

CROSS-LINGUISTIC PATTERNS OF META-DISOURSE: DISCIPLINARY SIMILARITIES AND SECTION-BASED DIFFERENCES

Djuddah Leijen, Helen Hint, Helena Lemendik,
Baiba Egle, Anna Ruskan, Christer Johansson

Abstract. This study examines metadiscourse markers across a corpus of Estonian and Lithuanian journal articles in the field of linguistics. We aim to 1) compare the global use of all the metadiscourse markers across the languages and texts, making distinctions between these languages and specific academic journals, and 2) to discern whether similar and/or different patterns can be identified across the languages and whether such patterns also manifest across various academic journals. We find that Estonian writers use self-mentions more frequently in *methods* sections than Lithuanian counterparts. Comparing journals, the Lithuanian journal *Kalbotyra* shows more transition markers, code glosses, and endophoric markers, while the Estonian *ERÜ aastaraamat* relies more on transition markers in results and discussion sections. Despite discipline similarities, variations emerge in specific sections and interpersonal categories across languages and journals. The study provides insights into metadiscourse patterns and their role in different languages and academic contexts, offering potential guidance for future research and practice in non-English academic writing.*

Keywords: metadiscourse, research articles, IMRaD, Estonian, Lithuanian

1. Introduction

In journal articles, metadiscourse markers play a crucial role in establishing a balance in the discourse, aligning the writer's intentions with the reader's perceptions and interactions within the propositional content of the text. The concept of metadiscourse encompasses various linguistic expressions that not only convey the primary content of the text but also guide and engage readers through the discourse, addressing their needs and facilitating both the coherence and cohesiveness

* This work is supported by the project "Bwrite: Academic Writing in the Baltic States: Rhetorical Structures through culture(s) and languages" (EMP475), funded by Iceland, Liechtenstein and Norway through the EEA Grants and Norway Grants.

The colored versions of Figures 1–4 have been published in the online edition of the article (<https://doi.org/10.5128/ERYa20.07>).

of the text (Crismore et al. 1993, Mauranen 1993, Vande Kopple 1985). The use of metadiscourse markers, while well documented in English, also presents notable patterns and variations across different languages (Dahl 2004, Fløttum et al. 2006, Mur-Dueñas 2011), genres (Ädel 2018, Hyland et al. 2022), and disciplines (Birhan 2021, Hyland et al. 2022). Hyland (2005) has been pivotal in shaping our current understanding and application of metadiscourse markers, particularly through the application of his interpersonal model of metadiscourse, which has been widely recognised and applied in studies investigating English academic writing, but also across a variety of other languages (Cao, Hu 2014, Peng, Zheng 2021).

The interpersonal model (Hyland 2005), see also Table 1, bifurcates metadiscourse into the interactive and the interactional dimensions, each serving a distinct yet intertwined role in creating academic discourse. The interactive dimension, encompassing transitions, frame markers, endophoric markers, evidentials, and code glosses, serves to guide the reader through the text. Conversely, the interactional dimension, which includes hedges, boosters, attitudinal markers, self-mentions, and engagement markers, mirrors the intention of the author to involve the reader in the text (Hyland et al. 2022, Hyland, Tse 2004). As such, at the macro level, i.e., at the whole text level, the use or lack of use of these metadiscourse markers can offer a snapshot of how the text and the author interact with the reader and, more specifically, where in the text, as has been highlighted by (Ruskan et al. 2023). Building upon this framework, the present study empirically explores the use of metadiscourse markers across two languages: Estonian and Lithuanian, found across various published journal articles in the discipline of linguistics to 1) compare the global use of all the metadiscourse markers across the languages and texts, making distinctions between these languages and specific academic journals, and 2) discern whether similar and/or different patterns can be identified across the languages and whether such patterns also manifest across various academic journals.

Exploring metadiscourse markers across different languages and academic journals requires a closer examination of the linguistic and discursive choices embedded within the editorial expectations of each respective journal context. Regarding journal articles subject to stringent editorial and language requirements, selecting texts that successfully pass through the peer review and editorial process may be influenced. Consequently, these texts may exhibit similarities in style, thereby validating the survivorship bias that can arise when we choose to investigate texts that have undergone a rigorous process of writing, reviewing, revising, and editing before publication. This process, particularly when assessed by a small group of stakeholders, may exhibit a specific metadiscursive footprint. Moreover, studies have indicated that stylistic expectations in language communities other than English are often higher, assuming a very high language proficiency and rhetorical style representative of the national culture and/or academic/discursive culture (Duszak 1994, Duszak, Lewkowicz 2008, Harbord 2018). Therefore, while we may have some understanding of the intricacies involved in constructing texts, the editorial process of journals, and the complexities of national cultures and languages, the evaluation of these metadiscoursal patterns at the macro level (i.e., patterns observed across the whole text), and the meso level (i.e., patterns observed across sections of a text) remain underexplored (see also Leijen et al., Forthcoming). Furthermore,

to our knowledge, no studies have compared how these patterns compare across languages and across journals published in other languages. Evaluating these patterns at the macro- and meso- levels across languages and journals may help us better understand whether the interpersonal model proposed by Hyland, specifically the bifurcation interactive and interactional dimensions, show distinctive editorial differences and/or language-related differences.

This study builds upon a prior investigation (Hint et al. 2022, Ruskan et al. 2023), which identified metadiscourse markers across 21 journal articles (seven journal articles belonging to three specified journals in the field of linguistics). This work is part of a larger project aiming to uncover the rhetorical structures of academic texts in Estonian and Lithuanian (Jürine et al. 2021, Leijen et al., Forthcoming). For this study, we combined the research on the use of metadiscourse markers across journal articles in Lithuanian (30 journal articles belonging to three specified journals in the field of linguistics) with the previously mentioned prior investigation of Estonian metadiscourse markers. The goal is to compare metadiscourse patterns of Estonian and Lithuanian and to assess the metadiscourse usage in comparable academic journal articles in the field of linguistics across these two languages. Additionally, this research endeavours to identify potential universal metadiscourse strategies that may transcend these languages and academic journal contexts. By offering a nuanced, cross-linguistic, and cross-academic journal perspective, this study contributes valuable insights to the existing knowledge on metadiscourse in languages other than English. These insights may potentially inform and enhance future research and practice.

