1. INTRODUCTION

Academic discourse continues to be one of the popular research areas in applied linguistics with cross-disciplinary language variation being “now one of the more fruitful lines of research” (Hyland 2011: 178). Argumentation and rhetorical choices, lexis and grammatical constructions have been studied extensively in an attempt to define characteristic features of specific disciplines and specific discourses. The differences and similarities in the expression of author stance in academic discourse have become one of the main topics in a rapidly growing body of research on both national and disciplinary academic identity.

The importance of studies of academic discourse has been emphasized by Becher & Trowler (2001: 46), who claim that “a detailed analysis of disciplinary discourse … can help not only to bring out characteristic cultural features of disciplines but also to highlight various aspects of the knowledge domains to which they relate”. The identified specific patterns typical of different science fields can contribute to the general understanding of how scientific knowledge is created and reported.

There are numerous studies which concentrate on the clearly contrasting hard and soft disciplines, the hard ones being the sciences and soft ones the humanities and social sciences. The humanities traditionally include fields such as the arts, theatre studies, philosophy, theology alongside architecture, history and linguistics while social sciences usually include fields such as economics, law, sociology, management and psychology. The range of subdisciplines that make up the humanities and social sciences suggests

---

1 Earlier versions of this article were presented at the 45th Meeting of the Societas Linguistica Europaea in Stockholm in August 2012 and the 11th Conference of the European Society for the Study of English in Istanbul in September 2012. Thanks are due to the European Union Structural Funds project “Postdoctoral Fellowship Implementation in Lithuania”, which funded Šinkūnienė’s research for this article, to the North-West University Potchefstroom, who funded Van Olmen’s research, and to Juana I. Marín-Arrese and Aurelija Usonienė, organizers of the seminar “Evidentiality and Epistemic Modality in Academic and Journalistic Discourse: Cross-linguistic Perspectives” at the ESSE conference in Istanbul.
that one can expect there to be considerable variation within the two and that one can obtain interesting results by analyzing and comparing not only very different disciplines but also sister disciplines. It is safe to say that biomedical discourse differs from that of the humanities but how (dis)similar are the soft disciplines themselves?

Although most research focuses on academic texts written in English, recent decades have seen an increase in academic discourse analysis of other languages. Such studies try to define specific features of specific cultural communities, frequently in contrast to Anglo-American scientific writing tradition. These cross-linguistic studies complement the cross-disciplinary ones in that they highlight universal features of academic discourse as well as point out specific features, typical of only some disciplines or some cultures. One question frequently raised by these studies has to do with the dominating factors shaping academic texts: is it the big (or national) culture or the small (disciplinary) culture that determines certain patterns of argumentation and author stance expression (cf. Atkinson 2004)? On a broader scale, mapping out argumentation patterns of different discourse communities not only helps to define their national academic identity features but also highlights those that are discipline-specific (cf. Fløttum et al. 2006, Lorés-Sanz 2011).

Empirical studies have shown that, notwithstanding disciplinary and national idiosyncrasies, all academic discourse is dialogic in nature. The construction of a scientific text has to be regarded as a dialogue between the author and the reader (Hyland 2008: 5), in which the projection of the author’s stance is of great significance. With the help of a range of linguistic items, the author communicates to the reader his or her attitude toward the claims that are made, by evaluating them, by strengthening or weakening them and by showing greater or smaller commitment. This makes epistemic modality central to academic writing (Hyland 2001: 291).

So it is not surprising that many studies on modality in academic discourse have concentrated on epistemic modality, i.e. the expression of the speaker’s assessment of the likelihood of the state of affairs (Palmer 2001: 7), as it clearly links with pragmatic phenomena such as hedging. There are fewer empirical studies that look at patterns of non-epistemic modality, i.e. the specification of conditions on the agent of the event (Palmer 2001:8). When these conditions are inherent to the subject (e.g. ability, internal necessity) or circumstantial, Palmer (2001: 10) uses the term „dynamic“ and provides the following examples:

(1) John can speak French. (ability)

(2) John will do it for you. (willingness)

When they have to do with permission, obligation, desirability, appropriateness and the like, the term „deontic“ is used. Such modal meanings can be illustrated by (3) and (4) from Palmer (2001: 10):

(3) and (4) from Palmer (2001: 10):
Modal verbs in English, with their multifunctional profile, are of special interest in academic discourse as they allow the author to strengthen or weaken his or her position.

An interesting marker in that respect is *must*. In its non-epistemic or “root” (Coates 1983: 10) meaning, *must* conveys a strong position on the part of the author, a position that does not accept alternative views. Epistemic *must* communicates strong certainty in the accompanying proposition. However, propositions modified by *must* are still more tentative than unmodified ones, which thus allows for the interpretation of this *must* as a hedge.

What is more, the status of *must* – and its cognates in Dutch and German – as purely epistemic is open to discussion. Another linguistic category within the context of which *must* is discussed is the category of evidentiality. Evidentiality is generally defined as “a linguistic category whose primary meaning is source of information” (Aikhenvald 2004: 3). Traditionally, evidentiality is divided into direct and indirect. Direct evidentiality refers to cases when the speaker has witnessed the situation himself or herself. Indirect evidentiality refers to either the speaker’s logical conclusion about the situation based on certain evidence available to him or her (the so-called inference) or on the information provided by others (the so-called reportative meaning).

For *must*, De Haan (2001) claims, on the basis of just one example, that it is purely epistemic (unlike Dutch *moeten*, for instance) but Nuyts (2001) and Cornillie (2009) point out that there is also an evidential aspect to its meaning. More specifically, the latter argue that the speaker’s judgment is based on a process of reasoning or, in other words, that *must* is inferential – like *moeten*, which has a purely reportative meaning as well. Mortelmans (2010) agrees and adds that the epistemic and evidential aspects of the modal verbs of necessity in Germanic are hard, if not impossible, to untangle (cf. van der Auwera & Plungian 1998: 86 too, who argue that the modal domain overlaps with the evidential domain and, more specifically, its inferential part in the area of epistemic necessity). To account for the (dis)similarities between English, Dutch and German, she proposes to follow Squartini’s (2008) distinction between types of inference: the process of reasoning can be based on situational/external evidence, on general knowledge or on conjecture.

This study attempts to describe the role of *must*, its Dutch cognate *moeten* and the Lithuanian necessity modal *turėti* ‘have to’ as well as to compare their frequency and distribution in two sister disciplines of academic discourse, i.e. the humanities and the social sciences. We believe that the genre of academic texts, in which evidence plays an important part, will shed new light on the debate about the epistemic-evidential
character of necessity modals and particularly on the types of inference with which the three verbs under examination can be associated.

2. CORPORA

The study is based on three different corpora: COCA or the Corpus of Contemporary American English (cf. http://corpus.byu.edu/coca), CAD or the Corpus of Academic Dutch (compiled for the present project) and CorALit or the Corpus Academicum Lithuanicum (cf. http://coralit.lt). Table 1 provides, for each of these corpora (or, rather, the parts used here), the total size and the size of the humanities and social sciences subcorpora.

Table 1: size of COCA, CAD and CorALit

<table>
<thead>
<tr>
<th>Discipline</th>
<th>COCA</th>
<th>CAD</th>
<th>CorALit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanities</td>
<td>11,311,012 w</td>
<td>440,557 w</td>
<td>2,028,906 w</td>
</tr>
<tr>
<td>Social sciences</td>
<td>8,518,419 w</td>
<td>458,046 w</td>
<td>1,480,548 w</td>
</tr>
<tr>
<td>Total</td>
<td>19,859,431 w</td>
<td>898,603 w</td>
<td>3,509,454 w</td>
</tr>
</tbody>
</table>

COCA is an online corpus of more than 450 million words of spoken, fiction, popular magazine, newspaper and academic journal articles in American English from 1990 to now. The humanities subcorpus includes texts from the arts, cultural studies and literary science while the social sciences subcorpus contains texts from psychology, education and sociology. CorALit\(^2\) is a large specialized, synchronic corpus of written academic Lithuanian from 2000 to 2009 covering five disciplines (i.e. the humanities, the social sciences, the biomedical sciences, technology and the physical sciences). The size of the corpus is roughly nine million words. The humanities subcorpus includes a wide range of academic genres such as textbooks, monographs and journal articles (the latter making up the bulk of the data) from the arts, cultural studies, philosophy, linguistics, history, theology, literary science and library studies. The social sciences subcorpus exhibits the same variation in genres and contains texts from law, political science, economics, psychology, education and management. Because of the differences with COCA, i.e. more genres, shorter time span but especially more subdisciplines, we realize that our comparison of English and Lithuanian will be far from perfect. This is not the case for Lithuanian and Dutch, as far as the variety of subdisciplines and the time span are concerned. The humanities subcorpus in CAD is made up of sixty journal articles, which have all been written by different authors between 2000 and the present. Its subdisciplines range from the arts and cultural studies over philosophy, linguistics and history to literary science and library studies.

