



**HAL**  
open science

# Teachers' Knowledge in Informatics-Exploring Educational Robotics Resources Through the Lens of Textual Data Analysis

Gabriel Parriaux, Christophe Reffay, Béatrice Drot-Delange, Mehdi Khaneboubi

## ► To cite this version:

Gabriel Parriaux, Christophe Reffay, Béatrice Drot-Delange, Mehdi Khaneboubi. Teachers' Knowledge in Informatics-Exploring Educational Robotics Resources Through the Lens of Textual Data Analysis. Jean-Philippe Pellet; Gabriel Parriaux. Informatics in Schools. Beyond Bits and Bytes: Nurturing Informatics Intelligence in Education, 14296, Springer Nature Switzerland, pp.126-138, 2023, Lecture Notes in Computer Science, 978-3-031-44899-7. 10.1007/978-3-031-44900-0\_10. hal-04384868

**HAL Id: hal-04384868**

**<https://uca.hal.science/hal-04384868>**




Submitted on 10 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Teachers' Knowledge in Informatics—Exploring Educational Robotics Resources Through the Lens of Textual Data Analysis

Gabriel Parriaux<sup>1</sup> , Christophe Reffay<sup>2</sup> , Béatrice Drot-Delange<sup>3</sup> ,  
and Mehdi Khaneboubi<sup>4</sup>

<sup>1</sup> University of Teacher Education, Lausanne, Switzerland  
`gabriel.parriaux@hepl.ch`

<sup>2</sup> University Bourgogne Franche-Comté, Besançon, France  
`christophe.reffay@univ-fcomte.fr`

<sup>3</sup> University of Clermont Auvergne, Clermont-Ferrand, France  
`beatrice.drot-delange@uca.fr`

<sup>4</sup> CY Cergy Paris University, Paris, France  
`mehdi.khaneboubi@u-cergy.fr`

**Abstract.** In this research, we are interested in the knowledge of primary and secondary teachers to teach informatics. Using pedagogical resources produced by them as a trace of their enacted Pedagogical Content Knowledge (ePCK), we perform Textual Data Analysis and Clustering to discover the topics they write about. Focusing on resources in educational robotics, we show that lexicon used by teachers is different depending on the robot they use. Reinert's clustering associates each robot with a separate cluster and a specific vocabulary. Multiple Correspondence Analysis (MCA) shows an opposition between lexicon found in resources using event-driven robots (Thymio and Ozobot) and in resources using sequential robots (Beebot and Bluebot). Event-driven robots tend to be more related with events and behavior notions, as sequential robots tend to be more related with first manipulations of an object and programming notions.

**Keywords:** PCK in informatics · enacted PCK · educational robotics · programming paradigms · Reinert's clustering · Textual Data Analysis · teacher education

## 1 Introduction and Context

This research constitutes the third part of a research project on teachers' PCK in informatics that has been conducted between 2021 and 2023. In the first article [4], we investigated educational resources produced by pre-service primary teachers to teach programming in France and Switzerland. The results showed

that the progression of activities and the choice of teaching methods were not completely thought through by future teachers.

In the second article [13], we focused on a subset of our initial educational resources including robots. We added to it a set of online resources on the same topic from more experienced teachers. We wondered whether the computing notions involved in the resources were the same in the case of event-driven and sequential robots. Having extracted from the corpus the lexicon that we identified as “Computer Science (CS)”, the results showed that it was significantly different according to the type of robot and the experience of the teacher.

In this third step, we want to explore further the potential of Textual Data Analysis techniques in the field of computing education. Unlike our previous research in which we focused on a portion of the lexicon that we manually extracted from the corpus, what would happen if we considered *the entire lexicon* and base our analyses on it? We propose to perform a new investigation on the same corpus of educational robotics resources, but this time using another method of Textual Data Analysis: Reinert’s clustering. Reinert’s clustering is an unsupervised divisive clustering technique [14] that, when applied to a corpus of texts, is able to extract some of their main topics, giving an idea of the mental universe embraced by their authors.

