

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

10

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

10

PETAR KEHAYOV

An Areal-Typological Perspective to
Evidentiality: the Cases of the
Balkan and Baltic Linguistic Areas



TARTU UNIVERSITY
PRESS

Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia

Dissertation is accepted for the commencement of the Degree of Doctor of Philosophy in Estonian Language on 6 May, 2008 by the Council of the Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia.

Supervisor: Professor emeritus Mati Ereht, PhD, University of Tartu, Estonia

Opponent: Professor Hannu Tammela, University of Tampere, Finland
Professor Alexandra Aikhenvald, La Trobe University, Australia

Commencement: Ülikooli 18, room 204, Tartu, on 30 June 2008 at 2 p.m.

Publication of this dissertation is granted by the Graduate School of Linguistics and Languages Technologies, University of Tartu

ISSN 1406–5657

ISBN 978–9949–11–891–5 (trükis)

ISBN 978–9949–11–892–2 (PDF)

Autoriõigus Petar Kehayov, 2008

Tartu Ülikooli Kirjastus

www.tyk.ee

Tellimus nr. 227

PREFACE

I cannot overstate by saying that this dissertation is the outcome of a great number of happy coincidences. Its prelude lies somewhere back in the 1990s – a time when I was trying to figure out what I wanted to do professionally. Though linguistics had been in my scope ever since the early 1990s, several years later I was almost sure that this field was not for me. The reason I was discouraged was that I was trained exclusively in the field of (traditional) historical linguistics, which – I constantly felt – was too much about rigid sound change rules and reconstructions of forms and too little about the semantics of these forms. It was as late as 2000, I think, when I first heard about the functional category of ‘evidentiality’. Becoming involved in the study of evidentiality gave me the opportunity both to satisfy my interest toward vague semantic issues and to take advantage of my Balkan background, to which I owe the cultural practice of being more explicit when talking about knowledge and experience than a speaker of English or Russian, for example.

The dissertation contains four articles published in the span of six years (2002–2008). This is a long period of time, especially for a young scholar, and the reader will certainly notice an evolution, both in terms of linguistic scope and theoretical sophistication. Some of my earlier statements seem inadequate or otherwise out-of-date to me now, and I have been eager to change or refine them. Compiling this dissertation provided me with a great opportunity to do this. It is worth stressing, however, that none of these changes or refinements has any crucial impact on the main standpoints defended in the articles as they were published.

There are so many people I want to thank for their support that I cannot put them in any order of significance. To those forgotten, I apologize in advance. First, I am indebted to my academic supervisor Mati Ereht for his ever-willing support for my work. It might come as a surprise to him how much I have learned from him, not only about the field of linguistics, but also about the ethics of scientific inquiry. I am grateful to Rogier Blokland for keeping me on the right track since the beginning of my studies at the University of Tartu. Without him I would not have become a linguist. My sincere thanks goes to my wife Kadri for her inexorable patience in discussing my ideas with me and in this way helping me to better articulate them. Her contribution to the preparation of this dissertation could hardly be overestimated. My parents Neti and Dimitar are responsible for encouraging me to look for a field that is challenging for me and not to care about social prestige. I am also thankful to Axel Holvoet, Kaur Mägi, Florian Siegl and Björn Wiemer for sharing their linguistic expertise and their ideas about evidentiality with me. Johan van der Auwera made my stay in the University of Antwerp enjoyable and fruitful, introduced me to the large playground of European functional and typological linguistics and helped me to establish the necessary contacts, for which I am truly thankful.

I am also obliged to Vahur Aabrams and Jouni Tapper for our exciting off-the-record conversations which often touched on the topic of this dissertation. Furthermore, I cannot forget to mention the occasional help of Andreas Ammann, Jelena Antuševa, Norbert Boretzky, Bert Cornillie, Victor Friedman, Mati Hint, Helen Plado, Ērika Krautmane, Liina Lindström, Rami Memushaj, Helle Metslang, Ewa Schalley, Jana Šteinberga, Reeli Torn, Szilard Tóth, Eva Toulouse, Ilona Tragel and Virve Vihman with various linguistic as well as extra-linguistic questions. Finally, I wish to thank the two independent reviewers of the dissertation – Alexandra Y. Aikhenvald and Martin Ehala for their highly useful remarks.

As for official support, I have benefited from grants no. 5202 and 7006, and from project no. TFLEE 0084 of the Estonian Science Foundation, and from the support of the Graduate School of Linguistics and Language Technology of the University of Tartu. I also wish to thank the Government of Flanders for the scholarship which made possible my stay at the Center for Grammar, Cognition and Typology of the University of Antwerp in the autumn 2005 and spring 2006. Last but not least, my gratitude goes also to the administrative staff at the Department of Estonian Linguistics for helping me deal with various practical problems through the past four years.

CONTENTS

LIST OF THE ARTICLES.....	9
CONVENTIONS.....	10
1. INTRODUCTION.....	12
1.1. Main goals of the study.....	12
1.2. Methods and material.....	14
2. EVIDENTIALITY AS A GRAMMATICAL CATEGORY.....	20
2.1. Key definitions and problems relevant to evidentiality.....	20
2.2. Types of evidentiality systems.....	22
3. GEOGRAPHICAL DISTRIBUTION OF EVIDENTIALITY SYSTEMS.....	25
3.1. Hotbeds of evidentiality in Europe and their relations to larger Eurasian patterns.....	25
3.2. Evidentiality systems in the Balkan and Baltic areas.....	26
4. MAJOR QUESTIONS DEALT WITH IN ARTICLES I–IV.....	29
4.1. Parameters for comparison of the evidentiality systems.....	29
4.1.1. Formal parameters.....	33
4.1.2. Semantic and pragmatic parameters.....	36
4.1.3. Structural availability parameters.....	40
4.1.4. A comparison with de Haan’s parameters.....	41
4.2. The relationship between evidentiality and epistemic modality.....	43
4.3. The historical development of evidentiality markers.....	44
5. SOURCES REFERRED TO IN THE INTRODUCTORY PART (SECTIONS 1–4).....	53
ARTICLES.....	59

CONCLUSIONS AND PERSPECTIVES FOR FURTHER RESEARCH.....	165
Results of the comparison of evidentiality systems.....	166
The relationship between evidentiality and epistemic modality.....	171
The historical development of evidentiality markers	172
SOURCES REFERRED TO IN THE ENTIRE DISSERTATION.....	174
SUMMARY IN ESTONIAN	184
APPENDIX 1: Aikhenvald’s classification of evidentiality systems	192
APPENDIX 2: Hotbeds of evidentiality in Europe	193
APPENDIX 3: Tabular data for parameter values	194
APPENDIX 4: Typological distances	196

LIST OF ARTICLES

- I. Kehayov, Petar 2002: Typology of Grammaticalized Evidentiality in Bulgarian and Estonian. – *Linguistica Uralica* 38, 2, 126–144.
- II. Kehayov, Petar 2004: Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Morfosüntaks ja distributsioon. – *Keel ja Kirjandus* 47, 11, 812–829.
- III. Kehayov, Petar 2004: Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Semantika. – *Keel ja Kirjandus* 47, 12, 895–914.
- IV. Kehayov, Petar 2008 (in print): Interactions between grammatical evidentials and lexical markers of epistemicity and evidentiality: a case-study of Bulgarian and Estonian. – V. A. Plungian, B. Wiemer (eds), *Lexikalische Evidenzialitäts-Marker in slavischen Sprachen*. Wiener Slawistischer Almanach, Sonderband 72, Wien.

CONVENTIONS

The glossing conventions are slightly different in each article, and therefore the reader is invited to consult the abbreviation lists provided in the end of the articles (note that the abbreviations used in articles II and III are listed together in the end of article III). The abbreviations used in the remaining part of the dissertation (sections 1–4, Conclusions and Appendices) are listed below. There is one exception, however: in those cases where the examples are not mine but have been taken from studies which already use morpheme-by-morpheme translations, I have preserved the original glossing system used by the author. In other cases, especially when glossed examples were translated from other languages, glosses are adapted to the system below, and this is explicitly mentioned (e.g. adapted from Meydan 1996: 133).

