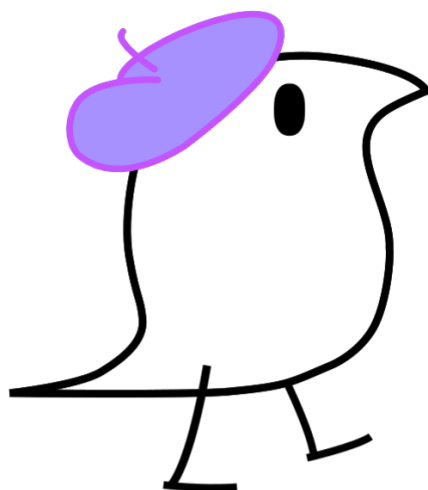


DokFest 2026

Book of Abstracts

Department of Linguistics
Stockholm University



Program

Welcome and lightning talks (C307): 14:00 - 14:45

Poster session 1 (SUBIC): 14:45–15:30

Eirini Akavalou	Multilingual study mentoring in Greek in primary school: A linguistic ethnographic study
Alice Bondarenko	Coexpression patterns of sociative, reciprocal and reflexive markers in African languages
Micaella Bruton	Adapting language models for Named Entity Recognition in historical sources
Patrik Hrabánek	Creaky voice across languages: A focus on Danish stød
Anna Klezovich	Open source Swedish Sign Language motion capture dataset — STS Mocap v1
Anastasia Panova	<i>Gila</i> – a universal word in Gawarbat?
Isabel Temmen	Investigating the shape bias in deaf children acquiring DGS (German Sign Language)
Crina Tudor	Adapting language models for Named Entity Recognition in historical sources
Qiang Xia	Phonaesthetics of constructed languages in a multilingual perspective

Coffee break: 15:30–15:45

Poster session 2 (SUBIC): 15:45–16:30

Stina Andersson	Parental repetitions as contingent input in early parent-child interaction: Timing, structural types, and associations with language development
Caroline Arvidsson	Turn initiations are subserved by higher-order processes outside of the perisylvian language network
Hillborg	Learning written German vocabulary in the signed mode: Challenges and strategies in the multimodal-multilingual classroom
Maike Beyer	
Helena Engström	Investigating posture effects on speech and voice production
Christoffer Forbes	The neurodevelopment of pragmatic processing in adolescence and beyond
Schieche	
Mariia Gvozdkova	Force constructions: syntactic variation across Slavic languages
Amanda Kann	More multilingual NLP with interlinear gloss corpora
Davide Lago	Spatiality in Yadgha prepositions
Korbinian Randl	Towards trustworthy Large Language Models

Poster session 1

Multilingual study mentoring in Greek in primary school: A linguistic ethnographic study

Eirini Akavalou

Department of Teaching and Learning, Stockholm University

The present research is based on an ethnographic study of multilingual language learning trajectories among Greek and Albanian-Greek students in Swedish schools, in which the role of Multilingual Study Mentoring (MSM)¹ will be considered. MSM is the pedagogical and linguistic support that is offered to newly arrived school students in Sweden. The inclusion of speakers from both groups enables an in-depth study of the sociology of language and language learning, also from a critical perspective, and how the speakers themselves construe the meaning of various language ecologies on a time-space scale. The topic focuses, on Multilingual Study Mentoring in Science and Social Science classrooms, including perspectives from students, Multilingual Study Mentors and classroom teachers, with respect to perceived participation, language and subject learning, and well-being. It is a longitudinal study based on a linguistic ethnographic approach (Blommaert & Dong, 2010; Copland & Creese, 2015; Hammersley & Atkinson, 2019) and its fieldwork will span three semesters. The data consists of language portraits (Coffey, 2015), classroom observations, as well as interviews with all three actors that participate in MSM. The topic draws ideas from the interdisciplinary framework of linguistic ecology (Kramsch & Vork Steffensen, 2008; Kramsch & Whiteside, 2008), pedagogical scaffolding (de Oliveira & Westerlund, 2022; Polias et al., 2017; Vygotskij et al., 1987) as the provision of educational support to newly arrived students is interconnected with, while aspects of well-being rely on the affective turn and emotional aspects of teaching (Zembylas, 2004, 2021). The research considers various aspects of students' migration background and multilingualism. Some preliminary results, that will be presented in form of case studies, show that translanguaging (Cenoz & Gorter, 2021; Fuster & Bardel, 2024) and transknowledging are dynamically present during MSM, and that the role of the multilingual study mentor is appreciated by both students and classroom teachers that consider it as a resource. The social and emotional realities of the newly arrived students is highlighted with focus on how the public system's support welcomes them upon their arrival to the new school. Lastly, the newly arrived students peer-support, interact and socialize in various languages on a daily basis that they consider also them as part of their own repertoire

