

Navigating texts and sources: Interactive metadiscourse in Estonian student writing

Helena Lemendik

Institute of Estonian and General Linguistics

University of Tartu

Jakobi 2

50090 Tartu, Estonia

E-mail: helena.lemendik@ut.ee

ORCID iD: <https://orcid.org/0009-0006-6529-3811>

<https://ror.org/03z77qz90>

Abstract. Metadiscourse helps academic writers guide readers through the unfolding text and its external sources. Although Baltic academic writing has attracted increasing attention, interactive metadiscourse in Estonian student writing has not been studied systematically. This study applies an Estonian adaptation of the interpersonal model of metadiscourse to a corpus of 40 Estonian-language bachelor's and master's dissertations from the humanities and social sciences. Potential markers were retrieved and classified semi-automatically, then validated contextually. Their distribution across fields, study levels and rhetorical sections was examined using normalised frequencies, multiple correspondence analysis and conditional inference trees. Evidentials and endophoric markers – the two most frequent categories – were also analysed by subcategory and in terms of their associated lexical items. Evidentials accounted for more than half of all interactive metadiscourse, followed by endophoric markers. Frame markers were the least frequent. Within the corpus, marker distributions varied most clearly across rhetorical sections: evidentials characterised literature reviews, while endophoric markers were concentrated in introductions and conclusions. Non-integral citations predominated, although integral citation was more prominent in some disciplinary and rhetorical contexts. Whole-text references were the most frequent endophoric subcategory, followed by previewing and reviewing references. The study provides the first systematic corpus-based account of interactive metadiscourse in Estonian university dissertations and contributes to research and instruction on Baltic academic writing.

Keywords: interactive metadiscourse, Estonian academic writing, university dissertations, rhetorical sections, endophoric markers, evidentials, reporting clauses

Submitted: 10/08/2026. **Accepted:** 26/09/2026

Copyright © 2026 Helena Lemendik. Published by Vilnius University Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1 Introduction

Over the past several decades, metadiscourse has become an established area of research within academic discourse studies, providing valuable insights into how writers organise texts, position themselves in relation to knowledge claims, signal their stance and establish relations with readers (Hyland 2017, 20; Hyland & Jiang 2024, 411). Effective metadiscursive choices can contribute to textual coherence and reader orientation by making relations among ideas, stages of the text, knowledge sources and writer-reader roles more accessible (Hyland et al. 2022, 2). In addition to providing a productive framework for examining academic texts, metadiscourse has pedagogical relevance: explicit attention to these resources can increase students' awareness of readers' needs and help them make more deliberate rhetorical choices in their own writing (Cheng & Steffensen 1996, 177; Hyland 2005, 85; 2017, 27).

Although studies of English-language academic writing continue to dominate the field, research has increasingly extended to other languages (Pearson & Abdollahzadeh 2023, 20), including those of the Baltic region. Lithuanian scholarship has made a particularly sustained contribution to this development. Studies by Šinkūnienė and colleagues have investigated hedging (Šinkūnienė 2008; 2011; 2012; 2014), boosting (Šinkūnienė 2014), authorial stance and reader engagement through personal pronouns (Šinkūnienė 2010; 2014; 2018; Linkevičienė & Šinkūnienė 2012), reformulation (Šinkūnienė 2014; 2019) and endophoric markers (Šinkūnienė 2014). Related work has addressed evidential markers (Ruskan 2012; Šinkūnienė 2014) and citation practices (Šinkūnienė 2017; Petrauskaitė & Šinkūnienė 2015). Much of this research has also adopted cross-linguistic and cross-disciplinary perspectives, thereby identifying both recurring features of academic discourse and rhetorical patterns associated more specifically with Lithuanian scholarly writing (Ruskan et al. 2023).

More recently, cross-Baltic collaboration intensified within the Bwrite project, *Academic Writing in the Baltic States: Rhetorical Structures through Cultures and Languages* (Jürine et al. 2021), to map Baltic academic writing traditions and develop methods for studying non-Anglophone academic writing. Outputs include comparisons of metadiscourse in Estonian and Lithuanian linguistics articles (Leijen et al. 2024), endophoric markers across the three languages of the Baltic states (Hint et al. 2024) and a review of Lithuanian academic discourse (Ruskan et al. 2023), alongside language-specific studies of Lithuanian (Ruskan & Maslauskienė 2023) and Latvian academic writing (Dubova et al. 2023). In Estonia, Hint et al. (2025) adapted Hyland's (2005) interpersonal model of metadiscourse to Estonian and developed language-specific annotation principles for expert research articles. Lemendik (2025) subsequently examined interactional metadiscourse in Estonian university dissertations, focusing on grammatical voice, stance and reader engagement. Estonian metadiscourse research now

covers both expert and student writing, but the interactive and interactional dimensions remain unevenly investigated.

To fill this gap, the present study builds on Hint et al.'s (2025) Estonian adaptation of Hyland's (2005) interpersonal model of metadiscourse and focuses on its interactive, reader-guiding dimension. The analysis covers five categories: transitions, frame markers, endophoric markers, evidentials and code glosses. All five are analysed quantitatively using descriptive statistics and exploratory multivariate methods, while more detailed analysis focuses on evidentials and endophoric markers, the two most frequent categories.

The two focal categories represent complementary dimensions of reader guidance. Evidentials establish intertextual relations by attributing information to external sources and linking it to the writer's own statements (Hyland 2005, 51; Burneikaitė 2006, 99). Endophoric markers establish intratextual and text-reflexive relations by guiding readers forwards or backwards within the current text or by referring to the text as a whole (Bunton 1999, 46; Hyland 2005, 51; Burneikaitė 2009, 12). Together, they show how students guide readers through both their dissertations and the source networks on which they draw. Because evidential and endophoric relations have heterogeneous realisations, the analysis also profiles the lexical items associated with them.

Against this background, the study has three aims: (i) to provide a quantitative account of the distribution of interactive metadiscourse across fields, study levels and rhetorical sections; (ii) to examine the internal patterning of evidentials and endophoric markers across the same dimensions; and (iii) to identify the lexical items most frequently associated with these two categories and use their distributions to characterise dominant functional patterns and the types of activity represented by the associated verbs.

Continuing the recent expansion of Baltic metadiscourse research into student writing, the present study contributes to a broader understanding of regional academic writing practices and provides the first systematic corpus-based account of interactive metadiscourse in Estonian university dissertations, establishing a basis for future comparisons of student academic writing in the Baltic region and beyond.

1.1 Interactive metadiscourse as reader guidance

In Hyland's (2005, 48–49) interpersonal model of metadiscourse, interactive resources help writers anticipate readers' processing needs by making explicit the relations among propositions, stages of discourse, parts of the current text, external knowledge sources and alternative formulations. The interactive dimension comprises transitions (*in addition; thus*), frame markers (*finally; to conclude*), endophoric markers (*noted above; in Section 2*), evidentials (*according to PN* (proper name); *PN states*) and code glosses (*such as*;

in other words). Although these resources are commonly characterised as organising discourse in response to readers' processing needs (Hyland 2005, 49; Hyland & Jiang 2018, 20), their function is broader than the linear organisation of discourse. In particular, evidentials also connect the current text with external knowledge sources (Nesi 2021, 2), so that interactive resources construct pathways through both the unfolding text and the knowledge sources on which it draws.

Metadiscourse is identified functionally and contextually rather than through a context-free inventory of linguistic forms (Hyland 2017, 18). A linguistic expression cannot therefore be classified solely on the basis of its lexical form. Burneikaitė (2009, 13), for example, shows that a noun such as *research* may function metadiscursively when it refers reflexively to the organisation of the current text, but propositionally when it denotes research activity in the represented world. The same principle applies to verbs. A research-act verb, for example, can contribute to intratextual guidance when anchored to a textual location, as in *in Chapter 3, X is analysed*. Contextual identification is therefore essential.

1.2 Evidentials, citation form and source framing

Evidentials establish intertextual relations by indicating that information in the current text originates in, or can be attributed to, another source. In academic writing, this relation is most visibly realised through citation (Hyland 1999, 342; Nesi 2021, 1). Citation is not merely a technical mechanism for documenting sources: it also enables writers to locate claims within disciplinary knowledge, allocate responsibility for propositions and negotiate relations between their own voice and those of other scholars (Hyland 1999, 342; Groom 2000, 20; Thompson 2005, 321).

Integral citations incorporate the author or source grammatically, whereas non-integral citations place the reference outside the syntax, usually parenthetically (Swales 1990, 148). Integral forms are therefore described as author- or source-prominent, whereas non-integral forms tend to foreground the information (Hyland 1999, 344; Thompson & Tribble 2001, 92). Citation form, however, does not determine citation function. Either form may, for example, attribute information, exemplify a claim, identify a method or concept that is used, compare sources, direct readers to further information or evaluate earlier work, and a single citation may perform more than one function (Petric 2007, 243–247).

Integral evidentials commonly involve reporting clauses, but source attribution may also be realised through nominal expressions referring to cited work, its findings or its approach, as well as through reporting adjuncts, including adpositional constructions (Thompson & Tribble 2001, 96; Charles 2006, 313; Nesi 2021, 5). Reporting verbs are particularly informative because they both characterise the activity attributed to the cited

source and may position the dissertation writer in relation to the reported proposition (Hyland 1999, 344; Deng & Li 2026, 02). Thompson and Ye (1991, 369) accordingly distinguish the denotational and evaluative dimensions of reporting verbs. Hyland (1999, 349) develops the denotational dimension through the categories of research acts, discourse acts and cognition acts: writers may represent what previous researchers did, what they said or wrote or what they thought or believed. Reporting meanings are nevertheless constructional rather than purely lexical: surrounding co-text may affect how a source and its proposition are represented (Charles 2006, 319, 326).