2. Data and method

The data used for this study comprises two languages, Estonian and Lithuanian, collected and coded as part of a larger research project investigating writing conventions and rhetorical structures, that is, writing traditions (Hint et al. 2022, Jürine et al. 2021, Leijen et al., Forthcoming). For this study, a corpus of academic journals was used across the two languages each contributing to a comprehensive analysis of all the metadiscourse markers contained and identified by Hyland's interpersonal model of metadiscourse (Hyland 2005). For a complete overview of the metadiscourse markers, see Table 1, and for a complete description of metadiscourse markers in Estonian, see (Hint et al., submitted). The Estonian sub-corpus comprises 21 journal articles with seven articles each from three journals: *Keel ja Kirjandus* ('Language and Literature'), *Eesti Rakenduslingvistika Ühingu aastaraamat* ('Estonian Papers in Applied Linguistics'), and *Emakeele Seltsi aastaraamat* ('Yearbook of the Estonian Mother Tongue Society'), totalling 89,224 words. The Lithuanian sub-corpus, somewhat larger than the Estonian sub-corpus, with a total word count of 136,443, encompasses 30 articles, ten from each of the following journals: *Kalbotyra* ('Linguistics'), *Lietuvių kalba* ('The Lithuanian Language'), and *Taikomoji Kalbotyra* ('Applied Linguistics').

Table 1. An interpersonal model of metadiscourse by Hyland (2005: 49)

Category	Function	Examples
Interactive	Help to guide reader through the text	Resources
Transitions	express relations between main clauses	<i>in addition; but; thus; and</i>
Frame markers	refer to discourse acts, sequences, or text stages	<i>finally; to conclude; my purpose is</i>
Endophoric markers	refer to information in other parts of the text	<i>noted above; see Fig; in section 2</i>
Evidentials	refer to information from other texts	<i>according to X; Z states</i>
Code glosses	elaborate propositional meanings	<i>namely; e.g.; such as; in other words</i>
Interactional	Involve the reader in the argument	Resources
Hedges	withhold commitment and open dialogue	<i>might; perhaps; possible; about</i>
Boosters	emphasize certainty or close dialogue	<i>in fact; definitely; it is clear that</i>
Attitude markers	express writer's attitude to proposition	<i>unfortunately; I agree; surprisingly</i>
Self-mentions	explicit reference to author(s)	<i>I; we; my; me; our</i>
Engagement markers	explicitly build relationship with reader	<i>consider; note that; you can see that</i>

To ensure a thorough and accurate representation of metadiscourse markers across the languages and journals, the initial phase of the data annotation was close reading and manual annotation of a small portion of academic journal articles in the field of linguistics from the corpus. Furthermore, both language teams negotiated their language-specific understanding of what constituted metadiscourse markers in their languages based on a shared understanding of English language descriptions (see Table 1). The annotated markers spanned various linguistic units, including constructional, lexical, and grammatical elements, as well as punctuation marks. Subsequently, a meticulous manual annotation phase was performed, which involved a detailed review and, where necessary, adjustment of the annotated data (for a comprehensive overview of the full data coding process see Hint et al., Forthcoming). The manual annotation was conducted by at least two annotators for each language. To ascertain the reliability of the annotations, Cohen's kappa interrater reliability measures were applied, revealing a high degree of agreement between annotators across all categories (0.900 and above for both languages). Any discrepancies and disagreements encountered during this phase were collaboratively discussed and resolved, ensuring a refined and consistent annotation across the dataset, thereby bolstering the reliability of the corpus of metadiscourse across the two languages.

Given the objective to map the overall pattern of metadiscourse use on the macro level (i.e., across the whole journal article), the text section variable was annotated based on the classic IMRaD structure of a research article (Sollaci, Pereira 2004) as much as the articles would allow. However, not all research articles in our corpus adhered to a clear IMRaD structure. For example, in many cases, in addition to the *introduction*, the journals would have an additional section to cover the theoretical framework labelled *literature review* in Estonian journals or labelled *theory* in Lithuanian journals, ostensibly covering the same type of information. Additionally, in some instances the *results* and *discussion* sections were combined

into a single section, *results and discussion*, serving both functions. Therefore, and regardless of the actual section headings in the research article, we used the unified labels ‘*introduction*’, ‘*literature review/theory*’, ‘*method*’, ‘*results and discussion*’, and ‘*conclusion*’, as much as possible. To conduct a comparative analysis across the journal articles contained within the languages, and across the languages, we collapsed some of the journal article sections. For example, in some instances where there was a separate section for *results* and *results and discussion*, the *results* would be included in the *results and discussion* section. Furthermore, some sections which were coded in the original corpus were excluded from the analysis of the cross-language and cross-journal article analysis, such as *titles*, *footnotes*, and *acknowledgments*. Table 2 provides an overview of the included sections for analysis of the language specific journals.

Table 2. Journals and included sections for comparative analysis

Journals		Sections
Estonian	<i>Keel ja Kirjandus</i> (KK)	Introduction; Literature Review; Method; Results; Results & Discussion; Discussion; Conclusion
	<i>Eesti Rakenduslingvistika Ühingu aastaraamat</i> (ERÜ)	Introduction; Literature Review; Method; Results; Results & Discussion; Discussion; Conclusion
	<i>Emakeele Seltsi aastaraamat</i> (ESA)	Introduction; Literature Review; Method; Results; Results & Discussion; Discussion; Conclusion
Lithuanian	<i>Kalbotyra</i> (K)	Introduction; Theory*; Data and Method; Results, Results & Discussion; Discussion; Conclusion
	<i>Lietuvių kalba</i> (LK)	Introduction; Theory*; Data and Method; Results & Discussion; Conclusion
	<i>Taikomoji Kalbotyra</i> (TK)	Introduction; Theory*; Data and Method; Results & Discussion; Conclusion

**Theory*, in the Lithuanian data, is the same as *literature review* in the Estonian data.

To provide a more in depth understanding of the articles within the specified journals and to elucidate the extent to which variations can be accounted for, we provide information of the journals pertaining to any editorial requirements. This includes highlighting the explicitness of editorial guidelines related to accepted article types and anticipated styles, the provision of editorial assistance, and the extent of editorial editing post-article acceptance. This information serves to ascertain whether editorial expectations might instigate a particular survivorship bias within published articles, and subsequently, within the language-specific corpus.

