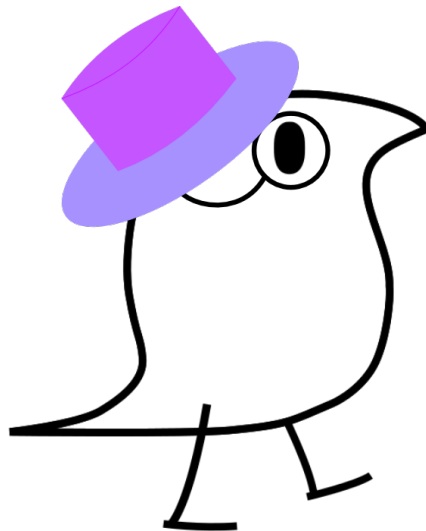


DokFest 2025

Book of Abstracts

Department of Linguistics
Stockholm University



Program

Welcome: 13:00 - 13:05

Session 1 - Computational linguistics: 13:05–14:00

Micaella Bruton	Augmenting the past: Toward an automated pipeline for historical cipher analysis & decryption
Crina Tudor	Prompting the past: Exploring zero-shot learning for Named Entity Recognition in historical texts using prompt-answering LLMs
Amanda Kann	Does translation affect word order proportions?

Break: 14:00–14:15

Session 2 - Typology and sign language: 14:15–15:10

Anastasiia Panova	Possessive suffixes in Gawarbat: A story of marginalization
Maike Beyer	Language teaching and learning contexts with immigrated deaf and hard-of-hearing multilingual learners
Alice Bondarenko	Together in a heap: a look at sociative adverbs in Swedish

Coffee break and poster session: 15:10–15:40

Charlotte Stinkeste	Should chatbots use anthropomorphic expressions like "I think"?
---------------------	---

Session 3 - Experimental methods: 15:40–16:35

Stina Andersson	Touch and gesture during the first year and later language skills
Qiang Xia	How does transmission delay affect turn-taking in remote conversations?
Christoffer Forbes Schieche	Neural correlates of pragmatic processing within and beyond adolescence

Session 1 - Computational linguistics

Augmenting the past: Innovative techniques towards enhancing the availability of historical cryptographic texts and keys

Micaella Bruton

Department of Linguistics, Stockholm University

In the field of historical cryptography, the complexity and uniqueness of ancient ciphers can often pose significant challenges. The fixed amount of historical plaintext/ciphertext data means that the development and performance of text alignment, key detection, cipher-type detection, and decryption systems are severely limited. My approach to resolving this issue is two-fold: firstly, to develop a collection of historical language models capable of generating plaintext, and secondly, to establish a Python library for encrypting plaintext using various historical cipher types. 82 historical language n-gram models are being developed, initially with a sole focus on the 17 languages included in The HistCorp Collection of Historical Corpora and Resources (Pettersson & Megyesi, 2018); the number of models developed for each language varies depending on the amount of data available, with each model spanning no more than 100 years. The intended python library, ChronoFidelius, is currently in development and only supports homophonic digit ciphers, but will be extended to include additional historical cipher types. Features typically encountered in real-world historical ciphertexts, such as nulls or transcribed errors, can also be optionally simulated. Together, these methods enable the controlled, on-the-fly generation of plaintext/ciphertext and key data, which will support the development of various historical cryptographic models. This capability not only advances the field of digital humanities but also opens new avenues for the accurate interpretation of historical documents.

Prompting the past: Exploring zero-shot learning for Named Entity Recognition in historical texts using prompt-answering LLMs

Crina Tudor

Department of Linguistics, Stockholm University

This work investigates the application of prompt-answering Large Language Models (LLMs) for the task of Named Entity Recognition (NER) in historical texts. Historical NER presents unique challenges due to language change through time, spelling variation, limited availability of digitized data (and, in particular, labeled data), and errors introduced by Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) processes. Leveraging the zero-shot capabilities of prompt-answering LLMs, we address these challenges by prompting the model to extract entities such as persons, locations, organizations, and dates from historical documents. We then present an extensive error analysis of the model output in order to identify and address potential weaknesses in the entity recognition process. The results show that, while such models display ability for extracting named entities, their overall performance is lackluster. Our analysis reveals that model performance is significantly affected by hallucinations in the model output, as well as by challenges imposed by the evaluation of NER output.

Does translation affect word order proportions?

Amanda Kann

Department of Linguistics, Stockholm University

Translated texts tend to differ structurally from same-language original texts in various ways, and automatic classifiers trained on syntactic features can identify translations with high accuracy (Volansky et al., 2015). The strength of translation effects in a given text may also depend on the text's original language – "source language interference" – or on the relationship between the source and target languages. These properties suggest that using translated text in quantitative syntactic typology or language modelling may yield unrepresentative results.

In this talk, I will present preliminary findings from a quantitative study of subject/verb word order variation in a multiparallel corpus of European Parliament speeches in 21 European languages. This corpus contains original speeches in all 21 languages and sufficient amounts of translated data for all 420 possible language pairs, enabling cross-linguistic comparisons across the entire language sample without potential confounding variation from text type differences.

