# HEDGING MODAL ADVERBS IN SLOVENIAN ACADEMIC DISCOURSE

Jakob LENARDIČ, Darja FIŠER
Faculty of Arts, University of Ljubljana; Jožef Stefan Institute

*Lenardič, J., Fišer, D. (2021): Hedging modal adverbs in Slovenian academic discourse. Slovenščina 2.0, 9(1): 145–180.*

*DOI:* https://doi.org/10.4312/slo2.0.2021.1.145-180

This paper first presents a comparative analysis of modal adverbs in doctoral theses in the humanities and social sciences on the one hand, and in natural and technical sciences on the other from the 1.7-billion-token corpus of Slovenian academic texts KAS (Erjavec et al., 2019a). Using a randomized concordance analysis, we observe the epistemic and non-epistemic usage of the modal adverbs and show that epistemic adverbs are more characteristic of the humanities and social sciences theses. We also show that the non-epistemic dispositional meaning of possibility, which is most commonly used in natural and technical sciences theses, is not used as a hedging device. In the second part of the paper we compare the usage of a selected set of modals in bachelor's, master's and doctoral theses in order to chart how researchers' approach to stance-taking changes at different proficiency levels in academic writing, showing that the observed increase in hedging devices in doctoral theses seems to be less a function of an increased proficiency level in academic writing as such and more the result of conceptual differences between undergraduate and postgraduate theses, only the latter of which are original research contributions with extensive discussion of the results.

**Keywords:** epistemic modality, root modality, hedging, semantics, pragmatics, corpus linguistics

## 1 INTRODUCTION

Modal expressions offer an interesting insight into academic discourse because they can pragmatically function as *hedges* (Lakoff, 1972; Hyland, 1996, 1998), which are used by authors to present their claims with varying degrees of tentativeness. In academic writing, hedging is a particularly important pragmatic device, as it "enables writers to express a perspective on their statements, to present unproven claims with caution, and to enter into a dialogue with their audiences" and is therefore an "important means by which professional scientists confirm their membership in research communities" (Hyland, 1996, pp. 251–252).

In related work, which has primarily focused on English academic discourse, it is often shown that hedging is more characteristic of humanities and social sciences rather than natural and technical sciences (Hyland, 1998; Takimoto, 2015), which reflects the general idea that humanities and social sciences are more interpretative and less rooted in empirical research than natural and technical sciences (Takimoto, 2015). In this paper, we try to confirm whether this is also the case for Slovenian academic discourse on the basis of the doctoral theses in the *KAS corpus of Slovenian academic writing* (Erjavec et al., 2019a).[1] We present a quantitative analysis of the most frequent modal adverbs that display epistemic and possibly non-epistemic meanings and then conduct a randomized concordance analysis to determine whether the modals that pragmatically serve as hedging devices are also used more frequently in the humanities and social sciences.

Apart from cross-disciplinary comparisons, hedging in academic discourse has also been studied from the perspective of its developmental trajectory (Hyland, 2004; Lancaster, 2016) where it is compared between early forms of academic writing such as (under)graduate research papers on the one hand and published academic writing on the other in order to chart how researchers' approach to stance-taking changes as they gain experience in academic

1 This paper is an extended version of the conference paper Lenardič and Fišer (2020). We have employed a more fine-grained classification of epistemic modality, which has allowed us to take additional evidential/assumptive modals into consideration as well. Furthermore, we now also compare the prominence of hedging in PhD theses with hedging in bachelor's and master's theses on the basis of a relevant subset of the analysed modals.

writing (Aull and Lancaster, 2014). We contribute to this line of research by comparing a subset of the most frequent modal adverbs between the doctoral theses on the one hand and the bachelor's and master's theses in the *KAS corpus* (Erjavec et al., 2019a) on the other, namely, the subset of those modals that invariably play a hedging role in terms of discourse pragmatics and thus correspond to the authors' stance taking.

The paper is structured as follows. In Section 2, we lay out the relevant linguistic theory on modality and present the pragmatic notion of hedging. In Section 3, we discuss previous treatments of modality in Slovenian linguistics as well as related work on corpus-based treatment of hedging in academic discourse. In Section 4, we present the corpus we used for our analysis from the perspective of the extra-linguistic metadata relevant for our purposes as well as discuss the selection criteria of the modal adverbs that we have analysed. In Section 5, we present and discuss the results. In Section 6, we conclude the paper.

## 2 THEORETICAL FRAMEWORK

### 2.1 Epistemic and Non-Epistemic Modalities

Modality has been defined in many different ways in the literature, but it is perhaps von Fintel (2016, p. 21) who most succinctly summarizes the notion:

> Modality is a category of linguistic meaning having to do with the expression of possibility and necessity. A modalized sentence locates an underlying or prejacent proposition in the space of possibilities [...] *Sandy might be home* says that there is a possibility that Sandy is home. *Sandy must be home* says that in all possibilities, Sandy is home.

Modality thus evaluates a proposition from the perspective of the gradient from possibility to necessity. Notions such as *possibility*, *likelihood*, and *necessity*, which are logically related by entailment, are also referred to as the modal force (Kratzer, 2012). Aside from this, modality is polysemous and the usual linguistic distinction is made between epistemic modality on the one hand and non-epistemic modality on the other (Palmer, 2014), the latter of which is usually referred to as root modality (Coates, 1983) or circumstantial modality (Kratzer, 2012). In this paper, we use the term root modality.

Epistemic modality encompasses the speaker's judgement about the truth of the proposition (Palmer, 2014, p. 50). A modal like *mogoče* in sentence (1) is

epistemic, expressing that the speaker is not completely certain that the prejacent i.e. unmodalised proposition *Ana je doma* "Ana is home" is true.[2]

(1) Ana je *mogoče* doma.
"Ana is possibly home."

By contrast, root modality also evaluates the proposition in the domain of possibility (and necessity), but, unlike epistemic modality, does not tie the evaluation to the speaker's knowledge. An example of a non-epistemic modal is *lahko* in sentence (2).

(2) Ta program se *lahko* namesti na Windows.
"This program can be installed on Windows."

Here, *lahko* is not used to indicate the speaker's knowledge about the truth of the expressed proposition but rather to attribute possible qualities to the subject NP *ta program* "this program".

A single modal often allows for more than one reading that is contextually determined. For instance, *lahko* in sentence (3) has an epistemic reading that can be paraphrased as "It is possible that Ana is at home or at school" and a root meaning that denotes permission that Ana is granted by someone else ("Ana is allowed to stay at home or in school"), which is typically disambiguated by the context it appears in.[3] This motivates the manual concordance analysis of the Slovenian modal adverbs that will be presented in Section 5.2.

(3) Ana je *lahko* doma, *lahko* pa je v šoli.
"Ana may be at home or school."
"Ana can be at home or school."

Finally, many root modal expressions display prominent meta-discursive usage, as in the case of reader-oriented meta-commentary clauses like the one in example (4). Such use along with the purely epistemic meaning often corresponds to the pragmatic notion of hedging (Hyland, 1996, 1998; Grabe and Kaplan, 1997), which we introduce in Section 2.2.

2 For ease of exposition, we use simple constructed linguistic examples to showcase the relevant semantic characteristics of modality in this section.

3 The modal meaning involving obligation/permission is referred to as *deontic modality* by Palmer (2014).

(4) Kot *lahko* vidimo iz rezultatov …
"As can be seen from the results…"

### 2.2 Hedging – a Pragmatic Strategy

In linguistics, Lakoff (1972, p. 471) was the first to use the term *hedges* to refer to "words whose meaning implicitly involves fuzziness – words whose job is to make things fuzzier or less fuzzy". Lakoff (1972)'s basic concept is further explicated by Hyland (1996, p. 251), who claims that hedges are "any linguistic means used to indicate either (a) a lack of complete commitment to the truth of a proposition, or (b) a desire not to express that commitment categorically". Additionally, hedging not only involves markers of tentativeness but is typically extended to include rhetoric communicative strategies, e.g., politeness, by means of which the author implicitly includes the addressee in the discourse her or she is presenting (Grabe and Kaplan, 1997, p. 154).

Hyland (1996)'s definition of hedging overlaps quite significantly with that of epistemic modality defined in the previous section, but there is an important difference: a hedge is not a lexical property that holds of a specific category like modality, but rather a pragmatic device that can in principle hold for any lexical category given the suitable communicative context.

In terms of grammatical categories, hedging corresponds not only to modal verbs or adverbs, but also to other lexical categories such as the use of certain reporting verbs that indicate the author's tentativeness (e.g., *we believe that*) as well as syntactic strategies such as the use of the passive rather than the active voice to syntactically omit the otherwise entailed agent of the verbal event (Rizomilioti, 2006, p. 56) or the use of inclusive plural pronouns to help establish rapport between the reader and the writer (Hyland, 1996).

## 3 RELATED WORK

### 3.1 The Slovenian Modal System

Slovenian linguists generally discuss Slovenian modals either in relation to highly specialised topics in theoretical linguistics or in the context of applied and descriptive comparative linguistics. Theoretical linguists usually focus on discussing the formal properties of individual selected modal lexemes;

for instance, Marušič and Žaucer (2016) propose a syntactic explanation why the modal adverb *lahko* is a positive-polarity item (i.e., it cannot syntactically co-occur with negation), while Hladnik (2015, p. 86) discusses the fact that the lexeme *da*, which is syntactically a subordinator, triggers an epistemic meaning in relative clauses (e.g., *človek, ki* da *pride* "the person who *supposedly* is coming"). In applied/comparative linguistics, researchers usually use the modals as a springboard for studying broader pragmatic topics; for instance, Pisanski Peterlin (2015) discusses how Slovenian epistemic modals are used in English–Slovenian translation in comparison to original Slovenian texts in order to determine how epistemic modality is influenced by language transfer, while Pihler Ciglič (2017) compares the use of assumptive modals like *morda* with related lexemes in American Spanish in the context of literary translations.

