

## Epistemic qualifications of the English marker *likely* and its equivalents in Lithuanian

**Anna Ruskan**

Department of English Philology

Vilnius University

Universiteto g. 5

LT-01513 Vilnius, Lithuania

Tel.: +370 5 2687 228

E-mail: [anna.ruskan@flf.vu.lt](mailto:anna.ruskan@flf.vu.lt)

### Abstract

The current study focuses on the epistemic qualifications realised by the English adjective and adverb *likely* and its equivalents in Lithuanian *panašu* ‘likely, it seems’ and *tikėtina* ‘believable, likely’, which derive from the semantic domain of comparison and belief. The aim of the study is to identify the functional similarities and differences of the markers in terms of their frequency, syntactic features (Complement-Taking-Predicates (CTPs), adverbials), functions, collocational profile and type of discourse (academic, newspaper). The English and Lithuanian data were drawn from the monolingual corpora, namely the Corpus of Contemporary American English (COCA), Corpus of the Contemporary Lithuanian Language (CCLL), Corpus of Academic Lithuanian (CorALit) and the bidirectional parallel corpus ParaCorp<sub>EN→LT→EN</sub>. The quantitative and qualitative findings reveal that the closest cross-linguistic CTP and adverbial equivalents are *likely* and *tikėtina* ‘believable, likely’ because they are most frequent in formal registers (academic, newspaper discourse) and display similar collocational profiles and contexts of use. In contexts with explicit evidence and argumentation, they may acquire evidential functions. Although the CTP and adverbial *panašu* ‘likely, it seems’ shares similarities with *likely* and *tikėtina* ‘believable, likely’ in expressing the author’s degree of probability, it shows a different semantic profile from the latter due to its conceptual link with the original meaning of similarity and appearances. The study shows how markers that derive from the semantic domain of comparison vary in functional distribution in present-day English and Lithuanian and introduces their functional equivalents deriving from a different semantic domain.

**Keywords:** epistemicity, evidentiality, epistemic modality, epistemic evidential overlap, complementation, adverbial

## 1 Introduction

Over the last decade there have been a number of cross-linguistic studies dealing with the formal and functional comparison of evidential and epistemic modal markers. Much attention has been devoted to the modal verbs *must* (Mortelmans 2010; Šinkūnienė & Van Olmen 2012), *seem* (Johansson 2001; de Haan 2007; Aijmer 2009; Usonienė & Šinkūnienė 2013), epistemic and evidential adverbials and their equivalents in other Germanic, Romance, Baltic and Slavic languages (van der Auwera *et al.* 2005; Mortensen 2006; Simon-Vandenbergen & Aijmer 2007; Usonienė & Šolienė 2010). The current study focuses on the epistemic qualifications realised by the English adjective and adverb *likely* and its adjectival and adverbial equivalents in Lithuanian *panašu* ‘likely, it seems’ and *tikėtina* ‘believable, likely’. Etymologically the adjectival forms of *likely* and *panašu* ‘likely, it seems’ relate to the semantic domain of comparison. *Likely* is derived from the adjective *like*, meaning “having the same characteristics or qualities as some other person or thing; of approximately identical shape, size, etc., with something else; similar” (NSOED 1993, 1588) and the adjective forming suffix *-ly*, which originally has the same meaning as *like*. As mentioned by Besnard (forthcoming), the meaning of similarity is “reduplicated” in *likely*. The Lithuanian neuter adjective *panaš-u* is related to the verb *panėšti į ką* ‘resemble sb/sth’ (SEJL 2007, 423) and its meaning of similarity is apparent when it functions predicatively: *Tai panašu į vasarą* ‘It is **similar** to/**looks like** summer’. The translational correspondences of *likely* in the parallel corpus ParaCorp<sub>EN→LT→EN</sub> (Šolienė 2012, 2015) allow for another Lithuanian equivalent to be established, namely *tikėtina* ‘believable, likely’, which derives from the verb of cognition *tikėti* ‘believe, expect’.

Although these markers, except for *tikėtina* ‘believable, likely’, have been investigated in both English (Quirk *et al.* 1985; Biber *et al.* 1999; Bamford 2005; Mindt 2011; Besnard forthcoming) and Lithuanian (Tekorienė 1990; Akelaitis 2011; Ruskan 2012), little attention has been paid to their formal and functional features across languages and types of discourse. The aim of the present study is to identify the functional similarities and differences of the English marker *likely* and its Lithuanian equivalents *panašu* ‘likely, it seems’ and *tikėtina* ‘believable, likely’ within the category of epistemicity (Boye 2010, 2012) in academic and newspaper discourse and to examine the links between their functional distribution and the semantic domains they derive from. The markers are compared in terms of their frequency, structural features (Complement-Taking-Predicates (CTPs) followed by *that* or *to*-complement clauses, adverbials), functions, collocational profile and types of discourse. Since they have been mainly considered intra-linguistically, it is necessary to explore their functional variation from a cross-linguistic perspective.

## 2 The category of epistemicity and previous accounts of *likely* and its Lithuanian equivalents

The category of epistemicity relates to the author's justificatory support for the proposition and comprises the subcategories of evidentiality and epistemic modality (Boye 2012, 20). The former refers to the ways the author makes access to information for the proposition, e.g. direct and indirect (reportive, inferential, unspecified) evidence, while the latter covers the meaning areas of modal certainty, probability (strong, weak, unspecified) and epistemic possibility (Boye 2010)<sup>1</sup>. The category of epistemicity shows both internal meaning relationships within each subcategory (e.g. direct and indirect evidence; certainty and probability) and external meaning relationships across the subcategories (e.g. indirect evidence and probability).

In terms of evidential epistemic meaning overlap, the most widely discussed markers are the verbs *must* (van der Auwera & Plungian 1998; Mortelmans 2000; 2010; Ramat & Topadze 2007; Squartini 2008; de Haan 2009; Cornillie 2009; Šinkūnienė & Van Olmen 2012; Fetzer 2014; Marín-Arrese 2015) and *seem* and their cognates in other Germanic and Romance languages (Johansson 2001; de Haan 2007; Aijmer 2009; Usonienė & Šinkūnienė 2013). Close contextual analysis of these verbs reveals that either their evidential or epistemic meaning component is foregrounded depending on the context. For example, the evidential dimension of *must* in academic discourse becomes apparent when the source of evidence is explicitly stated in the context and the verb co-occurs with a subject expressed by a personified non-agentive noun (e.g. *these studies*), stance markers *clearly* and *actually* and argumentative markers *nevertheless* and *thus*, while its epistemic meaning is prominent in contexts without explicit evidence and stance markers (Fetzer 2014, 352). Similarly, Wiemer and Kampf (2012, 15–17) claim that the evidential dimension of epistemic modal markers becomes salient when the source of information can be retrieved from the micro or macro-linguistic context.

The explicitness of evidence in the context of epistemic modal markers is not unanimously treated as proof of their evidential status. In de Haan's view (2009, 272), it is far from clear whether *must* can be treated as evidential in contexts with explicit evidence. The scholar maintains that evidential markers assert evidence for the proposition, while

---

<sup>1</sup> Boye (2012, 19–21) refers to evidence as “epistemic justification” and to authorial commitment as “epistemic support”. The author opts for this terminological choice because these terms contribute to better comprehension of the relationship between evidentiality and epistemic modality within the category of epistemicity, which is defined in terms of “justificatory support”. “Epistemic justification” is further subdivided into “direct, inferential and reportive justification”, whereas “epistemic support” is graded on the epistemic scale as “full support, partial support and neutral support”. The present study will refer to evidential and epistemic modal meanings by using traditional terms as in Boye (2010).

epistemic modal markers evaluate it, and *must* is the verb that evokes the evaluation but not assertion of evidence. Boye (2001, 99) also adheres to the view that the Danish cognate *mätte* primarily encodes epistemic modal meaning. Similarly, Marín-Arrese (2015, 217) claims that *must* should be categorised as an epistemic modal, though “with evidential nuances derived from its conclusional force”. The present study aligns with the view that the explicitness of contextual evidence may serve as a sign of evidential meaning and status of the marker (Mortelmans 2000; Wiemer & Kampf 2012; Fetzer 2014). The distinction between modal and evidential markers based on “evaluation” and “assertion” of evidence seems to be quite complex. In contrast to reportive markers, inferential evidential markers contribute to both evaluation and assertion of evidence.

