

# Multiple Perspectives in Linguistic Research on Baltic Languages



# Multiple Perspectives in Linguistic Research on Baltic Languages

Edited by

Aurelija Usonienė, Nicole Nau  
and Ineta Dabašinskienė

**CAMBRIDGE  
SCHOLARS**

---

P U B L I S H I N G

Multiple Perspectives in Linguistic Research on Baltic Languages,  
Edited by Aurelija Usonienė, Nicole Nau and Ineta Dabašinskienė  
This book first published 2012

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Copyright © 2012 by Aurelija Usonienė, Nicole Nau and Ineta Dabašinskienė and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-3645-1, ISBN (13): 978-1-4438-3645-6

# TABLE OF CONTENTS

Acknowledgements .....	vii
------------------------	-----

Introduction .....	1
Aurelija Usonienė, Nicole Nau and Ineta Dabašinskienė	

## **Part I: Corpus-based Contrastive Studies**

Epistemic Necessity in a Parallel Corpus: Lithuanian vs. English .....	10
Audronė Šolienė	

What Can Child Language Tell Us about Language Development? A Case for Case Study in Lithuanian and Russian .....	43
Maria Voeikova and Ineta Dabašinskienė	

## **Part II: Specialised and Professional Discourse in Lithuanian**

‘Good Language’ and Insecure Speakers: A Study into Metalinguistic Awareness of TV and Radio Journalists in the Context of Language Monitoring in Lithuania .....	76
Loreta Vaicekauskienė	

The Discourse of Food Promotion: A Case Study of Lithuanian Bread Packages.....	104
Jūratė Ruzaitė	

Adverbials as Hedging Devices in Lithuanian Academic Discourse: A Cross-disciplinary Study.....	137
Jolanta Šinkūnienė	

## **Part III: New Perspectives on Grammatical Categories in Baltic Languages**

Reportive Evidentiality and Reported Speech: Is There a Boundary? Evidence of the Latvian Oblique .....	170
Joanna Chojnicka	

The Specifying Existential Sentence Type in Lithuanian: A Problem Statement.....	193
Violeta Kalėdaitė	
Lithuanian Impersonal Modal Verbs REIKĖTI ‘need’ and TEKTI ‘be gotten’: A Corpus-based Study .....	206
Erika Jasionytė	
The Semantics and Distribution of Latvian Reflexive Verbs .....	229
Andra Kalnača and Ilze Lokmane	
Do We Need to Count Numerals as a Part of Speech in Lithuanian?.....	257
Loïc Boizou	
Contributors.....	282
Index.....	285

## ACKNOWLEDGEMENTS

The papers for this volume were selected from participants in the 43rd Annual Meeting of the Societas Linguistica Europaea held in Vilnius, 2010. All contributions were reviewed and commented on by the editors and two external reviewers. The editors wish to express their gratitude to the linguists whose constructive criticism and advice has been important and helpful when improving the content and form of the papers: Juana I. Marín Arrese (Madrid), Johan van der Auwera (Antwerpen), Sturla Berg-Olsen (Oslo), Bert Cornillie (Leuven), Wolfgang U. Dressler (Wien), Jens Normann Joergensen (Copenhagen), Daniel Van Olmen (Antwerpen), Jurgis Pakerys (Vilnius), Vladimir Plungian (Moscow), Mario Squartini (Torino), Maria Voeikova (Saint Petersburg), Bernhard Waelchli (Bern/Stockholm), and Bjoern Wiemer (Mainz).

The organizational and technical aspects of the preparation of the 43rd Annual Conference of the Societas Linguistica Europaea (SLE <http://www.societaslinguistica.eu/>) held at the Faculty of Philology, Vilnius University, 2-5 September 2010 were partially sponsored by the Research Council of Lithuania. Acknowledgements are due to the Research Council of Lithuania (Committee of Humanities and Social Sciences: Project No. LIT-1-14) for providing partial financial support.

Most grateful thanks go to Mark Fearon for his very professional help in the editing of this volume.





## INTRODUCTION

AURELIJA USONIENĖ, NICOLE NAU  
AND INETA DABAŠINSKIENĖ

The Baltic languages have been the object of linguistic research for more than 150 years. In the second half of the 19th century, studying Lithuanian and Latvian was seen as important by the leading linguists in Europe, and both languages played their part in the advancement of linguistic theory and the development of linguistic methodology. Lithuanian became famous for its conservative character, which made it especially useful in historical-comparative Indo-European studies, and August Schleicher's Lithuanian grammar (1856-1857) was a standard work of reference and a model for other grammars. Schleicher's contemporary August Bielenstein's study of Latvian dialects influenced the methodology and terminology of dialectology, a newly developing field at that time. Among other things, Bielenstein coined the term "isogloss" (Händler and Wiegand 1982; see also Kļaviņa 2008). As these examples show, the study of Lithuanian and Latvian was carried out in an international context, where language specific and general linguistic research were in close contact and supplemented each other.

Unfortunately, during the 20th century the situation changed. For some time, the Baltic languages did not attract much notice in linguistic circles outside of their homelands. The only exception was historical-comparative linguistics, where Lithuanian, but not Latvian, remained an object of study. Baltic linguistics in Lithuania and abroad was mostly engaged with diachronic studies and concentrated on topics such as the development of the Baltic accentual system, the reconstruction of the Baltic verbal system, or Baltic dialectology from a historical-comparative perspective. Synchronic linguistics in both Lithuania and Latvia was mainly concerned with the codification of grammar and lexicon as part of normative language planning. The lack of broader perspectives and a more varied approach to language can be partly explained by the geopolitical situation, which cut off linguists from the Baltic countries from the rapid development of the field that their colleagues in the Western World

experienced. Only in the 1990s, after the reestablishment of the independent states, could linguists from Latvia and Lithuania begin to catch up with new developments. At the same time interest in the modern Baltic languages by non-native linguists grew, and opportunities for their study increased. As a consequence, research on Lithuanian and Latvian has broadened considerably in recent years. It now includes sociolinguistic perspectives, first and second language acquisition research, corpus-linguistics, pragmatics, discourse analysis, and other branches and approaches that are well established in contemporary linguistics. Within the synchronic description of grammar, functional-typological as well as formal approaches have been added to the traditional (predominantly normative) way of dealing with grammatical phenomena.