Evaluative meanings in source-framing expressions can be described broadly as positive (*advocate, argue*), neutral (*address, cite*), tentative (*believe, hypothesise*) or negative (*attack, condemn*) (Thompson & Ye 1991, 372; Hyland 1999, 350; Deng & Li 2026, 05), but engagement-oriented accounts may offer finer distinctions between acknowledgement, distance, endorsement and challenge (Martin & White 2005, 134; Liardét & Black 2019, 40–41). In engagement-oriented accounts, an expression such as *according to PN* typically acknowledges an external voice without overtly specifying the current writer's alignment, whereas a construction such as *PN claims that* more readily distances the writer from responsibility for the attributed proposition. Other constructions, such as *PN demonstrates that*, may endorse the reported proposition as well founded (ibid.). These, however, are context-sensitive evaluative potentials rather than fixed lexical meanings, since negation, modification, contrast and subsequent commentary may alter the positioning of the construction (Thompson & Ye 1991, 372; Charles 2006, 323). Recent reporting-verb research likewise treats denotation and evaluation as related but analytically separable dimensions (Deng & Li 2026, 05). In the present study, verb denotation is analysed systematically, whereas evaluative potential is considered only for selected cases in which the positioning is sufficiently clear.

1.3 Endophoric markers and intratextual navigation

Endophoric markers establish intratextual relations by directing readers to the current text as a whole or to identifiable locations within it. By explicitly presenting the dissertation or one of its parts as an object of reference, endophoric markers help readers locate, anticipate or recover information in the unfolding text (Bunton 1999, 44–46; Burneikaitė 2009, 12; Lahodová Vališová 2024, 24). Previous research has differentiated endophoric references according to such dimensions as direction (backward- and forward-looking), target (e.g. section, chapter, whole dissertation or non-linear element such as a table), textual scope (ranging from a sentence or paragraph to an entire chapter or dissertation) and distance (from references to an adjacent text segment to those operating across sections or chapters) (Bunton 1999, 47–48; Burneikaitė 2009, 13). These dimensions are especially relevant to dissertations, whose length increases the need to maintain readers' orientation across chapters and sections.

The present study distinguishes three broad functions: previewing references direct readers towards forthcoming discourse, reviewing references direct them back to preceding discourse and whole-text references construe the current dissertation or study as a discourse object. Here, whole-text references are included within the endophoric category rather than being divided between endophoric and frame markers. The boundary between announcing the purpose of a dissertation and describing what is done in the dissertation or study may be difficult to apply consistently (cf. Bunton 1999, 48–49; see also Burneikaitė 2009, 12). Statements concerning purpose, organisation and research activity may use similar self-referential constructions. For example, *the aim of this dissertation is to analyse X* both announces the purpose of the discourse and presents the dissertation itself as a broader textual unit within which *X* is investigated and discussed, and from which information about it can be retrieved.

Like evidentials, endophoric markers are realised through heterogeneous constructions. Text-denoting nouns such as *dissertation*, *chapter*, *section* and *part* may combine with verbs that present, discuss, refer to or locate material, as in *this dissertation provides an overview of X*; *as was mentioned in the previous section* or *X can be found in Part 2*. Endophoric relations may also be signalled by vague or deictic locational expressions such as *above*, *below*, *following* and *previous(ly)*. Examining the lexical items associated with these markers can therefore reveal recurrent ways in which Estonian student writers preview, review and represent the dissertation as a whole.

2 Data and methodology

This section describes the corpus and the procedures used to identify, annotate and analyse interactive metadiscourse, including the operationalisation of evidentials and endophoric markers.

2.1 Data

The research data comprise Estonian-language dissertations defended at the University of Tartu between 2018 and 2022 in the fields of humanities and arts (HUM) and social sciences (SOC). The corpus was stratified by field, study level and year. It contains 10 bachelor's (BA) and 10 master's (MA) dissertations from HUM and the same numbers from SOC, amounting to 40 texts in total. More specifically, two BA and two MA dissertations were selected from each field for each of the five years. BA and MA dissertations were included as study-level subcorpora within the same broad genre of university dissertation writing, allowing comparison across degree levels, while HUM and SOC were selected to provide a broad field contrast. The corpus was not further stratified by individual discipline: within each broad field, the representation of individual disciplines was not fixed in advance.

The dissertations were obtained from the University of Tartu's institutional repository, DSpace¹. Within each field-by-level-by-year group, texts were selected randomly after the application of two eligibility criteria: Estonian as the language of the dissertation and the presence of an identifiable research-report macrostructure. For the section-based analysis, the dissertation had to contain distinguishable theoretical, methodological and results/discussion components. Results and discussion could appear as separate chapters or as a combined chapter. Both arrangements were treated as a single section type. All dissertations also contained introductions and conclusions.

The same corpus of 40 dissertations was previously analysed by Lemendik (2025) with respect to interactional metadiscourse. The present study reuses the same textual material to investigate the complementary interactive dimension, but the analytical data were collected specifically for this purpose: interactive marker occurrences were retrieved and contextually validated anew, and the category-specific variables and lexical annotations are specific to the present analysis. The shared corpus design provides a basis for future comparison between the two dimensions.

The texts were manually pre-processed to remove irrelevant material. The excluded material comprised title pages, authors' declarations, foreign-language abstracts, lists of terms, reference lists, appendices and other formatting elements. Tables and figures were removed together with their captions, notes and labels. Longer quotations, illustrative linguistic examples and formulae were also deleted, whereas footnotes were retained. Following pre-processing, the corpus comprised 389,104 words and 24,034 sentences. The word distribution is relatively even across the two fields, with HUM accounting for 51% and SOC for 49% of the corpus. BA dissertations constitute 42% of the words and MA dissertations 58%, reflecting the generally greater length of master's-level work.

2.2 Candidate retrieval and contextual annotation

Metadiscourse was identified semi-automatically. Each sentence was transferred to Excel with metadata for field, study level, rhetorical section and text identification code. A Python script (Van Rossum & Drake 2001) retrieved marker candidates and assigned initial category and subcategory labels. The inventory drew on the Estonian adaptation of Hyland's (2005) model developed by Hint et al. (2025), supplemented with orthographic variants, synonyms and dissertation-specific items such as *bachelor's/master's dissertation*. The search used 132 items – 104 one-word lexemes and 28 multiword constructions – and yielded 20,013 interactive-marker candidates. Every occurrence was then manually validated against the contextual criteria outlined in Section 1.1 (see also Khabbazi-Oskouei 2013, 94). Propositional uses were removed and category labels revised where necessary. One researcher annotated the dataset without

¹ On 25 November 2024, the University of Tartu's DSpace repository was renamed the Academic Digital Archive (ADA).

independent double-coding. The author had participated in developing and applying the Estonian annotation principles in Hint et al. (2025), where inter-rater agreement was high (Cohen's kappa 0.900 overall and 0.938 for interactive metadiscourse). This continuity supports consistent application of the underlying category system.

Following Hint et al. (2025, 182), transitions were annotated using narrower operational criteria: simple clause- or sentence-linking connectives were excluded, and candidates were retained only in more elaborate expressions relating arguments, such as *seega võib öelda, et* 'thus it can be said that' rather than the isolated form *seega* 'thus'. Importantly, Estonian has a relatively flexible word order, which is strongly influenced by information structure (Lindström 2017, 547). In such a system, overt lexical signalling may be especially valuable for reader guidance, since relations between arguments should not be expected to be recoverable consistently from linear ordering alone.

Table 1 summarises the five interactive metadiscourse categories in Hyland's (2005) model, following the Estonian operationalisation developed by Hint et al. (2025) and the further adaptations made in the present study.

Category	Principal reader-guiding function	Examples	Treatment in the present study
transitions	make relations between propositions or arguments explicit	<i>seega võib öelda, et</i> 'thus it can be said that'	addition, comparison and consequence subcategories; simple local connectives excluded
frame markers	signal discourse acts, sequences or text stages	<i>esiteks</i> 'firstly'; <i>kokkuvõttes</i> 'to conclude'	text stage and sequence subcategories; whole-text references assigned to the endophoric category
endophoric markers	direct readers within the current dissertation or represent the dissertation as a textual whole	<i>järgmises peatükis</i> 'in the next chapter'; <i>käesolevas töös</i> 'in the present dissertation'	previewing, reviewing and whole-text subcategories
evidentials	attribute information to external textual sources	<i>PN leiab, et</i> 'PN finds that'; (<i>PN 2020</i>)	integral and non-integral evidentials subcategories
code glosses	reformulate, explain, exemplify or elaborate meanings	<i>teisisõnu</i> 'in other words'; <i>näiteks</i> 'for example'	elaboration, reformulation and exemplification subcategories; abbreviations not annotated

Table 1. Interactive metadiscourse categories and their operationalisation in the present study

After contextual validation and the removal of irrelevant occurrences, 8,216 of the 20,013 candidate markers remained. The difference between the two totals reflects

both the strong context-dependence of metadiscourse and the inclusive nature of the automatic search. Evidentials required the least correction: only approximately 0.3% of automatically assigned evidential candidates were rejected. This precision depended on the careful specification of potential citation patterns. Manual validation reduced the candidate totals by 31% for code glosses, 76% for endophoric markers, 82% for transitions and 85% for frame markers.

2.3 Operationalisation of evidentials and endophoric markers

All five interactive categories were included in the overall frequency analysis; the procedures described below concern the two focal categories. In the discussion below, the term *marker* refers to the complete expression performing the relevant metadiscursive function in context, whereas *associated lexical item* refers to the single verb, noun, adposition or other item selected as the principal lexical anchor of that relation. One associated lexical item was recorded for each integral evidential and eligible endophoric marker; when several candidates occurred, the most functionally central item was selected.

Following the distinction outlined in Section 1.2, evidentials were coded as integral when the cited author or source formed a constituent of the running sentence and as non-integral when the reference was syntactically detached in parentheses. This coding enabled comparison of source prominence across subcorpora and rhetorical sections. Non-integral evidentials received no lexical annotation because they normally lacked an overt source-framing expression. Verbs associated with integral evidentials were classified as verbs denoting research acts, discourse acts or cognition acts (Thompson & Ye 1991, 372; Hyland 1999, 349; Deng & Li 2026, 05). Evaluative positioning was not coded exhaustively because it may depend on the wider construction rather than on the verb alone (see Section 1.2).

