Gaining Insight by Transforming Between Temporal Representations of Human Interaction

Kristine Lund

UMR 5191 ICAR, CNRS, Ecole Normale Supérieure
University of Lyon, France
kristine.lund@ens-lyon.fr

Matthieu Quignard

UMR 5191 ICAR, CNRS, Ecole Normale Supérieure
University of Lyon, France

David Williamson Shaffer

University of Wisconsin – Madison

ABSTRACT: Recordings of human interaction data can be organized into temporal representations with different affordances. We use audio data of a learning-related discussion analyzed for its low-level emotional indicators and divided into four phases, each characterized by an overarching emotion. After arguing for the relevance of emotion to learning, we examine this original analysis with the help of three different representations, transforming the data between them in order to connect micro- and macro-levels of analysis and give meaning to these connections. The first is a FRIEZE representation showing the temporal distribution of the low-level indicators of emotion as well as the phases. The second is an epistemic network analysis with an aggregated representation that shows how the pattern of associations among indicators of emotion differs between phases. The third is a transcription of the original data that re-anchors the aggregation back into the temporal interaction, giving it meaning. This is a methods paper, and if the findings are not specifically focused on measuring learning, the data do concern a student narrative of interactions with her teacher. More importantly, the stage is set for giving meaning to micro- and macro-connections in pedagogical contexts, with a view to automated analyses.

Keywords: Temporal representations, emotion, collaborative knowledge construction, connecting micro- and macro-analyses

NOTES FOR PRACTICE

- It's a common approach to analyse and visualize interaction data over time. However, relationships between features in these interactions have not been well measured or modelled, especially between levels of interaction over which features interact (for example, between an initiation-response-feedback loop and a whole conversation, or between phases in a conversation that have been characterized in a particular way and low-level descriptions of how language is co-constructed).
- This methods paper demonstrates the significance of this segmentation issue by showing how different pairs of emotion indicators characterize different phases of an interaction.

- Our method can show the relationship between how students interact together and the objectives of a series of pedagogical sequences. In other words, we can evaluate the extent to which the goals of each sequence are being met by the details of what students accomplish by talking and carrying out actions together.

- In terms of implementation, although our emotion indicators were analysed by hand, some of them may be coded automatically. Ultimately, automatic tagging of a corpus will depend on the nature of the phenomena the method focuses on.

1 MANIPULATING TEMPORAL REPRESENTATIONS TO ANALYZE HUMAN INTERACTION

Researchers form their analyses of human interaction based on recordings of action, and, critically, on different representations that both illustrate the data in these recordings and generate new knowledge regarding them (Lund & Suthers, 2013; Dyke, Lund, Suthers, & Teplovs, 2013). Here, we build on work by Dyke, Lund, and Girardot (2009) who argue that there is a cyclical process in which a research question guides the capture of data, which is represented in some way, and then interpreted, leading to the creation of a new representation, and beginning the cycle again (cf. Figure 1)
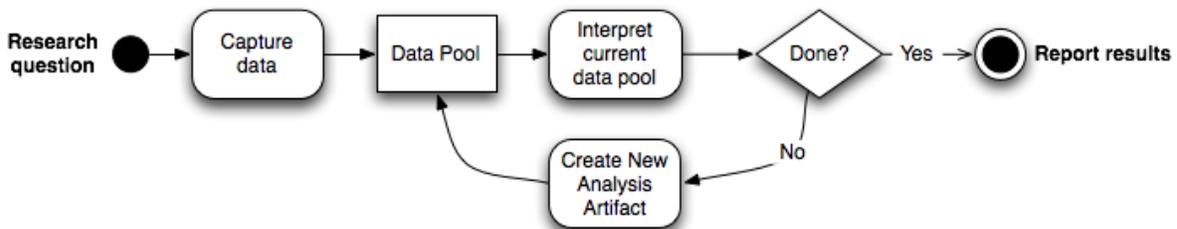

Figure 1. The cyclical process of data capture and analysis (Dyke et al., 2009).

We extend this work by looking at the interaction of three different forms of temporal representations: 1) the original transcription of the data; 2) a representation that highlights the co-occurrence of indicators of emotion in human interaction that uses a timeline format (Quignard, Ursi, Rossi-Gensane, André, Baldauf-Quilliatre, 2016); and 3) a representation that aggregates indicators over particular phases of human interaction (Shaffer, Collier, & Ruis, 2016). We explore how moving between these three representations — in other words, creating additional analysis artifacts — gives us insight into the underlying structure of the data. In order to illustrate the interest of the innovative method we propose, we focus on how the different timescales of the representations interact with one another to triangulate and elaborate our understanding of what these students are thinking and feeling as they discuss their work in school. The main goal is to establish a quantitative measurable link between two types of analyses (currently manual, but potentially automatable). The first analysis occurs on the transcription of a face-to-face human interaction and tags micro-level linguistic phenomena that characterize emotion. The second analysis splices this interaction into phases, each of which are characterized on a macro-level as embodying a particular emotion. The representation that aggregates indicators allows us to connect these two manual analyses by showing which low-level indicators characterize each phase. But in order to make sense of this, we need to go back to the context and explain how low-level

indicators can give rise to macro-level characterizations. If we can do this, we have a potentially powerful method ready to be tested in a variety of contexts, be they learning contexts or otherwise. Finally, we can look into automating the analyses at the micro-level. As it may not be obvious to know which low-level indicators are the most relevant for characterizing a particular phase, such automation allows us to test a great number of indicators for their pertinence.

1.1   The Role of Emotion in Collaborative Knowledge Construction

The context we chose to illustrate translating between these temporal representations also needs to be argued for in terms of its interest for learning. In this section then, we argue that emotion is important for collaborative knowledge construction in order to set the stage for the type of insights that moving between temporal representations can furnish. Although educational and developmental psychology have recognized in recent decades that competence is not only a cognitive behaviour, but also involves the social skill of displaying that behaviour (Perret-Clermont, Perret, & Bell, 1991), research focusing on how emotion cuts across the cognitive/social divide is somewhat more recent (Baker, Andriessen, & Lund, 2009).

In interactional linguistics, emotion is a language-based micro-social phenomenon that appears during action that is co-defined and co-managed between participants (Mondada, 2001; Quignard, et al., 2016). In this framework, indicators of emotion are identified within their temporal contexts in order to understand when they are present, when they are absent, and what their duration is. The goal is to see which indicators occur together during participant interaction in a particular moment at any given time.

In psychology, emotions or affects are known to have three main components (Cosnier, 1994): 1) a psychological feeling, 2) an expressive, behavioural component, and 3) associated bodily, physiological reactions. Emotions are positive or negative, have a degree of arousal, and a duration. In addition, people are more or less aware of them and can more or less control them. In this framework, indicators of emotion can be recorded as introspective answers in response to questions, observed in a way similar to indicators used in interactional linguistics, or physiologically measured (e.g., rate of heartbeat, sweating, shaking, face colouring, etc.).