Our research questions are

- RQ1 *What are the main themes that emerge from the textual analysis of educational resources for CS?*
- RQ2 *What kind of relation can be established between the lexicon and the type of robot used in the resources?*
- RQ3 *Based on the analysis of the lexicon, what other relations can be seen between the clusters, grade levels, programming paradigms and teachers’ expertise?*

## 2 Related Work

Several models have been proposed to represent the knowledge of teachers. Among them, Pedagogical Content Knowledge (PCK) presented by Shulman in 1986, is one of the most fruitful. Built in reaction to a context in the US in which specific knowledge of the subject matter seemed to be left apart in teacher education, PCK is defined by Shulman as “a pedagogical knowledge that goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching” [16, p. 9].

The original model was enriched by Magnusson, Krajcik and Borko for the teaching of science [9] who considered 5 components of PCK: orientations toward science teaching, knowledge of the curriculum, knowledge of students’ understanding, knowledge of assessment and knowledge of instructional strategies.

A community of researchers in science education was built around the PCK model and proposed in a second iteration the Refined Consensus Model (RCM) of PCK in 2017 [2], which presents three different domains of PCK: collective PCK (cPCK), personal PCK (pPCK) and enacted PCK (ePCK).

Enacted PCK (ePCK) is the teacher’s knowledge in action, “the specific knowledge and skills utilized by an individual teacher in a particular setting, with a particular student or group of students, with a goal for those students to learn a particular concept, collection of concepts, or a particular aspect of the discipline” [2, pp. 83–84]. It is not restricted to the moment of teaching in class with the students and also includes the planning and reflecting on the instruction by the teacher. According to Henze and Barendsen [6], ePCK is the part of the personal PCK that is active at a certain moment during teaching. Its dynamic aspect makes it not so easy to elicit and observe. Researchers interested in studying the enacted PCK of teachers can rely on classroom observations, but can also consider a pedagogical resource prepared by a teacher as a trace of its enacted PCK. This is the choice we made in studying pedagogical resources of school teachers who teach programming, they constitute a trace of their planning activity, a part of their ePCK.

Despite the fact that Textual Data Analysis is not a mainstream method in the field of education [5], we could identify some research where computational methods were applied to text in the context of education. Some of them have methodological aims and explicitly attempt to show that it’s possible to use computational techniques to analyze textual data in education [1, 15]. Some research use written data, for example open-ended questions in questionnaires to teachers or students [11, 18], curricula [11] or teacher resources [1]. Other research use transcription of oral data, for example interviews [15] or discourse in class [8, 10].

Few researchers employ Reinert’s clustering in the context of education [5], but there exists a good example related to science education where Reinert’s clustering is used to analyze teachers’ representations of the investigation procedure, confronting clustering of an open-ended question in a teachers’ survey and clustering of the content of the curriculum in science [11].

In the field of computing education, we found even less research using Textual Data Analysis, but we can mention one study using topic modeling to analyze students reviews on computer science MOOCs [3] and another study using Clustering technique of CS1 students’ programming codes to identify a group of students with difficulties [17].

### 3 Data Collection and Methodology

Data collected for our research has two origins: the first part consists of 59 educational resources produced by pre-service teachers during workshops about the teaching of informatics in France and Switzerland. Those resources, dedicated to activities in educational robotics, were part of a more important corpus composed during the first phase of our research. More details are given about them in [4]. The second part of the data consists of 61 online resources on the same topic from more experienced teachers.

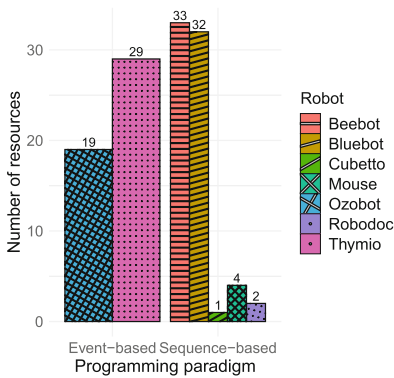
Our resources consist of pedagogical scenarios serving as preparation for a school lesson with one or more activities: teachers describe the learning objectives of the lesson, the roles of teachers and pupils, a schedule and the material

needed. Some resources also contain feedback about the difficulties encountered, either anticipated or experienced. For the pre-service teachers, the resources produced were used to validate one of their academic courses and their form differed depending on their university: a written scenario in a PDF document of 2 to 15 pages for some of them; a PowerPoint presentation explaining the scenario in the context of an oral exam; a video presenting the scenario in the same context, from which we extracted and transcribed the audio.