Endophoric markers were coded according to the three subcategories introduced in Section 1.3. All explicit references to the current dissertation or study as a whole were assigned to the endophoric category rather than divided between endophoric and frame markers. Classification was based on contextual function rather than verb denotation; thus, a research-act verb such as *analyse* could occur in an endophoric construction when the activity was anchored to a textual location. References to non-linear elements, such as tables, figures, diagrams, formulae and numbered linguistic examples, were excluded because opportunities for their use varied substantially across dissertations and could allow a few visually dense texts to disproportionately affect the frequencies. As verbs associated with endophoric markers do not attribute activities to external sources, they were not assigned to the reporting-verb taxonomy described above. Instead, their contextual meanings were described more broadly as representing, for example, discourse activity (*välja tooma* ‘point out/highlight’), research activity (*analüüsima* ‘analyse’), cognition (*arvama* ‘think’) or textual organisation (*koosnema* ‘consist of’).

2.4 Analytical procedures

The validated dataset was analysed in several stages. First, raw and normalised frequencies (occurrences per 10,000 words) of the five interactive metadiscourse categories were calculated. The distributions were examined across study level (BA and MA), field (HUM and SOC) and rhetorical section.

Multiple correspondence analysis, or MCA (Lê et al. 2008), was then performed on all 8,216 validated interactive metadiscourse occurrences to explore the joint patterning of marker category, textual location (rhetorical section) and field-by-level group. For this analysis, field and study level were combined into a single four-level variable comprising HUM_BA, HUM_MA, SOC_BA and SOC_MA. The resulting variable represents joint field-by-level profiles and does not separately estimate the effects of field and study level. Field and study level were examined separately in the subsequent conditional inference tree analyses (see below). MCA was used as an exploratory method for visualising associations among categorical variables.

The two focal categories of evidentials and endophoric markers were subsequently examined using separate conditional inference trees (Hothorn et al. 2006). In both analyses, field, study level and rhetorical section were entered as explanatory variables. The trees were used to examine how the relevant subcategory distributions were partitioned by field, study level and rhetorical section. All analyses were conducted in R (R Core Team 2021) using *FactoMineR* (Lê et al. 2008) and *factoextra* (Kassambara & Mundt 2020) for the MCA, and the *ctree()* function in *partykit* for the conditional inference trees (Hothorn et al. 2006).

Finally, the associated lexical items were ranked by frequency, and selected examples were examined to illustrate their typical uses.

3 Results

This section presents the overall distribution of interactive metadiscourse, followed by detailed analyses of evidentials and endophoric markers. Code glosses, transitions and frame markers are considered together in the final subsection.

3.1 Overall distribution and multivariate patterning

The corpus contained 8,216 validated interactive metadiscourse markers, corresponding to 211 occurrences per 10,000 words. Evidentials were by far the most frequent category (n=4,386; 53.4%; 112.7 per 10,000 words), followed by endophoric markers (n=1,379; 16.8%; 35.4 per 10,000), code glosses (n=1,018; 12.4%; 26.2 per 10,000), transitions (n=996; 12.1%; 25.6 per 10,000) and frame markers (n=437; 5.3%; 11.2 per 10,000). Evidentials and endophoric markers together thus accounted for just over 70% of all

interactive resources in the corpus, underscoring the prominence of navigation between external sources and the student writer’s own text.

The overall normalised frequency was higher in bachelor’s than in master’s dissertations (226 vs. 200 per 10,000 words) and in the social sciences than in the humanities (234 vs. 190). Nevertheless, the category hierarchy remained stable across the subcorpora. Evidentials accounted for 53–54% of markers across the two study levels; 50% in HUM, and 56% in SOC. Endophoric markers accounted for 17% at both study levels; 16% in SOC, and 18% in HUM. The clearest field-level contrast concerned evidentials, which were more frequent in the social sciences than in the humanities (131 vs. 95 per 10,000 words). Transitions were also more frequent in social science and bachelor’s dissertations, whereas code glosses’ frequencies varied comparatively little.

Rhetorical section produced a much clearer differentiation (Figure 1). Interactive metadiscourse was densest in introductions (343 per 10,000 words) and literature reviews (311), followed by methods sections (194), conclusions (178) and results/discussion sections (128). The *other* category (i.e. other textual locations, such as headings and footnotes) contained substantially less metadiscourse (39). The composition of the categories also changed with section function. Evidentials constituted 71% of the markers in literature reviews, whereas endophoric markers accounted for 46% in introductions and 45% in conclusions. Methods and results/discussion sections displayed more mixed profiles, with no single category reaching the same degree of dominance.

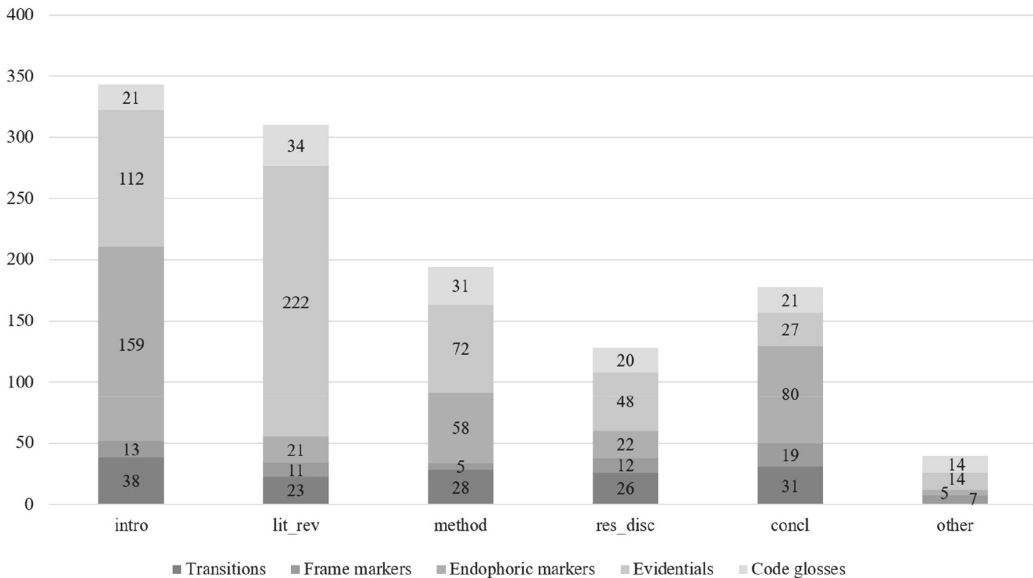


Figure 1. Distribution of interactive metadiscourse across rhetorical sections; per 10,000 words

The MCA (Figure 2) provides a joint representation of these patterns. Dimension 1 separates a source-oriented profile on the left from a text-internal, navigational profile on the right. Evidentials occupy the same side of the dimension as literature reviews, whereas endophoric markers are positioned with introductions. Methods and conclusions also occur on the positive side of the dimension. This configuration closely reflects the section-specific distributions reported above.

Dimension 2 further separates introductions and endophoric markers from results/discussion, transitions, code glosses and frame markers. The second dimension may therefore represent a contrast between overt argument linking or discourse framing resources and text-reflexive navigation.

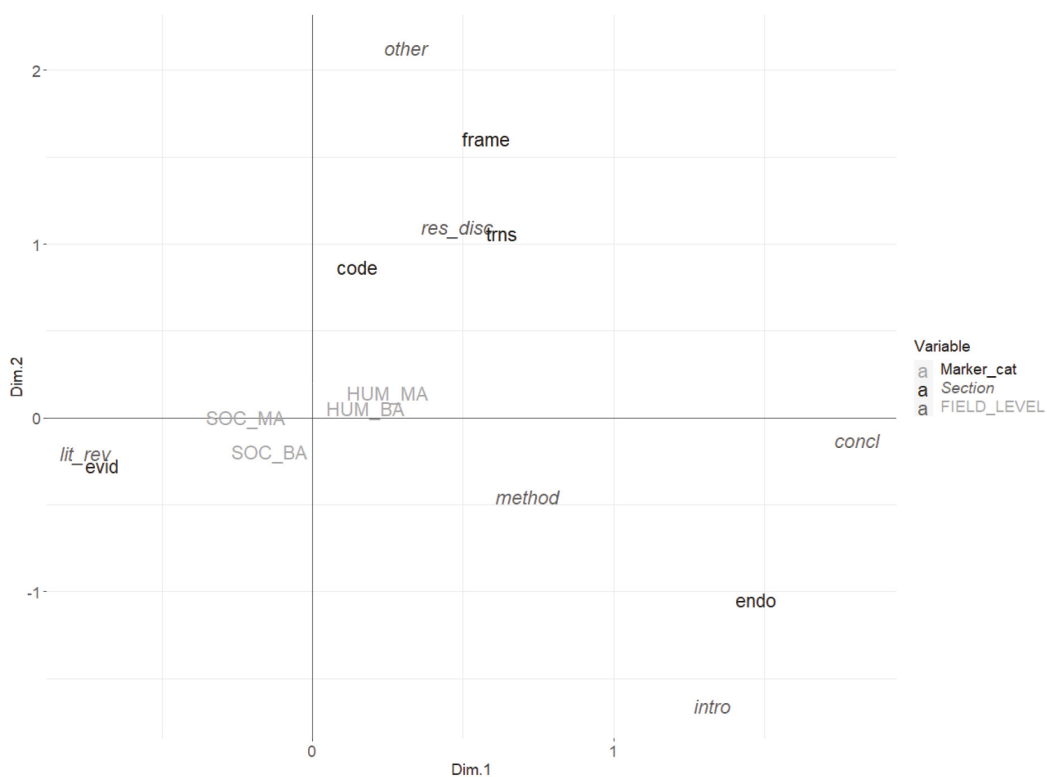


Figure 2. Multiple correspondence analysis of marker category, rhetorical section and field-by-level group

The four field-by-level groups cluster near the origin, suggesting that their profiles are less distinctive than the section-based profiles. The *other* textual-location category is positioned far from the origin. Its extreme coordinates are most likely due to the category's low frequency (Greenacre 2013), and it is therefore not interpreted substantively. Overall,

the MCA is consistent with the descriptive finding that rhetorical section and marker category structure the data more clearly than field-by-level grouping.