\(^2\) The detailed information on the CorALit compilation and design are provided in Usonienė et al. (2011).
The social sciences subcorpus in CAD is a similar collection of journal articles from law, political science, economics, psychology, education, management and sociology.

3. ENGLISH

3.1. FREQUENCY

Table 2 first gives the total number of occurrences of verbal must in COCA. On the basis of a sample of 500 cases per discipline, the percentage of irrelevant occurrences in quotes, titles and the like was calculated and detracted from the first figures. The estimated number of relevant cases is provided in the second column. The third column, finally, presents the number of relevant occurrences of must per 10,000 words. The data show that must is slightly more frequent in the social sciences than in the humanities.

Table 2: frequency of must in COCA

<table>
<thead>
<tr>
<th>Discipline</th>
<th># total</th>
<th># relevant</th>
<th>10,000 w</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanities</td>
<td>8,100</td>
<td>7,760</td>
<td>6.86</td>
</tr>
<tr>
<td>Social sciences</td>
<td>7,685</td>
<td>7,408</td>
<td>8.70</td>
</tr>
<tr>
<td>Total</td>
<td>15,785</td>
<td>15,168</td>
<td>7.64</td>
</tr>
</tbody>
</table>

For the rest of the analysis, we use the initial samples of 500 occurrences, excluding the irrelevant cases and including an additional thirty-nine cases as a compensation. The distribution of meanings found in these samples is presented in Table 3, which gives –from left to right– the raw frequency, the percentages and the estimated frequency per 10,000 words (which is calculated as follows: the total number of relevant cases in Table 2 that can be assumed to have a certain meaning on the basis of the percentages in Table 3, divided by the total number of words in the subcorpus mentioned in Table 1 and multiplied by 10,000). Note that “non-root” is used to refer to must as an epistemic-evidential verb. It is clear from the figures that, in spite of the difference in frequency between the humanities and the social sciences, must functions in the same way in both disciplines.

Table 3: distribution of meanings of must in COCA

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Humanities</th>
<th></th>
<th>Social sciences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># raw</td>
<td>%</td>
<td>10,000 w</td>
<td># raw</td>
</tr>
<tr>
<td>Root</td>
<td>464</td>
<td>92.80</td>
<td>6.37</td>
<td>464</td>
</tr>
<tr>
<td>Non-root</td>
<td>36</td>
<td>7.20</td>
<td>0.49</td>
<td>36</td>
</tr>
</tbody>
</table>

Our results are in line with previous research, which has shown that non-root must
does not occur very frequently in written English but rather is a feature of spoken language. In De Haan’s (2008) analysis of the Brown Corpus, only 16% of the attestations of must are non-root. In the same vein, Coates (1983: 48) observes: “The fact is that in written language (…) and in language written to be spoken (V) and in formal spoken language (T), Root MUST occurs more frequently than Epistemic MUST. But in informal spoken language, that is, normal everyday adult conversation, the inverse is true: Epistemic MUST is preponderant.” Studies on hedging in English academic texts also note the tendency for the non-root must to be infrequent. Hyland (1998: 108) reports only six cases of epistemic must acting as a hedge in his corpus of research articles of 75,000 words. Similarly, Varttala (2001: 115) mentions only eight instances of must with a hedging function in his analysis of a corpus of research articles of roughly 175,000 words representing three disciplines.

Must is often analyzed just in terms of the contrast root modality (necessity and obligation) versus epistemic modality (logical necessity or confident inference) (e.g. Coates 1983). But scholars like Palmer (2001) and Collins (2009) make the additional distinction between dynamic root modality and deontic root modality (though the difference is not always so clear, cf. Palmer 1990: 170). The latter, for instance, notes that, in his comparison of the modal verbs in the Australian and British components of the International Corpus of English and in an American English counterpart compiled by himself, the dynamic cases account for only 6.3% of all instances of must. A look at 100 root cases in the humanities and the same number of root cases in the social sciences confirms this observation. In both disciplines, just ten out of 100 root attestations are dynamic. Sentence (5) is a case in point.

(5) He told me that, out of decency, he omitted one detail, namely that the male familiar is believed to have an extraordinarily long penis that he must carry on his shoulders to be able to walk. (H65)

3.2. CORRELATIONS AND DISCOURSE FUNCTIONS

Let us first draw attention to the fact that, though third person subjects are by far the most common type, second and especially first person pronouns are not infrequent in the English data, especially in comparison with Lithuanian academic discourse, in which first and second person forms of turėti are very rare (see Section 5). In both the humanities and the social sciences, roughly 10% of the subjects are first person pronouns and, in all cases but one, must has a root meaning. Quite frequently, we is exclusive in the sense that it does not refer to people in general, but to a specific discourse community to which the author(s) belong such as teachers, writers or educators. This use of we in combination with root must creates a strengthened sense of
community: the modal verb serves as a call for action or as an indication of in-group values or of some code of conduct. Consider examples (6) to (8).

(6) If we want our students to be successful, we must show them that practicing can be a valuable, positive experience. (H358)

(7) As teachers, we must understand that our audiences will not condone excuses for weak or underprepared concert performances. (H225)

(8) As writers, we can’t become paralyzed at the thought of rejection. We can’t fear it, or seek to avoid it. Rather, we must confront it head on, charge into it with reckless abandon. We must look at rejection like a ball player looks at striking out, that thin line between trying and succeeding, a line we must cross as many times as necessary. (H141)

We can be used inclusively too, i.e. it can refer to people in general. In this case, the combination with must contributes to the persuasiveness of the text. Sentence (9) can serve as an example.

(9) But until Western political leaders recognize that Somali outlaws are not merely an annoyance but a deadly peril to international law and stability that must be defeated at all costs, we must expect that the toll of blood and treasure exacted by this new breed of pirate will continue to grow. (H33)

Perhaps somewhat surprisingly, must also co-occurs with the second person pronoun, especially in the humanities, in which you accounts for 8% of the occurrences. Example (10) is indicative of these cases.

(10) For diagnostic assessment, you need to be mindful of the sequence of skills you are building in your students. If you are planning to teach a particular skill, you must first check on the skills the students must have mastered in order to learn the new one. (H374)

The second person pronoun combines with root must and is used to give advice to an unidentified reader. The writer assumes a position of authority in fields such as writing and music, about which the reader is thought to want to know more.

Previous research has shown that modal meanings correlate with syntactic features. One of the strongest correlations is that between non-root must and the perfect (cf. Coates 1983 and Wärnsby 2003), as in (11). In the humanities, thirteen of the fifteen cases with a perfect are non-root. One of the two exceptions is (12). The picture for the social sciences is similar.

(11) However, especially from a distance, their massive appearance conjures up the might of the Roman Empire, so the facades must have deterred nomadic attackers while the forts performed their real function -- administration of the empire’s southern limits. (SS184)
Anyone who knows anything about classical African art as it is constructed for market purposes knows this: more than any other type of object, an African mask, to be “authentic” by the standards of connoisseurs, must have been used, worn in the context of performances of a ritual nature. (H488)

Another feature traditionally associated with non-root must is existential there, as in (13). In three of the eight cases in the humanities, however, must expresses root modality. Sentence (14) is a case in point. In the social sciences, one of the two cases has a root meaning.