These positive developments, however, did not disseminate new results to a wider linguistic community as much as would be desirable. Very little of modern research carried out in Baltic linguistics is presented internationally. Many linguists from Latvia and Lithuania still publish mostly in their mother tongue, and only a small part of their work has been translated into other European languages while the rest of it is inaccessible to most linguists from other parts of the world. With the present contribution we hope to change this situation and to acquaint an international audience with some recent work in Latvian and Lithuanian linguistics.

The present volume demonstrates the diversity of approaches in current linguistic research on Baltic languages. Data are discussed from different methodological perspectives and findings interpreted in the light of different theoretical approaches. However, several aspects unite the contributions and reflect trends in current research on Lithuanian and Latvian. First, they all show clearly that the period of academic isolation has been overcome: each piece of research presented here is embedded into the international discussion of the respective field or on the topic under consideration. Secondly, the authors are explicit about the methodology they use, including methods of data collection, which have significantly advanced with the application of modern technologies. Linguistic studies based on a researcher's individual introspection have been substituted for more advanced corpus-based and corpus driven analyses which rely upon collective intuition. Although these methods also have a few widely discussed flaws, corpora that are up to modern standards offer researchers huge amounts of authentic language data, the analysis of which enhances the reliability and objectivity of their findings.

The importance of corpora in contemporary linguistic research is reflected in several of the papers: the studies by Šolienė, Šinkūnienė,

Jasionytė, as well as Voeikova and Dabašinskienė are corpus-based in a narrow sense, including quantitative analyses; Chojnicka investigates a self-compiled corpus of parliamentary speeches; Kalėdaitė, and Kalnača and Lokmane draw their examples from online-corpora of Lithuanian and Latvian, respectively. Boizou's critique of the traditional treatment of numerals as a word class has been inspired by problems encountered with tagged corpora of Lithuanian and automatic syntactic analysis. Furthermore, several papers reflect the increasing consideration of the specifics of different styles and genres. Investigations of linguistic phenomena are carried out not for "language X" as an idealized abstract entity, but based on texts of a specific genre. The spectre of genres investigated in this volume reaches from academic discourse (Šinkūnienė) and literary fiction (Šolienė) over Latvian parliament speeches (Chojnicka) to texts found on Lithuanian bread packages (Ruzaitė).

Some new ideas and methods in linguistics were first spread among scholars of English or other foreign languages. Contrastive studies and the comparison of Lithuanian and Latvian data with data from English and other well researched languages in general, have doubtlessly enriched the field and given fresh impetus to traditional Baltic linguistics. This can also be seen in several of the papers collected here (besides the contrastive study by Šolienė, most clearly in the contributions by Kalėdaitė, Ruzaitė, and Šinkūnienė). Furthermore, the papers by Chojnicka, Jasionytė, Kalnača and Lokmane, and Voeikova and Dabašinskienė demonstrate the interplay and mutual benefit of (typological) comparative studies and the description of individual languages.

All the studies in this volume explore new ways of describing Latvian or Lithuanian from a synchronic, non-normative point of view. Boizou openly questions the usefulness of the traditional parts-of-speech, one of the corner stones of traditional grammar, for effective linguistic description. Vaicekauskienė in turn challenges Lithuanian prescriptivism from a sociolinguistic point of view. About 20 years ago Rinholm (1990, 296) concluded his discussion of Lithuanian language planning by stating: "At present the dominant linguists are firmly in control of the strictly formulated and well-guarded standard norms." While it may seem that this statement is still true today, upon closer investigation it becomes evident that a change of attitudes is in progress, and the broader approach to Lithuanian and Latvian that is demonstrated by the contributions to this volume is part of this change.

The chapters of this book are divided into three thematic parts. Part one presents two studies that investigate Lithuanian in contrast to another language, based on parallel or comparable corpora. In this way they

contribute to the understanding of the expression and the acquisition of grammatical categories in general.

In the first paper, *Epistemic necessity in a parallel corpus: Lithuanian vs. English*, **Audronė Šolienė** investigates the expression of epistemic necessity in a corpus of Lithuanian and English novels and their translations into the respective other language. The axes of contrast are laid on the auxiliary and adverb strategies in English as opposed to the corresponding modal verb and adverb/particle strategies in Lithuanian. The author compares the preferred use of certain means of expression for epistemic and non-epistemic modality in translation equivalents throughout the corpus and finds that the use of adverbials for the expression of epistemic modality is significantly more frequent in Lithuanian than in English. Her research results support theses put forward by other scholars about the correlation between the degree of grammaticalization of a modal expression and the frequency of its use with an epistemic meaning. They also fit with the postulation (for example, by van der Auwera, Schalley & Nuyts 2005) of an east-west cline of languages preferring adverbs in the Eastern part of Europe as against a preference for modal verbs in languages of Western Europe.

**Maria Voeikova and Ineta Dabašinskienė** in their article *What can child language tell us about language development? A case for case study in Lithuanian and Russian* pursue the line of research in language acquisition studies which singles out the category of case to be one of the most complex grammatical categories. The acquisition of Russian and Lithuanian as morphologically complex languages shows similar developmental patterns. However, data comparison of one Russian and one Lithuanian child has shown the same order of emergence but a different average distance between forms in the two languages: in Russian this distance is significantly higher. This result corroborates the main hypothesis of the study that case distinctions in Lithuanian are more significant for the expression of syntactic relations and therefore must be acquired earlier than in Russian. It corresponds to the less inflecting status of Russian compared to Lithuanian on the scale of the ideal language types from agglutinating to isolating. These features reveal a more synthetic way of morphological marking in Lithuanian compared to a more analytic marking in Russian. The slight typological difference between the two languages is not the only explanation for the different speed of the acquisition process: the individual differences of the children and the characteristics of the input should also be taken into consideration.

The second part of the volume unites three analyses of professional and specialized discourse from different research perspectives. The first is

a contribution to sociolinguistics and a critical analysis of language planning in Lithuania, the second takes a discourse analytical approach, while the third presents corpus-based research on specialized discourse.