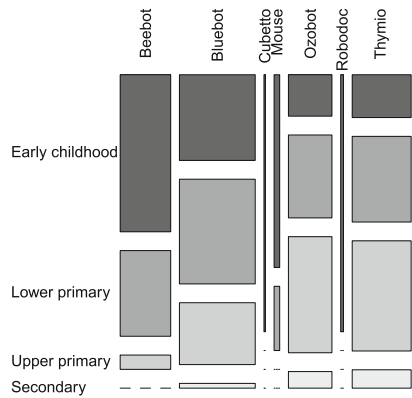
In order to enrich our corpus, we decided to augment it with 61 pedagogical scenarios available online. We followed a formalized process to use a search engine with keywords that would let us discover different resources for each robot, filtered the first results to download only resources that we could clearly identify as pedagogical scenarios. The documents were written documents in PDF between 4 and 20 pages. We have less information about the authors of these online documents than we have for our pre-service teachers, but we categorize them as “experienced” because we could see going through the documents that they were often well elaborated and that some of them were clearly affiliated to an educational academy.

We are conscious that the format of our resources is quite different depending on the context and it can be considered as a limitation of our research. But at the same time, the main goal of the documents was the same: presenting the pedagogical scenario of a school activity with robots. They are all textual and contain a lexicon that we can valuably analyze with Textual Data Analysis methods.

Finally, our corpus is composed of a total of 120 educational resources in French describing activities in educational robotics created either by novices or by experienced teachers.



**Fig. 1.** Resources classified by type of robot and programming paradigm.



**Fig. 2.** Resources classified by type of robot and grade level.

Figure 1 shows the number of resources per robot in our corpus and their repartition between event-based and sequential. Event-based robots have sensors and actuators, they are mostly programmed in an event-based mode: various events are associated with specific behaviors; sequential robots do not have information from their environment and are programmed in a sequential mode: a sequence of instructions that are executed from a beginning to an end. Figure 2 presents the repartition of resources by robots and by grade level. We see that there are very few resources related to secondary education.

Processing of the data, pretreatment and analyses are performed in R [12]. The first step consists of extracting all the text from the resources. Having classified them according to categorical variables like the type of robot (sequential or event-driven), teachers' experience (novices or experienced), the grade levels and so on, we add those as metadata about the documents.

Before pretreatment, the whole corpus is composed of 248,813 occurrences and a vocabulary of 54,819 different words.

In order to perform Reinert's clustering, the corpus is split into segments of around 40 words, trying to maintain together words in complete sentences. We obtain 4,784 segments of text.

Then, we perform the lemmatization of the text, an operation that replaces every inflected form of a word into its unique normalized form. In the later process, forms that appear too few times in the corpus will be ignored in the analysis. Without lemmatization there is the risk to lose certain words where every inflected form would be counted as a different word, which is especially true for verbs in French. We use SPACYR, a wrapper for SPACY Python's library in R, with `fr_dep_news_trf`, a French transformer pipeline as a model, to process it.

Tokenization is then performed, which means the transformation of continuous text of the segments into individual tokens or words. The whole text is split into single tokens, with the space character as a natural separator. Punctuation, symbols and URLs are removed. We remove stopwords so that we can focus on notions which appear mostly through substantives or verbs. We use a custom list of 449 stopwords, as some of the common lists provided in R packages were rather too short (157 words for the standard) or too long (687 words for the list called `stopwords-iso`).

Finally, we compose an exclusion list along the classification process to remove words that are appearing in clusters and that make no sense for the analysis (for example abbreviations for school years, name of institutions or schools). It is composed of 116 words.