¹Studiehandledning is the term used in Swedish.

Coexpression patterns of sociative, reciprocal and reflexive markers in African languages

Alice Bondarenko

Department of Linguistics, Stockholm University

Reciprocal ('each other') markers can also coexpress other situation types, such as reflexive ('oneself'), as in (1), and sociative ('together'), as in (2).

- (1) Yoruba [yor] (Atlantic-Congo; Awoyale 1986)

*Won rí **ara** won*

they saw **body** their

'They saw **themselves**' OR 'They saw **each other**' (OR 'They saw their bodies')

- (2) Tagdal [tda] (Songay; Benitez-Torres 2021:88, 92)

- a. ***nám**-m_əgra*

REC-understand

'understand **each other**'

- b. ***nəm**-áṭṭeema*

REC-sit

'sit **together**'

Out of these two coexpression types, reciprocal-reflexive coexpression is well understood in a cross-linguistic perspective (see e.g. Heine & Miyashita 2008, Nedjalkov 2007). Comparatively less is known about reciprocal-sociative coexpression. It is known to be attested in various, genealogically diverse, language families but the cross-linguistic frequency is not known. Further, it has been suggested that the two coexpression types are mutually exclusive and that it is unlikely to find three-way coexpression (combining reflexive, reciprocal and sociative in one marker) (Nedjalkov 2007, Kemmer 1993).

This study investigates the coexpression patterns of reciprocal, reflexive, and sociative markers across a genealogically diverse sample of African languages. The sample includes 91 languages from 22 language families, chosen through geographically clustered Diversity Value sampling (Sjöberg 2023) in order to maximize genealogical diversity.

The results suggest that reciprocal-sociative coexpression is almost as widespread as reciprocal-reflexive coexpression in African languages, but that the two types are complementary distributed. The etymology of these markers also differ; while reciprocal markers of reflexive origin frequently originate in nouns meanings 'body' or 'head', this is not the case for reciprocal-sociative markers.

Automated decryption and key extraction of historical ciphers

Micaella Bruton

Department of Linguistics, Stockholm University

Historical ciphertexts present a challenging setting for computational analysis: they are typically unsegmented, noisy, and frequently employ homophonic substitution with variable-length codes. This project investigates automated methods for analyzing and decrypting such texts, as well as extracting their underlying encryption keys, with the long-term goal of developing an end-to-end pipeline for analyzing real-world ciphertext images.

To support this research, I developed *HistCiph*—a large-scale synthetic dataset of historical plaintext–ciphertext pairs spanning 10 languages and texts from the 12th–19th centuries—and the *ChronoFidelius* toolkit—enabling the encryption of plaintext utilizing historical cipher methods. Building on these resources, extensive ciphertext–plaintext alignment experiments were conducted across on sequences of up to 1,050 characters, demonstrating that LSTM models with attention significantly outperform transformer-based architectures in this low-resource, long-sequence setting. These alignment methods enable reliable key extraction even under noisy conditions and in the absence of spacing or punctuation.

Current work focuses on automated decryption utilizing LSTMs with attention. In a controlled setting where ciphertexts share variants of a larger key, models achieve near-perfect decipherment accuracy, even with errors and variable-length codes. In a more challenging setting with fully independent and overlapping keys, preliminary results range from 20–80% F1, depending on factors such as noise and code variability.

