Profiling dialogues: Multi-trait mapping of televised argumentative exchanges

GÁBOR Á. ZEMPLÉN
Dept. of Argumentation Theory and Marketing Research, Institute of Business Economics, Eötvös Loránd University, Budapest, ELTE GTI

Morals and Science 'Lendület' Research Group, Hungarian Academy of Science, MTA BTK FI
zemplen@gti.elte.hu

A 24 minute televised discussion was analysed in a teaching module on how parties obey or violate norms of rational discussions. The research project studied the dynamics of the argumentative exchange, and constructed a multi-trait map of the situational meso-context. The exploratory multi-modal data gathering and handling method provided a multi-channel map of over 25 traits and 700 data points. The dataset can be linked to linguistic (micro-context) analysis, and can be utilised in institutional (macro-context) studies.

KEYWORDS: actio, complexity, derailment, dual-process, fallacy, framing, multi-modal argumentation, non-verbal communication, strategic manoeuvring, televised debate

1. INTRODUCTION: EXPLORATORY ANALYSIS OF ARGUMENTATIVE DYNAMICS

Televised debates are of the more significant sources of information for citizens concerning political, ethical, environmental, economic, policy-related or scientific issues. These debates, Q&As, as well as various other forms of persuasion dialogues inform citizens and shape their views concerning things past and things to come, as Aristotle delineates deliberative and forensic kinds of speech (Ar., *Rh.,* 1358b 1-3.).

These debates are also important fora for communication because they provide some of the most salient examples of handling disagreements and conflicts. As such, they have the potential to shape the norms of argumentative cultures in the general public. It is hard to imagine a well-developed democracy where there are *no* televised

discussions that even approach *any* model of a critical / rational / reasonable discussion, but deliberative cultures can turn out to be fragile 'ecosystems'. If argumentation theorists are interested in whether more-and-more or less-and-less televised discussions display characteristics of a critical / rational / reasonable discussion, they need tools to size up the improvement or demise. That is, they need tools to measure development in macro-context, changes in argumentative cultures.

Research on the normative, argumentative aspects of these communicative interactions has started to fuse in the last decades with empirical studies of persuasion, research in social and cognitive psychology, and research on gesture. As such, little of the methodology has become standardized, and various approaches are developed to address current research questions and future research needs. Most of the research is on the micro-context, where a link to argumentation and an interface with computing can readily be made.

The approach outlined in the paper was developed for the micro- and meso-level. The target of the analysis was a Hungarian televised debate between a pro-government journalist and a vice-rector of a University (Péter Csermely – György Fábri, HírTV, 24.02.2013). The main topic of the discussion was the politics of education: an intensifying debate that in late 2012 resulted in student protests, and some university lecture halls 'taken over' by spontaneous student organizations. As a result, the Hungarian government changed some of the short term plans for reforming higher education in the country. The exploratory community-based research started soon after the interview (2013). The data gathering and handling exercise was explored during 6 years of analysis and reanalysis of a televised debate by groups of students (14 to 40 per semester, 151 in total).

The project was designed to study how in a 24 minute discussion parties obey or violate norms of rational discussions. The interview was used to improve observation skills and methodological awareness, and from the second year the investigation focused on finding a 'bottom up' empirical approach to map the dynamics of debates. The aim was to create and study a data-set and how it can be used to enrich analysis of norm-violations. The project was officially terminated when the first student complained that the interview was 'old', in 2018[1]. After the intervention sequence I started analysing the

[1] The module on multi-trait content-analysis was developed for Communication and Media Studies M.A. students in 4x90 min/week, in 14 week courses on 'Rhetoric and Dialectic' and 'Business Communication'. The group sizes varied (19, 32, 40, 27, 14, 19, in total 151 students, appr. 10% dropout rate), and gender ratio was on average 20/80 % male/female. Key methodological steps were also tested with high-ability non-specialist B.A. and M.A. students in

data in 2019, with help from students, and colleagues, especially Mihály Héder, PhD; Erika Hlédik, PhD; Kristóf Kovács, PhD.

The research module constructed a language of redescription of televised debates that 1) expanded the targeted (usually linguistic) range of phenomena in the analysis, and that 2) provided a topography of a debate with adequate resolution, adding non-verbal, somatic responses, language-related gestures to the data-set as well as potential editorial actions (choice of specific shooting angles, like OTS, OSS).

The studio setup (see Figure 1) allowed for good non-verbal analysis (no props, no table, participants in chairs), the participants were experienced, so one could assume that the performance in the debate is predominantly controlled, and that the participants have ample experience to use situations to the best of their interests, and the debate was sufficiently long (over 20 minutes).

Figure 1 – The studio setup (Péter Csermely reporter left, György Fábri vice-rector, right).

The most important theoretical underpinnings of the approach will be discussed in more detail in Section 2., which gives a short introduction to the theoretical motivation behind the exploratory multi-trait research. Section 3. describes the didactic setting of the experiment. With respect to the extended pragma-dialectical theory of strategic manoeuvring (SM) the aim was to help students recognize the

---

Economics in a one semester elective course, and the same traits were tested with 1st year B.A. students prior to the introduction of the relevant theoretical concepts (see in more detail in Zemplén, 2014).

transitions that occur as the dialectical aims are hampered in a critical discussion, as when the "rhetorical aim has gained the upper hand at the expense of achieving the dialectical goal" (van Eemeren & Houtlosser, 2009, p. 5). Section 4. provides an introduction to the research-tool and to some of the potentials of the exploratory methodology to map argumentative exchanges, and Section 5. summarizes the paper.

2. RATIONALE FOR MULTI-TRAIT ANALYSIS OF MULTI-MODAL ARGUMENTATION

Fusing the study of 'embodied' agents with normative theories in discourse analysis presents a relatively novel and promising strand of research. My original theoretical interest at the time of developing the exercise was linked to the hope that it is possible to improve the fit between a broadly understood dialectical model and a suitable rhetorical theory in a way that justice is done to a (satisfactorily large) number of insights from social psychology and persuasion research. In the development of the multi-trait exercise I had an inspiration from studying equivalent framing, but quickly noticed some issues of concern when I appreciated the complexity of the inquiry.

The inspiration came from the study of cases, where informational content appears equivalent, yet the rhetorical effects are not, as in specific cases of framing[2]. With colleague Gergely Kertész, we assumed that several such effects do not violate the pragma-dialectical rules (or PD-rules) presupposed as necessary for reasonable discussions (first order conditions, see van Eemeren & Grootendorst, 2004, pp. 187-195). Although equivalent frames have the same information content, picking one of the variants in specific settings might be considered as manipulative by the other party and may even produce what is known as a boomerang-effect (Kruglanski & Higgins, 2007, p. 267). If in certain contexts the argumentative use of appeal framing can be considered as manipulative, then it is possible that a party quits the kind of argumentative discourse preferred by the PD theory because the party identifies a presentation device used by the other party as manipulative[3]. Can a critical discussion derail without violating the first order rules? Our initial view was that there might be derailments that are not fallacies. This might sound like an oxymoron, as

---

[2] A well-known example for success rate framing: "this surgical procedure has 90% survival rate" vs. failure rate framing "this surgical procedure has 10% mortality rate".

