

Grammatical and Pragmatic Aspects of Slovenian Modality in Socially Unacceptable Facebook Comments

Jakob LENARDIČ

Institute of Contemporary History

Kristina PAHOR DE MAITI

Faculty of Arts, University of Ljubljana; Institute of Contemporary History; CY Cergy Paris University

This paper investigates the grammatical and pragmatic uses of epistemic and deontic modal expressions in a corpus of Slovenian socially acceptable and unacceptable Facebook comments. We propose a set of modals that do not interpretatively vary in their modality type in order to enable robust corpus searches and reliable quantification of the results. We show that deontic, but not epistemic, modals are significantly more frequent in socially unacceptable comments, and specifically that they favour violent discourse. We complement the quantitative findings with a qualitative analysis of the discursive roles played by the modals. We explore how pragmatic communicative strategies such as hedging, boosting, and face-saving arise from the underlying syntactic and semantic properties of the modal expressions, such as the modal force and clausal syntax.

Keywords: corpus linguistics, modality, syntax, semantics, pragmatics, hate speech

Lenardič, J., Pahor de Maiti, K.: Grammatical and Pragmatic Aspects of Slovenian Modality in Socially Unacceptable Facebook Comments. Slovenščina 2.0, 11(1): 33–68.

1.01 Izvirni znanstveni članek / Original Scientific Article

DOI: <https://doi.org/10.4312/slo2.0.2023.1.33-68>

<https://creativecommons.org/licenses/by-sa/4.0/>



1 Introduction

Hate speech and other forms of socially unacceptable discourse have a negative effect on society (Delgado, 2019; Gelber and McNamara, 2016). For instance, calls to action targeting specific demographics on social media have been shown to lead to offline consequences such as real-world violence (Siegel, 2020). Linguistically, socially unacceptable attitudes are often disseminated in a dissimulated form, using pragmatic markers which superficially lessen the strength of intolerant claims or violent calls to action; nevertheless, the discursive markers of such dissimulated discourse are still not well known (Lorenzi-Bailly and Guellouz, 2019), especially outside of English social media.

In this paper, we investigate how Slovenian modal expressions contribute to the dissimulation of unacceptable discourse on social media at the interface between grammar (that is, syntax and semantics) on the one hand and discourse pragmatics on the other. We first perform a quantitative analysis, where we look at how the use of epistemic modals, which convey the speaker's truth commitment, and the use of deontic modals, which convey how the world ought to be according to a set of contextually determined circumstances, differ between unacceptable and acceptable discourse in the case of Slovenian Facebook comments obtained from the *FRENK* corpus (Ljubesić et al., 2021).

What follows is a qualitative discussion of all the observed modals.¹ We first discuss how the meaning of deontic necessity, which corresponds to some kind of obligation that needs to be fulfilled by the agent of the modalized proposition, can have a secondary pragmatic meaning that is akin to face-saving observed with epistemic modals and that arises with syntactically impersonal modals. We then discuss the only deontic likelihood modal, which is the adverb/particle *naj* ("should") co-occurring with a verb in the indicative mood, and discuss its usage

1 This paper extends our previous proceedings paper (Lenardič and Pahor de Maiti, 2022) along three dimensions. First, we make the quantitative analysis in Section 5 more precise by calculating the statistical significance for all the pairwise frequencies rather than just the overall differences. Second, in Section 6, we no longer exclusively discuss necessity modals but also the logically weaker modals – that is, those denoting likelihood or possibility. Third, we take into account an additional modal – that is, *dovoliti* –, which was omitted from the proceedings paper, while also taking into account the possible aspectual variation in verb forms.

in violent discourse from the perspective of its implicit subject-oriented semantics (Stegovec, 2019). Lastly, we discuss the deontic possibility modals denoting permission and their interaction with negation, which yields a lack of permission reading.

We then turn to the epistemic modals. We show that certainty and likelihood epistemics are primarily used to achieve a face-saving effect in pragmatics, and that they often occur in ironic contexts. We discuss how irony affects face-saving and its interaction with the underlying epistemic modality, claiming that what is being communicated with epistemics in such cases is often not truth commitment but the attenuation of controversial or impolite claims. For epistemic possibility modals, we also discuss their so-called concessive role in discourse (Palmer, 2014) in relation to the interaction between face-saving and the underlying epistemic meaning.

The paper is structured as follows. Section 2 presents the semantic and pragmatic properties of epistemic and deontic modals, while Section 3 presents some of the related corpus-linguistic work on modality in socially unacceptable discourse. Section 4 describes the make-up of the *FRENK* corpus in terms of the subtypes of socially unacceptable discourse and the criteria for the selection of the analysed modals. Section 5 presents the quantitative analysis, wherein epistemic and deontic modals are compared between the acceptable and unacceptable supersets in *FRENK*. Section 6 presents the qualitative analysis, where deontic and epistemic necessity modals are discussed in relation to the way grammar interfaces with the pragmatics. Section 7 concludes the paper.

2 Theoretical background

2.1 Semantic assumptions

Modal expressions are sentential operators that interpret a preajcent proposition within the irrealis realm of possibility (Kratzer, 2012). In terms of their lexical semantics, modal expressions are underspecified in the sense that they only encode the so-called modal force, which ranges from possibility via likelihood to necessity (Kratzer, 2012; von Stechow, 2006). The possibility, likelihood or necessity expressed by the modal is interpreted

relative to what goes on in the actual world, but this meaning component is not, in contrast to modal force, lexically encoded by the modal but rather determined by the linguistic or even extra-linguistic context.

In this paper, we look at two such contextually determined interpretations of modal expressions – the epistemic interpretation on the one hand and the deontic interpretation on the other (Coates, 1983; Kratzer, 2012; Palmer, 2014; von Stechow, 2006). Epistemic modals tie the evaluation of the possibility, likelihood, or necessity to the speaker’s knowledge about the actual world. For instance, the possibility adverb *morda* in (1), taken from the *FRENK* corpus, has the reading which says that there is a possibility that the referents of the indefinite subject *nekaj jih* (“some of them”) will stay in the country. This possibility reading is epistemic as it conveys that the speaker is not sure whether the possibility of their staying will actually turn out to be the case.²

- (1) [N]ekaj jih bo *morda* ostalo v naših krajih.
 “Some of them will *possibly* stay in our country.”

By contrast, deontic modals do not tie the evaluation of possibility or necessity to the speaker’s knowledge but to some contextually determined authority, such as a set of rules, the law, or even the speaker (Palmer, 2001, 10). An example of a deontic modal is the verb *dovoliti* in example (2), again taken from *FRENK*. This verb also denotes possibility in terms of modal force, so the deontic possibility reading roughly translates to *they should not be given the possibility* (i.e., be allowed) *to change our culture*.

- (2) [S]eveda se jim ne sme *dovoliti* da bi spremenil *naso* (*sic*) kulturo.
 “They should not be *allowed* to change our culture.”

A single modal can have different readings in terms of modality type. This is, for instance, the case with the necessity modal *morati*, where the epistemic reading in (3a) conveys that the speaker is certain (i.e., epistemic necessity) that whomever they are referring to is a *bona fide* Slovenian. By contrast, the deontic reading in (3b) says that what needs to be necessarily done is preparing for the competition.

2 For ease of readability, the modal under scrutiny is typeset in italics.

- (3) a. Ta *mora* biti pravi Slovenec, ni dvoma.
“He *must* be a *bona fide* Slovenian, no doubt about it.”
b. Pripraviti se bodo *moral*i tudi na konkurenco, ki je zdaj še nimajo.
“They *must* also prepare for the competition which they do not have.”

(Roeder and Hansen, 2006, p. 163)

Lastly, note that lexical semantic notions relating to modal force, such as possibility and necessity, should not be conflated with related interpretative notions, such as uncertainty and certainty, which, however, are not lexically entailed. To exemplify, while the adverb *mogoče* and the related predicative adjective *mogoč* are both invariably possibility modals, they need not necessarily always express uncertainty under the epistemic reading. This can be seen when they are paired up with negation, which is interpreted below the modal in the case of the adverb *mogoče*, yielding the uncertainty reading, and above the modal in the case of the adjective *mogoč* (inflected for neuter *-e* because of the subjectless syntax of the matrix clause), yielding the certainty reading, as shown by the paraphrases of (4).³

- (4) a. *Mogoče* ni bila dovolj socialna.
“It is *possible* that she was not sociable enough.”
a. [N]i *mogoče*, da bi bil islam na enaki stopnji kot Zahod.
“It is not *possible* that Islam is on the same level as the West.”