At the crossroads of argumentation studies and research in computer supported collaborative learning, our own work has specified the social and cognitive functions of emotions in argumentation (Polo, Lund, Plantin, & Niccolai, 2016). The social function of emotion refers to sociocognitive processes of collective reasoning. One example concerns Mercer and colleagues' categories of exploratory, cumulative, and disputational talk (e.g., Fernández, Wegerif, Mercer, Rojas-Drummond, 2002), where the group is respectively aligned on a consensual footing, a constructively critical footing, or a competitive footing (Polo, Plantin, Lund & Niccolai, 2016). The cognitive function of emotion refers to the cognitive and discursive process of schematization (Grize, 1996, 1997) where a discourse object is both characterized and appraised in ways that cast light on particular aspects of the object, thereby producing a representation of it that is not neutral. This kind of emotional schematization can therefore function

cognitively in that "emotions highlight; they make things stand out; they are sources of salience" (Lipman, 2003, p. 129).

Such an analysis makes it possible to track indicators of emotion across different timescales (Polo, Plantin, Lund, & Niccolai, 2016). For example, the kinds of low-level indicators that we track here are visible at the micro-level, manifest in moment-to-moment interactions through the use of prosody (e.g., rate of speech, syllabic emphasis) but at a more meso-level, it is also possible to characterize a sequence of interaction in terms of its emotional colouring. In the model we developed concerning the functions of emotional entities in the sociocognitive activity of collective reasoning, we connect an analysis of moment-to-moment emotions with the quality of an argumentative sequence, depending on how the debate is emotionally framed. In this paper we take a first step towards a formal model of this process, but instead of looking at argumentative sequences, we look at narratives of experienced emotions. In doing so, we seek to understand the micro-level through which emotion is built and expressed while connecting this to the larger time scale over which emotions develop and are shared.

This brief review highlights different ways emotion can be measured and also gives an indication of how it is both a socio-interactive and content-based phenomenon (see also Baker, Järvelä, & Andriessen, 2013). In this paper, we take the analyses of emotional indicators of human interaction that our three temporal representations make possible and relate them to the phases of activity that make up our corpus. Although our corpus is not a pedagogical task, per se, such an approach allows us to show how it is possible to meaningfully relate micro- and macro-analyses of emotion and content in human interaction, thus setting the stage for doing similar work on a corpus derived specifically from a learning context.

1.1.1 Modelling epistemic frames

The analysis above suggests that representations that show the temporal relationships among emotional events are a key part of socio-interactive and content-based analyses of emotions. There are several approaches to this.

One, which we use in what follows, is called FRIEZE. Similar to tools such as Tatiana (Dyke et al. 2009) and CORDTRA (Hmelo-Silver, Liu, & Jordan, 2009), FRIEZE provides researchers with a graphical representation of annotations as they occur in temporal context. Researchers can selectively show and hide indicators to observe temporal patterns in their occurrence, and to see how patterns of indicators change over time.

Here we investigate combining this approach with a model of discourse based on previous work in cognition and expertise. Studies suggest that a critical part of thinking about ideas and concepts is not just the presence or absence of ideas in discourse, but about how concepts, skills, and habits of mind are related to one another. For example, Bransford, Brown, and Cocking (1999) showed that experts have an extensive and complex organization of their knowledge about a domain. Chi, Feltovich, and Glaser (1981) showed that novice physics students have a different organization of their knowledge than more

expert students. DiSessa (1988) and Shaffer (2012) go further, suggesting that solving simple problems in a domain does not necessarily require understanding more than basic concepts. However, higher-level performance can only be achieved when concepts are linked to one another systematically. Shaffer (2012) describes this systematic web of connections as epistemic frame, and argues that patterns of association characterize groups of people who share similar ways of solving complex problems.

Based on this idea, epistemic network analysis (ENA) analyzes epistemic frames, or patterns of connections in student understanding, by looking at the co-occurrence of markers in discourse, which can be text, gestures, vocalizations, or actions that express some idea, emotion, or state in a learner. ENA uses network analytics to model the pattern of connections within discourse (Shaffer, Collier, & Ruis, 2016; Shaffer et al., 2009; Shaffer & Ruis, 2017). That is, ENA measures the structure of the connections in talk and action by grouping utterances into stanzas, or collections of related utterances and analyzing the co-presence of markers within each stanza.

Ideas and emotions are connected within topics or activities; however, expressions in discourse are also temporally contingent. During discussions, for example, students build understanding by "saying" and replying to "what is said" (Wells, 1999). As Bakhtin (1986) and Smagorinsky (2011) argue, speech typically addresses speech and anticipates a response, so is always "derivative" of previous contributions to shared discourse. Following this line of work, Dyke, Kumar, Ai, and Rosé (2012) and Suthers and Desiato (2012) propose sliding window analyses to model this temporal dimension of connections. A sliding window computes a value for sections of an activity (e.g., three turns of talk in Dyke et al. (2012). The window slides in the sense that a value is computed for each utterance based on the preceding lines of talk.

The method of using sliding windows to construct network models is described in more detail in Siebert-Evenstone et al. (2016). But briefly, ENA constructs a network model of each turn of talk, indicating the connections between that turn of talk and the previous turns in the sliding window. The network models for turns of talk for each person or group are then summed into a cumulative network that shows the strength of association between discourse markers over some period of time. As a result, it produces a representation of connectivity that is integrated over time. In this paper, we leverage this representation, along with two others, in order to gain insight about 1) how low-level indicators of emotion are related to each other in a given temporal segment and 2) how these indicators relate to activity phases in a narrative about schoolwork, where the activity phases have also been given an emotional colouring at a more macro-level.

1.2   Context of Study

Our corpus — called *Jaune Fluo* (Fluorescent Yellow) — was originally analyzed in the French Orféo project (Quignard et al., 2016) in order to observe the emotional character of verbal sequences and to analyze how this character emerges or declines through verbal, non-verbal, and interactional mechanisms (Plantin, Doury, & Traverso, 2000). The corpus — part of Meal Conversations Amongst

Students — is in French and is available from the CLAPI database.[1] It concerns a conversation between two students, M and J, during their lunch. J narrates her experience of one of her teachers giving graded work back to her while criticizing her in front of her peers.

1.3   Methodology

In this section, we describe how we first harnessed previously transcribed and analyzed audio data from another project that had been annotated using methods from interactional linguistics and put into the FRIEZE timeline (see below) where annotations were visible. Annotations were carried out on both a micro- and macro-level, in both cases characterizing phenomena with an emotional colouring. We relate how this data was used as input to an epistemic network analysis and how in light of the results we obtained, we returned to the timeline as well as to the transcribed audio data in order to interpret our findings.