At the end of this pretreatment process, we obtain a matrix called *document-feature matrix* crossing 2,599 words and 4,784 segments which serves as a basis for the Reinert's clustering.

Reinert's clustering is an unsupervised divisive clustering that gets applied to the document-feature matrix. It means that it performs a bipartition of the segments iteratively. Segments are divided into two groups with the aim of building clusters that are as homogeneous as possible, keeping together the most similar

segments, and as heterogeneous as possible between them, with groups of segments as different as possible. Similarity is established through the presence of similar words in the segments using a Pearson’s chi-squared test to compute the distance between the two clusters.

Reinert’s method proposes a double clustering that crosses the results between two simple clustering performed with different minimal sizes of segments to obtain more robust classes. Trying various settings, it is up to the researcher to determine the number of clusters which seems the most relevant in terms of interpretation.

Having compared different analyzes, we decide to perform a double Reinert’s clustering with a `min_segment_size = 10` for the first clustering and a `min_segment_size = 15` for the second. We keep eleven clusters. As is the case with double classification, no dendrogram and no visible hierarchy between the clusters is produced. We obtain a graph of the eleven clusters with a list of words that are strongly associated with them.

## 4 Results and Discussion

### 4.1 Results from Reinert’s Clustering

A plot of the classification with the 11 clusters is displayed in Fig. 3, showing for each cluster the 30 tokens most significantly associated with. The number of unclassified segments is 821. It’s an important number compared with the total of 4,784 segments (around 17%), but even though we decide not to force the classification of the remaining segments in the existing clusters, as this is reputed to potentially alter the homogeneity of the clusters.

Here is a list of the main clusters, with a title and some of the words that appear as strongly associated to—or overrepresented in—each cluster. An interface in the RAINETTE package lets us access the segments containing the words of a cluster easily, giving the necessary context to let us interpret better its meaning. The title is given according to our interpretation. Clusters 2, 3, 4 and 9 are small clusters with less than 200 segments and can be ignored.

- Cluster 1—Goals: *domain, grade, language, competence, objective, pedagogical, learning* (1,029 segments)
- Cluster 5—Movement: *erase, move forward, memory, left, right, turn, rotate, beebot* (203 segments)
- Cluster 6—Path: *path, route, cover, draw, time* (230 segments)
- Cluster 7—Sharing: *institutionalization, pupil, groups, collective, sharing* (531 segments)
- Cluster 8—Programming: *instruction, square, start, program, programming, check, bug, bluebot* (537 segments)
- Cluster 10—Events: *sensor, detect, behavior, object, event, explore, thymio* (543 segments)
- Cluster 11—Dancing: *movements, dance, tablet, code, app, ozobot* (515 segments)