Future work will integrate handwritten text recognition (HTR) to process ciphertext images directly, enabling the development of an end-to-end pipeline taking ciphertext images as input and outputs both decrypted plaintext and recovered encryption keys.

Creaky voice across languages: A focus on Danish stød

Patrik Hrabánek

Department of Linguistics, Stockholm University

Numerous languages of the world display various uses of phonation types (e.g., modal, breathy, creaky), ranging from phonetically induced sociolinguistic markers (e.g., English, Hungarian), prosodic and turn-taking functions (e.g., Finnish, Slovene), to phonologically conditioned contrasts (e.g., Jalapa Mazatec, Danish). In the first half of my poster, I document creaky voice – a phonation type pertaining to voiced segments. Creaky voice has been shown to manifest in distinct ways based on the presence or absence of various acoustic features, and accordingly, several subtypes of creaky voice have been identified (Keating et al., 2015; 2023). Creak-based phonation contrasts in a language are typically cued in multiple dimensions, such

as creak duration, creak location within or across segments, and prominence of its acoustic features (e.g., low F0 values, irregularity of F0, spectral tilt, Harmonics-to-Noise Ratio). Furthermore, creaky voice often co-varies with tone/pitch accent where non-modal phonation serves as an additional cue to tone identification in perception (e.g., Tone 3 in Mandarin, *ngã* tone in Vietnamese). An interesting case where creaky voice presents a phonologically contrastive distinction is Danish *stød*, a suprasegmental unit of Danish phonology in which creaky voice has historically been described as its primary feature, although much variation in production of *stød* has been documented. Consequently, there still exists debate concerning which acoustic features listeners primarily attend to in perception. The second half of my poster covers *stød* and presents findings related to its production and perception.

Open source Swedish Sign Language motion capture dataset — STS Mocap v1

Anna Klezovich

Division of Speech, Music and Hearing, Royal Institute of Technology

Sign languages are a under-resourced domain for natural language processing tasks. In comparison for spoken speech generation it is common to use tens of thousands of hours, e.g. VoxPopuli (Wang et al. 2021) is 400k hours. In addition to that, sign languages have quite a lot of video data, e.g. How2Sign, OpenASL, RWTH-PHOENIX-Weather datasets, but very little motion capture data. Essentially this creates a domain gap, because videos contain only 2D data that needs to be lifted, and there is not enough high quality 3D data. I am going to advertise our new 4.1 hours long open source Swedish Sign Language (STS) motion capture dataset (STS Mocap v1) and present some of the sign language generation experiments that we did on this dataset. For more details please refer to our LREC 2026 paper: Klezovich, A., Mesch, J., Henter, G.E., and Beskow, J. (2025). How much Data is Enough Data? A New Motion Capture Corpus for Probabilistic Sign Language Generation.

Gila – a universal word in Gawarbati?

Anastasia Panova

Department of Linguistics, Stockholm University

Gawarbati is an under-described Indo-Aryan language spoken in the Pakistan-Afghanistan border area. In this poster presentation, I will describe one of the most frequent words occurring in the Gawarbati corpus – *gila*.

On the one hand, it is used a plural marker on animate nouns, for example, the phrase *muma gila* [uncle PL] means ‘uncles’. Additionally, it is a marker of associative plural, expressing the meaning ‘X and other people associated with X’. For example, the same phrase *muma gila* [uncle PL] can also be interpreted as ‘the uncle and his family/relatives’ or ‘the uncle and his friends’.

On the other hand, *gila* is used as a placeholder similar to English *whatchamacallit*. For example, in the sentence *Xalaq ta nu bikilia **gila** ketiau, ta nu a uana istima l ketiau* ‘Let the people do the thing in their fields, let them use their water’, *gila* substitutes a noun phrase which comes later in the next clause. Further, it is used similar to English *thing* in such constructions as “the thing is...” and “this/the X thing”.

In the poster, I discuss possible diachronic relations between different uses of *gila* in Gawarabati and compare it to associative plural markers and placeholders in other languages.