Haven’t they already suffered enough from colonization, war, droughts, economic decline, and a totalitarian regime? Why is AIDS added to that long list? There must be someone out there plotting against them and enjoying their pain! (H62)

Writing is an exhausting, consuming, draining business, and after a stretch of constant giving, there must be a restoration period. (H458)

To conclude, Palmer (1990: 73-74) notes that must quite frequently combines with certain verbs denoting „the act of conversation“ such as say, admit, confess and mention and when the speaker uses these, he or she is actually performing the act „I must admit = I do admit“. Consider example (15).

I must acknowledge the value that Maria Elisa Christie and Clarissa Kimber added through their conscientious and thoughtful reviews. (SS100)

It is rather remarkable that this type of collocation is very rare in the data: (15) is the only case in the social sciences and the humanities too contain just a few cases resembling it, typically when must combines with the passive, as in (16).

The new letters, it must be said, do not add much. (H194)

Such expressions serve as a helpful tool to guide the reader through the text and to attract his or her attention. Generally, root must combines with the passive in approximately 20% of the cases in the humanities and in roughly 30% of the cases in the social sciences. One specific construction which is of interest here is exemplified in (17).

This article examines some of these beginning efforts and provides a close examination of international adoption as one small but plausible option. It must be clearly understood that adoption is not being discussed as the only or even a priority option. (SS8)

Just like in (16), the author employs must and the passive for expository purposes, i.e. to emphasize some important piece of information and to manage the reader’s interpretation of the text. Somewhat surprisingly, this use does not occur very often in the English data.
3.3. EPISTEMICITY AND EVIDENTIALITY

As mentioned before, the epistemic and inferential meanings of the Germanic modals of necessity are hard, if not impossible, to separate. The inferences with which must can be used range from situational, i.e. based on perceivable evidence, over generic, i.e. based on the speaker’s knowledge of the world, to conjectural ones, i.e. based solely on his or her own reasoning. In (18), for instance, there are external indications for the (situational) inference that the woman is home. In (19), the (generic) inference follows not only from the observation that the man had a heart attack but also from the general knowledge that people in good health do not normally die from being surprised. In (20), there is no external evidence or world knowledge that seems to motivate the (conjectural) inference that it is the postman at the door. It is based on the speaker’s thinking only.

(18) Her car is in the driveway. She must be home.
(19) What? He had a heart attack at his own surprise party?! He must have had a heart condition.
(20) [The doorbell rings.] It must be the postman. (cf. Mortelmans 2010: 12)

It is not always easy to distinguish these three types of inference in corpus data. But it will not come as a surprise that, in academic discourse, all but a few attestations of must are situationally and/or generically inferential. The process of inferring things from data and from world knowledge (which the reader is assumed to share) is central to the academic enterprise. Example (21) is indicative of approximately two fifths of the cases of must, in that the evidence on which the inference is based is mentioned explicitly. Most of the other instances resemble (22), in which more general knowledge is involved in the deduction. More conjecture-like cases such as (23) are rare and marked. Note also that there are no significant differences between the humanities and the social sciences.

(21) If this is so, the relative deprivation of blacks must have grown in the United States from 1939 to 1985, since the proportion of black prisoners in those years rose from 26 percent to 46 percent of the total prison population. (SS176)
(22) To build these structures, people must have had knowledge of engineering and design, and a large, stable work force. Until now, it was assumed they lived in huts made of tree trunks and leaves. (SS503)
(23) It was cool how they could be really loud one minute and the next be soft. They also did it without. It must be hard to remember all of the music. (H542)

Non-root must is typically used as a hedge. On the basis of specific evidence or general knowledge, the writer thinks that it is very likely that something is or, very
often in the data, was the case. A high degree of confidence is not the same as being certain, though. The use of *must* still allows for the possibility that something is or was not the case.

4. DUTCH

4.1. FREQUENCY

Table 4 first presents the total number of attestations of verbal *moeten* in CAD, then the number of relevant cases (i.e. not in titles and the like) and the number of relevant cases per 10,000 words.

Table 4: frequency of *moeten* in CAD

<table>
<thead>
<tr>
<th>Discipline</th>
<th># total</th>
<th># relevant</th>
<th>10,000 w</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanities</td>
<td>531</td>
<td>499</td>
<td>11.33</td>
</tr>
<tr>
<td>Social sciences</td>
<td>770</td>
<td>770</td>
<td>16.81</td>
</tr>
</tbody>
</table>

The comparison of Table 4 to Table 2 reveals one similarity and one difference. The Dutch corpora are like the English ones in that both modal verbs of necessity occur more frequently in the social sciences than in the humanities. However, *moeten* has a much higher rate of occurrence than *must* (which is line which previous findings by Mortelmans 2009: 7 and Van Olmen 2011: 497). One of the main reasons for this difference is that *moeten*, unlike *must*, has past tense and infinitival forms. Another important reason is that it is more or less the only commonly used necessity modal in Dutch whereas *must* has a number of “competitors” including *need (to)*, *have (got) to* and *should*. Accordingly, one would expect *moeten* to express more meanings more frequently than *must*. Let us first have a look at the distribution of root and non-root modality. Table 5 gives –from left to right– the raw frequency, the percentages and the frequency per 10,000 words.

Table 5: distribution of meanings of *moeten* in CAD

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Humanities</th>
<th>Social sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># raw</td>
<td>%</td>
</tr>
<tr>
<td>Root</td>
<td>468</td>
<td>93.79</td>
</tr>
<tr>
<td>Non-root</td>
<td>31</td>
<td>6.21</td>
</tr>
</tbody>
</table>

Like *must*, *moeten* does not convey non-root modality very frequently in academic discourse. What is surprising in view of the literature, which suggests that non-root *moeten* is much less common than non-root *must* (e.g. Mortelmans 2010: 7), is that
the rate of occurrence of non-root moeten is actually slightly higher than that of non-root must in the humanities. In Dutch, unlike in English, the social sciences differ considerably from the humanities and, not containing so many cases of non-root moeten, the former are more in line with previous research than the latter (according to Nuyts et al. 2010: 20, for instance, moeten only expresses epistemicity and/or evidentiality in one out of 100 attestations in spoken Dutch). Possible reasons for the differences between English and Dutch and between the Dutch humanities and social sciences here are discussed in Section 4.3.

Moeten, like must, can also be analyzed in terms of dynamic versus deontic root modality. If we translate Nuyts et al.’s (2010) fine-grained classification of modal meanings into Palmer’s (2001) terminology and if we follow their “progressive count” method (cf. Byloo & Nuyts 2011: 48), their results could be paraphrased as follows: in spoken Dutch, moeten is dynamic in roughly one third of the cases and deontic in roughly two thirds. A look at 100 root cases in the humanities and the same number of root cases in the social sciences reveals that the social sciences again exhibit more or less the expected distribution, i.e. 72 deontic versus 28 dynamic cases, whereas the humanities have an even distribution of the two root meanings. Note that the fact that dynamic modality is a relatively strong meaning in moeten, unlike in must (English probably uses need (to) and have to instead), is part of the explanation for the difference in frequency between the two necessity modals.

4.2. Correlation and Discourse Functions

Let us first consider the fact that, unlike in the English data, non-third person subjects are extremely rare here. Moeten combines with we in only 3% of the cases in both the humanities and the social sciences and it always has a root meaning. Extracts (24) and (25) are cases in point.

(24) De eigenlijke nood van het wonen berust erop dat de stervelingen het wezen van het wonen immer steeds weer zoeken, dat ze het wonen nog moeten leren. Als we Heidegger ernstig nemen, moeten we aannemen dat er een onoverbrugbare afstand bestaat tussen moderniteit en wonen. Dat is in ieder geval de overtuiging van Massimo Cacciari in Eupalinos or Architecture. (H25)

‘The actual necessity of residence lies in the fact that mortals look for the essence of residence again and again, that they still have to learn to reside. If we take Heidegger seriously, we must assume that there is an unbridgeable gap between modernity and residence. That is the conviction of Massimo Cacciari in Eupalinos or Architecture anyway.’

