YOUNG STUDENTS' LANGUAGE REPERTOIRES WHEN SOLVING PROBABILITY TASKS

Andreas Eckert and Andreas Ebbelind

Linnaeus University, Sweden

This study investigates how language expressing uncertainty can support learning in probability. Using Systemic Functional Linguistics, we analysed classroom discourse during a five-lesson sequence with sixth-grade students. Video transcripts were coded for modality to examine how teachers and students articulated certainty and conjecture. Early lessons featured low modality linked to personal beliefs, while later lessons connected modal expressions to data-based reasoning. Teacher strategies shifting from mirroring to challenging and grounding statements in evidence proved crucial in fostering probabilistic thinking. Findings underscore the role of language as a resource for teaching probability and call for further research on linguistic approaches to statistical uncertainty.

INTRODUCTION

Teaching students to draw conclusions and make decisions based on data has received increased attention in educational research and school curricula worldwide (Langrall et al., 2017). In everyday language, this can be summarised as making generalisations using probabilistic language and data as evidence (Makar & Rubin, 2009). This probability-based reasoning distinguishes statistical thinking from other mathematical domains by making it inherently non-deterministic and by relying on reasoning with uncertainty (Groth, 2013). Students can use terms like "probably," "maybe," "unlikely," and "it could be" to informally express different levels of uncertainty.

Research in probability learning has frequently documented both the spoken and written language used by children and their teachers. This focus on language is also evident across many subfields of mathematics education, where the primary emphasis is often placed on understanding mathematical concepts, with limited attention given to the language that facilitates this understanding. However, Morgan et al. (2014) critique this approach as overly simplistic, noting that " naïve conceptions of language as a transparent means of transmission of ideas from speaker to listener have been seriously challenged by current thinking about communication" (p. 847). They further identify a second, equally problematic assumption: viewing "language as a barrier to learning that must be overcome" (p. 846).

There is evidence of increasing attention to language and its conceptualisation within the field of probability teaching and learning. For instance, Ben-Zvi et al. (2012) shifted from a naïve view of language as a transparent medium to a more nuanced understanding of language as intertwined with activity, exemplifying research that

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 379-386. PME.

initially treats language as a transparent window into understanding. However, in later work, Ben-Zvi et al. (2012) presented a more complex view of language development, observing how it evolves through students' engagement with increasingly sophisticated probability tasks. Language can convey different levels of certainty with modal verbs such as "must" and "could" (Wagner et al., 2015). It is then up to fellow students or the teacher to interpret that probabilistic language in terms of the degree of certainty it represents and whether it reflects the (un)certainty of the speaker or what the data actually supports. There has, however, been a gap in the research on how to guide the negotiation of probability ideas in real classroom practices (Eckert & Nilsson, 2015) and how probabilistic language develops in student-teacher interactions. With this in mind, we now ask ourselves how uncertainty language can be utilised as a resource to create learning opportunities on the relationship between samples and sample spaces.

Literature background

Mathematical concepts are often perceived as objective truths, but Voigt (1996) argues that this may not be the case for school mathematics. Instead, students' interpretations of mathematical concepts can vary, shaped by their prior experiences. However, social interaction also shapes students' interpretations, as language serves a function in every situation (Halliday & Matthiessen, 2004). The concept of chance exemplifies this ambiguity, as it can be understood in different ways. This provided many opportunities to explore how it was discussed and negotiated in the classrooms in this study. Chance is an informal concept often taken to be synonymous with randomness. In Swedish, the participants' native language, "chance" and "randomness" are the same word, potentially complicating their understanding. Informally, chance is described as an accident or effect whose cause is unknown, or perhaps unknowable (Steinbring, 1993). It contrasts with deterministic events, which are predictable. Chance-based explanations can replace more deterministic ones when the latter are too complex. The use of probability as a predictive tool, instead of deterministic laws, ties the informal concept of chance to probability (Steinbring, 1991). This probability-based reasoning distinguishes statistical thinking from other mathematical domains (Groth, 2013). Hence, the use of language to signal this distinction, a language of uncertainty, is important for young learners.