Grammatical abbreviations

.	A form in the object-language is rendered by several elements in the glosses (including cases where the form is segmentable, but its segmentation is considered irrelevant).
-	morpheme boundary
—	The metalanguage does not have single word equivalent.
1, 2, 3	first person, second person, third person
ACC	accusative
ACT	active
ADM	admirative
AOR	aorist (tense)
COMP	complementizer
DEF	definite
DEM	demonstrative
F	feminine
ILL	illative
IMPF	imperfect (tense), imperfective (aspect)
INDIR	indirect (evidentiality)
INE	inessive
INF	infinitive
INTERJ	interjection
LOC	locative
M	masculine
PASS	passive
POSS	possessive
PRS	present
PST	past

PTCP	participle
REP	reported
SG	singular

Language codes

Alb	Albanian
Arum	Aromanian
Blg	Bulgarian
Bsk	Bashkir
Che	Chechen
Chv	Chuvash
Circ	Circassian
Est	Estonian
Gag	Gagauz
GAlb	Gheg Albanian
God	Godoberi
Grk	Greek
Ing	Ingush
Kaz	Kazakh
Ko	Komi
Lit	Lithuanian
Liv	Livonian
Ltv	Latvian
Ma	Mari
Mac	Macedonian
Mglr	Megleno-Romanian
Nog	Nogai
Rmn	Romani
Rom	Romanian
S/C	Serbian/Croatian
TAlb	Tosk Albanian
Tat	Tatar
Trk	Turkish
Tsz	Tsez
Udm	Udmurt

I. INTRODUCTION

I.1. Main goals of this study

This dissertation addresses, from a comparative typological perspective, the functional category of ‘evidentiality’, which designates the type of evidence on which speakers ground their statements. The object of comparison are the evidentiality systems of seven languages spoken in South-Eastern and North-Eastern Europe.

As any other typological enterprise, current cross-linguistic research on evidentiality is confronted with three main challenges. The first has to do with the commonness or rarity of different functional or formal types, in this case with the commonness or rarity of different types of evidentiality systems and their relevant features. This challenge is handled by quantitative methods. The second relates to the precise contents of whatever is referred to as ‘evidential’, or, in other words, to the location of evidentiality among the other functional categories and its relationships with them. The second challenge is therefore conceptual and classificatory by nature. While the first two challenges are essentially timeless, the third concerns the diachronic development of the evidentiality systems. We can not only compare evidentiality systems, but also their non-evidential sources and also the paths leading from non-evidential source structures to evidential target structures.

Interestingly enough, the most wide-ranging typological works on evidentiality so far (Chafe and Nichols 1986; Willett 1988; Guentchéva 1996a; Aikhenvald and Dixon 2003; Aikhenvald 2004; Guentchéva and Landaburu 2007) deal in great detail with the second and the third challenge while the first challenge, which should be the principal one in any typological inquiry conducted in Greenbergian spirit,¹ has received almost no attention. A big leap forward towards filling this gap in the world-wide quantification of evidentiality was taken with the appearance of *The World Atlas of Language Structures* (Haspelmath, Dryer, Gil and Comrie 2005), which contained two chapters on evidentiality written by Ferdinand de Haan (see De Haan 2005b, 2005c). These two chapters contained altogether nine evidentiality distinctions (three semantic and six coding), which were checked for 418 languages.

The research goals of this dissertation roughly correspond to the three challenges posed to modern research on evidentiality.

The first goal is to compare evidentiality systems according to a set of pre-defined parameters for comparison and to measure the typological distances

¹ See for example Whaley (1997: 23), who stresses the establishment of a relevant quantificational basis for typology as a main contribution made by Joseph Greenberg to linguistics.

between them. I will apply 19 such parameters for comparison relevant for the coding, semantic and syntagmatic properties of evidentiality. The language sample comprises languages from two widely acknowledged convergence areas: the Balkan and Baltic linguistic areas. The languages belonging to the Balkan area are Bulgarian, Macedonian, Albanian and Turkish, and the languages belonging to the Baltic area are Estonian, Latvian and Lithuanian. Despite the superficial similarity to de Haan's research program, the aim of this study is not exactly the same. Like any other wide-ranging typological research, de Haan's study aims at an evaluation of typological universality of features using data from various languages. In other words, for de Haan feature universality is *explanandum* and languages are *explanans*. In contrast, this study uses feature-analysis to evaluate the typological distance between evidentiality systems of any two languages in the sample, which means that here mutual distances between languages are *explanandum* and features are *explanans*. In this respect, the first aim of the present study differs from the goals of mainstream typology, conforming more to the goals of areal linguistics and dialectology (see van der Auwera 1998a: 20 for a similar affiliation).

The second goal of this dissertation falls under the broad heading 'status of evidentiality as a functional category'. Of major interest here are the relationships between evidentiality and other functional categories, such as epistemic modality. I will discuss collocations of grammatical markers of evidentiality and lexical markers of epistemicity or evidentiality and draw conclusions concerning the boundaries of the functional domain of evidentiality and its relationship with the functional domain of epistemic modality. It will be shown that even languages which historically have had a definitional role in the promotion of evidentiality as a universal functional category provide some evidence for the categorical affinity between evidentiality and epistemicity.

The third goal of the study is to contribute to the diachronic research on evidentiality. The main focus is probably on the most common strategy of grammatical evidentiality coding in Eurasia, – one using participial or other non-finite verb form with past reference. The study focuses on the Estonian evidential construction with past participle comparing different hypotheses about its origin. New synchronic evidence from Estonian permits us to single out the probable sources of this construction from the improbable ones.

The following discussion will refer to these three goals respectively as a) a comparison of evidentiality systems and assessment of typological distances between them, b) the status of evidentiality as a functional category, and its relationship with epistemic modality in particular, and c) the historical development of evidentiality markers.

1.2. Methods and material

Before going into methodological details, the present study shall be placed in the perspective of current linguistics according to the following three criteria: research field, theoretical affiliation and basic method.

The research field of this study is, by and large, the overlapping area of areal linguistics and language typology (see Dahl and Koptjevskaja-Tamm 2001 for the relationship between these frameworks). This is due to the fact that I will not only compare languages from one linguistic area, but languages from two discontinuous areas, and thus not only languages, but also areas, enlarging this way the object of comparison. The relatively narrow geographical scope (the Balkan Peninsula and the eastern coast of the Baltic Sea) and the use of a great number of parameters for comparison allows me to measure precisely the mutual distances between varieties, a purpose compatible with that of areal linguistics and dialectology.

Concerning the theoretical affiliation, this study attempts to remain theory-neutral as far as possible. Nonetheless, the perspective adopted is functional and therefore one can place it within the framework of functional typology (see Croft 2003: 2). Unlike many typological studies, however, which rely on the premises and constructs of cognitive linguistics, I will not presuppose any theoretical models of cognitive reality. This way I will adhere to Haspelmath's view, according to which a functional explanation of linguistic phenomena does not need a cognitively realistic description of language, but can make it with a phenomenological description, which "is agnostic about what the speakers' mental patterns are" (Haspelmath 2004: 569).

Adhering to another view, according to which typology is a theory-neutral method, not a theory in itself (see e.g. Whaley 1997: 7; Nichols 2007), I consider typological comparison to be the basic method of this research. I will compare languages according to a set of pre-defined theory-neutral criteria. Each specific criterion forms a parameter for comparison of language systems. Judging from the ratio between the number of languages compared and the number of parameters for comparison, this study can be described as a small sample typology with a relatively large number of parameters for comparison, and opposed to most of the current typological research applying a small number of parameters to a large language sample (cf. for example De Haan 2005b, 2005c). The tension between the linguistic scope (size of the sample) and the explanatory depth can be discerned throughout the articles included in this dissertation. The first and the last article (I and IV, respectively) compare in detail the evidentiality systems of Bulgarian and Estonian, whereas the intermediate two articles (II and III) deal with a greater number of languages where the comparison is much more mechanical. Considering that articles I and IV deal with only two languages, it is probably not entirely appropriate to characterise their approach and their contribution as typological. It would

perhaps be more appropriate to associate them with the field of contrastive linguistics. However, this is not a significant transgression from the methodological affiliation postulated above as typology and contrastive studies use similar type of analytical comparison to highlight linguistic variation.

Articles II and III were published in Estonian and in order to make their contribution accessible to a wider audience, I will discuss their contents in greater detail. The method used and the results obtained in these articles need some refinement. In addition, I will apply an entirely new data processing method, which will help to draw more general conclusions about the typological distances between evidentiality systems. This method retrieves typological distances between languages (or between their structural elements) in terms of a quantum of shared features. The refinements will be addressed at length in Section 4.1., in the conclusions of this study, and in Appendices 3 and 4. In the remaining part of this section I will give some preliminary information about the general procedure and will introduce the new method.

Articles II and III applied 32 parameters for comparison of evidentiality systems. Each parameter was stated as a particular feature of evidentiality systems. Such features are for example the use of the infinitive as an evidential marker or the ability of the formally marked term of the evidentiality system to express inferential meanings. In Section 4.1. I will revise the information about different parameter values with linguistic data acquired in the years after the publication of these articles. I will also critically evaluate the adequacy of these parameters – their mutual consistency relations, the possibility to collapse two (or more) parameters into one, the information available for their values in different languages, and the dispersability of each parameter into a fixed number of discrete values. I will also apply a new criterion to ensure that all parameters are universally relevant for the expression of evidentiality and to avoid areal idiosyncrasies conditioned by other structural factors. If a feature occurs only in one of the areas (the Balkan or Baltic area respectively), then it is added to the checklist only if it is attested in at least one language outside of the given area. This criterion is termed ‘typological relevance’. This entire procedure will reduce the number of parameters for comparison to 19.