However (and to our knowledge), no one has yet attempted a comprehensive typological study of the general syntactic and semantic properties of the Slovenian modal system in the context of descriptive Slovenian linguistics on par with Palmer (2014)'s work on English modal auxiliaries. What is especially noteworthy in relation to modal adverbs is that the Slovenian reference grammar *Slovenska slovnica* (Toporišič, 2004) only lists them as examples of the particle word class, but does not devote any attention to their syntactic characteristics nor to a more fine-grained semantic classification that would disentangle notions such as the modal force from the modal base for a given modal. As we will see in Section 4.2, such an uncomprehensive classification of modal adverbs in the reference grammar seems to have, at least from the perspective of syntactic consistency, also negatively affected the morphosyntactic tagging in Slovenian corpora, which is based on the reference grammar, as modal lexemes that are syntactically adverbs seem to be arbitrarily assigned to either the adverb or the particle classes.

In our paper, we take into account the fact that modals display a complex semantics. Although our primary aim is to investigate academic discourse, we nevertheless believe that certain aspects of our study, such as the rate at which a modal conveys a particular modal reading (Section 5.2), also positively contribute to the general understanding of the lexical-semantic characteristics the Slovenian modal system. However, a more comprehensive description of the

modal system, which should also compare the use of Slovenian modality in registers other than academic discourse, goes far beyond the scope of this paper.

### 3.2 Modal Adverbs and Hedging in Academic Discourse – Cross-Disciplinary Comparisons

In related work on hedging in academic discourse, researchers (Hyland, 1998; Rizomilioti, 2006; Pisanski Peterlin, 2010; Takimoto, 2015, a.o.) have generally taken into account all of the major categories that can in principle be used to hedge discourse, such as modal auxiliaries, modal and non-modal (e.g., approximators) adverbs and adjectives, and lexical verbs.

For instance, Takimoto (2015) analyses how hedges corresponding to 5 syntactic categories (adverbs, adjectives, auxiliaries, nouns, and verbs) are used across 4 different natural sciences disciplines and 4 humanities/social sciences disciplines, showing that "70% of all hedges and boosters were found in humanities and social sciences" (2015, p. 103) and that philosophy contains "almost 5.3 times as many hedges and boosters as electrical engineering" (*ibid.*).[4] Similarly, Rizomilioti (2006, p. 64) compares the use of hedging between a 200,000 token corpus of journal papers in literary criticism and a comparable corpus of papers in biology, showing that there are more adverbs of uncertainty in the literary criticism corpus than in the biology corpus.

Given the high degree of lexical polysemy and the consequent likelihood that not all of the observed lexemes in the studied corpus function as hedges, a prominent strategy to filter out irrelevant data relies on the close reading of all the concordances that potentially correspond to hedges in order to single out only the relevant occurrences. For this to be possible, the corpora used in the related literature are often quite small, generally consisting of 100,000–500,000 tokens and around 50–60 research articles (Thompson, 2000; Pisanski Peterlin, 2010; Hyland, 1998; Rizomilioti, 2006; Takimoto, 2015).

Nevertheless, despite such a strategy of close reading, the epistemic and non-epistemic notions of possibility seem conflated in some of the related

4 Some authors use the term *boosters* to describe those hedges that convey the author's certainty rather than tentativeness; since our analysis, presented in Section 5.1, does not show prominent differences between hedges and boosters, we use *hedges* as a general term for expressing both tentativeness and certainty.

work. For instance, Piqué-Angordans et al. (2002), who survey how English modal auxiliary verbs (e.g., *can*, *may*, *should*) vary between their epistemic and root/deontic senses across 3 corpora of research articles in medicine, biology, and literary criticism, provide the following 2 examples as expressing epistemic modality in their corpus of research articles in medicine (2002, p. 53):

(5) Tricyclic antidepressants, however, *can* also have significant adverse effects, such as arrhythmias, postural hypotension, sedation, dry mouth, constipation, confusion, and urinary retention.

(6) The quantities of the factors *could* limit the amount of renin mRNA that can be produced, even under conditions of normal salt loading and in the absence of pharmacological interventions.

While the use of *could* in sentence (6) undoubtedly expresses an epistemic judgement, i.e., that the authors are not certain whether the "quantities of the factors" do in fact "limit the amount of renin mRNA", the use of *can* in sentence (5) plays a different i.e. non-epistemic modal role, in contrast to Piqué-Angordans et al. (2002)'s claim.[5] That is, *can* in (5) simply expresses that "tricyclic antidepressants" have properties that can cause adverse effects under certain undefined conditions. As we will see in Section 5.2, the distinction between the two meanings is crucial from the perspective of hedging; we will claim that only expressions of possibility like that in (6) but not in (5) constitute this pragmatic strategy.

We therefore attempt to make our quantitative analysis of the modals more precise by making such a distinction between the modality types introduced in Section 2.1, arguing that only those instances of possibility expressed by the modals that correspond either to epistemic modality or to the meta-discursive usage function as hedges, whereas non-epistemic meanings of possibility that correspond to dispositional ascriptions do not.

5 This sentence is taken from the introduction of the paper by Rowbotham et al. (1998), where the co-text affirms that the use of *can* here is not meant to convey the authors' epistemic judgement. It is also worth noting that Portner (2009, p. 30) claims that *can* is never used epistemically (e.g., *It can be raining* does not seem to admit an epistemic reading unless it is negated).

Our corpus, which we introduce in Section 4.1, is also significantly larger than those in the related literature, consisting of approximately 1.7 billion tokens. Because close reading of such a large corpus was not a feasible approach for us and because we wanted to reduce the amount of irrelevant data that in part arises from the often unpredictable lexical polysemy,[6] we limit our analysis to a single word class, i.e., modal adverbs, which can be queried systematically via its morphosyntactic tag and at the same time arguably constitute the most prominent category for expressing sentential modality in Slovenian.

### 3.3 Modal Adverbs and Hedging in Academic Discourse – Between Academic Stages

In another major strand of related work (e.g., Aull and Lancaster, 2014; Aull et al., 2017; Crosthwaite et al., 2017), it is shown that there are prominent differences in the use of markers of stance between early and advanced academic writing. For instance, Aull and Lancaster (2014) survey the distribution of English approximative hedges (e.g., *generally*, *evidently, somewhat*) in the context of research papers written by students at US universities, comparing them between 3 corpora: first, a corpus of argumentative essays by first-year undergraduate students (abbr. *FY*); second, a corpus of upper-level essays by third-year students and graduate students (abbr. *UP*); and third, published scholarly writing from peer-reviewed journals in the academic subcorpus of

6 It is also often quite unclear whether research that observes hedging across multiple word classes (and broader syntactic patterns) takes into account the idiosyncratic grammatical features of a category that distinguish it from others and could serve as potential caveats for studying pragmatic effects. An example of this is modal adjectives. Modality in NP-modifying adjectives exhibits sub-sentential semantic scope (Portner, 2019), which means that it does not take scope over the asserted proposition in contrast to prototypical modals but rather over an implicit proposition that is presupposed in the semantics of the noun phrase (DeLazero, 2011).
Crucially, what is then hedged in such cases is a non-overt claim; for instance, *možno* in a sentence like *To so možne analize* "These are the possible analyses" takes scope over a non-overt presupposed proposition in the noun phrase *možne analize*, with the resulting modalised meaning being either something like *these analyses might be correct* (epistemic) or *these analyses can be correct under certain circumstances* (root), which however is not something that is asserted by the original sentence. Since the modalised proposition is thus non-overt, it is often quite unclear if and how the claim is being hedged in such cases. None of the reviewed related work on hedging that looks at modal adjectives takes this into account.

the *Corpus of Contemporary American English* (abbr. *COCAA*). It is shown that the frequency of such approximative hedges increases between all three corpora: from 109.5 per 100,000 words in the *FY sub*corpus to 173.5 in the *UP* subcorpus, that is a 58% increase from *FY*, and finally to 203.8 per 100,000 words in *COCAA,* that is an 86% increase from *FY* (Aull and Lancaster, 2014, p. 162).

Interpreting this increase observed in American English academic writing, Aull and Lancaser (*ibid.*) claim that students are "often encouraged to take a 'critical stance' with regard to others' arguments" and that a "highly attitudinal, forceful, and assertive stance is less valued in advanced student writing than stances that are implicitly attitudinal [...] or open to other views in the surrounding discourse" (*ibid.*, p. 155). Similarly, Aull et al. (2017, p. 32) claim that published academic writing more prominently displays "qualified and circumscribed arguments" than the writing of incoming college students. In sum, advanced writers use hedge to obviate a forceful, asserted stance by more frequently using hedging devices.

However, such an increase in hedging from less mature to more advanced writing is not necessarily a universal trend. Crosthwaite et al. (2017), who compare the use of stance expressions between learner and professional research reports in dentistry, observe that hedging in their dentistry professional corpus is *less* frequent than in the learner corpus. This is precisely the opposite of the results reported by Aull and Lancaster (2014). In the second part of the paper, we therefore attempt to determine this trend for Slovenian academic writing by comparing the frequency of hedging adverbs between Slovenian bachelor's, master's, and doctoral theses, which are the final works signalling the completion of each of the three major stages of tertiary education in Slovenia.