(25) Hierbij moeten we wel vermelden dat de gestandaardiseerde regressiecoëfficiënt voor deze variabele halveert van 0,184 naar 0,097 na toevoeging van lijstfactoren in het vijfde regressiemodel. (SS11)
'Hereby, we must mention that the standardized regression coefficient for this variable decreases by half from 0.184 to 0.097 after adding list factors in the fifth regression model.'

In (24), the first person plural subject pronoun stands for people in general and can easily be replaced by men ‘one’. In (25), by contrast, it clearly refers to the author, who presents him- or herself as being forced to mention some important piece of information. No other non-third person subjects occur in the data. Even though there were some instances of you in the English and Lithuanian data, these appeared mainly in instructional contexts; such type of texts did not appear in the Dutch data probably because it is more restricted in size. Still the use of the second person pronouns do not seem to be very typical of academic discourse generally. Hyland notes that you and your are rare in academic writing, “probably because they imply a separation between writer and reader” (Hyland 2008: 9)

A related, syntactic feature of moeten is that it combines with passive infinitives in roughly one third of the cases in the humanities and roughly two fifths of the cases in the social sciences. The Dutch modal is similar to its English counterpart in that passives infinitives are typical of root modality (note that, like for non-root must, the correlation between non-root moeten and the perfect is almost 100%) but the percentage of cases with a passive infinitive is about twice as high in Dutch as in English. Consider examples (26) and (27).

(26) Een eerste vraag die moet worden beantwoord is: wat maakte de betrokkenheid van de ‘traditionele’ clan leiders bij de opbouw van een staat in noordwest Somalí mogelijk en wat hield ze in? (H34)

‘A first question that must be answered is: what made the involvement of the “traditional” clan leaders in the formation of a state in North West Somalia possible and what did it entail?’

(27) De term ‘pestbeleid’ moet dan ook steeds begrepen worden als de retrospectieve poging om specifieke maatregelen te isoleren binnen dit groter geheel. (H103)

‘That is why the the term “bullying policy” must always be understood as the retrospective attempt to isolate specific measures within this bigger whole.’

Many instances are like in (26) in that the passive construction allows the author, who is the person raising and answering the question, to remain in the background. But in quite a few instances, such as (27), the passive seems to be employed to avoid using we ‘people in general’, including the hearer in a way: it is everyone who needs to understand the term in the right way.

The passive is also often attested in combination with verbs denoting the act of conversation in a broad sense, though such verbs occur with first person plural
subjects too (cf. Section 3.2). English academic discourse exhibits constructions such as (28) and (29) as well but they seem to be slightly more frequent in CAD. They make up roughly 5% of the cases of moeten (which always express root modality) in the humanities and the social sciences.

(28) Hierbij moet worden opgemerkt dat, in tegenstelling tot kinderopvang, familiale factoren ook bepaald worden door genetische factoren. (SS79)

‘Hereby, it must be noted that, in contrast to day care, family factors are also determined by genetic factors.’

(29) Wel moeten we nogmaals benadrukken dat onze factoren niet verwijzen naar kenmerken van de kiezers, maar naar kenmerken van de kandidaten op de kieslijsten. (H1)

‘However, we must emphasize one more time that our factors do not refer to characteristics of the voters but to characteristics of the candidates on the lists.’

As mentioned before, such formulas are employed for expository purposes. The writer uses them to guide the reader through the text and, in (28) and (29), to draw his or her attention to some relevant piece of information (cf. the fact that family factors are determined by genetic factor as well in 28) and, perhaps most importantly here, to manage the manner in which he or she understands the text (e.g. the possible misinterpretation of the factors in 29).

To conclude this section, we want to explore potential explanations for the differences between the humanities and the social sciences within the domain of root modality. It is clear that the high frequency of deontic instances in the social sciences can, in part, be attributed to the law articles. Referring to legal obligations and the like, the authors are forced to use deontic moeten quite often. Extract (30) can serve as an example: it is full of requirements which the arbitrating council should meet to function in a lawful way. Another source of many deontic cases is the subdisciplines of education and political science, which abound with recommendations directed at educators or at the government. Extract (31) is a case in point: the authors (indirectly) tell Flemish university colleges and universities to be clearer about their expenditures for promotional campaigns.

(30) Art. 15.1 van de Spaanse arbitragewet (2003) vereist dat, behoudens akkoord van partijen en indien er niet volgens billijkheid uitspraak mag worden gedaan, de alleen-zetelende arbiter een jurist moeten zijn. Bij een arbitraal college moet minstens één arbiter een jurist zijn. Tot een wetswijziging in 2011 vereiste dit artikel dat bij nationale arbitrage de arbiters niet alleen jurist maar zelfs advocaat moeten zijn. (SS36)

‘Article 15.1 of the Spanish arbitrating law (2003) stipulated that, excepting an agreement of the parties and in case it is not allowed to pass judgment on the basis of good sense, the sole arbiter must be a legist. In an arbitrating council, at least one of
the arbiters must be a legist. Until the modification of the law in 2011, this article stipulated that, in national arbitrating, the arbiters must be not only legists but even lawyers.

(31) Zo menen sommigen onder hen dat de uitgaven hieromtrent beperkt zijn, terwijl de Vlaamse hogescholen en universiteiten in 2005 samen al meer dan drie miljoen euro van hun budget besteedden aan promotiecampagnes. Ons inziens moet er dringend meer duidelijkheid worden gecreëerd. (SS17)

‘For instance, some of them believe that the expenditures in this area are limited while, together, the Flemish university colleges and universities spent more than three million euros of their budget to promotional campaigns in 2005. In our opinion, more clarity must be provided.’

The reason for the higher number of dynamic cases in the humanities is harder to give. One striking difference is that, in the humanities, about two fifths of the dynamic instances are past tense forms, which make up approximately four fifths of the total number of past tense forms of moeten, while, in the social sciences, only about 10% of the dynamic instances are past tense forms, making up just a quarter of the total number of past tense forms of moeten. Extract (32) is a typical example of such a dynamic use of past tense moeten in the humanities.

(32) Dit vereiste improvisatie bij de inrichting, zoals de vele kaats- of tennisbanen die tot theater werden omgevormd. Bij de oprichting van de almozeniersschouwburg in 1661 was dat niet anders. De almozeniers moeten genoegen nemen met een ruimte in het gildehuis van de oude Voetboog, aan de Grote Markt. (H64)

‘This required improvisation in the construction, like the many bouncing or tennis courts that were transformed into theaters. This was the same for the establishment of the almoners’ theater in 1661. The almoners had to settle with a space in the guild house of the Oude Voetboog, at the Grote Markt.’

The modal verb in (32) is indicative of most of the other instances in that it is used to refer to some circumstantial necessity in the past, i.e. the establishment of a theater by some other institution forcing the almoners to move out of their own building, and in that it occurs in the subdisciplines of history and literary science, in particular when it takes a more historical turn.

4.3. EPISTEMICITY AND EVIDENTIALITY

Moenen, like must, has epistemic and evidential meanings that are often hard to disentangle. The Dutch verb differs from its English equivalent in two important respects, however. On the one hand, moeten can be used in a purely evidential way. Sentence (33) is a case in point. In this example, the speaker is not indicating that he or she considers it likely that it is a pretty girl. Rather, he or she is indicating that it is said
that it is a pretty girl. In other words, moeten has a reportative function here, which is also evident from its optional co-occurrence with the grammaticalizing evidential marker naar verluidt ‘reportedly’.

(33) Het moet een mooie meid zijn, (naar verluidt).
   ‘It is said that it is a pretty girl, (reportedly).’

On the other hand, although it is similar to must in combining epistemic and inferential meanings, moeten appears to be restricted to situational and generic inferences. Mortelmans (2010) notes that the verb is fine in contexts such as (34) and (35) (as discussed above, the speaker in the former case infers that she is home from the external evidence that her car is in the driveway while the speaker in the latter case infers that he had a heart condition from the news that he or she just received and from the general knowledge that healthy people do not normally get a heart attack at their surprise party), the translations of (18) and (19). But she observes that it is not acceptable in contexts such as (36), the translation of (20). With conjectural inferences, i.e. those based solely on the speaker’s own thinking, Dutch uses zullen ‘shall, will’ and the modal particle wel instead.

