

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Espinoza, Catalina; Pikkarainen, Toni; Viiri, Jouni; Araya, Roberto; Caballero, Daniela; Jiménez, Abelino; Gormaz, Raúl

Title: Analyzing teacher talk using topics inferred by unsupervised modeling from textbooks

Year: 2020

Version: Published version

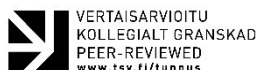
Copyright: © 2020 the Author(s)

Rights: CC BY-SA 4.0

Rights url: <https://creativecommons.org/licenses/by-sa/4.0/>

Please cite the original version:

Espinoza, C., Pikkarainen, T., Viiri, J., Araya, R., Caballero, D., Jiménez, A., & Gormaz, R. (2020). Analyzing teacher talk using topics inferred by unsupervised modeling from textbooks. *FMSERA Journal*, 3(1), 4-17. <https://journal.fi/fmsera/article/view/79631>



ANALYZING TEACHER TALK USING TOPICS INFERRED FROM UNSUPERVISED MODELING FROM TEXTBOOKS

Catalina Espinoza¹, Toni Pikkarainen², Jouni Viiri², Roberto Araya¹, Daniela Caballero¹, Abelino Jiménez¹ Raúl Gormaz¹

¹Centre for Advanced Research in Education (CIAE) – University of Chile

²University of Jyväskylä

ABSTRACT

We propose a method that automatically describes teacher talk. The method allows us to describe and compare classroom lessons, as well as visualizing changes in teacher discourse throughout the course of a lesson. The proposed method uses a machine learning model to infer topics from school textbooks. Certain topics are related to different contents (e.g. kinematics, solar system, electricity), while others are related to different teaching functions (e.g. explanations, questions, numerical exercises). To describe teacher talk, the machine learning method measures the appearance of the inferred topics throughout each lesson. We apply the proposed method to a collection of transcripts from physics lessons, as well as discussing the potentialities of integrating the proposed method with other kinds of automatic and manual classroom lesson descriptions.

INTRODUCTION

Classroom talk is important for learning interactions to take place. Therefore, analyzing the content and quality of teacher and student talk has been an active area of educational research (e.g., Mortimer & Scott, 2003). Examples of content analysis include counting the number of questions, describing which concepts are dealt with, and identifying who is talking. Analyzing the quality of teacher and student talk includes looking at what kind of questions are posed, as well as whether the talk is dialogic or interactive. The results of these studies are often published as frequencies and cumulative counts. However, learning does not occur momentarily at the end of each lesson, but over time. The frequency descriptions do not reveal the temporal dynamics of teaching and learning and, consequently, many researchers have underlined the importance of temporal

analysis of teaching-learning events (e.g. Mercer, 2008; Knight, Wise, & Chen, 2017; Csanadi, Eagan, Kollar, Shaffer, & Fischer, 2018).

Working with automatic descriptors of classroom talk (i.e. extracted automatically by a computer) is more cost-effective than traditional observational methods. It is also faster and can be used for online processing of teacher talk. Therefore, using automatic descriptors may help focus traditional methods on the most critical aspects of the dynamics of a lesson. The potential of automatic content analysis has already been acknowledged in the field of education (Rosé et al., 2008; Araya et al., 2012; Wang, Pan, Miller, & Cortina, 2014; Scheihing, Vernier, Born, Guerra & Cárcamo, 2016, Donnelly et al., 2017). For instance, Blanchard et al. (2015) evaluated different automatic speech recognition systems when analyzing teacher talk during noisy lessons. Kelly, Olney, Donnelly, Nystrand, and D’Mello (2018) analyzed teacher questions in classroom situations. They used automatic speech recognition (ASR), natural language processing (NLP), and machine learning to train computers to recognize whether teacher questions were authentic or rhetorical.

There are several automatic methods for summarizing text (Gambir and Gupta, 2017). In general, both unsupervised and supervised machine learning methods need large amounts of data. This is also the case for topic models, a class of unsupervised machine learning methods which infer topics as groups of words that occur together in a text. Topic models have been applied in a variety of contexts and have been preferred in cases where interpretability and speed are essential (Boyd-Graber, Hu & Mimno, 2017).

