SIGTYP 2022

The 4th Workshop on Research in Computational Typology and Multilingual NLP

Seattle/Online, 14 July 2022

S I G T Y P

Overview

6

Papers

4

Extended abstracts

5

Shared task papers

Workshop Modules

Typology

Multilingual Representations

New Corpora and Databases

Shared Task

Keynote Talks

Linguistic Trivia

Schedule

https://sigtyp.github.io/ws2022-sigtyp-schedule.html

Invited Speakers

Kristen Howell

Graham Neubig

Isabel Papadimitriou

Shared Task: Prediction of Cognate Reflexes

Chair: **Johann-Mattis List**

The task: Since the reflex for the BELLY cognate set is missing in English, we could try and predict it from known correspondences to German and Dutch.

Cognate Set	German	English	Dutch
ASH	a ʃ ə	æ ʃ	ɑ s
BITE	b ai s ə n	b ai t	b ɛi t ə
BELLY	b au x	-	b œi k

Families: Tshanglic, Bai, Sino-Tibetian, Sui, Uto-Aztecan, Afro-Asiatic, Dogon, Japonic, IE, Atlantic-Congo, Chapacuran, W.Kho-Bwa, Berta, Palaung

Organizers

Ekaterina Vylomova, University of Melbourne

Hila Gonen, University of Washington

Jonas Pfeiffer, New York University

Edoardo Ponti, The University of Edinburgh

Alexey Sorokin, Moscow State University

Harald Hammarström, Uppsala University

Ryan Cotterell, ETH Zurich

Andrey Shcherbakov, The University of Melbourne

Gabriella Lapesa, University of Stuttgart

Sabrina Mielke, Johns Hopkins University

Pranav A, Dayta AI

Ritesh Kumar, Bhim Rao Ambedkar University

Program Committee

Johannes Bjerva, Aalborg University
Emily Ahn, University of Washington
Miriam Butt, University of Konstanz
John Mansfield, The University of Melbourne
Daan van Esch, Google AI
Elisabetta Ježek, University of Pavia
Paola Merlo, University of Geneva
Joakim Nivre, Uppsala University
Robert Östling, Stockholm University
Ivan Vulić, the University of Cambridge
Richard Sproat,   Google Japan
Željko Agić, Corti
Agnieszka   Falenska, University of Stuttgart
Michael Hahn, Stanford University

Edoardo Maria Ponti, MILA
Alexey Sorokin, Moscow State University
Tanja Samardžić, University of Zurich
Kemal Kurniawan, The University of Melbourne
Aryaman Arora, Georgetown University
Samopriya Basu, The University of North Carolina at Chapel Hill
Badr M. Abdullah, Saarland University
Guglielmo Inglese, KU Leuven
Olga Zamaraeva, University of Washington
Nianwen Xue, Brandeis University
Borja Herce, The University of Zurich
Chinmay Choudhary, National University of Ireland, Galway
Bradley Hauer, The University of Alberta

SIGTYP 2022: Best Paper Award

Please vote for the Best Paper here:
https://forms.gle/A3jvcJY12ZsH8uVw5
(till July, 14th 4PM Seattle Time)!

SIGTYP 2022: Principles

The Virtual part is Free for everyone: no fees for the workshop and free access to all platforms (Zoom, RocketChat)

Accessible from everywhere in the world before and after the workshop.
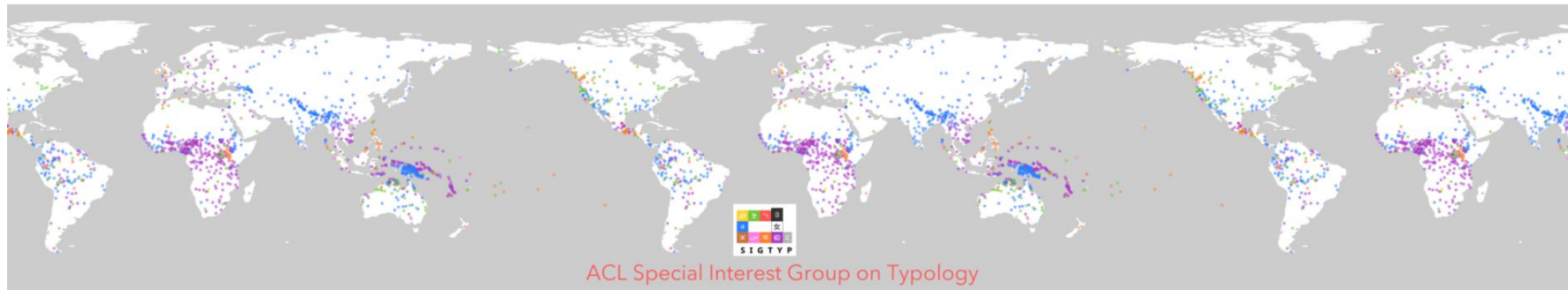
SIGTYP Website: sigtyp.github.io