(34) Haar auto staat op de oprit. Ze moet thuis zijn.
   ‘Her car is in the driveway. She must be home.’

(35) Wat? Hij heeft een hartaanval gekregen op zijn eigen verrassingsfeestje?! Hij moet een hartzekte gehad hebben.
   ‘What? He had a heart attack at his own surprise party?! He must have had a heart condition.’

(36) [De deurbel gaat.] Het zal de postbode wel zijn.
   ‘[The doorbell rings.] It must be the postman.’

No cases of reportative moeten are attested in our corpus. This result is not entirely unexpected as academic discourse can be assumed to more faith in external evidence and general world knowledge than in hearsay. Examples of situational and more generic inferences are given in (37) and (38). In the first example, the inference that the text is corrupt follows directly from some perceivable piece of evidence, i.e. the absence of a particular element. In the second example, the inference seems to be based more on general rules of logic: if A includes B, B shares the properties of A.

(37) En bovendien ontbreekt in de overgeleverde tekst het tweede lid van de vergelijking die door meere wordt geïmpliceerd. De tekst moet corrupt zijn. (H329)
‘And, what is more, in the text that has been handed down, the second part of the comparison, which is implied by meer, is missing. The text must be corrupt.’

(38) Aangezien de “monitorial citizen” een specificatie is van de “rights-conscious citizen”, moet hij ook positief staan tegenover meer burgerrechten en een grotere vorm van maatschappelijke gelijkheid. (SS593)

‘Since the “monitorial citizen” is a specification of the “rights-conscious citizen”, he must also be in favor of more civil rights and a greater form of societal equality.’

As mentioned before, the humanities contain many more cases of non-root moeten than the social sciences. This high rate of occurrence, which is even a little strange in light of previous research on Dutch, can (again) be attributed to the subdisciplines of history and literary science, particularly that of a more historical nature. Moeten is typically used by authors to make claims about the past, for which they have some proof but about which they cannot be absolutely sure by definition. Consider (39) and (40). In the latter example, for one, the speaker infers that Gezelle, a nineteenth-century poet, was intrigued by a certain medieval story from the fact that he translated it into more modern Dutch.

(39) ‘Hetzelfde moet voor bakkersovens hebben gegolden. Andere panden aan het Noorderinde mochten namelijk een opvallend groot aantal bakkers tot hun successieve bewoners rekenen.’ (H194)

‘The same must have held for bakers’ ovens. After all, other houses at the North End included a remarkably large number of bakers among their successive residents.’

(40) Alleen al die korte hertaling waaraan hij zich waagde, maakte onze tijdrovende zoektocht doorheen jaargangen van Rond den heerd succesvol. Het Reynaertverhaal moet Gezelle hebben geboeid. (H240)

‘That short translation which he dared to make, if nothing else, has made our time-consuming search in volumes of Rond den heerd a success. The Reynaert story must have fascinated Gezelle.’

Actually, the English data too are full of such contexts, even in the social sciences. It seems that, as (21) and (22) suggest, the texts subsumed under the heading of this discipline in English just happen to take a more historical perspective than those in Dutch (supporting evidence comes from the fact that non-root moeten never combines with the perfect in the Dutch social sciences). Note, finally, that the “expected” lower frequency of non-root moeten in contrast to must, as attested in the social sciences, could also be explained in terms of general communication preferences. With respect to non-root must and German non-root müssen, which closely resembles non-root moeten in frequency and usage, Mortelmans (2010: 21) cites House (1996), who argues that English prefers an indirect and hearer/reader-oriented style
of communication while German favors a direct and content-related style of communication. For markers of epistemicity, this difference means that they “will be more commonly used in the English texts, because they contribute to making texts more indirect as well as more addressee-oriented” (Becher et al. 2009: 129). Dutch might be closer to German than to English in this respect.

5. LITHUANIAN

5.1. FREQUENCY

There are two multifunctional modal verbs in Lithuanian which can express both root and non-root modality and thus “cover the whole domain of modality from dynamic to epistemic” (Holvoet 2009: 206). The verb *galėti* ‘can/may’ expresses possibility, the verb *turėti* ‘have to’ expresses necessity. Unlike the modal auxiliaries in English, the two Lithuanian modal verbs display all the properties of regular verbs. Therefore, the distinction is more of a semantic than of a grammatical nature.

Even though there are only two modal verbs to express non-root possibility/necessity, different semantic nuances can be conveyed by their different forms. Table 6, suggested by Usonienė (2004: 43), lists possible English equivalents of the forms of the two modals in Lithuanian expressing various degrees of certainty.

<table>
<thead>
<tr>
<th>English</th>
<th>Lithuanian</th>
</tr>
</thead>
<tbody>
<tr>
<td>might</td>
<td><em>galėtų</em> ‘can/may.SBJV.3’</td>
</tr>
<tr>
<td>may</td>
<td><em>gali</em> / <em>(galejo)</em> ‘can/may.PRS.3’ <em>(can/may.PST.3’)</em></td>
</tr>
<tr>
<td>could</td>
<td><em>turi</em> / <em>(turėjo)</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>can hardly</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>He</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>should</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>be</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>right.</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>ought to</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>would</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>will</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
<tr>
<td>must</td>
<td><em>turėtų</em> ‘have.SBJV.3’ būti teisus.</td>
</tr>
</tbody>
</table>

Alongside its modal meaning, the verb *turėti* has also retained its original meaning of possession (Holvoet 2007: 43).

*Turėti* is an extremely frequent word in the Lithuanian language. According to the Frequency Dictionary of Written Lithuanian (Utka 2009), based on a morphologically annotated corpus of one million words, *turėti* is the eighteenth most common
word and the third most common verb, after *būti* ‘to be’ and *galėti* ‘can/may’. The verb is also frequent in academic discourse. Its form *turi*³ ‘have.PRS.3’ is the thirtieth (with a total of 17,582 tokens) out of 380,000 types in the nine million word corpus of academic Lithuanian (CoraLit). Keeping in mind the multifunctionality of *turėti*, one might be interested to see whether it is the modal meaning or the meaning of possession which is dominant in academic discourse.

Due to the high frequency of *turėti*, only its positive finite forms have been selected for analysis. All the finite forms were extracted from the corpus automatically and afterward analyzed manually in order to discard any irrelevant cases and to separate the occurrences with possessive and modal readings. Just like for English and Dutch, irrelevant occurrences are cases in quotes, titles, examples and the like. Occurrences of *turėti* with a modal meaning have been classified into root and non-root. Table 7 presents –from left to right– the raw frequency, the percentages and the frequency per 10,000 words of the relevant instances of *turėti*.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Humanities</th>
<th></th>
<th>Social sciences</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># raw</td>
<td>%</td>
<td>10,000 w</td>
<td># raw</td>
</tr>
<tr>
<td>Possessive</td>
<td>2,263</td>
<td>47.60</td>
<td>11.15</td>
<td>2,131</td>
</tr>
<tr>
<td>Root</td>
<td>2,330</td>
<td>49.01</td>
<td>11.46</td>
<td>3,434</td>
</tr>
<tr>
<td>Non-root</td>
<td>161</td>
<td>3.39</td>
<td>0.79</td>
<td>76</td>
</tr>
<tr>
<td>Total</td>
<td>4,754</td>
<td>100</td>
<td>23.40</td>
<td>5,641</td>
</tr>
</tbody>
</table>

We can see from Table 7 that the total number of possessive cases of *turėti* in the two disciplines is 4,394 while modal *turėti* amounts to 6,001 cases. The ratio of modal versus possessive meanings shows that modal meanings in the two analyzed academic disciplines are dominating (roughly 58% versus 42%). This fact is in contrast to *turėti*’s behavior in fiction, in which the situation is nearly the opposite. Possessive *turėti* accounts for 61% and modal *turėti* for 39% of all cases of the verb (Šolienė 2012).