Topic models can compute a numerical representation of the topics discussed during a lesson. A topic inferred by a topic model is defined by a large list of words, sorted by a degree of belonging to the topic. To determine if a topic was discussed in a lesson, none of the words that define the topic are strictly necessary. Instead, a topic is recognized by the presence of a combination of words with a high degree of belonging, a combination of words that may vary in different contexts. For example, to recognize a topic related to kinematics, there is no need for the words “motion” or “velocity”. The topic might instead be deduced by the presence of other words such as: “position”, “change” and “time”. This feature makes the topic models suitable for describing teacher talk, as it can adapt to different variations of language.

One way of obtaining descriptors of teacher talk would be to train a topic model using lesson transcripts. However, the viability of this approach is limited as lesson transcripts are difficult to obtain. School textbooks, on the other hand, are widely available and can be found in different languages. Textbooks provide a large collection of written texts in which content is labelled by subject and grade. Since textbooks are usually aligned with the national curriculum, topic models make it possible to compare teaching across different countries.

For these reasons, we propose training a topic model using school textbooks in order to find a characteristic set of topics that can be used to describe the lesson

transcripts. To describe teacher talk, we quantify the appearance of the inferred topics throughout a lesson. We show that this approach can be used to describe a single lesson or to compare different lessons, as well as describing the temporal development of a lesson.

Our research questions (RQ) are the following:

- RQ1: What kind of automatic lesson descriptors does the topic model capture?
- RQ2: How do these descriptors differentiate between lessons?
- RQ3: How does the automatic lesson descriptor provide insights into the temporal development of a lesson?

Answering these questions will help us understand the scope and limitations of the proposed method for describing speech in learning environments. RQ1 aims to determine what kind of content analysis can be automated. RQ2 aims to answer whether the method is able to capture and explain differences between lessons, and thus whether it is suitable for monitoring changes in teacher practice. RQ3 aims to capture speech patterns during lessons that could later be related to student learning gains in future research.

METHOD

Data collection

The proposed method has two sources of data: school textbooks and classroom lesson transcripts. We collected nineteen audio recordings from physics lessons: fifteen from lower secondary lessons and four from upper secondary lessons. The lessons were transcribed using an Automatic Speech Recognition (ASR) system developed by Aalto University (Kronholm et al., 2017, Caballero et al., 2017). The content of the selected school textbooks needs to be aligned with the content of the classroom lesson transcripts. We therefore collected a total of 31 physics school textbooks available in Finnish: *Fysiikka* (1,2,3,5,6,7,8,9), *Kuutio* (7), *Ilmiö* (7-9), *Physica* (1,2,3,4,5,6,7,8,9), *Fotoni* (1,2,3,4,6,7,9), *FYKE*, *FYKE* exercise book (1,2), *Aine ja energia* and *Titaani*.

Data analysis

Figure 1 outlines the three-stage method for obtaining automatic descriptions of teacher talk. In the first stage, the collected lesson transcripts and school textbooks are pre-processed. In the second stage, an unsupervised topic model is trained using the pages from the textbook. This produces a set of topics that characterizes the information in the textbooks. In the third stage, the pre-processed lessons are described using the topic model. Following this, the lesson descriptions are visualized in order to answer the research questions. We describe each of these stages below.