[3] We took the boomerang effect as a possible perlocution of the communicative move although there are no externaliseable commitments of the speech act performed that might contradict the pragma-dialectical norms (Kertész & Zemplén, 2010, pp. 2073-6).

pragma-dialecticians often treat 'fallacy' and 'derailment' as co-referent, but in this broader view derailment could be used for *any communicatively or interactionally dysfunctional move*[4] that hampers the full realization of critical reasonableness, including certain actions that block the parties from reaching the dialectical aim of the discussion. Second order conditions play a rather limited role in most discussions of the PD theory[5], and we did not pursue this line of theoretical work, but noted that in this approach *any move* can include any behavioural phenomena[6] where it is plausible to think that the action is linked to derailment, and that many of these acts might not show up in the analytical overview of the reconstruction.

2.1 Units of analysis & temporality

Many researchers in argumentation studies aim at the description of argumentation as a social activity, and prescriptive models give norms and regulations along which the functional aims (like the resolution of a difference of opinion) are easier to achieve. Modern theories are increasingly process- (as opposed to product-) oriented, however, at the level of technical analysis and linguistic (discourse analytical) foundations they generally rely on some form of pragmatic background theory that is fundamentally product-oriented, individuating discourse elements and classifying them. The pragma-dialectical school in Amsterdam addresses derailments in the extended theory as tokens of specific types (instantiations of PD-rule-violation), which type is bound to a specific discussion-stage. Parsing up interactions and pairing behavioural elements (mapping speech acts) with abstract relata, like

---

[4] In van Eemeren's view "exploiting the possibilities of presentational variation in strategic maneuvering [...] boils down [...] to 'framing' one's argumentative moves in a communicatively and interactionally functional way" (van Eemeren, 2010, p. 117).

[5] "It is important to bear in mind that the pragma-dialectical procedure deals only with "first order" conditions for resolving differences of opinion on the merits…" (van Eemeren, 2010, p. 35), also referring to *compulsions* after Barth and Krabbe. See also: "To some extent, everyone who wants to satisfy the second-order conditions can do so, but in practice, people's freedom is sometimes more or less severely limited by psychological factors that are beyond their control, such as emotional restraint and personal pressure." (van Eemeren & Grootendorst, 2004, p. 189).

[6] The 'phenomena' studied might be distributed over time, etc. For the distinction between data and phenomena see (Bogen & Woodward, 1988), which was found productive in an earlier analysis of syntactic microvariation, to find subpopulations with different grammaticality judgements using Euclidean distance and Ward's clustering algorithm and two-way analysis of variance (ANOVA) (Gervain & Zemplén, 2005).

discussion stages is not fundamentally different from individuating argument schemes, locating fallacies, testifying to the dominantly taxonomic and set-theoretical motivation widespread in current approaches to argumentation. Most normative analyses move towards a reconstruction that is an *atemporal* product, e.g. a syllogistic reconstruction, a list of argument schemes, etc. In the analytical overview of a pragma-dialectical analysis the sequence of speech-acts might be rearranged, for example. When temporality is included (as in a reconstruction of dialogue moves), it is generally in the sense of 'sequencing'. This is a rather significant limitation if we want to focus on *people* as opposed to *statements* when studying arguments.

For a multi-modal rhetorical analysis, such frameworks become restrictive at some point. Fine temporal resolution is very important in the study of both nonverbal communication, and many aspects of the use of voice, and the methodology should arch from *actio qualities* to presentational devices, to give an account of both the speaker's style, the *energy* of her/his gestures and voice, and a normative evaluation of the strategic manoeuvres. In the development of the methodology, I tried to take to heart the warning by Robert Rosen: "Any question becomes unanswerable if we do not permit ourselves a universe large enough to deal with the question" (Rosen, 1998, p. 2.). As the temporality of the *being* stands in stark contrast to the *proposition*, one of the preliminary decisions was to focus on repetitive actions and treat them as temporal events. The debate is transcribed as an n-dimensional universe, consisting of n types of action, where occurrences of the tokens of the specific types have temporal properties and can have various additional properties.

Another crucial decision made was to assume various processes that influence performance. A heated debate significantly affects the neuro-endocrine system, and some reporters use techniques (including interruptions) that increase the likelihood of certain speech events in the performance of the interviewed (e.g. switching off, or non-grammatical sentence production). The 'real debate' is in real time, and arguers are complex systems with interacting components, and with respect to the internal organization (micro-level description) open systems. Embodied agents do all sorts of things when engaged in communicative exchanges, including unconscious coordination between participants, the chameleon effect. Already for a broad rhetorical analysis not everything that the agent does translates well to *actio*[7], a

[7] Consider e.g. "Actio differs from nonverbal communication in general in that actio is performed in a rhetorical situation with the intention to be persuasive." (Gelang & Kjeldsen, 2011).

movement can be made as part of *actio* and/or as part of *stress-relief* (glitches, manipulators, etc.).

Acknowledging 'embodiment' is acknowledging that taking part in a debate is a (multi-)goal directed action. The approach can and probably should rely on a significantly broader notion of function than the one developed by pragma-dialecticians, expressed in the metatheoretical commitment of 'functionalization'. In PD elements are ordered to distinct issues, stages, and are normatively evaluated with respect to +/- fulfilling their functions (not violating norms, and therefore not hindering a resolution of a difference of opinion). For the analysis assuming a dual-process model of the participants[8], we can easily locate research questions: Are there not techniques to hamper the optimal functioning of the reflective system or to deliberately promote malfunction of the rational agent?[9] Should not there be somatic responses to perceiving norm-violations? Or are there ways of specifically triggering (if possible, via *actio)* reflexive processes, 'compulsions' that influence dialectical and rhetorical performance?

The broader functionalization of 'elements of discourse' can enable a multi-modal rhetorical analysis, but this is also the point where it is easy to lose the foothold: saying that things like waving hands or changes in pitch are relevant for a normative analysis is like opening Pandora's box far too wide. How to study the *dynamics* of argumentation in a way that the analysis can be related to traditional discourse-analysis? When we attempt a 'bottom up' mapping of an argumentative exchange, we need to accept that what we find may well be considered 'noise', and some points on our map and many of our conclusions (derived from an analysis of the data) may very well be artefacts.

Can one extend or improve a theory when one does not know what exactly to include in the empirical domain? Unless the data are somehow theoretically interpreted, even if statistically significant, they mean little. It was assumed that if we start to collect data, some of these may be linked to social regulations, and some to homeostasis, some to

---

[8] Dual-process frameworks were outlined to the students, based on (Lieberman, 2003). The approach was linked to ongoing research (Hodgkinson et al., 2008, Stanovich & West 2000, Mercier & Sperber, 2009), assuming differentiability between slower, 'critical', reflective belief-generating processes and reflexive information processes - generally pre-linguistic, somatic, reactive responses.