ACL Special Interest Group on Typology

SIGTYP

THIS IS THE WEBSITE OF THE ACL SPECIAL INTEREST GROUP ON TYPOLOGY (SIGTYP)

The purpose of SIGTYP is to promote the study of linguistic typology in the ACL community and to provide members of the ACL with a means of exchanging news of recent research developments in linguistic typology and multilingual NLP. SIGTYP also organizes an annual workshop collocated with a *ACL conference.

SIGTYP 2022 Workshop Page

sigtyp.github.io/ws2022-sigtyp-schedule.html

SIGTYP2022 — JULY,14th — 508-Tahuya/Hybrid

We kindly invite **everyone** to join the virtual part of SIGTYP 2022! Below you may explore papers, slides, and recorded talks. All discussions are happening in our Rocket.Chat. Each paper is provided with its own discussion channel. On our Rocket.Chat and Google Group you may also find a single Zoom link that will be used during the day of the workshop.
SIGTYP2022 Proceedings are now available here and on the ACL Anthology website.
Please vote for **the Best Paper** here: https://forms.gle/A3jvcJY12ZsH8uVw5 (till July, 14th 4PM Seattle Time)!

Time zone: America/Seattle

8:30 – 8:40 Opening Session

✳ **Keynote Talk** ✳

8:40 – 9:30 Kristen Howell: Grammar Inference for Local Languages. Leveraging Typology for Automatic Grammar Generation
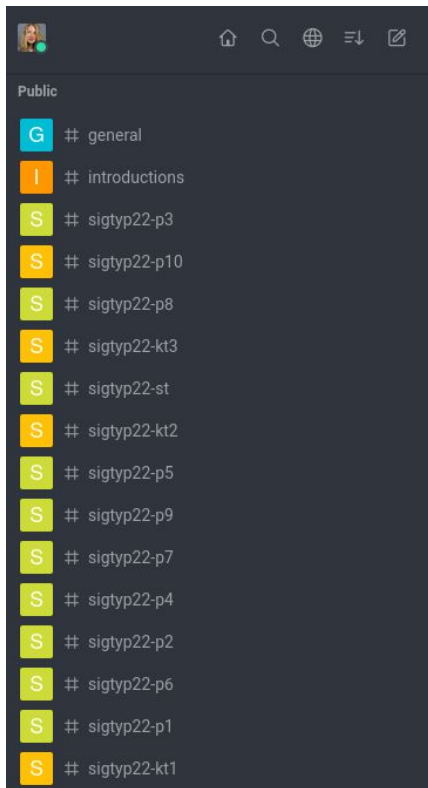
✳ **Multilingual Representations (Long Talks)** ✳

09:30 – 09:45 Multilingualism Encourages Recursion: a Transfer Study with mBERT

09:45 – 10:00 Cross-linguistic Comparison of Linguistic Feature Encoding in BERT Models for Typologically Different Languages

10:00 – 10:10 Coffee Break

SIGTYP 2022 Rocketchat: sigtyp.inf.ethz.ch

Public

- G # general
- I # introductions
- S # sigtyp22-p3
- S # sigtyp22-p10
- S # sigtyp22-p8
- S # sigtyp22-kt3
- S # sigtyp22-st
- S # sigtyp22-kt2
- S # sigtyp22-p5
- S # sigtyp22-p9
- S # sigtyp22-p7
- S # sigtyp22-p4
- S # sigtyp22-p2
- S # sigtyp22-p6
- S # sigtyp22-p1
- S # sigtyp22-kt1

Home

Welcome to SIGTYP 2022!

View the workshop schedule for module and talk times, and links to each paper, video, and paper RocketChat channels:
• SIGTYP 2022 schedule

The same Zoom link will be used for all sessions:
• Zoom link

There are Rocket.Chat channels for each keynote, the shared task, and for each paper, linked here:

Keynotes:
• Kristen Howell
• Graham Neubig
• Isabel Papadimitriou

Shared Task:
• The SIGTYP 2022 Shared Task on the Prediction of Cognate Reflexes

The Rocket.Chat desktops apps for Windows, macOS and Linux are available to download here.

The native mobile app, Rocket.Chat, for Android and iOS is available from Google Play and the App Store.