The fact that (4a) has a weaker interpretation than (4b) is thus a compositional effect of the different relative scopes of negation, even though at their core both modals still express possibility, as is also indicated by the paraphrases. Being lexically entailed, the force is a stable semantic property of a modal expression, which is why we will refer to a modal such as *mogoče* as a “possibility modal” rather than an “uncertainty modal” (even though it turns out that the adverbial use of *mogoče* always expresses uncertainty and never certainty in *FRENK*) through the rest of the paper.

³ While example (4a) is from *FRENK*, (4b) is from the Slovenian reference corpus *Gigafida 2.0* (Krek et al., 2019), as such negated instances of the adjectival form are not present in *FRENK*.

2.2 Modals pragmatically

Because of the intensional semantics of modals and the contextual variability of the way in which the modal force is interpreted, modals are able to play several (often overlapping) roles in discourse. In functionalist terms, they are important mainly from the perspective of the interpersonal dimension of communication (Halliday, 1970).

Interpersonally, epistemic modals are used in both positive and negative politeness strategies to satisfy the positive face needs of the speakers and the addressee, i.e., the need to be liked by the interlocutor, as well as their negative face needs, i.e., the need to act independently (Brown and Levinson, 1987). Epistemic modals show the following three pragmatic uses (Coates, 1987) related to Brown and Levinson's (1987) Politeness Theory. First, they are used as part of the negative politeness strategy to save the addressee's negative face, when for instance the speaker tries to facilitate open discussion by not assuming the addressee's stance on the conversational issue in advance. Second, epistemic modals can be used as an *addressee-oriented* positive politeness strategy, which involves the preservation of the positive image of the addressee and prevents them from feeling inferior to the speaker. Finally, they are used as part of a *speaker-oriented* positive politeness strategy, which involves the preservation of the positive image of the speaker by enabling the smooth withdrawal from a statement that can be perceived as a boast, threat, or similar.

Related to such politeness strategies, modals fulfil the conversational role of so-called hedging or boosting devices (Hyland, 2005). Epistemic modals function as hedges when the speaker uses them to reduce their commitment to the truth of the propositional content – i.e., to signal their hesitation or uncertainty with regard to what is being expressed, which is a type of face-saving strategy in and of itself (Gonzalez Garcia, 2000; Hyland, 1998a). In terms of modal force, it is weak epistemic modals denoting possibility that typically correspond to hedges, though certain necessity modals can also acquire such a function in certain contexts, as we will show in the qualitative analysis.

Strong epistemic modals, which express the certainty or high commitment of the speaker to the truth of the utterance, typically function

as boosters and are used by the speaker to convince his or her audience, make his or her utterance argumentatively stronger, close the dialogue for further deliberation (Vukovic, 2014), stress the common knowledge and group membership (Hyland, 2005), and so forth. Such boosters can also be used manipulatively to boost a claim that is otherwise controversial or highly particular (Vukovic, 2014).

Deontic modality also fulfils interpersonal roles in communication. Because deontic modals express notions such as obligation and permission, they have to do with negotiating social power between an authority and the discourse participant to whom the permission is granted or obligation imposed upon (Winter and Gärdenfors, 1995). Deontic statements often involve a power imbalance between interlocutors (which is especially evident in cases when it is not in the interest of the agent to fulfil the obligation), so the use of deontic modals is often paired up with other pragmatic devices denoting politeness or face-saving. Politeness is thus “an overarching pragmalinguistic function that can be overtly or covertly marked in deontic and epistemic modal utterances” (Gonzalvez Garcia, 2000, p. 127).

3 Related work on modality in socially unacceptable discourse

The linguistic and pragmatic characteristics of modality in online socially unacceptable discourse have not yet been extensively explored in the literature. One exception is the work done by Ayuningtias, Purwati, and Retnaningdyah (2021), who analyse YouTube comments related to the 2019 Christchurch mosque shootings in New Zealand. They find that clauses with deontic modals outnumber those with epistemic modals, and that the main discursive strategy of commenters in socially unacceptable comments is to use deontic modals to incite violent action against members of the Muslim community.

Other corpus linguistic studies investigate modal markers from the perspective of stance. Chiluya (2015), for example, analyses the stance expressed in the tweets of two radical militant groups, Boko Haram and Al Shabaab. Among other stance-related elements, she investigates the use of hedges (including weak epistemic modals) and

boosters (including strong epistemic modals). The results show that boosters are more frequent than hedges, although their overall frequency in the data was low. According to the author, the low frequency of hedges shows that radicalist discourse does not exhibit the tendency to mitigate commitment, which goes hand in hand with the slightly higher presence of boosters that are used as a rhetorical strategy to support (possibly unfounded) statements and to influence, radicalize and win over their readers by projecting assertiveness.

Another study on stance in this context is by Sindoni (2018), who looks at the verbal and multimodal construction of hate speech in British mainstream media. She analyses epistemic modal operators (among other related devices) in order to uncover the writer's stance and attitude towards the content conveyed in the news item. She finds that modality is strategically used to present the author's opinions as facts, while the opinions of others are reported as hypotheses and assumptions.

4 The *FRENK* corpus

4.1 Corpus make-up

For this study, we have used *FRENK*, a 270,000-token corpus of Slovenian Facebook comments of mostly socially unacceptable discourse (Ljubešić et al., 2019). The Facebook comments in the *FRENK* corpus concern two major topics – migrants, generally in the context of the 2015 European migrant crisis, and the LGBTQ community, mostly in the context of their civil rights – and are manually annotated for several different kinds of discourse.⁴ The annotations distinguish whether the discourse is aimed towards a target's personal background, such as sexual orientation, race, religion, and ethnicity, or their belonging to a particular group, such as political party. They also distinguish the type of the discourse itself, which falls into four broad categories, one being acceptable discourse and the others different kinds of socially unacceptable discourse (Pahor de Maiti et al., 2019, p. 38):

- Acceptable discourse
- Socially unacceptable discourse

4 The annotations are performed on the comment level while also taking into account the features of the entire discussion thread.

- Offensive discourse, which corresponds to abusive, threatening or defamatory speech that is targeted towards someone on the basis of their background or group participation.
- Violent discourse, which contains threats or calls to physical violence and is often punishable by law (Fišer et al., 2017, p. 49).
- Inappropriate speech, which contains offensive language but is not directed at anyone in particular.

For our study, we have created two subsets of comments: the *acceptable subset* containing comments tagged as *acceptable*, and the *unacceptable subset* containing comments tagged as *offensive*, *violent* or *inappropriate*. This decision is based on the frequency distributions shown in Table 1.

Table 1: *The make-up of the FRENK corpus in terms of socially (un)acceptable discourse*

Subcorpus	Tokens	
Acceptable	92,922	34%
Offensive	143,948	53%
Inappropriate	1,471	1%
Violent	8,789	3%
Not relevant	24,572	9%
Σ	271,702	100%

The FRENK subcorpora are uneven in terms of size, with the violent and inappropriate sets contain significantly fewer comments than the acceptable and offensive sets. Because violent discourse is generally less frequent than offensive discourse in linguistic corpora,⁵ it is difficult to annotate automatically (Evkoski et al., 2022), so one of the crucial features of *FRENK* is the fact that the annotations into discourse type were done manually, employing eight trained annotators per Facebook comment (Ljubešić et al., 2019, p. 9). Note that about 9% of the Facebook comments are marked as *not relevant*, which refers to comments with incorrect topic classification (*ibid.*, 5).

⁵ This is also a result of the EU Code of conduct and terms of service of social media platforms, according to which content deemed illegal due to its hateful character needs to be taken down.

The latest, that is, version 1.1, of the *FRENK* corpus, which also includes texts in Croatian and English, is available for download from the CLARIN.SI repository (Ljubesić et al., 2021). However, the online version, which is accessible through CLARIN.SI's NoSketch Engine concordancer and which we have used for the purposes of this paper,⁶ is not yet available to the public.

4.2 The modals analysed in the study

Table 2 shows that there are 13 modal expressions used in the study. We have selected the modals using two criteria.

The first criterion is the modal's tendency towards a single modal reading. As discussed in Section 2.1, modals are in principle ambiguous in terms of the modality type. However, corpus data show that certain modals have an overwhelming preference for a single reading. For instance, while the modal auxiliary *morati* can theoretically have both the epistemic and deontic interpretations (Roeder and Hansen 2006, pp. 162–163), as was shown in (3), the epistemic reading (3a) is actually extremely rare in attested usage, and in the case of the *FRENK* corpus completely non-existent.⁷ Similarly, whenever the adverb *naj*_{IND} is used in the indicative rather than conditional mood (glossed with the subscript IND in Tables 2 and 4 and through the rest of the paper), its meaning is always some shade of the deontic reading (command, wish, etc.). Thus, all the modals in Table 2 are either unambiguously deontic or unambiguously epistemic, so they function as a robust set for testing how deontic and epistemic modality manifests itself in different types of discourse without confounding examples with unintended interpretations.