It is important to note that timescales and representations are related, but they are separate dimensions. Timescales are part of the initial problem space. Data are coded in terms of timescales, on both the micro-level, in terms of words or events (e.g., pauses, interjections) and on the macro-level in terms of phases that have been given a characterization. On the other hand, this same data is represented in different ways, depending on the tool used. FRIEZE chooses to use rectangles to represent both micro-words/events and macro-phases on a timeline. In this way, one can see which micro-words/events make up which macro-phase. ENA process micro-words/events and macro-phases differently. It produces one network for each phase and can also produce a network that illustrates how one phase differs from another in terms of how the micro-words/events characterize them.

1.3.1   Harnessing previously analyzed data

The original analysis of the *Jaune Fluo* corpus — carried out by linguists — was divided into four successive phases using a collaborative, grounded analysis approach. First, each of the nine researchers involved was asked to analyze with his/her own method and research habits the corpus excerpt and explain how (s)he would divide the entire interaction in terms of phases that could be given an emotional colouring. These research habits involved studying and annotating a transcription that had been done using a set of conventions designed to render visible particular forms of human interaction. Each phase that was proposed had to be justified by giving the set of indicators or clues by which the emotional colouring could be assessed.

Second, researchers held a data session to share their respective analyses. They reached a consensus on how to divide the corpus, choosing four successive phases with the following emotional colouring (fear, guilt, shame, and pride). They thus agreed on both the type of emotional colouring and the borders of

---

[1] CLAPI database in Lyon, France: http://clapi.ish-lyon.cnrs.fr – *Corpus: Repas Conversations Entre Étudiants Lyon 2006. Enregistrement: Repas français: vraiment vraiment désolée*. As supporters of open data, we provide the direct link to the raw corpus as well: http://clapi.ish-lyon.cnrs.fr/V3_Feuilleter.php?num_corpus=47

each phase. The group gathered and organized all the markers in a global annotation grid (Quignard et al., 2016) — cf. Table 1.

Third, a danger of circularity was identified that the researchers countered by adapting their method. This circularity involved researchers experiencing empathy with J and M and, as a result, annotating a particular segment as being highly emotional in a particular way and ignoring another segment, even if both segments had equal numbers of markers for emotion. In order to avoid a biased approach because of empathy, the researchers decided to perform a new, systematic annotation of the entire excerpt with a restricted set of indicators (lines 1–8 of the preceding table). Each person in the group was in charge of annotating only one category of markers (i.e., one person analyzed the words, pauses, and overlaps; a second analyzed speech rate and salience; a third laughter; a fourth the discursive markers, etc.). The goal of this task was to obtain a multi-annotated transcript with the least possible dependency between categories.

Table 1. Grid of the annotation types used to analyze the corpus in the Orféo project

N	Category	Indicators
1	Prosody and body talk	Salient syllables, high or slow word rate
2	Interactivity	Pauses, latching turns, overlapping turns
3	Non-verbal vocal productions	Breaths, vocalizations, muffled sounds, laughter
4	Discourse markers	Agreement, disagreement, phatic markers, interjections, hesitation
5	1st and 2nd person markers	Possessives, verb marks, pronouns
6	Macrosyntactic segments	Central vs. peripheral (pre, post or embedded) syntactic constructs
7	Turn construction units	Simple, composed or aborted units
8	Repeated segments	First, second, … n-th mention of the same verbal segment
9	Remarkable syntactic constructions	Non-verbal phrases, binary structures
10	Discourse	Reported talk, rhetorical style
11	Commitment	Modalization, intensification, exaggeration
12	Lexicon	Specific lemmas related to emotion, affects, feeling

The fourth and final phase of the task for this initial group of researchers in the Orféo project was to investigate the relationships between subsets of markers and emotional or noticeable phenomena, by the use of a representation (called FRIEZE), specifically designed for it. FRIEZE gives a graphical representation of annotations in time, and a way for researchers to select subsets of indicators in order to observe how they behave in relation to each other as well as with respect to specific interactional phases or noticeable turns (Are indicators simultaneous? Do they appear in conjunction or in opposition? Etc.). The representation also gives a way to calculate an interactive curve, which integrates (as a sum) the number of indicators present and selected at a given time. This curve provides insight regarding which segments in the excerpt are more or less impacted by indicators than on average.

This last stage was the most complex for the researchers for two reasons. First, a large number of indicators were in play and this meant that the scope of the analysis was much broader than what the researchers were used to in their regular practice. Second, the timeline formed from the verbal content was not a data representation with which the linguists were familiar. The group therefore focused on very short and targeted linguistic phenomena. They did not return to the four phases delimited at the beginning (fear, guilt, shame, and pride) in order to connect their annotations to this emotional colouring.

Here then, we propose to do just that; we pick up this project in its third phase at the stage of the annotated excerpt and we investigate which low-level emotional indicators and indicator patterns are the most relevant with respect to the four phases that were given emotional colouring. To that end, we more specifically focused the indicator set on phenomena that are simple to identify, and that can theoretically be recognized automatically. These are also indicators that closely overlap with the transcription conventions in Polo, Lund, Plantin, and Niccolai (2016). They are shown in Table 2, below.

Table 2. Low-level emotional indicators

n	Category	Indicator	Code
1	Prosody, body talk	Elongated syllables	ELONG
2	Prosody, body talk	Salient syllables	SALIENT
3	Prosody, body talk	High word rate	HIWORDRATE
4	Interactivity	Speech overlap	OVERLAP
5	Non-verbal vocal productions	Vocalization	VOCAL
6	Non-verbal vocal productions	Laughter	LAUGHTER
7	Discourse markers	Exclamations, interjections	EXCLAM

The annotation had been carried out in the Orféo project on a time-aligned transcript in which each word has a determined position in time (starting and ending time). Since the indicators are attributed to words, we extracted a large table from the annotation software Praat (Boersma & Weenink, 2017) that had the position and extension in time of all the indicators. We put this into column form also indicating in which emotional phase the indicators occurred. These annotations were used as input in order to generate representations from both FRIEZE and ENA.

1.3.2 Preparing the corpus for ENA

ENA has been used to model the structure of connections among coded elements in a range of learning analytic contexts (e.g., Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015; Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Hatfield, 2015; Knight, Arastoopour, Shaffer, Shum, & Littleton, 2014; Orrill, Shaffer, & Burke, 2013; Quardokus Fisher, Hirshfield, Siebert-Evenstone et al., 2016; Svarovsky, 2011). The specific method with which ENA does this is described in detail elsewhere (Shaffer, 2014; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017; Siebert-Evenstone et al., 2016), but in brief, ENA creates network models by computing the co-occurrences of codes in recent temporal context — that is, within a *stanza*. A stanza is defined for each utterance in the data as that utterance (known as the *referring utterance*) plus some number of *preceding utterances* that come immediately before the referring

utterance. The co-occurrence of concepts in recent temporal context is a good indicator of connection (Dorogovtsev & Mendes, 2013; i Cancho & Solé, 2001; Landauer, McNamara, Dennis, & Kintsch, 2007; Lund & Burgess, 1996; Newman, 2004), and so ENA creates adjacency matrices that quantify the co-occurrence of coded elements in the data. The resulting adjacency matrices are normalized and embedded in a high-dimensional space. A dimensional reduction is performed, and the nodes of the networks — which represent the coded elements in the data — are placed in the space using an optimization algorithm such that the centroid of each network corresponds to the location of the network in the dimensional reduction. The weight (thickness and saturation) of the edges connecting the nodes corresponds to the relative frequency of co-occurrences for each pair of codes.