[9] "Through development, socialization, and individuals' learning of social rules, the reflective system gains control over the reactive system via several cognitive (e.g., response inhibition, shifting) and neural mechanisms (fronto-parietal network). However, this control is not absolute; hyperactivity within the reactive system can override the reflective system **...**" (Xavier et al., 2006).

both, and there might be many forms of 'complusions' that influence both the dialectical performance and the rhetorical effect. The broader than usual data-sampling was used to extend a primarily linguistic/pragmatic theory by tapping into the multi-modal spatio-temporal reality of arguing agents.

A third decision was to focus on intraindividual variation as well as the aggregated data of the two participants, so as to get some form of mapping of the dynamics of the televised exchange. With respect to gesture research, a significant difference is that the behavioural data are (at least comparatively) raw, as opposed to traditional gesture-analysis, that tends to transcribe gestures into a sign-system, or interprets them in the pragmatics of the activity, etc. As the focus was to study *internal dynamics of a debate,* the approach departed from standard discourse analysis methodologies. And much of standard psychology, too. As Molenaar noted a few years before the experiment started:

> Psychological processes like cognitive information processing, perception, emotion, and motor behavior occur in real time at the level of individual persons. Because they are person-specific, these processes differ from variables occurring in a population of human subjects—variables such as sex, socioeconomic status, or experimental condition (so called between-subject variables). Much psychological research is concerned with variation at the level of the population. However, whenever person-specific processes are involved, and in so far as these processes are nonergodic (i.e., obey person-specific dynamic models and/or have nonstationary statistical characteristics), their analysis should be based on intraindividual variation. (Molenaar & Campbell, 2009, p. 116.)

2.2 Context & Audience heterogeneity

Televised debates are elements of public discourse, some with significant impact in large populations. The televised debate is edited: the audience receives a more or less dramatized version of the actual debate (with cuts, text messages on screen, close-ups, or other media content). The audience of a televised debate is heterogeneous, and so are their viewing conditions (viewing angle, attention paid to either auditory or visual stream). This is the usual scenario, the 'input', that, on the long run shape what is often referred to as deliberative culture.

If we believe that argumentation has to go multi-modal (see e.g. Groarke 2014), it is unclear how much 'context' is relevant for a study of multi-modal strategic maneuvering? A dialectical analysis usually reduces context (and even what counts as content), but one can, in

principle, take into account the visual stream. As recent research shows, the concept of strategic manoeuvring can include the study of the non-verbal responses to fallacies and impolite exchanges, or responses to an opponent's nonverbal disagreement. In one of the pioneering papers examining strategies for responding to fallacious moves Weger and co-workers studied strong nonverbal indicators of disagreement during an opponent's speech „reconstructed as a rational response to the activity type in so far as it represents an attempt to rebut an opponent's arguments while the opponent is making them" (Weger et al., 2013, p. 196). The research ranked response types on perceptions of speaker likeability, and one of the noteworthy results was the heterogeneity of the audience. The statistical analysis showed that even with a careful experimental setup, the +/- rating of the 'move' differed among groups of speakers: „participants in the high verbal aggressiveness group rated the speaker in the ask moderator condition more positively on composure than in the headshake condition." (p. 193). The same study also found that "For participants in the low verbal aggressiveness group, the speaker was perceived to be significantly less composed in the ask moderator condition than in the direct request condition with no other pairs of conditions producing significant differences." (p. 193).

This type of research eminently pursues the study of a micro-level exposition of the multi-modal concept of strategic maneuvering. It creates an 'experimental scenario', and measures audience-response. The audience response of *populations* is the target, and generalization often remains an issue[10]. The aim is to find 'types' of people.

The research introduced in the paper assumed that participants in the research are not (just) sources of raw data, but also 'containers' of valuable observations and tacit knowledge. To construct a partial map from a holistic perspective, I tried not to narrow down the analysis to just one set of modalities, as this can lead to missing much of the interaction between the different modalities (Gelang, 2013).

To turn individual variation into a research asset it was considered that research subjects have various expertise, some through training, and some through tacit knowledge. When watching the video there was ample variation in what individuals considered as a salient feature of the behaviour of the participants, so it was assumed, that their various non-specialist expertise could be put to use. This approach situates the research in a contextual (and constructivist) didactic framework, and the aim is to utilize the individual differences between the perceivers.

---

[10] Consider (Seiter et al., 2009, p. 9), on whether data form one sex are likely to generalize to the other sex.

In the data-gathering phase individuals could focus on any well-delineable and repeated feature, where they assumed to have above-average expertise (eye- or hand-movement, actio-qualities, performance errors, or editorial decisions). These features were developed into traits (the specifics will be discussed in section 4.,)[11] and data was collected, registering the time of occurences of the tokens of the trait, and optional (additional) qualitative information. Each trait can be assigned to a channel (auditory/visual/editorial), and each registered token of a trait can be assigned to one of the speakers. Tokens can have other properties assigned to them, for example an auditory trait (Tone/Pitch Change) can have tokens with values (Up/Down), just as a visual trait (Audience; Other/Viewers, see Figure 4).

The units of the analysis are occurrences of various types of behaviours, not just utterances (e.g. complex questions), but also movements, gestures or performance errors. As students were not restricted on label use, some of the traits discussed later have unconventional labels, but they came with 2-3 line descriptions of the trait and specific notes on ambiguous cases, and decisions on whether to include or not similar instances. One result of the research is that it explicates some 'lay' assumptions of a generally (partially) tacit knowledge-domain.

The 'trait-analysis' exercise took intuitively significant traits, specified them, and registered occurrences of the tokens (but generally not the duration). The methodology to trace changes in argumentation-related activity allowed for mapping the dialogue in multi-dimensional space, as each observer contributed to a unidimensional description of the diadic interaction. In this dialogue-rendering the 'events' are not simply ordered, put in a sequence, but are temporally positioned. The tokens can be analysed with respect to their alignment with other tokens (with a grain size of 1 second). The distribution of tokens can also be studied in given segments of the dialogue, and 'phenomena' can include frequency changes, or various patterns (co-occurrence, inhibition-excitation). In the discussion of the data (Section 4.) there are examples both for topical segmentation (dialogue-segments of a discussion around a specific topic/issue), and for artificial segmentation (8 minute segments of the dialogue). The specific (operationalized)

[11] Traits are used to give an externalized reconstruction of artefact-human knowledge-mobilization processes leading to – among other things – belief-revision, changes in mental states. This analysis is theoretically linked to the 'trait'-analysis of scientific theories in another paper (Zemplén 2017), where I used 'traits' of Newtonian diagrams to show the heterogeneous uptake of the theory and to argue that to reconstruct the epistemic content of a theory we are not justified in neglecting the investigation of the pragmatic and rhetorical functions of visuals.

traits included less complex linguistic data (speech-breaks, marked changes in eye- or hand-position, pitch, or speech-speed), often quite numerous, as well as categories in linguistic theories (fallacies, abusive adjectives, *ad hominems*), often single digit occurrences.

