Using Network Text Analysis to Characterise Teachers' and Students' Conceptualisations in Science Domains

H.U. Hoppe[1], M. Erkens[1], G. Clough[2], O. Daems[1], A. Adams[2]
(1) Rhein-Ruhr-Institut für angewandte Systeminnovation (RIAS),
Duisburg, Germany
(2) The Open University, Institute of Educational Technology,
Milton Keynes, UK

Abstract. The ongoing EU project JuxtaLearn aims at facilitating the acquisition of science concepts through videos, especially also through creation of videos on the part of the learners. Learning Analytics techniques are used to extract and represent teachers' and students' concepts manifested in interactive workshops based on textual artefacts. First results of using the network text analysis method are available. This approach will be further used to pinpoint the teachers' specific perspectives and views and possibly the development of their conceptualisations over time.

Keywords: science learning, network text analysis, threshold concepts

1 Background

Focusing on "performance" as a learning mechanism, the recently started EU project JuxtaLearn aims at provoking student curiosity in science and technology through creative film making and editing activities. Available public video resources will be analysed for their potential to facilitate students' creative inspiration and further conceptual insight and understanding.

From a science education point of view, teaching and learning support in JuxtaLearn is guided by threshold concepts [1]. Previously identified threshold concepts are the basis for reinforcing deeper understanding and further creative production through scaffolding reflections focused on essential elements. To identify such concepts and to explore how these are understood and appropriated by teachers and students, a series of face-to-face workshops are being conducted. Learning Analytics techniques are used to extract structured representations of the underlying conceptual relations.

2 Approach

The ongoing series of teacher-student workshops in JuxtaLearn aims at envisioning pedagogical scenarios around certain scientific threshold concepts. In addition to two initial workshops with science teachers a third workshops also involved a group of 6

A-level students. Because the misconceptions surrounding thresh-old concepts are difficult to pin down, the workshop was structured to elicit a deeper understanding of the gaps in the students' knowledge through a role reversal in which the students taught the teachers. Textual documents produced in these workshops (transcripts and summaries) have been analysed using the AutoMap/ORA toolset for Network Text Analysis [2,3].

As a result of this analysis process, we have generated multimodal networks of categorised concepts. Categories are, e.g., pedagogical concepts, domain concepts, tools roles and actors. Based on first examples we claim that such networks can represent and characterize the specific foci of the workshops. This approach will be further used to pinpoint the teachers' specific perspectives and views and possibly the development of their conceptualisations over time.

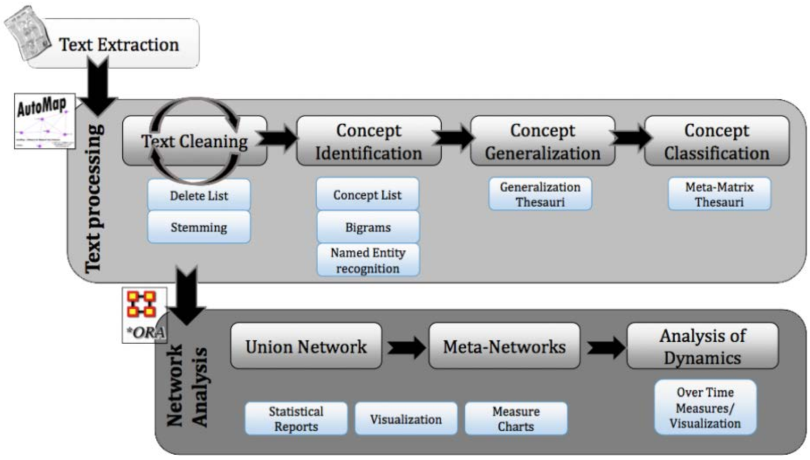

Figure 1: Analysis workflow

Figure one shows the process or workflow of Network Text Analysis. It is divided into three main parts: data selection and extraction, (pre-) processing and network analysis.

2.1 Data Selection/Extraction

We have extracted the textual data from workshop transcriptions of the audio recordings. The input documents comprised students' preparation notes and conversation transcripts during three role reversal lessons in the fields of chemistry, biology and physics and ensuing debriefings. The extraction was performed on all textual documents from one workshop at once as well as on the separate lessons.

2.2 Text Processing

The AutoMap (pre-)processing functions include text cleaning as well as identification, generalisation and classification of relevant concepts. The cleaning step includes the removal of non-relevant “stop words” (articles, auxiliary verbs etc.), a *kstemmer* to reduce words to their root stem and the detection of relevant concepts, e.g. by analysing the word frequency. The classification and generalisation steps assign different concept representations to the respective key concepts using a generalisation thesaurus. Connections (edges) are established if the corresponding terms appear within a sliding window of a given length that is run over the whole text.

Apart from *roles*, *general concepts* and *tools & technologies*, we have identified *agents* and *knowledge* as the most relevant concept categories. All these categories have been represented in an ontology-based meta-thesaurus. The *Agent* category represents all acting persons in the lessons; teachers have been labelled as T1 to T6, the researcher staff as R1 to R4 and students as S1 to S6. The *Knowledge* category represents discipline-specific topics associated with the lesson subjects.