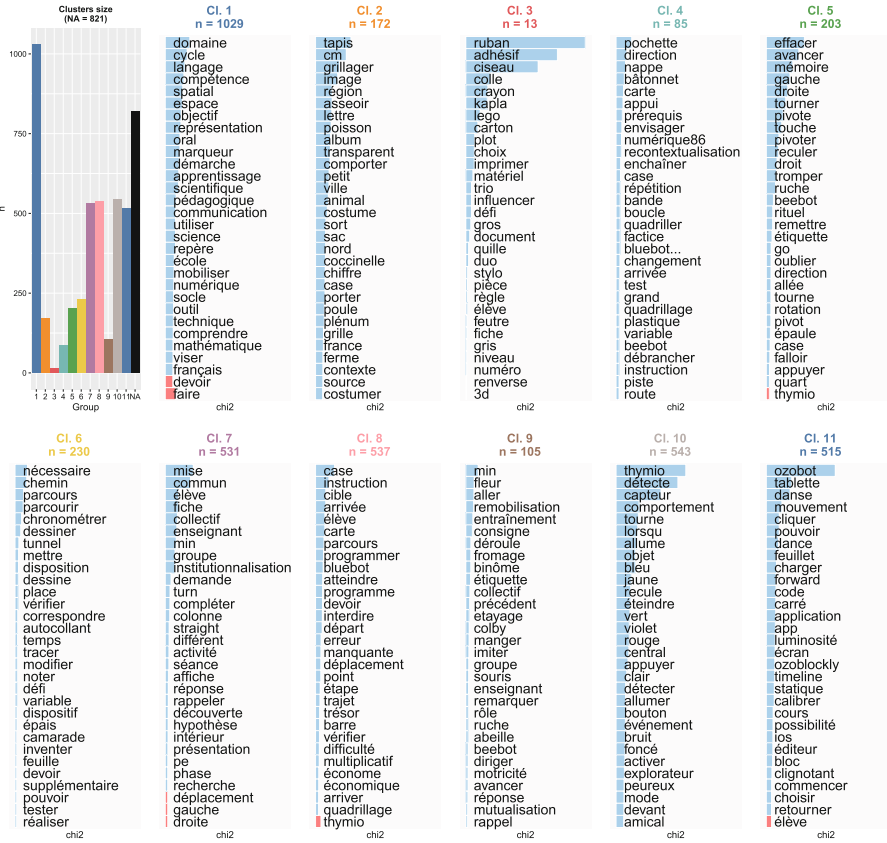


Fig. 3. Eleven clusters produced by Reinert’s clustering. Lexicon displayed in French. Image produced in R with RAINETTE package.

### 4.2 Results from Multiple Correspondence Analysis

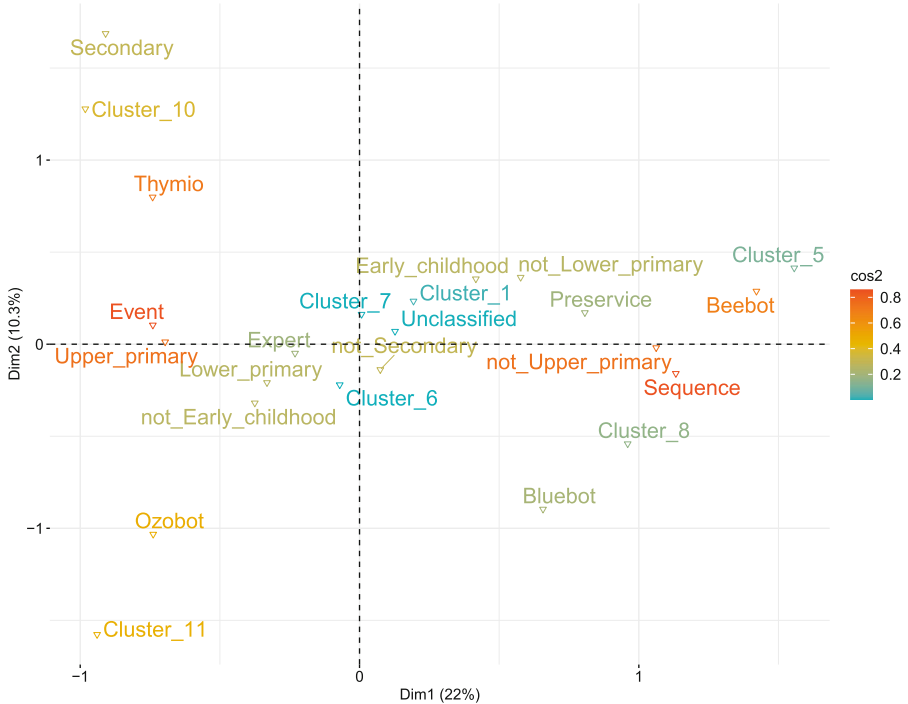
Once Reinert’s clustering has been performed, we execute a Multiple Correspondence Analysis (MCA) to explore the relations between our clusters and the different variables associated to the pedagogical resources in our corpus.

A contingency table is constituted with eight categorical variables in columns and the 4,784 segments of texts as individuals in rows. The categorical variables associated with the segments are: *cluster* in which the segment is classified, *programming paradigm* of robots, *name* of robots, *level of expertise* of teachers and four *grade levels* of pupils concerned by the activities (early childhood, lower primary, upper primary and secondary). As our resources sometimes cover more than one grade level, we had to define grade levels as four separate variables. To gain in readability and avoid unnecessary information, segments concerning robots that rarely appear in our corpus are removed. Segments classified in the



small clusters 2, 3, 4 and 9 are also removed. We end with a contingency table of 4,216 rows and 8 columns.