Let me now introduce the parameter values. In articles II and III each parameter is seen as a variable with four values:

- + the feature is present,
- (+) the feature is documented, but is not central to the category, is very infrequent or restricted only to some dialects,
- (–) the feature occurs only incidentally or the information about its occurrence is of doubtful credibility,
- the feature is not present,
- ? no information is available.

The values for each parameter in each language are presented in the tables in the end of articles II and III. In the period after the publication of these articles I

concentrated on looking for information about cases marked with (–) and ?, and in Tables 1–3 in Appendix 3 I have managed to eliminate the ambivalent value (–) collapsing its instantiations with (+) or –. As a result, the meaning of value – was expanded to ‘the feature is not present or extremely rare’.

Let me take as an example the feature ‘co-occurrence with morphological mood’. It checks whether evidentiality markers (which in the majority of the languages at hand are considered to be mood markers) can co-occur with other grammatical mood markers within the same predication unit. This is possible in Turkish and in Latvian indexed with +, in some restricted dialects of Lithuanian, which receives (+), but is not possible in the other languages which all receive –.

As a next step, I will process the tabular data using the “isopleth method” introduced in linguistic typology by van der Auwera (1998a, 1998b, 1998c, 1998d). The basic aim of this method is to draw a geographical map which outlines areas of convergence much better than the traditional isogloss maps. According to van der Auwera (1998d: 260) “[i]n these maps the demarcation lines called ‘isopleths’, mark off areas of languages displaying the same number of *plethora* of features, but not necessarily the same features.” Van der Auwera (1998d) wanted to find out which language or vernacular is closest to the epicentre of the Balkan linguistic area. He used 10 features (or isoglosses)² and a tripartite value system (‘1’ the feature is instantiated, ‘0.5’ the feature is instantiated to some extent, ‘0’ the feature is not instantiated). Van der Auwera ranked nine languages according to the number of features they instantiate. The resulting ranking list is as follows: Bulgarian (10), Tosk Albanian (9.5), Macedonian (9.5), Romanian (9), Gheg Albanian (7.5), Greek (6), Serbian/Croatian (5.5), Romani (5.5), Turkish (3). The isopleth map on Figure 1 groups together vernaculars that exhibit the same number of Balkan features, but not necessarily the same features. Each isopleth line corresponds to a certain score of features. The innermost line, for example, encloses the language with the highest score (10/10), which happens to be Bulgarian.

² These isoglosses comprise the most significant non-lexical Balkanisms: a stressed mid-to-high central vowel, a vowel inventory of *i*, *e*, *a*, *o*, *u* without phonological contrast of quantity, openness or nasalization, the merger of genitive and dative cases, a future tense based on reduced form of ‘want’ verb, clitic postposed definite article, evidentials, admiring or dubitative verb forms, infinitive loss, an analytic adjectival comparative, object doubling, the structure of numbers 11 to 19 following the pattern ‘one/two/etc.-upon-ten’.

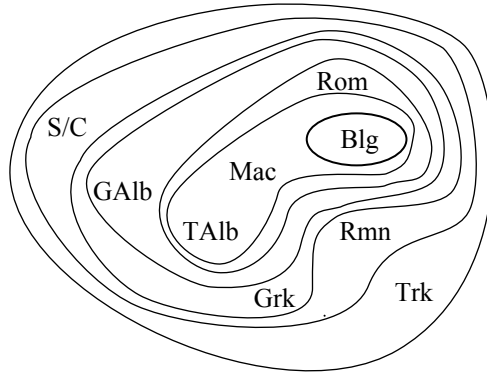


Figure 1. Balkanisms: an isopleth map based on 10 isoglosses (van der Auwera 1998d: 262)³

In this study, I will compare evidentiality systems from two distinct (i.e. non-contiguous) linguistic areas. In particular, I will measure the typological distance between each two evidentiality systems of the sample in terms of the number of features shared. As I am comparing seven languages, I need seven maps, each one illustrating the typological distances of six languages from the language on which the map is based. This is shown in Figure 2, in which Bulgarian, Macedonian, Albanian, Turkish, Lithuanian, Latvian and Estonian are compared according to some 10 hypothetical features. Unlike van der Auwera (1998d), I will not grant half points, which means that if I use the same three values (–, +), +) as in the comparison of the evidentiality systems, I should specify where exactly I draw the boundary between considering a feature present or not present. It seems reasonable to draw the boundary between values – and (+). Thus, two languages share a certain feature, if and only if both have +/(+) value, or both have the – value. Figure 2 illustrates a case in which the typological proximity to Estonian is measured and it turns out that Estonian shares 10 features with Latvian, 9 with Lithuanian, 7 with Bulgarian, 6 with Macedonian, 5 with Albanian and 4 with Turkish. Proximity can be illustrated with line weight and style, as is done in van der Auwera (1998a; 1998b):

³ The location of the languages on the map roughly corresponds to the geographical location of the languages in the Balkans.

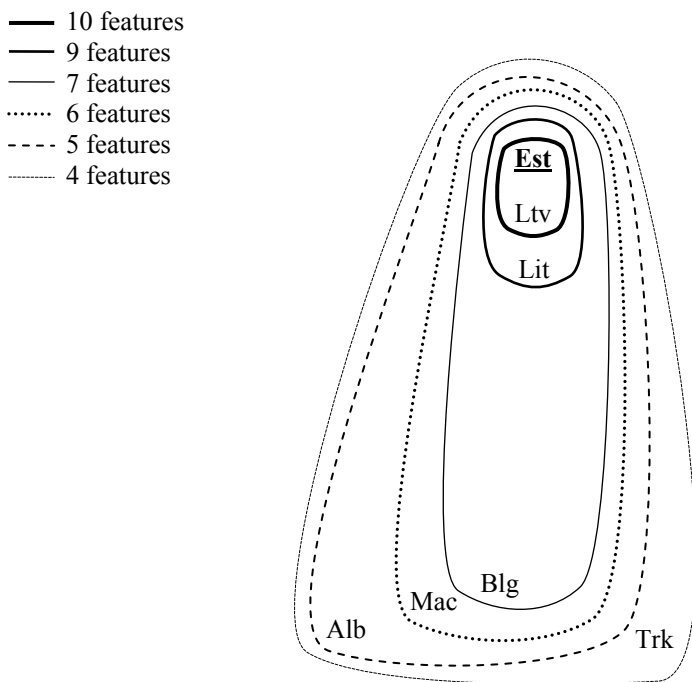


Figure 2. Illustration of an isopleth map showing the degree of proximity to Estonian⁴

As in van der Auwera's map, here each isopleth line corresponds to a certain common number of features (or isoglosses), which need not be the same for any two pairs of languages. Thus, the fact that Bulgarian shares 7 and Macedonian 6 features with Estonian does not mean that Bulgarian shares with Estonian exactly the same six features which Macedonian shares with Estonian. The features for Bulgarian and Estonian may well be different – what is compared is not features, but the *number* of features. Unlike van der Auwera (1998d), I will not operate with isogloss maps. Isoglosses can be directly derived from the value assignments for each feature in each language, but I will leave out this step because 19 isogloss maps (corresponding to 19 features) is too much; their overlay would only obscure the tendencies.

This method allows us to compare the typological distance between languages with their genetic and geographical distance. Unlike the traditional isogloss maps, it is a fairly simple heuristic for detecting the degree of structural similarity between language systems.

⁴ Here too the location of the languages on the map roughly corresponds to their geographical location in Eastern Europe.

Let me elaborate briefly on the relationships between the three types of distance: typological, genetic and geographical. I start from the assumption that typological distance is a dependent variable. If it turns out that typological distance coincides with genetic distance, then it can be explained in terms of genetic affinity. If, on the other hand, typological distance coincides with geographical distance, then it can be explained in terms of areal convergence. If typological distance coincides with both, genetic and geographical distance, then it is probably a derivative of both of them. Challenging questions arise only in cases in which typological distance coincides neither with genetic, nor with geographical distance. There are three types of explanation of such cases: (i) the observable typological distance is due to universal language-processing principles (see e.g. Whaley 1997: 47–49; van der Auwera 1998a: 15), (ii) the observable typological distance is due to previously unknown and thus “invisible” convergence area(s), (iii) the observable distance is due to a chance.