## 4 METHODOLOGY

### 4.1 The KAS Corpus of Academic Slovenian

The study presented in this paper has been carried out on the 1.7-billion-token *KAS corpus of Slovenian academic writing* (Erjavec et al., 2019a). The theses in the corpus were written between 2000 and 2018 at Slovenian universities

and other academic institutions.[7] The corpus is linguistically annotated and is also marked up for several extra-linguistic metadata categories that are tailored to the genre of academic theses, the most relevant for our purposes being the publisher and CERIF (Common European Research Information Format). The corpus is accessible online through the CLARIN.SI *noSketch Engine* concordancer,[8] which is an open-source version of *Sketch Engine* corpus query system.

The Publisher information corresponds to the institution or faculty where the thesis was defended. There are a total of 70 different publisher abbreviations, 55 of which are faculties of the Universities of Ljubljana, Maribor, Nova Gorica, and Primorska. The remaining 15 are research institutes with their own study programmes or private and semi-private colleges. The corpus represents a very diverse breadth of scientific (sub)disciplines, so each thesis has been assigned to (at least) one of the five top-level CERIF[9] categories: BIO(MEDICAL SCIENCES), HUM(ANITIES), PHYS(ICAL SCIENCES), SOC(IAL SCIENCES), and TECH(NOLOGICAL SCIENCES). Since the CERIF categories represent a generalised division of academic disciplines, they are particularly well-suited for comparative corpus analyses of academic genres, especially given the diverse disciplinary scope of the individual publishers included in the corpus.

The CERIF division of the theses in the *KAS* corpus is given in Table 1.

**Table 1:** *The five disciplinary subcorpora of KAS*

| CERIF | Size (in tokens and %) | |
|---|---|---|
| BIO | 100,514,116 | 7% |
| HUM | 150,634,867 | 10% |
| PHYS | 147,690,128 | 10% |
| SOC | 1,018,235,132 | 66% |
| TECH | 121,360,503 | 8% |
| ∑ | 1,538,434,746 | 100% |

7 The morphosyntactic annotation and lemmatisation of the corpus was performed with the ReLDI morphosyntactic tagger and lemmatizer (https://github.com/clarinsi/reldi-tagger), which gives an accuracy of 98.94% on the parts of speech and 94.27% on the complete morphosyntactic descriptions. For a comprehensive description of the corpus, see Erjavec et al. (2020).

8 https://www.clarin.si/noske/.

9 https://eurocris.org/services/main-features-cerif. Accessed on 16 June 2021.

As shown in Table 1, the five CERIF subsets of *KAS* are unequal in size, with the SOC(IAL SCIENCES) subset accounting for over half of the corpus. Consequently, we will provide frequency counts for our modal adverbs that are relativised to a million tokens. Furthermore, the total token size (1,538,434,746) listed in Table 1 is slightly smaller than that of the entire *KAS* corpus (1,699,097,710); this is because approximately 9% of the theses are assigned to multiple CERIF categories, while the texts that we take into account include all the theses with only one CERIF label.

In the first part of our analysis, we focus on the subcorpus of doctoral theses, *KAS-dr* (Erjavec et al., 2019c), which consists of 1569 doctoral theses, amounting to a total of 100 million tokens or roughly 7% of the entire *KAS* corpus. In the second half of our analysis, we compare the results obtained for the *KAS-dr* subcorpus with the subcorpora of master's (*KAS-mag*; Erjavec et al., 2019b) and bachelor's theses (*KAS-dipl*; Erjavec et al., 2019d), which contain 496,000,000 tokens (31% of the entire *KAS* corpus) and 1.1 billion tokens (72% of the entire *KAS* corpus), respectively. Because of this inequality in size, and because the theses are unequally distributed among the CERIF categories in all three subcorpora in roughly the same ratio as in Table 1 (i.e., SOC theses account for more than half of each subcorpus), we will again use normalized frequencies to compare the findings in the three subcorpora.

### 4.2 Modal Adverbs

The modal adverbs analysed in this paper are listed in Table 2. There are 6 adverbs that denote possibility (*lahko, mogoče, možno, morda, menda*, *morebiti*), 3 adverbs that denote likelihood (*najbrž, domnevno, verjetno*), and 3 adverbs that denote certainty (*nedvomno*, *zagotovo, gotovo*).

The modals were selected in the following way. We first extracted all the lemmas in the *KAS-dr* subcorpus that are morphosyntactically tagged as either adverbs or as particles. It is important to note that the Slovenian descriptive grammar *Slovenska slovnica* (Toporišič, 2004), which is the basis for the MULTEXT tagset[10] used by the *KAS* corpus (Erjavec, 2012), postulates that the particle is a separate word class. Toporišič (2004, pp.

10 https://www.sketchengine.eu/slovene-tagset-multext-east-v5.

**Table 2:** *The most frequent epistemic modal adverbs in the KAS-dr subcorpus*

| MODAL | Meaning | AF | RF |
|---|---|---|---|
| *lahko* | possibly | 296,311 | 2,920 |
| *verjetno* | likely | 12,958 | 128 |
| *morda* | possibly | 9,727 | 96 |
| *zagotovo* | certainly | 3,291 | 32 |
| *gotovo* | certainly | 3,152 | 31 |
| *nedvomno* | certainly | 2,534 | 25 |
| *mogoče* | possibly | 1,878 | 19 |
| *možno* | possibly | 1,346 | 13 |
| *najbrž* | likely | 1,082 | 11 |
| *domnevno* | likely | 969 | 10 |
| *morebiti* | possibly | 811 | 8 |
| *menda* | possibly | 315 | 3 |

*Note*. AF lists the absolute frequencies while RF lists the relative frequencies per 1 million tokens.

445–449) exceptionally defines the particle class solely in terms of its semantic rather than syntactic properties, claiming that the category is distinct from adverbs in that it consists of semantically abstract clausal modifiers (i.e., propositional operators) rather than event modifiers such as adverbials of manner or time. While most of the lexemes in Table 2 are tagged as adverbs in *KAS*, *morda*, *najbrž*, *morebiti*, and *menda* are tagged as particles, even though their syntactic distribution is prototypically adverbial. In other words, there are no categorical differences between *verjetno*, which is tagged as an adverb, and *najbrž*, which is tagged as a particle. For simplicity's sake, we thus refer to all the 12 lexemes in Table 2 as adverbs. From this extracted list of adverb and "particle" lexemes in the corpus, we selected all that semantically correspond to epistemic modals and are not stylistically marked; because of this latter criterion, we omitted the infrequent colloquial hearsay modals *bržda* "likely"*, baje* "possibly", *nemara* "likely", and *bojda* "possibly".

The 12 lexemes in Table 2 largely correspond to the epistemic modal adverbs identified for Slovenian by Pisanski Peterlin (2015, p. 31). However, in contrast to her approach, our selection criteria were stricter in that we excluded

those adverbs that are frequently ambiguous between a modal and non-modal (e.g., manner) interpretation.[11]

Such an ambiguous modal is *očitno* "apparently", as shown by the two possible paraphrases of example (7), taken from *KAS-dr*, where the first corresponds to a modal interpretation denoting the speaker's attitude towards the proposition while the other to a non-modal interpretation in which the adverb specifies the manner of the verbal event.

(7) Z naraščajočim deležem titana se je *očitno* zmanjšala količina ter velikost evtektičnih karbidov M7C3.
"It appears that with the increasing amount of titanium, the quantity and size of eutectic carbides M7C3 has decreased."
"With the increasing amount of titanium, the quantity and size of eutectic carbides M7C3 has decreased in an obvious manner/to a great degree."

Discounting such ambiguous adverbs reduces the amount of irrelevant data; that is, it ensures that our comparative analysis is not hindered by the noise due to polysemy.

## 5 THE RESULTS

### 5.1 Quantitative Analysis of Modal Adverbs Across Disciplines in Doctoral Theses

Table 3 compares the distribution of the 12 modal adverbs in focus between the humanities (i.e., HUM) and social sciences (SOC) disciplines in *KAS-dr* on the one hand and the biotechnical (BIO), physical sciences (PHYS), and technological (TECH) disciplines on the other. The size of HUM and SOC is 68,207,965 tokens in total, while the size of BIO, PHYS, and TECH is 39,679,476 tokens in total. The AF columns reports the absolute frequency and RF the relative frequency, which is normalised to 1 million tokens.

11 The adverb *lahko* also has a manner interpretation, i.e., "easily". However, this use is very rare – in our analysis of a randomized set of 250 concordance examples (see Section 5.2) for this adverb, there was only 1 example, given in (i), where *lahko* is used in its comparative form *lažje* and corresponds to the non-modal manner usage:
(i) […] zaradi česar *lažje* in pogosteje prihaja do sprememb v vrednostih indikatorjev.
"[…] because of which changes in the values of the indicators occur more frequently and more easily."

Based on a comparison of the relative frequencies, the modals in Table 3 are divided into two groups. The first group consists of the modals *lahko* ("possibly"), *verjetno* ("likely"), and *možno* ("possibly"). Each modal in this group is more frequent in the biotechnical, physical sciences, and technological sciences than in the humanities and social sciences, as indicated by the BPT:HS ratio reported in the fourth column. On the whole, this group is 1.1 times more frequent in BIO, PHYS, and TECH than it is in HUM and SOC.