Was the data gathered by professional experts? No. To ascertain that an utterance constitutes a specific fallacy, or to prove that an interlocutor has violated some norm of a process for 'rational resolution' requires contributory expertise on PhD level to be expected. Already the correct use of key concepts of pragma-dialectics is a specialist expertise that needs several semesters of exposure and practice to develop. But there is growing evidence that norm-violations are detectable by average citizenry, and some of the research participants (advanced M.A students) might be considered fairly reliable sources of information.

To cash in on individual differences, the module utilized 'instinctive' reactions to what was perceived and noted. Assumedly when participants were asked to pick the 'features' they tended to favour domains where their perceptual accuracy or grain size in expression was better. That is, people prefer 'traits', where they are good at picking the tokens (they have highly granular experiences). In moderately sized groups (20-40) it is expected that *some* students have fairly well developed discourse-analytic skills, and they tend to pick features that link to these skills. In short the data-collectors were not treated as reliable sources of data, providing expert analysis, but as members of the audience, who pay much more attention to some feature of the dialogue than an average viewer, and who are presumably better at registering occurrences of a feature than an average viewer, especially after three or four expositions.

The aim of the exercise was to develop a novel teaching tool to foster the uptake of content knowledge of argumentation, and to utilize existing tacit knowledge, inspired by work on types of expertise, and so called 'transmuted' non-specialist expertise (Collins & Evans, 2007). I now turn to a short description of the didactic scenario of the research, before returning to how the mapping game extends the *context* of the dialogue, incorporating various sources of data (ranging from editing practices to non-verbal behaviour or performance-errors during speech).

3. THE DEBATE ANALYSED

The same debate can be analysed in several ways, and below is a chart of the 'local contexts' of the module development. The time allotted to the analysis of the interview was between 30 – 60% of class time

(14x180 min). The grid below focuses on the 'content' of teaching in which the module was embedded (Table 1, from 2013 to 2018).

Year	Course aim	Didactic theme	Male %	Literature (chapter-presentation) + Reading	Student output
1	Norm violations in PD	breakdown of a rational discussion	36,84%	Fundamentals of argumentation theory: Mahwah, N.J.: L. Erlbaum. 1996	Groupwork, co-authored analysis (140 000 n) srt file (200 lines)
2	behavioural analysis	link behavioural analysis with normative models	25,00%	Selected essays in Rhetoric, Informal Logic	multi-trait description, essays
3	Meta-analysis	check the data, improve analytical skills	15,00%		individual research projects
4	Content-Knowledge	you think you know about argumentation?	7,41%	Handbook of Argumentation Theory, Springer, 2014.	short research plans
5	Nonverbal Communication	do you know how you move?	21,43%	APA Handbook Nonverbal Communication, 2016	short research plans
6	Dual and Group Processes	'Sleep on it!' : what strikes you second	15,79%	Blackwell Handbook of Social Psychology: Group Processes, 2008	'Second view' notes, individual research essay and presentation

Table 1 – Table of module runs, year 1-6.

The target set for Year 1 was a detailed study of the interview with a topic- and move-analysis of the debate. To assemble the document, first pairs of observers located fallacies / derailments / non-admissible questions in up to four minutes of the video. After the preliminary analysis working groups focused on 1) The institutional setting, media landscape; 2) Heterogeneous message-design: the range of addressees (also looking for conflicts/inconsistencies) 3) The asymmetries of the debate, gestures, meta-communication and argumentation. This document (around 140 000 n) was made available for students in later years at specific points in the module. Several norm violations were found, and recurrence of certain 'moves' suggested that the reporter manipulates and distorts the position, polarizes the debate.

A group in Year 1 produced an srt file (with over 200 entries and just under 2000 words) that could be viewed together with the original video[12]. As the original version was found too dense for viewing, Year 3 produced an abridged version (appr. 120 lines).

For Year 2 the debate was approached from a neo-behaviouristic perspective. The didactic reason for the exercise was to develop and assess observations skills, and to highlight the problem of categorisation for the analyst. Before discussing the outcome of the exercise I outline the multi-trait approach.

As described previously, the collaborative research to mapping argumentative exchanges utilized a data-sampling method in which observers individually picked perceived features of the argumentative performance of participants in the televised discussions after exposure to the videotaped Csermely-Fábri debate, a 20 min. latency period, and a request to pick (name and describe) a relevant feature that the individual thinks has about 5-20 occurrences in the dialogue. So an 'artificial' filter was introduced for the 'feature' selection, based on an estimate on the number of occurrences (the number of data points). The features were developed into 'traits' by the individuals, and the eventually developed 'trait-definitions' often significantly deviated from the original 'feature'[13].

A contextual-model was adapted to teaching argumentation in the hope that methodological reflection, training of observational skills help transmute non-specialist expertise to specialist expertise in some participants. The gathering of less complex data (speech-breaks, marked changes in eye-position, hand-gestures), often yielded a rich (and hence

---

[12] https://www.dropbox.com/s/20e8onutsqcy3mc/Versus%20-%20Mi%20folyik%20az%20egyetemen-%20-%20Hallgat%C3%B3i%20szerz%C5%91d%C3%A9s%20helyett%20tand%C3%ADj-%20-%202013.02.24-360p.mp4 SRT file currently in Hungarian.

[13] Participants had to develop a one page 'operationalisation' of the category, as counting occurrences hinged on 'precisely' how the feature was described.

cumbersome) harvest (Figure 2, Table 2), which helped students appreciate personal differences in memory-distortion, their own observer biases and various general methodological issues.

Audio/Video * 8 min. segments * Crosstabulation

			8 min. segments			
			1,0	2,0	3,0	Total
Fábri	Audio/Video	A	32	46	45	123
		V	86	91	65	242
		V*	1	1	0	2
	Total		119	138	110	367
Csermely	Audio/Video	A	44	53	44	141
		V	81	65	66	212
		V*	5	1	2	8
	Total		130	119	112	361
Total	Audio/Video	A	76	99	89	264
		V	167	156	131	454
		V*	6	2	2	10
	Total		249	257	222	728

Table 2 - The number of data points in 8 minute segments of the dialogue (the data points are assigned to channels: Auditory, Visual, and Editorial (V*, OTS)).