We used association plots to determine the patterns of the use of metadiscourse markers across the languages and across the journals. The *vcd* package (Meyer et al. 2003) in the free statistical software R (R Core Team 2022) was used for conducting the analysis and creating association plots. We further used the *reshape2* package (Wickham 2007) and *ggplot2* package (Wickham 2016) to further visualise the cross-linguistic and cross-journal article comparisons with heatmaps. Specifically, the heatmaps facilitate the analysis of the Pearson residuals presented in the association plots, revealing deviations from the expected statistical

independence of rows and columns, representing the sections (IMRaD type and variations) and metadiscourse markers in this study, respectively. The plot visualises the association and dissociation between the journal article sections and the metadiscourse category. To further highlight these variations and compare any editorial variation between the journals for each language, a baseline model was calculated. The baseline model basically consists of the combined journal articles representing a metadiscourse model, which you could expect when you observe the use of metadiscourse across a larger dataset. The three separate journals in the Estonian and Lithuanian sub-corpus are compared to the baseline metadiscourse model to determine the location of variation (for example, variation of metadiscourse marker use) or at the level of the journal article sections, i.e., *introduction*, *method*, etc. and finally, variation across languages.

3. Results and discussion

The first aim of this paper is to compare the prevalent use of metadiscourse markers across languages and texts, differentiating among these languages and their corresponding academic journals. The second aim is to underscore both the similarities and differences within individual languages and genres, as well as between them in the corpus. Before presenting the results as they relate to the aims, we checked the specific journal websites for specific information pertaining to aspects related to, 1) text length, 2) macro formatting guidelines related to IMRaD structures, 3) style guides, plus 4) specific linguistic style suggestions to add some additional contextual information which may help us to determine whether editorial guidelines of the specific journals may result in a specific pattern emerging in the metadiscourse. The results are shown in Table 3.

The majority of the journals do not provide an explicit guideline related to sections and language-specific guidelines, except perhaps the Estonian language journal *Keel ja Kirjandus* and the Lithuanian journal *Kalbotyra* which both explicitly state that the journal reserves the right to edit the article in cooperation with the author or requires editing by a native-language specialist. All other journal guidelines provide a broader guideline related to text style guides, such as formatting and reference style guides (i.e., APA, MLA, etc.).

3.1. Language-specific metadiscourse baseline models

To compare the distribution of metadiscourse markers across journal articles for each language, we explore the relationship between metadiscourse markers and the journal article sections coded in the corpus. Table 4 presents the percentages of the metadiscourse markers used for the purpose of this comparative analysis. The percentages represent metadiscourse markers in the entire sub-corpus and the percentages of all metadiscourse markers within each section of the journal articles. Generally, in both the Estonian and Lithuanian datasets, interactive markers – markers that assist in guiding the reader through the text, such as transition markers, code glosses, endophoric markers, evidentials, and frame

Table 3. Journal editorial guidelines pertaining to any stylistic or discoursal guidelines

Journals	Specific guidelines
KK	<ul style="list-style-type: none"> • Recommended length 40,000–50,000 characters (incl. spaces). • The journal reserves the right to edit the article in cooperation with the author, incl. shortening it if necessary. • Listed text style guidelines.
ERÜ	<ul style="list-style-type: none"> • Length up to 40,000 characters (excluding spaces). • The text should be structured in sections with decimal hierarchical numbering and section titles. Excessive use of footnotes must be avoided. • Instructions for abbreviations, citations (in-text), references, glosses, and text style are included.
ESA	<ul style="list-style-type: none"> • Recommended length 35,000–50,000 characters (incl. spaces). • Text style requirements and recommended structuring (only one level of sub-section), in-text citation instructions and reference guidelines.
K	<ul style="list-style-type: none"> • Recommended length 8,000 words; in exceptional cases the paper can be longer. • If the language of the paper is not a native language of the author(s), the paper should be proof-read by a native-language specialist to check its correctness. • Division into sections and subsections. • Other listed text style guidelines.
LK	<ul style="list-style-type: none"> • There are no requirements for the length of an article. • Recommended structure and form: the research question/problem, review of previous research on the subject, data and methods, research findings/results (evaluated and validated), evidence (documented), conclusions and references. • Listed text style guidelines.
TK	<ul style="list-style-type: none"> • The total length of the publication should range from 10,000 to 80,000 characters with spaces. The recommended volume of scientific reviews is 16,000 characters. In case the contribution exceeds the indicated length, it should be negotiated with the Editorial Board. • Other listed text style guidelines.

markers – constitute the majority of metadiscourse markers in the texts. This comprises approximately 75% of the Estonian data and 67.8% of the Lithuanian data. Furthermore, the most frequent marker in both datasets is transition markers (e.g., next, but, however, etc.), accounting for 32.3% and 22.7%, respectively. When examining the distribution of metadiscourse markers across the various sections (IMRaD) of the research papers, we observe that the majority of metadiscourse markers are located in the *results and discussion* section for both the Estonian and Lithuanian datasets. 54% of the metadiscourse markers are found in the combined *results* and the *results and discussion* sections in the Estonian sub-corpus, and 51.1% of the metadiscourse markers are found in the *results and discussion* section of the Lithuanian sub-corpus. The lowest number of metadiscourse markers in both datasets is found in the *discussion* section in Estonian and Lithuanian: 7.2% and 3.3%, respectively.

Table 4. Percentage distribution of Estonian and Lithuanian metadiscourse markers across the various sections of the journal articles

Sub-corpus	Metadiscourse markers	%	Metadiscourse markers represented in each section of the journals	%
Estonian	Transition markers*	32.3	Introduction	14.5
	Code glosses*	16.8	Literature Review	7.8
	Endophoric markers*	16.5	Data & Methods	8.1
	Evidentials*	5.7	Results	30.2
	Frame markers*	3.7	Results & Discussion	23.8
	Hedges°	5.9	Discussion	7.2
	Boosters°	6.5	Conclusion	8.5
	Attitudinal markers°	3.5		
	Engagement markers°	1.4		
	Self-mentions°	7.7		
Lithuanian	Transition markers*	22.7	Introduction	18.6
	Code glosses*	12.2	Theory	12.2
	Endophoric markers*	8.4	Data & Methods	6.6
	Evidentials*	22.0	Results & Discussion	51.1
	Frame markers*	2.5	Discussion	3.3
	Hedges°	9.4	Conclusion	8.1
	Boosters°	10.5		
	Attitudinal markers°	2.1		
	Engagement markers°	9.5		
	Self-mentions°	0.7		

* Interactive category: metadiscourse markers that help to guide readers through the text.