In her paper on *'Good language' and insecure speakers: a study into metalinguistic awareness of TV and radio journalists in the context of language monitoring in Lithuania* **Loreta Vaitėkauskienė** discusses a new issue in the field of critical language planning studies – the impact of restrictive language policy and institutionalized monitoring on the metalinguistic awareness of professional language users. She demonstrates how the dominant ideology of “correct language” influences the self-judgment of speakers and leads to insecurity. The research is based on structured qualitative interviews with a representative sample of Lithuanian TV and radio journalists and presents a comparison of the attitudes of the broadcast media and the official metalinguistic discourse. It shows that although journalists operate with a more elaborated system of values than merely ‘purity’ and ‘correctness’, they abandon their values in the self-ratings and tend to assess their own linguistic behaviour only in terms of prescriptive criteria. Thus, interventions in language processes using powerful legislative arguments make Lithuania a good example of what consequences institutionalized models of prescription can bring in the long term. It can distort ownership relations of standard language and make speakers feel that they can merely produce a restricted standard code, which must be corrected by the experts. And in the worst case scenario an indication of linguistic insecurity can develop even among well-educated and professional members of the speech community. The study can serve as a basis for comparative research in other speech communities with different degrees of institutionalization of language ideologies and language monitoring.

**Jūratė Ruzaitė's** paper on *The discourse of food promotion: a case study of Lithuanian bread packages* deals with the discourses of bread promotion in Lithuania. Food narratives have evolved into complex systems of discourses that reflect national/cultural and global ideologies of food; therefore, investigations of food discourse have become an important area in different interdisciplinary studies that can reveal how the users and producers of such discourses construct shared knowledge about food-related values. The author analyzes the discourse of Lithuanian bread packages by taking into account their verbal and visual aspects from the perspective of the interdisciplinary theory of multimodality. Only an analysis of a whole combination of visual and verbal aspects can lead to an exhaustive account of how bread is perceived culturally, and how it integrates the central features of global food discourses. Interestingly, the

results show that bread packages also serve as institutional narratives, which makes them even more ideology-based. The integrative analysis of bread promotion strategies in one language helps to reveal how cultural myths about food are created and how global ideologies (highlighted in other investigations) become salient in cultural ideologies.

**Jolanta Šinkūnienė** addresses issues of the expression of author stance in academic discourse in her paper *Adverbials as hedging devices in Lithuanian academic discourse: a cross-disciplinary study*. Investigations of author stance expression are one of the prominent trends in current research on academic discourse, quite often specifically addressing the issue of disciplinary differences in the choice of certain linguistic items. This article deals with hedging, one of the pragmatic realizations of author stance, in a large synchronic Lithuanian academic corpus (CorALit) comprising five major science areas. The results of the analysis confirm some previous observations about clear differences between soft and hard science fields, at the same time outlining specific features of Lithuanian researchers' academic identity. The relatively low frequency of adverbials acting as hedging devices point to specific argumentation patterns within Lithuanian academic discourse. Extending the research on author stance to languages other than English helps to highlight some universal disciplinary trends as well as peculiar features typical to specific discourses and cultural communities. Placed within the context of similar cross-linguistic studies it contributes to the rich mosaic of pragmatic and intercultural aspects characteristic of academic discourse worldwide.

The five studies in the third part of the volume present new perspectives on grammatical and lexical categories. Until recently, the respective categories in Lithuanian and Latvian had been described almost exclusively within the framework of traditional grammar.

In her paper entitled *Reportive evidentiality and reported speech: is there a boundary? Evidence of the Latvian oblique* **Joanna Chojnicka** addresses the problem of the relationship between evidentiality and reported discourse by investigating the use of a Latvian verb-form that is difficult to categorize. In Latvian traditional grammar it is classified as a mood, but several researchers have pointed out the inadequacy of this classification. The current discussion, to which Chojnicka makes a substantial contribution, regards the oblique either as an expression of the category of evidentiality or as a marker of reported speech. Analyzing the use of the oblique in Latvian parliamentary speeches, Chojnicka postulates a cline between these two functions and describes the characteristics of utterances containing the oblique at several positions on this cline. She demonstrates how such an approach may lead to a better understanding

and a more accurate description of the category and its interaction with other means of expression in Latvian. Furthermore, the article contributes to the general understanding of evidentiality and its borders.

**Violeta Kalėdaitė** in her article *The specifying existential sentence type in Lithuanian: a problem statement* investigates grammatical, semantic and discourse related properties of existential sentences with relative clause extension. These sentences contain a form of the verb *būti* 'be' in its exclusively existential sense (equivalent to English *there is*), a subject NP and a relative clause. A corpus search shows that the pattern, which is known in various languages, is also very common in Lithuanian. Features special to the Lithuanian construction include genitive marking of plural subject, and an infinitival variant that lacks a formal subject. Kalėdaitė shows further that in texts, the investigated type of existential sentences is typically used to introduce a new topic or a new referent and to link, often by contrasting, topics or referents.

**Erika Jasionytė's** paper on the *Lithuanian impersonal modal verbs REIKĖTI 'need' and TEKTI 'be gotten': a corpus-based study* is devoted to a synchronic analysis of two impersonal modal verbs expressing the notion of necessity. Due to limited corpus-based synchronic research into the Lithuanian modal system, the paper aims to fill the gap and provide a detailed account of the semantics and use of two modal verbs of necessity. The purpose of the study is to establish to what extent the Lithuanian verbs *reik(ė)ti* 'need' and *tekti* 'be gotten' have developed modal meanings. The paper deals with the frequency of modal and non-modal use of the verbs, which can be indicative of the level of grammaticalization that linguistic expressions are undergoing. The findings of the corpus-based analysis carried out support cross-linguistic observations that impersonal modals are less polyfunctional (than personal modals) and they do not develop epistemic values.