Finally, it should be specified how the information about the parameter values was collected. I have used the following sources:

- 1) personal knowledge⁵;
- 2) different reference and pedagogical grammars;
- 3) various narrow-scope papers and monographs on evidentiality and related topics;
- 4) language informants;⁶
- 5) Internet.⁷

⁵ My personal knowledge of the relevant languages is as follows: Bulgarian (native), Estonian (very good), Macedonian, Latvian (passive), Turkish, Lithuanian (basic), Albanian (none).

⁶ Language guides were occasionally used for Turkish, Albanian, Latvian and Lithuanian. I have not used questionnaires or applied any other elicitation techniques. Native speakers or specialists in these languages were consulted only in case the sources mentioned in points 1–3 did not contain information about certain parameters.

⁷ I have used Google to look for particular forms or collocations of forms in Bulgarian, Macedonian and Estonian.

2. EVIDENTIALITY AS A GRAMMATICAL CATEGORY

2.1. Key definitions and problems relevant to evidentiality

The linguistic term ‘evidential(ity)’ is quite recent compared to other grammatical notions such as ‘tense’, ‘aspect’, and ‘modality’. The first who used it in its modern sense as a label of generic grammatical category marking the source of information was Jakobson in 1957 (reprinted in 1971). Since Jakobson the notion of evidentiality has gone through several rounds of scholarly discrimination and refinement, and therefore it makes sense to start with a definition of evidentiality which is, first, recent, and second, captures linguistic evidence from as many languages as possible. Such is Aikhenvald’s definition of evidentiality which is based on the study of over 500 languages. It states: “[in] about a quarter of the world’s languages, every statement must specify the type of source on which it is based – for example, whether the speaker saw it, or heard it, or inferred it from indirect evidence, or learnt it from someone else. This grammatical category, whose primary meaning is information source is called ‘evidentiality’.” (Aikhenvald 2004: 1). Of crucial importance here is the fact that Aikhenvald does not define the category of evidentiality only semantically, but also on the basis of its formal properties by saying that evidentiality is grammatical category. This restriction excludes all lexical evidential markers from the extension of the cross-linguistic category of evidentiality.

It is also clear from the above definition that the notion of evidentiality covers various types of evidence which specify the way in which the information is acquired. Accordingly, it has become a convention in the literature on evidentiality to talk about evidential specifications such as ‘visual’ (when information is acquired through seeing, e.g. (1a) from Tuyuka), ‘auditory’ (when information is acquired through hearing, e.g. (1b) from Koasati), ‘inferential’ (when information is traced down on the basis of physical or other evidence, e.g. (1c) from Eastern Pomo), ‘assumed’ (usually based on common knowledge, e.g. (1d) from Tsafiki) and ‘reported’ (based on second- or third-hand hearsay, e.g. (1e) from Oksapmin).

- (1) a) *diiga apé-wi*.
soccer play-VIS.3SG.MASC.PAST
‘He played soccer (I saw him play).’ (Barnes 1984, quoted in De Haan 2005a: 384)
- b) *nipó-k aksóhka-ha*
meat-SUBJ char-AUD
‘It sounds like the meat is charring.’ (Kimball 1991, quoted in De Haan 2005a: 390)

- c) *bé·k-al* *p^ha·bé-k-ine*
 3pl-PATIENT burn-PUNCTUAL-INFERENTIAL
 ‘They must have gotten burned’ (I see circumstantial evidence – signs of a fire, bandages, burn cream) (McLendon 2003, quoted in Aikhenvald 2004: 53)
- d) *Manuel ano fi-n-ki-e*
 Manuel food eat-NOMN-VCLASS:do-DECL
 ‘Manuel ate’ (he always eats at eight o’clock and it’s now nine o’clock) (Dickinson 1999, 2000, quoted in Aikhenvald 2004: 54)
- e) *Haperaapnong mahan kuu gaamin tit*
 Haperap:to over.there women husband.and.wife one
pipaa-ri
 went-REP
 ‘A husband and wife went (reportedly) over there to Haperap’
 (Lawrence 1987, quoted in Aikhenvald 2004: 46–47)

These specifications (or distinctions) can be viewed as evidential primitives, which are grouped in a different way in the evidentiality systems of the different languages. It is common for example, that languages subsume ‘visual’ and ‘auditory’ under the heading ‘sensory’ realized with a single marker, or subsume ‘inferential’, ‘assumed’ and ‘reported’ under another broad heading commonly named ‘non-firsthand evidentiality’. I have tried to use Aikhenvald’s term-labels in this dissertation, the only exceptions being that the label ‘indirect’ is occasionally used instead of ‘non-firsthand’, and the label ‘quotative’ is sometimes used instead of ‘reported’.

Before proceeding with the demonstration of different types of evidentiality systems, I shall briefly discuss the question about the relation between evidentiality and the adjacent semantic domain of epistemic modality. This is crucial, because it boils down to the question of what is and what is not an evidential (see Aikhenvald 2004: 3–11 who devoted an entire chapter to this question). One can distinguish between two major viewpoints concerning this relation. According to the first, evidentiality is not a subcategory of modality, or in other words modality and evidentiality are distinct (non-overlapping) functional categories. The primary function of evidentials is to mark the way in which information was acquired without necessarily expressing the speaker’s certainty about the truth of the proposition (see Aikhenvald 2004: 3). Works advocating this view include Jakobson (1971), Kozintseva (1994), De Haan (1999), Lazard (1999, 2001), Nuyts (2001: 27–28), DeLancey (2001) and Aikhenvald (2003, 2004: 7). According to the second viewpoint, evidentiality and the expression of certainty/reliability are inherently related and the domains of evidentiality and modality are at least partly overlapping (see e.g. Lyons

1977: 799–800; Chafe and Nichols 1986: vii; Willett 1988: 52; Dendale and Tasmowski 2001: 343; Palmer 2001: 8–9; Dahl 1985: 148, 190; Kiefer 1994; Boye 2006: 21).

2.2. Types of evidentiality systems

Since the seminal anthology edited by Chafe and Nichols (1986) and Willett’s comprehensive cross-linguistic survey (1988), the major effort in the research of evidentiality has been concentrated on two issues: i) classification of evidential specifications (or evidential values), and ii) classification of evidentiality systems, or in other words classification according to the strategies used by languages to arrange these evidential values in paradigms. A brief account of the cross-linguistically most frequent evidential values was provided in Section 2.1. In the section at hand, I will expand on the second issue. I will briefly introduce the major types of evidentiality systems using Aikhenvald’s classification (Aikhenvald 2003, 2004), which is by far the most comprehensive existing classification of evidentiality systems.

Aikhenvald distinguishes between two-term (A), three-term (B), four-term (C), and five-(or-more)-term evidentiality systems (D). Each of the types is further split up into subtypes according to the values of the terms. Macrotype A comprises systems distinguishing between two evidential values (or choices). Such are A₁ FIRSTHAND AND NON-FIRSTHAND, A₂ NON-FIRSTHAND VERSUS ‘EVERYTHING ELSE’, A₃ REPORTED (OR ‘HEARSAY’) VERSUS ‘EVERYTHING ELSE’, A₄ SENSORY EVIDENCE AND REPORTED (OR ‘HEARSAY’) and A₅ AUDITORY VERSUS ‘EVERYTHING ELSE’. Example 2 from Ngiyambaa (Donaldson 1980, quoted in Aikhenvald 2004: 34–35) demonstrates type A₄ with ‘sensory’ term covering all senses (cf. (2a) instantiating hearing) and ‘reported’ term covering information acquired by hearsay, cf. (2b):

- (2) a) *gabuga:-gara=lu* *ɲamumiyi*
 egg+ABS-SENS.EV=3ERG lay+PAST
 ‘It’s laid an egg’ (by the sound of it)
- b) *bura:y-dja=lu* *ga:-y-aga*
 child+ABS-REP=3ABS bring-CONJ.M-IRR
 ‘It’s said that she is going to bring the children’

Three-term systems involve at least one sensory value and include five attested types: B₁ ‘DIRECT (OR VISUAL), INFERRED, REPORTED’, B₂ ‘VISUAL, NON-VISUAL SENSORY, INFERRED’, B₃ ‘VISUAL, NON-VISUAL SENSORY, REPORTED’, B₄ ‘NON-VISUAL SENSORY, INFERRED, REPORTED’, and B₅ ‘REPORTED,

QUOTATIVE, ‘EVERYTHING ELSE’’. Example (3) from Wanka Quechua (Floyd 1997, 1999, quoted in Aikhenvald 2004: 43) exemplifies type B₁:

- (3) a) *Chay-chruu-mi achka wamla-pis walashr-pis*
 this-LOC-DIR.EV many girl-TOO boy-TOO
alma-ku-lkaa-ña
 bathe-RFL-IMPF.PL-NARR.PAST
 ‘Many girls and boys were swimming’ (I saw them)
- b) *Daañu pawa-shra-si ka-ya-n-**chr**-ari*
 field finish-PART-EVEN be-IMP-3-INFR-EMPH
 ‘It (the field) might be completely destroyed’ (I infer)
- c) *Ancha-p-**shi** wa’a-chi-nki wamla-a-ta*
 too.much-GEN-REP cry-CAUS-2 girl-1p-ACC
 ‘You make my daughter cry too much’ (they tell me)

Four-term systems include the following three subtypes: C₁ ‘VISUAL, NON-VISUAL SENSORY, INFERRED, REPORTED’, C₂ ‘DIRECT (OR VISUAL), INFERRED, ASSUMED, REPORTED’ and C₃ ‘DIRECT, INFERRED, REPORTED, QUOTATIVE’. Eastern Pomo sentences in (4) (McLendon 2003, quoted in Aikhenvald 2004: 52–53) exemplify type C₁:

- (4) a) *mi-p-al p^ha-bé-k-a*
 3.sg.-male-PATIENT burn-PUNCTUAL-DIRECT
 ‘He got burned’ (I have direct evidence, e.g. I saw it happen)
- b) *bi-Yá p^ha-bé-k^h-**ink**’e*
 hand burn-PUNCTUAL-SENSORY
 ‘I burned my hand’ (I feel the sensation of burning in my hand)
- c) *bé-k-al p^ha-bé-k-**ine***
 3pl-PATENT burn-PUNCTUAL-INFERENTIAL
 ‘They must have gotten burned’ (I see circumstantial evidence – signs of a fire, bandages, burn cream)
- d) *bé-k-al p^ha-bé-k^h-**le***
 3pl-PATENT burn-PUNCTUAL-REPORTED
 ‘They got burned, they say’ (I am reporting what I was told)

Evidentiality systems that contain five or more distinctions are rather rare. Witness (5) from Tariana (Aikhenvald 2004: 2–3) encoding the distinction between VISUAL, NON-VISUAL SENSORY, INFERRED, ASSUMED and REPORTED:

- (5) a) *Juse ifida di-manika-ka*
 José football 3sgnf-play-REC.P.VIS
 ‘José has played football (we saw it)’
- b) *Juse ifida di-manika-mahka*
 José football 3sgnf-play-REC.P.NONVIS
 ‘José has played football (we heard it)’
- c) *Juse ifida di-manika-nihka*
 José football 3sgnf-play-REC.P.INFR
 ‘José has played football (we infer it from visual evidence)’
- d) *Juse ifida di-manika-sika*
 José football 3sgnf-play-REC.P.ASSUM
 ‘José has played football (we assume this on the basis of what we already know)’
- e) *Juse ifida di-manika-pidaka*
 José football 3sgnf-play-REC.P.REP
 ‘José has played football (we were told)’

3. GEOGRAPHICAL DISTRIBUTION OF EVIDENTIALITY SYSTEMS

In a dedicated world map, Aikhenvald (2004: 303) outlines three large loci of grammatical evidentiality. These are the Eurasian area (also known as ‘Balkan-West Asia evidentiality belt’ (see DeLancey 2001; Lazard 2001) or ‘pan-Eurasian evidential perfect’ (see Nichols 1986: 253)) with a probable ‘epi-centre’ in Turkic (Aikhenvald 2004: 291), the North American area (another belt extending from Alaska to the southeast, reaching the south-eastern states of the US), and the South American area (comprising several Amazonian and Andean languages). Other, narrower and more isolated areas include New Guinea and Central Australia (Aikhenvald 2004: 288–293, 303). The most comprehensive evidentiality systems (with five or more distinctions) are found in the Americas and in New Guinea (see Aikhenvald 2004: 60–63).

3.1. Hotbeds of evidentiality in Europe and their relations to larger Eurasian patterns

Within Europe, grammatical evidentiality is clearly an Eastern European phenomenon; the westernmost language with a grammatical evidentiality system in Europe is Albanian.⁸ Within these boundaries we can identify four hotbeds of grammatical evidentiality (see Appendix 2):

- 1) The eastern coast of the Baltic Sea (Estonian, Livonian, Latvian and Lithuanian (see Stolz 1991; Klaas 1997; Wälchli 2000; Holvoet 2007: 81–105);
- 2) the Balkans (Bulgarian, Macedonian, Torlak Serbian, Albanian, Turkish (including Gagauz), Megleno-Romanian, Aromanian and some dialects of Balkan Romani (see Friedman 1986, 1994, 2003; Lindstedt 2000b);
- 3) the Caucasus (in order to remain within the geographical borders of Europe, this would include only languages spoken on the northern slopes of the Greater Caucasus Mountains or to the north of them. This includes languages like Chechen, Ingush, Godoberi, Tsez, Circassian and Nogai (see Dobrushina and Tatevosov 1996: 94–97; Chirikba 2003; Johanson 2003; De Haan 2005b; Comrie and Polinsky 2007));
- 4) the Volga-Kama area which can be extended to northeast to include Komi and to southeast to include Kazakh. This area comprises Mari, Komi,

⁸ This of course holds only if we, following Aikhenvald (2004), exclude from consideration the modal auxiliaries which in many European languages (e.g. Dutch *moeten*, German *sollen* or Polish *mieć*) have evidential extensions. (I am grateful to Björn Wiemer for this remark)

Udmurt, Chuvash, Tatar, Bashkir and Kazakh (see Serebrennikov 1960: 58–60, 119–120, 160–170, 259–265; Leinonen 2000; Johanson 2003; Siegl 2004). It is beyond doubt that evidentiality is grammaticalized to a different extent in the languages listed above, and that some of these languages have evidentiality strategies, where other categories have evidential overtones, rather than evidentiality proper, where the language has a dedicated evidential paradigm (see Aikhenvald 2004: 105 about this distinction). Nonetheless, these four areas are generally acknowledged to be rich in evidentiality systems (see for example Aikhenvald 2004: 288–291). If we zoom out the area in scope, we will see that the last three areas are connected with the larger Eurasian area (see Appendix 2) that has been labelled Balkan-West Asia evidentiality belt (see DeLancey 2001: 370). Notably, these three areas comprise at least one Turkic language and as a rule evidentiality is attested earlier in Turkic than in non-Turkic coterritorial languages.⁹ Furthermore, even the Baltic evidentiality area may be linked to this Eurasian area if one reconstructs a Volga-Oka linguistic area (see Helimski 2003: 160–161) including the extinct Eastern Baltic (Eastern Galindian), Mordvin, Mari and at least three extinct languages (Muroma, Meščera and Merya) which can be located taxonomically at the split of the Proto-Finno-Saamic and Volgaic branches of Finno-Ugric. Very little is known about the grammatical structures of these extinct languages, but they are all genetically related to languages which have grammatical evidentiality systems (see also a similar remark by Lindstedt (2000a: 375) concerning the ties between these areas). However, this would be too far-fetched a connection, because the arguable Volga-Oka area did not include any Turkic language whatsoever. Moreover, the only Turkic language spoken in the Baltic area, Karaim, does not have a grammatical evidentiality system (see Johanson 2003).

As the history of language contacts in Eastern Europe is beyond the purview of this study, in the next section I will instead narrow the scope, focusing on the Baltic and Balkan areas without assuming any historical connections between them.

3.2. Evidentiality systems in the Balkan and Baltic areas

There is no one-to-one correspondence between the popular notion of ‘Balkan Sprachbund’ and the sample of Balkan languages chosen for this study. As already noted in Section 3.1., the Balkan languages which are generally agreed to have grammatical evidentiality systems are Bulgarian, Macedonian, Torlak Serbian, Turkish, Albanian, Megleno-Romanian, Aromanian and some dialects

⁹ The only exception to this rule seem to be West Caucasian evidentiality systems (see Chirikba 2003: 266–267; Aikhenvald 2004: 289).

of Balkan Romani. In contrast, most of the accounts of the Balkan Sprachbund would exclude Standard Turkish and insert Greek and Romanian and sometimes Ladino (Judezmo) to the list (cf. for example Schaller 1975, Lindstedt 2000b and Asenova 2002).

Along similar lines, what is referred to as the (Circum-)Baltic linguistic area varies (see e.g. Haarmann 1976: 106–116; Mathiassen 1996; Helimski 2003: 158–159; Dahl and Koptjevskaja-Tamm xv–xx), but is on any account larger than the Baltic hotbed of evidentiality (Lithuanian, Latvian, Livonian and Estonian), including also languages like Finnish, Vote, Swedish, Baltic German etc.