Comparing originals in a given language to translations into the same language reveals a cross-linguistically consistent regularizing effect of translation - original texts generally contain a higher proportion of the non-dominant word order type than translations do, regardless of the degree of word order freedom observed in the source language.

Session 2 - Typology and sign language

Possessive suffixes in Gawarbati: A story of marginalization

Anastasiia Panova

Department of Linguistics, Stockholm University

Gawarbati is one of the understudied Indo-Aryan languages spoken in the Hindukush region. In this talk, I outline the use of possessive suffixes in possessive noun phrases in Gawarbati as evidenced by data from the Gawarbati spoken corpus.

In the languages of the Hindukush region, the semantic relationship of possession within noun phrases is typically encoded by a genitive or oblique case marker on the possessor. However, languages of the Western Hindukush tend to show double marking in possessive noun phrases. For example, in Shumashti (Indo-Aryan), in the noun phrase *to-no r-u* (you.sg-gen head-pos.2sg) *your head* (Morgenstierne 1945: 255), the possessive relation is marked twice, by a genitive case marker on the possessor and by a possessive suffix on the possessee. In the literature (Di Carlo 2011; Liljegren 2023), the presence of possessive suffixes has been discussed as one of the features of a linguistic (sub-)area in the Western Hindukush, involving several Indo-Aryan and Nuristani languages.

Upon closer examination, we find that possessive suffixes in the Western Hindukush languages behave rather differently. In some languages (for example, in Shumashti), possessive suffixes have a well-formed paradigm and are used with all types of nouns. In other languages, possessive suffixes show various kinds of restrictions. One of the latter languages is Gawarbati. Possessive suffixes in Gawarbati have gaps and syncretism in their paradigm, they are restricted to kinship terms, and even in kinship terms they are optional. I hypothesize that possessive suffixes in Gawarbati used to play a more prominent role in the grammatical system but were later marginalized.

Language teaching and learning contexts with immigrated deaf and hard-of-hearing multilingual learners

Maike Beyer

Department of Linguistics, Stockholm University

Deaf and hard-of-hearing migrants arrive in Germany with highly diverse backgrounds and language experiences. Upon arrival, they face the challenge of learning (spoken/written) German and German Sign Language (DGS) almost simultaneously. However, teaching these newly arrived migrants, henceforth referred to as immigrated deaf and hard-of-hearing multilingual learners (IDML; Cannon & Marx, 2023), remains an under-researched area.

Teachers often lack curricular and instructional guidelines to support IDML in language acquisition and development. They report a strong need for information on language acquisition and development, didactic strategies, and adapted teaching materials for IDML (Begon, 2023; Marx & Urbann, 2022). Using a linguistic ethnographic approach, including video recordings, field notes, and interviews, I examine how teachers navigate the lack of instructional guidelines by using different strategies while teaching IDML. I document how they respond to the linguistic diversity in language teaching and learning situations.

In this presentation, I will demonstrate the heterogeneity of semiotic repertoires among IDML in schools for deaf and hard-of-hearing students. Drawing on examples from my data, I will show how teachers support language learning processes in these settings. I will also briefly present my ongoing thesis project, focusing on language teaching and learning in these contexts.

Together in a heap: a look at sociative adverbs in Swedish

Alice Bondarenko

Department of Linguistics, Stockholm University

Words meaning together, or sociative markers, frequently have several meanings in addition to joint action, as pointed out by Nedjalkov (2007: 18). Apart from studies of a few larger languages (see e.g. Moltmann 2004 for English *together* and Zalaznjak & Šmelev 1999 for Russian *vmeste*), the semantics and polysemy of sociative adverbs have not attracted much attention.

Swedish has several adverbs that can express joint action but differ in their range of other meanings. In this talk, I present an ongoing corpus study on the adverbs *tillsammans*, *ihop* and *samman*. Using a sample of 500 examples for each adverb from a social media corpus (Borin et al. 2012), I analyze the syntactic and semantic properties of each adverb and argue that the degree of boundedness to the predicate is directly correlated to the semantics of the adverbs.

Session 3 - Experimental methods

Touch and gesture during the first year and later language skills

Stina Andersson

Department of Linguistics, Stockholm University

A child's activity level, curiosity and own initiative to interaction has been identified as important for future development, not only in the motoric area but also in relation to language. However, the relation between gross and fine motor skills and later language remains inconclusive, and positive effects of fine motor skills have mainly been found during the second year. Some studies have shown that parental response to communicative acts from their children depends on the child's motoric behaviour, e.g. if the child is standing or sitting.

This study is inspired by the concept of cascading effects, as proposed in dynamic systems theory, which assumes that development in one area can influence achievements in other areas. It can be argued that active infants, by gaining faster control over motor systems, may receive a developmental advantage in language and communication, as motor systems are essential for symbolic interaction such as gestures and speech. Accordingly, this study investigates whether motoric behaviour during the first year of life in interaction with a parent can predict language competence at four years of age. Additionally, it explores the extent to which interaction context and amount of parental speech influence these potential relationships.