6 <https://www.clarin.si/noske>

7 With the exception of *potrebno*, the frequency counts were performed on lemmas, as this is sufficient for distinguishing the part of speech as well. For instance, the lemma *mogoče* corresponds to the adverbial forms, whereas the lemma *mogoč* corresponds to the adjectival ones; however, the adjectival form when used predicatively is consistently ambiguous between the non-epistemic and epistemic interpretations, see Lenardič and Fišer (2021) for discussion and examples. In the case of *potrebno*, we queried the word form, as the lemma *potreben* also yields attributive uses (which are irrelevant because we are focusing on sentential modality), whereas the word form *potrebno* reliably yields the predicative uses.

Table 2: The analysed modals; AF stands for absolute frequency

Modal	Syntax	Modality	Force	AF
<i>naj</i> _{IND}	Adverb	Deontic	Likelihood	886
<i>morati</i>	Verb	Deontic	Necessity	489
<i>treba</i>	Adjective	Deontic	Necessity	306
<i>smeti</i>	Verb	Deontic	Possibility	150
<i>verjetno</i>	Adverb	Epistemic	Likelihood	123
<i>mogoče</i>	Adverb	Epistemic	Possibility	92
<i>potrebno</i>	Adjective	Deontic	Necessity	65
<i>dovoliti</i>	Verb	Deontic	Possibility	60
<i>morda</i>	Adverb	Epistemic	Possibility	46
<i>najbrž</i>	Adverb	Epistemic	Likelihood	29
<i>ziher</i>	Adverb	Epistemic	Necessity	25
<i>dopustiti</i>	Verb	Deontic	Possibility	19
<i>zagotovo</i>	Adverb	Epistemic	Necessity	16
Σ				2245

The catenative possibility verbs *dovoliti* and *dopustiti* also have the imperfective forms *dovoljevati* and *dopuščati*. As they are rarer than their perfective variants (i.e., 55 instances of *dovoliti* vs. five instances of *dovoljevati* and 11 instances of *dopustiti* vs. eight instances of *dopuščati*), they are counted in Table 2 under the morphologically less complex perfective forms. However, the aspectual distinction does not affect, at least in the *FRENK* data, the modality type, which stays deontic; the imperfective form in (5b) only seems to trigger or emphasize the continuous interpretation of the permission.

- (5) a. Sloveniji ne bomo *dovolili* nastanka nasilnih band.
 “We won’t *allow* violent gangs to form in Slovenia.”
 b. Problem je v politiki, katera *dovoljuje* islamizacijo nase dezele
 “The problem is in politics, which continues to *allow* the
 Islamization of our land.”

The imperfective form of *dopustiti* is said to have an epistemic (or rather, doxastic) interpretation (Močnik, 2019), as shown in the constructed example in (6), which conveys that the speaker is not certain where the referent of the null subject of the embedded clauses is located.

(6) *Dopuščam*, da je notri, in *dopuščam*, da je zunaj.

“He *might* be inside and he *might* be outside.”

(Močnik, 2019, p. 422)

However, in *FRENK* all instances of imperfective *dopuščati* convey the deontic interpretation of permission, as shown in the corpus example in (7). As with *dovoliti* and *dovoljevati*, the aspectual distinction does not affect the modal interpretation, which invariably stays deontic.

(7) [S]amo Slovenija *dopušča* izdajalcem i[n] koloborantom, da serjejo po državi

“Only Slovenia *allows* traitors and collaborators to shit all over the country.”

The second criterion concerns the fact that some lexemes known to convey modal interpretations also occur frequently with a superficially similar propositional meaning that, however, is not modal. On such case is the adverb *itak*, as in example (8), also taken from *FRENK*.⁸

(8) Krscanstvo pa *itak* izvira iz istih krajev kot islam in juduizem (*sic*).

“*Of course*, Christianity comes from the same place as Islam and Judaism.”

This adverb differs from, for example, the certainty adverb *zagotovo* in that it does not convey the speaker’s degree of certainty,⁹ but

8 With these two criteria as filters, it does not appear that many modals have been excluded from our study – that is, in Slovenian, ambiguous modals do not outnumber the unambiguous ones, at least not considerably. For instance, most of the modals discussed by Roeder and Hansen (2007) in their overview of Slovenian modal expressions are included in our study (the two exceptions being *lahko*, which is consistently ambiguous between epistemic, ability i.e. “dynamic”, and deontic readings, and *utegniti*, which is ambiguous between ability and epistemic readings). Note that the findings presented in this paper are not generalizable to the ambiguous modals like *lahko*; one would first have to manually annotate (a subset of) them to determine how frequent their different readings are, which is beyond the scope of the present paper.

9 *Zagotovo* has the synonym *gotovo*; we have excluded it from our overview because it is too frequently used in the non-modal sense, as in (1), which is mostly typical of non-standard Slovenian.

(1) Postrelit in *gotovo*.

“Shoot them all – that’s the end of it.”

rather simply intensifies whatever he or she knows to be actually the case (the historical-geographic source of Christianity). Because such non-modal readings are usually as frequent as the modal meaning in attested usage, we have omitted them from our study.

Lastly, note that in terms of syntactic category the modals in Table 2 do not constitute a homogeneous set. While most modals are syntactically adverbs (e.g., *morda*, *ziher*), some are verbs selecting for finite clausal complements, such as *dovoliti* in (2) and *dopuščati* in (7), verbs selecting for non-finite complements, such as *morati* in (3), and predicative adjectives (of the syntactic frame *It is necessary to*) selecting for non-finite complements, such as *treba* (see the examples in Section 6.1.1). However, such syntactic differences have no bearing on the modal interpretation – in all cases, the modals remain sentential operators that take semantic scope over the proposition denoted by the clause.

5 Quantitative analysis

5.1 The distribution of the modals between acceptable and unacceptable discourse

Tables 3 and 4 show how the Slovenian modals are distributed between the acceptable and unacceptable subsets for the unambiguously epistemic and deontic modals, respectively. The unacceptable subset brings together the three subtypes – offensive, inappropriate, and violent – introduced in Section 4.1. The acceptable and unacceptable sets contain 92,922 and 154,208 tokens, respectively.

In the epistemic set (Table 3), half of the modals – that is, the possibility modal *mogoče* and the necessity modals *ziher* and *zagotovo* – are more frequent in the corpus of unacceptable discourse, while the remaining 3 modals – that is, the possibility modal *morda* and the logically synonymous likelihood modals *najbrž* and *verjetno* – are more frequent in the subset of socially acceptable discourse. Overall, the six epistemic modals are 1.2 times more frequently used in acceptable discourse than they are in unacceptable discourse.

Table 3: The distribution of epistemic modals in the FRENK corpus; AF stands for absolute frequency and RF for relative frequency, normalized to a million tokens

Modal	Acceptable		Unacceptable		A/U	LL	P	DIN
	AF	RF	AF	RF				
<i>verjetno</i>	52	559.6	66	428.0	1.3	2.1	0.1508	13.3
<i>morda</i>	24	258.3	19	123.2	2.1	5.8	0.0156	35.4
<i>mogoče</i>	29	312.1	55	356.7	0.9	0.3	0.5581	-6.7
<i>najbrž</i>	12	129.1	13	84.3	1.5	1.1	0.2898	21.0
<i>zagotovo</i>	3	32.3	13	84.3	0.4	2.7	0.1011	-44.6
<i>ziher</i>	8	86.0	15	97.3	0.9	0.1	0.7791	-6.0
Σ	128	1,377.4	181	1,173.8	1.2	1.9	0.1676	7.9

The distribution is reversed in the set of unambiguously deontic modals (Table 4). As shown in the U/A column, all modals, save for the possibility verb *smeti* (“to allow”), are more characteristic of unacceptable rather than acceptable discourse, with the deontic necessity adjective *treba* and deontic likelihood adverb *naj_{IND}* (“should”) showing the largest preference for the unacceptable set. Overall, the seven deontic modals are 1.3 times more frequently used in socially unacceptable discourse than they are in acceptable discourse.