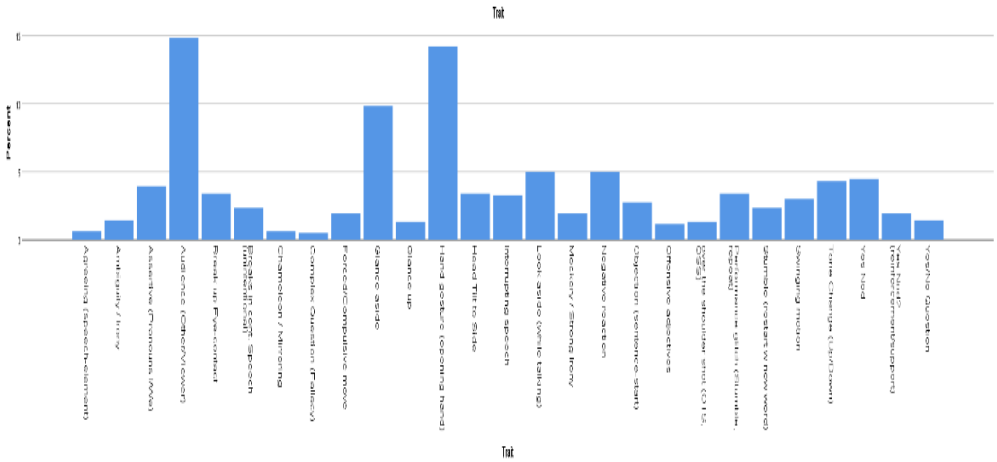

Figure 2 - The distribution of the gathered over 700 data-points (trait-frequency).

4. RESULTS AND MULTI-TRAIT MAPS

Although absolutely no precautions were taken, the mapping exercise produced a rather 'balanced mapping'. Parsing the dialogue into topics of various lengths (established in Year 1), the average resolution of the data is roughly 2 seconds / data point for all of the topics, with an average 0,50838 trait/s (Table 3).

Topic start (s)	0	82	278	431	724	985	1199	1280	1400
stop (s)	75	276	424	718	975	1197	1279	1397	1432
Topic net (s)	75	194	146	287	251	212	80	117	32
trait/s	0,44	0,58	0,54	0,50	0,57	0,45	0,63	0,47	0,53

Table 3 – Average data for a given topic / sec.

This so far suggested no major failure in design (e.g. coders get tired and/or lazy by the end of the coding exercise). In spite of this relatively even distribution of the data, several of the traits had uneven token-distributions, and this was also true for a number of traits with large numbers of tokens (Figure 3). The preliminary analysis of the aggregated data of the two speakers showed that trait-saturation might be an interesting property for the investigation.

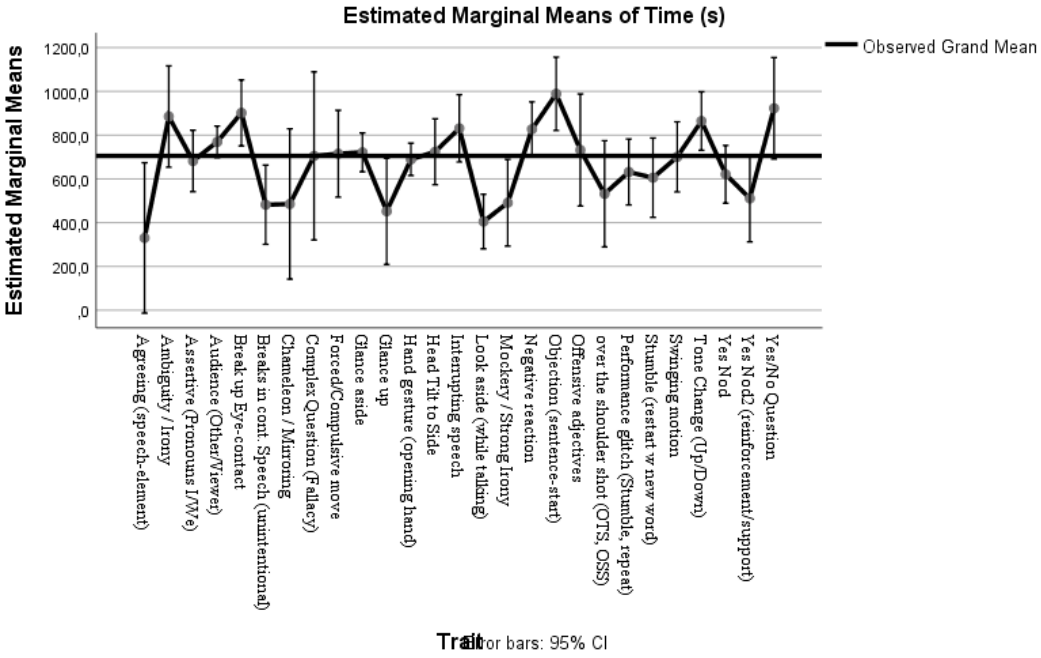

Figure 3 - Estimated marginal means (SPSS 25.0)

A good starting point for the analysis of the functional complexity is the list of 'early' and 'late' traits (Table 4):

Trait	Mean	Trait	Mean
Agreeing (speech-element)	330,200	Objection (sentence-start)	988,762
Look aside (while talking)	405,026	Yes/No Question	922,727
Glance up	452,100	Break up Eye-contact	901,462
Breaks in cont. Speech (unintentional)	482,056	Ambiguity / Irony	884,818
Chameleon / Mirroring	485,400	Tone Change (Up/Down)	864,394
Mockery / Strong Irony	491,000	Interrupting speech	830,960

Table 4 - Means for early and late traits.

Note that the data is structured, and so some traits appear to be balanced on Figure 3, like picking the audience (Other party/Towards viewers), but of course to map further imbalances an in depth analysis can assign the speakers to the tokens, and any of the specific values that the token has, not just the time of occurrence (Figure 4).

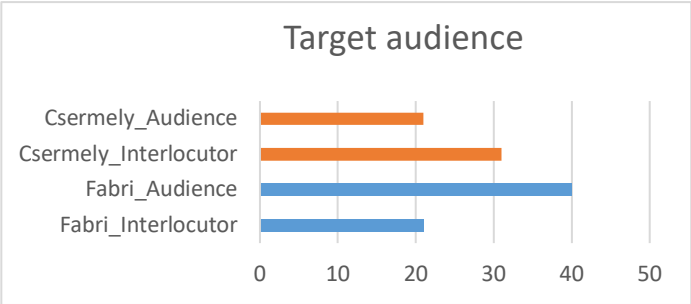

Figure 4 – The number of tokens of the trait 'Audience' in the dialogue, assigned to speaker and target 'Towards viewers/Other party'.

In the rest of this section, I show two ways of visualizing the results, and use the data aggregated for the speakers, keeping the analysis of the individual speakers (profiling) to a minimum (section 4.3). Can we assume that aggregated data (for both speakers) can be informative on their own right? If it makes sense to talk about intraindividual changes, the dialogue's aggregate (A/V) data shows the combined dynamics of the speakers.

First I outline the high temporal resolution visualization that maintains individual data-points (with the times of occurrences) in *dialogue bundles*. The second approach parses the data-set and assigns items to dialogue-fragments, that allow for more traditional analysis and comparison of sets of data.