° Interactional category: metadiscourse markers that involve the reader in the text.

To determine the patterns of metadiscourse marker usage across both languages and academic journals, we employed association plots. The plot visualizes the relationships between journal article sections and the metadiscourse categories, both in terms of association and dissociation. In the association plots, a red cell indicates under-representation (i.e., frequencies lower than expected under the assumption of independence), whereas a blue cell indicates over-representation (i.e., frequencies higher than expected under the assumption of independence). The base of each bar represents the degree of support for the association, while its height is proportional to its significance. In this context, significance denotes a pattern that deviates significantly from random variation.

Figure 1 presents the metadiscourse baseline models for Estonian and Lithuanian in each journal section of the corpus. In the *introduction* section, the Estonian metadiscourse baseline model exhibits a positive association with the use of evidential markers (e.g., *according to*) and self-mentions (e.g., *I*), while it demonstrates a negative association with transition markers (e.g., *in addition*). In the *literature review* section, there is a positive association with evidential markers but a negative

association with self-mentions. In the *method* sections, there is generally a positive association with self-mentions and negative associations with transition markers, endophoric markers (e.g., *in section one*), hedges, and boosters. In the *results* sections, there are generally positive associations with transitional markers and endophoric markers but a negative association with evidential markers, hedges, engagement markers, and self-mentions. In contrast, the *results and discussion* section exhibits positive associations with engagement markers and attitudinal markers but negative associations with frame markers and self-mentions. The *discussion* section of articles shows a small positive association with boosters and a small negative association with self-mentions. Finally, the conclusion section displays a small positive association with transitional markers and hedges, as well as negative associations with endophoric markers and evidentials.

The Lithuanian metadiscourse baseline model offers a more conventional overview of the journal sections. For the *introduction* section, a large positive association is observed with the use of evidential markers and frame markers, but more negative associations with transitional, endophoric, hedges, boosters, and engagement markers. In the *theory* section, there is also a positive association with the use of evidential markers and code glosses. More negative associations are found among the interactional markers, such as hedges, booster, and engagement markers, but also with endophoric markers. In the *data and method* section, we find small positive associations among the interactive markers frame markers and code glosses, and small negative associations amongst the interactional markers, hedges, and boosters. In the *results and discussion* section, there are large negative associations with evidential and frame markers, as well as a small negative association with self-mentions. We find more positive association amongst transitional, endophoric markers, hedges, boosters, and engagement markers. Overall, the *discussion* section exhibits a positive association with markers found in the interactional category: hedges, boosters, attitudinal markers, and self-mentions, but a negative association with evidential markers. Similarly, the *conclusion* section also demonstrates a positive association with the use of interactional markers: hedges, boosters, and self-mentions, but a negative association among the interactive category markers: evidentials, endophoric markers, and code glosses.

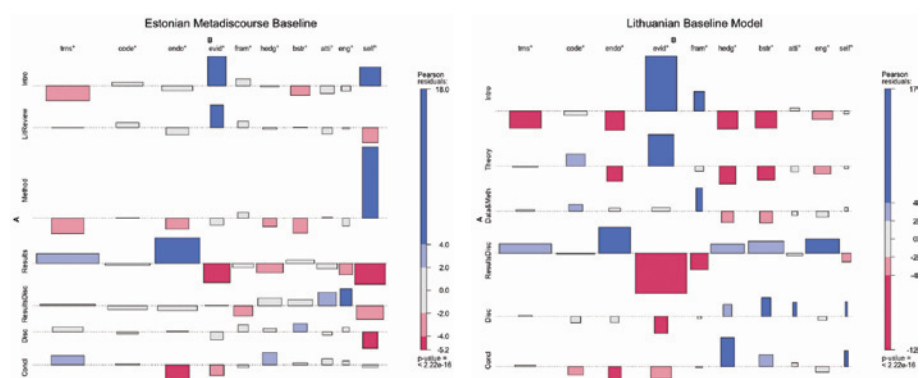


Figure 1. Estonian and Lithuanian metadiscourse baseline models (* – interactive category (guiding the reader through the text); ° – interactional category (involving the reader in the text))

Overall, the metadiscourse baseline models of the use of metadiscourse markers across the various sections of the text indicate significant variations. The most significant variation is observed within the interactional category (involving the readers in the text). More specifically, there is a negative association with metadiscourse marker use in the Estonian corpus within the *results and discussion* sections, while a more positive association is observed for the use of interactional category markers in comparable sections in the Lithuanian corpus. This observation might imply that Lithuanian journal articles in linguistics exhibit a more positive trend of writers engaging readers in the text in comparison to Estonian writers.

3.2. Metadiscourse models of Estonian academic journals versus baseline model

Since the corpus also distinguishes data from three different journals in both the Estonian and Lithuanian corpus, we conducted further comparisons between the specific metadiscourse patterns of these journals and how they compare to the metadiscourse baseline model. The assumption is that if a specific journal shows a large degree of variation to the baseline, we may assume that this variation is because of the type of research these journals accept, which invariably has a different discursual pattern. Alternatively, the journal may have a strong editorial preference or pattern which is different from the baseline. As such, we calculated the metadiscourse model for each journal for each language and compared these to the baseline model using heatmaps.

Figure 2 shows the association plots and the heatmap of differences between the Estonian journals: KK, ERÜ, and ESA, and the Estonian baseline model. On the left side of the figure, the association plots for the specific journals are presented. On the right side, the heatmaps of the differences which presents the calculation of the expected values for each count of data in comparison to the baseline count. Cells with residuals which are close to 0, which means that the observed count in the specific journal and the expected values are close, are coloured white. Positive residuals, coloured red, indicate that the observed values in the specific journal are greater than expected in the baseline model, while negative residuals, coloured blue, indicate that the observed values in the specific journal are less than what would be expected based on the baseline model. Additionally, since the heatmap calculates residuals from the bottom left to the top right, the sections in the heatmap are presented in reverse order.