This dissertation discusses to a different extent the evidentiality systems of (in order of decreasing attention) Estonian, Bulgarian, Latvian, Lithuanian, Turkish, Albanian and Macedonian. The evidentiality systems of these languages are relatively well studied. The reasons to exclude the remaining languages with evidentiality systems (Torlak Serbian, Megleno-Romanian, Aromanian, Balkan Romani and Livonian) from the study are as follows: first, the properties of their evidentiality systems are not so well recorded in the literature (i.e. for many parameters information is missing), and second, in the majority of these languages (Megleno-Romanian, Aromanian and Balkan Romani) evidentiality systems have emerged only recently as a result of language contact with the neighbouring languages with greater speech communities (see Friedman 1994, 2003 on the Balkan languages; see also Ikola 1953: 49 on possible Latvian influence on the Livonian evidentiality system). I assume that their typological distance from the adjacent languages would coincide with their geographical distance from them, and therefore believe that their inclusion would not have resulted in any interesting discoveries concerning the areal patterns of evidentiality coding.

The description of the 19 parameters on which the language comparison is based in Section 4.1. will illuminate the main formal, semantic and syntagmatic characteristics of the seven evidentiality systems under concern. For a more comprehensive description of these evidentiality systems see GBE II (351–365), Friedman (1986) and Nitsolova (2006) for Bulgarian, Koneski (1967: 380–381) and Friedman (1988b) for Macedonian, Meydan (1996) and Slobin and Aksu (1982) for Turkish, Fiedler (1966) and Duchet and Pernäska (1996) for Albanian, Kask (1984: 220–285), Toomet (2000) and Sepper (2002) for Estonian, Endzelin (1922: 757–761) and Holvoet (2007: 81–105) for Latvian and GLJ (231–235) and Holvoet (2007: 81–105) for Lithuanian. In the remaining part of this section a very concise description of the semantics and the markedness patterns of these systems will be presented.

In Aikhenvald's classification Balkan Slavic (Bulgarian and Macedonian) evidentiality systems are somewhat hesitantly classified as A₁ systems (see Aikhenvald 2004: 264, 40, 288–289, 298), the Albanian system is classified as A₂ (ibid. 280, 288), Turkish as A₂ (ibid. 30), Estonian and Latvian as A₃ systems (ibid. 33) and the Lithuanian system is left unclassified.

With regard to the markedness patterns of these systems, the formal and the functional markedness do not always coincide (see Aikhenvald 2004: 70–75). Such is the situation in Balkan Slavic evidentiality systems where the formally marked term is NON-FIRSTHAND. This term is encoded by the Slavic *l*-participles (formed from both aorist and imperfective stems) and (in some cases) by the omission of the finite auxiliary. It is remarkable, however, that it is the opposite term (FIRSTHAND) that seems to be functionally marked in this dichotomy, while NON-FIRSTHAND is functionally unmarked (see Friedman 2003). In the remaining languages, formal and functional markedness seem to coincide. In Turkish, the functionally marked term NON-FIRSTHAND is formally coded by the suffix *-miş*. In Albanian, the same functionally marked term is coded by inverted perfect and pluperfect forms consisting of a shortened participle and the conjugated auxiliary ‘have’ agglutinated to it. In Estonian and Latvian, the term REPORTED is marked by present and past participles occurring in an auxiliary-less construction. The present participles (and the future participle in Latvian) occur in a petrified form. Other ways to encode this functionally marked term include using infinitives or other nominalizations. Finally, the Lithuanian evidentiality system, which was not classified in Aikhenvald’s study, is probably of type A₂ where the term NON-FIRSTHAND can be encoded with two entirely different strategies: 1) inflected active participles occurring in an auxiliary-less construction, and 2) passive participles (derived from intransitive verbs) occurring in an auxiliary-less construction with sentence-initial actor argument in genitive.

4. MAJOR QUESTIONS DEALT WITH IN ARTICLES I-IV

The major questions dealt with in articles I–IV correspond to the three main goals of the dissertation as stated in Section 1.1.

The first goal concerns a comparison of the evidentiality systems of the languages of the Balkan and Baltic areas and an assessment of the typological distances between them. Although these issues are approached already in article I (only for Bulgarian and Estonian), an explicit discussion of the parameters used for evaluation of the typological distances is presented in articles II and III, both of which end with a tabular summary showing the parameter values for each evidentiality system. The description of the relevant parameters in these two articles is, however, incomplete, and some of the parameters need revision.

The second goal concerns the status of evidentiality as a functional category, and the relationship between evidentiality and epistemic modality in particular. Although also touched upon in articles I–III, the major contribution for revealing the nature of this relationship is made in article IV.

The last goal concerns the historical development of evidentiality markers. Although this topic is touched upon in articles I–III, it is discussed at length in an independent study by Kehayov and Siegl (Kehayov and Siegl 2007). This study is not included in the present dissertation for two reasons. First, a considerable part of it deals with languages which do not fall within the scope of the present dissertation, and second, it is a co-authored study in which the discussion of the internal evidence for the development of evidentiality marking in Estonian is my contribution. In Section 4.3., I will test different hypotheses concerning the development of past participles to evidentiality markers against the synchronic evidence from Estonian. For a more comprehensive discussion on this topic see Kehayov and Siegl (2007).

4.1. Parameters for comparison of the evidentiality systems

One of the main methodological issues in typology concerns the selection of parameters for comparison. In this study, I have applied the following criteria. First, I have trusted the intuitions of descriptive studies of particular evidentiality systems to mention properties that are considered relevant and non-trivial. For example, I have not encountered a single descriptive study explicitly stating that an evidential marker is used in declarative sentences, and therefore occurrence in declaratives was not selected as a parameter. Second, an evidentiality system is characterised in terms of other structural phenomena only if these phenomena exist in all languages at hand. For instance, in order to

test whether evidentiality coding is based on present participles, it had to be previously assured that all languages have a class of present participles. Or before I checked whether evidentials co-occur with morphological mood, I had to first check whether all languages at hand have morphological mood marking. In the revised parameter-set comprising 19 parameters, I apply a third criterion – the one of typological relevance, already introduced and motivated in Section 1.2.

In article II, I apply 17 parameters for comparison, each one defined as a specific feature of the morphosyntax of evidentiality coding or concerning the syntagmatic restrictions governing the use of the evidentiality markers. These features are listed as follows:

- 1) use of a past participle as a marker of evidentiality,
- 2) use of a present participle as a marker of evidentiality,
- 3) use of an infinitive as marker of evidentiality,
- 4) use of other deverbal noun as a marker of evidentiality,
- 5) zero-auxiliary as a marker of evidentiality distinctions,
- 6) agglutination of a compound tense form into a single evidential form,
- 7) use of a petrified morpheme,
- 8) occurrence of a distance particle,
- 9) switching between different auxiliaries as a marker of evidentiality,
- 10) suppletion, i.e. switching between different stems of the same auxiliary as an additional device used in the marking of evidentiality,
- 11) inversion of the main verb form and the auxiliary as a marker of evidentiality,
- 12) evidentiality coding on the first unit in compound verb forms,
- 13) voice alternation used as evidentiality strategy,
- 14) compound tense forms degraded to simple tense forms with evidential meaning,
- 15) the marked evidential form occurs in interrogative sentences,
- 16) the marked evidential form occurs in exclamatory sentences, and
- 17) the marked evidential form occurs in imperative sentences.

This list is enhanced in the next article (III) with the following 15 semantic parameters. For clarity, these parameters are numbered here consecutively starting from 18:

- 18) the formally marked evidential term co-occurs with “deontic” mood (imperative, optative, debitive or other),
- 19) the formally marked evidential term co-occurs with “epistemic” mood (potential, conditional-irrealis or other),
- 20) the formally marked evidential term covers reported evidentiality,
- 21) the formally marked term covers inferential evidentiality,
- 22) the formally marked term can have mirative meaning, i.e. expression of surprise over suddenly discovered state of affairs,

- 23) the magnitude of the evidentiality distinction varies according to the time scheme, i.e. the magnitude of evidentiality distinctions vary in simple past, perfect, pluperfect etc.,
- 24) existence of tools for coding a tripartite distinction with two marked evidential terms and one unmarked term,
- 25) particular evidential meaning (such as hearsay, inferential or other) can be specified out of a more general evidential term,
- 26) use of devices marking increase in cognitive distance from the information source,
- 27) the formally marked term is used to express generic statements,
- 28) the formally marked term is used by the speaker in reports of his own dreams,
- 29) the term marking indirect evidentiality has endophoric uses, i.e. uses in which the speaker refers to his own internal states,
- 30) the evidential forms are conventionalized in folk narratives (fairy tales, legends etc.) as a basic form of the predicate,
- 31) the evidential forms are conventionalized in historical writings, and
- 32) the evidential forms are used as discourse markers conveying distrust and irony about previously communicated statements.