4.1 Dialogue bundles

One way to study a dialogue is to map recurring elements in a category, and see how they are distributed over time. A simple visualization of the aggregated data of numerous traits (with 20+ tokens) shows full saturation (x=1) at time of the appearance of the last token, in seconds (Figure 5).

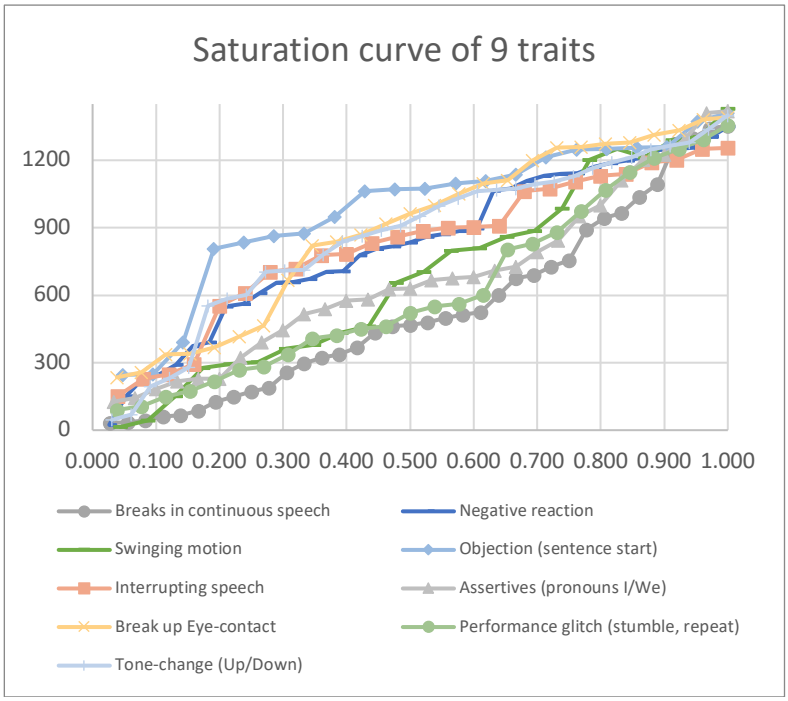

Figure 5 – Saturation of traits with large number of tokens, the backbone of the dialogue bundle. X axis: saturation (100%=1,000), Y axis: time (in seconds, 5 minute grid)

To make the meso-level dynamics more visible, a visualization below shows, how 'deviant' a token is (assuming equal spacing). The deviation from expected occurrences for traits with non-linear saturation curve shows how quick saturation phases ('bursts') deviate to the right, and stagnation (no occurrence) phases deviate to the left (Figure 6).

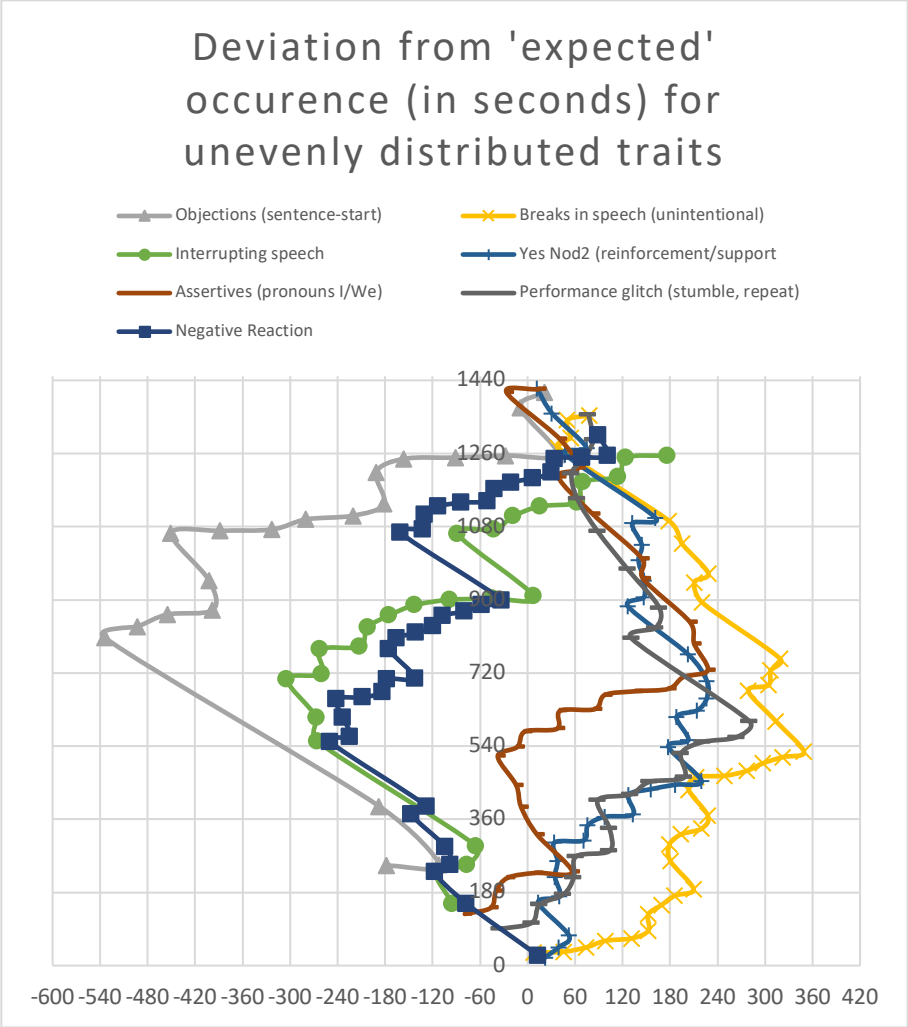

Figure 6 – Finer grained topography of the bundles showing the deviation from expected position assuming equal spacing of tokens (in seconds). X axis: deviation (in seconds, one minute grid). Y axis: temporal position of token (3 min. grid). Quick saturation phases deviate to the right, stagnation (no occurence) drifts to the left.

Some traits, like unintentional breaks in speech deviate to the right (many early occurrences), objections, negative reactions deviate to the left (most pronounced near the end of the debate). Once we look at the distribution of the less numerous traits, we find similar patterns: 'mockery and strong irony' drop in the second half of the debate, as well as nods of reinforcement and support, while 'Yes/No Question' increases, as does 'Ambiguity / Irony' (around midtime).

In this 'embodied' perspective the debate is some *process*. With the dialogue bundles both verbal and non-verbal presentational devices can be mapped, as well as compulsions, and we might be one step closer to mapping meso-level debate-dynamics. The 'tone' of the debate is bound to be reflected both on the utterance level, and on the level of hand gestures. But it would certainly be great if we could find same way of parsing or segmenting the data in time. Can we find sets of behaviours that travel together easily (something like speaker profiles, 'personas')? From the second year various working units supported the hypothesis that speakers can switch or tweak their style of interaction, including the frequency of derailments. It would be helpful to locate triggers that can be linked to the emergence of these profiles. See Figure 7 and segment (1) in the forensic / apologia phase of the debate (blaming the leadership of the University for student abuse of data on other students).