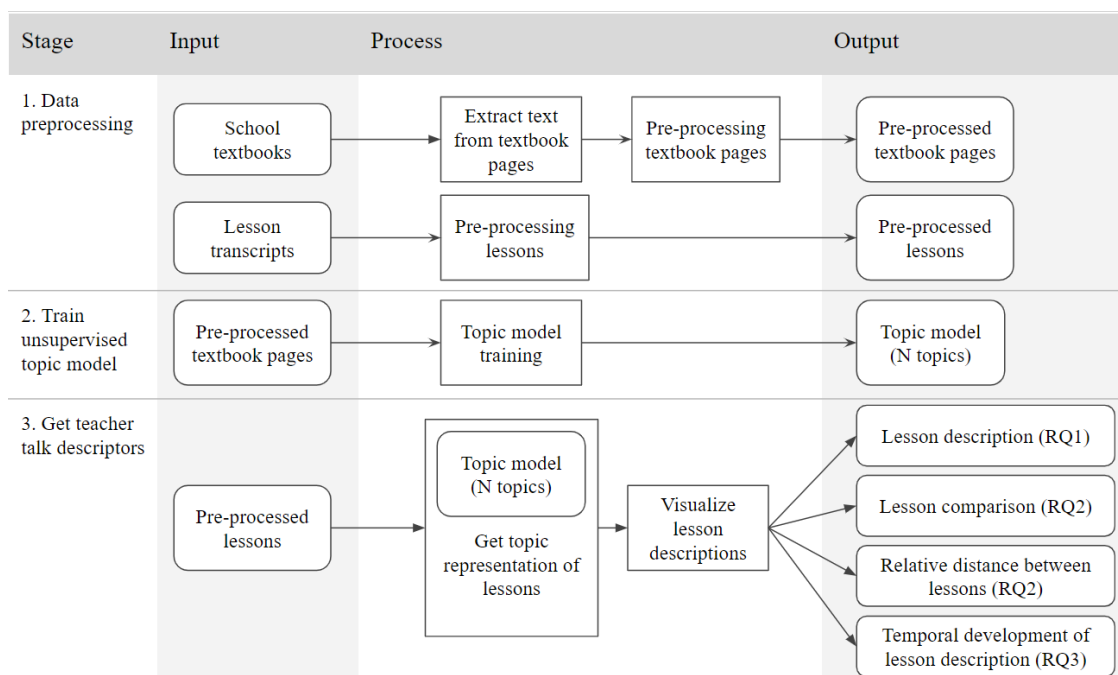


Figure 1. Stages for obtaining automatic descriptions of teacher talk from classroom lesson transcripts.

1. Data pre-processing

The school textbooks and classroom lesson transcripts are pre-processed independently, though following a similar process. The main difference is that classroom lessons transcripts are assumed to be already in a digital text format (".txt"). On the other hand, the school textbooks can be collected in several formats. Therefore, in order to pre-process them, the text from the textbook pages must first be digitalized. This was done using the Python libraries *PyPDF2*, *docx* and *tesseract*, to extract the text from ".pdf", ".doc" and ".jpg" files, respectively. Another difference between the pre-processing of lesson transcripts and textbook pages is that we only select the textbook pages where content or exercises are presented within a context. Therefore, the introduction and supplementary material (e.g. glossary) in each textbook was ignored. Similarly, only textbook pages with more than 20 words were considered. Other than this, the method for pre-processing the two sources of data is the same.

Both the textbook pages and the lesson transcripts contain groups of words that work as synonyms. To improve the performance of the topic model, words from each of these groups must be replaced by labels. Table 1 shows examples of these groups, as well as the label used to replace the words in the textbooks and lesson transcripts.

Table 1. Examples of groups of words that are replaced by a label.

Label	Words
NUMBER	A regular expression of any number
NAME	A list of Finnish first names
EXPLAINING	Eli, kuten, perustuu
EXAMPLE	Esimerkiksi, esimerk
TIME	Aika, aikaa, aikaan, ajasta, ajan
CALCULATE	Laskekaa, laskee, laskemaan, laskea
SUM	Summa, summia
CONCEPT	Käsite, käsitteitä, käsitteinä, käsitteenä, käsitteet

General techniques for pre-processing text are then applied. Any words that are not labels are transformed to lower case. Most symbols are removed, with only semantically-significant symbols being kept (such as the question mark and arithmetic symbols). Stop words are removed (i.e. common words that do not add any meaning to the text, e.g. ei, heidän, niinä). A stemming process is applied at the end of the pre-processing. This process aims to return the stem of each word, transforming similar words (e.g. plurals or verb conjugations) into a single term. The *Nltk* Python library (Bird, Klein & Loper, 2009) was used for the removal of stop words, as well as the stemming process. *Nltk* is a text processing library with support for several languages, including Finnish.

2. Training an unsupervised topic model

An unsupervised topic model is a machine learning technique capable of finding a set of topics in a collection of documents. In this study, we use a Latent Dirichlet Allocation model (LDA) (Blei, Ng and Jordan, 2003). LDA is a commonly used topic model that characterizes each document as a mixture of topics. The LDA topics are lists of words ordered by their degree of belonging to the topic. More specifically, LDA topics are defined as probability distributions over the vocabulary (i.e. all the terms included in the training data). LDA models work with the documents as a bag of words (i.e. the order in which the words appear in the document is irrelevant).

To train the LDA model we used the *Gensim* Python library (Rehurek & Sojka, 2010). The LDA model is trained to obtain 60 topics from the pre-processed

textbook pages. Usually, topic models are used to describe the same kind of documents that are used for training. However, in this case, we trained the LDA model using textbook data and then used the model to describe classroom lesson transcripts.