In the following contribution **Andra Kalnača and Ilze Lokmane** explore *The semantics and distribution of Latvian reflexive verbs*. The object of their investigation are Latvian verbs containing the reflexive postfix *-s*. On the content side these verbs are united by a common component 'by oneself (itself), independently'. Following the model of Geniušienė's (1987) typology of reflexive verbs, the authors compile the first exhaustive classification of these verbs in Latvian based on syntactic relations and semantic roles, and which includes further grammatical and semantic features such as modal and aspectual meanings. While the proposed classification has the form of a strictly hierarchic tree, the authors are aware of similarities among classes at different places in the tree and envisage the final classification as a network rather than a tree.

The similarities and differences of autocausative, decausative and passive reflexive verbs are explicitly discussed.

In the final contribution to this volume **Loïc Boizou** challenges traditional grammatical description by asking *Do we need to count numerals as a part of speech in Lithuanian?* The author discusses alternatives to the traditional word class numerals, as the latter is too heterogeneous and too general to allow the formulation of syntactic rules. Based on the framework of functional generative description and other models that build on the notion of dependency, Boizou proposes a multilayered model that distinguishes a morphological and a syntactic level of analysis and includes several modules. He demonstrates that such a model leads to simpler rules, and shows how it can improve the description of intricate constructions in Lithuanian such as the numeral complex and the pronominal use of cardinals. Being more stringent than models based on traditional grammar, the proposed model is better suited for applications such as the grammatical annotation of corpora and automatic syntactic analysis.

This volume, we believe, will be in line with the present state of the art of contemporary linguistic research, and will enrich international research by adding new data from two modern Baltic languages, Latvian and Lithuanian.

## References

- Geniušienė, Emma. 1987. *The typology of reflexives*. Berlin, New York: Mouton de Gruyter.
- Händler, Harald & Herbert Ernst Wiegand. 1982. Das Konzept der Isoglosse: methodische und terminologische Probleme. In Walter Besch et al., eds. *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung. Erster Halbband*. Berlin, New York: de Gruyter, 501-527.
- Kļaviņa, Sarma. 2002. Die Werke von August Bielenstein - eine bedeutende Quelle der Indoeuropäistik des 19. Jahrhunderts. *Res Balticae* 8, 151-168.
- Rinholm, Helge. 1990. Continuity and change in the Lithuanian standard language. In: István Fodor & Claude Hagège, ed. *Language reform*. Volume V. Hamburg: Buske, 269-300.
- van der Auwera, Johan, Ewa Schalley & Jan Nuyts. 2005. Epistemic possibility in a Slavonic parallel corpus: A pilot study. In Björn Hansen & Petr Karlík, eds. *Modality in Slavonic languages. New perspectives*. München: Sagner, 201-217.



## **PART I:**

### **CORPUS-BASED CONTRASTIVE STUDIES**

# EPISTEMIC NECESSITY IN A PARALLEL CORPUS: LITHUANIAN VS. ENGLISH<sup>1</sup>

AUDRONĖ ŠOLINĖ

## Abstract

The article examines the qualitative and quantitative parameters of equivalence between the realizations of epistemic necessity in English and Lithuanian. The focus of the contrast lies on the auxiliary and adverb strategies (van der Auwera et al. 2005) in English (*will, would, shall, should, ought to, must* vs. *probably, certainly, indeed, surely, definitely, presumably, no doubt/undoubtedly, for sure, for certain, decidedly, undeniably*) as opposed to the corresponding modal verb and adverb/particle strategies in Lithuanian (*turėti* ‘have to’ vs. *tikriausiai* ‘certainly/most probably’, *greičiausiai* ‘doubtless’, *veikiausiai* ‘certainly/most probably’, *turbūt* ‘probably’, *matyt* ‘evidently’, *būtinai* ‘necessarily’, *ko gero(-a)* ‘very likely/most probably’, *neišvengiamai* ‘inevitably’, and *be abejo/be abejonės/neabejotinai* ‘no doubt/undoubtedly’). The aim of the corpus-based study is to find out which means of expression are preferable in the two languages. The paper will also look at the frequency of epistemic and non-epistemic use of the modal expressions in the original and in translation.

**Keywords:** epistemic modality, modal verbs, epistemic adverbials, corpus-based analysis, frequency, contrastive analysis

---

<sup>1</sup> I am sincerely grateful to Johan van der Auwera (the University of Antwerp), Bert Cornillie (the University of Leuven), Nicole Nau (Adam Mickiewicz University of Poznan) and Aurelija Usonienė (Vilnius University) for reading my manuscript and offering their kind advice. Thanks are also due to Elom Bell for improving my English. Responsibility for any remaining errors or shortcomings is mine alone.

## 1. Introduction

Modality is a complex category encompassing more than one field and discipline, including morphology, lexicon, syntax and pragmatics, and fine-grained cross-linguistic differences are difficult to discover by introspection or analysis of contrived examples. The corpus-based approach adopted in this study helps to reveal patterns which would be difficult to find otherwise. The possibility of combining comparable and parallel corpora, thus taking advantage of the specific merits of both types, allowed me to pin down the correspondences between the formal and functional features in the source language (SL) and target language (TL) texts and establish parallels between them.

The paper focuses on the quantitative and qualitative parameters of equivalence between the lexical exponents of epistemic necessity in English and Lithuanian. The axis of contrast is laid on the auxiliary and adverb strategies (see van der Auwera, Schalley & Nuyts 2005): a probe will be made into the cross-linguistic distribution and behavior of the central English modal auxiliaries of necessity (*will, would, shall, should, ought to, and must*) and the modal adverbials (*probably, certainly, indeed, surely, definitely, presumably, no doubt/ undoubtedly, for sure, for certain, decidedly, undeniably*) as opposed to the corresponding Lithuanian modal verb *turėti* ‘have to’ and the adverbials<sup>2</sup> *tikriausiai* ‘certainly/most probably’, *greičiausiai* ‘doubtless’, *veikiausiai* ‘certainly/most probably’, *turbūt* ‘probably’, *matyt* ‘evidently’, *būtinai* ‘necessarily’, *ko gero(-a)* ‘very likely/most probably’, *neišvengiamai* ‘inevitably’, and *be abejo/ be abejonės/ neabejotinai* ‘no doubt/undoubtedly’, e.g.:

- (1) *He must have broken the window!* (BrD)<sup>3</sup>

<b>Tikriausiai</b>	jis	išdauž-ė	langą.
probably.ADV	he	break-PST.3	window.ACC

---

<sup>2</sup> As no consensus has been reached so far regarding the distinction between the word classes of modal particles and adverbs in Lithuanian linguistics, I will be referring to Lithuanian modal expressions like *ko gero* ‘most likely’, *turbūt* ‘probably’, etc. as epistemic stance adverbials (Biber et al. 1999: 854).