3.2 Evidentials: citation integration and lexical source framing

The analysis of evidentials considers both the distribution of integral and non-integral forms and the formal and lexical choices through which external sources are introduced.

3.2.1 Distribution of integral and non-integral evidentials

Of the 4,386 evidentials, 2,893 (66.0%) were non-integral and 1,493 (34.0%) integral. Non-integral citation was therefore the default form, but the relative frequency of integral citation varied across subcorpora and rhetorical sections. In the humanities, integral evidentials occurred at rates of 23 per 10,000 words in BA dissertations and 24 in MA dissertations; the corresponding rates in the social sciences were 55 and 52. Non-integral rates varied less, ranging from 74 per 10,000 words in the humanities to 77 in the social sciences. Both types were most frequent in literature reviews, where the combined evidential frequency reached 222 per 10,000 words. Their relative balance, however, differed by section. Integral evidentials represented approximately 27% of evidentials in introductions, 30% in literature reviews and 28% in methods sections, but 52% in results/discussion sections and 44% in conclusions. Thus, external sources were cited most densely in literature reviews, while sources were most likely to be grammatically foregrounded in the later interpretive sections of the dissertation.

In the conditional inference tree (see Figure 3), rhetorical section provided the first partition. Results/discussion and conclusion sections formed one branch, while introductions, literature reviews, methods sections and other textual locations formed the other. In the later sections of dissertations, humanities evidentials remained predominantly non-integral, whereas integral citation formed a slight majority in social science bachelor's dissertations and a clear majority in social science master's dissertations. In introductions, literature reviews, methods sections and other textual locations, non-integral citations predominated in every field-by-level group. It was most pronounced in humanities bachelor's dissertations and least pronounced in social science bachelor's dissertations. Study level therefore did not exert a uniform effect: its relation to citation form depended on both field and rhetorical section.

3.2.2 Formal and lexical realisation of evidentials

All non-integral evidentials were parenthetical. The standard author-date pattern accounted for 2,593 occurrences (about 90%). Less frequent variants included repeated-source markers such as *samas* 'ibid.' and *ibid.*, attention-directing elements such as *vt* 'see' and *vrd* 'cf.', sometimes accompanied by *nt* 'e.g.' and brief comments directing readers to additional related sources (Examples 1 and 2).

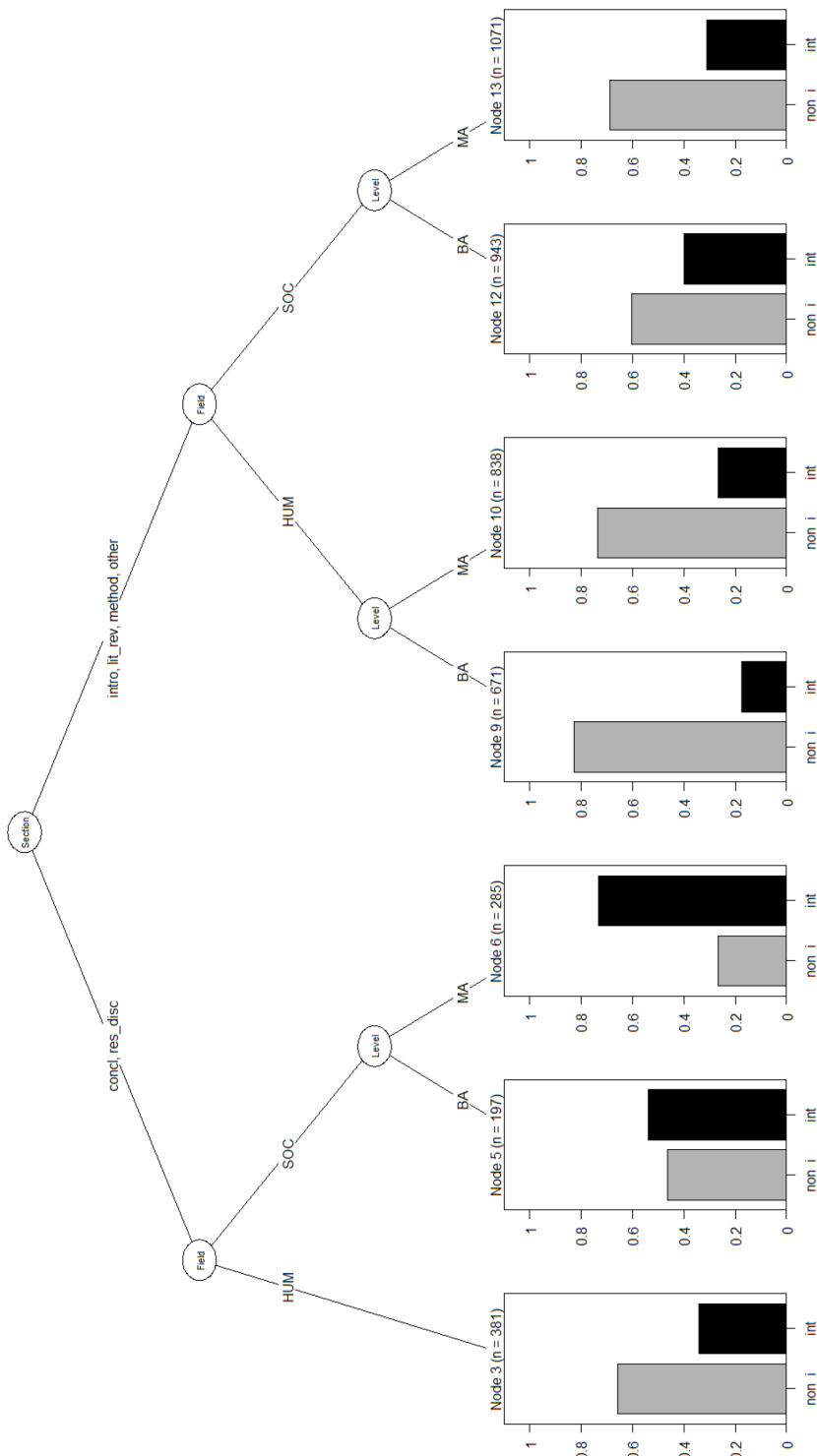


Figure 3. Conditional inference tree for integral vs. non-integral evidentials. Terminal bars show category proportions; n indicates the number of evidential occurrences in the node

- (1) *Üldiselt nimetavad mehed loetelukatses vähem värvinimesid kui naised (vt nt Uusküla 2011: 172; ka Uusküla, Sutrop 2011: 194).* (HUM_BA_2018_2)

‘In general, men produce fewer colour terms than women in the listing task (see, e.g. Uusküla 2011: 172; also Uusküla & Sutrop 2011: 194).’

- (2) *(ibid.: 35, sel teemal ka Brown 2000)* (HUM_MA_2018_2)

‘(ibid.: 35, on this topic, see also Brown 2000)’

The lexical analysis was restricted to integral evidentials. Across the 1,493 occurrences, 193 different associated lexical item types were identified. Verbs were the dominant realisation, accounting for 1,051 occurrences (70.4%), followed by source-framing adverbial or adpositional expressions (n=269; 18.0%) and nouns (n=171; 11.5%). Two adjectival occurrences (*sarnane* ‘similar’) were also identified.

The most frequent verbal items were *välja tooma* ‘point out/highlight’, *leidma* ‘find/consider’ and *kirjeldama* ‘describe’. The most frequent non-verbal source frames were *sõnul* ‘according to/in the words of’, *järgi* ‘according to/on the basis of’ and *kohaselt* ‘according to/as stated by’. The most frequent associated nouns were research-related, referring either to the cited work as a whole or to particular aspects of it, such as *lähenemine* and *käsitlus* ‘approach’, *mudel* ‘model’ or *tulemus* ‘result’. The most frequent associated lexical items of each type are listed in Table 2.

Verbs	Adverbs/Adpositions	Nouns
277 <i>välja tooma</i> ‘point out/highlight’	107 <i>sõnul</i> ‘according to/in the words of’	34 <i>uurimistöö/uurimus/uuring</i> ‘research/study’
109 <i>leidma</i> ‘find/consider’	62 <i>järgi</i> ‘according to/on the basis of’	19 <i>mudel</i> ‘model’
51 <i>kirjeldama</i> ‘describe’	29 <i>kohaselt</i> ‘according to/as stated by’	14 <i>lähenemine</i> ‘approach’
29 <i>selguma</i> ‘emerge/become evident’	15 <i>arvates</i> ‘in [PN’s] opinion’	14 <i>tulemus</i> ‘result’
27 <i>kirjutama</i> ‘write’	11 <i>põhjal</i> ‘based on’	11 <i>käsitlus</i> ‘approach’
27 <i>uurima</i> ‘investigate’		
27 <i>kinnitama</i> ‘confirm’		

Table 2. Most frequent associated lexical items of different types among evidentials

Discourse-act verbs formed the largest denotational group (n=627; 59.7%). The most frequent were *välja tooma* ‘point out/highlight’, *kirjeldama* ‘describe’, *kirjutama* ‘write’, *lisama* ‘add’, *esile tooma* ‘point out/highlight’, *väitma* ‘claim’, *märkima* ‘note’, *käsitleva* ‘address’ and *rõhutama* ‘emphasise’. Research-act verbs (n=194; 18.5%), including *selguma* ‘emerge/become evident’, *uurima* ‘investigate’, *näitama* ‘show’, *välja tulema* ‘be revealed’ and *loomaa* ‘create’, represented research procedures

or findings. Cognition-act verbs, such as *arvama* ‘think’, *olema seisukohal* ‘hold the view’ and *uskuma* ‘believe’, were considerably less frequent ($n=28$; 2.7%). A further substantial group of verbs had context-dependent or overlapping denotational profiles ($n=202$; 19.2%). For example, *leidma* ‘find/consider’, *kinnitama* ‘confirm’, *kasutama* ‘use’ and *viitama* ‘refer to/cite/indicate’ could belong to more than one denotational group, depending on the context (cf. Thompson & Ye 1991, 370). Such overlaps were not subdivided further, as the analysis sought to identify broad patterns in the types of activity represented rather than to classify every occurrence exhaustively.