3. Description of teacher talk using the topic model

Generally-speaking, the LDA model computes a vector for a document, where each dimension corresponds to a topic probability. This vector is a probability distribution. The sum of all topic probabilities is therefore one. Each topic probability represents the likelihood of associating the text in the document with the corresponding topic. For our study, we obtain 60-dimensional topic probabilities for classroom lesson transcripts using the 60-topic LDA model. These vectorial representations of discourse are what we call automatic lesson descriptors. It is a powerful, easy-to-interpret numerical description that allows for a quantitative comparison between lessons. Furthermore, it provides a tool for identifying the most important topics, as well as the proportion in which they are combined in a lesson.

To answer RQ1, we used the method to obtain a summarized description of the lessons. We calculated the topic probabilities for the lessons using the 60-topic LDA model. As mentioned previously, each LDA topic is a list of words that is sorted by the degree to which they belong to the topic. To interpret the topic probabilities as lesson descriptors, we check the top words of the topics with the highest probability). Depending on the topics that emerge from different lessons, we can understand what kind of content analysis can be automated.

To answer RQ2, we compared the automatic lesson descriptors in order to understand which topics were differentiators and how these topics varied between lessons. We compared lessons of similar and different content. Furthermore, we also calculated the relative distance between the lessons to see if the one-on-one differences could induce a coherent grouping of lessons. Since the topic representation of lessons is a probability vector, Euclidean distance can be used as a metric to cluster lessons by their descriptors. Following this idea, we used the *Scipy* Python library (Jones, Oliphant, Peterson, et al., 2001) to compute agglomerative hierarchical clustering (Müllner, 2011). Hierarchical clustering is divided into stages, with the number of stages depending on the number of elements to be joined. Initially, each lesson belongs to its own cluster. At each stage, the closest pair of clusters is linked. To measure the distance between clusters we use the 'weighted' method, in which the distance between an old cluster and a new cluster formed by the union of sub-clusters, is the arithmetic mean of the distance from the old cluster to all the sub-clusters. At the end of the clustering process, the lessons are linked in order of similarity. If the topic representation is suitable for detecting differences between lessons, then the clustering method should reveal a defined structure that relates similar lessons.

To answer RQ3, we described the development of topics during lessons. We split the lessons into time intervals and then tracked the topic probabilities for each

interval. We split the lesson transcripts into 10-line intervals. Each line of the transcript corresponds to 10 seconds of the original audio recording. Therefore, the length of each interval is 100 seconds. The method of topic representation for an interval is essentially the same as for the whole lesson. However, visualizing 60-topic probabilities for all of the intervals in a lesson is not viable. We first selected the most important topics as the ones with the highest probability for the whole lesson. We then visualized the development of the most important topic probabilities throughout the course of the lesson.

RESULTS

Description of a lesson

Table 2 shows an example of some of the 60 topics that the model obtained from the collection of Finnish physics textbooks. Each column represents a topic. The header contains the topic ID and labels that have been manually added in parentheses. The rows contain the words with the highest degree of belonging to each topic. The English words in capitals refer to the groups of words presented in Table 1.

Some topics are coherent. For instance, Topic 16 (direct current) includes words such as current, voltage, resistance, circuit, battery and lamp. On the other hand, Topic 37 (question) includes the word “kuva” (picture) which is not a question word, unlike the Finnish word “kuink” which is the stem of “kuinka” (how). However, “kuva” (picture) may appear in the textbook as part of a sentence such as “How heavy is the object in the picture?” and consequently belong to topic 37 (question).

Table 2. Example of seven topics and the top 6 words for each.

Topic 16 (direct current)	Topic 37 (question)	Topic 38 (atom)	Topic 45 (explain)	Topic 50 (electricity)	Topic 59 (radiation)	Topic 60 (calculate)
sähkövir	?	elektron	EXPLAINING	ph	säteily	NUMBER
jännit	kuink	ytim	FOR EXAMPLE	silmuk	aine	+
resistans	mite	ATOM	TIME	napo	ionisoiv	l
virtapiir	kuva	proton	muun	sähkövirt	gammasäteily	sarj
paristo	a	ydin	aine	vaihtojännit	radioaktiivis	ratkaisu
lampu	NAME	elektro	kuite	led	röntgensäteily	yhtälö

Figure 2 shows an example of a topic representation for a lesson on electricity, obtained using the 60-topic LDA model. There are four topics that stand out from the rest: Topic 16 (current), Topic 37 (question), Topic 45 (electricity) and Topic 50 (calculation).

