# Investigating Teachers’ Talk: Challenges of Classroom Language Data Generation

Irina Johansson Carlén

Malmö University

This presentation is about discussing affordances and constrains of a tentative set of data generation methods regarding classroom language. This is a part of an early-stage PhD project, which investigates which types of registers that characterises mathematical classroom language within the early school years, focusing on the teacher’s use of language registers. The presentation poses questions about data generation consisting of teachers’ talk in whole class instruction that could be analysed quantitatively using corpus and SFL based analyses. Data will be generated and treated through three levels: initially, classroom lessons are recorded and transcribed; further on, transcription texts are treated through a series of suggested corpus linguistics software, aiming to produce a final data set fit for qualitative SFL analysis tools.

## Introduction

Teachers’ choice of language registers introduces and is a role model for how students communicate and reason with and about mathematics. Teachers’ language registers scaffolds students’ exploration of different school academic registers. This means that the language choices that teachers make influence students’ establishment of mathematical language (Erath et al., 2021).

There are many ways of linguistically describing the mathematical language: through perspectives, layers and/or domains (Erath et al., 2021). Despite various explanation models, mathematical language still inhibits inherent tensions (Planas et al., 2023), concerning students’ emerging ability to position themselves in the different registers. It should be noted that the teachers assist in this development simultaneously. Simply put, teachers move in a continuum between all layers bridging these registers. A common way of investigating language use is through a functional perspective based for instance on systemic-functional linguistics (SFL) (Herbel-Eisenmann & Otten, 2011).

This raises questions about 1) what type of register characterises the teachers’ talk in the mathematics classroom in the early school years and 2) what types of linguistic processes characterises teachers’ semantic registers during teaching a) mathematical communication and b) mathematical reasoning?

## Tentative data generation design

To investigate the above-mentioned questions, the plan is to use three levels of data generation. The first level is to record teachers’ talk during the introduction and final summary of a lesson. There are basically two ways to do this: by audio and video recordings, or by audio recordings only, where the teacher wears a microphone during the session. However, these alternatives raise different questions regarding ethics. Further, on the one hand, the amount of data collected by video recordings will be richer as it includes the teacher’s non-verbal instructions. On the other hand, the processes of participation in all types of audio-visual research are complex; by applying core issues of anonymity, especially regarding children, video recording makes the anonymisation of individuals problematic if not impossible.

The second level of data generation is selection and transcription of the recorded material. At this level, there are questions regarding the amount of transcribed material needed for validity in relation to the amount of time in hand. This issue is related to the efficiency of different software for transcription.

The third level concerns construction of a corpus. I plan on using AntConc, a freeware for concordance and text analysis. This corpus, consisting of raw data, will form the foundation of a range of mixed method analysis tools, mainly based on SFL (McEnery & Hardie, 2012).

Based on SFL theory, different analytical approaches could include identifying lexical bundles (Herbel-Eisenmann & Otten, 2011), the use of processes, as well as extracting specific words connected to the mathematical discourse (Biber et al., 2002).

I would like to discuss:

1. What kind of affordances and constrains could the suggested data generation methods provide?
2. Which non-verbal aspects will be the most important to regard in relation to multiple mathematical registers?

## References

Biber, D., Reppen, R., & Conrad, S. (2002). Developing linguistic literacy: perspectives from corpus linguistics and multi-dimensional analysis. *J Child Lang*, *29*(2), 458-462; discussion 489-494. https://doi.org/10.1017/s0305000902235345

Erath, K., Ingram, J., Moschkovich, J., & Prediger, S. (2021). Designing and Enacting Instruction That Enhances Language for Mathematics Learning: A Review of the State of Development and Research. *ZDM: Mathematics Education*, *53*(2), 245-262. https://doi.org/10.1007/s11858-020-01213-2

Herbel-Eisenmann, B. A., & Otten, S. (2011). Mapping Mathematics in Classroom Discourse. *Journal for Research in Mathematics Education*, *42*(5), 451-485.

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics. Method, theory, and practice*. Cambridge: University Press.

Planas, N., Adler, J., & Mwadzaangati, L. (2023). What Is Mathematics Teaching Talk For? A Response Based on Three Sites of Practice in Mathematics Education. *ZDM: Mathematics Education*, *55*(3), 521-534. https://doi.org/10.1007/s11858-022-01452-5