The frequent source-framing constructions also differed in their evaluative potential. Expressions such as *sõnul*, *järgi* and *kohaselt* ‘according to’ typically acknowledge a source without overtly signalling alignment, whereas *arvates* ‘in [PN’s] opinion’ more explicitly frames the proposition as the source’s viewpoint and separates it from the dissertation writer’s own voice. The most frequent reporting verb, *välja tooma* ‘point out/highlight’, gives the attributed proposition greater prominence by representing the source as drawing attention to a particular point, although it does not necessarily signal agreement. By contrast, *väitma* ‘claim’ may create greater distance, whereas *kinnitama* ‘confirm’ and *näitama* ‘show’ may have stronger endorsing potential (see also Thompson & Ye 1991, 372; Charles 2006). As evaluative positioning was not coded exhaustively, these patterns are treated as tendencies rather than quantified contrasts.

3.3 Endophoric markers: linear navigation and whole-text reflexivity

Endophoric markers are examined in terms of the distribution of previewing, reviewing and whole-text references, together with their formal realisations and lexical profiles.

3.3.1 Distribution of endophoric subcategories

The 1,379 endophoric markers comprised 638 whole-text references (46.3%), 503 previewing references (36.5%) and 238 reviewing references (17.3%). The ordering was stable across the four field-by-level groups: whole-text references were generally the most frequent, previewing references the second most frequent and reviewing references the least frequent.

Sectional distributions were considerably more distinctive. Introductions had the highest endophoric density, with previewing and whole-text references occurring at almost identical rates (78 and 81 per 10,000 words). Conclusions displayed the clearest retrospective profile: whole-text references reached 56 per 10,000 words and reviewing references 21. Methods sections combined whole-text reference (28) with previewing (24), while literature reviews and results/discussion sections contained substantially fewer endophoric markers overall (Figure 4).

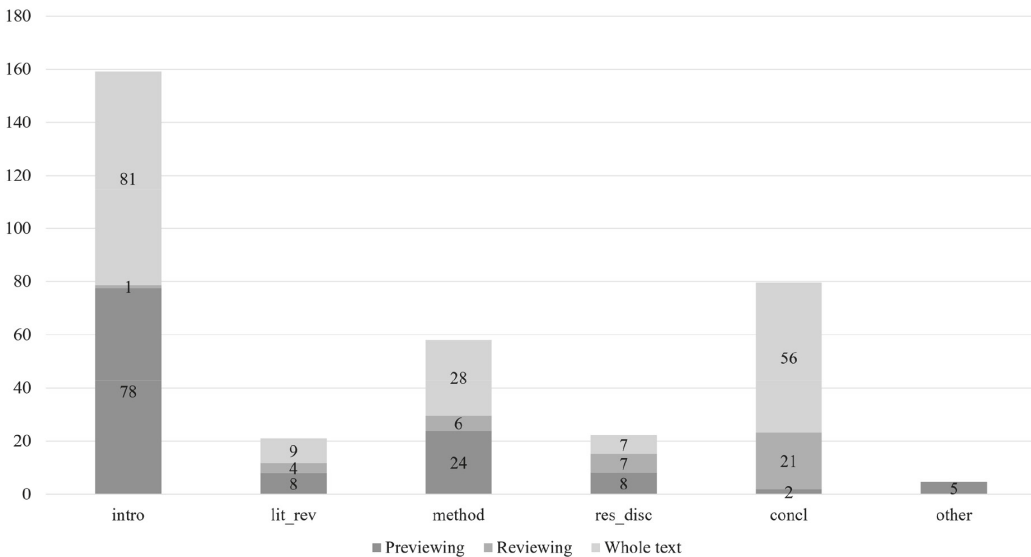


Figure 4. Distribution of endophoric markers across rhetorical sections; per 10,000 words

The conditional inference tree (see Figure 5) again selected section as the first partition, and study level did not appear in the tree. Conclusions formed a distinctive terminal group in which whole-text references prevailed, followed by reviewing references and a small proportion of previewing references. Introductions and other textual locations were characterised by an almost even division between whole-text and previewing references, with virtually no reviewing. Literature review and methods sections also favoured whole-text and previewing markers, although reviewing was somewhat more visible. Results/discussion sections were the only context in which field produced an additional split. Humanities texts divided almost evenly between previewing and reviewing references, with fewer whole-text markers, whereas social science texts showed a stronger whole-text profile. The tree is therefore consistent with a predominantly section-based pattern in this corpus, with an additional field split confined to results/discussion sections.

3.3.2 Formal realisations and lexical profiles of endophoric markers

Overall, endophoric relations were realised primarily through text-denoting noun phrases, with deictic expressions providing additional directional cues. Whole-text references centred on nouns such as (*bakalaureuse/magistri*)*töö* ‘(bachelor’s/master’s) dissertation’ (n=501), followed by *uurimus/uurimistöö* ‘study/research’ (n=109), often accompanied by *käesolev* or *siinne* ‘present/current’. Previewing markers likewise predominantly named a forthcoming chapter, subsection or part (n=345), referred to an appendix (n=104) or used forward-pointing expressions such as *edaspidi* ‘later’ or *allpool* ‘below’ (n=40). Reviewing markers relied mainly on noun phrases identifying

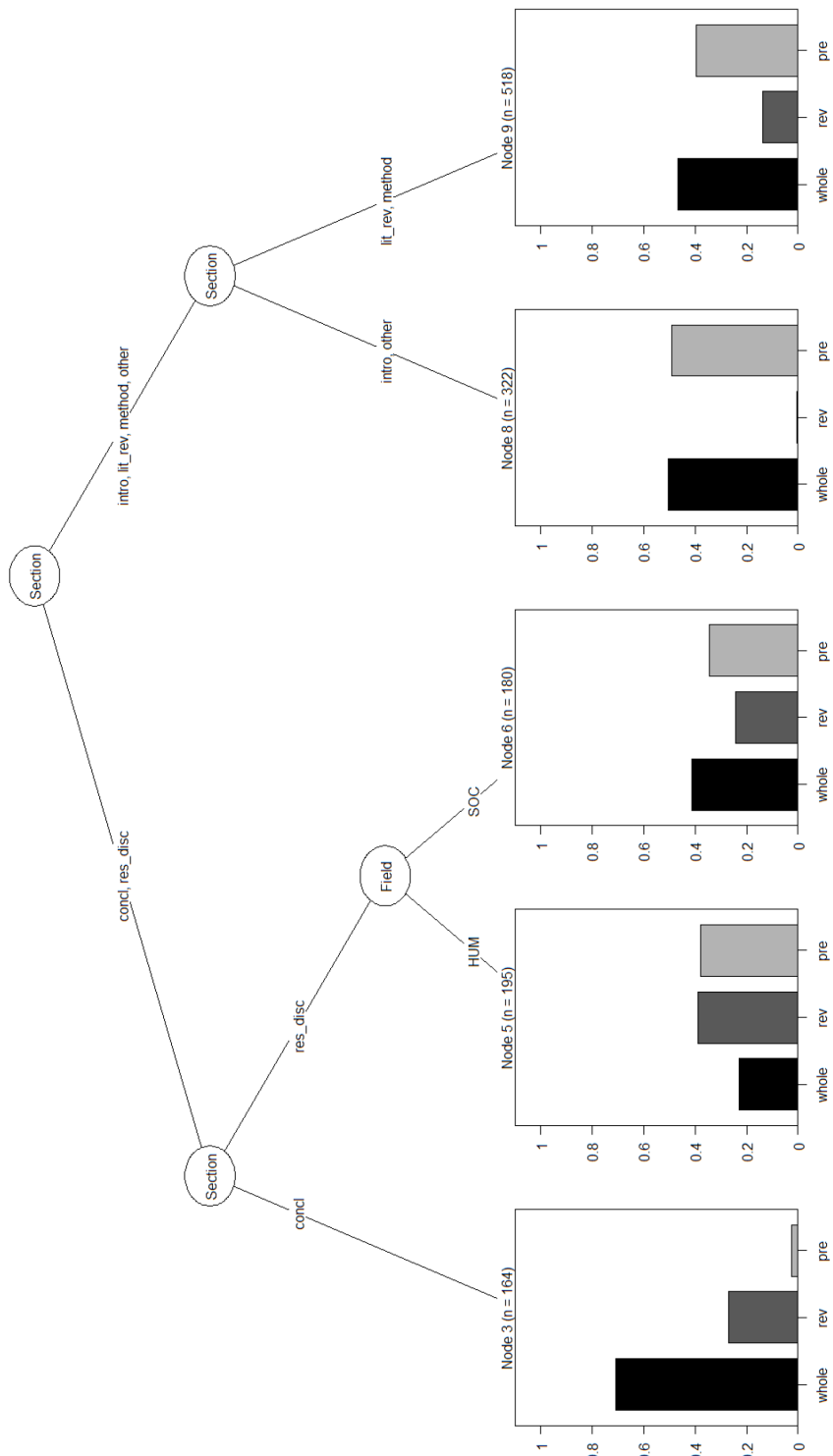


Figure 5. Conditional inference tree for whole-text, previewing and reviewing endophoric markers. Terminal bars show subcategory proportions; n indicates the number of occurrences in the node

an earlier textual unit, such as a chapter, subsection or part (n=160), supplemented by backward-pointing expressions such as *eespool/eelpool* ‘above’ (n=44) and formulaic clauses such as *nagu öeldud* ‘as stated’ (n=14).