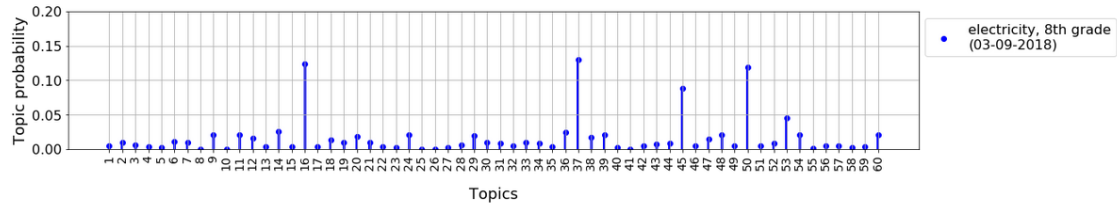


Figure 2. Automatic teacher talk descriptor for a lesson on electricity.

Description of two similar lessons

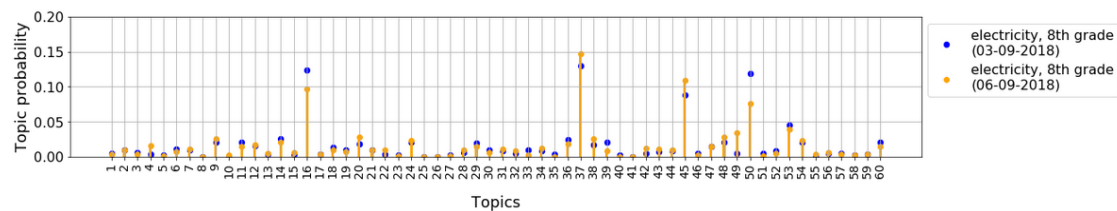


Figure 3. Comparison of topic representations for two lessons on similar content.

Figure 3 shows an example of the topic representations for a pair of lessons on electricity. The lessons were taught by the same teacher to the same class on different days. The blue dots represent the topic probabilities for the lesson on electricity shown in Figure 2. The orange dots represent the topic probabilities for the lesson on electricity taught to the same class three days later. The same four topics stand out from the rest: Topic 16 (direct current), Topic 37 (question), Topic 45 (explain) and Topic 50 (electricity), although there are some differences. Topic 37 (question) and Topic 45 (explain) have a higher probability in the second lesson. Conversely, Topic 16 (direct current) and Topic 37 (electricity) have a higher probability in the first lesson.

Description of two different lessons

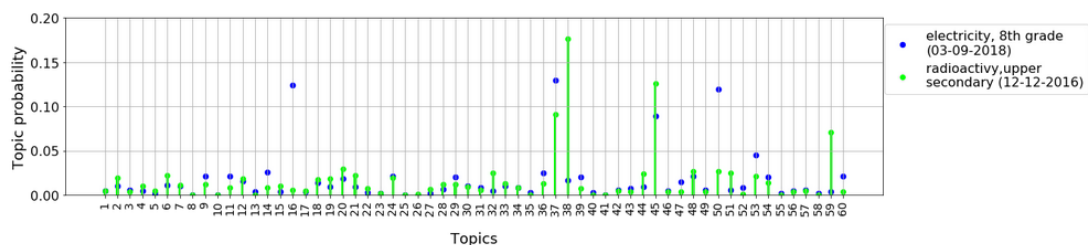


Figure 4. Comparison of topic representations for two lessons on different content.

Figure 4 shows an example of the topic representations for a lesson on electricity and a lesson on radioactivity. The blue dots represent the topic probabilities for the lesson on electricity shown in Figure 2. The green dots represent the topic

probabilities for a lesson on radioactivity, taught by a different teacher to an upper secondary class. For the lesson on radioactivity, four topics stand out from the rest: Topic 37 (question), Topic 38 (atom), Topic 45 (explaining) and Topic 59 (radiation). The two lessons both reveal a high probability for Topic 37 (question) and Topic 45 (explaining). However, Topic 45 (explaining) has a higher topic probability in the lesson on radioactivity and a lower topic probability for Topic 37 (question). The main difference between the two lessons are the two topics related to their respective content.