1.3.3 Comparison of network models

We can compare the pattern of co-occurrences for any two units of analysis (for example, two groups of students, or students during two different timespans) by subtracting the two network graphs. In this study, we compared the fear phase and the pride phase using a moving stanza window size of 20 lines. We chose a window size of 20 lines because the data was segmented with 1 line per word or event (e.g., pause, interjection…). The resulting corpus had 1088 lines and lasted 245 seconds in total, or 0.22 seconds per line on average. A 20-line stanza thus represents, on average, 4.4 seconds, or enough to cover at least one conversational exchange. The resulting *difference network* shows lines in pink when the relative frequency of co-occurrence of codes was stronger in the fear phase and yellow when the relative frequency of co-occurrence of codes was stronger in the pride phase. The weight of the lines in the difference network representation corresponds to the magnitude of the difference between the relative frequency of co-occurrence of codes between the fear phase and the pride phase.

1.3.4 Re-evaluating in light of results from ENA

Once specific patterns of indicators have been located by ENA, it is desirable to return to the FRIEZE representation to determine the interactional context in which they have been produced. FRIEZE enables researchers to deselect all indicators except those that form the given pattern. One can access the interactional context in transcribed form by dragging the mouse around the indicator in question. If one clicks the zone area, the sound is played. Note that an algorithm operating on the transcription software could automatically support this operation. Such an algorithm could then elaborate a collection of occurring patterns, directly targeted by ENA.

2   RESULTS

In this section, we describe the insights that each of our temporal representations make possible. We show how these representations allow for descriptions of emotional indicators of human interaction and how this relates to particular activity phases, where each phase has an emotional colouring.

2.1 Temporal Distribution of Indicators

Figure 2 shows the fear phase (top) and the pride phase (middle) where the indicators are noted by colour and duration along a timeline.[2] The legend shows the colour associated with which indicator. One can hover over an indicator (i.e., coloured box or rectangle) in the FRIEZE representation in order to bring up the speech associated with it (Quignard, et al., 2016).
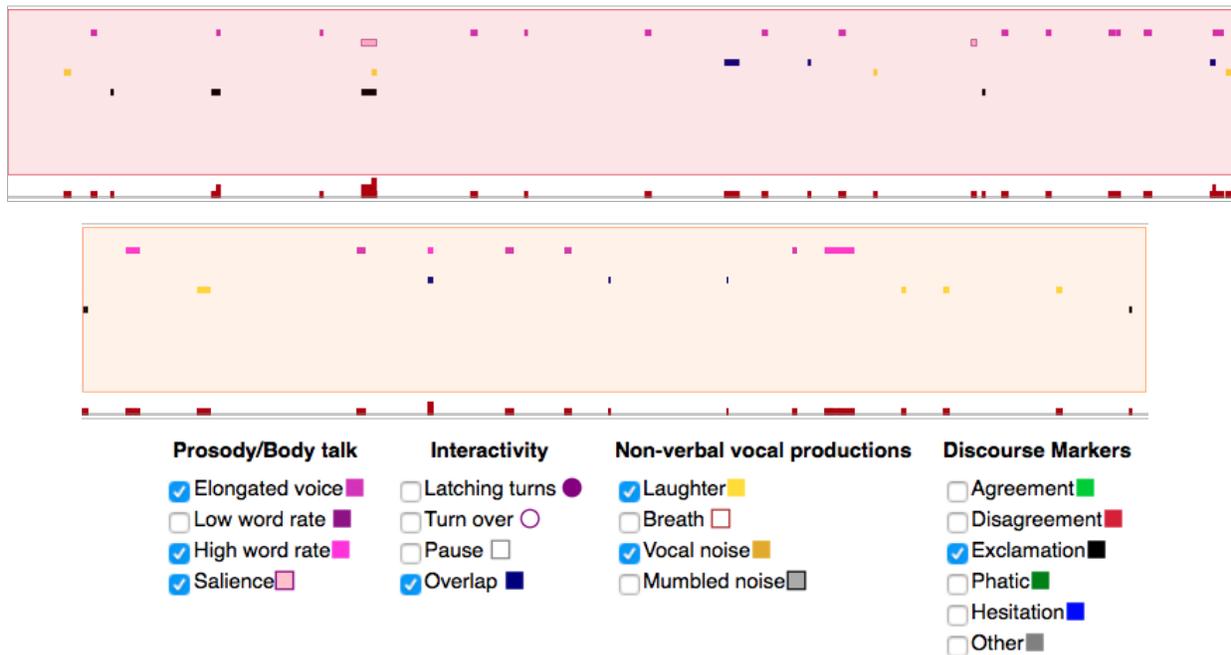

Figure 2. Fear phase (pink) and pride phase (yellow).

In the fear phase — where the speaker narrates how afraid she is of her teacher — the FRIEZE shows us vowel lengthening (elongated voice in the legend; top line, in pink) and prosodic markings (salience in the legend, also in pink), overlap is next down (dark blue), then laughing (light yellow) and vocalizations (dark yellow) are found on the same line. Lowest down, but still in the coloured rectangle, are exclamations (black). At the very bottom, just above a double line is a histogram that shows the accumulated number of indicators for that moment. In the pride phase — where the speaker turns her story of being shamed by the teacher into a positive experience — the FRIEZE also shows us that vowel lengthening occurs, and a high word rate is present, as well as prosodic markings (top line in different shades of pink). There is also speech overlap (second line, dark blue), laughter (third line, yellow), and finally exclamations (fourth line down in black). The FRIEZE thus shows similarities and differences between the two phases, but does not show which combinations of indicators are most characteristic of each phase. We thus translated the coded data into an aggregated temporal representation using epistemic network analysis (ENA).

---

[2] The FRIEZE representation is accessible here: http://perso.ens-lyon.fr/matthieu.quignard/orfeo/jauneFluo/timeline.html The FRIEZE code can be found here: http://perso.ens-lyon.fr/matthieu.quignard/Freize/timeline.js

2.2   How Patterns Differ Among Phases

ENA models the number of times a pair of actions co-occur within a given temporal window, and represents the results as a network graph. Two phases of the conversation that have been given an emotional colouring can be compared by computing a difference graph that shows which actions co-occur more frequently in each event. In this phase then, we create additional representations that give new meaning to our data.