In the end of articles II and III, each evidentiality system is assigned a value for each parameter and the results are presented in relevant tables. The conclusions obtained after the analysis of assignments are illustrated as two hypothetical evidentiality systems which can be regarded as two extremes between which the evidentiality systems of Bulgarian, Macedonian, Albanian, Turkish, Estonian, Latvian and Lithuanian could be located.

As already mentioned in Section 1.2., the information and the scholarly expertise acquired in the years after the publication of articles II and III call for the revision of some parameters. The remaining part of this section and most of the last section (Conclusions and perspectives for further research) constitute a supplement to these two articles. In what follows, I will discuss the reasons to reduce the number of parameters from 32 to 19. Some parameters will be collapsed, others excluded. After that, each parameter from the new set will be defined and supplied with an example from one of the languages studied.

Starting from the beginning of the above list, parameters 3 and 4 are collapsed into one parameter, which can be stated as ‘use of nominalizations other than participles as evidential markers’. The motivation behind this is that the use of infinitives and other nominalizations (apart from participles) in the coding of evidentiality is so marginal that it is better if they are subsumed under a common parameter. Parameter 6 (agglutination of compound tense form into a single evidential form) is excluded, because it only represents a coalescence of two previously more independent forms into one and seems thus to reflect a universal grammaticalization pattern, which has nothing to do with evidentiality. The characterization ‘petrified morpheme’ in parameter 7 is

deleted and the more adequate ‘dedicated morpheme’ is left. Loss of inflection is also a general feature of structural change, but ‘dedicated’ is relevant for the distinction between evidentiality proper and an evidentiality strategy. Parameters 9 and 10 are excluded for similar reasons. None of them seem to be relevant or specific for evidentiality. They reflect the general tendency in South Slavic to combine different auxiliaries or to use suppletion for marking of complex temporal-aspectual meanings. Parameter 11, also excluded, seem to be an idiosyncratic development of Albanian as I am not aware of any other language that uses the same coding strategy. Next, parameter 12 is excluded because it seems to pertain not only to evidentiality marking but also to any kind of TAM-marking. Along similar lines, the development in parameter 14 seems rather trivial. Reported and other indirect evidentials always have complex reference and if they are historically derived from tense, then it is natural that they develop from compound tenses (such as the perfect or pluperfect). It is also natural for the reference point marker to be omitted during the process of (re-)grammaticalization because the new forms do not have time reference but evidentiality as their basic meaning. Parameter 16 is excluded from the set because exclamatory sentences often encode irony and surprise and therefore it can be considered as a function of ‘reported’ and ‘mirative’ meanings of evidentiality systems which are listed as parameters 20 and 22 respectively. Parameter 17 remains but is specified; its new form is ‘evidential form is used to express reported commands’. Parameters 18 and 19 are collapsed to one parameter labelled ‘co-occurrence with morphological mood’ because some moods (e.g. conditional and subjunctive moods in various languages) are difficult to be classified according to the distinction between ‘deontic-like’ and ‘epistemic-like’ mood. The limitation ‘morphological’ is necessary as it checks the morphotactic restrictions governing the co-occurrence of evidential and mood markers. Parameter 23 is excluded as it is not entirely clear to me whether one could postulate the same paradigmatic set of tenses across the distinction between evidentials and non-evidentials: in many cases evidentials seem to have their own time reference (see Aikhenvald 2004: 99–103, 261). Parameters 26 and 8 are collapsed to one parameter as they denote two aspects (functional and formal) of the same phenomenon. Parameter 29 is rephrased as follows: “evidential forms are used with sensory and mental state verbs as objectivizers of speaker’s feelings”. Following Plungian (2001: 352), I used the term ‘endophoric’ for this feature in Table 2 of article III. As this term commonly denotes certain types of intralinguistic referring expressions (such as anaphora) it is infelicitous for the description of the cases at hand.¹⁰ The definition ‘objectivizer of speaker’s feelings’ is unambiguous, and thus more satisfactory. Parameter (31) is excluded because it is a matter of artificial standardization rather than of natural developments. The last parameter (32) is

¹⁰ Duchet and Pěrnaska use the term *endopathique* (1996: 35).

excluded because anaphoric uses are a function of reported evidentiality (parameter 20) often motivated by irony.

As a next step, I will discuss each parameter from the new set in detail. The discussion will be structured as follows: first, each parameter will be strictly defined; second, references for all languages exhibiting a positive value for the parameter will be added¹¹; third, if necessary I will add references showing that there are languages possessing the feature outside of the Balkan and Baltic areas (as already stated in Section 1.2, such evidence verifies whether the given parameter meets the criterion of ‘typological relevance’); fourth, the positive value of the parameter will be exemplified. I will distinguish between three types of parameters: i) parameters relevant to the formal expression of evidentiality, ii) parameters relevant to the semantics and pragmatics of evidential markers, and iii) parameters relevant to the structural availability of evidential markers. Finally, in order to keep my exposition coherent with the previous body of research, I will briefly compare my parameters with de Haan’s parameters (see De Haan 2005b, 2005c).

4.1.1. Formal parameters

Parameter 1: Use of a past participle as a marker of evidentiality.¹² This morphosyntactic strategy is attested in Turkish (e.g. Slobin and Aksu 1982), Bulgarian (e.g. GBE II: 351–352), Macedonian (e.g. Friedman 1988b), Albanian (e.g. Duchet and Pěrnaska 1996: 31), Lithuanian (e.g. GLJ: 232–233), Latvian (e.g. Holvoet 2001a) and Estonian (e.g. Muižniece, Metslang and Pajusalu 1999). Witness (6) from Bulgarian:

- (6) *Toj otišāl* *na svadba.*
 he go.PST.PTCP.M.SG to wedding
 ‘Allegedly, he went to a wedding.’

Parameter 2: Use of a present participle as a marker of evidentiality. This feature is attested in Lithuanian (e.g. GLJ: 232–233), Latvian (e.g. Holvoet 2007: 82) and Estonian (e.g. Kask 1984: 244–273). Typological relevance is guaranteed by the habitual present participle of the verb ‘say’ in Udege, which is grammaticalized as evidential marker (Nikolayeva and Tolskaya 2001, quoted in Aikhenvald 2004: 272). Witness example (7) from Lithuanian:

¹¹ References concerning the non-occurrence of a feature in a language will not be presented. However, in all cases the non-occurrence of a feature is confirmed.

¹² Thus characterization does not necessarily imply any directionality of change, such as participle → evidentiality marker.

- (7) *Jis rašqs laiška.*
 he write.ACT.PRS.PTCP.M.SG letter.ACC
 ‘Reportedly, he is writing a letter.’

Parameter 3: Other nominalization used as a marker of evidentiality. This includes different kinds of non-finite verb forms (infinitives, converbs and various verbal nouns) except from participles (i.e. verbal adjectives). The feature is attested in Estonian (e.g. Erelt, Metslang and Pajusalu 2006) and in some restricted subdialects of Latvian (e.g. Rudzīte 1964: 376). The feature is relevant due to its occurrence in Purépecha (Taraskan) where infinitive is a common marker of evidentiality in traditional narratives (Chamereau 2000, quoted in Aikhenvald 2004: 118). Witness (8) from Estonian:

- (8) *Ta olla pulmas.*
 s/he be.INF wedding.INE
 ‘Reportedly, s/he is at the wedding.’

Parameter 4: Omission (or loss) of the finite auxiliary in evidential past. This definition embraces two cases: in the first (Turkish, Bulgarian, Lithuanian, Latvian and Estonian), a finite copula/auxiliary is still used in the corresponding indicative compound tense forms, whereas in the second (Macedonian), the copula is lost in the corresponding indicative past tense. The situation is also complicated by the fact that there is an ongoing discussion with regard to some of the languages whether the copula is just omitted in evidential past or deleted, i.e. whether its syntactic position is present or absent (see Holvoet 2007: 96). However, all these cases are similar and can be arranged along a single cline of language change: from presence to omnipresence (omission) and finally to absence (loss) of the copula. This feature is attested in Turkish (e.g. Lewis 1960: 90–93; Johanson 1998), Bulgarian (e.g. GBE II: 351–353), Macedonian (e.g. Friedman 2003: 194), Lithuanian (e.g. Holvoet 2007: 81–105), Latvian (Holvoet 2007: 81–105) and Estonian (e.g. Kask 1984: 243; Künnap 1994). Witness (9) from Latvian:

- (9) *Jana vakar Ø atnākusi mājās.*
 Jana yesterday Ø come.PST.PTCP.F.SG home
 ‘Reportedly, Jana came home yesterday.’