An informal approach uses everyday language to make statistical discussions more accessible to students (Dolor & Noll, 2015). Students can use terms like "probably," "maybe," "unlikely," and "it could be" to informally express different levels of uncertainty in statistical inference (Makar et al., 2011). The idea is that statistical inference could be approached informally in education as reasoning that generalises from data, using probabilistic language. Introducing students to probabilistic language is a challenge, though (Braham & Ben-Zvi, 2019). The uncertainty could stem from one's capacity, contextual factors in the data collection, or the behaviour of random samples. The statistical uncertainty in the behaviour of random samples can also be quantified, e.g., in terms of sampling variability. Learning how to formulate correct inferences entails exploring these different types of uncertainty (Braham & Ben-Zvi,

2019). Starting in the work of Makar et al. (2019), we investigate how emergent uncertainty language can be utilised as a resource to create learning opportunities.

THEORY

Over the past thirty years, research in mathematics education has increasingly emphasised the role of language as a medium for teaching and learning. In this paper, we follow this trajectory and employ elements of Halliday's Systemic Functional Linguistics (SFL), with a focus on modality (Halliday & Matthiessen, 2004). SFL conceptualises language as an activity or practice rather than as an entity with a predetermined meaning.

Modality refers to language that conveys different levels of certainty, such as modal verbs like "must" and "could" (Wagner et al., 2015). Wagner et al. (2015) illustrate this through several examples "It must be six" shows a higher level of certainty (high modality) than "It could be six" (low modality). Statements without modal verbs (e.g., "It is six") indicate the highest level of certainty, known as root modality, because they do not consider other possibilities. Modal verbs can carry various meanings. For instance, it can denote ability ("You can finish the race"), permission ("You can be excused"), inclination ("I can help you"), or probability ("It can be a six"). These meanings can also include direction; for example, in the probability case, the modal verb distinguishes between "It can be a six" and "It can't be a six", both with root modality. Probability instruction introduces another level of complexity, since the content-specific language and reasoning inherently involve uncertainty, unlike other mathematical domains (Groth, 2013). Students can indicate uncertainty in their own ability to draw correct conclusions, and also recognise that a conclusion based on statistical data has both contextual and statistical uncertainty.

Our study examines how a teacher and pupils in the classroom use language to express modality in probability contexts. The analysis is based on a lesson sequence containing five lessons focusing on probability in a Swedish context. These lessons will be explained further in the results section. We specifically focus on how they articulate uncertainty and conjecture throughout the lesson sequences, rather than on correctness. This attention to language use helps reveal the diverse ways children make sense of and communicate uncertainty. The five lessons were video-recorded and transcribed. Both authors collaboratively reviewed the lessons, identifying instances of language related to probability. Each of these instances was coded in relation to modality as exemplified above. The transcripts were employed to verify the content. Subsequently, patterns within and across lessons were examined to identify similarities and differences. Finally, the modalities were contextualised in terms of their relevance.

RESULT

As previously stated, the lesson series consisted of a 5-lesson intervention on probability in year 6 (12-year-olds). These lessons were led by their usual teacher, who had several years of experience teaching math at the primary level but none in probability theory. The Swedish syllabus for year 4–6 includes random events, chance,

and risk based on observations, simulations, and comparison of probability across different random experiments. It is up to the teacher to decide how much time is spent on the content during the three years. The overall learning goal of the lesson series was to become able to reason about an unknown sample space based on frequency data. Lesson 1 focused on setting up the problem, using a random generator (an opaque bottle with coloured balls) and a horse race game. Students were asked to predict outcomes and reflect afterwards. Lesson 2-4 was about exploring the relationship between relative frequencies and sample spaces using a bottle similar to the one from lesson 1, but clear to make the sample space visible. The students were asked to reflect on issues of randomness, sample size, absolute and relative frequencies, etc. The teacher revisited the opaque bottle in lesson 5 and led the class through data collection and a whole-class discussion on drawing a conclusion about the sample space based on their collected data using probability language. For more details on the actual tasks, c.f. Eckert and Nilsson (2015).