The adjective and adverb *likely* has been primarily viewed as a marker of epistemic modality expressing truth/knowledge (Quirk *et al.* 1985, 1224), certainty/likelihood (Biber *et al.* 1999, 673; Biber 2006, 93) or probability calculated from “solid impartial evidence” (Bamford 2005, 22). The evidential basis of *likely* is mentioned in Besnard (forthcoming), who claims that probability denoted by *likely* derives from “observation”. On the scale of epistemic certainty it has been assigned a medium degree of probability (Quirk *et al.* 1985; Huddleston & Pullum 2002; Carretero 2002; Marín-Arrese 2009). Another important meaning component of *likely* as well as other epistemic modal markers is (inter)subjectivity. As Marín-Arrese (2007, 791; 2009, 246) reports, markers of epistemic modality encode only the author’s subjective attitude towards the proposition and leave the author’s perspective implicit, while evidential markers may encode both subjective and intersubjective attitudes and make the author’s perspective more explicit. The (inter)subjective dimension of *likely* has received a number of interpretations. Wierzbicka (2006, 269) states that in contrast to *probably*, the adverbial *likely* acquires a subjective tone because it modifies propositions that cannot be verified and contain the author’s subjective impressions. The scholar makes similar assumptions about the subjective adjectival use of *likely*. Following Nuyts (2001), the adjectival *likely* should be qualified as an intersubjective marker indicating the shared nature of evidence. However, Carretero (2002, 21) assigns the adjectival *likely* to neutral markers, which do not indicate whether commitment is authorial or shared by other people.

The syntactic status of *likely* has also received some consideration in the literature. The construction *likely to* is considered within the group of marginal auxiliary verbs expressing some degree of doubt (Quirk *et al.* 1985, 236) or adjectives controlling a *that* or *to*-infinitive complement clause (Biber *et al.* 1999, 673; 717). The adverbial *likely* is found within the group of content disjuncts (Quirk *et al.* 1985, 143) or modal adjuncts denoting medium degree of probability (Huddleston & Pullum 2002, 768). Biber *et al.* (1999, 854) emphasise that *likely* is used adverbially when it collocates with the

modifiers *very, quite, most*. It should be noted that *likely* is rarely used as an adverb in British English and Australian English, except for the mentioned collocations; its adverbial use is more common in American English (Wierzbicka 2006, 268).

Similarly, the Lithuanian markers *panašu* ‘likely, it seems’ and *tikėtina* ‘believable, likely’ can be used as adjectives followed by a *that* or *to*-infinitive clause and thus be qualified as CTPs or they may acquire adverbial status when used parenthetically or in responses to questions. Most attention has been devoted to the adjectival use of *panašu* ‘likely, it seems’, which has been regarded as a marker of likelihood (Tekorienė 1990, 136; Akelaitis 2011). Its adverbial status and evidential epistemic meaning overlap have been discussed in fiction and academic discourse (Ruskan 2012, 2013). Despite the fact that there are a number of studies dealing with the functions and syntactic status of adjective-based CTPs in Lithuanian (Usonienė 2012, 2013, 2015; Ruskan 2012, 2013, 2015), none of the studies have addressed the distribution and functions of *tikėtina* ‘believable, likely’ in present-day Lithuanian.

### 3 Data and methods

The methodology adopted in the current research is corpus-based. The data were obtained from the academic and newspaper sub-corpora of the Corpus of Contemporary American English (COCA), Corpus of Academic Lithuanian (CorALit) and the newspaper sub-corpus of the Corpus of the Contemporary Lithuanian Language (CCLL). In order to compare the overall frequencies of the markers in different types of discourse in the COCA and CCLL, reference was made to the sub-corpora of magazines, fiction and spoken discourse. To complement findings of the monolingual corpora, data were also drawn from the bidirectional parallel corpus ParaCorp<sub>EN→LT→EN</sub> (Šolienė 2012, 2015). The size of the corpora and sub-corpora relevant to the study is provided in Table 1.

Discourse	Corpora	Number of words
EN Academic	COCA	103,421,981
LT Academic	CorALit	8,670,613
EN News	COCA	105,963,844
LT News	CCLL	42,836,803
Fiction	ParaCorp <sub>EN→LT→EN</sub>	4,921,627

Table 1. The size of the corpora and sub-corpora

Both the academic sub-corpus of the Corpus of Contemporary American English and Corpus of Academic Lithuanian comprise texts from a variety of academic disciplines,

such as the humanities, medicine/biomedicine, the physical sciences, the social sciences and the technological sciences. The former includes texts published between 1990 and 2015 and the latter contains texts from the years 1999 to 2009. The newspaper sub-corpora of the Corpus of Contemporary American English and Corpus of Contemporary Lithuanian Language comprise texts from daily American (e.g. *The New York Times*, *Atlanta Journal-Constitution*, *The Washington Post*, *The Houston Chronicle*, *The Denver Post*, etc.) and Lithuanian central and local newspapers (e.g. *Lietuvos Rytas*, *Lietuvos Aidas*, etc.). The former includes texts from the years 1990 to 2015, the latter from 1990 to 2008 (see Utkā *et al.* 2012). The bidirectional parallel corpus ParaCorp<sub>EN→LT→EN</sub> consists of translations of English and Lithuanian original fiction texts. As the size of the corpora and sub-corpora differs, the quantitative findings are discussed with reference to the normalised frequency per 10,000 words. More detailed quantitative and qualitative analysis of *likely*, *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ was based on their use as CTPs followed by a *that* or *to*-complement clause (1)–(2) and adverbials (3)–(4) in academic and newspaper discourse:

- (1) *But given the weaknesses of both Labour and the Tories, the Liberal Democrats are **likely to** be part of the next government no matter who is prime minister.* (COCA, ND)
- (2) *After all, if a person is researching vaccination, it is highly **likely that** they have already come across those in an Internet search.* (COCA, AD)
- (3) *More deals will get done and sale prices for buildings **likely** will go up.* (COCA, ND)
- (4) <...> *autorius, **panašu**, pamiršo savo paties žodžius <...>* (CorALit, H)  
‘<...> the author, **it seems**, has forgotten his own words <...>.’

The query was made for *likely* directly followed by *to* or *that*-complement clauses as in (1)–(2) or separated from the complement clause by two, three or four words:

- (5) *Those from well-off families are eight times more **likely than** those in poverty **to** earn a bachelor’s degree by age 24–77 percent vs. 9 percent, one study found.* (COCA, ND)

The markers were qualified as adverbials when they were syntactically independent from the host clause they modified and showed positional flexibility (Dehé & Kavalova 2007; Brinton 2008). Functioning as adverbials, they can occur in clause-initial, -medial or -final position. The adverbial use of *likely* was identified by looking at its colligations. The query was made for *likely* following all types of verbs and cases where *likely* did

not display adverbial use were eliminated. In Lithuanian, the adverbial use of *tikėtina* ‘believable, likely’ and *likely* was identified by manual analysis. The study excluded cases where the markers under study functioned as predicatives (6)–(7) or attributes (8):

- (6) *Both sides say that without a permanent truce, another round of war is likely* (COCA, ND).
- (7) *Ar tai gali būti panašu į realybę?* (CorALit, P)  
‘Can it be **similar** to reality?’
- (8) *Having described the market, the analysis then seeks to evaluate the transaction’s likely effect within that market.* (COCA, AD, Soc)

The structural and functional distinction between the CTP and predicative use of the markers becomes especially apparent in the case of the Lithuanian neuter adjective *panašu* ‘likely, it seems’. When this adjective does not take a *that*-clause as its complement, it denotes similarity, as is apparent from (7).