The result of MCA can be seen in Fig. 4 for axes 1 and 2. Fifteen factors or axes are produced. First axis retains 22% of inertia and second axis 10.3%.



**Fig. 4.** Multiple Correspondence Analysis (MCA) showing clusters and variables related to documents, axes 1 and 2. Image produced in R with EXPLOR package.

### 4.3 Discussion

Looking at the results from Reinert's clustering, we observe that out of the seven main clusters created, five clusters are oriented toward CS content for primary and secondary (5, 6, 8, 10 and 11) and contain lexicon about movements on a grid, following paths, programming, events and machine; one cluster is related to general goals of CS teaching (1) and one is oriented toward pedagogical practices (7), with lexicon that has to do with collaborative learning and sharing knowledge with the whole class.

This let us answer RQ1: *What are the main themes that emerge from the textual analysis of educational resources for CS?* Performing clustering of our corpus, we can extract seven main clusters or topics. According to the PCK

model of Magnusson, five clusters (5, 6, 8, 10 and 11) are related to *knowledge of the curriculum*, one (1) is related to *orientations toward CS teaching* and one (7) is related to *knowledge of instructional strategies*.

We also have elements to answer RQ2, *What kind of relation can be established between the lexicon and the type of robot used in educational resources for CS?* We were able to identify the lexicon related to each of the robot present in our corpus: four clusters are significantly linked to one of the robots, since they contain the name of the robot among their overrepresented words. Cluster 5—*Manipulation* is related to *Beebot* and contains words having to do with the movement and first manipulation of an object, which could come close to direct manipulation or direct control according to Kalaš [7]; cluster 8—*Programming* is connected with *Bluebot*, with words rather related to programming notions and activities described in a way that could point to a computational control of an object [7], so in some way a more advanced phase in the process of learning CS compared to cluster 5; cluster 10—*Events* is associated with *Thymio* and contains words rather related to machines, events and behavior; and finally cluster 11—*Dancing* is associated with *Ozobot*, with words rather related to the idea of a dance. As vocabulary conveys ideas and notions, then our analysis says something about the notions used by the authors of our resources in relation with each robot.

Maybe these associations between robots and topics can appear as trivial, especially for the people who have some experience with them. But we think that it's interesting to consider that this classification did not emerge from a pre-established conceptual framework that we would have applied to the corpus; rather it emerged completely inductively from the data itself, just from a simple Textual Data Analysis where we *count words*.

Further on, this technique has revealed a difference between Beebot and Bluebot that we would not have thought about. Those robots are very similar in terms of affordance (Bluebot seems to be a kind of updated version of Beebot). However, in our resources, we see that the vocabulary associated with the two robots is somehow different, which shows that teachers connected different notions to the activities they presented for Beebot and Bluebot. Segments of text related to Beebot describe rather the first manipulations of a robot and those related to Bluebot a more advanced level of programming.

Then, MCA on Fig. 4 offers an interesting view on our data as it represents the attractions and oppositions between our variables: clusters, robots, programming paradigms, teachers' expertise and grade levels. The first observation we make is the proximity between the four clusters and the four robots to which they were associated through Reinert's clustering: cluster 10 is close to Thymio, cluster 5 is close to Beebot, cluster 11 is in the vicinity of Ozobot and cluster 8 is not far from Bluebot. We see it as a confirmation that there exists an association between these four clusters, the notions they convey, and the robots to which they are associated.

To answer RQ3, *Based on the analysis of the lexicon, what other relations can be seen between the clusters, grade levels, programming paradigms and teachers'*

*expertise?*, we can interpret the meaning of the axes of the MCA. We see on the first axis an opposition between *event-driven robots* on the left (Thymio, Ozobot) and *sequential robots* on the right (Beebot, Bluebot). This is consistent with the results that we had in our previous research and it shows that, considering the entire lexicon of our resources, there is an opposition in terms of vocabulary used by teachers between resources with activities on event-based robots and sequential robots. The second axis is a bit more difficult to interpret, as it opposes on one side Thymio, its cluster 10, secondary, Beebot and early childhood to Ozobot, its cluster 11, lower primary, Bluebot and its cluster 8. So there seems to be an opposition between resources for secondary or early childhood (Thymio could be for both, Beebot certainly for early childhood) and resources for primary (Ozobot and Bluebot).