An associated lexical item was assigned to 1,265 out of 1,379 endophoric markers. The remaining cases mostly occurred either in parentheses without additional lexical co-text or in phrases containing multiple automatically identified markers (e.g. *töö teoreetiline osa* ‘the theoretical **part** of the **dissertation**’) and therefore received no separate lexical annotation. Verbs predominated (n=1,034; 81.7%), followed by nouns (n=186; 14.7%), imperative forms and their abbreviated equivalents, such as *vaata* and *vt* ‘see’ (n=28; 2.2%), and adverbial/adpositional expressions (n=17; 1.3%).

The lexical profiles differed across the three subcategories. Table 3 lists the most frequent associated lexical items in whole-text, previewing and reviewing references.

Whole text	Previewing	Reviewing
102 <i>eesmärk</i> ‘aim’	50 <i>andma ülevaadet</i> ‘provide an overview’	34 <i>välja tooma</i> ‘point out/highlight’
53 <i>kasutama</i> ‘use’	36 <i>kirjeldama</i> ‘describe’	22 <i>ütleva</i> ‘say’
23 <i>selguma</i> ‘emerge/become evident’	31 <i>välja tooma</i> ‘point out/highlight’	17 <i>käsitleva</i> ‘address’
19 <i>keskenduma</i> ‘focus’	25 <i>käsitleva</i> ‘address’	17 <i>mainima</i> ‘mention’
18 <i>uurima</i> ‘investigate’	22 <i>vt</i> ‘see’ (imperative)	12 <i>kirjeldama</i> ‘describe’
18 <i>koosnema</i> ‘consist of’		8 <i>selguma</i> ‘emerge/become evident’

Table 3. Most frequent associated lexical items by endophoric subcategory

Whole-text references were led by *eesmärk* ‘aim’ and verbs such as *kasutama* ‘use’, *selguma* ‘emerge/become evident’, *keskenduma* ‘focus’, *uurima* ‘investigate’ and *koosnema* ‘consist of’. These items present the dissertation as both a purposeful research project and a textual product by stating aims and focus, describing methods, reporting overall findings and outlining organisation.

Previewing references favoured verbs representing discourse activity, such as *andma ülevaadet* ‘provide an overview’, *kirjeldama* ‘describe’, *välja tooma* ‘point out/highlight’, *käsitleva* ‘address’, *tutvustama* ‘introduce’, *selgitama* ‘explain’ and *esitama* ‘present’. Reviewing references were characterised by verbs of retrospective mention and reporting, including *välja tooma* ‘point out/highlight’, *ütleva* ‘say’, *käsitleva* ‘address’, *mainima* ‘mention’, and *kirjeldama* ‘describe’.

Given the prominence of *eesmärk* ‘aim’ among whole-text references, aim constructions were examined more closely. They were concentrated in introductions and conclusions, although individual formulations could recur across sections (Example 3).

- (3) [Intro:] *Käesoleva bakalaureusetöö eesmärk on analüüsida holonüümia ja meronüümia ehk osa-terviku suhteid Eesti Wordnetis.* [Method:] *Töö eesmärk on [...].* [Concl:] *Bakalaureusetöö eesmärk oli [...].* (HUM_BA_2019_2)

[Intro:] ‘The aim of the present bachelor’s dissertation is to analyse holonymy and meronymy, that is, part-whole relations, in Estonian Wordnet.’ [Method:] ‘The aim of the dissertation is [...].’ [Concl:] ‘The aim of the bachelor’s dissertation was [...].’

A closer examination of this subset identified three recurrent rhetorical orientations in the construction of dissertation aims. Inquiry-oriented aims construed the purpose of the study as generating or establishing knowledge, using verbs such as *välja selgitama* ‘find out’, *uurima* ‘investigate’, *tuvastama* ‘identify’ and *kaardistama* ‘map’. Text-oriented aims construed the purpose as organising or communicating the content of the dissertation, using verbs such as *andma ülevaadet* ‘provide an overview’, *kirjeldama* ‘describe’ and *käsitlema* ‘address’. Contribution-oriented aims foregrounded an intended scholarly or practical outcome, using expressions such as *andma soovitusi* ‘make recommendations’, *panustama* ‘contribute’ and *optimeerima* ‘optimise’.

Across the three endophoric subcategories, the most frequent expressions formed a relatively conventional and rhetorically restrained repertoire, generally describing aims, procedures, textual organisation or findings rather than explicitly promoting the dissertation’s novelty or contribution. Less frequent verbs and constructions nevertheless foregrounded contribution, novelty or intended impact more explicitly, as in *tõstatama* ‘raise [an issue]’, *heitma valgust* ‘shed light’, *andma panust* ‘make a contribution’ and *pakkuma uut lähenemist* ‘offer a new approach’.

3.4 Code glosses, transitions and frame markers: distribution and frequent realisations

Among 1,018 code glosses, elaboration was the most frequent function (n=502; 49.3%), followed by reformulation (n=277; 27.2%) and exemplification (n=239; 23.5%). The multifunctional *ehk* ‘or/same as’ was prominent in both elaborative (n=234) and reformulative uses (n=261), while *tähendama/see tähendab* ‘mean/that is’ was frequent in elaboration (n=145). All exemplifying code glosses were realised by *näiteks* ‘for example’, whereas explicit reformulations with *teisisõnu/teiste sõnadega* ‘in other words’ were rare (n=6).

Transitions (n=996) were distributed fairly evenly between additive (n=366; 36.7%) and consequential relations (n=357; 35.8%), while comparative or contrastive relations were somewhat less frequent (n=273; 27.4%). The most common forms were *lisaks* ‘in addition’ (n=132), *seega* ‘thus’ (n=123), *samuti* ‘also’ (n=109), *seetõttu/mistõttu/seepärast* ‘therefore/as a result’ (n=104) and *teisalt* ‘on the other hand’ (n=100).

Frame markers (n=437) primarily signalled sequence (n=281; 64.3%), particularly *esimene/esiteks* ‘first(ly)’ (n=94) and *teine/teiseks* ‘second(ly)’ (n=67). Text-stage markers were less frequent (n=156; 35.7%), with *järgmine/alljärgnev* ‘next/following’ (n=99) and *kokkuvõttes/kokkuvõlvalt* ‘in conclusion/summary’ (n=56) predominating.

4 Discussion and conclusions

The clearest finding is that interactive reader guidance in the present corpus of Estonian university dissertations is structured more strongly by rhetorical section than by field or study level. Bachelor’s and social science dissertations have higher overall rates, but category ordering remains stable across these subcorpora. Rhetorical sections, by contrast, show distinct profiles: literature reviews are source-oriented, introductions and conclusions text-reflexive and methods and results/discussion sections more mixed. These section-based differences are visible in the normalised frequencies and MCA, and rhetorical section was selected for the first partition in both conditional inference trees. Field and study level appeared in more limited contrasts. Taken together, the analyses suggest more pronounced section-based patterning within this corpus, but they do not establish that rhetorical section has a statistically stronger association with marker distribution in Estonian university dissertations generally. However, this section sensitivity accords with Estonian and Baltic research articles (Leijen et al. 2024, 124; Hint et al. 2025, 59), suggesting that some research-report conventions may extend across languages and between expert and student genres.

The higher aggregate rates in bachelor’s and social-science dissertations should not be interpreted as straightforward developmental or disciplinary effects. The field contrast was concentrated particularly in evidentials, including integral evidentials, while study-level differences varied across field and rhetorical section. Given the broad field groupings and exploratory design, the study cannot determine whether these differences reflect disciplinary conventions, students’ writing experience or a combination of the two. Nor should the higher rates be taken as evidence of better or more reader-oriented writing: more metadiscourse is not intrinsically better, since both excessive and insufficient use may hinder readability, and its effectiveness depends on the rhetorically appropriate use of particular resources (Steffensen & Cheng 1996, 163; Crismore & Abdollahzadeh 2010, 202).

Evidentials dominate the corpus, but their frequency is largely driven by literature reviews. This finding aligns with previous studies that have identified this section’s central role in academic writing (cf. Thompson 2005, 317; Ahn & Oh 2024, 6). Two thirds of evidentials are non-integral, indicating a general preference for proposition-prominent citation in which the source is backgrounded syntactically. This is a formal finding about information-prominence, not direct evidence that citation is uncritical

or used only to display knowledge. Citation form – integral or non-integral – does not determine rhetorical function, and both forms can support attribution, synthesis, comparison, application or evaluation (Hyland 1999, 356; Thompson 2005, 321; Petrić 2007, 242; Samraj 2013, 302).

However, this pattern resembles the predominance of non-integral evidentials reported for Lithuanian linguistics research articles, although the proportion is somewhat lower in the present dissertation corpus (66% compared with 76% in Ruskan and Maslauskienė 2023). A recent study of Estonian students' citation practices (Tammepõld 2025) likewise found a preference for non-integral citations, which accounted for 75% of citations in bachelor's dissertations and 72% in master's dissertations. Tammepõld's study, however, covered all major disciplinary domains and used a substantially larger corpus. Comparisons must therefore remain cautious, as the datasets differed in genre, disciplinary coverage and size. Nevertheless, all three suggest that parenthetical source attribution is a common resource in local academic writing.

The present data also show that source prominence varies within the dissertation. Integral citation becomes proportionally more common in results/discussion and conclusion sections, especially in social science master's dissertations. One interpretation could be that writers are more likely to foreground authors or studies when comparing findings, interpreting results or positioning their findings or conclusions in relation to previous work. Source prominence therefore appears to be section- and field-sensitive and should not be reduced to a simple developmental marker. A future functional analysis of the citations would be needed to determine whether these integral constructions perform more comparison, evaluation or source synthesis.

Endophoric marker distributions likewise varied systematically across rhetorical sections of the dissertations. Introductions combine whole-text reference with previewing, methods remain forward-oriented and conclusions combine whole-text reference with reviewing. Dissertation writers repeatedly establish aims, focus, procedures and organisation, then revisit the completed work and its findings in conclusions. This whole-text and prospective orientation parallels Estonian research articles, whereas Lithuanian articles showed more reviewing and fewer whole-text references (Hint et al. 2024, 57). Despite genre and disciplinary differences, the dissertation pattern therefore resembles the prospective and whole-text orientation of Estonian expert writing more closely than the retrospective profile reported for Lithuanian articles.