Table 4: The distribution of deontic modals in the FRENK corpus

Modal	Acceptable		Unacceptable		U/A	LL	p	DIN
	AF	RF	AF	RF				
<i>naj_{IND}</i>	227	2,442.9	583	3,780.6	1.5	33.1	8.6×10^{-9}	-21.5
<i>morati</i>	151	1,625.0	292	1,893.6	1.2	2.4	0.1238	-7.6
	87	936.3	197	1,277.5	1.4	6.0	0.0139	-15.4
<i>smeti</i>	41	441.2	60	389.1	0.9	0.4	0.5364	6.3
<i>potrebno</i>	24	258.28	41	265.9	1.0	0.0	0.9101	-1.5
<i>dovoliti</i>	18	193.7	38	220.5	1.1	0.7	0.3939	-12.0
	4	43.1	15	97.3	2.3	2.4	0.12	-38.6
Σ	552	5,940.5	1226	7,924.5	1.3	33.4	6.5×10^{-9}	-14.5

Statistically, we have tested the differences in pairwise frequencies for all modals, as well as the overall differences between the unacceptable and acceptable sets in both the epistemic (Table 3) and deontic (Table 4) modals. We have used the log-likelihood statistic, which

“establish[es] whether the differences [between pairwise frequencies in two corpora with different sizes] are likely to be due to chance or are statistically significant” (Brezina, 2018, pp. 83–84). The formula for calculating the log likelihood statistic is given in (9), where the observed values $O_{1,2}$ correspond to the absolute frequencies of a modal in the unacceptable and acceptable sets. The loglikelihood value, labelled LL, is given in the antepenultimate column in each table.

$$(9) 2 \times (o_1 \times \ln(\frac{O_1}{E_1}) + o_2 \times \ln(\frac{O_2}{E_2}))$$

In the epistemic set (Table 3), only one out of the six modals shows a statistically robust difference between the acceptable and unacceptable sets at $p < 0.05$ – that is, *morda*, whose LL value is 5.8 and $p = 0.0156$. The overall greater occurrence of epistemic modals in the acceptable set (AF = 128 tokens, RF = 1,377.4 tokens/million) than in the unacceptable set (AF = 181 tokens, RF = 1,173.7 tokens/million) is statistically unreliable; LL = 1.9, $p = 0.1676$.

In the deontic set (Table 4), only two out of the seven modals show a statistically robust difference – the likelihood modal *naj*_{IND} (LL = 33.1, $p = 8.6 \times 10^{-9}$) and the necessity modal *treba* (LL = 6.0, $p = 0.0139$). However, in contrast to the epistemic set, the overall greater occurrence of deontic modals in the unacceptable set (AF = 1226 tokens; RF = 7,924.5 tokens/million) than in the acceptable one (AF = 552 tokens; RF = 5,940.5 tokens/million) is statistically significant at the same cut-off point; log likelihood = 33.4, $p = 6.5 \times 10^{-9}$.

Using the online tool *Calc* (Cvrček, 2021), we have also calculated the Difference Index (DIN) – an effect-size metric – for all the pairwise differences in frequency. The DIN values, which are given in the final columns of Tables 3 and 4, are calculated using Fidler and Cvrček’s (2015, p. 230) formula in (10), where RF_1 and RF_2 are the respective relative frequencies of the modals in the acceptable and unacceptable sets.

$$(10) \text{DIN} = 100 \times \frac{RF_1 - RF_2}{RF_1 + RF_2}$$

The lines in (11) say how the DIN values are to be interpreted (*ibid.*).

- (11) a. DIN = -100: the word is present only in corpus 2 (i.e., unacceptable set) but not in corpus 1 (i.e., acceptable set)
 b. DIN = 0: the word occurs equally often in corpora 1 and 2 (i.e., in the acceptable and unacceptable sets)
 c. DIN = 100: the word is present only in corpus 1 (i.e., acceptable set) but not in corpus 2 (i.e., unacceptable set)

The highest DIN values belong to the modals *morda* (35.4) and *najbrž* (21.0), indicating their greatest preference for the acceptable set, while the lowest belong to *zagotovo* (-44.6), *dopustiti* (-38.6) and *naj_{IND}* (-21.5), indicating the modals' preference for the unacceptable set.

Note that out of these five modals with high/low DIN values, only *morda* and *naj_{IND}* also show statistically robust differences according to the log-likelihood test, while *dopustiti*, *zagotovo*, and *najbrž* show statistically unreliable differences. It is likely that the insignificant differences are due to the relatively small size of the *FRENK* corpus, and the attendant fact that all modals save for *naj_{IND}* have quite low absolute frequencies (e.g., 19 frequencies of occurrence for *dopustiti* in Table 4). We do note, however, that even if we were to disregard *naj_{IND}* as a possible statistical confounder, the overall difference in the deontic set would remain significant, LL = 6.7, $p = 0.009$.

5.2 Comparison with Previous Work

The findings presented in the previous subsection are related to those in the literature (see Section 3) as follows. Just like in Ayuningtias, Purwati, and Retnaningdyah's (2021) work on socially unacceptable discourse in YouTube comments, our deontic modals outnumber epistemic modals in both the acceptable and unacceptable sets (e.g., 1,226 deontic modals vs. 181 epistemic modals in the unacceptable set). Second, both modals of epistemic necessity in Table 3 – that is, *zagotovo* and *ziher* (“certainly”) – differ from most of the weaker modals, like *morda* (“possibly”) and *najbrž* (“likely”), in that they are more frequent in unacceptable discourse; this is similar to the finding by Chiluwa (2015), who shows that strong epistemic modals are more frequent than weak ones in the case of tweets by radical militant

groups. However, and in contrast to Chiluya (2015), our statistically robust finding is not the difference in modal force, but rather the difference in modality type, as discussed above.

6 Qualitative analysis

6.1 Deontic modals in violent discourse

In Section 5, it was shown that deontic modals are more typical of unacceptable discourse than they are of acceptable discourse, a finding that was shown to be statistically robust.

Table 5: *The distribution of deontic modals between the offensive and violent subsets of FRENK; the frequencies are relative and normalized to a million tokens*

Modal	Acceptable	Violent	Offensive
<i>treba</i>	936.3	4,437.4	1,083.7
<i>potrebno</i>	258.3	568.9	243.1
<i>dovoliti</i>	183.0	341.3	213.2
<i>smeti</i>	441.2	682.7	405.7
<i>morati</i>	1,625.0	1,479.1	1,910.4
<i>naj</i> _{IND}	2,442.9	6,371.6	3,647.2
Σ	5,893.7	13,881.0	7,503.3

To look at the pragmatics of deontic modals and their discursive role in relation to socially unacceptable discourse, let's first recall from Section 4.1 that the socially unacceptable discourse in the *FRENK* corpus is further subdivided into several subtypes. Here we focus on two – offensive discourse on the one hand and violent discourse on the other. It turns out that all of the surveyed deontic modals, with the exception of the auxiliary *morati*, are consistently more frequent in violent discourse than in offensive discourse.¹⁰ This is shown in Table 5, where for instance *treba* is almost four times as frequent in the violent-speech subset (RF = 4,437.3 tokens per million) as it is in the offensive subset (RF = 1,083.7 tokens per million).

¹⁰ A caveat for comparison, of course, is that the violent subset is much smaller than the acceptable and offensive sets (see Table 1), which is why we report the general trend rather than focus on specific pairwise differences.

6.1.1 Deontic necessity

Let us begin with the observation that *treba* and *morati* are synonymous, possibly completely so, in terms of modal logic, as both entail necessities in terms of modal force and invariably have a deontic reading that has to do with a contextually determined obligation. However, despite the synonymy, *treba* is by far more frequent in violent speech than it is in offensive speech, while *morati* is the only deontic modal that is more prominent in offensive than in violent speech.

The difference in the distribution of the two synonymous modals can be tied to the fact that they vastly differ in their communicative function, which crucially is observable within the same subset. Put plainly, the chief difference is that *treba* occurs in considerably more hateful statements than *morati*, even though the statements all qualify as violent hate speech rather than offensive speech in that some kind of incitement towards violence is expressed in the modalized statement.

For instance, let's first consider some typical examples with *treba* from the violent subset:

- (12) a. To golazen *treba* zaplinit, momentalno!!!!
 “These vermin *must* be gassed at once!”
- b. Pederčine je *treba* peljat nekam in postrelit.
 “Faggots *must* be taken somewhere and shot.”
- c. Ni *treba* par tisoč Voltov, dovolj je 220, da ga strese in opozori, da bo čez par metrov stražar s puško.
 “We don't *need* a couple of thousand volts; 220 is enough to electrocute them and warn them that, a couple of metres further on, an armed guard is waiting.”