## 4 Findings

This section provides an overview of the overall frequencies of *likely*, *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ in the COCA, CorALit, CCLL and ParaCorp<sub>EN→LT→EN</sub>, their patterns of use, functional distribution and collocational profiles.

### 4.1 Overall frequency and patterns of use

The overall frequencies of *likely* in the COCA and ParaCorp<sub>EN→LT→EN</sub>, given in Table 2, show that it is most commonly used in academic and journalistic discourse (magazines and newspapers). The tendency of *likely* to occur in formal or semi-formal registers is also emphasised by Bamford (2005, 20), who mentions that *likely* is frequent in the written part of the British National Corpus and academic lectures found in the Michigan Corpus of Academic Spoken English. The overall frequencies of the Lithuanian markers in the CorALit, CCLL and ParaCorp<sub>EN→LT→EN</sub>, provided in Table 3, indicate that *tikėtina* ‘believable, likely’ is also most frequently used in academic discourse, and in newspaper discourse it is the third most frequent marker. However, the distribution of *panašu* ‘likely, it seems’ varies across types of discourse, which will be explained by its functional profile discussed in 4.2. It is most frequent in fiction and spoken Lithuanian and the third most common marker in newspapers and magazines. The frequencies of *likely* and *tikėtina* ‘believable, likely’ in the ParaCorp<sub>EN→LT→EN</sub> confirm the tendency of the markers to occur primarily in formal registers. Both markers are as infrequent in English and Lithuanian original fiction texts and their translations as in fiction in the

monolingual corpora (the COCA and CCLL). *Panašu* ‘likely, it seems’ in the parallel corpus is almost as frequent as in newspapers and magazines.

<b>Discourse</b>	<b>COCA, fr/10,000</b>
academic	3.66
magazines	2.24
newspapers	1.89
spoken	1.10
fiction	0.62
	<b>ParaCorp<sub>EN→LT→EN</sub>, fr/10,000</b>
fiction (EN → LT)	0.54
fiction (LT → EN)	0.49

Table 2. The normalised frequency of *likely* in the COCA and ParaCorp<sub>EN→LT→EN</sub>

<b>Discourse</b>	<b>CorALit, fr/10,000</b>	
	<i>tikėtina</i> ‘believable, likely’	<i>panašu</i> ‘likely, it seems’
academic	0.56	0.07
	<b>CCLL, fr/10,000</b>	
magazines <sup>2</sup>	0.40	0.20
newspapers	0.30	0.20
fiction	0.14	0.40
spoken	0.04	0.44
	<b>ParaCorp<sub>EN→LT→EN</sub>, fr/10,000</b>	
fiction (LT → EN)	0.06	0.23
fiction (EN → LT)	0.16	0.29

Table 3. The normalised frequency of *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ in the CorALit, CCLL and ParaCorp<sub>EN→LT→EN</sub>

Quantitative analysis of the CTP use of *likely*, *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’, displayed in Table 4 below, shows some cross-linguistic differences in the complement clauses preferred in the two languages. In English, *likely* controls a *to*-complement clause more frequently than a *that*-complement clause, whereas in Lithuanian both *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ are followed mainly by a *that*-clause.

<sup>2</sup> This sub-corpus also comprises autobiographies, memoirs, diaries, etc.



	Academic discourse CTPs				Newspaper discourse CTPs			
	that-clause		to-clause		that-clause		to-clause	
	rf	f/10,000	rf	f/10,000	rf	f/10,000	rf	f/10,000
<i>likely</i>	2303	0.22	24509	<b>2.37</b>	503	0.05	10523	<b>0.99</b>
<i>tikėtina</i> 'believable, likely'	359	<b>0.41</b>	7	0.00	1014	<b>0.24</b>	12	0.00
<i>panašu</i> 'likely, it seems'	40	0.05	–	–	449	0.10	0	0.00
<b>Total</b>	2702	0.68	24516	2.37	1966	0.39	10535	0.99

Table 4. Complement patterns and their frequency in the academic and news sub-corpora and corpora

The common use of the complementation pattern with a *to*-infinitive clause, referred to as subject-to-subject raising construction (Biber *et al.* 1999, 731; de Haan 2007, 133–134), is also apparent in the use of the verbs *seem* and *appear* (Usonienė 2003; Aijmer 2009; Usonienė & Šinkūnienė 2013), evidential passive matrices *be thought to*, *be said to*, etc., (Noël 2002) and their equivalent constructions in Danish (Ørnsnes 2011) and Dutch (Coleman & Noël 2012, 22–23). In Lithuanian, few cases of the CTPs *tikėtina* 'believable, likely' and *panašu* 'likely, it seems'<sup>3</sup> were attested in this complementation pattern:

- (9) *Anot ekspertų, šiemet tikėtina sulaukti virusų, skirtų mobiliesiems įtaisams, kišeniniams kompiuteriams bei naujiems telefonams.* (CCLL, ND)  
 'According to experts, this year we are **likely to come across** viruses for mobile applications, handheld PCs and new smartphones.'
- (10) *Su ilgesiu laukiame šilumos ir grįžtančių paukščių. Panašu šiais metais sulaukti ir daugiau naujo.* (CCLL, Mag)  
 'We are longing for warmer weather and birds returning. This year we are **likely to experience** something new.'

A complementation pattern with the infinitive clause was also found in the use of the participial *-ma* construction, as shown in Usonienė (2016, 36):

- (11) <...> *Lietuvoje parengtų slaugos specialistų kvalifikacija užsienio šalyse įvertinama puikiai. Neseniai iš Vokietijos, kur atliko praktiką privačioje klinikoje, grįžo trys*

<sup>3</sup> It was found only in magazines but not newspaper discourse.

*studentės. Jos parsivežė garantinius raštus, kad, baigusios studijas, yra laukiamos sugrįžti nuolatiniam darbui.* (CCLL, ND)

‘Nursing qualifications obtained in Lithuania are highly valued abroad. Recently, three students, who have completed their apprenticeship in a private clinic in Germany, have come back with a letter confirming that after graduation they are **expected to be offered** a permanent position.’

In both languages, the CTP pattern with a *to* or *that*-complement clause is more frequent than adverbial use of the markers, the percentage of which is provided in Table 5. The CTPs followed by a complement clause, except for *tikėtina* ‘believable, likely’, are more frequent in academic discourse than newspapers, which may be determined by the formality of register. In newspaper discourse, CTPs *likely* and *panašu* ‘likely, it seems’ make up 55% of the overall use of the markers. The CTP *likely* occurs in the complementation pattern less frequently in newspapers than in academic discourse because in the former the adverbial *likely* is quite frequent; it makes up 22% of the overall use of the marker. However, unlike in English, the less frequent use of the CTP *panašu* ‘likely, it seems’ in the complementation pattern is not compensated by its adverbial use. In newspapers, it is used quite frequently as a marker of comparison, as illustrated in (7) in Methods and Data.