In Lesson 1, students and the teacher primarily use positive low modality when discussing the experiment. The teacher asks for predictions using formulations such as "which colour do you think will win," which contain a low positive modality, and the students respond with corresponding low positive modality, such as "I think red will win." In many cases, the teacher confirms the student's formulation by revoicing it, for example, "You think red will win," without commenting on or investigating the degree of uncertainty underlying the low modality. When the lesson begins to focus on seeking explanations for the results of the investigation, namely that blue seems to win more frequently than the other colours, the teacher continues to ask questions with low positive modality through formulations such as "Why do you think it turned out that way?" The students respond with various possible explanations, using low-positive modality, beginning with "I think that…". Only when a student suggests that the result may be due to chance does the language use change. When a student exclaims "Chance!", the teacher asks "Is it chance?" with positive root modality, and the student then follows up with statements characterised by *"it is" rather than "I think"*. In Lesson 1, the use of high versus low modality does not appear to be directly related to the degree of uncertainty because of the underlying data, but rather to one's personal belief in the certainty of the statement. The language is characterised by mirroring between the teacher's and the students' use of different modalities. The students are not challenged to become either more concrete or more general in their reasoning, but are consistently mirrored and confirmed. A possible explanation is that the teacher is communicating uncertainty about the content or where the lesson is headed rather than about the generalisation based on the small sample, and that the students mirror that uncertainty.

In Lesson 2, a variation of low and high, as well as positive and negative, modality occurs, and the teacher now sometimes varies the modality in responses to students. One example occurs in a sequence in which the teacher and student discuss the contents of the transparent bottle compared to those of the opaque one. A student says, "It can't

be the same bottle," and the teacher responds with the same negative root modality: "In what way can it not be the same bottle then?" When the student replies that it feels like there are more balls, based on the weight, this is coded as positive root modality, but the teacher's response, "Okay, you thought there were more blue ones in the other bottle?" is expressed with a lower modality. The variation in modality gives the student the opportunity to specify what in the data grounds the certainty of the statement. The student follows up with "There were quite a lot of blue ones many times", and then consistently uses *think* rather than *is* in the reasoning. The teacher then summarises the student's reasoning with "since we got so many blue ones last time, there must have been more blue ones in that bottle," which returns the student to their original high positive root modality, with the added grounding in data. These variations in modality were scant but occurred on three occasions, all of which created opportunities to ground statements in observations/data and thereby communicate the degree of certainty/uncertainty underlying the modality.

In Lessons 3 and 4, the focus is largely on procedure: students are to perform calculations and construct bar charts. When the teacher asks for statistical reasoning, i.e., how well the bar charts represent the contents of the bottle, the word *feel* is used, as in "Is there any of them that you feel matches better with how it looks in the bottle?" The students mirror this language use and express themselves in terms of what they think. This does not generate opportunities for students to reason using data and probability language. One possible explanation is that *feel* is based on a subjective opinion rather than an expression of uncertainty in a statistical argument, and thus the students mirror the type of reasoning implied by the teacher's question.

Language of low modality reappear at the second half of Lesson 4. The class has the combined investigation results, 350 observations, displayed on the board. Despite the results not fully corresponding to the distribution of colors in the bottle, i.e. one third of each, the teacher uses her authority and states that relative frequency is still useful through the claim that "the law of large numbers says that the more observations we have, the greater the chance that we get a picture that corresponds to reality." indicating that her certainty/belief in the mathematics has increased. As the discussion continues, it instead focuses on how relative frequency is a useful tool in conducting a statistical investigation. When a student states that one could investigate students' favourite ice cream flavour, the student uses low modality with words such as 'would' and 'maybe'. The modality relates to the use of the method rather than an expression that the result of such an investigation would involve a certain degree of uncertainty. One interpretation could be that the student is attempting to express uncertainty about whether this is actually a useful method, i.e., statistical reasoning, for saying something about reality.