The second group consists of 9 modals, that is *morda* ("possibly"), *zagotovo* ("certainly"), *gotovo* ("certainly"), *nedvomno* ("certainly"), *mogoče* ("possibly"), *najbrž* ("likely"), *domnevno* ("likely"), *morebiti* ("possibly"), and *menda* ("possibly"). Each modal in this group is more frequent in the humanities and social sciences than in the biotechnical, physical, and technological sciences; on the whole, this group is 2.2 times more frequent in the humanities and social sciences.

**Table 3:** *Modal adverbs in KAS-dr across academic disciplines*

| | HUM, SOC | | BIO, PHYS, TECH | | | | | |
|---|---|---|---|---|---|---|---|---|
| MODAL | AF | RF | AF | RF | BPT:HS | LLV | *p* | DIN |
| *lahko* | 194,386 | 2,850 | 119,639 | 3,015 | 1.1 | 234.167 | 0.0000 | –2.817 |
| *verjetno* | 8,635 | 127 | 5,089 | 128 | 1.0 | 0.539 | 0.4627 | –0.649 |
| *možno* | 760 | 11 | 713 | 18 | 1.6 | 82.812 | 0.0000 | –23.45 |
| ∑ | 203,781 | 2,988 | 125,441 | 3,161 | 1.1 | 247.631 | 0.0000 | –2.825 |

| | HUM, SOC | | BIO, PHYS, TECH | | | | | |
|---|---|---|---|---|---|---|---|---|
| MODAL | AF | RF | AF | RF | HS:BPT | LLV | *p* | DIN |
| *morda* | 8,028 | 118 | 2,123 | 54 | 2.2 | 1198.072 | 0.0000 | 37.497 |
| *zagotovo* | 2,655 | 39 | 844 | 21 | 1.9 | 257.012 | 0.0000 | 29.329 |
| *gotovo* | 2,695 | 39 | 568 | 14 | 2.8 | 590.887 | 0.0000 | 46.811 |
| *nedvomno* | 2,223 | 33 | 448 | 11 | 3.0 | 518.854 | 0.0000 | 48.542 |
| *mogoče* | 1,449 | 21 | 593 | 15 | 1.4 | 54.460 | 0.0000 | 17.406 |
| *najbrž* | 891 | 13 | 227 | 6 | 2.2 | 142.948 | 0.0000 | 39.088 |
| *domnevno* | 665 | 10 | 173 | 4 | 2.5 | 102.498 | 0.0000 | 38.199 |
| *morebiti* | 821 | 12 | 187 | 5 | 2.4 | 160.011 | 0.0000 | 43.726 |
| *menda* | 306 | 4 | 12 | 0 | 6.0 | 202.431 | 0.0000 | 87.369 |
| ∑ | 19,733 | 289 | 5,175 | 130 | 2.2 | 2994.528 | 0.0000 | 37.855 |

To check for statistical significance, we have tested the individual distributions using *Calc: Corpus Calculator* (Cvrček, 2021), an online statistical tool that offers a module for evaluating whether the difference between a pair of absolute frequencies is statistically significant. We report the log-likelihood values (LLV) for each pair of frequencies and the associated $p$ values calculated by the module, where the cut-off point for significance is $p < 0.05$. The calculation of the log-likelihood score is based on Andrew Hardie's implementation of Ted Dunning's (1993) original formula (Václav Cvrček, p.c.) and is as follows:

$$2 \times (O_1 \times \ln(\frac{O_1}{E_1}) + O_2 \times \ln(\frac{O_2}{E_2}))$$

where $O_1$ and $O_2$ are the observed absolute frequencies and $E_1$ and $E_2$ the expected frequencies. In Table 3, all the differences in the absolute pairwise frequencies are significant except for *verjetno*; LLV = 0.539, $p = 0.4627 > 0.05$.

However, as noted by Fidler and Cvrček (2015, p. 226), a problem of large corpora is that the $p$-value of a test does not take into account the practical importance (effect size) of the difference – i.e., "the larger the amount of data, the higher the likelihood that the resulting difference is significant" (2015, p. 227). To take the effect size into account, Table 3 also reports the Difference Index (DIN; also calculated by *Calc*) in the last column. DIN is calculated with the following formula (2015, 230):

$$100 \times \frac{RelFq(AF_{\text{HUM,SOC}}) - \text{RelFq}(AF_{\text{BIO,PHYS,TECH}})}{RelFq(AF_{\text{HUM,SOC}}) + \text{RelFq}(AF_{\text{BIO,PHYS,TECH}})}$$

The values of DIN range from –100 to 100, where –100 would mean that the word is present only in BIO, PHYS, and TECH; 0 would mean that the word occurs equally often in HUM and SOC on the one hand and BIO, PHYS, and TECH on the other, and 100 would mean that the word occurs only HUM and SOC.

In Table 3, the DIN values for all the 3 modals in the first group are negative, which reflects the fact that they occur more frequently in PHYS, SOC, and TECH. The –2.825 score for the overall difference for this group reflects the small BPT:HS ratio. Conversely, the DIN scores for the second group are much higher, where the overall difference between HUM and SOC on the one hand

and BIO, PHYS, and TECH on the other has a DIN score of 37.855, reflecting the much higher HS:BPT ratio in this group.

### 5.2 Comparison of Epistemic and Non-Epistemic Usage Across Disciplines

In order to gain more insight into the pattern observed in the previous section, according to which 9 out of the 12 analysed modal adverbs occur most frequently in the humanities and social sciences in *KAS-dr* while the remaining adverbs are more prominent in the biotechnical, physical, and technological sciences, we have manually classified a randomized set of 250 concordance examples for each of the 12 adverbs into one of the three categories:

a) epistemic modality;

b) meta-discursive root modality; or

c) dispositional root modality.

The results of the concordance analysis are presented in Table 4.[12] It shows that the distribution of epistemic and non-epistemic meanings of the adverbs generally follows the distribution of the modals between the academic disciplines (Table 3). Eight modals, namely *morda, najbrž, zagotovo, nedvomno, domnevno, gotovo, morebiti*, and *menda*, are used almost exclusively to denote epistemic modality. The modal *mogoče* is also used mostly as an epistemic modal (60% of the concordance). Crucially, all these modal adverbs are precisely those which are more frequently used in the humanities and social sciences (cf. the second group in Table 3). By contrast, the modals *možno* and *lahko*, which are more prominent in natural and technical sciences, infrequently convey the epistemic meaning (11% of the concordances in the case of *lahko* and 2% of the concordances in the case of *možno*). An exception is the modal *verjetno*, which despite its purely epistemic meaning is

12 Note that, in Table 4, the number of included concordances for each modal is not always exactly 250, like 248 in the case of *možno*. The lower number in these cases is due to a few instances of incorrect part-of-speech tagging in the corpus (e.g., some syncretic premodifying adjectives, like *možno* in the accusative/instrumental NP *možno analizo* "possible analysis", are incorrectly tagged as adverbs); we have discarded such irrelevant occurrences from our analysis. Furthermore, *menda* had the largest number of irrelevant examples (i.e., 49), all of which were sentences in which the modal was used in a quoted context, so it did not reflect the author's perspective.

**Table 4:** *The epistemic/root distribution of the modal adverbs in KAS-dr*

| MODAL | EPISTEMIC | | META-DISCURSIVE | | DISPOSITION | |
|---|---|---|---|---|---|---|
| | **Freq.** | **%** | **Freq.** | **%** | **Freq.** | **%** |
| *lahko* | 25 | 11% | 105 | 42% | 117 | 47% |
| *verjetno* | 250 | 100% | 0 | 0% | 0 | 0% |
| *možno* | 6 | 2% | 9 | 4% | 233 | 94% |
| *morda* | 240 | 96% | 7 | 4% | 0 | 0% |
| *najbrž* | 250 | 100% | 0 | 0% | 0 | 0% |
| *zagotovo* | 243 | 100% | 0 | 0% | 0 | 0% |
| *nedvomno* | 250 | 100% | 0 | 0% | 0 | 0% |
| *mogoče* | 150 | 60% | 3 | 1% | 97 | 39% |
| *domnevno* | 250 | 100% | 0 | 0% | 0 | 0% |
| *gotovo* | 245 | 98% | 5 | 2% | 0 | 0% |
| *morebiti* | 250 | 100% | 0 | 0% | 0 | 0% |
| *menda* | 201 | 99% | 2 | 0% | 0 | 0% |

more prominent in the natural and technical sciences. In the remainder of this section, we take a closer look at the results of the annotation process for each of the three categories and relate the use of modality to the notion of hedging that was introduced in Section 2.2.

#### 5.2.1 Epistemic Modality

Let us first take *morda*, which is used as an epistemic modal in 240 (96%) of the randomized concordances and only in 7 (4%) as a non-epistemic modal in the meta-discursive sense, as being representative of the group that is almost exclusively epistemic. Sentence (8), which is taken from a thesis defended at the Faculty of Social Sciences at the University of Ljubljana, exemplifies this epistemic usage.

(8) *Morda* je to eden od razlogov, da znanstvena skupnost ni bila uspešna pri svojem "programu" izboljšanja javnega razumevanja znanosti in znanstvene pismenosti.
"Perhaps this is one of the reasons that the scientific community wasn't successful in implementing their proposed program for improving the public understanding of science and scientific literacy."

Pragmatically, this corresponds to Hyland (1996, pp. 256–257)'s notion of an *accuracy-based hedge*, as it is used by the writer to denote their uncertainty about the validity of the proposition in the example; i.e., that whatever is denoted by the demonstrative *to* "this" in the main clause is indeed one of the reasons for the lack of success on part of the scientific community.