It's interesting to note that teachers' expertise doesn't stand out in the MCA, as its values are situated close to the center of the plan. As we tend to consider the PCK of novices to be quite different from the PCK of more experienced teachers [9], it's a bit surprising to see that it's not a factor of differentiation in our corpus. We also note that the clusters not clearly associated with a robot (1, 6 and 7) do not play a distinctive role in the MCA.

To conclude this discussion, we think that applying computational methods to the analysis of textual data in the field of CS education is a valuable approach, because it lets us go through an important quantity of texts that would be time consuming to manage manually. Reinert's clustering helps us understand, through the clusters it produces, what topics are constituting the mental universe of teachers. It is inductive and these topics are extracted from the data itself, without having a predefined model for the interpretation. Associated with Multiple Correspondence Analysis, it gives the possibility to interpret the relation between specific variables on the lexicon and the clusters produced.

The knowledge of teachers is not an easy thing to observe: as we cannot reach it directly, we can only observe its manifestation in certain contexts or objects. Pedagogical resources produced by teachers are one of them. They constitute a trace of their planning activity and, as such, of their enacted PCK. Performing a textual analysis of the content of such resources gives access to the individual words that make them up. As knowledge is related to notions, the words that appear through textual analysis of resources say something about the knowledge of their authors.

## 5 Conclusion

In this research, we apply Reinert's clustering and Multiple Correspondence Analysis (MCA) to the text of pedagogical resources in the field of educational robotics considered as traces of the enacted PCK of teachers. Through the analysis of the resulting clusters and the words significantly related to them, we are able to show the main topics that teachers write about when they create resources. Here is the way we can answer our three research questions.

For RQ1 about the main themes that emerge from the textual analysis of our educational resources, Reinert's clustering creates seven main clusters representing five topics that we would classify as *knowledge of the curriculum* according to Magnusson's model of PCK, one topic related with *orientations toward CS teaching* and one topic that would be part of the *knowledge of instructional strategies*.

Regarding RQ2 about the kind of relation that can be established between the lexicon and the type of robot used in the resources, Reinert's clustering and MCA let us characterize it and say that Beebot is associated with the first manipulation of an object. Bluebot is associated with programming. Thymio is associated with events and behavior. Ozobot is associated with a dance.

Concerning RQ3 about the other relations that can be seen between the clusters, grade levels, programming paradigms and teachers' expertise, MCA shows an opposition between event-driven and sequential robots.

As a limitation for our research, we only investigated one domain of PCK: enacted PCK, and only through the analysis of the trace of teachers' planning activity. For a broader view on teachers' PCK, it would be interesting to investigate the whole pedagogical cycle (planning – enactment – reflection) and the relations that can appear between personal PCK (pPCK) and enacted PCK (ePCK) in these contexts [6].

For this reason, we wish to extend our research, using the methods of Textual Data Analysis, to other kinds of context as teachers' interviews and classroom transcripts.