The quantitative analysis reveals both differences and similarities in the use of the verb in both disciplines. Scholars in the social sciences employ *turėti* one and a half times more frequently than scholars in the humanities (normalized frequencies of about 38 versus 23 cases per 10,000 words). Another clear difference is in the distribution of possessive and modal meanings. While researchers in the humanities use *turėti* as a verb of necessity and possession in more or less equal proportions (52% versus 48%), researchers in the social sciences clearly favor the use of *turėti* as a modal verb (62% versus 38%). Possible explanations for those patterns are provided

---
³ CorALit is not morphologically annotated or lemmatized. Therefore, all word forms are counted separately as separate types here.
in Section 5.2. It should also be noted that the non-root modal meaning of *turėti* is equally scarce in both fields (3%\(^4\) in the humanities and 1% in the social sciences).

A look at the quantitative data of the three languages is suggestive of certain disciplinary trends. Researchers in the social sciences employ the necessity modals under consideration more frequently than scientists in the humanities. Likewise, in all three languages, the non-root meaning is quite rare (under 8%) in both disciplines. However, while the Dutch corpus showed a much denser use of *moeten* in comparison to *must* in the English data, the Lithuanian corpus exhibits an even higher rate of occurrence of *turėti*, especially in the social sciences. If we deduct the possessive cases, the normalized frequency of the modal *turėti* per 10,000 words is 12.28 in the humanities and 23.71 in the social sciences. Like in Dutch, this can be partly explained by the variety of the forms that *turėti* has (past, frequentative and future tense, subjunctive). Though *turėti* is not the only modal of root necessity in Lithuanian, it is semantically weaker than, for instance, *privalėti* ‘be obliged’ and thus seems more suitable for academic discourse, in which authors might be willing to avoid sounding too imposing.

Certain aspects of the use of the verb are very comparable in both disciplines and point toward common trends in Lithuanian academic discourse. Tables 8 and 9 show the frequency information of various finite forms of *turėti* in the social sciences and the humanities. A detailed account of the distribution of the various forms is provided only for the third person, as this form is significantly more common than the first and second person forms.

Table 8: distribution of positive finite forms of *turėti* in the social sciences

<table>
<thead>
<tr>
<th>Form</th>
<th>Total # raw 10,000 w</th>
<th>Possessive # raw 10,000 w</th>
<th>Root # raw 10,000 w</th>
<th>Non-root # raw 10,000 w</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Sg/Pl total</td>
<td>126 0.86</td>
<td>46 0.31</td>
<td>80 0.54</td>
<td>0 0</td>
</tr>
<tr>
<td>2nd Sg/Pl total</td>
<td>39 0.27</td>
<td>30 0.20</td>
<td>9 0.07</td>
<td>0 0</td>
</tr>
<tr>
<td>3rd Sg/Pl total</td>
<td>5,476 36.99</td>
<td>2,055 13.89</td>
<td>3,345 22.60</td>
<td>76 0.52</td>
</tr>
<tr>
<td>PRS: tur-i</td>
<td>3,599 24.31</td>
<td>1,537 10.39</td>
<td>2,061 13.92</td>
<td>1 0.01</td>
</tr>
<tr>
<td>PST: tur-ėjo</td>
<td>601 4.06</td>
<td>375 2.53</td>
<td>216 1.46</td>
<td>10 0.07</td>
</tr>
<tr>
<td>FRQ: tur-ėdavo</td>
<td>7 0.05</td>
<td>1 0.01</td>
<td>6 0.04</td>
<td>0 0</td>
</tr>
<tr>
<td>SBJV: tur-ėt</td>
<td>1,164 7.86</td>
<td>72 0.49</td>
<td>1,027 6.94</td>
<td>65 0.44</td>
</tr>
<tr>
<td>FUT: tur-ės</td>
<td>1,05 0.71</td>
<td>70 0.47</td>
<td>35 0.24</td>
<td>0 0</td>
</tr>
</tbody>
</table>

We can see from Tables 8 and 9 that the first and the second person forms are very rare in both disciplines. This finding is in line with previous studies on personal pronoun usage in Lithuanian academic discourse (Šinkūnienė 2010), which suggests that Lithuanian scholars are not keen on overtly marking their presence in scientific texts, especially in comparison to English researchers. The infrequency of the first and

\(^4\) But note that about one or two percent of modal *turėti* in both disciplines are indeterminate or ambiguous and that, in the present analysis, these cases are subsumed under the heading of root modality.
second person forms results in the absolute dominance of the third person forms in both disciplines (93% in the humanities and 97% in the social sciences). In fiction, the third person forms are also dominant but the difference is not so dramatic, with the third person forms making up 67% of the total (Šolienė 2012).

It is somewhat surprising to find second person forms in CoraLit, just like in the English data. They are infrequent and most of them (i.e. 31 occurrences out of 52 cases) occur in one textbook in the social sciences, entitled *Darbo paieškos vadovas* ‘a guide to job hunting’. Directly addressing the reader creates a personal and interactive atmosphere, especially in combination with the first person pronoun representing the author, as in (41).

(41) Norėdami pasiekti viršukalnę (susirasti darbą), jūs turite numatyti savo kopimo etapus ir kelią (veiksmų planą). Taip pat mes jums rekomenduojame numatyti ne tik galutinį tikslą – įkopti į viršukalnę, bet ir tarpinius tikslus <...>. (SS138)

‘If you wish to reach the top of the mountain (to find a job), you must foresee the stages of climbing and the route (action plan). We also recommend you to plan not only the ultimate goal which is to reach the top of the mountain, but also intermediate goals <...>.’

Yet, even though the use of second person pronouns facilitates the involvement of the reader into the discourse, its combinations with *turėti* also emphasize the writer’s authority. This is especially clear in (41), in which a single author uses the first person plural, presumably to speak on behalf of the more experienced part of the community of experts to which she belongs. As mentioned before, the infrequency of such forms in the analyzed data suggests that this is not a typical feature of Lithuanian academic discourse and that their use is determined either by the specificity of the topic or the style of the author.

Academic Lithuanian and academic Dutch, with their low percentages of non-third person forms in combination with necessity modals, appear to be on a par if compared to academic English. This is hardly surprising, though. The overt expression of author
stand via personal pronouns seems to be an idiosyncrasy of the anglo-american writing tradition: personal pronouns are reported to be employed to a much lesser extent in academic writing of other cultures, such as Italian (Molino 2010), Spanish (Mur-Dueñas 2007), German, Russian, Bulgarian and French (Vassileva 1998).

As far as the third person forms in the Lithuanian data are concerned, the present tense form is dominant in both the humanities (55%) and social sciences (66%), which seems to be typical of the verb regardless of the type of discourse it occurs in. This form is the most frequent one in general written Lithuanian language (Utka 2009) and in fiction (Šolienė 2012). In the rest of the analysis, we will concentrate on the third person forms.

5.2. CORRELATIONS AND DISCOURSE FUNCTIONS

Though the focus of the study is on the modal use of the modal verbs of necessity, a few notes on the patterns of possessive turėti will also be provided, as it is quite frequent in Lithuanian academic discourse.

In spite of the fact that researchers in the social sciences employ possessive turėti more than researchers in the humanities (roughly 14 versus 11 cases per 10,000 words), usage patterns are similar in both disciplines. In all of its third person forms, the verb turėti recurrently combines with abstract rather than concrete nouns. Combinations with the nouns ‘influence’, ‘possibility’, ‘right’ and ‘significance’ account for a quarter of the cases of possessive turėti ‘have.PRS.3’ and for 38% of the cases of possessive turėjo ‘have.PST.3’ in the humanities. Sentence (42) is a typical example.

(42) Meno sociologijos atsinaujinimui tiesioginę įtaką turėjo bendra to meto dvasinę įvykių įtaką bei ypač audringi 1968-ųjų gegužės socialiniai, kultūriniai ir politiniai įvykiai <...> (H711)

‘General spiritual atmosphere of those times and especially the tempestuous social, cultural and political events had a direct influence on the renewal of art sociology.’

Similar percentages of combinations of possessive turi ‘have.PRS.3’ (32%) and turėjo ‘have.PST.3’ (37%) with the above-mentioned nouns are observed in the social sciences.