Investigating the shape bias in deaf children acquiring DGS (German Sign Language)

Isabel Temmen & Pamela Perniss

University of Cologne

Referential mapping is a critical process in language development, where children learn that words/signs connect to real-world referents. The process is supported by a range of communicative strategies and attentional biases (Perniss et al., 2017). One such bias is the shape bias, which refers to the extension of labels to new objects based on shape similarity (Landau et al., 1988). Novel noun generalization (NNG) tasks have provided experimental evidence for the shape bias. In these tasks, children are familiarized with an unknown target object (e.g. This is a dax) and must identify another instance of the object from other unknown objects (Show me the dax). Findings have shown that children (from 18-24 months) consistently select the object that shares the target object’s shape (Landau et al., 1988). However, the development of the shape bias has been shown to interact with cultural and linguistic factors, e.g. whether a language has a strong count/mass distinction (Kucker et al., 2019; Yoshida & Smith, 2001).

The degree to which languages encode shape in their grammar may also modulate the shape bias. Shape encoding is particularly pervasive in sign languages due to the high potential for iconicity (Taub, 2001). To date, there has been no research on the development or existence of a shape bias in children acquiring a sign language. In this study, we address this research gap by investigating whether deaf children acquiring German Sign Language (DGS) exhibit a shape bias and whether it is influenced by shape-encoding in DGS.

In order to do so, we must first better understand the patterns of shape encoding in the lexicon. To this end, we conducted pre-studies with adult deaf DGS signers (n=20 per study). In pre-study 1, participants were asked to name 50 common and early-acquired object signs, and then, for each sign, to produce as many other signs as possible with the same handshape as well as any classifier handshapes that can be used for the object. In pre-study 2, the object signs were rated for iconicity and for shape similarity to their referents. Finally, in pre-study 3, participants provided shape similarity ratings for objects whose signs share a handshape. The data from the pre-studies will allow us to construct sets of novel signs for use in the NNG task, based on correspondences between handshape and object shape in the DGS

lexicon. One set of novel signs will encode information about object shape in a way typical for DGS; the other set of novel signs will encode no information about shape. Our aim is to provide a more nuanced understanding of shape as a factor that can facilitate vocabulary growth (Smith et al., 2002) by investigating the shape bias in relation to (visually-iconic) linguistic encoding of shape.

Adapting language models for Named Entity Recognition in historical sources

Crina Tudor

Department of Linguistics, Stockholm University

Named entity recognition (NER)—the task of automatically identifying and labeling entities such as names of people, places, organizations, and other key entities in text—is particularly difficult to apply to historical sources. Such texts often contain inconsistent spelling, errors introduced during digitization (OCR noise), and very limited annotated data for training. Recent work has explored the use of large generative language models in a zero-shot setting (i.e. without task-specific training), but these approaches have generally produced unsatisfactory results for historical NER, especially when dealing with noisy and multilingual archival material.

An alternative strategy is domain-adaptive pre-training (DAPT), where models are further trained on domain-specific data before being fine-tuned for a specific task. While DAPT has shown clear benefits in some low-resource contexts, its effectiveness under realistic archival conditions—where data varies widely in language, time period, and quality—has not been thoroughly examined.

In this study, we investigate how well DAPT supports NER in historical texts by comparing five BERT-based models that differ in size, multilingual capacity, and prior domain exposure. Importantly, we use raw historical newspaper data without extensive filtering, in order to better reflect the kinds of material researchers encounter in archives. We evaluate performance across five languages included in the HIPE-2022 shared task, using standard F1 scores to measure accuracy.

We also explore how well these models transfer to new languages by testing them on a historical Romanian NER dataset that was not included during training. This allows us to examine whether familiarity with historical language or broader multilingual coverage plays a more important role in enabling models to generalize across languages.