2.2.1   Interactional indicators

Figure 3 illustrates emotional interactivity during the conversation. Note that in this model, emotional interactivity does not necessarily occur between interlocutors, but can also designate interactivity within a single utterance or between utterances by the same speaker.
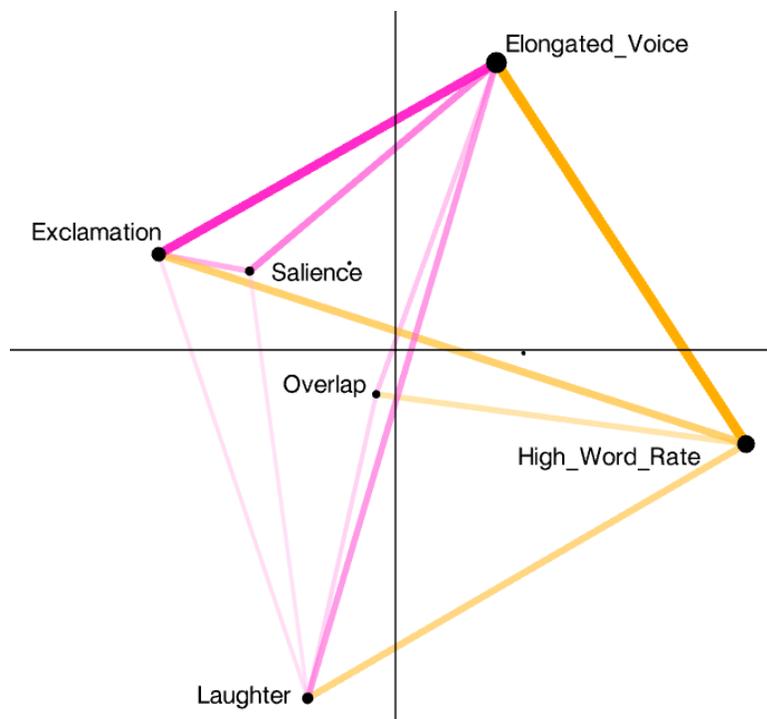

Figure 3. ENA difference graph comparing the fear (pink) and pride (yellow) phases in terms of interactional indicators.

We can now characterize the difference between the phases based on the extent to which interactional indicators are associated, and thus we can obtain a more global comparison of the phases. Figure 3 shows the extent to which different associations of indicators characterize the fear phase (pink) or the pride phase (yellow). The fear phase is characterized most strongly by elongated voice and exclamation. In the pride phase, in contrast, elongated voice is more associated with a high word rate. In the fear phase, where the speaker is narrating how frightening her teacher is, different hypotheses could explain this connection. For example, the *speaker's* elongated prosody could mean that she is waiting for her listener to give her clues that she is following the story (e.g., "you seeeee?"). And her exclamations

could point to her level of excitement (e.g., "I know!"). Alternatively, the *listener's* elongated prosody and exclamations could point to reactions to the narrative (e.g., "Reeeeaaally?").

The pride phase also has elongated prosody, but instead of being strongly connected to exclamations, in this phase elongation is most strongly connected to high word rate levels. As before, different hypotheses could explain this. Perhaps this phase is marked by more participant engagement, more interactivity, and more animation, as illustrated by the high word rate levels, as would be the case if the word rate levels include interaction between interlocutors. If they don't include interaction, they could indicate a high excitability in the speaker. Because of the segmentation of data by time slices rather than utterances, we did not distinguish between speakers while running the models in ENA. Thus, we cannot evaluate these hypotheses unless we return to the FRIEZE representations and the original transcript. However, even if we had distinguished among speakers, we still need to return to the FRIEZE representation in order to find interactional meaning.

2.3   Returning to the FRIEZE Representation

At this stage, we created another new analytic artifact where each sequence in which indicator pairs occur is put into boxed form. Figure 4 shows us where these associations of indicators occur within two of the phases: fear and pride.
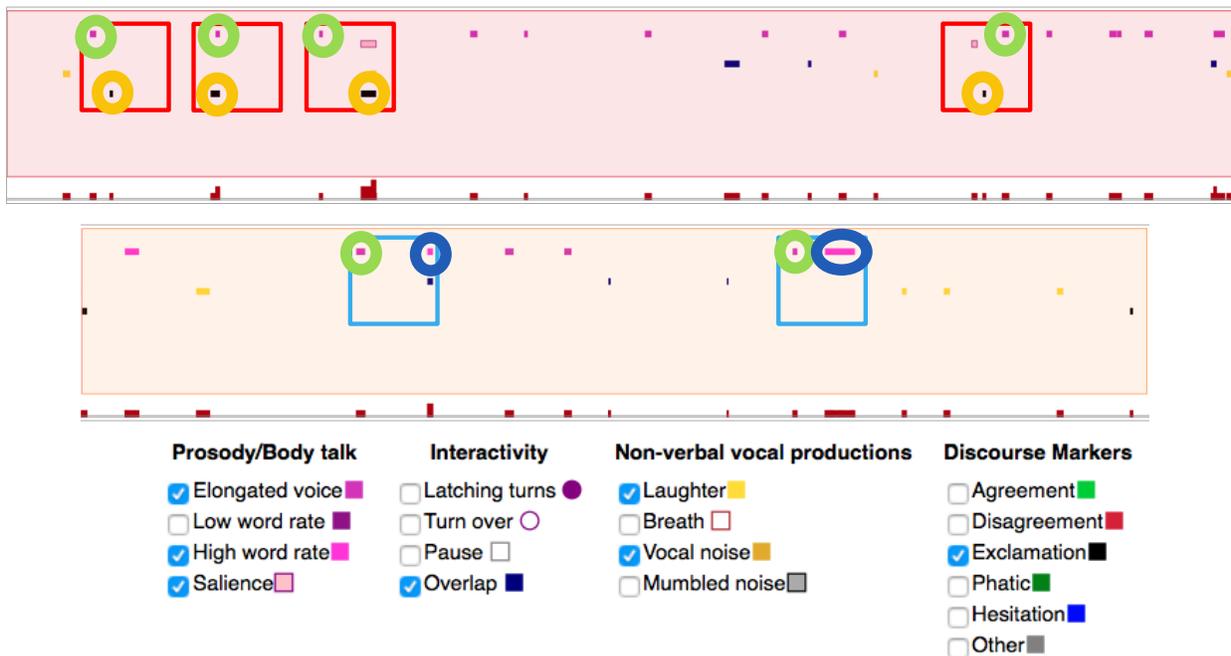

Figure 4. The FRIEZE representation annotated to show patterns of temporal connection (red rectangles) between verbal lengthening (green circles) and exclamations (orange circles) in the fear phase and patterns of temporal connection (blue rectangles) between verbal lengthening (green circles) and verbal acceleration (blue circles) in the pride phase.

This now gives us a means to return to the original data and, having located instances that characterize each phase, re-examine the evidence and re-anchor our interpretation of these representation-driven findings. Specifically, the FRIEZE representation lets us return directly to the audio data at each point in time to review the linguistic and cognitive markers of interest.