Relative distance between lessons

Figure 5 shows the result of using hierarchical clustering with the topic representation of the collected physics lessons. The lessons are labelled according to the content of the lesson and an ID to identify them. The blue lines indicate distance between clusters of lessons that are above a threshold of 0.1. The lines that are not blue represent the four clusters of lessons with an intra-cluster distance of less than 0.1. We selected this threshold because it generates the largest number of clusters that encapsulate lessons of similar content. The lessons 'temperature_1' and 'radioactivity_2' are the only ones that do not belong to a cluster. Most of the electricity lessons are grouped together and are closer to each other than the lessons from other clusters.

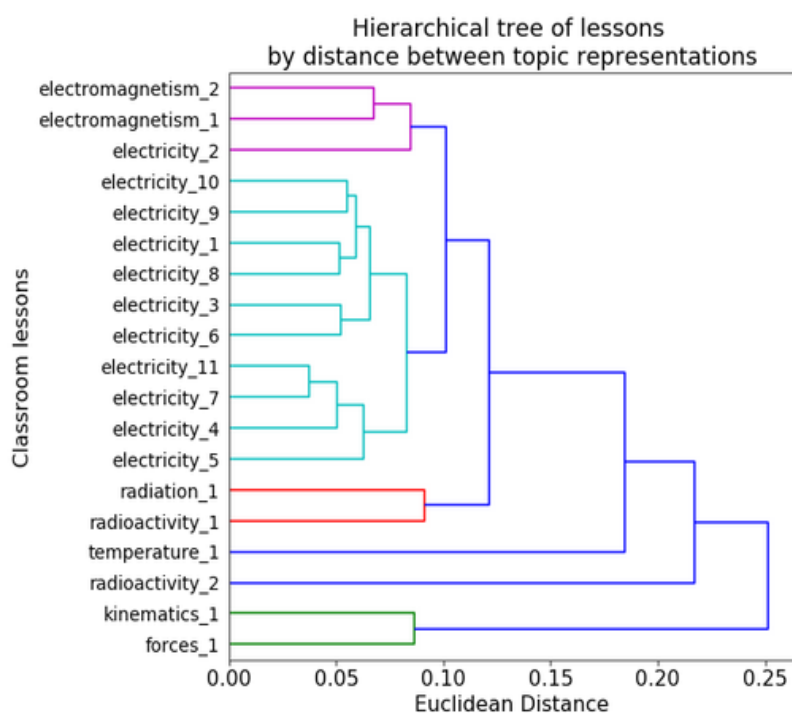


Figure 5. Hierarchical tree of physics lessons by distance between topic representations.

Temporal description of lessons

Figure 6 shows the temporal development of the topic probability over the course of the electricity lesson shown in Figure 1. The lesson is divided into twenty-four

intervals. For each interval, the probability of the topics with the highest probability is shown. The topics related to content (Topic 16 (direct current) and Topic 50 (electricity)) are grouped together. This reduces the amount of information in the figure, allowing us to easily identify which intervals in the lesson have the highest or lowest proportion of concepts relating to electricity. There are three moments during the lesson when the topics with the highest probability were not related to electricity: at the beginning of the lesson, during intervals 16 and 17, and at the end of the lesson. The importance of the topics is distributed evenly between intervals 4 and 15. Following this, there are then two intervals in which there is a decrease in the probability of topics relating to electricity, with a slight increase in Topic 37 (question). Between intervals 18 and 21 there is another peak in topics relating to electricity. Topic 37 (question) is the most important topic in the two final intervals.

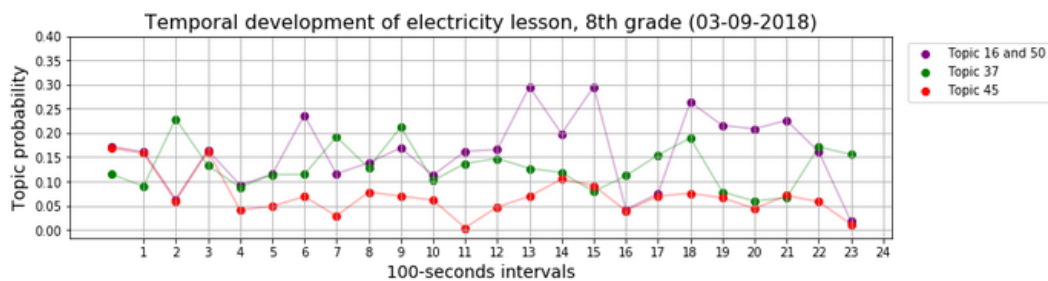


Figure 6. Time development of topic representations throughout the course of a lesson on electricity.