	CTP %		ADV %	
	ND	AD	ND	AD
<i>likely</i>	55	71	22	12
<i>tikėtina</i> ‘believable, likely’	79	75	7	10
<i>panašu</i> ‘likely, it seems’	55	67	4	7

Table 5. Percentage of CTP and adverbial distribution in academic and newspaper discourse

Although *likely* functions as an adverbial in standard British English when it collocates with the modifiers *most*, *more*, *very*, in informal American English the collocations are not a defining criterion of its adverbial use (OALD 2010, 892). The data from the corpora show that even in formal American English *likely* may be used as an adverbial without the modifiers mentioned above. In most cases CTPs and adverbials realised by *likely*, *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ do not show marked functional differences as shown in 4.2. Adverbials may take scope over a phrase (12)–(13) or occur in responses to questions (14)–(15):

(12) <...> groups must organize and focus their struggle for change on some target, **most likely** the state or the regime. (COCA, AD, H)

- (13) *Vieni svarsto apie Konstitucijos ratifikavimą be Prancūzijos (tikėtina, tuomet ir be Didžiosios Britanijos), kiti – apie ES plėtrą.* (CCLL, ND)  
 ‘Some are considering the ratification of the Constitution without France (**likely** then also without Great Britain), while others are thinking about the expansion of the EU.’
- (14) <...> *wonder what is going to happen to them. Nothing, most likely.* (COCA, ND)
- (15) *Ar gali 71 Seimo nario koalicija sukelti Vyriausybės griūtį? Visai tikėtina.* (CCLL, ND)  
 ‘Can a coalition of 71 MPs lead to the collapse of the Government? **Quite likely.**’

Further functional distribution of CTP and adverbial use of the markers under study is illustrated in 4.2.

#### 4.2 Functional distribution

The quantitative and qualitative findings reveal that the closest cross-linguistic CTP and adverbial equivalents are *likely* and *tikėtina* ‘believable, likely’ because they are both most frequent in formal registers, as shown in 4.1, and display similar collocational profiles and contexts of use:

- (16) *Historically, wars have been fought between states in accordance with the rules of war outlined in the Geneva Conventions. In those conventional wars casualties were among combatants rather than civilians. In contrast, today’s wars are **more likely** to be within states where civilians, not combatants sustain the most casualties.* (COCA, ND)
- (17) *Creativity comes in spurts, and it is **likely** that you **will have** swarms of ideas at just the time that you are too busy to work on them. If you jot them down for later reference, you may find, by the time you get around to them, that they have developed themselves in your subconscious.* (COCA, AD, Soc)
- (18) ***Tikėtina**, kad stabili dauguma Seime ir Vyriausybėje **leis** dabartiniam ministrui įgyvendinti galiojančius įstatymus, **tobulins** ir **priims** naujus teisinius aktus.* (CCLL, ND)  
 ‘It is **likely** that the majority in the Seimas and the Government **will let** the minister implement valid laws and **improve** and **accept** legal acts.’
- (19) ***Labiausiai tikėtina**, kad ateityje didžiausių ekspertinių žinių bagažu **disponuos** geriausiai ES interesams atstovaujančios verslo grupės.* (CorALit, Soc)  
 ‘It is **most likely** that in the future business groups that best meet the EU’s interests **will have** considerable professional expertise.’

In the examples above, the English and Lithuanian CTPs express the degree of the author’s commitment to the proposition, i.e. they indicate that the states or events described are probably true. Since they assess the chances of the proposition being true without evoking any particular evidence, they are qualified as markers of epistemic modality. The epistemic dimension of *likely* and *tikėtina* ‘believable, likely’ is signalled by their collocations with the modifiers *more* (16) and *most* (19) that denote gradability (Usonienė 2002, 145–146; Bamford 2005, 24), the frequencies of which are provided in Tables 6 and 7. In newspaper discourse, modifiers of gradability make up 34% of the overall CTP use of *likely*, whereas in academic discourse they make up as much as 54% of its overall CTP use, which reflects the argumentative nature of discourse in which degrees of likelihood are specified. In Lithuanian, the modifiers of the CTP *tikėtina* ‘believable, likely’ make up 29% of its overall CTP use in newspapers and 22% in academic discourse. A lower percentage of the modifiers of gradability collocating with *tikėtina* ‘believable, likely’ may be explained by the use of alternative means specifying degrees of probability, such as the adverbials *greičiausiai* ‘most likely, probably’, *veikiausiai* ‘most likely, probably’, *tikriausiai* ‘most likely, probably’, *greičiau* ‘more likely’ and *veikiau* ‘more likely’<sup>4</sup>. In contrast to *likely* and *tikėtina* ‘believed, likely’, *panašu* ‘likely, it seems’ co-occurs with modifiers of gradability (*more* and *very*) in few cases. In newspapers, these collocations make up 4% of the overall CTP use of *panašu* ‘likely, it seems’, while in academic discourse no cases with the modifiers were attested.

Modifiers	CTP <i>likely</i> + <i>to</i> -clause/ <i>that</i> -clause	
	ND	AD
<i>more</i>	2314	10020
<i>less</i>	766	2775
<i>very</i>	52	188
<i>most</i>	543	1171
<i>least</i>	78	203
	<b>3753 (34%)</b>	<b>14357 (54%)</b>

Table 6. Most frequent modifiers of the CTP *likely* + *to*-clause/*that*-clause in newspaper and academic discourse

<sup>4</sup> Originally the adverbials *greičiausiai*, *veikiausiai*, *tikriausiai* ‘most likely, probably’ are the superlative degree forms of *greitai* ‘fast’, *veikiai* ‘fast, strongly’ and *tikrai* ‘exactly’, and *greičiau* and *veikiau* ‘more likely’ are the comparative degree forms of *greitai* ‘fast’ and *veikiai* ‘fast, strongly’.

Modifiers	CTP <i>tikėtina</i> 'believable, likely' + <i>that</i> -clause	
	ND	AD
<i>mažai</i> 'little'	101	19
<i>labai</i> 'very'	68	27
<i>labiausiai</i> 'most'	52	11
<i>visai</i> 'quite'	41	4
<i>labiau</i> 'more'	37	18
	<b>299 (29%)</b>	<b>79 (22%)</b>

Table 7. Most frequent modifiers of the CTP *tikėtina* 'believable, likely' + *that*-clause in newspaper and academic discourse

As can be seen in (17)–(19), *likely* and *tikėtina* 'believable, likely' modify propositions containing future-tense forms *will have* (17), *leis* 'will let', *tobulins* 'will improve', *priims* 'will accept' (18) and *disponuos* 'will have' (19), which express epistemic prediction. On the one hand, the co-occurrence of future forms with *likely* and *tikėtina* 'believable, likely' may emphasise the synergy of the markers, the purpose of which is to strengthen the effect of epistemic qualification; on the other hand, the functions of *likely* and *tikėtina* 'believable, likely' and the modal verb *will* and its equivalents in Lithuanian may display different shades of epistemic qualification. While the latter express epistemic prediction, the former qualify its degree of certainty (cf. Hoye 1997, 118). Both means highlight the hypothetical nature of the situations illustrated. The meaning of epistemic probability is also highlighted by the Lithuanian translational correspondences of *likely* in the ParaCorp<sub>EN→LT→EN</sub>, such as the adverbials *greičiausiai* 'most likely, probably', *veikiausiai* 'most likely, probably', *tikriausiai* 'most likely, probably', *gal* 'perhaps', *ko gero* 'quite likely', *greičiau* 'more likely', the verb *galėti* 'can, may' and the CTP or adverbial *tikėtina* 'believable, likely' as shown below:

(20) *The Prime Minister, well known to be strong on personal friendship and loyalty, weak on political instinct, was **likely** to hang on to Garmony until he was forced out.* (ParaCorp<sub>EN→LT</sub>)

‘Premjeras, kuris garsėjo lojalumu ir stipriu polinkiu į asmenines draugystes, bet silpna politine intuicija, **greičiausiai** iš paskutiniųjų įsikibęs laikysis Garmonio tol, kol pats bus išsiūstas.’

(21) *It is far **more likely**, in fact, that the photographs will be snipped down the middle – indeed, knowing David, they already have been, last night.* (ParaCorp<sub>EN→LT</sub>)

‘Tiesą pasakius, daug **labiau tikėtina**, kad nuotraukos bus sukarpytos perpus – kadangi gerai pažįstu Deividą, tai bijau, kad jis spėjo jas sukarpyti dar vakar.’