Similarly, *menda* and *domnevno* are also used mainly as epistemic modals in the sense that they convey the author's uncertain about what they are claiming. However, in contrast to *morda*, the adverbs *menda* and *domnevno* are additionally used to signal that the claim is an assumption, possibly one that is shared within the author's research community.[13] Sentence (9), which is taken from a thesis defended at the Faculty of Arts at the University of Maribor, exemplifies this usage:

(9) Klun je nato v svojem govoru zavrnil očitke, da je bil pobudnik interpelacij, kot je to *menda* trdil Schwegel.
"In his speech, Klun then denied the accusations that he was the instigator of the interpellations, as was supposedly claimed by Schwegel."

In this example, the writer uses *menda* to signal that it is not universally certain whether Schwegel indeed claimed that Klun had been the instigator of whatever the interpellations were, but that it is merely assumed that he made the claim; because *menda* thereby conveys the author's uncertainty (although with an additional assumptive meaning lacking with *morda*), its role in terms of hedging is also accuracy-based in Hyland (1996)'s terms.

All the epistemic examples with the remaining modals (which we do not exemplify here due to space constraints) also function as similar accuracy-based hedges, where the sole semantic and pragmatic difference is in the modal force of the lexeme in question; that is, a modal like *najbrž* "likely" denotes a greater degree of the speaker's commitment to the truth of the proposition than *morda* or *morebiti* "possibly".

13 As Pihler Ciglič (2017) notes, there is an on-going debate in the literature whether evidential/hearsay modals like *menda* and *domnevno* constitute a category that is distinct from other epistemic modals. We follow Palmer (2001) and von Fintel and Gillies (2007) in assuming that the evidential adverbs we analyse are an epistemic subtype since they invariably signal the speaker's uncertainty. In any case, this is a complex issue that hinges on quite a few technical and formal assumptions about modality; see Portner (2009, section 4.2.2) for a good overview of this issue.

#### 5.2.2 Meta-Discursive Root Modality

Sentence (10), taken from a thesis defended at the Faculty of Pedagogy at the University of Ljubljana, exemplifies one of the few cases of the non-epistemic meta-discursive use of *morda*.

(10) Zato lahko *morda* na tem mestu poudarim strinjanje z Banduro (1997), da je samoučinkovitost precej povezana s samouravnavanjem […]
"This is why I can (perhaps) emphasise my agreement with Bandura (1997) that self-effectiveness is related to self-regulation."

In contrast to its epistemic use in (8), *morda* in this sentence clearly does not denote the writer's uncertainty and could be freely omitted from the sentence without a change in the propositional truth-commitment. It is rather used as part of a meta-discursive strategy with which the writer "acknowledge[s] the reader's role in ratifying knowledge" (Hyland, 1996, p. 258), in the sense that the lexical meaning of possibility, which is inherently entailed by the modal, "subtly hedges the universality of a writer's claim by implying that a position is an individual interpretation" (*ibid.*).

Such meta-discursive use is most prominent with the modal *lahko*, having been observed in 105 (42%) out of a total 250 of the randomized set of concordances. The sentence in (11), which is taken from a thesis from the Biotechnical Faculty at the University of Ljubljana, exemplifies this usage.

(11) Zaključimo *lahko*, da alkidni premazi na osnovi organskih topil izkazujejo nižje kontaktne kote na obeh substratih kot vodni akrilni premazi […]
"We can conclude that alkyd coatings on the basis of organic solvents show smaller contact angles on both substrates than aqueous acrylic coatings…"

In all the 105 examples with the meta-discursive use of *lahko*, the modal adverb is used with directive verbs that are inflected for the so-called inclusive plural, like *zaključimo* "we conclude" in example (11). According to Takimoto (2015, p. 99), the use of "inclusive pronouns (e.g., *we*) […] enables the writers to produce more interpersonal signals to the readers, which may allow the writers to share contexts with the readers and draw on their assumed belief

specific to a particular field of study". In other words, the inclusive inflection emphasises the meta-discursive use of *lahko* as a hedge that is reader-oriented rather than accuracy-oriented (Hyland, 1996). Note that the remaining modals which are also used in this meta-discursive role (*mogoče, možno, morda, zagotovo, morebiti*, *menda*) do not pattern with the inclusive plural inflection (cf. example (10), where the first person is used) as consistently, which may possibly correlate with the fact that their use in this role is much less frequent in comparison to *lahko*, this being the de-facto modal for expressing meta-discursive commentary.

#### 5.2.3 Dispositional Root Modality

Finally, we turn to the dispositional root modality of *lahko*, *mogoče*, and *možno*. Sentence (12), which is taken from a thesis defended at the Faculty of Medicine at the University of Ljubljana, exemplifies this meaning with the modal *možno*, which is by far the most frequently used in this sense (233 or 94% examples), while sentence (13), which is from a thesis in the former Faculty of Electrical Engineering, Computer Science and Information Sciences at the University of Ljubljana, contains the modal *mogoče*, which is used in the dispositional sense in 97 (39%) of the concordance examples.[14]

(12) Upliniti je *možno* najrazličnejšo biomaso (les, oglje, kokosove olupke, riževe lupine).
"It is possible to gasify many kinds of biomass (wood, charcoal, coconut peels, rice husks)."

(13) Celoten grafični vmesnik je zasnovan tako, da ga je *mogoče* hitro prilagoditi potrebam metode […]
"The entire GUI is designed in such a way that it can be easily tailored to the needs of the method."

14 In standard descriptive Slovenian linguistics, the lexemes *možno* and *mogoče* are usually referred to as adverbs in sentences like (12) and (13); see, e.g., the *Dictionary of Standard Slovenian* entry for *možno* (Bajec et al., 2014). Note, however, that in both examples *možno* and *mogoče* require that the VP be infinitival. It would therefore be more precise to analyse the two lexemes as predicative adjectives, on par with those heading extrapositional *it*-constructions in English like *It is possible to*+$VP_{inf}$ (Van linden and Davidse, 2009). Conversely, adverbs in clausal adjunct positions are unable to govern the syntactic properties of other sentential constituents in such a way.

In such cases, the modals are used to denote possibility in its root non-epistemic sense. This kind of modality is not concerned with the knowledge or attitude of the writer (as in the case of epistemic modals and those used in the meta-discursive sense), but is rather used to convey the characteristic properties (i.e., the disposition) on the basis of which the underlying subject NP can be used in some way; for instance, example (13) says that the GUI is such that it is possible to tailor it to the needs of whatever is the method in question.

Palmer (2014, p. 38) claims that such subject-oriented modality is actually "not strictly a kind of modality at all, modality being essentially subjective", and that such modals are used "to make purely objective statements about the subject of the sentence" (*ibid.*). From the perspective of pragmatics, it does not seem that such dispositional modals actually constitute hedging of any kind given that they are used to convey objective properties of what the authors are describing in a given example. It should be noted that Hyland (1998, p. 5) claims that "hedges are the means by which writers can present a proposition as an opinion rather than a fact: items are only hedges in their epistemic sense, and only when they mark uncertainty". Examples (12) and (13) do not involve the speaker's opinion one way or the other; hence, they are not hedges. Lastly, we note that *možno* is used the most frequently in the BIO, PHYS, and TECH disciplines out of all the observed modals (see Table 3). We speculate that because it is used almost exclusively as a non-attitudinal dispositional modal, it is also well suited for the natural sciences, which are generally objective in that they deal "with numerical data, which is more likely to generate a more precise picture of their findings" Takimoto (2015, p. 95) than, e.g., the presumably more subjective and less empirical humanities.[15]

#### 5.2.4 Discussion

With the manual concordance analysis, we have shown that adverbs which mainly convey epistemic modality (and thus pragmatically function as

15 We do note, however, that the empirical vs. non-empirical divide partially transcends the distinction between humanities/social sciences on the one hand and natural/technical sciences on the other, but is rather influenced by the methodological framework adopted by the researcher. Thus, a thesis in a humanities discipline may be more concerned with empirical data than other theses in the same discipline.

accuracy-based hedges) are exactly those that are more frequent in the humanities and social sciences in our corpus. This result is generally consistent with related studies that compare the use of adverbial hedging between humanities disciplines on the one hand and natural sciences on the other. For instance, Takimoto (2015, p. 105) shows that, in his corpus, the English adverbs of epistemic possibility are used two times more frequently in the humanities than they are in the natural sciences. Similarly, Rizomilioti (2006, p. 64) shows that adverbs of uncertainty are used 1.2 times more frequently in her literary criticism corpus than in her comparable biology corpus, whereas the difference we have shown is even greater – on average, all the mainly epistemic modals (except for *verjetno*) in our corpus are 2.2 times more frequent in the humanities and social sciences.

Lastly, a note on *verjetno*: this modal is on average the most frequent in natural sciences discourse despite its purely epistemic meaning, as shown in Tables 3. We speculate that this is because *verjetno* does not seem to be completely synonymous with *najbrž*, which also entails likelihood. *Verjetno* seems to have a stronger evidential meaning, in the sense that it conveys that the speaker has some empirical evidence for judging the given proposition as likely, whereas *najbrž* seems more rooted in introspective speculation. A similar claim has been made for the distinction between the certainty modal auxiliaries in English, where the "difference between *will* and *must* is that *will* indicates what is a reasonable conclusion, while *must* indicates the only possible conclusion on the basis of the evidence available" (Palmer, 2014, p. 57).