The data consists of 22 parent-child dyads video-recorded in interaction at 6, 9 and 12 months of age. These children were selected from a larger group of 65 children, based on their productive language scores (measured by CDI) at 48 months: 11 children with high scores and 11 with low scores. Child gestures, child touch at parent, parental speaking time and interaction context was annotated and related to CDI scores at 48 months of age.

The results reveal differences in motoric behaviour between the high and low CDI score groups. Child use of deictic gestures and handling of objects as well as amounts of parental vocalizations were positive predictors of later productive language, while child use of comforting touch was a negative predictor. These findings suggest that a child's interactional use of gesture and touch during the first year of life can have long-term implications for language development.

How does transmission delay affect turn-taking in remote conversations?

Qiang Xia

Department of Linguistics, Stockholm University

Although conversational turn-taking has been studied extensively for decades, less is known about how it is managed and adapted in remote conversations, where latency caused by signal transmission is unavoidable and pervasive. Turn-taking is a highly coordinated process in which interlocutors take turns speaking with minimal gaps or overlaps by entraining to each other and anticipating turn endings based on various temporal cues (e.g., speech rate; see Sack et al. 1974; Duncan, 1972; Wilson and Wilson, 2005; Heldner and Edlund, 2010; Levinson and Torreira, 2015; Stivers et al., 2009, etc.). However, when these temporal properties are out of sync or unpredictable due to factors such as transmission delay or packet loss, turn-end anticipation and speaker entrainment become challenging. This raises the question of whether speakers use different turn-taking mechanisms in remote conversations than in undelayed interactions, and how they adapt to transmission delays to maintain conversation dynamics. We hypothesize that longer transmission will result in longer post-speech gaps experienced by speakers from their own perspective, leading to disruptions in conversational rhythm and more frequent overlapping speech. To test this, a pilot study was conducted. Two pairs of speakers engaged in a survival task at three delay levels: 0 ms, 400 ms, and 800 ms. Each conversation lasted about 15 minutes. The preliminary analysis, focusing on floor-transition time and turn-taking frequency, is discussed.

Neural correlates of pragmatic processing within and beyond adolescence

Christoffer Forbes Schieche

Department of Linguistics, Stockholm University

While several fMRI studies have investigated the processing of pragmatic phenomena in adult populations, less is known about the neural correlates of pragmatic processing during adolescence. We recently finalized the acquisition of fMRI data from an adolescent sample (13–18 y.o.) who participated in an indirect speech act (ISA) paradigm. This group is further subdivided into a younger (*Young*, 13–15 y.o.) and older (*Mid*, 16–18 y.o.) subgroup. Additionally, we have data from young adults (*Old*, 18–36 y.o., see Bendtz et al., 2022) who previously completed the same paradigm, which is also split into two groups based on pragmatic task performance outside the scanner. This design allows us to examine not only neural correlates of pragmatic processing during adolescence, but also its development across adolescence and into adulthood.

Initial results show similarities between adolescents and adults, as well as with prior similar studies (e.g., Asaridou et al., 2019; Bašnáková et al., 2014). Across both groups, we observed increased activity within the language and theory of mind

networks when listening to indirect versus direct speech acts. However, we also found group differences outside these networks. Specifically, activity differed between the adolescent subgroups in a cluster in the left posterior intraparietal sulcus. We further observed differences between adolescents and young adults in a cluster spanning the bilateral posterior cingulate cortex. On the other hand, we found similarities between the Young group and adults with poorer pragmatic task performance. In this talk, I will present and discuss these findings.

Poster session

Should chatbots use anthropomorphic expressions like "I think"?

Charlotte Stinkeste

Division of Speech, Music & Hearing, KTH Royal Institute of Technology

AI chatbots become the new front line of digital interaction, the way they speak may matter just as much as what they say. This ongoing study investigates how chatbot conversational style and topic of interaction influence user perceptions and behaviors. Adopting a 3×2 between-subjects experimental design, comparing a chatbots conversational style (machine-like, human-like formal, and human-like friendly) and conversation topic (transactional and social), we examine the effects of these variables on the following outcomes: chatbots perceived warmth, competence, trust, tone appropriateness, anthropomorphism, security, and social presence, as well as disclosure tendency, financial behavior, and enjoyment. Drawing on prior work in human-computer interaction and social cognition, this study explores how anthropomorphic cues particularly linguistic oneness shape user expectations and reactions within different conversational contexts.

The literature suggests that human-like conversational styles increase perceived warmth, social presence, and trust, while formal tones are perceived as more competent and secure in transactional settings. Friendly tones, although potentially less competent in formal contexts, may enhance user enjoyment and self-disclosure, especially in social interactions. Moreover, alignment between conversational style and topic (e.g., formal tone in transactional tasks) is expected to increase tone appropriateness and trust. We hypothesize that conversational style and topic interact to affect users perceptions, with human-like styles (particularly friendly) performing best in social topics and formal styles yielding higher ratings in transactional ones.

Beyond informing chatbot design, these findings provide linguists with empirical insights into how stylistic variation in languageformality, tone, and anthropomorphic cues shapes humans social perception, behavior, and interactional alignment in human-machine communication.