Phonaesthetics of constructed languages in a multilingual perspective

Qiang Xia

Department of Linguistics, Stockholm University

Constructed languages (conlangs) provide a powerful source for testing the relative impact of universal versus language-specific effects of sound symbolism. Conlangs are typically designed to elicit specific impressions, such as Star Trek’s Klingon (harsh) or Tolkien’s Quenya (pleasing). Although these conlangs were crafted for an English-speaking audience, previous studies have shown that they still elicit the intended impressions among German native speakers when presented in a neutral tone. To test the cross-linguistic stability of these impressions, we conducted a series of online rating experiments. 56 stimuli from 14 conlangs were rated using a 7-point semantic differential scale (pleasant–unpleasant). The stimuli were recorded by two German native speakers in a neutral voice. The experiment was translated into 14 natural languages (e.g., Arabic, English, Georgian, Italian, Cantonese and Mandarin Chinese). The results from 903 participants indicate that certain conlangs employ relatively universal phonaesthetic cues that consistently produce the intended impressions across languages. Others, however, appear more dependent on the listener’s linguistic background. These differences are reflected in how phonetic and phonemic features contribute to ratings.

Poster session 2

Parental repetitions as contingent input in early parent-child interaction: timing, structural types, and associations with language development

Stina Andersson

Department of Linguistics, Stockholm University

To fully or partly repeat a young child's utterances is a well-documented parental habit. These repetitions constitute an aspect of contingent parental input, as they both occur promptly after and are semantically linked to the child's utterance. A high frequency of repetitions during the child's second year has been positively related to language development. However, much remains unexplored regarding changes over time in amount and nature of repetitions, and how repetitions relate to later language skills. Additionally, it remains unclear whether this differs across different types of repetitions. Furthermore, the more precise temporal properties of repetitions remain unknown. The study investigates repetitions of child utterances in video recorded dyadic parent-child interaction (N=20) at five time points: 1;0, 1;6, 2;0, 2;6 and 3;0. Four structural types were examined: exact, exact-plus, partial and partial-plus repetitions. Longitudinal associations between repetitions and child language development were examined using vocabulary and morphosyntactic measures. The results showed that repetitions across all investigated ages predominantly occurred in the first utterance following the child's source utterance, and started within 1000 ms of its end. Exploratory results suggested that positive associations between repetitions and later vocabulary extend to ages 2;6 and 3;0 as well. Mixed-effects models indicated that structural types follow different developmental trajectories, both in terms of frequency at different ages and in the shape of their trajectories over time. This suggests that parents adapt the use of repetitions to the child's developmental level. Furthermore, these findings highlight that treating repetitions as a single composite measure may obscure important differences between types, resulting in a less nuanced representation of parental input.

Turn initiations are subserved by higher-order processes outside of the perisylvian language network

Caroline Arvidsson¹, Ekaterina Torubarova², André Pereira², and Julia Uddén^{1,3}

¹Department of Linguistics, Stockholm University, ²Division of Speech, Music and Hearing, Royal Institute of Technology, ³Department of Psychology, Stockholm University

Conversation is arguably one of the most ubiquitous contexts for producing language. It is characterized by turn-taking, in which speakers rapidly alternate between speaker and listener roles, while simultaneously planning their upcoming turn. Neurobiological models of speech planning have implicated regions within the left-hemispheric dominant classical perisylvian language network (e.g., the inferior frontal gyrus, the superior temporal lobes, and the parietal cortex), as well as lower-level sensory and motor areas. These models, however, are grounded in the production of single linguistic units (e.g., words) in non-interactive contexts. Consequently, the neural substrates subserving the initiation of conversational turns are largely unknown.

In this fMRI investigation, we asked whether turn initiations would activate areas outside the classical perisylvian core language network. We hypothesized that the initiation of conversational turns would engage cortical regions associated with higher-order processing outside of the perisylvian language network. We analyzed two different conversational datasets, in which participants engaged in unscripted conversations via an audio-video link with a confederate outside the scanner: (1) the Neuroengage dataset (Torubarova et al., 2025, 28 participants, comprising 14 hours of conversation); and (2) the Multimodal Corpus of Bidirectional Conversation during fMRI Scanning (Rauchbauer et al., 2020, 24 participants, 4.8 hours of conversation).