The MCA positioned transitions, frame markers and code glosses relatively close to one another and to results/discussion sections, suggesting that they share a broad discourse-relational profile in the present corpus. Compared with evidentials and endophoric markers, these resources generally organise relations among propositions, discourse

stages or formulations within the unfolding discourse rather than directing readers to an external source or an explicitly identified location elsewhere in the dissertation. This proximity does not, however, remove the functional distinctions among them. Borderline cases nevertheless arose during manual validation. For example, *esiteks*, *teiseks* and *kolmandaks* ‘firstly’, ‘secondly’ and ‘thirdly’ both sequence points and introduce additional items, but were classified as frame markers when ordering was their dominant function. Similarly, *näiteks* ‘for example’ was treated here as an exemplifying code gloss, although other frameworks may group illustration with transitions. Aull (2015, 142), for example, includes sequencing and illustration among transition-marker functions. At a broader interpretive level, the results thus suggest three complementary modes of interactive reader guidance: discourse-relational and elaborative guidance through transitions, frame markers and code glosses; intertextual source attribution through evidentials; and intratextual navigation through endophoric markers.

Earlier Estonian research using broader transition criteria found transitions to be the largest metadiscourse category in expert linguistics articles (Hint et al. 2025, 170), while Lithuanian linguistics articles have also shown transitions and evidentials to be the most frequent interactive resources (Ruskan & Maslauskienė 2023, 97). The category hierarchy observed here must, however, be interpreted in light of the operational decisions detailed in Section 2.2, especially the narrow transition criteria, the classification of whole-text references as endophoric and the exclusion of references to non-linear elements. It should therefore not be treated as directly comparable evidence of a different writing tradition.

The lexical profiles of integral evidentials point to a largely conventional and relatively neutral repertoire. Student writers most often represent sources through discourse acts, especially *välja tooma* ‘point out/highlight’, while *sõnul* ‘according to’ and related expressions acknowledge external voices without overtly specifying alignment. Research-act verbs introduce reported procedures and findings, while items such as *kinnitama* ‘confirm’ and *näitama* ‘show’ may have stronger endorsing potential. More distancing or epistemically tentative lexical items, such as *väitma* ‘claim’ and *oletama* ‘assume’, are less frequent, and no unambiguous instances of the writer overtly challenging a cited source were identified. However, the absence of explicitly negative reporting verbs should not be equated with an absence of criticality, since evaluation may also be conveyed through other means, such as negation, countering or contrastive relations, and expressions of support or rejection elsewhere in the surrounding text (Thompson & Ye 1991, 372; Martin & White 2005, 159). This broader view is consistent with Mur-Dueñas’s (2012, 406) treatment of critical acts as functional rather than grammatical units that may extend across clauses or sentences. In her analysis of finance research article introductions, negative evaluation was more often directed impersonally at the literature or disciplinary community than at the work of individual scholars. Moreover,

as noted in Section 3.2.2, these patterns are best understood as tendencies in source framing rather than a complete account of student criticality.

The lexical profiles of endophoric markers, on the other hand, reinforce the dual status of the dissertation as research activity and textual product, while a smaller set of aim formulations additionally foregrounds intended contribution. Within aim constructions, inquiry-oriented formulations represent the project as an investigation, text-oriented formulations present it as an organised discourse and contribution-oriented formulations foreground an intended outcome. More broadly, terms for findings and verbs representing textual representation reinforce the relationship between the dissertation as a research project and as a textual product. This duality may also explain why the boundary between frame markers and whole-text endophoric markers is difficult to maintain consistently.

The one-item lexical annotation applied here offers a useful middle level between aggregate category counts and exhaustive close analysis of every occurrence. It makes a large number of occurrences comparable, reveals the dominant vocabulary associated with each function and identifies frequent lexical items and contexts that merit closer inspection. It does not, however, capture every contribution made by a multiword construction or its wider argumentative co-text. The results therefore describe dominant lexical associations rather than complete phraseological analyses.

For academic writing instruction, the most useful outcome is not a prescriptive list of approved expressions but a section-sensitive functional map of choices. Corpus examples can show how literature reviews manage dense source attribution, how introductions preview and represent the dissertation as a whole and how conclusions review the completed text and its findings. Within these rhetorical contexts, students can compare how the choice between integral and non-integral citation distributes attention between information and source; how verbs represent saying, thinking and research activity; how source-framing expressions may acknowledge, distance or endorse external voices; and how previewing, reviewing and whole-text references support readers at different stages of an extended text. The aim should not be to replace one form with another, but to expand students' control over the available alternatives and their rhetorical consequences. Corpus examples from Estonian dissertations would make such instruction responsive to local writing practices rather than relying on imported English-language formulae.

Several limitations constrain the conclusions. The corpus comprises 40 dissertations from one university and two broad disciplinary domains. The findings therefore cannot be generalised to Estonian academic writing as a whole. Because marker occurrences are nested within individual dissertations, the MCA and conditional inference trees should be interpreted as exploratory descriptions of patterning in this corpus rather than as inferential tests comparing the strength of the associations between marker distribution

and rhetorical section, field or study level. The material was annotated by one researcher, and the reliability of the study-specific adaptations – particularly the associated lexical items and verb-denotation categories – was not independently assessed. Finally, the operational decisions described above limit direct cross-study comparisons.

Future research should expand the corpus to additional disciplines and doctoral writing. Matched comparisons of student and expert texts within the same disciplines would help determine whether the observed patterns are associated with writing experience, genre or disciplinary practice. A fuller constructional analysis of source framing would complement the present one-item lexical annotation and build on calls to examine the phraseological realisation of metadiscursive functions beyond individual lexico-grammatical features (Mur-Dueñas 2011, 3075). Systematic coding of evaluative positioning would likewise give useful information on citation practices. Corpus-based analysis could also be complemented by qualitative methods, e.g. text-based interviews in which writers discuss their own texts and the considerations underlying their metadiscursive choices, which cannot be inferred from textual patterns alone. Finally, comparison with corresponding Baltic student corpora using harmonised annotation would further develop regional research on academic writing.

Abbreviations

BA – bachelor’s dissertation	method – methods section
code – code gloss	non_i – non-integral evidential
concl – conclusions section	other – other textual location
endo – endophoric marker	PN – proper name
evid – evidential	pre – previewing endophoric marker
frame – frame marker	res_disc – results/discussion section
HUM – humanities and arts	rev – reviewing endophoric marker
int – integral evidential	SOC – social sciences
intro – introduction section	trns – transition
lit_rev – literature review section	whole – whole-text endophoric marker
MA – master’s dissertation	

References

- Ahn, Chae-Young & Sun-Young Oh. 2024. Citation practices in applied linguistics: A comparative study of Korean master’s theses and research articles. *Journal of English for Academic Purposes* 69, article 101369. <https://doi.org/10.1016/j.jeap.2024.101369>
- Aull, Laura. 2015. *First-year university writing. A corpus-based study with implications for pedagogy*. New York: Palgrave Macmillan.

- Bunton, David. 1999. The use of higher level metatext in Ph.D theses. *English for Specific Purposes* 18 (1), S41–S56. [https://doi.org/10.1016/S0889-4906\(98\)00022-2](https://doi.org/10.1016/S0889-4906(98)00022-2)
- Burneikaitė, Nida. 2006. Evidentiality in graduate student writing: A study of Lithuanian students' master's theses in English. *Kalba ir kontekstai* 1, 97–105.
- Burneikaitė, Nida. 2009. Endophoric markers in linguistics master's theses in English L1 & L2. *Žmogus ir žodis* 3, 11–16.
- Charles, Maggie. 2006. Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* 25 (3), 310–331. <https://doi.org/10.1016/j.esp.2005.05.003>
- Cheng, Xiaoguang & Margaret S. Steffensen. 1996. Metadiscourse: A technique for improving student writing. *Research in the Teaching of English* 30 (2), 149–181.
- Crismore, Avon & Esmaeel Abdollahzadeh. 2010. A review of recent metadiscourse studies: The Iranian context. *Nordic Journal of English Studies* 9 (S2), 195–219. <https://doi.org/10.35360/njes.223>
- Deng, Liming & Xinyue Li. 2026. Reporting verbs across genre sets and research paradigms: A corpus-based analysis of denotation and evaluation in applied linguistics research articles. *Frontiers in Psychology* 17, article 1799996. <https://doi.org/10.3389/fpsyg.2026.1799996>
- Dubova, Agnese, Diāna Laiveniece, Baiba Egle & Dzintra Lele-Rozentāle. 2023. Metadiskursa analīzes teorija latviešu zinātnisko tekstu pētniecībai. [Metadiscourse analysis theory in Latvian scientific text research]. *Linguistica Lettica* 32, 170–195. <https://doi.org/10.22364/lingualet.32>
- Greenacre, Michael. 2013. The contributions of rare objects in correspondence analysis. *Ecology* 94 (1), 241–249. <https://doi.org/10.1890/11-1730.1>
- Groom, Nicholas. 2000. Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. *Patterns and perspectives: insights into EAP writing practice*. Paul Thompson, ed. Reading: Centre for Applied Language Studies. 14–25.
- Hint, Helen, Helena Lemendik, Christer Johansson & Djuddah Leijen. 2025. Decoding metadiscourse markers in Estonian academic texts: A language-specific perspective. *Written Communication* 42 (1), 152–192. <https://doi.org/10.1177/07410883241286901>
- Hint, Helen, Anna Ruskan, Helena Lemendik & Baiba Egle. 2024. “At the end of this study, see the following discussion”: Endophoric markers in Estonian, Latvian, and Lithuanian research articles. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian Papers in Applied Linguistics* 20, 45–62. <https://doi.org/10.5128/ERYa20.03>
- Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15 (3), 651–674. <https://doi.org/10.1198/106186006X133933>