2.4 Zooming in on Emotional Indicator Patterns

In accordance with language science norms for reporting on transcriptions of corpora, the French language is kept for reference. The English translation appears afterwards. We look at our two types of co-occurring indicator patterns, the first characterizing the "fear" phase (a verbal lengthening and an exclamation) and the second characterizing the "pride" phase (a verbal lengthening and a verbal acceleration). As noted above, there are four occurrences of the first pattern and two occurrences of the second for this three-minute extract.

2.4.1 The "fear" phase

The first three examples of verbal lengthening coupled with an exclamation (all shown in Figure 3.) occur in the initial exchanges of this interaction (1a–1b, 2, and 3). In the first example (cf. Figure 4), each speaker enunciates one part of the two-part indicator pattern. As mentioned previously, the person telling the story is J and M is the listener. *Italics* signify verbal lengthening and **bold text** signifies an exclamation. When a word is both lengthened *and* is an exclamation, it appears in ***italicized bold***.

J (1a)	t'auras *jamais* 5
M (1b)	**ah** si avec ma prof c'est possible hein parce que je sais qu'elle nous attend avec une batte de baseball derrière la porte le jour de la rentrée hein
J (2)	**ah là** ***là***
M	c'est tout à fait possible... une interro qui était censée être facile la moyenne de la classe 7
J (3)	***ouais***/
M	facile hein/
…	

J (1a)	you'll *never* get 5
M (1b)	**eh** yeah I will with my prof it's possible ya know because I know she's waiting for us behind the door with a baseball bat the day classes begin
J (2)	**oh my** ***gosh***
M	it's totally possible... an exam that was supposed to be easy the class average was 7
J (3)	***yeah***/
M	easy huh/

Figure 5. A sequence including three instances (1a–1b, 2, and 3) of characteristic indicator pairs of the "fear" phase: a verbal lengthening and an exclamation).

We can note several things. First, the indicator pattern can occur within one speaker's utterances (examples 2, 3) or across speakers (example 1a–1b). Although we could have set it up to do so, ENA did not make this distinction when it calculated which pairs of indicators characterize a particular

interaction phase. The indicator pattern can also occur simultaneously for one word or group of words (3). This is simply reflected by the fact that one word or group of words can be coded in multiple ways (here a word is both lengthened *and* serves as an exclamation). Example (2) is related to this phenomenon in that the group of three words is an exclamation ("ah là là"/"oh my gosh"), but only the last word is lengthened. Finally, the order in which the indicator pattern appears is not taken into consideration by ENA, although again, we could have set up the ENA model to do so. A verbal lengthening can occur before an exclamation (such as in example 1a–1b), or the exclamation ("mais"/"I mean") can occur before the verbal lengthening (*euh, um*) See example 4 in Figure 6 below.

J (4)	et du coup elle va nous latter je pense qu'on peut le dire comme ça en même temps c'est vrai que… quand elle m'a envoyé le mail il paraît qu'elle a cassé des gens **mais** vraiment *euh* dur donc quand elle leur a dit tu sais les corrections à faire et moi tu sais elle a commencé le mail en nous disant bon il est pas mal votre poster pour un sujet compliqué parce que enfin comment représenter la culture mouais
M	ouais/ (rire)
J (4)	and then she's going to kick our behinds I think that we can put it like that at the same time it's true that… when she sent me the e-mail I heard that she gave really bad grades to people **I mean** really *um* bad so when she told them you know the corrections that had to be done and I you know she started the e-mail by saying so your poster isn't so bad for such a complicated subject because representing culture yeah
M	yeah/ (laughter)

Figure 6. A sequence where an exclamation ("mais"/"I mean") occurs before a verbal lengthening ("euh"/"um").

Recall that, given the indicator pattern that characterized the first phase of the interaction we studied, we hypothesized that when a speaker is narrating the fear she feels about how her teacher is going to treat her, her elongated prosody could mean that she is waiting for her listener to give her clues that she is following along. We also hypothesized that her exclamations could point to her level of excitement in telling her story. Alternatively, this type of indicator pattern could illustrate how her listener was attending to the narrator. Going back to the interaction shows us a combination of these. Speaker J systematically performed the vowel lengthening but it was not in order to incite listening. Rather, *she* was the listener at this point in the dialogue and was showing heightened attention to similar fear-inducing story elements. J also enunciated the exclamations in this paired indicator pattern, with the exception of 1b, pronounced by M. In a phase with an emotional colouring of fear, indicators of verbal lengthening and exclamation make sense for both speaker and listener, as they may signal emphasis (examples 1, 2 and 3, in Figure 5 and the exclamation in 4 in Figure 6). In Figure 6, however, the vowel lengthening leads to word searching during storytelling (example 4).

2.4.2 The "pride" phase

The two instances of the indicator pattern vocal lengthening/high word rate that characterized this phase are shown below (cf. Figure 7). As for the "fear" phase, *italics* signify verbal lengthening. Underlined text signifies a high word rate. In the first example, the same speaker enunciates both of the

two-part indicator pattern (J says "que"/"because" in a lengthening manner and J also accelerates the words "et puis faut"/"and then you need." In the second example, J lengthens the word "euh"/"um" and in the same turn accelerates the phrase "le seul truc c'est que la photo elle a de gros pixels quoi"/"the only thing you know is that the photo has huge pixels."

Recall that we hypothesized that the pride phase could be more participant engaged, more interactive, and more animated, as illustrated by the high word rate levels — that is — if these word rate levels included interaction between interlocutors. Here though, we see that J — the narrator — is again the interlocutor picked up by the algorithm. We also hypothesized that if one speaker spoke the utterances modelled, this could illustrate a high excitability. This hypothesis seems to be confirmed by our data. How does this relate to the pride phase, per se? In the pride phase, according to the content, the narrator and the listener take the negative emotions of fear, guilt, and shame, and turn J's experience into one of pride. Such a transformation is compatible both with verbal lengthening and excitability in the context of the content. The interlocutors are giving arguments about the quality of some of the other students' work in relation to J's, and this helps to make J feel better about her own work.

J (1a)	une photo c'est toujours un peu délicat parce *que* les pixels c'est pas beau quoi
M	hm ben il faut vraiment trouver la la=
J (1b)	= et puis faut
M	le ton qui fait que ça se voit pas trop non plus et que ça
_____	
J (1a)	a photo is always a bit more delicate *because* the pixels are not beautiful you know
M	hm well you have to really find the the=
J (1b)	= and then you need
M	the tone so that you can't see it as much and then that

M	quand tu mets de trucs par dessus après c'est trop chargé moi je trouve
J (2)	ça dépend t'sais les autres ils en ont tu sais thierry john et camille ils ont fait un truc *euh* ils ont une photo et ça rend bien mais le seul truc c'est que la photo elle a de gros pixels quoi
M	oui
_____	
M	when you put things on top of it after it's too busy I think…
J (2)	it depends you know the others they have them you know tim james and carole they did a thing *um* they have a photo and it looks good the only thing you know is that the photo has huge pixels
M	yes

Figure 7. A sequence including two instances (1a–1b and 2) of characteristic indicator pairs of the "pride" phase: a verbal lengthening and a high word rate.