The chief linguistic characteristic of the *treba* examples boils down to lexical choice. The most prominent nominal collocates in the violent subset for the *treba* examples, calculated on the basis of the Mutual Information statistic, is *golazen* “vermin”, which can be seen in example (12a), where migrants are referred to as such.¹¹ According to Assima-

11 As an anonymous reviewer notes, such collocational analysis should be taken tentatively because of how small the violent subset is in *FRENK*. For this reason, the presented collocations might not be a property of violent Slovenian discourse in general, and could be limited specifically to the *FRENK* data.

kopoulos, Baider, and Millar (2017, 41), such metaphoric expressions “are an intrinsic part of the Othering process, and central to identity construction”.

In the case of animal metaphors such as *M IGRANTS ARE VERM IN*, migrants are conceptually construed and stereotyped as an invasive out-group that is maximally different from the in-group to which the speaker considers themselves to belong (*ibid.*). The other most prominent nominal collocates is *elektrika* (“electricity”); metaphors containing this lexeme or lexemes related to electricity (volts, to shock, etc.) often have implied reference, where the undergoers of the verbal event, i.e., migrants, are not directly mentioned, as shown in example (12c). Curiously, when the targets of violent speech are not migrants but members of the LGBT community, instead of metaphors like *golazen*, slurs such as *pedri* (“faggots”) are used, as in example (12b).

Notice that in example (12c), *treba* occurs in a negated sentence. Here, negation takes semantic scope over necessity, which means the semantic composition of necessity and negation in relation to the proposition is “it is not *necessarily* the case that *P*” rather than “it is *necessarily* not the case that *P*”. Pragmatically, negation in this example is interpreted in a similar manner to the so-called *metalinguistic negation* (Martins, 2020), as the commenter merely objects to the specific number of volts, but still condones the violent action i.e., the electrocution of migrants.

The examples with *morati*, on the other hand, are significantly less lexically charged, as shown in (13), and the statements framed in a more indirect way.

- (13) a. Vse Evropske države bi *morale* bolj grobo udarit po migrantih.
“All European countries should *have to* strike back more strictly against migrants.”
b. Kdo nas zaščiti a *moramo* mi tudi nabavit pištolo.
“Who will protect us? Do we also *have to* buy a gun?”
c. Evropa bi *morala* stopiti skupaj hermetično zapreti meje.
“Europe should *have to* come together and hermetically close the borders.”

Even when the *morati* examples convey that it is necessary that some kind of action be taken against migrants, as in example (13a),

the verbs used are such that they no longer convey explicit violent acts, such as *postreliti* (“to shoot”), *zapliniti* (“to gas”), and *stresti* (“to electrocute”) in the *treba* examples (12), but express non-violent acts, as in the case of the verbal phrase *zapreti meje* “close the borders” in (13c). Indeed, the calls to violent action with *morati* are significantly more tentative, as many of the cases of deontic *morati* are embedded under the conditional mood clitic *bi*, which leads to a composite meaning where the deontic necessity is interpreted as a suggestion rather a direct command, as in examples (13a) and (13c), which is also not the case with *treba*.

To sum up the discussion so far, we have observed that while *treba* and *morati* both convey deontic necessity (roughly an obligation that needs to be met), they occur in quite substantially different statements in terms of hateful rhetoric in the case of the same type of unacceptable discourse, i.e., violent speech. Further, *morati* is also the only deontic modal which is less typical of violent speech than it is of offensive speech.

We suggest that the difference is tied to the way the pragmatics of deontic modals interact with their core syntactic and semantic properties. As discussed in Section 2.2, pragmatically deontic modals fulfil the interpersonal function in communication. The interpersonal dimension has to do with the fact that the deontic necessity, i.e., obligation, is ascribed by the speaker to whoever corresponds to the agent of the verbal event in the modalized proposition; concretely, in the case of example (13a), the speaker says that it is European countries that have the obligation to strike back against migrants.

The chief difference between the *treba* (12) and the *morati* (13) examples, manifested in the discussed lexical differences, lies in this interpersonal pragmatic dimension, which is crucially influenced by the syntax of the expressions. *Treba* is an impersonal predicative adjective which, in contrast to *morati*, syntactically precludes the use of a nominative grammatical subject that would be interpreted as the agent in the modalized proposition (Rossi and Zinken, 2016). Consequently, all the statements in the *treba* set of examples are such that the agent has an undefined, arbitrary reference – for instance, it is unclear who is expected to “gas the vermin” in example (12a). What happens

pragmatically is that the subjectless syntax of the adjective *treba* allows the speaker to sidestep the ascription of obligation to a specific agent, thus largely obviating what is perhaps the core interpersonal aspect of deontic modality. This cannot be really avoided with *morati*, which is a personal verb that obligatorily selects for a grammatical subject in active clauses – in other words, because of its personal syntax, *morati* presents a bigger interpersonal burden on the speaker, as he or she needs to specifically name the person or institution that is required to fulfil the obligation.

Note that in the violent subset there is only one example where *morati* is used with the verb *dobiti* (“get”), which induces a passive-like interpretation (14). Here, the grammatical subject headed by *vsak* (“everyone”) is interpreted as the target of the violent action rather than the agent. It is telling that this is also the only example with *morati* which is closer in the use of lexically charged items (i.e., being “shot in the head” rather than “the closing of borders” in the previous examples) to the *treba* examples, as this passive-like construction also precludes the use of an agentive noun phrase (unless it is introduced by the Slovenian equivalent of the *by*-phrase, but there are no such examples in the corpus).

- (14) [V]sak, ki se približa našim ženskam in otrokom, *mora* dobiti metek v čelo.
“Everyone who gets close to our women and children *must* be shot in the head.”

In short, the interpersonal structure influences the degree of hateful rhetoric, in the sense that speakers are more ready to use degrading metaphors, slurs and violent verbal expressions when they can avoid ascribing the obligation to someone specific. We follow Luukka and Markkanen (1997) by suggesting that impersonality has a similar hedging effect to epistemic modals, in the sense that the unexpressed agent in impersonals introduces a degree of semantic vagueness to the proposition, as does uncertainty brought about by the epistemic reading. Thus, with *treba* both deontic imposition and epistemic face-saving meet in one and the same lexeme.

6.1.2 Deontic likelihood

There is only one modal expressing deontic likelihood (i.e., a suggestion) in our set of modals; that is, *naj*_{IND}, which is by far the most frequent word under consideration (see Table 2). In this section, we discuss the possible reasons as to why this modal significantly favours unacceptable discourse (Table 4), and just like *treba* specifically favours the violent subtype of socially unacceptable discourse (Table 5).

Let's start by briefly presenting a lesser-known fact of the Slovenian modal system, which is that *naj*_{IND} has an additional semantic component to it that is otherwise not present in *morati* or *treba*. This component, which is linguistically unexpressed (and thus hard to detect empirically, especially in corpus data), has to do with the fact that when a speaker/writer uses *naj*_{IND}, they not only convey a suggestion, but also the fact that it is the speakers themselves who are also the source of deontic authority (Stegovec, 2019). To see this, compare the constructed examples in (15), which are minimally different from one another save for the following two facts. On the one hand, in the last example (15c), *morati* is used instead of *naj*_{IND}. On the other hand, the subject of the adversative *ampak* clause is coreferential in (15a) and (15c) with the null subject of the initial clause but not in (15b).

- (15) a. #Rekel je, da *naj* grejo stran, ampak noče, da grejo.
 I intended: “He said they *should* go away but he doesn't want them to.”
- b. Rekel je, da *naj* gredo stran, ampak nočem, da grejo.
 “He said they *should* go away but I don't want them to.”
- c. Rekel je, da *morajo* iti stran, ampak noče, da grejo.
 “He said that they *have to* go away but he doesn't want them to.”

(Examples (15a) and (15b) from Stegovec, 2019, p. 60)

Because *naj*_{IND} obligatorily involves speaker control, (15a) expresses a contradiction and is therefore semantically ill-formed (labelled with #). By contrast, (15b) is semantically coherent as the subjects of the two coordinated clauses are not coreferential. Lastly, (15b) also does not express a contradiction even under a coreferential

reading of the subjects of *reči* and *ne hoteti*, which shows that *morati* does not exhibit such subject-oriented control of the deontic suggestion.

For *FRENK*, we propose that the covert presence of speaker control explains an empirical gap in morphosyntactic agreement; that is, the fact that in all examples in the violent discourse *naj*_{IND} exclusively patterns with a verb in the third person, as in (16). This is not the case for other inflected modals in violent discourse, which do pattern with first person agreement, as shown in (17). Relatedly, the *naj*_{IND} clauses always contain an unexpressed subject that refers to a vague, arbitrary agent who is either a member of the in-group or part of a depersonalized collective, e.g. arbitrary migrants. In both cases, the speaker/writer, who is the deontic source, is excluded from this vague group that is expected to carry out what the speaker suggests.