<sup>3</sup> The translationally related sentence pairs (LT→EN or EN→LT) given as examples in the paper come from the *PraCorp<sub>EN-LT-EN</sub>* and they have not been glossed (except some instances). Other Lithuanian sentences given as single instances have been glossed. The examples from the *ParaCorp<sub>EN-LT-EN</sub>* carry a reference code which accompanies an original sentence.

(2) *That **must have been** horrible for you.* (HN)

‘Tau	<b>tur-ėjo</b>	<b>būti</b>	išties	baisu.’
you.DAT	have- PST.3	be.INF	really	horrible

Contrastive analyses based on parallel and comparable corpus data (Aijmer 1997, 1999, Dyvik 1998, 2004, Johansson 2001, 2007, Simon-Vandenberg & Aijmer 2007, Mortelmans 2009 among others) show that in a cross-linguistic perspective the percentage of lexical correspondence in expressions of epistemic modality is not very high and different subsystems tend to interact. For instance, Aijmer’s (1999) study on the comparison of epistemic possibility in English and Swedish has shown that when there is a gap in the Swedish system of modal auxiliaries, it can be filled by a modal adverb. The research made by van der Auwera, Schalley & Nuyts (2005) also demonstrates that, despite the similarity of the available linguistic inventory for epistemic meaning realization in terms of the auxiliary and adverb strategies, speakers of the English and Slavonic languages do not use these strategies with equal frequency. The researchers link it up to the relatively high degree of grammaticalization of the English modal auxiliaries and their polyfunctionality. Moreover, the results of the investigation of the Slavonic translational equivalents could be looked at from an areal perspective: the authors describe an east-west cline of languages with more modal adverbs in the Eastern areal of the Slavonic languages and more modal verbs in the west.

A hypothesis is made that the proportion of lexical correspondence can be very low when dealing with realizations of grammatical categories cross-linguistically. The purpose of this parallel corpus-based study is to find out what means of expression are preferable in the given languages. The paper will also look at the frequency of epistemic and non-epistemic use of the modal expressions under analysis and will aim to determine whether there is any language-specific conceptualization of the strength of the speaker’s commitment to the factuality of his/her proposition and to what extent the speaker’s evaluation of the proposition is influenced by the interactional context of use and available evidence (Squartini 2008, Boye & Harder 2009).

## 2. Preliminaries

Before embarking on the cross-linguistic analysis of the lexical markers of necessity in English and Lithuanian and the description of the empirical data used in this study, a short explanation of some semantic concepts is due.

## 2.1 Epistemic and non-epistemic modality. Epistemic-evidential overlap

Modal meanings can be classified in a number of ways. Traditionally (cf. Palmer 2001), scholars distinguish three kinds of modality: epistemic, deontic, and dynamic, e.g:

- (3) Epistemic: *They may be in the office.* - *They must be in the office.*  
 Deontic: *They may/can come in now.* - *They must come in now.*  
 Dynamic: *They can run very fast.* - *I will help you.*

(Palmer 2003: 7)

For this study, a crucial distinction is the one between epistemic and non-epistemic modality (cf. Mortelmans, Boye & van der Auwera 2009). The latter encompasses dynamic and deontic modalities. This paper will not set deontic and dynamic modalities apart and, in case of contrast with epistemic modality, both of them will be referred to as non-epistemic. Epistemic modality is defined as dealing with the “evaluation of the chances that a certain hypothetical state of affairs under consideration (or some aspect of it) will occur, is occurring or has occurred in a possible world” (Nuyts 2001: 21), as in:

- (4) *I'm not... so good today... Tomorrow **will be better**.* (AM)

As far as evidentiality is concerned, the definition proposed in van der Auwera & Plungian (1998) offers a more or less uncontroversial characterization of this category: “Evidentiality concerns the indication of the source or kind of evidence speakers have for their statements” (van der Auwera & Plungian 1998: 57). The relation between the realms of evidentiality and epistemic modality is not always clear. Epistemic-evidential syncretism has been widely discussed in the literature (van der Auwera & Plungian 1998, Mortelmans 2000, Plungian 2001, Cornillie 2007, 2009, Squartini 2008 among others). Plungian (2001: 354) contends that the evidential value is always inherently present in the epistemic meaning: “while an evidential supplement can always be seen in an epistemic marker, the opposite does not always hold: not all evidential markers are modal in that they do not all necessarily imply an epistemic judgment”. In this paper I align myself with the views expressed in van der Auwera & Plungian (1998), who see the relation between evidentiality and epistemic modality as the one of overlap and include one particular type of evidentiality – inferential evidentiality – within the domain of epistemic

modality. The epistemic necessity and inferential evidentiality expressed by the English modal auxiliary *must* is the best known example of this overlapping domain, e.g.:

- (5) *From the fear in her touch Langdon sensed someone **must be approaching**...* (BrD)

I also align myself with the position expressed in Cornillie (2007) and his suggestion to extend the domain where epistemic necessity and inferential evidentiality meet. His study of Spanish (semi)-auxiliaries gives evidence that “the overlap category between modality and evidentiality should not be exclusively situated in the domain of necessity but can also include prediction” (Cornillie 2007: 116).

## 2.2. The expression of modal necessity in Lithuanian

There are no modal auxiliaries in Lithuanian that could be comparable to the Germanic and Romance modal auxiliaries in terms of the ‘NICE’ properties that are essential to the classification of a form as an auxiliary (cf. Palmer 1987: 14-21). The majority of Lithuanian modal verbs are fully conjugated lexical verbs and do not have any conspicuous morphological and morphosyntactic characteristics that would set them apart from the other verbs. The key modal verbs are *galėti* ‘can/may’ and *turėti* ‘have to’. Both verbs are polyfunctional, i.e. they can express both epistemic and non-epistemic (deontic and dynamic) modality.

