIM] It is unclear which capacity of models we test [TOM] Very difficult to evaluate, even manually. subtask 3 (MWE comprehension/disambiguation) Given a sentence and a span of a potential idiomatic expressions, classify it as idiomatic, literal or coincidental [GULSEN] There are some datasets for this task. Maybe the 3rd category complicates the things. [JOAKIM] [TOM] The same as SemEval 2022 (EN, PT, Galician). There are artefact issues (the models don't really pay attention to the context). subtask 4 (paraphrasing) Given a sentence, rephrase it so that there are no MWEs [AGATA] The input should be raw text, without a span. Objective: simplification of a text. [JOAKIM] The most natural tasks among (2, 3 and 4). Close to what people do with LLMs. Can we avoid doing manual evaluation? (LLM as judge) [TOM] His favorite [ALINE] They work with questionnaires for humans for this problem. There is a synonym dataset. Another task: collect sentences with synonyms of MWEs. [ALINE] Sometimes the simplest way to express a meaning is with a MWE. Questions: Which subtasks to choose? How to evaluate them?

AdMIRe extension

- Tom's wg3_meeting_2025-03-12_editslides
- wg3_meeting_2025-03-12_editTask website
- guidelines & notesData curation

From:
https://unidive.lisn.upsaclay.fr/ - **Universality, diversity and idiosyncrasy in language technology**
CA21167 COST Action

Permanent link:
https://unidive.lisn.upsaclay.fr/doku.php?id=wg3:wg3_meeting_2025-03-12_edit&rev=1741775954

Last update: **2025/03/12 11:39**