(22) **Quite likely** the person at the next table was a spy of the Thought Police, and **quite likely** he would be in the cellars of the Ministry of Love within three days, but a cigarette end must not be wasted. (ParaCorp<sub>EN→LT</sub>)

‘**Visai gali** būti, kad moteris prie gretimo stalo yra minčių policijos šnipė, taip pat **gali** būti, kad po trijų dienų jis bus Meilės ministerijos rūsiuose, bet cigaretės nuorūkos prarasti nevalia.’

Examples (16)–(19) found in the monolingual corpora and (20)–(22) in the parallel corpus illustrate the contexts in which *likely* and *tikėtina* ‘believable, likely’ do not evoke any specific evidence but express the probability of prediction. However, the data from the newspaper and academic corpora in the two languages show that there may be explicit evidence in the context of *likely* and *tikėtina* ‘believable, likely’ leading the author to a calculation of probability:

(23) *Inscriptional evidence* dates this neutralization from the first century onward; coetaneous or perhaps slightly afterward is the inscriptional *evidence* for the palatalization and assibilation of /kj/, e.g., *TERCIAE* for *TERTIAE* (179 C.E.) (Vnnen 1988:109). **Given these observations**, it is **likely** that /dj/ and then /gj/ were reducing to /y/ during the same time that /kj/ was adopting a palatalized sibilant articulation (COCA, AD, H)

(24) Yet the survey accounts only for people 16 and older, **and evidence suggests** that young people in poorer areas are **increasingly likely** to carry knives, and **increasingly likely** to use them. (COCA, ND)

(25) For example, **the report indicates**, along much of the East and Gulf coasts, including the Chesapeake Bay, it is most **likely** that there will be about a two-foot rise by the year 2100. But there is a 1 percent chance the sea level will rise four feet by then and 15 feet by 2200. (COCA, ND)

(26) **Tyrimo rezultatų analizė** leidžia teigti, jog 67% respondentų ekologiškus maisto produktus labiausiai linkę pirkti prekybos centruose. **Tikėtina**, kad spartėjant gyvenimo tempui, vartotojui svarbu viską, ko jam reikia, rasti vienoje vietoje, todėl dauguma apsiperka prekybos centruose. (CorALit, B)

‘**Analysis of the results** makes it possible to claim that 67% of respondents tend to buy eco food in shopping malls. It is **likely** that with the increased pace of life users may want to find everything they need in one place, so naturally most people go shopping in malls.’

In the examples above, the author resorts to written sources such as other studies, research and reports that ground the evaluation of probability. In (23) and (24), there is

explicit reference to evidence that justifies the proposition. As shown in Fetzer (2014), the co-occurrence of the verbs *seem*, *appear*, *must* and *may* with the noun *evidence* highlights their evidential contexts of use. In (23)–(26), the evidence is both asserted and evaluated and thus this use of *likely* and *tikėtina* ‘believable, likely’ can be treated as evidential epistemic. It is not possible to discern the dominance of one or another meaning component since they co-occur. Probability can be calculated not only from external sources of evidence, as in (23)–(26), but also from the author’s internal sources of evidence, marked by the *if*-clause, *as/since/because*-clause and the connectives *then*, *therefore* as in the following:

(27) *If an ensemble rehearsal produces average intensity levels above allowable exposure levels, it’s very likely that a large percentage of time is spent producing high intensity levels, and little or no time is spent producing music at low levels.* (COCA, ND)

(28) *As the government banned broadcasts of the rebel’s declaration on state radio and television, it is likely that most of the 20 million inhabitants of the world’s fourth largest island were unaware that a coup attempt had taken place.* (COCA, ND)

(29) *Mano žiniomis, dabar tokį politinį darinį palaikančiųjų sumažėjo, bet jų vis dar daugiau nei remiančiųjų buvusią koaliciją. Tad tikėtina, kad socialliberalų ir socialdemokratų sąjunga turėtų pelnyti didesnę žmonių pasitikėjimą nei ankstesnioji.* (CCLL, ND)

‘To my knowledge, the number of supporters of this political entity has decreased, but there are still more of them than there are supporters of the former coalition. **Therefore** it is **likely** that the union of Social Liberals and Social Democrats will earn the people’s trust more than the previous one.’

(30) *Lietuvoje ekonominio pobūdžio <...> sąjungų yra kelios dešimtys, todėl tikėtina, kad ateityje jų skaičius didės.* (CorALit, Soc)

‘In Lithuania, there are dozens of unions of an economic character <...>, **therefore** their number in the future is **likely** to increase.’

The latter linguistic cues emphasise the base of the authors’ inferential reasoning (Mortelmans 2000, 141–143; Aijmer 2008, 67; Alonso-Almeida & Cruz-García 2011, 69–70). The external and internal sources of evidence exemplified in (23)–(30) qualify *likely* and *tikėtina* ‘believable, likely’ as markers of strong probability because in the light of evidence provided alternative points of view hardly seem possible. The explicitness of evidence is also in line with Bamford’s (2005, 19) claim that “impartial evidence”, on the basis of which the degree of probability is calculated, is implicit in such genres

as academic lectures and more explicit in research articles. The explicitness of evidence and consequently its shared status unavoidably bring up the evidential dimension of the markers (Marín-Arrese 2009), which may be determined to some extent by the genres of academic and newspaper discourse, in which the author is particularly concerned with grounding his/her arguments. The diminished degree of the author's commitment can be observed when the CTP or adverbial *likely* and *tikėtina* 'believable, likely' co-occur with the markers indicating the author's assumptions:

- (31) *Some psychologists argue that these tools or codes stem from the experience of trauma itself; however, it **seems** more **likely** that these tools and codes are the conventions by which trauma is recognized in literary representation in the West (see among others, BenEzer and Herman). This is an undecidable aporia.* (COCA, ND)
- (32) *Reliance on self-reported data does, however, carry a risk of misclassification, because individuals with symptoms **may** be more **likely** to report involuntary tobacco exposure than those without symptoms. A second limitation is that the study is cross-sectional, and we have no data on the duration of tobacco exposure.* (COCA, AD, M)
- (33) *Those who lose their subsidy **would likely** be unable to afford coverage, raising the number of uninsured.* (COCA, ND)
- (34) *“Todėl visai **tikėtina**, kad eksporto augimo tempai **gali** kiek sumažėti”, – prognozuoja G. Nausėda. (CCLL, ND)*  
 “‘Therefore it is quite **likely** that the export growth rate **may** lessen slightly’, G. Nausėda predicts.’
- (35) *Jeigu bendradarbiavimo idėja su lenkais **įgautų** konkretesnį pagreitį, **tikėtina**, **paspartėtų** elektros tilto tiesimas, Lietuvai **atsivertų** galimybė pardavinėti naujo reaktoriaus gaminamą elektrą.* (CCLL, ND)  
 ‘If the idea of cooperation with the Poles **were developed** more quickly, it is **likely** the electricity bridge **would be created faster**, and Lithuania **would have** the opportunity to sell electricity produced by the new reactor.’

The CTP *likely* (31)–(32) collocates with the verbs *seem* and *may* and the adverbial *likely* (33) co-occurs with the verb *would*. Similarly, the reduced degree of probability denoted by the CTP (34) and adverbial (35) *tikėtina* 'believable, likely' is marked by the modal verb *gali* 'can, may' and the subjunctive forms *įgautų* 'would acquire', *paspartėtų* 'would be faster', *atsivertų* 'would open'. Thus contextual elements play a great role in highlighting the functional profile of *likely* and *tikėtina* 'believable, likely'. The CTP



and adverbial *panašu* ‘likely, it seems’ displays a similarity with *likely* and *tikėtina* ‘believable, likely’ in expressing the author’s degree of weak probability:

- (36) *Remdamosi konvencionalumo įverčiais galime manyti, kad jautresnės yra motinos, kurioms priimtinesnės visuomenės vertybės, svarbesnės socialinės normos, o socialinė adaptacija geresnė. Jautrios motinos, panašu, gali būti šiek tiek banalesnės, konformistiškesnės, o nejautrios – ne tokios praktiškos ir mažiau atsižvelgti į vyraujančias socialines normas.* (CorALit, Soc)

‘On the basis of conventional norms we can think that those mothers who accept society’s values, for whom social norms are more important and social adaptation is better, are more sensitive. Sensitive mothers, **it seems, can** be slightly **more banal, more conformist**, while insensitive ones are **not so practical** and **care less** about social norms.’