Let us now turn to root turėti. The main cross-disciplinary difference in the use of root turėti is the frequency with which the verb is employed. Researchers in the social sciences like this modal verb in its root meaning to a much greater extent than researchers in the humanities (23 versus 11 occurrences per 10,000 words). A closer analysis of the most frequent form, i.e. turi ‘have.PRS.3’, reveals that, out of the five subdisciplines making up the bulk of the subcorpus of social sciences, i.e. economics, education, psychology, political science and law, it is the texts in economics and law that display the highest frequency (20 occurrences per 10,000 words) of root
‘have.PRS.3’ In the subdiscipline of education, the normalized frequency of this form is 13 occurrences per 10,000 words whereas, in political science, it is 9 occurrences per 10,000 words. This marker is the most infrequent in psychology texts, in which its normalized frequency is just 3 occurrences per 10,000 words.

The high frequency of root turėti in law and economics can be explained by the nature of those disciplines. In economics, researchers quite frequently offer observations of a more hortative nature, which can be expressed by turėti, as in (43) and (44).

(43) Taigi įmonės finansininkas, norėdamas laiku sureaguoti į kylančią bankroto grėsmę, turėti nuolat stebėti įmonės finansinius rodiklius, o ypač mokumo, kapitalo rentabilumo ir jų kaitą. (SS291)
‘Therefore, in order to timely react to the threat of approaching bankruptcy, the accountant constantly must follow the financial indicators of the company, especially its solvency and profitability as well as their fluctuations.’

(44) Pastaroji [deflacija], straipsnio autoriaus nuomone, atspindi ekonomines tradicijas, kurios nebeatitinka iuolaikinio ekonomikos pažinimo lygio ir turėti būti keistina. (SS285)
‘The latter [deflation], in the opinion of the author of the article, reflects economic traditions, which no longer correspond to the standards of contemporary economics and therefore must be changed.’

Sentence (44) is interesting as an example of Lithuanian academic discourse in that it demonstrates the author’s attempt to avoid referring to himself directly (cf. in the opinion of the author of the article). Rather than choosing an explicit I think or in my opinion, the author resorts to naming himself in the third person, thus making the text less personal. Such strategic uses as well as the much more frequent use of impersonal constructions account for the lack of first person pronouns in Lithuanian academic discourse.

Not unexpectedly, in the subdiscipline of law, root turėti often refers to various legal obligations, placed on various institutions and individuals. Consider the following example.

(45) Tačiau vien to fakto, kad fizinis asmuo gauna autorinį atlyginimą, negalima laikyti lemiamu asmens veiklą pripažįstant ekonomine veikla. Tam, kad autoriaus veiklą būtų galima pripažinti ekonomine veikla, turėti būtų nustatytı visi minėti ekonominės veiklos požymiai. (SS1035)
‘However, the fact that an individual receives payments of royalties alone cannot be treated as the decisive factor in acknowledging an individual’s activity as economic activity. In order for an individual’s certain activity to be acknowledged as economic activity, all the outlined features of economic activities must be identified.’
As has been already mentioned in Section 4.2, similar contexts of use are attested for *moeten* in the Dutch data, which points to common trends at the level of subdisciplines.

The slightly more diverse selection of text types in CorALit has also enabled us to trace specific trends of the use of *turėti* in texts other than research articles. In the discipline of the humanities, for instance, 45% of the most frequent form *turi* 'have.PRS.3' are used in textbooks. In such texts, claims with *turėti* are quite general, often of an explanatory nature or with reference to the conventions that are valid and should be followed within the discipline. Consider (46) and (47).

(46) Žurnalistika *turi* plačiai ir tiksliai reikšti viešąją nuomonę, kuri žmonių gyvenime yra vienas svarbiausių tikros demokratijos požymių. (H412)

‘Journalism must broadly and precisely express public opinion, which in people’s life is one of the most important features of true democracy’.

(47) Visos žinios apie sportą, kūno kultūrą *turi* būti tampriai susietos su bendromis žiniomis apie žmogų ir apie bendrą kultūrą apskritai. (H2084)

‘All knowledge about sports, physical culture must be closely related to the general knowledge about the human being and generally about culture.’

In the textbook material in CorALit, the frequency of the third person forms of *turėti* amounts to 15 cases per 10,000 words in the humanities. In the social sciences, it is as high as 28 occurrences per 10,000 words, which is again suggestive of the stronger need to express obligation in this particular discipline.5

As for the combinations of *turėti* with „verbs of conversation“ to perform discourse-oriented functions, which have been attested in English and (more frequently) in Dutch, such uses are not very typical of *turėti*. The passive (e.g. *turi būti pastebėta/pasakyta* ‘it must be observed/said’), though grammatically and semantically possible, sounds somewhat unusual, and such constructions are not attested in the data. The only combinations of *turėti* with *say, observe, emphasize* and the like occur when *turėti* takes the first person singular or plural form, as in (48).

(48) Kaip mes beprirėmume pagrindines veikalo idėjas, *turime* priipazinti, kad šis rūpestis – nuoširdus. (H631)

‘Regardless of how we are going to accept the main ideas of the work, we must admit that this worry is genuine.’

There are a total of 13 such cases in the humanities and 7 cases in the social

---

5 Because of the different number/length of the textbooks in both disciplines, normalized frequencies per 10,000 words have been calculated only for the textbook data to make the comparison between the two disciplines meaningful. The textbooks in the humanities contain 453,494 words, the social sciences 109,942 words.
sciences. In light of the very high frequency of *turėti*, they are really marginal. Note, however, that this does not mean that Lithuanian researchers are less conscious about the rhetorical and pragmatic effects of their texts. This emphatic/expository function in Lithuanian academic discourse is performed by another verb of necessity, impersonal necessitive *reikia* ‘need’, as in (49).

(49) *Reikia pažymėti*, kad tradicinės kultūros raiškos ypatumai yra tiesiogiai susiję, priklausomi nuo jos specifikos, t.y. bendruomeninės prigimties ir bendruomeninių sklaidos formų. (H4613)

‘It needs to be noted, that the peculiarities of the traditional cultural expression are directly related, dependent upon its specificity, i.e. communal nature and communal forms of dissemination.’

A wide range of verbs of saying – such as *pasakyti* ‘say’, *pastebėti* ‘observe’, *pabrėžti* ‘stress’ and *pridurti* ‘add’ – combine with *reikia* ‘need’ to act as metadiscourse signals.

5.3. EPISTEMICITY AND EVIDENTIALITY

As mentioned before, non-root *turėti* is quite rare in academic Lithuanian and accounts for only 3% of the use of the verb in the humanities and 1% in the social sciences. Traditionally, non-root *turėti* is assigned to the domain of epistemic necessity and, to our knowledge, there exist no extensive accounts of the additional semantic properties of the verb.

It is clear from Tables 8 and 9 that it is primarily the subjunctive and the past tense forms of *turėti* which express non-root modality. The subjunctive form of *turėti* makes up 37% of all cases of non-root *turėti* in the humanities and as many as 86% in the social sciences whereas the past tense form is dominant in the humanities (58%) and is used to a much lesser extent in the social sciences (13%). It has been shown in Table 6 that the subjunctive of *turėti* is more tentative than its other forms. Therefore, its choice to present certain claims as being more tentative is not surprising. In the social sciences, the subjunctive form of *turėti* is not only frequently employed to present tentative predictions about the future, as (50) shows, but also acts as a marker of pragmatic politeness while talking about the potential success, value or contribution of a certain piece of work (e.g. a study, a book, an analysis) to society in general or to specific academic community, as in (51).

(50) Kita vertus, 2007–2013 m. ES struktūrinė parama Lietuva *turėtų* sumažinti finansinių išteklių stokos problemą daugelyje Lisabonos strategijos sričių. (SS1341)

‘On the other hand, the support of EU structural funds provided to Lithuania in 2007 – 2013 must reduce the problem of limited financial resources in many areas of the Lisbon Strategy.’
(51) Dėl šios priežasties manytina, kad publikacija turėtų būti naudinga tiek studijuo- jantiems baudžiamąją teisę, tiek teisininkams praktikams, kasdieniame savo darbe taikantiems baudžiamuosius įstatymus. (SS1056)

‘Due to this reason it can be considered that the publication must be equally useful to those who study criminal law and to lawyers practitioners, who apply criminal law in their everyday work.’