## References

1. Aguilar, J., Salazar, C., Velasco, H., Monsalve-Pulido, J., Montoya, E.: Comparison and evaluation of different methods for the feature extraction from educational contents. *Computation* **8**(2), 30 (2020)
2. Carlson, J., et al.: The refined consensus model of pedagogical content knowledge in science education. In: Hume, A., Cooper, R., Borowski, A. (eds.) *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, pp. 77–94. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-13-5898-2\\_2](https://doi.org/10.1007/978-981-13-5898-2_2)
3. Chen, X., Zou, D., Xie, H., Cheng, G.: What are MOOCs learners' concerns? text analysis of reviews for computer science courses. In: Nah, Y., Kim, C., Kim, S.H., Moon, Y.-S., Whang, S.E. (eds.) *DASFAA 2020. LNCS*, vol. 12115, pp. 73–79. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-59413-8\\_6](https://doi.org/10.1007/978-3-030-59413-8_6)
4. Drot-Delange, B., Parriaux, G., Reffay, C.: Futurs enseignants de l'école primaire: connaissances des stratégies d'enseignement, curriculaires et disciplinaires pour l'enseignement de la programmation. *RDST. Recherches en didactique des sciences et des technologies* (23), 55–76 (2021)
5. Emprin, F.: Les apports d'une analyse statistique des données textuelles pour les recherches en didactique : l'exemple de la méthode Reinert. *Annales de Didactique et de Sciences Cognitives. Revue internationale de didactique des mathématiques* (23), 179–200 (2018)
6. Henze, I., Barendsen, E.: Unravelling student science teachers' pPCK development and the influence of personal factors using authentic data sources. In: Hume, A.,

- Cooper, R., Borowski, A. (eds.) *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*, pp. 203–223. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-13-5898-2\\_9](https://doi.org/10.1007/978-981-13-5898-2_9)
7. Kalaš, I.: Programming in lower primary years: design principles and powerful ideas. In: Dagienė, V., Jasutė, E. (eds.) *Constructionism 2018 – Constructionism, Computational Thinking and Educational Innovation: Conference Proceedings (2018)*
  8. Wessel, L.: Vocabulary in learning processes towards conceptual understanding of equivalent fractions—specifying students' language demands on the basis of lexical trace analyses. *Math. Educ. Res. J.* **32**(4), 653–681 (2020)
  9. Magnusson, S., Krajcik, J., Borko, H.: Nature, sources, and development of pedagogical content knowledge for science teaching. In: Gess-Newsome, J., Lederman, N.G. (eds.) *Examining Pedagogical Content Knowledge*, vol. 6, pp. 95–132. Springer, Dordrecht (1999). [https://doi.org/10.1007/0-306-47217-1\\_4](https://doi.org/10.1007/0-306-47217-1_4)
  10. Mercer, N.: The analysis of classroom talk: methods and methodologies. *Br. J. Educ. Psychol.* **80**(1), 1–14 (2010). <https://onlinelibrary.wiley.com/doi/pdf/10.1348/000709909X479853>
  11. Prieur, M., Monod-Ansaldi, R., Fontanieu, V.: Réception des démarches d'investigation prescrites par les enseignants de sciences et de technologie. *RDST. Recherches en didactique des sciences et des technologies* (7) (2013)
  12. R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2022)
  13. Refay, C., Parriaux, G., Drot-Delange, B., Khaneboubi, M.: Robotics in primary education: a lexical analysis of teachers' resources across robots. In Keane, T., Lewin, C., Brinda, T., Bottino, R., (eds.) *Towards a Collaborative Society Through Creative Learning: IFIP World Conference on Computers in Education, WCCE 2022, Hiroshima, Japan, 20–24 August 2022, Revised Selected Papers*. Springer International Publishing (2023). <https://link.springer.com/book/9783031433924>
  14. Reinert, A.: Une méthode de classification descendante hiérarchique: application à l'analyse lexicale par contexte. *Cahiers de l'Analyse des Données* **8**(2), 187–198 (1983)
  15. Sherin, B.: A computational study of commonsense science: an exploration in the automated analysis of clinical interview data. *J. Learn. Sci.* **22**(4), 600–638 (2013)
  16. Shulman, L.S.: Those who understand: knowledge growth in teaching. *Educ. Res.* **15**(2), 4–14 (1986)
  17. Silva, D.B., Silla, C.N.: Evaluation of students programming skills on a computer programming course with a hierarchical clustering algorithm. In: *2020 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. IEEE, Uppsala, Sweden (2020)
  18. Sobczak, A., Debucquet, G., Havard, C.: The impact of higher education on students' and young managers' perception of companies and CSR: an exploratory analysis. *Corp. Governance: Int. J. Bus. Soc.* **6**(4), 463–474 (2006)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