For further help, please consult the documentation.
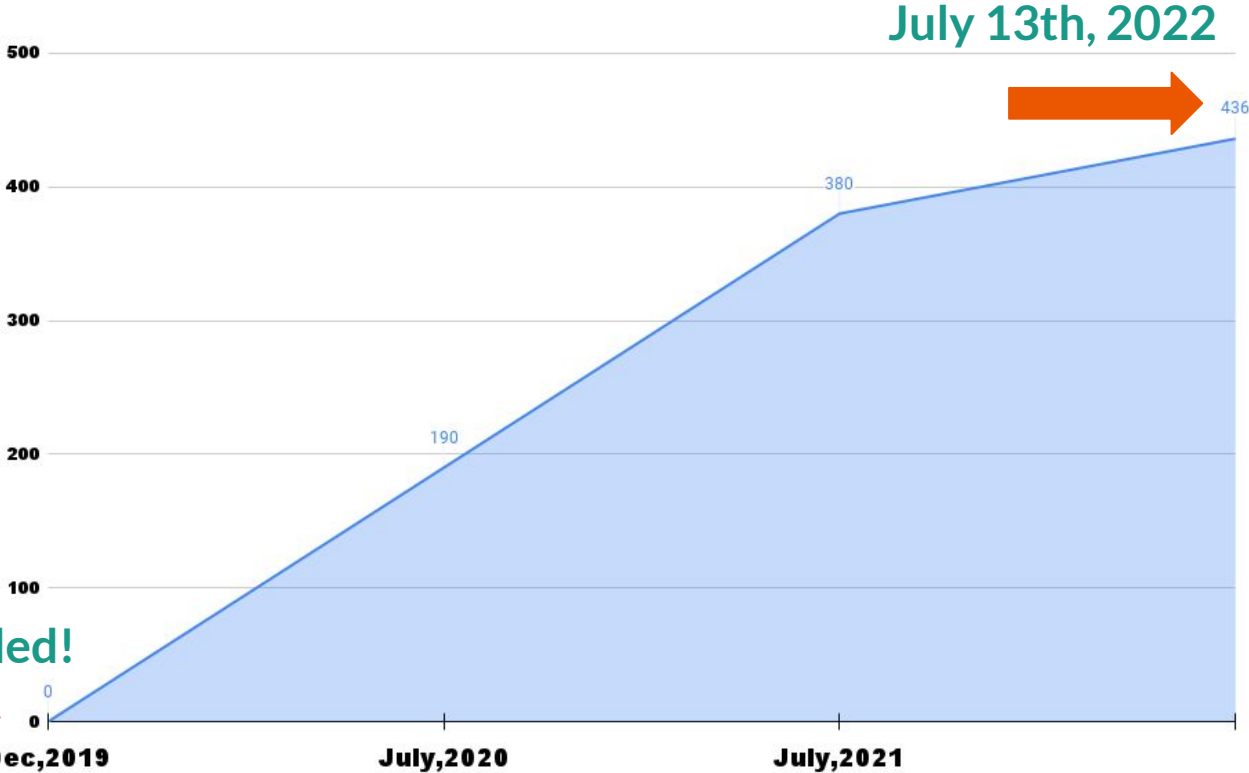
SIGTYP 2022 Hosts in Seattle

Hila Gonen

Jonas Pfeiffer

SIGTYP
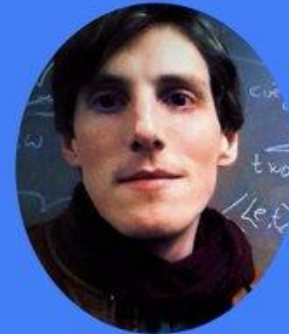
Our Members

SIGTYP Online Lecture Series: 30 lectures

https://sigtyp.github.io/lectures.html

SIGTYP Online Lecture Series: 30 lectures

[14/05/2021] Olga Zamaraeva: Typologically-driven Modeling of wh-Questions in a Grammar Engineering Framework

[04/06/2021] Jon Rawski: Typology Emerges from Computability

[11/06/2021] Tiago Pimentel: An Informative Exploration of the Lexicon

[18/06/2021] Maria Ryskina: Informal Romanization Across Languages and Scripts

[25/06/2021] Shruti Rijhwani: Cross-Lingual Entity Linking for Low-Resource Languages

[02/07/2021] David Inman: Conceptual Interdependence in Language Description, Typology, and NLP: Examples from Nuuchahnulth

[09/07/2021] Tuhin Chakrabarty: NeuroSymbolic methods for creative text generation

[16/07/2021] Sabrina Mielke: Fair Comparisons for Generative Language Models -- with a bit of Information Theory

[23/07/2021] Richard Futrell: Investigating Information-Theoretic Influences on the Order of Elements in Natural Language

[30/07/2021] Eleanor Chodroff: Structure in Cross-linguistic Phonetic Realization

[05/08/2021] Amit Moryossef: Including Signed Languages in NLP

[13/08/2021] Duygu Ataman: Machine Translation of Morphologically-Rich Languages: a Survey and Open Challenges

[20/08/2021] Ekaterina Vylomova: UniMorph and Morphological Inflection Task: Past, Present, and Future

[27/08/2021] Adina Williams: How Strongly does Grammatical Gender Correlate with the Lexical Semantics of Nouns?