It occurs in contexts where the author seems to have some doubts as to the veracity of the information conveyed in the proposition and the degree of probability is quite low. Although the author’s source for the claim is pointed out, namely *remdamosi konvencionalumo įverčiais galime manyti* ‘on the basis of conventional norms we can think’, the comparative constructions *banalesnės* ‘more banal’, *konformistiškesnės* ‘more conformist’, *ne tokios praktiškos* ‘not so practical’, *mažiau atsižvelgti* ‘care less’ foreground the evaluation of evidence, i.e. the meaning of epistemic probability. The lower degree of probability is also strengthened by the modal verb *gali* ‘can, may’. However, in a number of cases the Lithuanian CTP and adverbial *panašu* ‘likely, it seems’ displays a different semantic profile from *likely* and *tikėtina* ‘believable, likely’, which may reflect their different semantic development. In contrast to *likely* and *tikėtina* ‘believable, likely’, it may denote inferences based on both perceptual or conceptual evidence as in the examples below:

- (37) *Kauniečiai padrikai gynėsi, o atakos nebuvo logiškos. Panašu, kad per paskutines penkias minutes, kai skirtumas jau buvo šoktelėjęs iki dvidešimties taškų, jau niekas nebetikėjo, jog rungtynes galima išgelbėti.* (CCLL, ND)

‘The Kaunas defence **was erratic** and their attacks **were not logical**. **It seems** that during the last five minutes, when the difference jumped to around twenty points, nobody **expected** the match could be saved.’

- (38) *Tačiau, panašu, kad labai daug tiesos Prezidento žodžiuose, kuris prabilo apie aiškius oligarchijos elementus.* (CCLL, ND)

‘However, **it seems there is a lot of truth** in the President’s words, where he spoke about **clear oligarchic elements**.’

(39) *Tarp valdančiųjų partijų netgi egzistuoja neoficialus susitarimas – kokiai partijai kokie administracijos postai atiteks. Panašu, kad būsimieji konkursai – tik formalumas, nes postai seniai pasidalyti. Valdinkai nebetilps į Savivaldybę?* (CCLL, ND)

‘The governing parties even have an informal agreement – they know which party will take which administrative posts. **It seems** that future applications for positions are just a formality, as all posts have already been distributed. **Will there be any place for civil servants in the municipality?**’

(40) *Gdytis, nenustačius diagnozės, labai pavojinga. Panašu, kad opa yra kraujagyslinės kilmės, todėl patariu kreiptis į kraujagyslių chirurgą.* (CCLL, ND)

‘It is very dangerous to start treatment without a diagnosis. **It seems** that the ulcer is related to a blood vessel, so **I** would **advise** you to contact a blood vessel surgeon.’

In (37), the inference is drawn from visual evidence and (38) from auditory data, whereas in (39) and (40) the inference is based on conceptual sources. In the examples above, the CTP *panašu* ‘likely, it seems’ is not interchangeable with *tikėtina* ‘believable, likely’ due to the subjective interpretation of perceptual and conceptual evidence, which is foregrounded by the evaluative phrases *padrikai gynėsi* ‘defence was erratic’, *nebuvo logiškos* ‘were not logical’, *nebetikėjo* ‘nobody expected’, *labai daug tiesos* ‘a lot of truth’, *aiškius oligarchijos elementus* ‘clear oligarchic elements’, the rhetorical question (39) and the authorial *patariu* ‘I advise’ (40). The CTP and adverbial *panašu* ‘likely, it seems’ also occurs in contexts containing evaluation when it denotes hearsay:

(41) *Kuršių nerijoje didžioji dalis pinigų atsiduoda ne žuvimis, ne kopu smėliu ar pušų sakais, bet betonu. Beje, betonas, panašu, čia yra atgijęs. Nauji pastatai dygsta greičiau nei grybai.* (CCLL, ND)

‘On the Curonian Spit, the biggest part of the money is not benefiting the fish, or the sand of the dunes or the resin of the pines, but the concrete. By the way, **it seems**, the concrete is being renewed here. New buildings are springing up faster than mushrooms.’

(42) *Socialliberalai nusprendė ginti kariuomenės vado teises. Užkulisuose užvirė tikras žodžių mūšis. Panašu, kad socialliberalai nuogąstauja, kad socialdemokratų gviėšiasi didesnės įtakos nacionalinio saugumo ir užsienio politikos srityse.* (CCLL, ND)<sup>5</sup>

‘The Social Liberals decided to defend the rights of the head of the army. Behind

---

<sup>5</sup> I would like to thank Bert Cornillie for his comments on this example at the 49th Annual Meeting of the Societas Linguistica Europaea (University of Naples Federico II, Naples, 31st August – 3rd September, 2016).

the scenes a real war of words started. **Apparently** the Social Liberals **fear** that the Social Democrats **hanker after more influence** on national security and foreign policy.’

In (41), the author is ironic in saying that in one of the country’s most prestigious seaside resorts, money is benefiting the concrete because investment is being made in the construction of private houses. In (42), the report on the Social Liberals’ and the Social Democrats’ dispute over political influences also conveys evaluation (*nuogąstauja* ‘fear’, *gviešiasi didesnės įtakos* ‘hanker after more influence’). The evidential basis of *panašu* ‘likely, it seems’ is confirmed by its English equivalents in the ParaCorp<sub>EN→LT→EN</sub>:

(43) *And it looked like no one was popping in to express their outraged sympathy.* (ParaCorp<sub>EN→LT</sub>)

‘Buvo **panašu**, kad niekas nė neketina pasirodyti ir pareikšti savo pasipiktinimo bei užuojautos.

(44) *It seemed to be coming from the end of the hallway in front of them.* (ParaCorp<sub>EN→LT</sub>)

‘Buvo **panašu**, kad jis sklinda iš koridoriaus galo, kurio link jiedu ėjo.’

(45) “*Young man, apparently you are new here.*” (ParaCorp<sub>EN→LT</sub>)

‘Jaunuoli, **panašu**, kad jūs čia naujokas.’

In (43), the verb *looked like* points out visual evidence, whereas in (44) the verb *seemed* refers to auditory evidence. The adverbial *apparently* in (45) may be interpreted as an inferential or hearsay marker. Similar to (37)–(42), *likely* and *tikėtina* ‘believable, likely’ are not compatible with the contexts of use in (43)–(45) because they display an objective semantic prosody evoked by external or internal cognitive sources shared with the reader. The subjective semantic prosody of *panašu* ‘likely, it seems’ and its hearsay function explain the infrequent use of the marker in academic discourse and its frequent use in newspapers, fiction or spoken language, which contain more subjectivity and hearsay evidence. The inferential and hearsay functions of the CTP and adverbial *panašu* ‘likely, it seems’ may be explained by the conceptual link of the marker with its core meaning of similarity and appearances. In contrast, *likely* does not evoke the meanings of similarity or appearances to any great extent. In most contexts, it marks a degree of probability derived from external or internal cognitive sources of evidence and displays an objective semantic prosody. In a similar manner to *likely*, the CTP and adverbial *tikėtina* ‘believable, likely’, connected with the semantic domain of belief and expectation, expresses the degree of the author’s commitment to the proposition.