To see whether *verjetno* truly has a stronger evidential meaning than *najbrž*, we have used the Collocations tool in the *noSketch Engine*, with which *KAS-dr* can be queried online. This tool allows us to observe how the two keywords differ in the collocates (i.e., co-occurring lexemes) that they pattern with, thus revealing larger co-textual differences between them. In the BIO subset of *KAS-dr*, the top-ranking collocates of *verjetno*, based on the MI Score,[16] are words directly related to empirical phenomena in biomedicine, such as *nevroinvazije* ("neuroinvasion"), *nepatogen* ("non-pathogenic"), and *polieter* ("polyether"), while the top-ranking collocates of *najbrž* are non-empirical,

16 The MI score "expresses the extent to which words co-occur compared to the number of times they appear separately" (https://www.sketchengine.eu/guide/glossary/).

meta-discursive expressions like *učinki* ("effects"), *posledica* ("consequence"), and *dejavnikov* ("factors"). If *verjetno* truly has a stronger evidential meaning than *najbrž*, as is hinted at by its collocational profile, then it comes as no surprise that it is the most frequent in biomedical sciences, where empirical evidence abounds.

### 5.3 Comparison of Epistemic Modal Adverbs Across Academic Stages

In this section, we compare the use of hedging in bachelor's, master's, and doctoral theses in *KAS-dipl*, *KAS-mag*, and *KAS-dr*, respectively. We do this for the following 9 modal adverbs: *verjetno, morda, zagotovo, gotovo, nedvomno, najbrž, domnenvo, morebiti*, and *menda*. These are the modals that almost exclusively (i.e., in more than 96% of the analysed concordances; see Table 4) convey epistemic modality, as was discussed in the previous section.[17] Because of their epistemic meaning, these modals invariably constitute *accuracy-based hedges* (Hyland, 1996) in terms of discourse pragmatics. Consequently, their distribution across the three *KAS* subcorpora offers a window into how authors' stance in relation to truth commitment changes from early (i.e., bachelor's and master's theses) to more proficient academic writing (i.e., doctoral theses).[18] Their distribution across the disciplines is also independent of thesis type, which is shown in Table 5, where each modal (save for *verjetno* in *KAS-dr*) is more frequent in the HUM and SOC disciplines than in BIO, PHYS and TECH in all the three subcorpora of *KAS*.

In Table 6, we now compare the frequencies of the 9 hedging adverbs between the bachelor's theses in *KAS-dipl* and master's theses in *KAS-mag*. The size of *KAS-dipl* is 1,101,796,659 tokens, while the size of *KAS-mag* is 495,827,656 tokens.

The frequencies of all the hedging adverbs are generally stable in both the bachelor's theses in *KAS-dipl* and the master's theses in *KAS-mag*. Overall, there is a negligible 0.6% decrease in the frequency of hedging from bachelor's

17 This is also independent of thesis type; for instance, *morda* in *KAS-dipl* is used as an epistemic modal in 97% cases in a random sample, which is similar to its modal-sense distribution in *KAS-dr* in Table 4.

18 For this reason, we omit the modals *lahko*, *možno*, and *mogoče* in this section. That is, they are not used exclusively in their epistemic sense and thus do not always relate to the authors' stance; see also the discussion of *možno* in the previous section.

**Table 5:** *The relative frequencies of the modals normalized to a million tokens in the 3 KAS subcorpora*

| | *KAS-dipl* | | *KAS-mag* | | *KAS-dr* | |
|---|---|---|---|---|---|---|
| **MODAL** | **HS** | **BPT** | **HS** | **BPT** | **HS** | **BPT** |
| *verjetno* "likely" | 110 | 89 | 105 | 94 | 127 | 128 |
| *morda* "possibly" | 95 | 57 | 91 | 57 | 118 | 54 |
| *zagotovo* "certainly" | 50 | 33 | 49 | 34 | 39 | 21 |
| *gotovo* "certainly" | 34 | 18 | 30 | 15 | 40 | 14 |
| *nedvomno* "certainly" | 29 | 12 | 28 | 13 | 33 | 11 |
| *najbrž* "likely" | 12 | 7 | 10 | 6 | 13 | 6 |
| *domnevno* "likely" | 6 | 3 | 5 | 4 | 10 | 4 |
| *morebiti* "possibly" | 9 | 6 | 11 | 7 | 12 | 5 |
| *menda* "possibly" | 2 | 1 | 2 | 0 | 4 | 0 |
| ∑ | 347 | 226 | 331 | 230 | 396 | 243 |

**Table 6:** *Hedging adverbs in bachelor's theses (KAS-dipl) and master's theses (KAS-mag)*

| | *KAS-dipl* | | *KAS-mag* | | | | |
|---|---|---|---|---|---|---|---|
| **MODAL** | **AF** | **RF** | **AF** | **RF** | **LLV** | ***p*** | **DIN** |
| *verjetno* "likely" | 115,248 | 105 | 51,487 | 104 | 1.892 | 0.1690 | 0.364 |
| *morda* "possibly" | 93,030 | 84 | 41,983 | 85 | 0.228 | 0.6325 | −0.141 |
| *zagotovo* "certainly" | 49,783 | 45 | 22,932 | 46 | 8.520 | 0.0035 | −1.166 |
| *gotovo* "certainly" | 32,710 | 29 | 13,425 | 27 | 81.751 | 0.0000 | 4.601 |
| *nedvomno* "certainly" | 27,058 | 25 | 12,519 | 25 | 6.561 | 0.0104 | −1.387 |
| *najbrž* "likely" | 11,849 | 11 | 4,548 | 9 | 85.103 | 0.0000 | 7.938 |
| *domnevno* "likely" | 5,509 | 5 | 2,168 | 4 | 28.515 | 0.0000 | 6.695 |
| *morebiti* "possibly" | 9,028 | 8 | 4,853 | 10 | 97.841 | 0.0000 | −8.863 |
| *menda* "possibly" | 2,019 | 2 | 639 | 1 | 63.710 | 0.0000 | 17.42 |
| ∑ | 346,234 | 314 | 154,554 | 312 | 7.024 | 0.008 | 0.405 |

theses (314 tokens per million) to master's theses (312 tokens per million). We have again used the *Calc: Corpus Calculator* (Cvrček, 2021) tool to compare the absolute pairwise frequencies statistically. The log-likelihood values (LLV), the related *p* scores, and the difference indices (DIN) calculated by the tool are given in the last three columns in Table 6 (see also Section 5.1 for how the LLV and DIN values are calculated). All the differences are statistically significant except for *verjetno* (LLV = 1.892; $p = 0.1690 > 0.05$) and *morda*

(LLV = 0.228; $p$ = 0.6325 > 0.05). A negative DIN value indicates that the modal is more frequent in the second group (i.e., master's theses), while a positive value indicates that the modal is more frequent in the first group (i.e., bachelor's theses), though the closer the value is to 0, the less prominent is the difference. The DIN value for the overall difference (LLV = 7.024; $p$ = 0.008 < 0.05) is 0.405, which reflects the fact that the epistemic modal adverbs are generally used at roughly the same frequency in bachelor's theses and in master's theses.

In Table 7, we compare the use of hedging adverbs between the bachelor's theses in *KAS-dipl* and the doctoral theses in *KAS-dr*. The size of *KAS-dr* is 101,473,395 tokens.

**Table 7:** *Hedging adverbs in bachelor's theses (KAS-dipl) and doctoral theses (KAS-dr)*

| | ***KAS-dipl*** | | ***KAS-dr*** | | | | |
|---|---|---|---|---|---|---|---|
| **MODAL** | **AF** | **RF** | **AF** | **RF** | **LLV** | ***p*** | **DIN** |
| *verjetno* "likely" | 115,248 | 105 | 12,958 | 128 | 439.879 | 0.0000 | –9.943 |
| *morda* "possibly" | 93,030 | 84 | 9,727 | 96 | 137.020 | 0.0000 | –6.336 |
| *zagotovo* "certainly" | 49,783 | 45 | 3,291 | 32 | 374.346 | 0.0000 | 16.429 |
| *gotovo* "certainly" | 32,710 | 30 | 3,152 | 31 | 5.816 | 0.0159 | –2.262 |
| *nedvomno* "certainly" | 27,058 | 24 | 2,534 | 25 | 0.644 | 0.4221 | –0.836 |
| *najbrž* "likely" | 11,849 | 11 | 1,082 | 11 | 0.072 | 0.7880 | 0.427 |
| *domnevno* "likely" | 5,509 | 5 | 969 | 10 | 296.129 | 0.0000 | –31.268 |
| *morebiti* "possibly" | 9,028 | 8 | 811 | 8 | 0.465 | 0.4952 | 1.246 |
| *menda* "possibly" | 2,019 | 2 | 315 | 3 | 66.565 | 0.0000 | –25.762 |
| Σ | 346,234 | 314 | 34,839 | 344 | 242.231 | 0.0000 | –4.423 |

All the hedging adverbs (except for *zagotovo, najbrž*, and *morebiti*) are used more frequently in doctoral theses than in bachelor's theses. Overall, there is a 9.5% increase in the frequency of hedging from bachelor's theses (314 tokens per million) to doctoral theses (344 tokens per million). All the differences are statistically significant except for *nedvomno* (LLV = 0.644; $p$ = 0.4221 > 0.05), *najbrž* (LLV = 0.072; $p$ = 0.7880 > 0.05), and *morebiti* (LLV = 0.465; $p$ = 0.4952 > 0.05). The DIN value for the overall difference (LLV = 242.231; $p$ = 0.0000 < 0.05) between bachelor's and doctoral theses is –4.423, which reflects the fact that doctoral theses employ the adverbs more frequently. In

sum, while hedging adverbs are used almost equally frequently in bachelor's and master's theses, their use increases in doctoral theses.