In general, KK does not deviate too much from the baseline model. The most notable finding in the comparison is the negative residual (indicated in blue, suggesting a lower count in the specific journal compared to the baseline model) of transition markers in the *results* sections of KK. In other words, in our corpus, KK has fewer transition markers (*such as, but, in addition, and, etc.*) in the *results* section than one would expect based on the metadiscourse baseline model for Estonian. ERÜ, like KK, closely aligns with the metadiscourse baseline model, and the only noticeable deviation (indicated in blue) is also in the usage of transition markers. However, compared to KK, ERÜ exhibits a negative trend in using transition markers, specifically in the *results and discussion* section. Finally, ESA

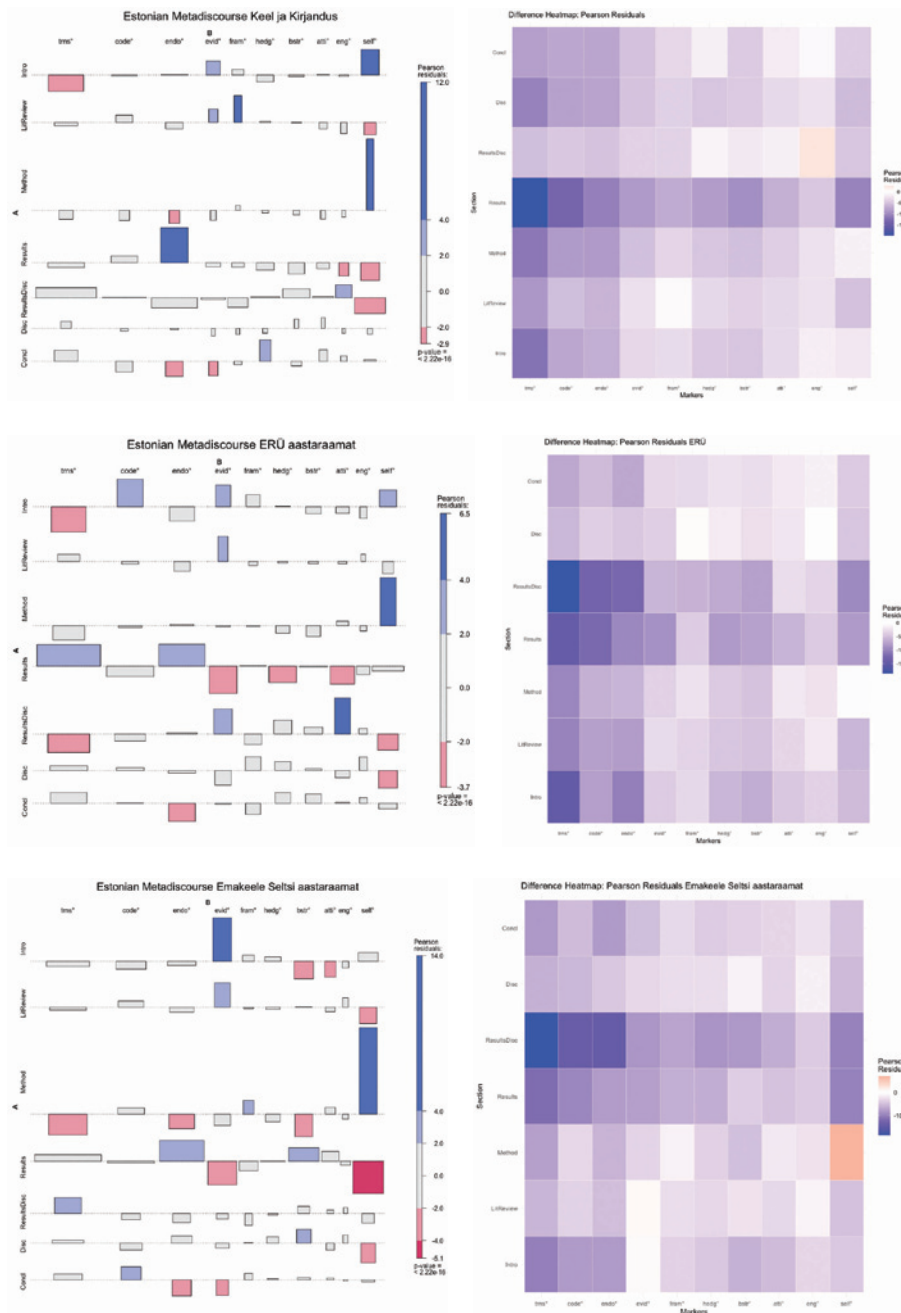


Figure 2. Estonian journal articles (KK, ERÜ, ESA) compared to the Estonian baseline model

shares commonalities with ERÜ in the negative residual for transition markers in the *results and discussion* section, but overall, not too much variation with the baselines. What sets ESA apart from the other journals and the baseline model is a slight inclination towards a positive residual (indicating a higher frequency in the journal compared to the baseline) of self-mentions (e.g., *I, we*, etc.) in the *methods* section, although this tendency is not very pronounced.

3.3. Metadiscourse models of Lithuanian academic journals versus baseline model

The same comparative analysis was carried out for the Lithuanian journals. Figure 3 shows the association plots and the heatmap of differences between the Lithuanian journals K, LK, and TK and the Lithuanian baseline model. Again, the association plots are presented on the right side of the figure and the heatmaps of differences (i.e., comparing the journal metadiscourse model with the Lithuanian baseline model), are presented on the right. Overall, the heatmaps do not display significant deviations from the specific journal article metadiscourse models in comparison to the Lithuanian metadiscourse baseline model. However, across the journals, again, the largest variation is observed in the *results and discussion* section. Among the three journals, K, LK, and TK, the most significant variation from the baseline model is observed in the usage of evidentials (referring to information from other texts, e.g., *according to X, Z states*), which generally appears less frequent (indicated by blue) compared to the baseline model. Given the nature of the *results and discussion* section, this is not surprising, but interesting, nonetheless. Specifically, you would expect a lower frequency of references to information from other texts in the presentation of *results*, but the inclusion of a *discussion* could require a few evidentials to be included. Comparatively, the journal LK, when contrasted with K and TK, appears to exhibit a lower residual for transition markers in the *results and discussion* section, as has also been noted in the Estonian metadiscourse models of the specific journals. Lastly, the metadiscourse model for the journal K, overall, exhibits some minor positive residuals (although only slightly) for certain markers (mainly in the interactional category involving readers in the text) in the *discussion* section, confirming an earlier observation when comparing the metadiscourse baseline models of Estonian and Lithuanian, which showed some variation in the association plots within the interactional category.

