

Continuous, Parallel, and Unsupervised Collection of Data on Social Interactions and Speech at Preschool: Issues and Potentials

Aurélie Nardy¹, Jean-Pierre Chevrot¹, H el ene Bouchet¹, Sicheng Dai², Marton Karsai³

¹Lidilem - Univ. Grenoble Alpes, Grenoble, France. ²Univ. Lyon, ENS de Lyon, Inria, CNRS, UCB Lyon 1, LIP UMR 5668, IXXI, Lyon, France. ³Department of Network and Data Science, Central European University, Vienna, Austria

Abstract

Our aim is to introduce a methodology for the collection of speech data from around 170 children and

30 teaching staff in a French preschool, as well as the frequency of social interactions. The technological devices used consist of small boxes worn by participants at school one week a month for 30 months. They contain two microphones for recording voices and a radio frequency identification system for reconstructing social interactions. The children regularly took language tests to assess their lexical and syntactic reception skills. The families completed a socio-demographic questionnaire. This innovative approach is part of an interdisciplinary project involving psycholinguists, sociolinguists, data scientists and ethologists. We collected 1,705,126 social interactions and almost 35,000 hours of audio recordings, of which just over 800 have been transcribed.

This dataset enables the exploration of issues related to the relationships between language development and children's socialization including how preschoolers aged 3 to 6 years influence each other in the process of language development, how teachers and assistants interact with the children and influence their language, and more broadly, how children's language and the network of relationships co-evolve.

To illustrate the potential of such a study, we will present a preliminary 4-month analysis of two first-year preschool classes (G1: ages 3-4) and two third-year classes (G3: ages 5-6), comprising 95 children. The analysis aims at assessing how class membership, sex, mono/multilingualism, parental education and occupation, as well as receptive levels of syntax and lexicon, shape social relationships during classtime and freetime. Our objective is to examine homophily in children, i.e., the principle whereby similar individuals preferentially associate with one another. While various dimensions of homophily have been identified in childhood, such as sex, personality traits, languages spoken, race, and disability (Rubin et al, 1994), the question of the influence of language level remains unanswered.

Our results highlight that the number of homophily dimensions decrease during preschool for both classtime and freetime. Interestingly, homophily based on syntax level, as measured by a comprehension test, is observed during class time in G1 and during freetime in G3.

References

Rubin, K. H., Lynch, D., Coplan, R., Rose-Krasnor, L., & Booth, C. L. (1994). "Birds of a feather...": Behavioral concordances and preferential personal attraction in children. *Child Development*, 65(6), 1778–1785.