While subjunctive turėti also occurs in pragmatic politeness contexts in the humanities, it is much more frequent in the discourse of history to make tentative judgments about certain archaeological finds or the interpretation of historical data, as in (52).

(52) Vėlyviausias M-L H grupės tipas tarp sągčių iš Stragnų turėtų būti iš fragmentų susidėjusi žalvarinė sagtis AR 38:1841+1631 (pav. 4:5). (H1181)

‘Out of buckles from Stragnai area, the type dated latest in the M-L H group must be the brass buckle AR 38:1841+1631 (picture 4:5) formed from fragments.’

In fact, it is hardly surprising that the past tense form of turėti, which is frequent in the humanities, but much rarer in the social sciences, predominantly occurs in history texts, in which evaluations of and assumptions about past events are made – a trend also observed in Dutch academic discourse. Probably, it is the risk to make firm claims about the past, even with evidence at hand, that makes researchers resort to other epistemic and epistemic-evidential markers that occur alongside non-root turėti quite frequently in the discourse of history. The semantic range of those markers is from high/medium certainty to low certainty. These markers include matyt ‘evidently’, greičiausiai ‘most probably’, a variety of constructions with the verbs manyti ‘think/suppose’ and spėti ‘guess’. Consider (53) and (54).

(53) Kaip matome, realiai aptarti 7 variantai. Matyt, 5 variantas turėjo būti skirstomas į a ir b porą. (H1156)

‘As we can see, in reality there are only 7 variants discussed. Evidently, variant 5 must have been divided into a and b subgroups’.

(54) Galima spėti, jog mokyklos ir bažnyčios perkėlimo darbai greičiausiai turėjo būti apmokėti iš Lietuvos departamento parapijų paaukotų pinigų, nes Karvaičių bažnyčios kasa tuo metu jau buvo pereikvota. (H312)

‘It can be guessed that the transfer works of the school and the church most probably must have been paid from the contributions from Lithuanian department parishes, because the account of Karvaičiai church at that time had already been overspent.’

The contextual analysis of turėti in academic discourse shows that, just like must in English and moeten in Dutch, turėti predominantly occurs in cases of situational and generic inferences. It is not infrequent that researchers explicitly mention the reasons
which triggered their specific deductions. This can be done with the help of linking words such as *vadinasi* ‘consequently’ and *nes/kadangi* ‘because’ and judgment words such as *sprendžiant* ‘judging from’ and the like, as in (54) and (55). But the basis of the writer’s reasoning can also be present in the context without explicit markers, as in (56).

(55) *Sprendžiant iš turinio ir stilias, eilių autorius turėtų būti Ernstas Moritzas Arndtas, tačiau šią prielaidą dar reikės pagrįsti nuodugniau.* (H1001)

‘Judging from the content and style, the author of the poems *must* be Ernst Moritz Arndt, however this assumption still needs to be grounded more extensively.’

(56) Savo emocijas skrupulingai slėpusio Basanavičiaus „įjausminimas“ *turėjo* būti nelengvas uždaviny. Apmaudo gaida girdėti epizoduose, kuriose Basanavičius nepateikia taip pageidaujamų patriotinių jausmų. (H1080)

‘The “turning on” of Basanavičius, who normally intensely hid his emotions, *must* have been a hard task. A hint of annoyance can be heard in the episodes, where Basanavičius does not show so much desired patriotic feelings.’

As for the generic inferences, Lithuanian researchers, just like their Dutch or English colleagues, frequently base their judgments on their professional knowledge or on general knowledge about the world. Sentence (57) is a case in point.

(57) *Gyva vaizdinga kalba para ytas tekstas turėtų būti patrauklus paprastam skaitytojui.* (H466)

‘The text written in lively vivid language *must* be attractive to an ordinary reader.’

In this example, it is common knowledge about reading preferences of ordinary readers which helps the author tentatively predict the potential success of the book.

It has to be noted that clear conjectural contexts do not appear to be typical for *turėti*. Even when there is no evidence whatsoever in the immediate context, the author’s assumptions still seem to be motivated by something more than just his or her own reasoning. Finally, just like *must* and *moeten*, when it is used in its non-root meaning, either on its own or with other softening expressions, *turėti* always acts as a hedge as the proposition is presented as likely but not absolute.

6. CONCLUDING REMARKS

The analysis of academic discourse in the humanities and the social sciences in three different languages suggests that traces of both disciplinary and national culture can be seen in scientific texts. Researchers in the social sciences seem to be more
“directive” than scientists in the humanities, irrespective of the language. The fact that the subcorpus of English social sciences does not contain any texts from law or from economics may account for its much smaller numbers for necessity modals than in its Dutch and Lithuanian counterparts. This tendency as well as the tendency for most of the non-root meanings to occur in historically oriented discourse, irrespective of language confirms the similarity of expression, not only in different disciplines but also in subdisciplines.

The observation that the necessity modals under consideration are generally rather rare in their non-root meaning (normalized frequencies ranging from 0.09 to 0.78 per 10,000 words in our analysis) may point toward some general tendencies within academic discourse. Even though these linguistic items perform the pragmatic function of hedging, researchers might be unwilling to qualify their statements with markers of near certainty. The fact that in all three languages the non-root meanings appear primarily in contexts with situational or generic inferences suggests that there is always quite reliable grounds for the claim. It might be that the majority of such evidence-based claims appear in scientific texts with no modification at all. This is speculative at best, as it might also be that there exists more popular alternatives with a comparable semantic potential to enable scientists to modify their claims.

Finally, the fact that non-third person forms in combination with necessity modals are very rare in Dutch and Lithuanian academic discourse while English researchers do not try to avoid first and second person pronouns with must, points toward culture-specific differences in the expression of author stance. While Lithuanian and Dutch scientists try to distance themselves from their claims with the help of passives and impersonal constructions, English scientists try to emphasize their involvement in the text or involve the reader, employing inclusive we and reader-oriented you.

Jolanta Šinkūnienė, Daniël Van Olmen

BŪTINYBĖS MODALINIAI VIEKSMAZODŽIAI ANGLŲ, OLANDŲ IR LIETUVIŲ MOKŠO KALBOJE: EPISTEMIŠKUMAS IR (AR) EVIDENTIALUMAS?

Sątrauka

Straipsnyje tyrinėjami modalinių veiksmazodžių must, moeten ir turėti vartosenos ypatumai humanitarinių ir socialinių mokslų tekstuose. Tirti tekstynai COCA (Corpus of Contemporary American English), CorALit (Corpus Academicum Lithuanicum) ir autorų specialiai sudarytas olandų mokslo kalbos tekstynas. Atliktą kiekvienės ir kokybės analizę parodė, kad mokslinioje kalbos tekstui yra dominuojantys reikšmingų modalinių veiksmazodžių vartojimas, o neepistemės nevykdomos veiksmazodžių reikšmė yra dominuojanti visų trijų analizuotų kalbų tekstuose. Anglakalbiai moks-
lininkai nevengia vartoti pirmojo ir antrojo asmens įvardžių, siekdami sukurti glaudesnį santykį su skaitytoju. Lietuvių ir olandų kalba parašytose tekstuose, atvirkščiai, asmeniniai įvardžiai beveik nevartojami, dažnesnės pasyvės ar beasmenės konstrukcijos.

Vartojami epistemine reikšme, analizuoti būtinių modalinių veiksmadžiai atlieka sąsivelnio funkciją, tačiau apskритai jų epistemine reikšme tyrinėtoje mokslo veiksmu visoje kalbomis yra reta. Epistemine reikšme pavartotų veiksmadžių kontekste analizė rodo, kad dažnai jai autoriai formuluoją propoziciją kaip labai tikėtina, remdami išorinius arba vidiniai žinių šaltinius.

REFERENCES

Mur-Dueñas, Pilar. 2007. ‘I/we focus on…’: A cross-cultural analysis of self-mentions in business