- (16) a. *Naj* postavljajo snajperiste na mejo
“They *should* put snipers on the border.”
b. Kr pelte si jih domov pa vam *naj* posiljujejo žene.
“Take them home. They *should* rape your wives if you do.”
- (17) a. [A] *moramo* tud mi nabavit pištolo [...]
“Must we buy a gun?”
a. [...] da *dovolimo* tem pedercinam [...]
“... that we *allow* these faggots.”

Pragmatically, the covertness of speaker control in *naj*_{IND} acts as a hedge much like the impersonal syntax acts as a hedge for the obligatory meaning of *treba*, as discussed in the preceding section. This is also the reason why *naj*_{IND} favours unacceptable discourse, and specifically the violent subtype. Even though *naj*_{IND} is semantically weaker than either *treba* or *morati* in terms of modal force, this weakness is counterbalanced by the fact that speaker control is uniquely entailed. But because the entailment is not linguistically evident outside of constructed examples like (15a), it is also exempt from the surface discourse, in contrast to a first person verb like *dovolimo* “we allow” in (17b), which overtly spells out the deontic source. Consequently, in examples like (16) the speakers (or rather the authors of the Facebook comments, in this case) use *naj*_{IND}

to express a suggestion that is imposed exclusively on others, with the speakers themselves being the covert deontic source. This also seems to explain the exclusive patterning of *naj*_{IND} with third-person agreement, which semantically excludes the speaker from the denotation of the null subject pronoun.

6.1.3 Deontic possibility

To wrap up our discussion of deontic modals, we finally turn to the three possibility verbs – *smeti*, *dovoliti*, and *dopuščati*.

There is one major difference in this set. The verb *smeti* always combines with negation in the violent subset, as in (18a). This is not the case with *dovoliti* and *dopuščati*, which appears both in negated and non-negated uses, as is exemplified with *dovoliti* in (17b) (preceding section) and (18b).

- (18) a. Te horde *ne smemo* spustit v državo.
 “We must *not allow* this horde into our country.”
 b. [P]olitiki *ne bojo dovolili*, da se jim zgodi kaj hudega.
 “The politicians will *not allow* that anything horrible happens to them”.

In (18a) and (18b), negation is interpreted above the modal, which pragmatically yields a reading in which permission is not given to “release this horde into the country” in (18a) and “allow anything to happen to the politicians” in (18b). Notice that negating *smeti* or *dovoliti* results in a stronger reading than in the examples in which *treba* is negated, as in (12c) in Section 6.1.1, where negation-over-necessity results in a lack-of-obligation interpretation.

However, the fact that *smeti* always patterns with negation in the *FRENK* violent data is likely not due to pragmatic considerations, but to lexico-syntactic ones, especially since the interpretation of modality in the examples with negated *dovoliti* and *smeti*, as in (18a) and (18b), is the same. If this is correct, negated *smeti* fills in a gap in the Slovenian lexical inventory, where the modal that is used to grant permission (and like *smeti* and unlike *dovoliti* does not select for a finite clause) – that is, *lahko* – is syntactically a positive-polarity item (Marušič and Žaucer,

2016),¹² which means that its distribution is limited to non-negated contexts, so the closest grammatically related word – that is, *smeti* – is used in negated contexts in its stead. That this is the case is suggested by the fact that the almost exclusive patterning with negation is not limited to the violent subset or even to the larger unacceptable superset, but holds for all occurrences of the modal within the corpus. Only five (3%) out of the 150 occurrences of *smeti* appear in non-negated clauses in the entire corpus, as in (19), whereas *dovoliti* and *dopustiti* are almost equally distributed between negated and non-negated clauses.

- (19) [A]li se *smeta* dva istospolno usmerjena uradno poročiti
“Are two people of the same sex *allowed* to get married?”

6.2 Epistemic modals in offensive and acceptable discourse

Epistemic modals are slightly more frequent in acceptable comments, although the difference is not statistically robust, as shown in Section 5. In order to further explore the possible differences and similarities in the use of epistemic modals between different types of comments, we look at their distribution in three subcorpora, namely in acceptable, offensive and violent comments, and the distribution is shown in Table 6. We find that epistemic modals are very infrequent in the violent comments (even unattested for *morda* “possibly” and *najbrž* “likely”) in contrast to deontic modals, which are more frequent almost across the board in the violent set (Table 5). On the other hand, the epistemic modals show a similar distribution between acceptable and offensive comments in contrast to violent ones. In fact, the main difference can be observed at the level of modal force. It can be observed that probability modals (*verjetno*, *najbrž*) are the only ones more frequently used in the acceptable comments, while certainty and possibility modals (*ziher*, *zagotovo* and *morda*, *mogoče*) appear more often in offensive comments. In order to investigate the possible reasons for such differences, we now look at the communicative functions that are realized by the investigated modals.

12 The modal *lahko* is exempt from our study because unlike *smeti* its interpretation varies between epistemic and deontic modality in the corpus data.

Table 6: The distribution of epistemic modals between the acceptable, violent, and offensive subsets of FRENK

Modal	Acceptable	Violent	Offensive
	258.3	0.0	169.3
<i>mogoče</i>	312.1	113.8	555.8
<i>verjetno</i>	559.6	341.3	451.6
<i>najbrž</i>	129.1	0.0	90.3
<i>ziher</i>	86.0	113.8	97.3
<i>zagotovo</i>	32.3	113.8	83.4
Σ	1,377.4	682.7	1,447.5

Note. The frequencies are relative and normalized to a million tokens.

6.2.1 Epistemic necessity

In offensive comments, the epistemic necessity modals *ziher* and *zagotovo* generally act as prototypical boosting devices, which means that in terms of their illocutionary force (Searle, 1975) they are used to emphasize the commenter’s certainty in the assertion, as in example (20a). The meaning of certainty is also often emphasized stylistically, such as by the capitalization of the modal and the string of repeated exclamation marks in (20a).

- (20) a. Begunca? Ekonomske migrante pa picke, ki se ne znajo borit za svoj kos zemlje *ZIHER* ne!!!!!!
 “Accepting a refugee? *CERTAINLY* not accepting economic migrants and cunts who don’t know how to fight for their piece of land!!!!!!”
- b. Eni bi še radi denar od Slovenije, lahko dobite enosmerno karto za vašo vukojebino. Tam med ovcami se boste *zagotovo* počutili bolj domače
 “Some would even like to get money from Slovenia. You can get a one-way ticket back to your shithole. There, you will *certainly* feel more at home among the sheep.”

In lexically charged examples like (20a) and (20b), boosting also contributes to the positioning and legitimization of the commenter as an authoritative member of an in-group that is exclusionary of the

migrant out-group. In addition to the stylistically marked typography, such exclusiveness is emphasized by the contemptuous argumentation and explicitly through functional pronominal elements that are involved in the process of Othering by excluding the commenter and their associates, such as *vašo vukojebino* “your shithole” in (20b). Hyland (1998b), building on previous work by He (1993) and Myers (1989), claims that such authoritative use of boosters strategically establishes solidarity between the in-group members by making it seem as though the truth were arrived at by consensus, which makes it less easy to dispute by the members of the out-group.

In the acceptable comments, the two epistemic necessity modals often occur in ironic statements. This is shown by example (21) with the necessity adverb *ziher*, where the ironic interpretation is conveyed not only by the modal itself but also by the use of the intensifying adverb *itak* (“of course”), exaggeration by means of the collective reading of the plural pronoun *vsi* (“everyone”), the use of the verb in the first-person *dejmo* (“let’s”), and the use of the emoticon at the end.

- (21) *Itak, dejmo vsi lagat, to je ziher prav :)*
“Of course, let’s all lie, that’s *certainly* the right thing to do :)”

Irony, being a sophisticated rhetorical device in terms of discourse stylistics, enhances the persuasive effect of the proposition (Gibbs and Izett, 2005). It allows the speaker to adopt an attitude of emotional composure or detachment (Attardo, 2000), which helps the commenter to save face by superficially positioning them as an authority and masking the deficits of the argumentation, thus making the comment less disputable and the commenter less exposed to potential criticism. In addition, irony works as a hedge by attenuating the direct criticism conveyed in the comment. This allows the speaker not only to protect their face but also the face of the target, since ironic criticism is accepted better or in a friendlier way than direct critiques (Gibbs and Izett, 2005). Brown and Levinson (1987, p. 212) also note that “even fairly blatant indirectnesses [such as irony] may be defensible as innocent”, as it enables the commenter to save face by distancing themselves from the proposition.

Although the boosting function of epistemic necessity modals predominates in the corpus data, there are few examples in which such modals are used as hedging devices even under a non-ironic reading. A case in point is example (22) from the set of offensive comments.