In Lesson 5, when students were to carry out an investigation to formulate a statistical inference about the distribution of balls in the opaque bottle, the language used varied in modality. In this lesson, the content of the statements also oscillates between the concrete, how the relative frequencies in the sample turned out, and generalisation,

how the relative frequency should be interpreted in relation to the contents of the bottle. Thus, creating learning opportunities on the relationship between samples and sample spaces.

For example, when a student says, “One ball is worth 2%,” the teacher responds, “One ball is worth 2%; how can we use that to understand approximately the distribution?” Both use positive root modality, but in this lesson, the teacher explicitly uses it to create opportunities for generalising reasoning from data. Similarly, when a student says, “I think there are two white [balls],” the teacher follows up with, “You think there are two white, how are you thinking then?” thereby creating an opportunity for the student to base the statement on the fact that there was a higher frequency of white balls in the sample.

The culmination of this shift in how utterances with different modalities are taken up occurs toward the end of the lesson, when the entire class’s combined sample of 350 observations from the opaque bottle is summarised on the board. They agree that 10 percentage points corresponds to one ball, since that was the relative frequency for red. A student says, “but then that becomes two more on one [white] and one more on one [blue],” and the teacher responds, “now you need to be a bit more precise.” She challenges the student’s vague generalisation and the data on which it is based. The student follows up with, “Since white won, I think it will be two more added to white,” which is to be added to the earlier estimate of three white balls in the bottle, resulting in a new estimate of a total of five white balls in the bottle based on data. The student’s statement has low positive modality, and the uncertainty is clearly linked to the generalisation based on data from the investigation. An additional example from another student who arrives at a similar conclusion is: “the white ones are 158, and that’s almost like 160, and 40 times 4 is 160, so then it should be 4… or four times as many white as red.” This statement also has low modality and appropriately expresses uncertainty through probability language in its generalisation based on the sample.

In summary, low modality was used in all five lessons, but it was not until the fifth lesson that the uncertainty was linked to statistical reasoning, whereas in earlier lessons it was more person-based. The results reveal clear patterns in language use and its progression across the lessons. The teacher predominantly used low-modality questions to encourage exploration and conjecture. High modality was used for classroom management and rule enforcement, signalling authority and certainty, while medium modality appeared in explanations of probabilistic concepts. Students mostly operated in low modality, expressing uncertainty through language such as “I think” or “maybe.” Occasional high modality reflected strong claims (e.g., “it must be,” “it cannot be”), and root modality occurred in spontaneous assertions of outcomes to questions. Across the five lessons, modality evolved from low modality toward a blend of low to root modality.

DISCUSSION

This study set out to examine how linguistic resources for expressing uncertainty, conceptualised through modality, can be used to create learning opportunities about the relationship between sample and sample space. The analysis reveals patterns in the use of modality by both teachers and students across a sequence of probability lessons, as well as qualitative shifts in how uncertainty is articulated over time. Initial results indicate that the teacher controls the use of modality, and the student responds, either by mimicking the teacher or by adjusting to challenges. By foregrounding language use rather than correctness, the study highlights interesting aspects of understanding probabilistic meaning-making as a discursive process.