In Section 3.3, we saw that related work done in the context of English academic writing reports significant differences in hedging between different stages of the writers' academic progress. Aull and Lancaster's (2017) report results similar to ours in Table 7 in that they also see an increase in the use of hedging devices from less mature forms of academic writing such as students' research papers to more mature forms such as published journal papers. They interpret this difference by claiming that advanced academic writers are more likely to avoid an assertive stance in presenting their research than less experienced writers, favouring an approach to writing that is "implicitly attitudinal" and "open to other views in the surrounding discourse" (*ibid.*).

We propose that this also explains why hedging adverbs are more frequent in Slovenian doctoral theses (Table 7) in comparison to bachelor's and master's theses (Table 6). Relatedly, we speculate that the lack of such an increase from bachelor's theses to master's theses is because bachelor's theses together with master's theses constitute a uniform group in relation to research content and academic maturity. That is, most of the master's theses in *KAS-mag* (roughly 80%) are post-Bologna-reform master's theses that are in terms of academic maturity similar to the pre-Bologna bachelor's theses, in the sense that they are not (post)graduate research dissertations in contrast to doctoral theses.

This difference is evidenced in the official guidelines for (post)graduate programmes that are based on Slovenia's Higher Education Act, in which the aims of post-Bologna master's theses are more broadly defined than those of doctoral theses. For instance, according to the guidelines of the Faculty of Economics at the University of Ljubljana,[19] a master's thesis must present results that are "either achieved by the candidate's independent research or his or her expert evaluation of previous work". By contrast, similar guidelines for doctoral studies specify the aims of a doctoral thesis in narrower terms, in that it must necessarily present an original scientific contribution.[20] It is fur-

19 See Article 4 in http://www.ef.uni-lj.si/media/document_files/katalog_info_jav_znacaja/PravilaOMagistrskihDelihBolonjskiMagistrskiPrograma.pdf. (Accessed on 4 January 2020.)

20 See Article 35 in https://www.pef.uni-lj.si/fileadmin/Datoteke/Pravni_akti/Pravilnik_o_podiplomskem_%C5%A1tudiju_3.stopnje.pdf. (Accessed on 4 January 2020.)

thermore noteworthy that, at the University of Ljubljana, doctoral students (but not bachelor's and master's students) are required to publish at least one scientific paper in a peer-reviewed scientific journal before they are allowed to defend their thesis.

Post-Bologna master's theses may thus include only a discussion and evaluation of related work and need not present original research, whereas doctoral students hedge their novel claims in order to "negotiate solidarity with a reader who [might] hold contrary points of view" (Aull and Lancaster, 2014, 154), a pragmatic goal that is especially important in the context of peer review. In other words, it is precisely because Slovenian doctoral students are expected to present novel research that they more frequently employ accuracy-based hedges like the surveyed modal adverbs than undergraduate students writing bachelor's or post-Bologna-reform master's theses.

We wanted to confirm this by comparing the pre-Bologna master's theses, which used to be scientific works, with the post-Bologna master's theses, which inherited the old university diploma status of the concluding requirement at the undergraduate level. Although the *KAS-mag* subcorpus is not marked up for metadata that would distinguish these two master's thesis types, it is possible to demarcate them by publication date. The Bologna reform started to be implemented in Slovenia in 2004, so all the theses prior to this date must necessarily correspond to the old pre-Bologna scientific master's thesis. The pre-Bologna master's programme was gradually phased out in the 2010s, and the master's students enrolled in this system had to defend their theses by the end of the academic year of 2015/2016; consequently, all the theses in the last two publication dates in the subcorpus – 2017 and 2018 – correspond to the post-Bologna master's theses. (Conversely, the master's theses published in the remaining period – especially after 2010 and before 2016 – may correspond to either variant and it is difficult to distinguish between the two given the lack of mark-up, although the post-Bologna theses seem to be in the majority.)

By limiting our query to these two periods (2001–2004 and 2017–2018) in *KAS-mag*, which has yielded 449 theses (17,819,133 tokens) in the pre-Bologna subset and 2647 theses (65,764,329 tokens) in the post-Bologna subset, we are able to determine whether the frequency of hedging adverbs

changes between post-Bologna master's theses published in 2017–2018 and the pre-Bologna theses published in 2001–2004. The comparison is shown in Table 8.

**Table 8:** *The relative frequencies of hedging adverbs (per one million tokens) in KAS-mag*

| | post-Bologna (2017–2018) | | pre-Bologna (2001–2004) | | | | |
|---|---|---|---|---|---|---|---|
| **MODAL** | **AF** | **RF** | **AF** | **RF** | **LLV** | ***p*** | **DIN** |
| *verjetno* "likely" | 6,890 | 105 | 2,261 | 127 | 60.428 | 0.0000 | –9.548 |
| *morda* "possibly" | 5,395 | 82 | 1,426 | 80 | 0.696 | 0.4039 | 1.240 |
| *zagotovo* "certainly" | 2,956 | 45 | 618 | 35 | 36.333 | 0.0000 | 12.893 |
| *gotovo* "certainly" | 1,186 | 18 | 961 | 54 | 586.587 | 0.0000 | –49.881 |
| *nedvomno* "certainly" | 943 | 14 | 713 | 40 | 392.534 | 0.0000 | –47.236 |
| *najbrž* "likely" | 457 | 7 | 167 | 9 | 10.424 | 0.0012 | –14.845 |
| *domnevno* "likely" | 308 | 5 | 25 | 1 | 47.438 | 0.0000 | 54.897 |
| *morebiti* "possibly" | 585 | 9 | 156 | 9 | 0.0314 | 0.8593 | 0.798 |
| *menda* "possibly" | 74 | 1 | 39 | 2 | 10.410 | 0.0013 | –32.09 |
| ∑ | 18,794 | 286 | 6,366 | 357 | 228.236 | 0.0000 | –11.116 |

*Note.* The 2017–2018 theses are all post-Bologna master's theses, while the 2001–2004 theses are all pre-Bologna master's theses.

The majority of the hedging adverbs (5 out of 9) are more frequent in pre-Bologna master's theses (the so-called scientific masters), especially *gotovo* (DIN = –49.881) and *nedvomno* (DIN = –47.236), which are three times more frequent in the pre-Bologna subset. The frequency of two of the hedging adverbs, *morda* (DIN = 1.24) and *morebiti* (DIN = 0.798), is stable in both subsets, and their differences are not statistically significant ($p > 0.05$). There are only two hedging adverbs, *zagotovo* (DIN = 12.893) and *domnevno* (DIN = 54.897), which are more frequent in the post-Bologna theses. In total, pre-Bologna master's theses published before 2004 employ the hedging adverbs 24% more frequently than the post-Bologna master's theses published after 2017, which is a an even greater difference (LLV = 228.236; $p = 0.0000 < 0.05$; DIN = –11.116) than the one observed from bachelor's theses to doctoral theses reported in Table 7. This confirms our hypothesis that hedging is more common in original scientific contributions as is the case with doctoral and the pre-Bologna master's theses, which are in Slovenia referred to as *znanstveni*

*magisterij* ("scientific master's degree"), in contrast to their post-Bologna counterparts, which are referred to as *strokovni magisterij* ("professional/ expert master's degree").

## 8 CONCLUSION

In this paper, we have first analysed modal adverbs in the 100-million-token *KAS subcorpus of Slovenian doctoral theses*, comparing their frequency and use between humanities and social sciences on the one hand and natural sciences and technical sciences on the other. As one of our main contributions to research on hedging, we have taken into account the fact that modals are in actual usage often unpredictably ambiguous between epistemic and non-epistemic readings, and argued that only those modals that either convey epistemic judgements or meta-discursive commentary also function as hedges, whereas those that express dispositional possibilities do not. On the basis of this distinction, we have shown that the modals that are mainly used in the epistemic sense (and that thereby constitute accuracy-based hedges displaying varying degrees of the authors' tentativeness about the truth of the proposition) are used more frequently in Slovenian doctoral theses in the humanities and social sciences rather than the natural and technical sciences, which is generally in line with the related work (e.g., Takimoto, 2015; Hyland, 1998).[21]

Next, we have compared the use of the exclusively epistemic modal adverbs in theses at different stages of university education: bachelor's, master's and doctoral theses. We have shown that such modals are more frequent in doctoral theses than in bachelor's and master's theses, which is in line with the increase in hedging observed by Aull and Lancaster (2014) from first-year undergraduate writing to published research articles in the context of

21 It is difficult to say to what degree this trend can be generalised to hedging expressions other than modal adverbs. A problem here, as mentioned in Section 2.2, is that hedging is a pragmatic strategy and not a linguistic property (in the narrow sense), which means that a hedge can correspond not only to virtually any of the (open class) lexical categories (i.e., adverbs, adjectives, lexical verbs, nouns), but to many syntactic devices as well (the use of voice, mood, impersonalisation devices, etc.). To study this would require a manual analysis of the texts in the corpus, whereas for *KAS*, which is a very large corpus that is not syntactically parsed, we could only rely on the MSD-tags assigned to the tokens. We therefore leave such an analysis for future work.

American English academia. We have argued that such an increase in hedging observed in Slovenian doctoral reflects an important conceptual difference between bachelor's and post-Bologna master's theses on the one hand and doctoral theses on the other – that is, it is only doctoral theses that are research dissertations whose primary aim is presentation of novel research, the careful and responsible interpretation and discussion of which often needs to be properly hedged. We have confirmed this hypothesis by comparing the pre- and post-Bologna master's theses, the status of which has changed with the Bologna process from what was once a scientific degree to what is now a professional degree.