- (22) [K]r k cerarju nej gredo *zihr* ma veliko stanovanje ... bedaki.
“They better go to the prime minister Cerar, he *surely* has a big flat ... assholes.”

The modal in (22) hedges the propositional content by invoking the presumed shared knowledge of the in-group, which concerns the size of the prime minister’s home. Here, the modal works as a face-saving device because it protects the speaker from the accusation of making an unfounded claim, as the modalized statement, despite entailing certainty, is still weaker than the unmodalized variant which would otherwise report that the speaker holds factual knowledge about the size of the prime minister’s apartment.

6.2.2 Epistemic likelihood

Likelihood modals predominantly convey the epistemic meaning of low certainty without encoding additional communicative meanings. Such neutral usage is exemplified in (23), where the commenter uses *verjetno* to express that, according to their knowledge, it is likely (but not certain) that the addressee does not have the relevant evidence.

- (23) O kredibilnosti posnetkov po katerih sodiš *verjetno* nimaš dokazov.
“You *likely* have no evidence to support the credibility of the recordings on which you base your judgments.”

Occasionally, the likelihood modals fulfil additional communicative functions. For instance, the sentence in (24) uses a simile to communicate an offensive comparison to a thorny plant, which is attenuated with the likelihood modal. The negative representation of the target is therefore less categorical, which helps to preserve face both on the side of the commenter and the addressee.

(24) [N]a sliki imaš lep cvetoč travnik[,] ti si pa *verjetno*
en blesav osat.

“The picture shows a nice blooming meadow, and you
are *probably* one stupid thistle.”

As in the case of epistemic necessity modals (see Section 6.2.1), there is frequent use of irony in the acceptable comments. For instance, example (25) has an ironic interpretation, as the commenter is facetiously suggesting that the addressee ask the reporters for concrete evidence on whatever it is that they are reporting on.

(25) Fajn, potem pa jim reci naj dokažejo, da se je to kar pišejo,
res zgodilo v Ljubljani, ker *najbrž* imajo dokaze, če so napisali
kar pač so.

“Great, then tell them to prove that this what they are
writing about really happened in Ljubljana, because they
probably have evidence if they wrote what they did.”

Irony plays two roles here, both contributing to the face-saving dimension of the discourse. On the one hand, the use of irony positions the commenter as an apparent authoritative source of knowledge on the quality of the news media outlet which helps protect the commenter against potential criticism (see also example (21)). On the other hand, the irony, due to its non-literal meaning, allows for defensibility on part of the speaker (Brown and Levinson, 1987). Since the modal is part of a larger ironic context, its primarily communicative function is not to convey knowledge about the proposition; rather, it simply contributes to the overall attenuating effect of the complex, non-literal, communicative strategy used by the commenter.

6.2.3 *Epistemic possibility*

Epistemic possibility modals are used in offensive discourse to hedge impoliteness, as shown by example (26), which contains the epistemic adverb *morda*.

(26) Vi ste očitni preobremenjeni z nestrpnostjo in nehumanostjo.
Jih *morda* vi pedenate?

“You are certainly preoccupied with intolerance and inhumanity. Is it *maybe* you who are looking after them?”

The second sentence in the example contains a question, but its communicative role is not to present an earnest query to the addressee, but rather functions as a rhetorical device. This is emphasized by the lexemes with strong negative connotations (e.g., *nehumanostjo* “inhumanity”) and a certainty-denoting adverb (e.g., *očitno* “clearly”) in the preceding sentence, which stands in stark contrast to the uncertainty that would be conveyed through questions and epistemic modal in a neutral, non-ironic context. What is happening here pragmatically is that the modal helps save the commenter’s face by allowing them to appear less assertive, as the possibility semantics brought about by *morda* lessens the impoliteness of the question by way of (superficial) tentativeness.

A unique characteristic of possibility expressions is that they are also used to convey the so-called concessive interpretation, which is shown in (27) with *mogoče*.

- (27) *Mogoče* sem malo staromodn, ampak tistemu, ki nekomu vzame življenje, ga je treba vzeti tudi njemu!
“*Maybe* I’m a bit old-fashioned, but whoever that takes somebody’s life should also lose theirs!”

There is a debate in the semantic literature as to whether such concessive uses of epistemic expressions even constitute modality at all, or if they instead play some other discursive role. Palmer (2014, p. 31), for instance, claims that they do not, as “the speaker does not indicate doubt about the proposition, but rather accepts it as true, in order to contrast one state of affairs with another”. By contrast, Baranzini and Mari (2019, p. 120) claim that “concessivity is to be understood at the discourse level and not as a meaning of the modal itself”, and that the concessive reading arises from the underlying epistemic modal meaning.

In concessive examples with *mogoče* like (27), we believe that modality is still involved in the semantics. What the commenter concedes is not necessarily that a certain state of affairs holds in the actual world; instead, they use the concessive clause to indicate that they allow for

the possibility of their being old-fashioned.¹³ From the perspective of the ongoing discourse, the concessive clause does not add anything to the at-issue meaning, but rather only adds to the interlocutors' conversational common ground (Green, 2017). In relation to face-saving, the concessive clause is used by the commenter to rhetorically agree with a possibility tied to the knowledge or belief state of the addressee, thereby establishing communicative rapport between them.

7 Conclusion

This paper has presented a corpus investigation of epistemic and deontic modal expressions in Slovenian Facebook comments in the *FRENK* corpus.

We have first proposed a set of Slovenian modals that show an overwhelming tendency towards a single modal reading. Because of such unambiguity, they constitute a robust set that allows for precise quantitative comparisons between different types of discourse without irrelevant confounding examples and for careful manual analysis of the corpus examples. Quantitatively, we have shown that deontic modals are a prominent feature of unacceptable discourse, and that they are especially prominent in discourse that concerns incitement to violent action, which is legally prosecutable.

In terms of discourse pragmatics, we have first shown that two deontic necessity modals, which are completely synonymous both in terms of force and modality type, nevertheless differ profoundly in the degree of hateful rhetoric in the same type of socially unacceptable discourse. We have shown that what makes a difference in such examples is the presence of impersonal syntax, which offers speakers the ability to linguistically obviate the ascription of the denoted obligation to a particular agent. We have suggested that this sort of face-saving strategy of ambiguity by way of impersonality correlates with the speaker's tendency to use dehumanizing language, such as slurs or degrading metaphors.

13 That the speaker does not concede that something is a fact is shown by the admissibility of the parenthetical clause in the constructed example in (1).

(1) *Mogoče sem staromodni (kdo bi vedel), ampak ...*
“*Maybe I’m old-fashioned (who knows), but ...*”

For deontic likelihood, we have claimed that the adverb/particle *naj* is well-suited for violent discourse because it implicitly conveys speaker control, and have proposed that this explains why *naj* exclusively occurs in comments in which the speaker expects others to carry out what is being suggested. For possibility modals, we have explored their interaction with negation, and showed how it leads to a stronger reading that involves the denial of permission.

Lastly, we have discussed the fact that epistemic necessity and likelihood modals do not only express truth commitment but also acquire additional communicative functions related to the face-saving strategy. In acceptable comments they often help signal irony, which pragmatically acts as a hedge. In offensive comments, they are used as boosters whereby they help position the speaker as an authoritative figure. Epistemic possibility modals also typically act as face-protecting hedges rather than just expressions of uncertainty, even when they occur in concessive clauses.

Acknowledgments

We would like to thank all three anonymous reviewers for their helpful comments and suggestions. The work described in this paper was funded by the Slovenian Research Agency research programme *P6-0436: Digital Humanities: resources, tools and methods* (2022–2027), the DARIAH-SI research infrastructure, and the national research projects *Z6-4616: Slovenian Verbal Valency: Syntax, Semantics, and Usage* (for author 1) and *N6-0099: LiLaH: Linguistic Landscape of Hate Speech* (for author 2).

References

- Assimakopoulos, S., Baider, F. H., & Millar, S. (2017). *Online hate speech in the European Union: a discourse-analytic perspective*. Cham: Springer Nature.
- Attardo, S. (2000). Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *Rask*, 12(1), 3–20.
- Ayuningtias, D. I., Purwati, O., & Retnaningdyah, P. (2021). The Lexicogrammar of Hate Speech. In Y. Wirza et al. (Eds.), *Proceedings of Thirteenth Conference on Applied Linguistics (CONAPLIN 2020)* (pp. 114–120). Indonesia: Atlantis Press.