The verb *galėti* ‘can/may’, cognate with *galia* ‘power’, is the major exponent of possibility. The key marker of epistemic and non-epistemic necessity is *turėti* ‘have to’, whose source domain is possessive constructions and whose pre-modal possessive meaning has been retained to the present day, e.g.:

- (6) *Jis            tur-i                    šunį.*  
       he            have-PRS.3                    dog.ACC  
       ‘He’s got a dog’

The epistemic and non-epistemic meanings of this verb can be illustrated in the following examples:

- (7) *Jis            tur-i                    būti      jau      namie.*  
       he            have-PRS.3                    be.INF    already    home  
       ‘He must be at home already’

- |     |                      |             |              |
|-----|----------------------|-------------|--------------|
| (8) | <b><i>Tur-iu</i></b> | <i>eiti</i> | <i>namo.</i> |
|     | have-PRS.1SG         | go.INF      | home.ADV     |
|     | 'I must go home'     |             |              |

Another less frequently used verb of modal necessity is *privalėti* 'be obliged to'. Like the key modal verbs *galėti* 'can/may' and *turėti* 'have to', it is a full conjugated lexical verb. Originally, as claimed in Holvoet (2009: 200), *privalėti* 'be obliged to' might have been used to express dynamic necessity (the noun *prievolė* denotes 'compulsory work or service'); however, in modern Lithuanian, in contrast to the two key modal verbs, *privalėti* 'be obliged to' encodes only deontic necessity, e.g.:

- |     |                           |             |              |
|-----|---------------------------|-------------|--------------|
| (9) | <b><i>Prival-au</i></b>   | <i>eiti</i> | <i>namo.</i> |
|     | have-PRS.1SG              | go.INF      | home.ADV     |
|     | 'I am obliged to go home' |             |              |

Traditional Lithuanian grammars distinguish not only personal (full) but also impersonal lexical verbs used to convey various modal meanings (Balkevičius 1998). There are a few verbs with a defective paradigm like *reikia* 'need', *teko* 'be gotten' that can be regarded as semi-modal auxiliaries. The latter are mainly the so-called verbal impersonals, i.e. third person or 'zero-coded forms' of the verb, which are uninflected verb stems unmarked for number and used in impersonal constructions, e.g.:

- |      |                           |                |  |
|------|---------------------------|----------------|--|
| (10) | <b><i>Reik-ia</i></b>     | <i>dirbti.</i> |  |
|      | need-PRS.3                | work.INF       |  |
|      | 'It is necessary to work' |                |  |
- 
- |      |                        |                     |             |                |
|------|------------------------|---------------------|-------------|----------------|
| (11) | <i>Jam</i>             | <b><i>tek-o</i></b> | <i>daug</i> | <i>dirbti.</i> |
|      | he.DAT                 | be gotten-PST.3     | a lot       | work.INF       |
|      | 'He had to work a lot' |                     |             |                |

In different languages modal adverbials are mostly used to convey epistemic overtones, with a very few and rare cases where they compete with modal verbs in the deontic and dynamic realms. One could refer to "a well known example in Slovenian with its modal particle *lahko* 'easily' covering all types of possibility, also deontic and dynamic" (Holvoet 2009: 204). However, Lithuanian cannot boast of such adverbials. In Lithuanian, unlike modal verbs, modal adverbials do not convey non-epistemic readings. They are used to indicate a low/high degree of speaker's confidence regarding the truth-value of the proposition. The use of epistemic adverbials allows the speaker to specify modal nuances

regarding the probability/likelihood of the proposition more precisely, whereas the range of epistemic modal verbs is narrower, cf.:

- (12) *Jis tur-ėjo gauti mūsų laišką.*  
 he have-PST.3 receive.INF our letter  
 ‘He must have received our letter’
- (13) *Jis ko gero/turbūt/veikiausiai gav-o*  
 he very likely/probably/certainly receive-PST.3  
*mūsų laišką.*  
 our letter  
 ‘He probably received our letter’

In Lithuanian, there are only a few modal adverbs morphologically marked as adverbs, i.e. having the most productive adverb forming suffix – (*i*)ai (Ambrasas 1997: 378). They are *tikriausiai/greičiausiai/veikiausiai* ‘most probably’ and *būtinai* ‘necessarily’ and they are used to express high probability, i.e. epistemic necessity. There has not been any detailed research carried out on the paths of grammaticalization and lexicalization of the Lithuanian particle *gal* ‘perhaps’ and the two modal words *turbūt* ‘probably’ and *galbūt* ‘maybe’; however, it is obvious that the marker of epistemic possibility *gal* is related to the verb *galėti* (‘can/may-INF’)<sup>4</sup>, while the words *turbūt* ‘probably’ and *galbūt* ‘maybe’ are the truncated forms of *turi* ‘have-PRS.3’ and *gali* ‘can/may-PRS.3’ in their modal meanings plus the existential verb *būti* ‘to be’<sup>5</sup>. The source of origin of some of the epistemic necessity adverbials is rather transparent; for instance, *matyt* derives from the infinitive *matyti* ‘see’, *be abejo/be abejonės/neabejotinai* are related to the verb *abejoti* ‘doubt’, *ko gero(-a)* is the genitive form of *kas gera*, literally meaning ‘the one which is good’ (Wiemer 2007: 195).

A variety of other modal realizations of necessity such as the expressions with adjectives, participles, nouns, the verbs of propositional

---

<sup>4</sup> Different authors seem to differ in their views regarding the origin of *gal* ‘perhaps/maybe’. For instance, the authors of the *Lithuanian Grammar* maintain that *gal* ‘maybe’ derives from *gali* ‘(you) can-PRS.2SG’ (Ambrasas 1997: 397), while Wiemer (2007) claims that “*gal* is the truncated form of *galėti* ‘can’ (probably derived from the PRS.3-form *gali* <...>)” Wiemer (2007: 195).