3.4. Metadiscourse models across languages and academic journals

To perform a cross-language comparison of metadiscourse markers, we included the sections from the language-specific metadiscourse baseline models that allowed for a direct comparison. In other words, we examined *introductions, literature review/theory, methods, results and discussions, discussions, and conclusions*. To make a cross-language specific journal comparison, we took the metadiscourse model of the Estonian journal ERÜ and the metadiscourse model of the Lithuanian journal K, two journals which in the corpus had comparable sections. The

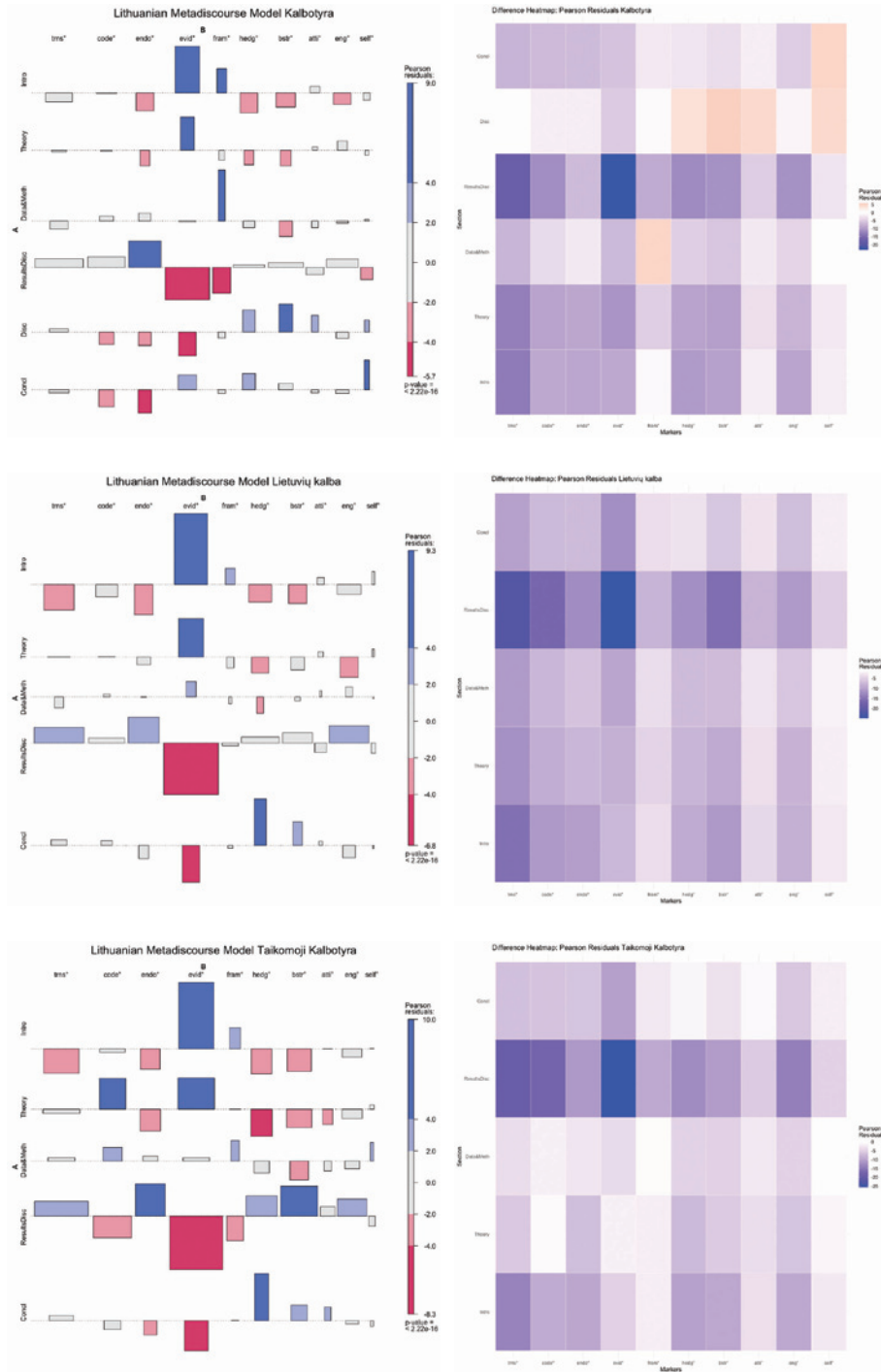


Figure 3. Lithuanian journal articles (K, LK, TK) compared to the Lithuanian baseline model

purpose of these comparisons is to determine whether the metadiscourse models exhibit any similarities, suggesting a more universal metadiscourse pattern across Estonian and Lithuanian, or if they demonstrate variations, either across sections of a journal, metadiscourse markers, or both.

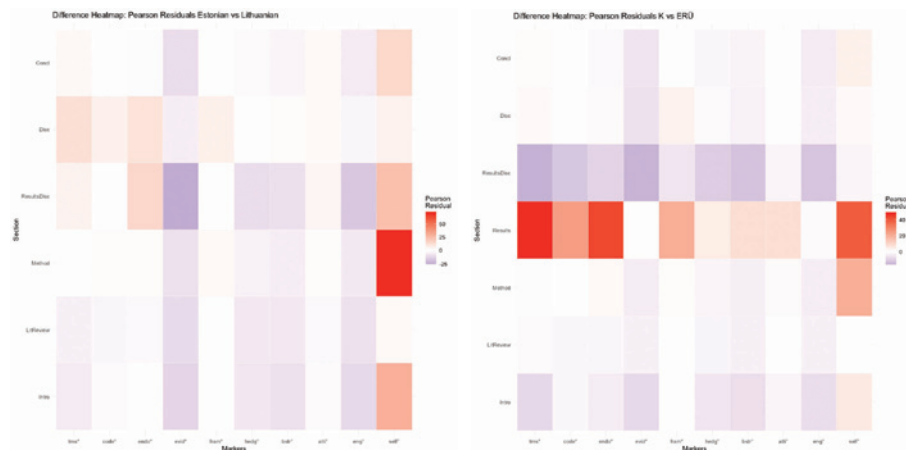


Figure 4. Estonian metadiscourse model compared to Lithuanian metadiscourse model and Lithuanian journal K compared to the Estonian journal ERÜ