2.3 Network Analysis

As a result of this analysis process, multimodal networks of categorised concepts are generated. Based on this first example, we claim that these networks can represent and characterise the specific foci of the workshops. Networks and derived measures are graphically represented using ORA.

3 First Results

Based on the complete workshop transcript, the resulting two-mode network (actors x knowledge) contains 16 *agent* nodes (black circular nodes) and 71 *knowledge* nodes (rectangular). In Figure 2, every subject area and corresponding sub-activity in the workshop (chemistry, biology and physics) forms a cohesive cluster in the overall network.

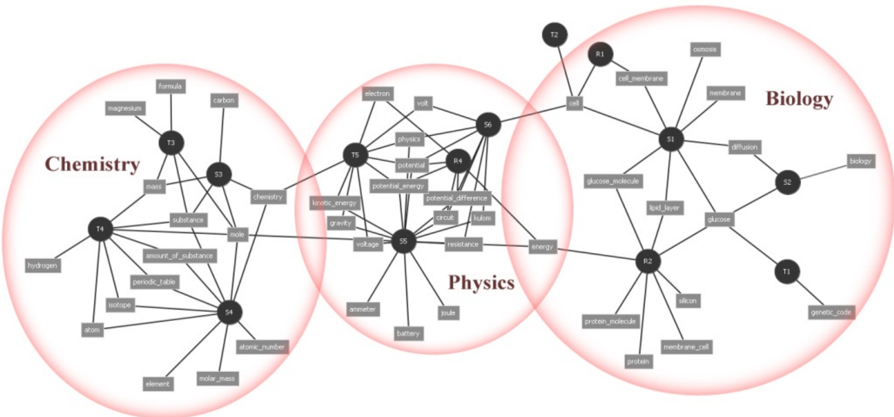

Figure 2: Example network (from complete workshop transcript)

The number of connections between one actor and surrounding topics (also called "degree") indicates the thematic richness of this actor's contributions. In this sense, S1, S4, S5 and S6 score better than S2 and S3. Also, we see that teachers were not much involved in the discussion in the biology domain (whereas researcher R2 was).

Another relevant structural property of the extracted network is the identification of concepts that bridge over between other concepts or between areas of discourse (here: the domains). A network measure that captures this bridging function is "betweenness centrality" (cf. [4] as a standard reference). Figure 3 shows the top 4 knowledge items ranked according to their betweenness centrality.

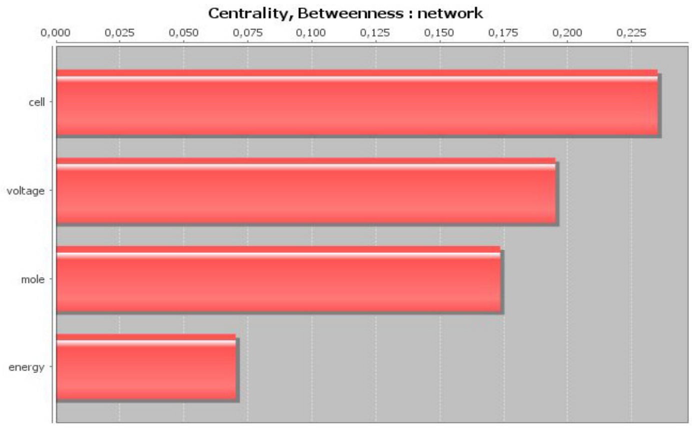

Figure 3: Knowledge items with highest betweenness centralities (from overall network)

Since the concepts *cell*, *voltage*, *mole* and *energy* build bridges between and within these clusters they seem to be of special interest. Notably, two of them, *cell* and *moles*, had already been identified as stumbling blocks in the prior identification of threshold concepts. A third stumbling block, potential difference seems to play a less central role as a connector between other concepts.

4 Outlook

From our examples, we see evidence for the claim that using network text analysis to extract relations between categorised items from textual artefacts can reveal underlying conceptualisations by humans in a meaningful way. In our future work we plan to elaborate on the following extensions and applications:

- The use of pencast recordings (using a LiveScribe[1] smartpen) as input. Here the transcription could be automatically generated.

[1] www.livescribe.com

- The comparative characterization of workshops based on extracted networks. Here, the question is if the networks capture relevant differences.
- Using networks for identification of misconceptions (probably in combination with other methods).
- Using networks as material for reflection with teachers and/or students.

References

1. Meyer, J.H.F. and Land, R. (2003). Threshold concepts and troublesome knowledge: linkages to ways of thinking and practising, In: Rust, C. (ed.), Improving Student Learning - Theory and Practice Ten Years On. Oxford: Oxford Centre for Staff and Learning Development (OCSLD), pp 412-424.
2. Carley, K. M. & Columbus, D. (2012). Basic Lessons in ORA and AutoMap 2012. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-12-107.*
3. Diesner, J. & Carley, K. M. (2004). Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a novel method for Network Text Analysis. *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations.* Harrisburg, PA: Idea Group Publishing.
4. Wasserman, S., & Faust, K. (1994). *Social Networks Analysis: Methods and Applications*. Cambridge: Cambridge University Press.