[03/09/2021] Kyle Gorman: On "Massively Multilingual" Natural Language Processing

[10/09/2021] Kyle Mahowald: "Deep" Subjecthood: Classifying Grammatical Subjects and Objects across Languages

[17/09/2021] Kayo Yin: Understanding, Improving and Evaluating Context Usage in Context-aware Machine Translation

[24/09/2021] Tanja Samardžić: Language Sampling

[01/10/2021] Antonia Karamolegkou and Sara Stymne: Language Choice for Cross-Lingual Dependency Parsing

[08/10/2021] Mathias Müller: Exploring a Sampling-based Alternative to Beam Search

[15/10/2021] Antonis Anastasopoulos: Extracting Linguistic Information from Text

[21/10/2021] Shauli Ravfogel: Linear Information Removal Methods

[29/10/2021] Claire Bowern: Linguistics and Voynichese

[05/11/2021] Sabine Weber: Using the Distributional Inclusion Hypothesis for Unsupervised Entailment Detection

[12/11/2021] Khuyagbaatar Batsuren: Understanding and Exploiting Language Diversity: Building Linguistic Resources from Cognate to Morphology

[19/11/2021] Aniello De Santo: Mathematical Linguistics and Typological Complexity

[26/11/2021] Aleksandrs Berdicevskis: "Typology, will you marry sociolinguistics?" asks NLP

[03/12/2021] Hagen Blix -- Postponed

[10/12/2021] Badr M. Abdullah: Capturing Cross-linguistic Similarity with Speech Representation Learning

[17/12/2021] Elizabeth Salesky: Looking beyond Unicode for Open-Vocabulary Text Representations

Coming back in 2022!

SIGTYP YouTube Channel

Next Edition: 2023

Open Review .net

2019

2020

2021

ACL 2019 (Florence, Italy)
First edition

EMNLP 2020 (Online)
Second edition

NAACL 2021 (Online)
Third edition (current)

(Hybrid)
Fourth edition

2022: Multilingual Representation Learning (MRL)

https://sigtyp.github.io/ws2022-mrl.html

8 December, Abu Dhabi (UAE)

Organizers

Duygu Ataman, NYU
Orhan Firat, Google
Hila Gonen, UW and Meta AI
Jamshidbek Mirzakhalov, Salesforce
Kelechi Ogueji, University of Waterloo
Sebastian Ruder, Google
Gözde Gül Şahin, Koç University

2022: MRL Shared Task on Multilingual Clause-level Morphology

https://sigtyp.github.io/st2022-mrl.html

8 December, Abu Dhabi (UAE)

Organizers

Omer Goldman, Bar Ilan University
Reut Tsarfaty, Bar Ilan University
Djame Seddah, INRIA Paris
Benjamin Muller, INRIA Paris and
Sorbonne University
Benoît Sagot, INRIA Paris
Hila Gonen, University of Washington and
Meta AI
Jamshidbek Mirzakhalov, Salesforce
Kelechi Ogueji, University of Waterloo
Francesco Tinner, University of Zurich
Duygu Ataman, New York University

Task 1: Inflection

In this task the input is verbal lemma (the form given as a lexicon entry) and a specific set of inflectional features. The task requires generating the desired output clause manifesting the features.

Examples

Languages	Input	Output
English	give IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)	I will give him to her
German	geben IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)	Ich werde ihn ihr geben
Turkish	vermek IND;FUT;NOM(1,SG);ACC(3,SG);DAT(3,SG)	Onu ona vereceğim
Hebrew	נתן IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM)	אתן אותו לה

Task 2: Reinflection

In this task the input is an *inflected clause*, accompanied by its *features*, and *a new set of features* representing the desired form. The task is to generate the desired output that will represent the desired features.