- Baranzini, L., & Mari, A. (2019). From epistemic modality to concessivity: Alternatives and pragmatic reasoning per absurdum. *Journal of Pragmatics*, 142, 116–138.
- Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage (Vol. 4)*. Cambridge: Cambridge University Press.
- Chiluwa, I. (2015). Radicalist discourse: a study of the stances of Nigeria's Boko Haram and Somalia's Al Shabaab on Twitter. *Journal of Multicultural Discourses*, 10(2), 214–235.
- Coates, J. (1983). *The Semantics of the Modal Auxiliaries*. London and Canberra: Croom Helm.
- Coates, J. (1987). Epistemic modality and spoken discourse. *Transactions of the Philological Society*, 85(1): 110–131.
- Cvrček, V. (2021). *Calc 1.03: Corpus Calculator*. <https://www.korpus.cz/calc/>. Last accessed: 20. 12. 2022.
- Delgado, R. (2019). *Understanding words that wound*. New York: Routledge.
- Evkoski, B., Pelicon, A., Mozetič, I., Ljubešić, N., & Kralj Novak, P. (2022). Retweet communities reveal the main sources of hate speech. *PLoS one*, 17(3), e0265602.
- 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.
- Fišer, D., Erjavec, T., & Ljubešić, N. (2017): Legal framework, dataset and annotation schema for socially unacceptable online discourse practices in Slovene. In Z. Waseem et al. (Eds.), *Proceedings of the first workshop on abusive language online* (pp. 46–51). Vancouver: Association for Computational Linguistics.
- Gelber, K., & McNamara, L. (2016). Evidencing the harms of hate speech. *Social Identities*, 22(3), 324–341.
- Gibbs Jr, R. W., & Izett, C. D. (2004). Irony as persuasive communication. In H. L. Colston and A. N. Katz (eds.): *Figurative Language Comprehension* (pp. 143–164). New York: Routledge.
- González García, F. (2000). Modulating grammar through modality: a discourse approach. *ELIA*, 1, 119–136.
- Green, M. (2017). Conversation and common ground. *Philosophical Studies*, 174(6), 1587–1604.

- Halliday, M. A. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of language*, 6(3), 322–361.
- He, A. W. (1993). Exploring modality in institutional interactions: Cases from academic counselling encounters. *Text-Interdisciplinary Journal for the Study of Discourse*, 13(4), 503–528.
- Hyland, K. (1998a). *Hedging in scientific research articles*. Amsterdam: John Benjamins Publishing Company.
- Hyland, K. (1998b). Boosting, hedging and the negotiation of academic knowledge. *Text & Talk*, 18(3), 349–382.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, 7(2), 173–192.
- Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives* (Vol. 36). Oxford: Oxford University Press.
- Krek, S. et al. (2019). *Corpus of Written Standard Slovene Gigafida 2.0*. <http://hdl.handle.net/11356/1320>. Slovenian language resource repository CLARIN.SI.
- Lenardič, J., & Fišer, D. (2021). Hedging modal adverbs in Slovenian academic discourse. *Slovenščina 2.0: empirical, applied and interdisciplinary research*, 9(1), 145–180.
- Lenardič, J., & Pahor de Maiti, K. (2022). Slovenian Epistemic and Deontic Modals in Socially Unacceptable Discourse Online. In D. Fišer & T. Erjavec (Eds.), *Proceedings of the Conference on Language Technologies and Digital Humanities* (pp. 108–116). Ljubljana: Institute of Contemporary History.
- Ljubešič, N., Fišer, D., & Erjavec, T. (2019). The FRENK datasets of socially unacceptable discourse in Slovene and English. In K. Ekštejn (Ed.), *International conference Text, Speech, and Dialogue. TSD 2019. Lecture Notes in Computer Science*, v. 11697 (pp. 103–114). Cham: Springer.
- Ljubešič, N., Fišer, D., Erjavec, T., & Šulc, A. (2021). *Offensive language dataset of Croatian, English and Slovenian comments FRENK 1.1*. <http://hdl.handle.net/11356/1462>. Slovenian language resource repository CLARIN.SI.
- Lorenzi-Bailly, N., & Guellouz, M. (2019). Homophobie et discours de haine dissimulée sur Twitter: celui qui voulait une poupée pour Noël. *Semen. Revue de sémio-linguistique des textes et discours*, 47. doi: 10.4000/semn.12344

- Luukka, M. R., & Markkanen, R. (1997). Impersonalization as a form of hedging. In R. Markkanen and H. Schröder (Eds.), *Hedging and Discourse, Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts* (pp. 168–187). Berlin: de Gruyter.
- Martins, A. (2020). Metalinguistic negation. In V. Déprez & M. T. Espinal (Eds.), *The Oxford Handbook of Negation*. Oxford: Oxford University Press.
- Marušič, F. L., & Žaucer, R. (2016). The modal cycle vs. negation in Slovenian. In F. L. Marušič & R. Žaucer: *Formal Studies in Slovenian Syntax* (pp. 167–192). Amsterdam: John Benjamins Publishing Company.
- Močnik, M. (2019). Slovenian ‘dopuščati’ and the semantics of epistemic modals. In *Proceedings of 27th Formal Approaches to Slavic Linguistics meeting (FASL 27)*. Michigan: Michigan Slavic Publications.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied linguistics*, 10(1), 1–35.
- Pahor de Maiti, K., Fišer, D., & Ljubešić, N. (2019). How haters write: analysis of nonstandard language in online hate speech. In J. Longhi and C. Marinica: *Proceedings of the 7th Conference on CMC and Social Media Corpora for the Humanities (CMC-Corpora2019)*. Paris: CLARIN K-Center for CMC.
- Palmer, F. R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.
- Palmer, F. R. (2014). *Modality and the English modals*. Oxon: Routledge.
- Roeder, C. F., & Hansen, B. (2006). Modals in contemporary Slovene. *Wiener Slavistisches Jahrbuch*, 52, 153–170.
- Rossi, G., & Zinken, J. (2016). Grammar and social agency: The pragmatics of impersonal deontic statements. *Language*, 92(4), e296–e325.
- Searle, J. R. (1975). A taxonomy of illocutionary acts. *Language, mind, and knowledge*, 7, 344–369.
- Siegel, A. A. (2020). Online hate speech. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy: The state of the field, prospects for reform* (pp. 56–88). Cambridge: Cambridge University Press.
- Sindoni, M. G. (2018). Direct hate speech vs. indirect fear speech. A multimodal critical discourse analysis of the Sun’s editorial “1 in 5 Brit Muslims’ sympathy for jihadis”. *Lingue e Linguaggi*, 28, 267–292.
- Stegovec, A. (2019). Perspectival control and obviation in directive clauses. *Natural Language Semantics*, 27(1), 47–94.
- von Fintel, K. (2006). Modality and Language. In D. M. Borchert (Ed.), *Encyclopedia of Philosophy – Second Edition* (pp. 20–27). Detroit: MacMillan Reference USA.

- Vukovic, M. (2014). Strong epistemic modality in parliamentary discourse. *Open Linguistics*, 1(1): 37–52.
- Winter, S., & Gärdenfors, P. (1995). Linguistic modality as expressions of social power. *Nordic Journal of Linguistics*, 18(2), 137–165.

Slovnični in pragmatični vidiki naklonskosti v slovenščini v družbeno nesprejemljivih komentarjih na Facebooku

V članku predstavimo slovnično in pragmatično rabo epistemskih in deontskih naklonskih izrazov v korpusu družbeno sprejemljivih in nesprejemljivih komentarjev v slovenščini, ki so bili objavljeni na platformi Facebook. Za potrebe analize oblikujemo seznam naklonskih izrazov, ki pomensko pripadajo zgolj eni vrsti naklonskosti, kar nam omogoča učinkovite in točne korpusne poizvedbe in zanesljivo interpretacijo kvantitativnih rezultatov. V članku pokažemo, da so deontski naklonski izrazi, ne pa tudi epistemski, statistično značilno bolj pogosti v družbeno nesprejemljivih komentarjih, pri čemer še posebej izstopajo v komentarjih z nasilno vsebino. Kvantitativne izsledke nadgradimo s kvalitativno analizo diskurzivne vloge naklonskih izrazov. V kvalitativnem delu tako raziščemo, kako se pragmatične sporazumevalne strategije, med njimi pragmatično omejevanje in ojačevanje pomena propozicije ter blaženje potencialne grožnje posameznikovi integriteti, sklapljajo s temeljnimi skladenjskimi in pomenoslovnimi značilnostmi naklonskih izrazov, na primer njihovo naklonsko stopnjo in stavčno skladnjo.

Ključne besede: korpusno jezikoslovje, naklonskost, skladnja, pomenoslovje, pragmatika, sovražni govor