## 5 Concluding remarks

The present study shows that the closest cross-linguistic epistemic equivalents are the English *likely* and Lithuanian *tikėtina* ‘believable, likely’. When used as CTPs and adverbials, they both display similar frequencies, collocational profiles, functions and semantic prosody in academic and newspaper discourse. Although their primary function is to assess the degree of probability in the proposition, they also occur in contexts with explicit evidence shared with the reader. The explicitness of evidence signals both the assertion and evaluation of evidence and thus emphasises the co-occurrence of epistemic and evidential meaning components in the context of the use of the markers. The difference between the CTP use of *likely* and *tikėtina* ‘believable, likely’ lies in their complementation patterns. While the former is commonly followed by a *to*-complement clause, the latter most frequently controls a *that*-clause, which adds to previous studies of diverse modes of expressing epistemicity (Usonienė & Šolienė 2010; Usonienė 2015).

Although the Lithuanian CTP and adverbial *panašu* ‘likely, it seems’ shares the meaning of probability with *likely* and *tikėtina* ‘believable, likely’, it displays a distinct functional semantic profile, which is highlighted by its different frequency, collocational preferences and semantic prosody in academic and newspaper discourse. In contrast to *likely* and *tikėtina* ‘believable, likely’, it may mark inferences drawn from both perceptual or conceptual evidence and acquire a subjective semantic prosody, which motivates its low frequency in academic discourse. Since only *panašu* ‘likely, it seems’ denotes inferences based on perceptual evidence, a conclusion can be drawn that this marker has preserved the original meaning of similarity and appearances to the greatest extent. *Likely* and *tikėtina* ‘believable, likely’ qualify propositions in terms of conceptual evidence that grounds the author’s degree of commitment. *Panašu* ‘likely, it seems’ may also point out hearsay, which explains the high frequency of the marker in newspaper discourse.

The high frequency of *likely* and *tikėtina* ‘believable, likely’ in academic discourse highlights the argumentative dimension of the markers. By expressing a strong degree of probability and making sources of evidence explicit, they strengthen the author’s argumentation. The diminished degree of probability denoted by *likely* and *tikėtina* ‘believable, likely’ may be signalled by their collocations with the verbs *seem*, *may* and *would*. *Panašu* ‘likely, it seems’ serves as a cautious strategy of the author’s argumentation due to its subjective authorial evaluation and lower degree of commitment. In terms of discourse, *likely* and *tikėtina* ‘believable, likely’ are more common in academic discourse than in newspapers, though no discursive differences are found in their functional distribution.

The functional semantic profile of *likely*, *tikėtina* ‘believable, likely’ and *panašu* ‘likely, it seems’ has been compared on the basis of data drawn from both monolingual and parallel corpora, which confirms the effectiveness of combining types of corpora in exploring meanings and functions across languages (Nöel 2002; Simon-Vandenberg & Aijmer 2007). The present study contributes to a better understanding of the evidential epistemic overlap of epistemic modal markers and functional similarities and differences of markers deriving from the same semantic domain and their functional equivalents deriving from a different semantic domain. The study aligns with the view that the evidential reading of a marker is triggered by the retrieval of the source of information from the micro or macro linguistic context (Wiemer & Kampf 2012, 15–17).

### Acknowledgements

The current study has been carried out within the project *Modality and Evidentiality in the Lithuanian Language* (Project No MIP-062/2014) funded by the Research Council of Lithuania. I would like to express my sincere gratitude to Aurelija Usonienė for insightful comments and remarks. All shortcomings or errors are but my own. Thanks are also due to Audronė Šolienė for help in data extraction from the ParaCorp<sub>EN→LT→EN</sub>.

### Abbreviations

AD	Academic discourse
ADV	Adverbial
B	Biomedical sciences
CTP	Complement-taking predicate
EN	English
f	frequency
H	Humanities
LT	Lithuanian
M	Medicine
Mag	Magazines
ND	Newspaper discourse
P	Physical sciences
rf	raw frequency
Soc	Social sciences

## Data sources

- COCA Corpus of Contemporary American English. Available at: <http://corpus.byu.edu/coca/>
- CorALit Lietuvių mokslo kalbos tekstynas (Corpus Academicum Lithuanicum). Available at: <http://www.coralit.lt/>
- CCLL Corpus of the Contemporary Lithuanian Language. Available at: <http://donelaitis.vdu.lt>
- ParaCorp<sub>EN→LT→EN</sub> Bidirectional Parallel Corpus of English and Lithuanian

## References

- Aijmer, Karin. 2008. Modal adverbs in interaction – *obviously* and *definitely* in adolescent speech. *The dynamics of linguistic variation: Corpus evidence on English past and present*. Terttu Nevalainen, Irma Taavitsainen, Päivi Pahta & Minna Korhonen, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company. 61–83.
- Aijmer, Karin. 2009. *Seem* and evidentiality. *Functions of Language* 16 (1), 63–88.
- Akelaitis, Gintautas. 2011. *Panašu* – naujas modalinis žodis? [*Panašu* ‘similar/likely’ – a new modal word?] *Gimtoji kalba* 11, 3–10.
- Alonso-Almeida, Francisco, Laura Cruz-García. 2011. The value of *may* as an evidential and epistemic marker in English medical abstracts. *Studia Anglica Posnaniensia* 46 (3), 59–73.
- Bamford, Julia. 2005. Subjective or objective evaluation? Prediction in academic lectures. *Strategies in academic discourse*. Elena Tognini-Bonelli & Gabriella Del Lungo Camiciotti, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company. 17–29.
- Besnard, Anne-Laure [forthcoming]. *Be likely to* and *be expected to*, epistemic modality or evidentiality? Markers of (non)commitment in newspaper discourse. *Evidentiality revisited: Cognitive grammar, functional and discourse pragmatic perspectives*. Juana Marín-Arrese, Gerda Haßler & Marta Carretero, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan, eds. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Biber, Douglas. 2006. *University language. A corpus-based study of spoken and written registers*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Boye, Kasper. 2001. Evidence for evidentiality in Danish. *Reflections on modality*. Henrik H. Müller, ed. Frederiksberg: Samfundslitteratur. (*Copenhagen Studies in Language*, Vol. 26). 81–127.

- Boye, Kasper. 2010. Semantic maps and the identification of cross-linguistic generic categories: Evidentiality and its relation to epistemic modality. *Linguistic Discovery* 8 (1), 4–22.
- Boye, Kasper. 2012. *Epistemic meaning: A crosslinguistic and functional-cognitive study*. Berlin: de Gruyter Mouton.
- Brinton, Laurel J. 2008. *The comment clause in English*. Cambridge: Cambridge University Press.
- Carretero, Marta. 2002. The influence of genre and register on epistemic modality in spoken English: A preliminary study. *Estudios Ingleses de la Universidad Complutense* 10, 11–41.
- Colleman, Timothy, Dirk Noël. 2012. The Dutch evidential NCI: A case of constructional attrition. *Journal of Historical Pragmatics* 13 (1), 1–28.
- Cornillie, Bert. 2009. Evidentiality and epistemic modality: On the close relationship between two different categories. *Functions of Language* 16 (1), 44–62.
- de Haan, Ferdinand. 1999. Evidentiality and epistemic modality: Setting boundaries. *Southwest Journal of Linguistics* 18, 83–101.
- de Haan, Ferdinand. 2007. Raising as grammaticalization: The case of Germanic SEEM-verbs. *Rivista di Linguistica* 19 (1), 129–150.
- de Haan, Ferdinand. 2009. On the status of ‘epistemic’ *must*. *Studies on English modality. In Honour of Frank Palmer*. Anastasios Tsangalidis & Roberta Facchinetti, eds. Bern: Peter Lang. 261–284.
- Dehé, Nicole, Yordanka Kavalova, eds. 2007. *Parentheticals*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Fetzer, Anita. 2014. Foregrounding evidentiality in (English) academic discourse: Patterned co-occurrences of the sensory perception verbs *seem* and *appear*. *Intercultural Pragmatics* 11 (3), 333–555.
- Hoye, Leo. 1997. *Adverbs and modality in English*. London, New York: Longman.
- Huddleston, Rodney, Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Johansson, Stig. 2001. The English verb *seem* and its correspondences in Norwegian: What seems to be the problem. *A wealth of English. Studies in honour of Göran Kjellmer*. Karin Aijmer, ed. Göteborg: Acta Universitatis Gothoburgensis. 221–245.
- Marín-Arrese, Juana I. 2007. Commitment and subjectivity in the discourse of opinion columns and leading articles. A corpus study. RAEL Volume: 1 Issue: *Different Approaches to Newspaper Opinion Discourse*, 82–98.
- Marín-Arrese, Juana I. 2009. Commitment and subjectivity in the discourse of a judicial inquiry. *Modality in English. Theory and description*. Raphael Salkie, Pierre Busuttil & Johan van der Auwera, eds. Berlin, New York: Mouton de Gruyter. 237–268.