Figure 4 shows the heatmap of differences between Estonian and Lithuanian. The heatmap on the left of Figure 4 shows the comparison of the metadiscourse baseline model of Estonian to the metadiscourse baseline model of Lithuanian. The heatmap on the right of Figure 4 shows the comparison of the metadiscourse model of the Lithuanian journal K and the Estonian journal ERÜ. In comparison to the language-specific comparison, the heatmaps show a less uniform comparison and similarity between the language-specific baseline model and language-specific journal. Initially, a direct comparison of the Estonian and Lithuanian models reveals less variation, with the only significant positive deviation observed in the use of self-mention (e.g., *I, we*, etc.) in the methods section. In other words, Estonian writers tend to include self-mentions much more in *methods* sections when compared to their Lithuanian colleagues. Interestingly, when comparing the Lithuanian journal K to the Estonian journal ERÜ, a different trend emerges. When comparing these two journals, the majority of positive variations, red-coloured cells, are observed in the *results* section, with K having more observations of transition markers, code glosses, and endophoric markers, within the interactive category of metadiscourse markers aiming to help guide the reader through the text, and more observations of self-mentions in the *results* section in comparison to ERÜ. The first three in the interactive category might be explained by the type of texts K publishes in their journal, which might contain more linguistic examples which would need more writer guidance such as, *in addition to ...*, *as seen in Table 2*, *e.g.*, and *in other words*. ERÜ, on the other hand, might publish articles which more frequently present results without the use of guidance by authors and reserves the transition markers for the *results and discussion* or *discussion* section, as one would expect.

4. Conclusion

Overall, the results of the cross-linguistic and journal comparisons underscore the value of employing metadiscourse markers. When comparing metadiscourse models across journals within a specific discipline, such as linguistics, the variation in the model is relatively small, with only certain sections of the journal displaying deviations from the general baseline model. It would be intriguing, however, to extend this analysis to metadiscourse models in journals from neighbouring and other disciplines, like social science or medical sciences. This exploration might reveal entirely different patterns or reinforce existing ones. Moreover, when comparing across languages, the comparison of disciplines may not yield significant variations. Instead, most variations could stem from specific cultural or rhetorical expectations. For instance, the perennial question of whether to use self-mentions such as ‘I’ or ‘we’ might manifest differently across languages. Nevertheless, despite the disciplinary similarities across languages, comparing similar discipline-specific journals across languages may reveal more significant variation within specific sections of a paper and across the various categories of the interpersonal metadiscourse model. As demonstrated in the comparison between Lithuanian and Estonian journals, specific strategies to guide readers through the text may be more essential for certain types of articles that are commonly accepted and published in those journals, effectively overcoming bias. To validate this assumption, a broader and more extensive selection of journals may be required.

References

- Ädel, Annelie 2018. Variation in Metadiscursive “You” Across Genres: From Research Articles to Teacher Feedback. – *Educational Sciences: Theory & Practice*, 18 (4), 777–796. <https://doi.org/10.12738/estp.2018.4.0037>
- Birhan, Amare Tesfie 2021. An exploration of metadiscourse usage in book review articles across three academic disciplines: A contrastive analysis of corpus-based research approach. – *Scientometrics*, 126 (4), 2885–2902. <https://doi.org/10.1007/s11192-020-03822-w>
- Cao, Feng; Hu, Guangwei 2014. Interactive metadiscourse in research articles: A comparative study of paradigmatic and disciplinary influences. – *Journal of Pragmatics*, 66, 15–31. <https://doi.org/10.1016/j.pragma.2014.02.007>
- Crismore, Avon; Markkanen, Raija; Steffensen, Margaret S. 1993. Metadiscourse in persuasive writing: A Study of texts written by American and Finnish university students. – *Written Communication*, 10 (1), 39–71. <https://doi.org/10.1177/0741088393010001002>
- Dahl, Trine 2004. Textual metadiscourse in research articles: A marker of national culture or of academic discipline? – *Journal of Pragmatics*, 36 (10), 1807–1825. <https://doi.org/10.1016/j.pragma.2004.05.004>
- Duszak, Anna 1994. Academic discourse and intellectual styles. – *Journal of Pragmatics*, 21 (3), 291–313. [https://doi.org/10.1016/0378-2166\(94\)90003-5](https://doi.org/10.1016/0378-2166(94)90003-5)
- Duszak, Anna; Lewkowicz, Jo 2008. Publishing academic texts in English: A Polish perspective. – *Journal of English for Academic Purposes*, 7 (2), 108–120. <https://doi.org/10.1016/j.jeap.2008.03.001>
- Fløttum, Kjersti; Dahl, Trine; Kinn, Torodd 2006. *Academic Voices: Across Languages and Disciplines. Pragmatics & Beyond New Series*, 148. Amsterdam–Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.148>