Turn initiations were defined as a 600 ms time window (based on the reported minimum latency of speech preparation; Indefrey and Levelt (2004)) preceding the participant’s turn. For each dataset, a general linear model was used with five conversational events as regressors: turn initiation, production, comprehension, silence, and fixation (looking at a fixation cross between conversations).

In both datasets, the contrast “turn initiation vs. fixation” revealed engagement of classical language and sensorimotor areas. This contrast also consistently, across datasets, generated clusters of activation in the midline structures of the frontal cortex, spanning from the medial superior frontal gyrus (SFG) to the medial prefrontal cortex (mPFC; e.g., the anterior cingulate cortex).

Based on these findings, we suggest that turn initiations are subserved by higher-order extra-linguistic processes localized in the SFG and the mPFC. These processes likely play a critical role in conversational language use, where speakers must continuously meet the temporal and pragmatic constraints governing timely and contextually appropriate conversational turn-taking. Finally, we argue that adequate models of speech planning and large-scale neurobiological accounts of language must incorporate the interactive context in which language processing naturally develops and operates, namely, conversation.

Learning written German vocabulary in the signed mode: Challenges and strategies in the multimodal-multilingual classroom

Maike Beyer

Department of Linguistics, Stockholm University

This poster presentation is based on data from the project “*Additional Language Acquisition of Deaf and Hard-of-Hearing Immigrant Students in Germany*” (2021–2024), which focuses on language learning contexts among immigrant deaf and hard-of-hearing multilingual learners (IDML). Using a linguistic ethnographic approach, the study draws on classroom video recordings and semi-structured teacher interviews collected in a German as a Second Language classroom.

Upon arrival in Germany, IDML acquire not only German as a new language but also learn German Sign Language (DGS). As vocabulary exercises are seen as a crucial first step in learning a new language (Beck et al., 2013), this study examines how IDML learn written German vocabulary. In deaf education, fingerspelling plays a unique role in language learning because it combines both the signed and written modes simultaneously (Holmström & Schönström, 2018). Learning fingerspelling, therefore, requires phonological awareness of the signed mode and the visual-orthographic awareness of written language.

Drawing on classroom interactions, the analysis highlights the challenges and strategies involved in these vocabulary-focused activities. The illustrative examples demonstrate the challenges students encounter in learning written German orthographic sequences and in developing phonological awareness of the DGS alphabet. Furthermore, students support each other by drawing on their L1 during vocabulary learning. Overall, the example in this poster presentation will demonstrate how IDML use multilingual-multimodal resources to make meaning and support the learning of written German vocabulary.

Investigating posture effects on speech and voice production

Helena Engström

Department of Linguistics, Stockholm University

Human speech occurs in a variety of different body positions. Conversation seemingly seamlessly flows while standing up from a chair, lying down on the beach or while swinging on a swing-set at the playground. However, these shifts in body position do affect human speech physiology because of the constant influence of Earth’s gravitational pull.

Previous research into posture effects on voice characteristics has mainly focused on phonation produced by classically trained singers (Sundberg et al., 1991), leaving many aspects of how speech production is affected by changes in body position in vocally untrained speakers understudied. By investigating compensatory behavior of speech produced in an upright, supine and inverted position, our purpose is to map together the integrated mechanisms of articulation, phonation and speech breathing.

This poster presentation presents the aims and experimental design of the project *Body posture as a window into biological foundations of speech* and pilot results centering around differences in breathing patterns and voice quality compared between positions (Engström et al., 2024), as previously presented at the Pan European Voice Conference 2026.

The neurodevelopment of pragmatic processing in adolescence and beyond

Christoffer Forbes Schieche

Department of Linguistics, Stockholm University

During adolescence, pragmatic development is still ongoing, yet this developmental trajectory is understudied, and its neural correlates are currently unknown. Previous fMRI studies in adults have shown that processing indirect speech acts engages the core language and Theory of Mind (ToM) networks to a greater extent than direct speech acts. However, individual differences in pragmatic ability have been linked to activity in regions outside these networks.