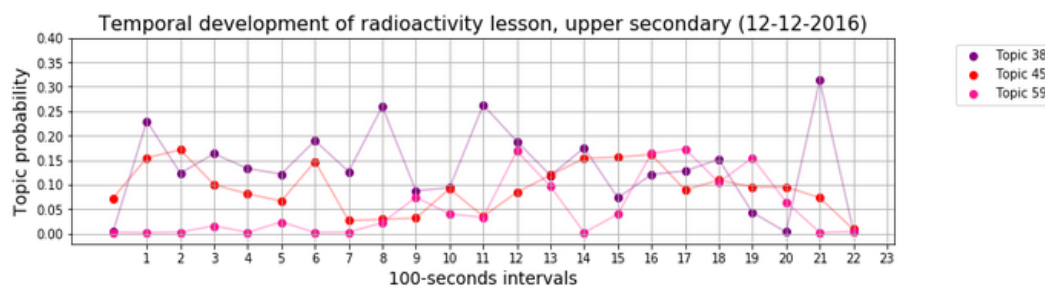


Figure 7. Time development of topic representations throughout the course of a lesson on radioactivity.

Unlike the lesson on electricity, the two topics related to the content of the lesson on radioactivity introduced in Figure 3 are more distinguishable: Topic 38 refers to atomic structure, while Topic 59 refers to radiation. Therefore, by computing the temporal development of the topics throughout the course of a lesson it is possible to visualize how the two topics are integrated in the teacher talk: whether these topics appear at the same time or are treated separately. Figure 7 shows that the first half of the lesson is dominated by Topics 38 and 45. It is only in the second half of the lesson that Topic 59 is integrated with the other topics, particularly during between intervals 16 and 19.

DISCUSSION AND CONCLUSION

In this study, we presented a method for automatically obtaining a summarized description of teacher talk. The method uses an unsupervised learning model to compute a set of topics that emerge from school textbooks. The inferred set of topics is then used to describe teacher talk from classroom lesson transcripts. The main assumption behind the proposed method is that a topic model trained with school textbooks provides a useful description of teacher talk.

We apply the proposed method using a set of Finnish physics textbooks and nineteen physics lessons on different topics. We obtained a list of 60 topics by training an LDA model. These LDA model topics are groups of words that have the highest probability of co-occurring in the textbooks. The topics are interpretable and refer to several different content items (e.g. electricity, atom) and teaching functions (e.g. explanations, questions, numerical exercises) (Table 2). Using the LDA model, we obtained a teacher talk descriptor: a vector with 60 topic probabilities. The teacher talk descriptor allows us to capture the content of the lesson and the proportion of general teaching discourse (Figure 2, Figure 3, Figure 4). We observe a level of consistency between the teacher talk descriptors: two similar lessons are characterized by similar topic probabilities (Figure 2), while two lessons on different topics are characterized by distinguishable topic probabilities (Figure 3). Moreover, the teacher talk descriptors make it possible to group a collection of lessons based on their similarities (Figure 4).

In addition to providing a description of entire lessons, we also use the teacher talk descriptors to show the temporal development of discourse over the course of a lesson. As expected, the method captures huge variations in the topics used during the lessons (Figure 5, Figure 6). In particular, topics related to content have a greater range of variability. In this sense, there are intervals where the topic probability is closer to zero and other, larger intervals where they are the main topics. Furthermore, the teacher talk descriptors reveal a potential for detecting intervals where concepts from different topics are integrated (Figure 6). More research is needed to validate the accuracy of this temporal description. This implies gathering a larger collection of classroom lesson transcripts.

One limitation of the proposed method is that it does not capture administrative teacher talk. As the topic model is trained with textbooks, there is a lack of classroom management discourse in the teacher talk descriptors. We therefore hypothesize that a more robust set of topics could be obtained if the topic model were trained using both textbooks and lesson transcripts. Another limitation is that although the teacher talk descriptor is composed of 60 topic probabilities, only the most important topics are used to interpret the characteristics of a lesson. Therefore, traditional observation methods are more suitable for obtaining a more detailed description of teacher talk. For example, one of the lessons on electricity was closer to the lessons on electromagnetism than to other lessons on electricity (Figure 4). The teacher talk descriptor for that lesson reveals a wider

distribution of the topics. It also reveals a lower probability of Topic 16 (direct current) than the other lessons on electricity. Although the descriptions provide an insight into why this lesson is separated from others, other questions arise: Was the lesson an introductory lecture? Did the teacher use a dialogic approach? Understanding why this lesson focused less on concepts relating to direct current would require a qualitative analysis.