An interesting concern is the development of meaning across the five lessons. In the initial lesson, classroom discourse is predominantly characterised by low modality, with both teacher and students producing tentative utterances expressing possibility, conjecture, and uncertainty. From an SFL perspective, these choices can be interpreted as linguistic realisations of a stance that keeps the space of meaning open. In this phase, uncertainty can function as a resource that legitimises speculation and supports collective inquiry. However, as in the example from lesson 1, the uncertainty in this case might also depend on the teacher not knowing how to proceed with the lesson. From the students' point of view, the modal verbs are the same even though the origin of the uncertainty changes. Comparing the patterns from lesson 1 to those of lesson 5, the teacher challenges students' choices of modal verbs to a greater extent. The teacher connects the inferences to the data at hand, asks students to become more explicit, and switches between modalities in the process. From an SFL perspective, these choices can be interpreted as linguistic realisations of a stance that steers the space of meaning toward statistical uncertainty. In this phase, uncertainty can function as a resource to elicit inferences that move beyond speculation towards generalisations based on data.

Similar to the results of Makar et al. (2011), the use of modal verbs oscillated between certainty (deterministic) and uncertainty (relativistic) statements, and between different types of uncertainty, similar to students in Braham and Ben-Zvi (2019). What is interesting, and adds to previous results, is how, by foregrounding language use as a system of choices rather than correctness, we also see how the modal verbs in some cases communicate uncertainty that does not have to do with statistics. Our students increasingly employed higher modality when drawing on numerical calculations. At the same time, the teacher continues to use low-root modality interrogatives, for example, how or what, to invite further reasoning and maintain dialogic openness. What could have been a communication of statistical uncertainty could just as well be interpreted as an uncertainty where the lesson is headed mathematically. We therefore invite further investigation into the linguistics of statistical uncertainty in lower primary school to disentangle how language can be used as a conscious resource to create learning opportunities on the relationship between samples and sample spaces.

References

Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM - the International Journal on Mathematics Education, 44*(7), 913–925. https://doi.org/10.1007/s11858-012-0420-3

Braham, H.M., & Ben-Zvi, D. (2019). Design for reasoning with uncertainty. In G. Burrill & D. Ben-Zvi (Eds.), Topics and trends in current statistics education research. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-030-03472-6_5

Dolor, J., & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal, 14*(1), 60–89. https://doi.org/10.52041/serj.v14i1.269

Eckert, A., & Nilsson, P. (2015). Introducing a symbolic interactionist approach on teaching mathematics: The case of revoicing as an interactional strategy in the teaching of probability. *Journal of Mathematics Teacher Education, 20*(1) 31–48. https://doi.org/10.1007/s10857-015-9313-z

Groth, R. E. (2013). Characterizing key developmental understandings and pedagogically powerful ideas within a statistical knowledge for teaching framework. *Mathematical Thinking and Learning, 15*(2), 121–145. https://doi.org/10.1080/10986065.2013.770718

Halliday, M., & Matthiessen, C. (2004). *An introduction to functional grammar* (4th edition). Arnold.

Langrall, C., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). Teaching and learning probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 490–525). NCTM.

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1–2), 152–173. https://doi.org/10.1080/10986065.2011.538301

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82–105.

Morgan, C., Craig, T., Schuette, M., & Wagner, D. (2014). Language and communication in mathematics education: an overview of research in the field. *ZDM, 46*(6), 843–853. https://doi.org/10.1007/s11858-014-0624-9

Steinbring, H. (1991). The concept of chance in everyday teaching: Aspects of a social epistemology of mathematical knowledge. *Educational Studies in Mathematics, 22*(6), 503–522. https://doi.org/10.1007/bf00312713

Steinbring, H. (1993). The context for the concept of chance: Everyday experiences in classroom interactions. *Acta didactica Universitatis Comenianae Mathematics,* (2), 20–32.

Voigt, J. (1996). Negotiation of mathematical meaning in classroom processes: Social interaction and learning mathematics. In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin & B. Greer (Eds.), *Theories of Mathematical Learning*. Lawrence Erlbaum Associates.

Wagner, D., Dicks, J., & Kristmanson, P. (2015). Students' language repertoires for prediction. *The Mathematics Enthusiast, 12*(1-3), 246–261.https://doi.org/10.54870/1551-3440.1346