In this study, we used an established fMRI paradigm contrasting indirect and direct speech acts to investigate the development of pragmatic processing in adolescents ($n = 51$) and adults ($n = 57$, Bendtz et al., 2022). Adolescents were divided into younger (13–15) and older (16–18) groups, and adults into low- and high-scoring groups, based on behavioral pragmatic measures.

We replicated prior adult findings in our adolescent sample, with increased activation in canonical language and ToM regions during indirect speech act processing. We also found age-related increases in activity in the posterior medial superior frontal gyrus (pmSFG), and a combination of age-related differences and adult individual differences in the left posterior intraparietal sulcus (pIPS) and the dorsal posterior cingulate cortex (dPCC). Notably, all three clusters were located outside of the classic perisylvian language network.

The increasing activity in the pmSFG may reflect a shift from more general ToM processes to more conversation-specific pragmatic representations. We discuss the findings in the pIPS and the dPCC in relation to their overlaps with the multiple demand and default mode networks, respectively. Similarities between young adolescents and adults with lower pragmatic ability in these regions suggest that delayed development may be linked to pragmatic difficulties in adulthood. Taken together, our findings indicate that pragmatic development is supported by multiple networks that may contribute to different aspects of pragmatic processing.

Force constructions: Syntactic variation across Slavic languages

Mariia Gvozdkova

Department of Slavic and Baltic Studies, Finnish, Dutch and German, Stockholm University

The semantic role of Force (e.g., 'wind', 'fire', 'lightning') was introduced as a subtype of the Effector role by Van Valin and Wilkins (1996), and was distinguished from prototypical agents and instruments in transitive clauses. Force is defined as an entity that instigates an action but not consciously or voluntarily. In sentences, forces can be marked by a prepositional phrase (e.g., all trains were cancelled due to a severe storm) and appear in syntactic constructions labeled as force constructions (e.g., Kittilä, 2005). These include active and passive sentences. One notable exception is a specific impersonal construction common to East Slavic languages and, to a lesser extent, Polish (Junghanns et al., 2017; Schlund, 2020). It consists of a verb in impersonal form, a force marked with the instrumental case, and an object marked with the accusative. The present thesis aims to describe force constructions by comparing their distribution across several Slavic languages based on web corpora for each language. Using a quantitative method, the frequency of each construction type will be calculated with a predefined set of force lexemes. Moreover, it will be tested whether there are tendencies to mark different forces according to their type: abstract forces (e.g., fear, fate), force phenomena (e.g., wind, wildfire), and episodic forces (e.g., wave, lightning). The preliminary hypothesis suggests that abstract forces tend to occur in active constructions, impersonal constructions are more common with episodic actors, whereas force phenomena show approximately equal frequency across different construction types.

More multilingual NLP with interlinear gloss corpora

Amanda Kann

Department of Linguistics, Stockholm University

Language data with structured, multilingually comparable annotation is a necessity for large-scale computational typology and many multilingual NLP applications. However, even the broadest resources, such as UniMorph and Universal Dependencies, contain data from fewer than 200 of the world's over 7000 languages.

Interlinear glossed text (IGT) examples are common in typological literature and descriptive grammars. They often contain rich morphosyntactic information, but have received little attention in multilingual computational linguistics. Recent efforts to compile IGT examples from various sources have resulted in massively multilingual IGT corpora, such as the PolyGloss corpus (Ginn et al., 2026), covering over 2000 languages. I will present two ongoing projects which make use of this resource for previously underexplored downstream tasks.

First, although glossed examples themselves are of limited use in traditional NLP pipelines, their components (morphological analysis and translations into a

metalanguage) can be leveraged to infer other types of annotation, such as parts of speech or dependency structure (Georgi, 2016). I am investigating modern annotation transfer and LLM-based methods for inferring semantic role and syntactic dependency labels from glosses, with the aim of extending multilingual language model benchmarks such as MultiBLiMP (Jumelet et al., 2026) to more languages without the need for additional annotation.