3   DISCUSSION

3.1   Contribution

Our contribution illustrates the interest of transforming data between different temporal representations in order to increase the interpretative power of analyses that strive to better understand the relationship between emotion, content, and interaction. The first representation of our

data shows how indicators of human interaction over two different phases of activity can temporally co-occur (cf. Figure 2), and the second (cf. Figure 3) aggregates those indicators over time in order to show what combinations of them occurring in a particular phase best characterize that phase as compared to other phases. Finally, looping back to the initial transcription (cf. Figures 4, 5, 6, and 7) in order to re-examine the indicators within their interactional context — and in relation to the content of the narrative — gives us insight as to how these indicators can be interpreted. We can also make sense of how these indicators characterize particular phases, given their phase-level description. In other words, we illustrated how human interaction can be characterized at different levels of granularity, and how these levels can be connected. Said yet differently, we linked micro-indicators of emotion (represented by FRIEZE) to macro-characterizations of emotion (also represented by FRIEZE) through the use of aggregation (ENA). We generated two types of new knowledge. First, we can see the interest of being able to connect micro- and macro-levels of analysis in this way, but only if we can successfully find meaning in these connections by going back to the transcription and giving interpretations to them. Second, once this method has been demonstrated, we are set to apply it to different contexts and further, connect micro-level indicators of one sort to macro-level indicators of *another* sort. In this paper, our analyses at micro- and macro-levels both dealt with emotions.

3.2   Limitations

The obvious limitation of this paper is that it is a single case study from a relatively small corpus of interaction data. The data was hand coded, and therefore the analysis as described here could not be accomplished at scale. However, we argue that automated coding is a separate issue from the analysis of coded data, and a critical step in the development of new analytical methodologies is to work with data at smaller scales, where the researchers can compare results from quantitative analysis to the grounded interpretations of data produced from line-by-line inspection (see Shaffer, 2017). Thus, while it is true that *this particular corpus* is relatively small, and that *this particular coding* on *this particular corpus* was done by hand, the methodological implications are widely applicable to the LAK community as a whole.

If we discuss limits more in terms of what we could have done differently, it is clear also from the analysis of excerpts from the transcripts — chosen by identifying segments of the interaction using the FRIEZE, based on patterns identified by the ENA analysis — that we could profitably change some of the choices we made in setting up the ENA model to better understand how emotional indicators interact in this corpus. For example, we could distinguish between intra- and inter-personal links between emotional markers; and, we could consider the temporal ordering of emotional markers in the discourse. We did not continue the iterative process of translating between temporal representations here, but the analysis above suggests that the different temporo-analytical scales are mutually refocusing, in the sense that each temporal representation has the power to orient us more productively to details visible at another scale in another representation. Put another way, the cyclical process of data capture and analysis described by Dyke et al. (2009) does not necessarily develop new

representations, but can also involve cycles of mutual-refinement among a set of representations working from the micro-level of a transcript to the global level of a discourse network.

Finally, how a window size is chosen deserves some attention. Our goal was to capture the recent previous temporal context of each line of data. We wanted this temporal context to be just large enough to find what is relevant to the line of data we were looking at, but not so large so as to introduce extraneous connections. And although we argue for this in section 1.3.3, the careful reader may suggest that our mean of 0.22 seconds is an arbitrary value. One can question whether the distribution of our event duration is Gaussian, thus warranting using such an average for window size. Although the distribution is not Gaussian, it is unimodal with a peak around 0.15 seconds (median value) and the geometrical mean is also 0.15. That said, the mean value we did use (0.22 seconds) fits very well into the peak of the distribution. But given this discrepancy, we re-calculated the difference graphs with a larger window size (i.e., 30 lines). We still obtained the same basic result, with elongated voice connecting to exclamation in the fear phase and to high word rate in the pride phase. Indeed, there is no natural language processing solution for knowing how far to go back in time in human interaction in order to pick up all of the prior referents to a give line of data. Our calculation for window size was designed to pick up one conversational exchange, but analysts should experiment with window size and check how and why results change, if they do.

3.3   Perspectives

In terms of perspectives, having established the interest of our methodology, we are currently performing similar analyses on data from a serious games context where students must work together using knowledge about mechanics and electronics in order to explore the terraforming of other planets. In addition to indicators of emotion, we will also analyze the competencies mobilized, and we will do so in relation to the pedagogical activity phases. We frame this as another instantiation of our method, but this time, we will perform two types of micro-level analyses. We will still use low-level indicators of emotion, but we will add low-level indicators of the competencies mobilized as students take on the *savoir-faire* of mechanics and electronics professions. Then, we will relate both of these low-level indicators to the specific pedagogical activity phases. We expect that low-level indicators of emotion will allow us to identify where engaged, collaborative problem solving is occurring because this is often characterized by heated arguing or celebrated agreement, both of which influence how learners may or may not publicly admit to change of opinion (Molinari & Lund, 2012). We will also be able to see how such emotion relates to the knowledge being discussed, and where this knowledge is situated in the pedagogical sequence.

4   ACKNOWLEDGMENTS

REFERENCES

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology*, *6*(1016). http://dx.doi.org/10.3389/fpsyg.2015.01016

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, *32*(3B), 1492–1501.

Baker, M., Andriessen, J., & Lund, K. (2009). Socio-relational, affective and cognitive dimensions of CSCL interactions: Integrating theoretical-methodological perspectives. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Proceedings of the 8th International Conference on Computer-Supported Collaborative Learning* (CSCL 2009), 8–13 June 2009, Rhodes, Greece (Vol. 2, pp. 31–33). International Society of the Learning Sciences.

Baker, M., Järvelä, S., & Andriessen, J. (Eds.) (2013). *Affective learning together: Social and emotional dimensions of collaborative learning.* London: Routledge.

Bakhtin, M. (1986). *Speech genres and other late essays.* Trans. Vern W. McGee. Austin, TX: University of Texas Press.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program]. Version 6.0.24. http://www.praat.org/

Bransford, J. D., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* *5*(2), 121–152. http://dx.doi.org/10.1207/s15516709cog0502_2

Cosnier, J. (1994). La psychologie des émotions et des sentiments, Paris, Retz; en version actualisée : Émotions et sentiments, Conférences Université Saint-Joseph, Beyrouth, 2004.

DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 47–70). Lawrence Erlbaum Publishers.

Dorogovtsev, S. N., & Mendes, J. F. F. (2013). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press.

Dyke, G., Kumar, R., Ai, H., & Rosé, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In J. van Aalst, B. J. Reiser, C.

Hmelo-Silver, & K. Thompson (Eds.), *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences* (ICLS '12), 2–6 July 2012, Sydney, Australia (Vol. 1, pp. 363–370). International Society of the Learning Sciences.