Figure 7 – Assignment of blame (Csermely, facing viewers)

(1) "…és esetleg több évfolyamnak megtakaríthatták volna ezt a rendkívül kellemetlen és megalázó élményt, amit nagyon sokan most élnek át…"

> "..and you might have saved several yeargroups this extremely unpleasant and humiliating experience that many people experience right now…"

So next I turn to a research exercise on whether episodes of the dialogue can contribute to some of the deviations in the dataset.

4.2 Towards delineating dialogue profiles

Studying changes during longer interviews may help to map 'phases' of the dynamics. Could specific triggers be responsible for changing the 'tone' of the debate, something like switching 'speaker profiles', or 'shifting gear'. For initial hypothesis-generation Yeargroup 2 was divided into four working units (N, E, S, W), and received the aggregated data (an excel file with all traits and the times of trait-occurrences as registered by peers) printed in one copy, and had access to the trait-descriptions. Two groups from the four differentiated tokens as belonging to speakers (N, S), while two groups disregarded the source (E, W). Two groups were instructed to first focus on reflective/dialectical/verbal traits, in general the higher level, more theoretical categories (S, W), while two groups focused first on somatic, reflexive traits, lower level traits (N, E).

Each working unit therefore started from a particular perspective with respect to data handling, and they could assume interaction across dimensions, or study simply frequency changes. The initial hypotheses were refined in class, discussed in the group, and as homework the groups handed in protocols of their hypotheses and data. Most units generated hypotheses that were concerned with the temporal distribution of tokens. Some of the noticed changes were:

- Given 5 minute intervals there is a marked increase of registered non-verbal traits between minutes 10-15, and a gradual decrease between minutes 15-20.
- After the 8th minute, coinciding with a change in topic, there are changes in tone and forced (unintentional) movement (increases from 5 to 12 and 2 to 13 in toto), and stops/breaks in speech as well as looks aside (decrease from 30 to 23 and 25 to 13 in toto).
- A specific glance aside by the interviewed and the fact that in the 15th minute both speakers talk for seconds at the same time was one of the most interesting focal points that students picked. There is a significant change in the frequency of some traits after this 'dual-talk' episode of the interview (+ reinforcement drops for reporter (from 0,33 to 0 / min) and interviewed (from 0,72 to 0,11), - reinforcement only increases for reporter (from 0,13 to 0,55 /min.).

Of the various possible artificial segmentations of the dialogue a simple trisection was used (8 minute fragments, Table 2, Figure 8). About half of the registered breaks in eye-contact and over half of the 'objections' took place in the last third of the debate. In the eristic (last) third of the dialogue performance errors dropped.

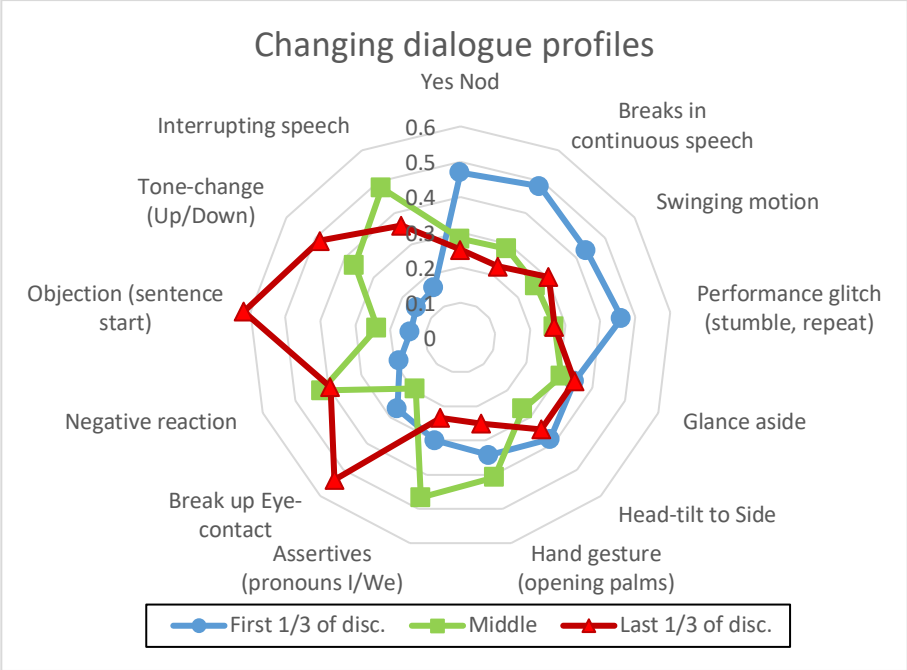

Figure 8 – The 8-minute segmentation of the dialogue. High frequency in a given 8-minute period is a point near the periphery, low frequency is a point close to the center.

Whether or not specific triggers might be responsible for changing the 'tone' of the debate, the temporal segments are quite different. Figure 8 shows how certain tokens of the traits tend to cluster in certain phases of the debate. To the right a number of traits possibly linked to stress-relief (dominant in the first third of the debate), to the left traits of agonistic dialogues.

4.3 A multi-trait map of a 24 minute interview

To assist micro-context analysis, the suitable resolution may depend on the actually investigated phenomena. Below is a representation of the debate, using a 3 minute grid and 50 sec intervals. Previously highlighted episodes occur at 380 sec and around 900 sec, possible

turning points for tweaking performance. The raw data-set allows one to look for the dynamics, the temporal development that link non-verbal behaviour to some linguistic traits, for example two types of irony: 'soft' (ambiguity) and 'strong' (mockery, ridicule).

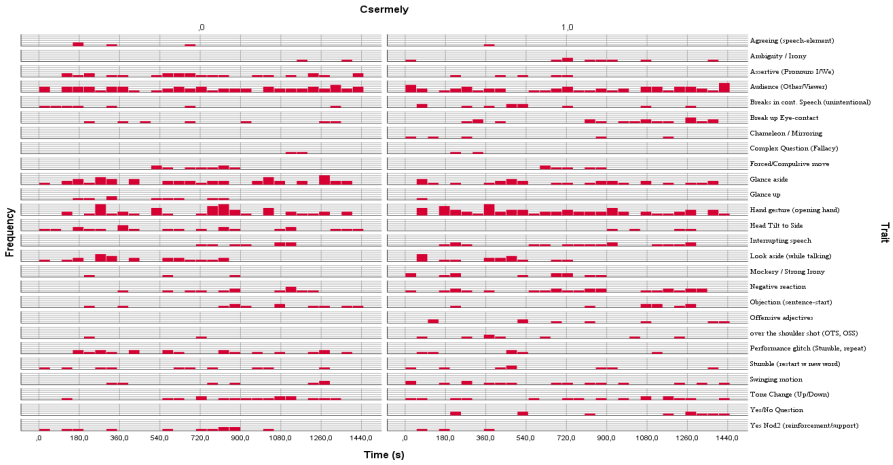

Figure 9 – Multi-channel map of the dialogue (speakers individuated, with 'raw' trait data, Fábri left, Csermely right; 50 sec. resolution, 3 min. grid).