- Harbord, John 2018. Language policy and the disengagement of the international academic elite. – *Science Editing*, 5 (1), 32–38. <https://doi.org/10.6087/kcse.115>
- Hint, Helen; Leijen, Djuddah A. J.; Jürine, Anni 2022. Eestikeelse akadeemilise teksti tunnustest. – *Keel ja Kirjandus*, 4, 327–353. <https://doi.org/10.54013/kk772a3>
- Hint, Helen; Lemendik, Helena; Johansson, Christer; Leijen, Djuddah A. J. (Forthcoming). Decoding metadiscourse markers in Estonian academic texts: A language-specific perspective. – *Written Communication*.
- Hyland, Ken 2005. *Metadiscourse: Exploring Interaction in Writing*. London–New York: Continuum.
- Hyland, Ken; Tse, Polly 2004. Metadiscourse in academic writing: A reappraisal. – *Applied Linguistics*, 25 (2), 156–177. <https://doi.org/10.1093/applin/25.2.156>
- Hyland, Ken; Wang, Wenbin; Jiang, Feng (Kevin) 2022. Metadiscourse across languages and genres: An overview. – *Lingua*, 265, 103205. <https://doi.org/10.1016/j.lingua.2021.103205>
- Jürine, Anni; Leijen, Djuddah; Sinkuniene, Jolanta; Laiveniece, Diāna; Johansson, Christer; Groom, Nicholas 2021. Academic writing in the Baltic States: Introducing the Bwrite project. – *Educare*, 1, 27–37. <https://doi.org/10.24834/educare.2021.1.3>
- Leijen, Djuddah; Hint, Helen; Jürine, Anni (Forthcoming). Academic writing tradition and how to model it. – C. Donahue, C. Gannett (Eds.), *International Writing Studies Research: Cultivating Collective Research Capacity through International Exchanges about Higher Education Writing Research*. The WAC Clearinghouse International Series.
- Mauranen, Anna 1993. Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. – *English for Specific Purposes*, 12 (1), 3–22. [https://doi.org/10.1016/0889-4906\(93\)90024-I](https://doi.org/10.1016/0889-4906(93)90024-I)
- Meyer, David; Zeileis, Achim; Hornik, Kurt 2003. Visualizing independence using extended association plots. – *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. <https://www.r-project.org/conferences/DSC-2003/Proceedings/MeyerEtAl.pdf> (27.3.2024).
- Mur-Dueñas, Pilar 2011. An intercultural analysis of metadiscourse features in research articles written in English and in Spanish. – *Journal of Pragmatics*, 43 (12), 3068–3079. <https://doi.org/10.1016/j.pragma.2011.05.002>
- Peng, Jian-E.; Zheng, Yuting 2021. Metadiscourse and voice construction in discussion sections in BA theses by Chinese university students majoring in English. – *SAGE Open*, 11 (2), 215824402110088. <https://doi.org/10.1177/21582440211008870>
- R Core Team 2022. *R: A Language and Environment For Statistical Computing*. R Foundation for Statistical Computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (19.3.2024).
- Ruskan, Anna; Hint, Helen; Leijen, Djuddah A. J.; Šinkūnienė, Jolanta 2023. Lithuanian academic discourse revisited: Features and patterns of scientific communication. – *Open Linguistics*, 9 (1), 20220231. <https://doi.org/10.1515/opli-2022-0231>
- Sollaci, Luciana B.; Pereira, Mauricio G. 2004. The introduction, methods, results, and discussion (IMRaD) structure: A fifty-year survey. – *Journal of the Medical Library Association*, 92 (3), 364–371.
- Vande Kopple, William J. 1985. Some exploratory discourse on metadiscourse. – *College Composition and Communication*, 36 (1), 82–93. <https://doi.org/10.2307/357609>
- Wickham, Hadley 2007. Reshaping data with the reshape package. – *Journal of Statistical Software*, 21 (12), 1–20. <https://doi.org/10.18637/jss.v021.i12>
- Wickham, Hadley 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Cham: Springer. <https://ggplot2.tidyverse.org>

METADISKURSUSE MUSTRITE KEELTEVAHELINE VÕRDLUS: VALDKONDLIKUD SARNASUSED JA ARTIKLIOSADE ERINEVUSED

**Djuddah Leijen¹, Helen Hint¹, Helena Lemendik¹,
Baiba Egle², Anna Ruskan³, Christer Johansson⁴**

Tartu Ülikool¹, Liepāja Ülikool², Vilniuse Ülikool³, Bergeni Ülikool⁴

Artikkel käsitleb metadiskursuse markereid eesti ja leedu keeleteaduslikes ajakirjades. Meie eesmärk on 1) võrrelda kõigi metadiskursuse markerite üldist kasutust tekstides keeliti, otsides nii keeltes kui ka ajakirjade kaupa ilmnevaid erinevusi, ning 2) leida keeliti võimalikud sarnased ja/või erinevad mustrid ja selgitada välja, kas need mustrid tulevad esile ka eri ajakirjades. Tulemused näitavad, et eesti autorid kasutavad artiklite meetodiosas enesele osutamisi (ingl *self mentions*) sagedamini kui leedu autorid. Ajakirjade võrdluses leidub leedu ajakirjas Kalbotyra rohkem sidususmarkereid (*transition markers*), täpsustavaid markereid (*code glosses*) ja tekstisiseseid viiteid (*endophoric markers*), samal ajal kui Eesti Rakenduslingvistika Ühingu aastaraamatus jäävad sellised lugejat juhatavad markerid rohkem tulemuste ja arutelu osadesse. Vaatamata valdkondlikele sarnasustele ilmneb siiski erinevusi teatud artikliosades ja interpersonaalse metadiskursuse kategooriates nii keeliti kui ka ajakirjade lõikes. Uurimus heidab valgust metadiskursuse kasutusmustritele ja nende rollile eri keeltes ja akadeemilistes kontekstides ning võib olla edaspidi suunanäitajaks mitteingliskeelsete akadeemiliste tekstide uurijatele ja praktikutele.

Võtmesõnad: metadiskursus, teadusartiklid, IMRaD, eesti keel, leedu keel

Djuddah Leijen (University of Tartu) is interested in a range of research topics, including research on text production, writing research, genre studies, PhD writing processes, peer review, and modelling writing structures.

Lossi 3, 51003 Tartu, Estonia
djuddah.leijen@ut.ee

Helen Hint (University of Tartu) is interested in a range of research topics, including the pragmatics of referential devices, writing studies, linguistic features of Estonian academic texts, and Estonian language didactics.

Jakobi 2, 51005 Tartu, Estonia
helen.hint@ut.ee

Helena Lemendik (University of Tartu) investigates the specific features of Estonian academic texts, focusing primarily on the linguistic aspects from a metadiscursive perspective.

Jakobi 2, 51005 Tartu, Estonia
helena.lemendik@ut.ee

Baiba Egle's (Liepāja University) research interests include the Latvian language of science, multilingualism in science, language policy, as well as various aspects of text linguistics and translation studies.

Lielā iela 14, 3401 Liepāja, Latvia
baiba.egle@venta.lv

Anna Ruskan's (Vilnius University) research interests include evidentiality, epistemic modality, grammaticalization, corpus linguistics and contrastive linguistics.
Universiteto str. 5, LT-01513 Vilnius, Lithuania
anna.ruskan@flf.vu.lt

Christer Johansson's (University of Bergen) research interests include experimental psycholinguistics, statistical data analysis, and computational linguistics. He investigates relevant factors for human language processing, with one goal being to aid language users through the design of language tools adaptable to individual needs. His current research focuses on codeswitching and how the interaction of vision and speech affects visual change blindness.
Sydnesplassen 7, 5007 Bergen, Norway
christer.johansson@uib.no