- Hyland, Ken. 1999. Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics* 20 (3), 341–367. <https://doi.org/10.1093/applin/20.3.341>
- Hyland, Ken. 2005. *Metadiscourse: Exploring interaction in writing*. London, New York: Continuum.
- Hyland, Ken. 2017. Metadiscourse: What is it and where is it going? *Journal of Pragmatics* 113, 16–29. <https://doi.org/10.1016/j.pragma.2017.03.007>
- Hyland, Ken & Feng (Kevin) Jiang. 2018. “In this paper we suggest”: Changing patterns of disciplinary metadiscourse. *English for Specific Purposes* 51, 18–30. <https://doi.org/10.1016/j.esp.2018.02.001>
- Hyland, Ken & Feng (Kevin) Jiang. 2024. Metadiscourse: The evolution of an approach to texts. *Text & Talk* 44 (3), 411–433. <https://doi.org/10.1515/text-2021-0156>
- Hyland, Ken, Wenbin Wang & Feng (Kevin) Jiang. 2022. Metadiscourse across languages and genres: An overview. *Lingua* 265, article 103205. <https://doi.org/10.1016/j.lingua.2021.103205>
- Jürine, Anni, Djuddah Leijen, Helen Hint, Jolanta Šinkūnienė, Diāna Laiveniece, Christer Johansson & Nicholas Groom. 2021. Academic writing in the Baltic states: Introducing the Bwrite project. *Educare* 1, 27–37. <https://doi.org/10.24834/educare.2021.1.3>
- Kassambara, Alboukadel & Fabian Mundt. 2020. *factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7*. <https://CRAN.R-project.org/package=factoextra>
- Khabbazi-Oskouei, Leila. 2013. Propositional or non-propositional, that is the question: A new approach to analyzing ‘interpersonal metadiscourse’ in editorials. *Journal of Pragmatics* 47 (1), 93–107. <https://doi.org/10.1016/j.pragma.2012.12.003>
- Lahodová Vališová, Marie. 2024. Endophoric signposting: A contrastive study of textual references in L2 Czech master’s theses and native English academic writing. *Miscelánea: A Journal of English and American Studies* 70, 15–40. https://doi.org/10.26754/ojs_misc/mj.20249976
- Lê, Sébastien, Julie Josse & François Husson. 2008. FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software* 25 (1), 1–18. <https://doi.org/10.18637/jss.v025.i01>
- Leijen, Djuddah A. J., Helen Hint, Helena Lemendik, Baiba Egle, Anna Ruskan & Christer Johansson. 2024. Cross-linguistic patterns of metadiscourse: Disciplinary similarities and section-based differences. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian Papers in Applied Linguistics* 20, 115–132. <https://doi.org/10.5128/ERYa20.07>
- Lemendik, Helena. 2025. “Teaduskeele tunnusteks on täpsus, neutraalsus ja impersonaalsus.” Kuidas siiski haakuvad juurdunud tavadega tänapäeva akadeemilised tekstid? [“The key features of scientific language are precision,

- neutrality, and impersonality.” Yet, how do contemporary academic texts align with these traditional expectations?]. *Eesti ja soome-ugri keeleteaduse ajakiri = Journal of Estonian and Finno-Uric Linguistics* 16 (1), 41–79. <https://doi.org/10.12697/jeful.2025.16.1.02>
- Liardét, Cassi L. & Sharyn Black. 2019. “So and so” says, states and argues: A corpus-assisted engagement analysis of reporting verbs. *Journal of Second Language Writing* 44, 37–50. <https://doi.org/10.1016/j.jslw.2019.02.001>
- Lindström, Liina. 2017. Lause infostruktuur ja sõnajärg. [Information structure and word order in the sentence]. *Eesti keele süntaks. [Estonian syntax]*. Mati Ereht & Helle Metslang, eds. Tartu: Tartu Ülikooli Kirjastus. 547–565.
- Linkevičienė, Nijolė & Jolanta Šinkūnienė. 2012. Asmeniniai įvardžiai mokslo kalboje. [Personal pronouns in Lithuanian scientific discourse]. *Kalbotyra* 64 (3), 78–102. <https://doi.org/10.15388/Klbt.2012.7662>
- Martin, J. R. & Peter R. R. White. 2005. *The language of evaluation: appraisal in English*. Palgrave Macmillan.
- 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>
- Mur-Dueñas, Pilar. 2012. Critical acts in published and unpublished research article introductions in English: A look into the writing for publication process. *International Advances in Writing Research: Cultures, Places, Measures*. Charles Bazerman, Chris Dean, Jessica Singer Early, Karen Lunsford, Suzie Null, Paul Rogers & Amanda Stansell, eds. Fort Collins: TheWAC Clearinghouse; Anderson: Parlor Press. 403–420.
- Nesi, Hilary. 2021. Sources for courses: Metadiscourse and the role of citation in student writing. *Lingua* 253, article 103040. <https://doi.org/10.1016/j.lingua.2021.103040>
- Pearson, William S. & Esmaeel Abdollahzadeh. 2023. Metadiscourse in academic writing: A systematic review. *Lingua* 293, article 103561. <https://doi.org/10.1016/j.lingua.2023.103561>
- Petrauskaitė, Rūta & Jolanta Šinkūnienė. 2015. Dar kartą apie intertekstualumą. Ką jis sako apie mokslinį tekstą? [Intertextuality in research writing revisited]. *Kalbotyra* 67, 67–85. <https://doi.org/10.15388/Klbt.2015.8943>
- Petrić, Bojana. 2007. Rhetorical functions of citations in high- and low-rated master’s theses. *Journal of English for Academic Purposes* 6 (3), 238–253. <https://doi.org/10.1016/j.jeap.2007.09.002>
- R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ruskan, Anna. 2012. Evidential adjectives in Lithuanian academic discourse. *Kalbotyra* 64 (3), 103–123. <https://doi.org/10.15388/Klbt.2012.7663>
- Ruskan, Anna, Helen Hint, Djuddah Arthur Joost Leijen & Jolanta Šinkūnienė. 2023. Lithuanian academic discourse revisited: Features and patterns of scientific

- communication. *Open Linguistics* 9 (1), article 20220231. <https://doi.org/10.1515/opli-2022-0231>
- Ruskan, Anna & Greta Maslauskienė. 2023. Metadiscourse in Lithuanian linguistics research articles: A study of interactive and interactional features. *Kalbotyra* 76, 90–120. <https://doi.org/10.15388/Kalbotyra.2023.76.6>
- Samraj, Betty. 2013. Form and function of citations in discussion sections of master's theses and research articles. *Journal of English for Academic Purposes* 12, 299–310. <https://doi.org/10.1016/j.jeap.2013.09.001>
- Steffensen, Margaret S. & Xiaoguang Cheng. 1996. Metadiscourse and text pragmatics: How students write after learning about metadiscourse. *Pragmatics and language learning* 7. Lawrence F. Bouton, ed. 153–170.
- Swales, John. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Šinkūnienė, Jolanta. 2008. Autoriaus pozicijos švelninimas: tarpdalykiniai ir tarpkalbiniai raiškos priemonių ypatumai. [Hedging in Lithuanian and English research articles: A cross-disciplinary and cross-linguistic study]. *Kalbotyra* 58 (3), 97–108. <https://doi.org/10.15388/Klbt.2008.7586>
- Šinkūnienė, Jolanta. 2010. Autoriaus pozicijos raiška asmeniniais įvardžiais rašytiniame akademiname diskurse. [Personal pronouns and authorial presence in written academic discourse]. *Filologija* 15, 124–141.
- Šinkūnienė, Jolanta. 2011. *Autoriaus pozicijos švelninimas rašytiniame moksliniame diskurse: gretinamasis tyrimas*. [Hedging in written scientific discourse: A cross-linguistic and cross-disciplinary study], ms. Humanitarinių mokslų daktaro disertacija. Vilnius: Vilniaus universitetas.
- Šinkūnienė, Jolanta. 2012. Adverbials as hedging devices in Lithuanian academic discourse: A cross-disciplinary study. *Multiple perspectives in linguistic research on Baltic languages*. Aurelija Usonienė, Nicole Nau & Ineta Dabašinskienė, eds. Newcastle upon Tyne: Cambridge Scholars Publishing. 137–167.
- Šinkūnienė, Jolanta. 2014. *Lietuviškojo humanitarinių ir socialinių mokslų diskurso ypatybės: Mokslo studija*. [Insights into Lithuanian academic discourse of the humanities and social sciences]. Vilnius: Vilniaus universiteto leidykla.
- Šinkūnienė, Jolanta. 2017. Citations in research writing: the interplay of discipline, culture and expertise. *Cross-linguistic correspondences: from lexis to genre*. Thomas Egan & Hildegunn Dirdal, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company. 253–270. <https://doi.org/10.1075/slcs.191.10sin>
- Šinkūnienė, Jolanta. 2018. The power of English: *I* and *we* in Lithuanian, Lithuanian English and British English research writing. *Intercultural perspectives on research writing*. Pilar Mur-Dueñas & Jolanta Šinkūnienė, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company. 59–79. <https://doi.org/10.1075/aals.18.03sin>

- Šinkūnienė, Jolanta. 2019. Reformulation markers in academic discourse: A cross-linguistic and cross-disciplinary study. *Jezikoslovlje* 20 (3), 531–553. <https://doi.org/10.29162/jez.2019.19>
- Tammepõld, Marri-Mariska. 2025. *Viitamine kui nähtav intertekstuaalsus ja evidentsiaalsus üliõpilaste akadeemilistes tekstides*. [Citing sources as manifest intertextuality and evidentiality in university students' academic writing]. Magistritöö. [Master's dissertation]. Tartu: University of Tartu.
- Thompson, Geoff & Yiyun Ye. 1991. Evaluation in the reporting verbs used in academic papers. *Applied Linguistics* 12 (4), 365–382. <https://doi.org/10.1093/applin/12.4.365>
- Thompson, Paul. 2005. Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes* 4 (4), 307–323. <https://doi.org/10.1016/j.jeap.2005.07.006>
- Thompson, Paul & Chris Tribble. 2001. Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology* 5 (3), 91–105.
- Van Rossum, Guido & Fred L. Drake, ed. 2001. *Python reference manual*. PythonLabs.