<sup>5</sup> Brinton & Traugott (2005) speak about fusion, coalescence, and lexicalization of Polish *może* and Lithuanian *gal* via the constructions *možet byt* ‘it may be’ and *galbūt* ‘it may be’, respectively (Brinton & Traugott 2005: 81).



attitude and the verbs of probability taking clausal complements will not be dealt with in the present paper.

### 3. Data and method

The method used in the research is non-experimental data collection; it is a contrastive analysis based on the data obtained from a self-compiled bidirectional parallel corpus – *ParaCorp<sub>EN-LT-EN</sub>*. The corpus design follows the model of the English-Norwegian Parallel Corpus (Johansson 2007). The *ParaCorp<sub>EN-LT-EN</sub>* was compiled from original English fiction texts and their translations into Lithuanian and original Lithuanian fiction texts and their translations into English. The size of the corpus is 1, 572, 498 words (see Table 1).

	Original	Translation	Total
ParaCorp <sub>EN→LT</sub>	486, 871	386, 640	873, 511
ParaCorp <sub>LT→EN</sub>	296, 759	402, 228	698, 987

**Table 1. Size of the two sub-corpora ParaCorp<sub>EN→LT</sub> and ParaCorp<sub>LT→EN</sub>**

The advantage of such a corpus model is that it allows multidirectional comparisons and can be used both as a parallel corpus and a comparable corpus (Johansson 2007: 11). It must be admitted, however, that there is an imbalance between the two sub-corpora. My aim was to compile a balanced bidirectional corpus; however, the matching of original texts in terms of size was difficult as the number and range of texts that have been translated from English into Lithuanian is far greater than those of translations from Lithuanian into English. A similar situation has been observed in other languages as well (cf. Čermák & Klégr 2004: 84, Johansson 2007: 13). Mainly due to this reason, the included literary texts vary in their length and number: the *ParaCorp<sub>EN→LT</sub>* includes full texts (6 novels and 2 short stories), whereas the *ParaCorp<sub>LT→EN</sub>* is comprised of both full texts and extracts (3 full text novels, 39 short stories and 14 extracts). Moreover, the English texts have been translated by 8 translators (5 women and 3 men); the Lithuanian texts have been translated by 19 translators (13 women and 6 men). Most of the texts included in the corpus were written, translated, and published in the period of 1980–2006. However, there are some texts that were published before 1980: the *ParaCorp<sub>LT→EN</sub>* includes the novel *Hstera (Esther)* by V. Kavaliūnas and some short stories, and the *ParaCorp<sub>LT→EN</sub>* includes G. Orwell's novel *1984*. The list of all the texts is in the Data Sources.

The texts are not lemmatized; the alignment used is based on the aligning tool LYGIA (developed at the Centre of Computational Linguistics of Vytautas Magnus University). The alignment was performed first at the paragraph level, then at the sentence level. Then, in order to generate concordance lines, I used the multilingual browser ParaConc (Barlow 1995). Though the search was automatic, the analysis of concordances was carried out manually, since the *ParaCorp<sub>EN-LT-EN</sub>* is not annotated.

Since the sub-corpora are of different size, the raw frequency numbers have been normalized per 1,000 words. Furthermore, in order to check whether the similarities and differences are statistically significant, I have also performed the log-likelihood (LL) test, which is commonly considered to be a more statistically reliable test than the chi-square test (cf. Dunning 1993). Frequencies of particular patterns and uses are of crucial importance to this study, since frequency can be an important factor in specification of meaning (Leech 2003, Simon-Vandenberg & Aijmer 2007).

The use of a parallel corpus makes it possible to establish cross-linguistic paradigms illustrating how an item in the source language is rendered by alternative variants in the target language. In addition, it is possible to establish equivalents by going from translations to sources. Though the use of parallel corpora in contrastive studies has been criticized by some scholars as giving only a mirror image of their source language (Teubert 1996) and being infected by *translationese* (Gellerstam 1996), my views are very much in line with the opinion expressed in Mauraanen (2002), Noël (2003) and Simon-Vandenberg & Aijmer (2007). They consider translations to be part of natural language in use. The output of translators “varies on a number of parameters, as does that of any language user, whether bilingual or monolingual” (Mauraanen 2002: 164); moreover, “by looking at correspondences in other languages one discovers subtle meaning distinctions which may go unnoticed if one considers monolingual data only” (Simon-Vandenberg & Aijmer 2007: 7). Thus, the combination of the data obtained from comparable and parallel corpora can contribute to providing a fine-grained picture of the items under study. Naturally, the fact that the language data in this analysis have been collected from fiction limits the scope of my conclusions to a certain extent, but one has to agree that a literary text encompasses a broad variation of linguistic style and contains a lot of dialogue, which is an ideal environment for epistemic modality.

#### 4. Criteria for the data selection.

##### Constraints on epistemic and non-epistemic meaning distinction

When the overall distribution of the modal necessity markers in the two languages was considered, the first step was to isolate all occurrences of the English modal auxiliaries (*will, would, shall, should, ought to, must*) and the modal adverbials (*probably, certainly, indeed, surely, definitely, presumably, no doubt/undoubtedly, for sure, for certain, decidedly, undeniably*) as opposed to the corresponding Lithuanian modal verb *turėti* 'have to' and the adverbials *tikriausiai* 'certainly/most probably', *greičiausiai* 'doubtless', *veikiausiai* 'certainly/most probably', *turbūt* 'probably', *matyti* 'evidently', *būtinai* 'necessarily', *ko gero(-a)* 'very likely/most probably', *neišvengiamai* 'inevitably', and *be abejo/be abejonės/neabejotinai* 'no doubt/undoubtedly'.

It must be noted that there have been two stages of selected data analysis carried out, and the sets of selected language data were not identical for each stage. In the first stage of the quantitative analysis, I aimed to investigate which linguistic markers of modal necessity (without any specification into epistemic and non-epistemic necessity) are prevalent in both languages. So, the first stage of the analysis was concerned with the overall occurrences of all modal necessity verbs (both positive and negative forms). As for the Lithuanian modal verb of necessity *turėti* 'have to', sentences in which it was used in non-modal possessive constructions have been discarded initially, e.g.:

- (14) *Ar asilq turi?* (MI1)  
'Do you **have** a mule?'