5. A ROAD LESS TRAVELLED?

Being aware of the 'experimenter's regress', the present contribution aimed to prove little in a theoretical sense. The exploratory module and the trait-registering exercise relied on a broader range of inputs than traditional discourse analysis, linked to the growing need to develop tools for multi-modal argument analysis and assessment (Tseronis 2018). Potentially of theoretical and practical significance, the method discussed in the paper outlines an approach that might be an asset to micro- and meso-context analysis, and also raises some issues on how to move from one context to another.

The multi-channel mapping of a debate is part of an attempt to develop a relational (as opposed to reductionist) approach to functional arguing organisms which allows the study of those qualities that we are trying to learn about, and not only those that we have the best structural descriptions of. To locate recurring somatic responses, language-related gestures, as well as various types of linguistic phenomena non-specialist expertise of the research participants was utilized. Some parsing of speech elements using abstract theoretical concepts was incorporated

in the analysis, providing a partial map of elements in an analytical overview.

The temporal scale of the study is in seconds, not years, but if it is possible to provide a rich in detail analysis of a 'real-time' debate, the approach can be fit to comparative studies (e.g. televised presidential debates to compare long term changes in particular argumentative cultures), and the study of long-term (macro-level) trends in deliberative cultures. Before it's too late.

ACKNOWLEDGEMENTS: I thank all the participants in the research, acknowledging their contribution, and the help from Erika Hlédik, Mihály Héder as well as fruitful discussions with Kristóf Kovács, Dorottya Egres, and the participants of the European Network for Argumentation and Public Policy Analysis (COST Action CA17132 – APPLY).

REFERENCES

Bogen, J., & Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review, 97*(3), 303-352. DOI:10.2307/2185445.

Collins, H. M.& Evans, R. (2007) *Rethinking Expertise*. Chicago: University of Chicago Press.

Gelang M. (2013) Towards a Political Actio. In: Poggi I., D'Errico F., Vincze L., Vinciarelli A. (eds) *Multimodal Communication in Political Speech. Shaping Minds and Social Action*. PS 2010. Lecture Notes in Computer Science, vol 7688. Springer, Berlin, Heidelberg.

Gelang, M, & Kjeldsen, J. (2011). Nonverbal communication as argumentation. In *Proceedings of the 7th conference of the International Society for the Study of Argumentation*, ed. F.H. van Eemeren, B. Garssen, D. Godden, and G. Mitchell, 567–576. Amsterdam: SicSat.

Gervain J & Zemplén G. A. (2005). "Focus Raising: A Paradigmatic Example of the Treatment of Syntactic Microvariation," in *Syntax and Variation*, ed. Cornips L. & Corrigan K. P. Amsterdam: John Benjamins.

Groarke, L. (2014). Going Multimodal: What is a Mode of Arguing and Why Does it Matter? Argumentation, 29(2), 133–155. DOI:10.1007/s10503-014-9336-0

Keevallik, L. (2018). What does embodied interaction tell us about grammar? Research on Language and Social Interaction, 51, 1–21. DOI:10.1080/08351813.2018.1413887

Kertész, G. & Zemplén, G. Á. (2011). The challenges of training critical discussants: Dialectical effectiveness in strategic maneuvering and in Science Education. *Proceedings of the 7th International Conference on Argumentation of the International Society for the Study of Argumentation*. Amsterdam, Sic Sat. 2070-2082.

Eemeren, F. H. van & Houtlosser, P. (2009). Strategic maneuvering. In F. H. van Eemeren (Ed.), *Examining Argumentation in Context: Fifteen studies on strategic maneuvering* (pp. 1-23), Amsterdam: John Benjamins

Eemeren, F. H. van & Grootendorst, R. (2004). *A Systematic Theory of Argumentation. The Pragma-Dialectical Approach.* Cambridge, England: Cambridge University Press.
Eemeren, F. H. van (2010). *Strategic Maneuvering in Argumentative Discourse*. Amsterdam: John Benjamins.
Kruglanski, A. & Higgins, E.T. (2007). *Social Psychology: Handbook of Basic Principles (2nd edition).* New York: Guilford Press
Mercier, H. & Sperber, D. (2009) "Intuitive and reflective inferences". In Evans, J. St. B. T. and Frankish, K. (Eds.) *In Two Minds: Dual Processes and Beyond*. Oxford University Press.
Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, 18(2), 112–117. DOI: 10.1111/j.1467-8721.2009.01619.x
Lieberman M. (2003). Reflexive and Reflective Judgment Processes: A Social Cognitive Neuroscience Approach. in *Social Judgements: implicit and explicit processes*. Eds. Forgas, Williams, von Hippel. Cambridge UP.
Rosen, R. (2000). *Essays on Life Itself*. New York: Columbia University Press.
Seiter, J. S.; Weger Jr., H., Kinzer H,J, & Jensen A.S. (2009) Impression Management in Televised Debates: The Effect of Background Nonverbal Behavior on Audience Perceptions of Debaters' Likeability, *Communication Research Reports*, 26:1, 1-11, DOI: 10.1080/08824090802636959
Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–726.
Tseronis, A. (2018) Multimodal argumentation: Beyond the verbal/visual divide. *Semiotica* 2018 (220):41-67.
Xavier N., et al. (2006) The Neurocognitive Mechanisms of Decision-making, Impulse Control, and Loss of Willpower to Resist Drugs *Psychiatry* (Edgmont). May 2006; 3(5): 30–41.
Weger Jr., H, Seiter, J,S,, Jacobs, K.A., & Akbulut, V. (2013) *Responses to an opponent's nonverbal behavior in a televised debate*. Audience perceptions of credibility and likeability. Journal of Argumentation in Context 2:2, 179–203. doi 10.1075/jaic.2.2.01weg
Zemplén, G. Á. (2014) From theoretical concepts to behavioural trait-analysis: utilizing tacit knowledge and transmuted expertise in Facebook-coordinated courses on Rhetoric and Dialectic. In *Ten years of Facebook: Proceedings of the Third International Conference on Argumentation and Rhetoric*, edited by G. Horváth, R. K. Bakó and É. Biró-Kaszás. Nagyváraden: Partium Press.
Zemplén, G. Á. (2019) Diagrammatic Carriers & The Acceptance of Newton's Optical Theory. *Synthese*. 196:9, 3577–3593. DOI: 10.1007/s11229-017-1356-5

APPENDIX

Upon request more data can be supplied on the multi-trait exercise.