In this study, the proposed method was applied to Finnish physics lessons. However, it can also be applied to lessons taught in different languages and based on different subjects. Gathering more lesson transcripts will enable us to validate and extend the potentialities of the proposed method. For example, it will be possible to train a topic model with data from textbooks and lessons, and then test whether this improves the differentiation of lessons by content, teacher or grade. Our ultimate goal is to be able to relate talk patterns in learning environments with student learning gains. We aim to use the method presented in this study to measure how deeply a topic is discussed, at what stage during the lesson, and with which other topics it is related. In the future, we hope to relate the temporal integration of topics with student learning gains.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding received from PIA-CONICYT Basal Funds for Centers of Excellence Project FB0003, as well as the AKA-EDU-01 grant from CONICYT. The authors are also thankful for the funding provided to project No. 294218 by the Academy of Finland.

REFERENCES

- Araya, R., Plana, F., Dartnell, P., Soto-Andrade, J., Luci, G., Salinas, E., & Araya, M. (2012). Estimation of teacher practices based on text transcripts of teacher speech using a support vector machine algorithm. *British Journal of Educational Technology*, 43(6), 837-846.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc."
- Blanchard, N., Brady, M., Olney, A. M., Glaus, M., Sun, X., Nystrand, M., ... & D'Mello, S. (2015, June). A study of automatic speech recognition in noisy classroom environments for automated dialog analysis. In *International Conference on Artificial Intelligence in Education* (pp. 23-33). Springer, Cham.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Boyd-Graber, J., Hu, Y., & Mimno, D. (2017). Applications of topic models. *Foundations and Trends® in Information Retrieval*, 11(2-3), 143-296.

- Caballero, D., Araya, R., Kronholm, H., Viiri, J., Mansikkaniemi, A., Lehesvuori, S., ... & Kurimo, M. (2017, September). ASR in classroom today: automatic visualization of conceptual network in science classrooms. In *European Conference on Technology Enhanced Learning* (pp. 541-544). Springer, Cham.
- Donnelly, P. J., Blanchard, N., Olney, A. M., Kelly, S., Nystrand, M., & D'Mello, S. K. (2017, March). Words matter: automatic detection of teacher questions in live classroom discourse using linguistics, acoustics, and context. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 218-227). ACM.
- Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review*, 47(1), 1-66.
- Jones, E., Oliphant, T., & Peterson, P. (2014). SciPy: Open source scientific tools for Python. <http://www.scipy.org/>
- Kelly, S., Olney, A., M., Donnelly, P., Nystrand, M., & D'Mello, S., K. (2018). Automatically Measuring Question Authenticity in Real-World Classrooms. *Educational Researcher*, 47(7), 451-464
- Knight, S., Wise, A.F., & Chen, B. (2017). Time for Change: Why Learning Analytics Needs Temporal Analysis. *Journal of Learning Analytics*, 4(3), 7-17.
- Kronholm, H., Caballero, D., Mansikkaniemi, A., Araya, R., Lehesvuori, S., Pertilä, P., Virtanen, T., Kurimo, M., & Viiri, J. (2017). The automatic analysis of classroom talk. In *Proceedings of the annual FMSERA symposium 2016* (pp. 142-151). Finnish Mathematics and Science Education Research Association (FMSERA). Retrieved from <https://journal.fi/fmsera/article/view/60940/27049>.
- Mercer, N. (2008). The seeds of time: why classroom dialogue needs a temporal analysis. *Journal of the Learning Sciences*, 17(1), 33-59.
- Mortimer, E. F., & Scott, P. (2003). *Meaning making in science classrooms*. Milton Keynes: Open University Press.
- Müllner, D. (2011). Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*.
- Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.
- Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-supported Collaborative Learning*, 3(3), 237-271.

Scheihing, E., Vernier, M., Born, J., Guerra, J., & Cárcamo, L. (2016). Classifying discourse in a CSCL platform to evaluate correlations with teacher participation and progress. *arXiv preprint arXiv:1605.07268*.

Wang, Z., Pan, X., Miller, K. F., & Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Computers & Education, 78*, 115-123.