- Marín-Arrese, Juana I. 2015. Epistemicity and stance: A cross-linguistic study of epistemic stance strategies in journalistic discourse in English and Spanish. *Discourse Studies* 17 (2), 210–225.
- Mindt, Ilka. 2011. *Adjective complementation: An empirical analysis of adjectives followed by that-clauses*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Mortelmans, Tanja. 2000. On the ‘evidential’ nature of the ‘epistemic’ use of the German modals *müssen* and *sollen*. *Belgian Journal of Linguistics* 14 (1), 131–148.
- Mortelmans, Tanja. 2010. *Epistemic must and its cognates in German and Dutch. The subtle differences*. Unpublished manuscript. Antwerp: University of Antwerp.
- Mortensen, Janus. 2006. *Epistemic and evidential sentence adverbials in English and Danish: A comparative study* [PhD dissertation]. Roskilde: Roskilde University.
- Noël, Dirk. 2002. *Believe-type matrix verbs and their complements: Corpus-based investigations of their functions in discourse* [PhD dissertation]. Ghent: Ghent University.
- NSOED. 1993. Brown, Lesley, ed. *The new shorter Oxford English dictionary*. Volume 1. Oxford: Clarendon Press.
- Nuyts, Jan. 2001. *Epistemic modality, language and conceptualization: A cognitive-pragmatic perspective*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- OALD. 2010. Wehmeier, Sally, Colin McIntosh, Joanna Turnbull, Michael Ashby, eds. *Oxford advanced learner’s dictionary*. Oxford: Oxford University Press.
- Ørsnes, Bjarne. 2011. Passives and evidentiality: Danish reportive passives and their equivalents in German. *Acta Linguistica Hafniensia* 43 (1), 21–59.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London, New York: Longman.
- Ramat, Anna Giacalone, Manana Topadze. 2007. The coding of evidentiality: A comparative look at Georgian and Italian. *Rivista di Linguistica* 19 (1), 7–38.
- Ruskan, Anna. 2012. Evidential adjectives in Lithuanian academic discourse. *Kalbotyra* 64 (3), 103–123.
- Ruskan, Anna. 2013. *Nemorfologinio evidencialumo raiška ir turinys lietuvių kalboje: bevardės giminės būdvardžiai ir prieveiksmiai*. [The expression and contents of non-morphological evidentiality in Lithuanian: The case of neuter adjectives and adverbs], (ms.). Humanitarinių mokslų daktaro disertacija [PhD dissertation]. Vilnius: Vilniaus universitetas.
- Ruskan, Anna. 2015. Evidential adverbials in Lithuanian: A corpus-based study. *Kalbotyra* 67, 104–130.
- SEJL. 2007. Smoczyński, Wojciech. *Słownik etymologiczny języka litewskiego*. Wilno: Uniwersytet Wileński, Wydział Filologiczny.



- Simon-Vandenberg, Anne-Marie, Karin Aijmer. 2007. *The semantic field of modal certainty. A corpus-based study of English adverbs*. Berlin: Mouton de Gruyter.
- Squartini, Mario. 2008. Lexical vs. grammatical evidentiality in French and Italian. *Linguistics* 46 (5), 917–947.
- Šinkūnienė, Jolanta, Daniel Van Olmen. 2012. Modal verbs of necessity in academic English, Dutch and Lithuanian: Epistemicity and/or evidentiality? *Darbai ir dienos* 58, 153–181.
- Šolienė, Audronė. 2012. Epistemic necessity in a parallel corpus: Lithuanian vs. English. *Multiple perspectives in linguistic research on Baltic languages*. Aurelija Usonienė, Nicole Nau & Ineta Dabašinskienė, eds. Newcastle upon Tyne: Cambridge Scholars Publishing. 10–42.
- Šolienė, Audronė. 2015. Multifunctionality of modal markers: Lithuanian epistemic adverbials *gal* and *galbūt* ‘perhaps/maybe’ vs. their translational correspondences. *Kalbotyra* 67, 155–176.
- Tekorienė, Dalia. 1990. *Bevardės giminės būdvardžiai*. [Neuter adjectives]. Vilnius: Mokslas.
- Usonienė, Aurelija. 2002. Evaluation in perception. *Kalbotyra* 51 (3), 143–152.
- Usonienė, Aurelija. 2003. Extension of meaning: Verbs of perception in English and Lithuanian. *Meaning through language contrast*. Katarzyna M. Jaszczolt & Ken Turner, eds. Amsterdam, Philadelphia: John Benjamins Publishing Company. 193–220.
- Usonienė, Aurelija. 2012. Komplementiniai predikatai ir jų multifunkcionalumas. [Complement-taking predicates and their multifunctionality]. *Darbai ir dienos* 58, 223–232.
- Usonienė, Aurelija. 2013. On the morphosyntactic status of complement-taking predicate clauses in Lithuanian. *Acta Linguistica Hafniensia: International Journal of Linguistics* 45 (1), 73–99.
- Usonienė, Aurelija. 2015. Non-morphological realizations of evidentiality: The case of parenthetical elements in Lithuanian. *Contemporary approaches to Baltic linguistics*. Peter Arkadiev, Axel Holvoet & Björn Wiemer, eds. Berlin: Mouton de Gruyter. 437–463.
- Usonienė, Aurelija. 2016. *Reikšmės pasaulis: tekstynais paremti semantiniai kalbų tyrimai*. [The world of meaning: Corpus-based semantic studies of language]. Vilnius: Akademine leidyba.
- Usonienė, Aurelija, Jolanta Šinkūnienė. 2013. A cross-linguistic look at the multifunctionality of the English verb *seem*. *English modality: Core, periphery and evidentiality*. Juana I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera, eds. Berlin: De Gruyter Mouton. 281–316.

- Usonienė, Aurelija, Audronė Šolienė. 2010. Choice of strategies in realizations of epistemic possibility in English and Lithuanian: A corpus-based study. *International Journal of Corpus Linguistics* 15 (2), 291–316.
- Utkā, Andrius, Mažvydas Petkevičius, Jolanta Kovalevskaitė. 2012. “*Dabartinės lietuvių kalbos tekstyno*” statistinė analizė. [Statistic analysis of the Corpus of the Contemporary Lithuanian Language]. Kaunas: Vytauto Didžiojo universiteto leidykla.
- van der Auwera, Johan, Vladimir Plungian. 1998. Modality’s semantic map. *Linguistic Typology* 2, 79–124.
- van der Auwera, Johan, Ewa Schalley, Jan Nuyts. 2005. Epistemic possibility in a Slavonic parallel corpus – a pilot study. *Modality in Slavonic languages. New perspectives*. Björn Hansen & Petr Karlík, eds. München: Sagner. 201–217.
- Wiemer, Björn, Veronica Kampf. 2012. On conditions instantiating tip effects of epistemic and evidential meanings in Bulgarian. *Slověne: International Journal of Slavic Studies* (2), 5–38.
- Wierzbicka, Anna. 2006. *English: Meaning and culture*. Oxford: Oxford University Press.