Lithuanian epistemic stance adverbials *tikriausiai* 'certainly/most probably' and *greičiausiai* 'doubtless' can also function as adverbials of manner, meaning 'in the most precise way' and 'in the quickest way' respectively. These instances have been eliminated from the initial data-set as well, e.g.:

- (15) *Jo slaugomieji ligoniai greičiausiai pasveikdavo.* (SB)  
'The patients he cared for recovered **the quickest**.'

The second stage of the analysis was a combination of both quantitative and qualitative analysis. At this stage, the emphasis has been laid on the distinction between the epistemic and non-epistemic use of the

modal realizations under study. The initial language data-set has been filtered further and analyzed taking into account the criteria and constraints described in Sections 4.1. – 4.4.

#### 4.1. Negative sentences and sentence fragments

Elliptical sentences where the proposition was not expressed by means of a fully fledged sentence and negative sentences, as in the examples under (16) and (17), have not been included into the analysis of the distribution of epistemic and non-epistemic meanings:

(16) *Probably Frank Jibben.* (MI)

(17) *...one must not make martyrs.* (MI)

#### 4.2. Adverb/adverbial scope specification

For the English adverbs, I only included uses with sentential scope, as in example (18), since it is only in these contexts that there is a choice between the adverbial and the auxiliary strategies. Such sentences as (19) have been discarded:

(18) *... Da Vinci was **certainly** aware of that fact.* (BrD)

(19) *I've been telling David to see him for months, **probably** years.*  
(HN)

As far as the Lithuanian epistemic stance adverbials are concerned, their position in the sentence is free and they can also have variable scope – from entire clauses to (parts of) noun phrases (henceforth NPs). Only the cases where the epistemic stance adverbials were not integrated into clausal syntax and had a clausal scope have been counted as in (20) and (21):

(20) *... jis **tikriausiai** mirė.* (MI)  
'He must be dead.'

(21) ***Tikriausiai** tai buvo vasaros virusas.* (JV1)  
'It seemed like a summer virus.'

### 4.3. Subject specification

Coates (1983), amongst others, suggested that expressions of epistemic modality can be differentiated from non-epistemic expressions in terms of certain contexts and environments they appear in. Cornillie's (2007) study of Spanish (semi)-auxiliaries showed that lexical and modal *prometer* 'promise' and *amenazar* 'threaten' show different preferences for subjects (Cornillie 2007: 135-149). In this paper, the singled out sentences have been considered from the point of view of how much the subject specification can determine the epistemic reading of a sentence. The hypothesis is that 1<sup>st</sup>/2<sup>nd</sup> person verb forms can rarely receive an epistemic reading, which is in accordance with Heine's (1995: 26) and Coates' (1983: 97) observations. The impersonal *you* in its generic sense is very common as the subject in sentences yielding a non-epistemic reading, e.g.:

- (22) *Apskritai, lagery niekuomet nesakoma, kaip ką turi daryti, - turi pats susiprasti.*(SB)

'The iron rule is that no one ever tells anyone how to do anything in camp – **you have to figure it out** by yourself.'

Special attention was given to the Lithuanian 3<sup>rd</sup> and 2<sup>nd</sup> person forms *turi* 'have-PRS.3' and *turi* 'have-PRS.2SG' since they are homographs. The two forms differ in their accentuation in speech, whereas in the written language, it is the context of use that has to be taken into account. Some sentences with the predicate *turi* contain an overt NP in the subject position as in (23) or a pronoun as in (24), which eases the ascription of *turi* to a third or second person form group.

- |      |                            |              |             |                |
|------|----------------------------|--------------|-------------|----------------|
| (23) | <i>Žmogus</i>              | <i>tur-i</i> | <i>daug</i> | <i>dirbti.</i> |
|      | man.NOM                    | have-PRS.3   | much        | work.INF       |
|      | 'A person must work a lot' |              |             |                |

- |      |                       |              |             |                |
|------|-----------------------|--------------|-------------|----------------|
| (24) | <i>Tu</i>             | <i>tur-i</i> | <i>daug</i> | <i>dirbti.</i> |
|      | you                   | have-PRS.2SG | much        | work.INF       |
|      | 'You must work a lot' |              |             |                |

When the sentences contained no overt subject, their interpretation needed a slightly wider context than a one-sentence frame to establish co-referential links, e.g.:

- (25) *Tur-i*                      *daug*      *dirbti*.  
       have-PRS.3            much      work.INF  
*Jis*      *tur-i*                      *pastatyti*              *namą*.  
       he      have-PRS.3            build.INF              house.ACC  
       ‘He must work a lot. He must build a house’

Moreover, an important but not decisive factor in distinguishing between epistemic and non-epistemic meaning seems to be the animacy of the subject. For example, sentences (23) and (24) with animate subjects clearly receive a non-epistemic interpretation, but this is not always the case. There is no doubt that the factors determining the epistemic or non-epistemic interpretation converge. For instance, sentence (26) has an animate subject, but still is epistemic, as it contains a stative verb (the correlation of stativity and epistemic modality is discussed below).

- (26) *My husband must be at home already.*

A number of scholars associate use of an inanimate subject, an impersonal *it/this* and existential *there* with an epistemic reading (Coates 1983, Heine 1995, Wärensby 2004). Bybee et al. (1994), for example, observe that the presence of impersonal *it* in a subject position precludes a non-epistemic reading in cases like (27):

- (27) *It must be easier to live beautiful lives when you're posh.* (HN)

In the same vein, Wärensby (2004) claims that sentences containing the existential subject *there* are unquestionably epistemic and “this correlation may be explained by appealing to the fact that sentences containing introductory subjects are existential in nature and always involve some copular or state predicate” (Wärensby 2004: 179), e.g.:

- (28) *Harry could hear running water; there must be a stream somewhere close by.* (RJK)

Thus, the singled out sentences have been considered from the point of view of how much the subject specification can determine an epistemic reading of a sentence.