In our future work we would like to extend our analysis of the modals in the *KAS-dr* subcorpus to classes such as epistemic adjectives and verbs, while taking special care to properly account for the way their unique semantics interacts with the pragmatics. This will enable us to further ascertain whether expressions of epistemic modality are really more characteristic of humanities and/or social sciences disciplines across the board, as claimed by Takimoto (2015) and Hyland (1998), or whether they are a quirk of a specific word class, such as adverbs, as is claimed by Rizomilioti (2016). Furthermore, there might be prominent differences in the frequency of hedging between different parts of a thesis; for instance, the section dedicated to the discussion of results might contain many more hedging devices than the section dedicated to the research methodology (see also Thompson 2000 for precisely such findings for English). We also leave this for future work, as the *KAS* corpus is not annotated for thesis sections, nor is any other available Slovenian corpus.

Lastly, the extra-linguistic metadata in the *KAS* corpus also includes author-related information such as the name of the student and the advisor of the thesis. The second analysis presented in this paper could therefore be extended by taking into account how the use of hedging devices, such as epistemic adverbs, changes not only from undergraduate to (post)graduate theses in general, but also in the case of individual authors who first wrote a bachelor's or a master's thesis and then went on to pursue a doctoral degree. This would provide an even greater insight into the developmental trajectory of young Slovenian researchers as they advance through the higher educational system.

**Acknowledgments**

We would like to thank Maja Miličević Petrović for help with the statistics, and all the anonymous reviewers for their helpful comments. The work described in this paper was funded by the Slovenian Research Agency within the national research programme *Slovene Language – Basic, Contrastive, and Applied Studies* (P6-0215) and within the national basic research project *Slovene Scientific Texts: Resources and Description* (J6-7094, 2014-2017).

## REFERENCES

Aull, L. L., & Lancaster, Z. (2014). Linguistic Markers of Stance in Early and Advanced Academic Writing: A Corpus-based Comparison. *Written communication, 31*(2), 151–183. doi: 10.1177/0741088314527055

Aull, L. L., Bandarage, D., & Miller, M. R. (2017). Generality in student and expert epistemic stance: A corpus analysis of first-year, upper-level, and published academic writing. *Journal of English for Academic Purposes, 26*, 29–41. doi: 10.1016/j.jeap.2017.01.005

Bajec, A., et al. (Eds.). (2014). Možno (lexicographic entry). In *Slovar slovenskega knjižnega jezika*.

Coates, J. (1983). *The Semantics of the Modal Auxiliaries*. London and Canberra: Croom Helm.

Crosthwaite, P., Cheung, L., & Jiang, F. K. (2017). Writing with Attitude: Stance expression in learner and professional dentistry research reports. *English for Specific Purposes*, *46*, 107–123. doi: 10.1016/j.esp.2017.02.001

Cvrček, V. (2021). *Calc v1.02: Corpus Calculator*. Czech National Corpus. Retrieved from https://www.korpus.cz/calc/

DeLazero, O. E. (2011). On the Semantics of Modal Adjectives. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 87–94. Retrieved from https://repository.upenn.edu/pwpl/vol17/iss1/11/

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, *19*(1), 61–74.

Erjavec, T., Fišer, D., & Ljubešić, N. (2019a). *Corpus of Academic Slovene KAS 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1244

Erjavec, T., Fišer, D., & Ljubešić, N. (2019b). *Corpus of Academic Slovene (MSc/MA theses) KAS-mag 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1266

Erjavec, T., Fišer, D., & Ljubešić, N. (2019c). *Corpus of Academic Slovene (doctoral theses) KAS-dr 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1265

Erjavec, T., Fišer, D., & Ljubešić, N. (2019d). *Corpus of Academic Slovene (BSc/BA theses) KAS-dipl 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1267

Erjavec, T., Fišer, D., & Ljubešić, N. (2020). The KAS corpus of Slovenian academic writing. *Language Resource and Evaluation*. doi: 10.1007/s10579-020-09506-4

Erjavec, T. (2012). Mutext-East: Morphosyntactic Resources for Central and Eastern European Languages. *Language Resources and Evaluation*, *46*, 131–143. doi: 10.1007/s10579-011-9174-8

Fidler, M., & Cvrček, V. (2015). A Data-Driven Analysis of Reader Viewpoints: Reconstructing the Historical Reader Using Keyword Analysis. *Journal of Slavic Linguistics, 23*(2), 197–239. Retrieved from https://www.jstor.org/stable/24602151

von Fintel, K. (2006). Modality and language. In D. M. Borchert (Ed.), *Encyclopedia of Philosophy – Second Edition* (pp. 20–27). Detroit: MacMillan Reference USA.

von Fintel, K., & Gillies, A. (2007): An opinionated guide to epistemic modality. *Oxford studies in epistemology*, *2*, 32–63.

Grabe, W., & Kaplan, R. B. (1997). On the writing of science and the science of writing: Hedging in science text and elsewhere. In J. S. Petöfi (Ed.), *Hedging and Discourse* (pp. 151–167). De Gruyter, Berlin and New York.

de Haan, F. (2001). The Relation Between Modality and Evidentiality. *Linguistic Reports*, *9*, 201–216.

Hladnik, M. (2015). *Mind the Gap: Resumption in Slavic Relative Clauses*. LOT Publications. Retrieved from https://www.lotpublications.nl/mind-the-gap-resumption-in-slavic-relative-clauses

Hyland, K. (1996). Talking to the Academy: Forms of Hedging in Science Research Articles. *Written Communication*, *13*(2), 251–281. doi: 10.1177/0741088396013002004

Hyland, K. (1998). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins.

Hyland, K. (2004). Patterns of engagement: Dialogic features and L2 undergraduate writing. In L. Ravelli & R. A. Ellis (Eds.), *Analysing academic writing: Contextualized frameworks* (pp. 5–23). London, UK: Continuum.

Kratzer, A. (2012). The notional category of modality. In *Modals and Conditionals: New and Revised Perspectives* (pp. 27–69). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199234684.003.0002

Lancaster, Z. (2016). Expressing stance in undergraduate writing: Discipline-specific and general qualities. *Journal of English for Academic Purposes*, *23*, 16–30. doi: 10.1016/j.jeap.2016.05.006

Lakoff, G. (1972). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, *2*(4), 458–508. Retrieved from https://www.jstor.org/stable/30226076

Lenardič, J., & Fišer, D. (2020). Epistemic modal adverbs in Slovenian academic discourse. *Proceedings of the Conference on Language Technologies and Digital Humanities* (pp. 34–41).

Van Linden, A., & Davidse, K. (2009). The clausal complementation of deontic-evaluative adjectives in extraposition constructions: a synchronic-diachronic approach. *Folia Linguistica*, *43*(1), 171–211. doi: 10.1515/FLIN.2009.005

Marušič, F., & Žaucer, R. (2016). The modal cycle vs. negation in slovenian. In F. Marušič & R. Žaucer (Eds.), *Formal Studies in Slovenian Syntax* (pp.167–192). Amsterdam: John Benjamins. doi: 10.1075/la.236.08mar

Palmer, F. R. (2001). *Mood and Modality* (2nd ed.). Cambridge: Cambridge University Press.

Palmer, F. R. (2014). *Modality and the English modals*. Abingdon-on-Thames: Routledge.

Pihler Ciglič, B. (2017). Evidencialna branja prislova dizque v nekaterih različicah ameriške španščine in njegove ustreznice v slovenščini. *Ars & Humanitas*, *11*(2), 85–103. doi: 10.4312/ars.11.2.85-103

Piqué-Angordans, J., Posteguillo, S., & Andreu-Besó, J. V. (2002). Epistemic and Deontic Modality: A Linguistic Indicator of Disciplinary Variation in Academic English. *LSP & Professional Communication*, *2*(2), 49–65.

Pisanski Peterlin, A. (2010). Hedging Devices in Slovene-English Translation: A Corpus-Based Study. *Nordic Journal of English Studies*, *9*(2), 171–193. doi: 10.35360/njes.222

Pisanski Peterlin, A. (2015). So prevedena poljudnoznanstvena besedila v slovenščini drugačna od izvirnih? Korpusna študija na primeru izražanja epistemske naklonskosti. *Slavistična revija*, *63*, 29–44. Retrieved from https://srl.si/ojs/srl/article/view/COBISS_ID-57701986

Portner, P. (2009). *Modality*. Oxford: Oxford University Press.

Rizomilioti, V. (2006). Exploring Epistemic Modality in Academic Discourse Using Corpora. In *Information Technology in Languages for Specific Purposes*, Educational Linguistics, *7*, 53–71. Boston, MA: Springer. doi: 10.1007/978-0-387-28624-2_4

Rowbotham, M., Harden, N., Stacey, B., Bernstein, P., & Magnus-Miller, L. (1998). Gabapentin for the Treatment of Postherpetic Neuralgia: A Randomized Controlled Trial. *JAMA*, *280*(21), 1837–1842. doi: 10.1001/jama.280.21.1837

Takimoto, M. (2015). A Corpus-Based Analysis of Hedges and Boosters in English Academic Articles. *Indonesian Journal of Applied Linguistics*, *5*(1), 95–105. doi: 10.17509/ijal.v5i1.836

Thompson, P. (2000). Modal Verbs in Academic Writing. In B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis – Proceedings of the Fourth International Conference on Teaching and Language Corpora* (pp. 305–328).

Toporišič, J. (2004). Slovenska Slovnica. Maribor: Založba Obzorja.