Large multilingual IGT corpora can also be used to train language models for the task of automatic gloss prediction. While previous rule-based and neural approaches to this task have required large amounts of glossed training data in the target language, multilingual pretraining appears to improve glossing accuracy in more realistic lower-data settings, even for languages which are not well represented in the training dataset (Ginn et al., 2024). I am evaluating the glossing performance of such models on a previously unseen language, Gawarbatī [gwt], using varying amounts of monolingual finetuning data.

Spatiality in Yadgha prepositions

Davide Lago

University of Naples L’Orientale

This presentation provides an account of spatial encoding in Yadgha ([ydg] ISO 639-3), an endangered Eastern Iranian language spoken in the Lutkoh Valley of north-western Pakistan. The analysis draws on the corpus of data collected for my doctoral thesis. The study focuses on the morphosyntactic and semantic properties of the core prepositions *nə*, *də*, *trə*, *pə*, and *skə/sə*. These prepositions encode fine-grained distinctions based on visibility, elevation, and contact: *nə* marks a horizontal motion to a location; *də* marks location in a non-visible or occluded area; *trə* marks location in a visible area; *pə* encodes location in an upper or elevated domain; and *skə/sə* denotes location on a surface with contact with an implicit instrumental/comitative sense. I show that these forms interact systematically with enclitic demonstratives (proximal, medial, distal) and adverbs/postpositions, yielding complex spatial configurations that integrate deixis, topology, and elevation:

batʃa	ʃui	dəi	ʃahr	χavt
batʃa	ʃu-i	də=i	ʃahr	χavt
king	go.PST-3SG	in1=MED.DEM.DIR.SG	city	descend.PST-3SG
<i>n</i>	<i>v</i>	<i>prep=dem</i>	<i>n</i>	<i>v</i>

‘The king descended in the (near, non-visible) city’

Furthermore, Yadgha employs a set of motion verbs that lexicalise directionality and complement the prepositional system: verbs of motion (‘go’, ‘come’), downward motion (‘descend’) contribute to the expression of both static and dynamic spatial relations. The presentation examines how these resources combine to produce a highly articulated spatial domain. By situating Yadgha within the broader typology of spatial deixis and locative marking in the Hindu Kush-Karakoram (Halfmann 2024; Heegård & Liljegren 2018), the presentation highlights both inherited

patterns and language-specific innovations, contributing new empirical evidence to the description of Iranian spatial systems.

Towards trustworthy Large Language Models

Korbinian Randl

Department of Computer and Systems Sciences, Stockholm University

Large Language Models (LLMs) have become foundational to modern digital applications, powering systems such as AI-based chatbots (e.g., ChatGPT), content generation tools, and automated software debugging. At their core, these models are large statistical systems trained on the task of next-token prediction, generating text one token at a time (Brown et al., 2020). Despite its simplicity, this autoregressive objective has proven remarkably effective: LLMs produce syntactically coherent text and are even able to encode semantic information from their training data. This makes them strong candidates for the future of text-based machine learning.

However, their capabilities come with limitations. LLMs remain largely opaque “black boxes,” and are prone to hallucinations, i.e. outputs that sound plausible yet are factually incorrect. This is particularly due to their next-token training objective and when encountering inputs outside the distribution of their training data (Perković et al., 2024; Reddy et al., 2024). Such unreliability poses significant risks in safety-critical domains like healthcare or food regulation, where incorrect conclusions or fabricated references can lead to harmful decisions. Additionally, the substantial computational resources required for training and deploying LLMs raise concerns about their environmental sustainability.

My research examines where current LLMs fall short in terms of trustworthiness and how they can be improved using knowledge-enhanced prompting. Specifically, I demonstrate that LLMs are generally unable to faithfully explain their own reasoning, often producing explanations that sound credible but are not faithful to the underlying model behaviour (Randl et al., 2024b, 2025) (Randl et al., 2024b, 2025). I also investigate strategies for selecting few-shot examples in prompt-based classification, showing that conformal-prompting-based selection can reduce resource consumption without sacrificing predictive performance (Randl et al., 2024a). Finally, I study end-to-end explainability in RAG pipelines, revealing that encoder-only retrievers rely largely on keyword-matching heuristics and that generator models exhibit primacy bias (Randl et al., 2026).