Examples

Languages	Input	Output
English	I will give him to her IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) IND;PRS;NOM(1,PL);ACC(2);DAT(3,PL);NEG	We don't give you to them
German	Ich werde ihn ihr geben IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) IND;PRS;NOM(1,PL);ACC(2,SG);DAT(3,PL);NEG	Wir geben dich ihnen nicht
Turkish	Onu ona vereceğim IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) IND;PRS;PROG;NOM(1,PL);ACC(2,SG);DAT(3,PL);NEG	Seni onlara vermiyoruz
Hebrew	אתן אותו לה IND;FUT;NOM(1,SG);ACC(3,SG,MASC);DAT(3,SG,FEM) IND;PRS;NOM(1,PL,MASC);ACC(2,SG,MASC);DAT(3,PL,FEM);NEG	אנחנו לא נותנים אותך להן

2022: Field Matters (at COLING)

https://field-matters.github.io

Mid October, Gyeongju, Republic of Korea

Organizers

Oleg Serikov, HSE University,
Elena Klyachko, Institute of Linguistics RAS, HSE University
Francis Tyers, Indiana University
Valentin Malykh, Huawei
Timofey Arkhangelskiy, University of Hamburg
Tatiana Shavrina, AIRI, SberDevices
Ekaterina Vylomova, University of Melbourne
Anna Postnikova, HSE University
Ekaterina Neminova, HSE University

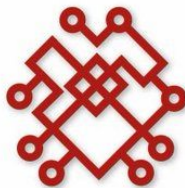# 2022: Field Matters (at COLING)

Field linguists provide structural data about the rapidly diminishing diversity of languages

Field Matters workshop
- brings the needs of field linguists and the community of NLP practitioners together
- develops up-to-date NLP tools for fieldwork

Shared Task on processing the fieldwork recordings: <u>diarization</u> of native <u>speakers</u> and <u>transcription of speech</u> on 10 low-res languages

Workshop papers deadline: **Aug, 1**
Shared task deadline: **Aug, 10**

See the **website** for details on speakers, topics, Shared Tasks!
<u>field-matters.github.io</u>

Follow us on **twitter** @field_matters
- updates on the Shared Task (evaluation phase begins: July, 15)

- workshop news and notifications

- field linguistic NLP content

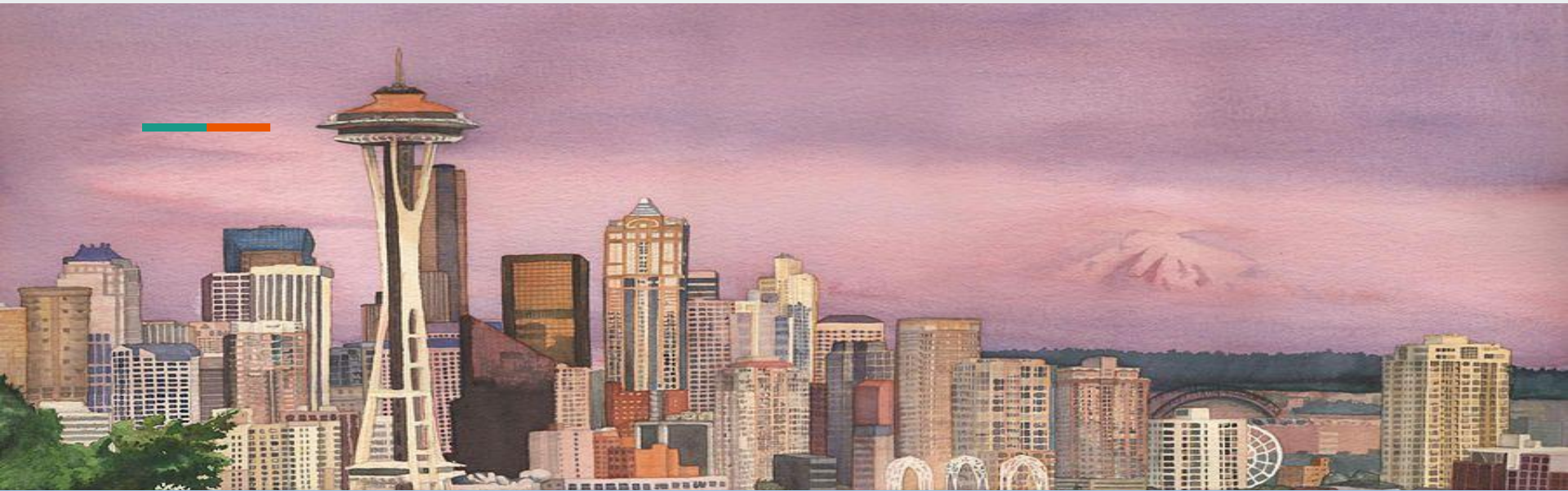# Network

Sponsor

Thank you,
enjoy the workshop!