Dyke, G., Lund, K., & Girardot, J.-J. (2009). Tatiana: An environment to support the CSCL analysis process. In C. O'Malley, D. Suthers, P. Reimann, & A. Dimitracopoulou (Eds.), *Proceedings of the 8th International Conference on Computer-Supported Collaborative Learning* (CSCL 2009), 8–13 June 2009, Rhodes, Greece (Vol 1, pp. 58–67). International Society of the Learning Sciences.

Dyke, G., Lund, K., Suthers, D. D., & Teplovs, C. (2013). Analytic representations and affordances for productive multivocality. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions*. In C. Hoadley & N. Miyake (Series Eds.), Computer Supported Collaborative Learning Series: Vol. 15 (pp. 639–658). New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3_33

Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2002). Reconceptualizing "scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *Journal of Classroom Interaction*, *36*(2/1), 40–54.

Grize, J.-B. (1996). *Logique naturelle et communication*. Paris: Presses universitaires de France.

Grize, J.-B. (1997 [1990]). *Logique et langage*. Paris: Ophrys.

Hatfield, D. L. (2015). The right kind of telling: An analysis of feedback and learning in a journalism epistemic game. *International Journal of Gaming and Computer-Mediated Simulations*, *7*(2), 1–23.

Hmelo-Silver, C. E., Liu, L., & Jordan, R. (2009). Visual representation of a multidimensional coding scheme for understanding technology-mediated learning about complex natural systems. *Research and Practice in Technology Enhanced Learning, 4*(3), 253–280. http://dx.doi.org/10.1142/S1793206809000714

Cancho, R. F., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*(1482), 2261–2265. http://dx.doi.org/10.1098/rspb.2001.1800

Knight, S., Arastoopour, G., Shaffer, D. W., Shum, S. B., & Littleton, K. (2014). Epistemic networks for epistemic commitments. Paper presented at the International Conference of the Learning Sciences (ICLS '14), 23–27 June 2014, Boulder, CO, USA.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Lipman, M. (2003). *Thinking in education*. Cambridge, UK: Cambridge University Press.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. http://dx.doi.org/10.3758/BF03204766

Lund, K., & Suthers, D. D. (2013). Methodological dimensions. In D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, & N. Law (Eds.), *Productive multivocality in the analysis of group interactions.* In C. Hoadley & N. Miyake (Series Eds.), Computer Supported Collaborative Learning Series: 15, 21–35, New York: Springer. http://dx.doi.org/10.1007/978-1-4614-8960-3

Molinari, G., & Lund, K. (2012). How a power game shapes expressing opinions in a chat and in an argument graph during a debate: A case study. In J. van Aalst, B. J. Reiser, C. Hmelo-Silver, & K. Thompson (Eds.), *The Future of Learning: Proceedings of the 10ᵗʰ International Conference of the Learning Sciences* (ICLS '12), 2–6 July 2012, Sydney, Australia (Vol. 2, pp. 232–236). International Society of the Learning Sciences.

Mondada, L. (2001). Pour une linguistique interactionnelle, *Marges Linguistiques*, 1, mai, 142–162.

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*(5), 56131. http://dx.doi.org/10.1103/PhysRevE.70.056131

Orrill, C., Shaffer, D. W., & Burke, J. (2013). Exploring coherence in teacher knowledge using epistemic network analysis. Presented at the American Educational Research Association Annual Conference (AERA 2013), 27 April–1 May 2013, San Francisco, CA, USA.

Perret-Clermont, A.-N., Perret, J.-F., & Bell, N. (1991). The social construction of meaning and cognitive activity in elementary school children. In L. Resnick, J. Levine, & S. Teasley. *Perspectives on socially shared cognition* (pp. 41–62). Hyattsville, MD: American Psychological Association.

Plantin, C., Doury, M., & Traverso, V. (2000). *Les émotions dans les interactions*. Lyon: Presses Universitaires de Lyon.

Polo, C., Lund, K., Plantin, C., & Niccolai, G. (2016). Group emotions: The social and cognitive functions of emotions in argumentation. *International Journal of Computer Supported Collaborative Learning, 11*(2), 123–156. http://dx.doi.org/10.1007/s11412-016-9232-8

Polo, C., Plantin, C., Lund, K., & Niccolai, G. (2016). Group emotions in collective reasoning: A model. *Argumentation.* http://dx.doi.org/10.1007/s10503-016-9407-5

Quardokus Fisher, K., Hirshfield, L., Siebert-Evenstone, A. L., Arastoopour, G., & Koretsky, M. (2016). Network analysis of interactions between students and an instructor during design meetings. *Proceedings of the Annual Conference of the American Society for Engineering Education* (ASEE 2016), 26–29 June, New Orleans, LA, USA (17035). ASEE.

Quignard, M., Ursi, B., Rossi-Gensane, N., André, V., Baldauf-Quilliatre, H., et al. (2016). Une méthode instrumentée pour l'analyse multidimensionnelle des tonalités émotionnelles dans l'interaction. Congrès Mondial de Linguistique Française 2016, Tours, France. http://dx.doi.org/10.1051/shsconf/20162715004

Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. Paper presented at the 12ᵗʰ International Conference of the Learning Sciences (ICLS '16), 20–24 June 2016, Singapore. http://dx.doi.org/10.22318/icls2016.82

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, & S. Barab (Eds.), *Games learning, and society: Learning and meaning in the digital age* (pp. 403–433). Cambridge, UK: Cambridge University Press.

Shaffer, D. W. (2014). *User guide for epistemic network analysis web version 3.3*. Madison, WI: Games and Professional Simulations Technical Report 2014-1.

Shaffer, D. W. (2017). *Quantitative ethnography.* Madison, WI: Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45. http://dx.doi.org/10.18608/jla.2016.33.3

Shaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E. A., … Mislevy, R. J. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *The International Journal of Learning and Media, 1*(1), 1–21. http://dx.doi.org/10.1162/ijlm.2009.0013

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics (pp. 175–187).* Society for Learning Analytics Research. http://dx.doi.org/10.18608/hla17.015

Smagorinsky, P. (2011). *Vygotsky and literacy research: A methodological framework* (Vol. 2). Rotterdam, Netherlands/Boston, MA: Sense Publishers.

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *Proceedings of the 45th Hawaii International Conference on System Sciences* (HICSS-45), 4–7 January 2012, Maui, HI, USA (pp. 3368–3377). IEEE Computer Society. http://dx.doi.org/10.1109/HICSS.2012.274

Svarovsky, G. N. (2011). Exploring complex engineering learning over time with epistemic network analysis. *Journal of Pre-College Engineering Education Research*, *1*(2), 19–30. http://dx.doi.org/10.5703/1288284314638

Wells, G. (1999). Dialogic inquiry: Towards a sociocultural practice and theory of education. Cambridge, UK: Cambridge University Press.