6.11.24, 16.00-17.00 Uhr, IfL SR rechts

Pablo Faria - *PhD in Linguistics at the State University of Campinas. From 2008 to 2023, he investigated topics in language acquisition and computational linguistics. From 2024 onwards, his research interests comprise the topics of pragmatics, communication, and language acquisition, using theoretical, experimental and computational approaches to describe pragmatic aspects of speakers' linguistic knowledge and verbal interactions, and also the development of pragmatics in children. He also conducts projects and university extension courses on the topic of Nonviolent Communication and is currently a candidate for certification in this topic by the Center for Nonviolent Communication (USA).*

Computational Modeling of the Distributional Learning of Syntactic Categories

A child learning a language has to figure out what the syntactic, or part-of-speech, categories in her language are and assign words to one or more of them. The question we aim to answer here is how much of this learning can be accomplished through the distributional analysis of utterances. To this end, replications of Redington, Chater and Finch' (1998) and Mintz's (2002) computational models were conducted and tested on Brazilian Portuguese input data, obtained from publicly available corpora of both child-directed and adult-to-adult speech. In general, our results seem to align with Redington et al.'s and Mintz's ones. Nonetheless we have found some differences worth attention. In this talk, I will present and discuss our implementations and some of our results.

Fernando Generoso - *Master Student at Universidade Estadual de Campinas. Bachelor in English/Portuguese translation at PUC-Campinas and currently pursuing a master's degree in Language and Thought, with a focus on Language Acquisition at Universidade Estadual de Campinas. Research interests: Language acquisition, computational linguistics, neurolinguistics and cognition.*

Predictive model of anaphoric pronoun occurrence in Portuguese and English based on child corpora

The research project aims to analyze the acquisition of anaphoric pronouns in monolingual children during the early years of life, comparing the use of these linguistic mechanisms in Brazilian Portuguese and English. Data collection will be conducted in child corpora such as CHILDES and CIL, and the Python programming language will be used to create a predictive model for anaphora. The focus of the research will be on anaphoric pronouns in Portuguese and English, covering the pairs [ele(s)/ela(s)] and "isso/esse", [he/she], [it/they], and [this/that], with an emphasis on children between 2 and 4 years old. After data collection, statistical analyses will be conducted using the AntConc software to compare the behaviors of anaphoric processes between the two languages, highlighting similarities and differences. Morphological segmentation of words w1 and w-1 will be performed to observe the frequency of occurrence of each word class in the antecedent and

subsequent positions to an anaphoric pronoun, both in children's and adults' data. The results will be used to feed the predictive model based on the Lappin and Leass algorithm, aiming to evaluate its ability to identify anaphoric pronouns and its accuracy.