Parameter 5: Existence of a dedicated evidential morpheme. This parameter has a positive value in Albanian (e.g. Fiedler 1966; Duchet and Pernäska 1996), Latvian and Estonian (see Klaas 1997 for both languages). Consider (10), also from Latvian:

- (10) *Jana es-ot mājās.*
 Jana be-REP home
 ‘Reportedly, Jana is home.’

Parameter 6: The participle of the auxiliary/copula verb is used as a “distance particle” (a term initially suggested by Johanson 1998: 146). Here the language user extends the verbal complex with non-finite forms of the auxiliary in order to increase the communicative and attitudinal distance between himself and the reported information.¹³ This feature is present in Turkish (e.g. Johanson 1998), Bulgarian (e.g. Maslov 1981: 270–277), in certain varieties of Macedonian (e.g. Koneski 1967: 482; personal knowledge¹⁴), Albanian (e.g. Fiedler 1966), Lithuanian (e.g. Gronemeyer 1997: 98) and very marginally also in Latvian and Estonian (which are nevertheless given a negative value for this parameter). Witness (11) from Bulgarian. In (11b) the speaker is less certain about the truth of the underlying proposition than in (11a), and by adding the past participle of the auxiliary wants to avoid responsibility:

- (11) a) *Toj živeel mnogo dobre.*
 he live.PST.PTCP.M.SG very well
 ‘Reportedly, he lives very well.’
- b) *Toj bil živeel mnogo dobre.*
 he be.PST.PTCP.M.SG live.PST.PTCP.M.SG very well
 ‘Reportedly, (with lower commitment) he lives very well.’

Parameter 7: The use of morphosyntactic material marking voice to encode evidential distinctions. This restriction includes two possible cases. In the first, attested in Macedonian, passive participles participate in evidentiality marking in active constructions (i.e. constructions which are not characterised by a diathetic shift in argument structure). In the second, typical for Lithuanian, a variety of the entire voice construction functions as a marker of evidentiality. The feature is attested in Macedonian (e.g. Friedman 1988b), Lithuanian (e.g. Holvoet 2007: 90–104) and less significantly in Latvian (e.g. Holvoet 2001c: 122–131). Consider the Lithuanian example in (12):

¹³ Aikhenvald (2004: 157-158) calls this ‘conceptual distance’.

¹⁴ This strategy seems to be quite rare in Macedonian as it is rarely mentioned in the literature. However, one can encounter occurrences like *Kostovski izjavil deka Vladata treba da mu gi vrati akciiite na JSP, bidejki vo toj slucaj JSP namesto so minusno saldo na smetkata bil imal* (=be.PST.PTCP have.PST.PTCP) *pozitiva od 1,3 milioni evra*. ‘Kostovski announced (indefinite past) that the government should give back the shares to JSP, in which case instead of a negative balance JSP **would reportedly have** a positive balance amounting to 1.3 million euro.’ (<http://listserv.buffalo.edu/cgi-bin/wa?A2=ind0704a&L=maknws-l&P=625>, accessed on 15 February 2008)

- (12) *Jo rašo-ma laišk-as.*
 he.GEN.SG write-PASS.PRS.NOM.NT letter-NOM.SG
 ‘He is evidently writing a letter.’ (Ambrazas 1994, quoted in Gronemeyer 1997: 103)

4.1.2 Semantic and pragmatic parameters

Parameter 8: Formally marked evidential term covers the expression of reported evidentiality. The feature is attested in Turkish (e.g. Aksu-Koç and Slobin 1986), Bulgarian (e.g. Nitsolova 2006), Macedonian (e.g. Friedman 1997, 2003), Albanian (e.g. Duchet and Pernäska 1996), Lithuanian (e.g. Gronemeyer 1997), Latvian (e.g. Holvoet 2001a) and Estonian (e.g. Erelt (ed) 2003: 107). Consider (13) from Estonian:

- (13) *Ta ole-vat pulma läinud.*
 s/he be-REP wedding.ILL go.PST.PTCP
 ‘Reportedly, s/he has gone to a wedding.’

Parameter 9: Formally marked evidential term covers the expression of inferential evidentiality. The feature is attested in Turkish (e.g. Aksu-Koç and Slobin 1986), Bulgarian (e.g. Izvorski 1997), Macedonian (e.g. Lindstedt 2000a: 374–378), Albanian (e.g. Friedman 2003) and Lithuanian (e.g. Holvoet 2007: 84–85). Witness (14) from Turkish:

- (14) *El-im-i kes-miş-im*
 hand-1SG.POSS-ACC cut-INDIR-1SG
 ‘I must have cut my hand.’ (adapted from Aksu-Koç and Slobin 1986: 160)

Parameter 10: Formally marked evidential term covers the expression of mirative meanings. In other words, the formally marked term is used to mark speaker’s surprise, or as Aksu-Koç and Slobin (1986) put it, to convey that the speaker’s mind is not prepared to assimilate new information. This feature is attested in Turkish (e.g. Aksu-Koç and Slobin 1986), Bulgarian (e.g. Guentchéva 1996b: 54), Macedonian (e.g. Friedman 2003), Albanian (e.g. Duchet and Pernäska 1996) and Lithuanian (e.g. Holvoet 2007: 85). Example (15) is from Albanian:

- (15) *Sa bukur folke shqip!*
 how well speak.ADM.2SG Albanian
 ‘How well you speak Albanian!’ (adapted from Eintrei 1982: 111)

Parameter 11: Existence of means for expression of a tripartite distinction with an unmarked member. In such case the language has a strategy to mark evidential opposition of the type [UNMARKED / + feature / – feature]. The feature is attested in Turkish (e.g. Friedman 1978: 113) and Macedonian (e.g. Friedman 1986: 177–179). Similar oppositions seem to exist in Comanche (Charney 1993, quoted in Aikhenvald 2004: 50) and in Chemelhuevi (Munro 1978, quoted in Aikhenvald 2004: 51), although on the basis of the evidence presented in Aikhenvald’s work no decisive claims can be made about the precise nature of evidential distinctions in these languages. Witness (16) from Macedonian (based on Friedman 1986: 177):

- (16) a) *beše pravil* (UNMARKED)
 be.IMPF.3SG do.ACT.PST.PTCP
 ‘he had done (it)’ (with no reference to the source of information)
- b) *imaše praveno* (FIRSTHAND)
 have.IMPF.3SG do.PASS.PST.PTCP
 ‘he had (it) done’ (the speaker confirms it)
- c) *imal praveno* (NON-FIRSTHAND)
 have.PST.PTCP do.PASS.PST.PTCP
 ‘he is said to had done (it)’

Parameter 12: A subtype of evidentiality can be specified out of a more general evidential term. In all cases, this involves specification of reported meaning. Two different strategies can be distinguished. In Balkan Slavic and in Turkish, the effect is achieved through adding the participle of the auxiliary verb to the participle of the main verb which is already marked for non-firsthand meanings. The second possibility, instantiated in Lithuanian, involves a situation in which two evidential strategies, one employing active and the other passive participles, which initially have had identical functions, are in process of functional divergence whereby the first (active participles) show specialisation to mark reported evidentiality. The feature is attested in Turkish (e.g. Slobin and Aksu 1982: 194), Bulgarian (personal knowledge), Lithuanian (e.g. Holvoet 2007: 91), and shows limited occurrence in some varieties of Macedonian¹⁵. Witness the Turkish example in (16), adapted from Slobin and Aksu (1982: 194):

- (16) a) *Kemal gel-miş*
 Kemal come-PST.INDIR
 ‘Kemal has reportedly/evidently come.’

¹⁵ Rare sequences with two *l*-participles (one of auxiliary and the other on main verb) in Macedonian as the one shown in the example in footnote 14 seem to convey only reported evidentiality (another example can be found in Koneski 1967: 482).

- b) *Kemal gel-miş- (i)miş*
 Kemal come-PST.INDIR- be-PST.INDIR
 ‘Kemal has reportedly come.’

Parameter 13: Formally marked evidential term encodes generic statements. This means that the evidential marker is used in expressions of epistemic necessity derived from general knowledge (such as for example *Everyone has to die one day*). This feature is attested only in Albanian, but seems to be typologically relevant due to its occurrence in languages like Ladakhi (e.g. Palmer 2001: 38–39). Witness the following Albanian sentence:

- (17) *Po ja që fati qenka fat.*
 and INTERJ COMP destiny.DEF be.PRS.ADM.3SG destiny
 ‘You see, destiny is a destiny.’ (adapted from Duchet and Përnaska 1996: 37)

Parameter 14: Formally marked evidential term can be used when the speaker refers to his own dreams. As a rule, evidential forms are used in specific contexts in which the speaker wants to stress the other-worldliness of the dream situation. This feature is instantiated to varying extent in Turkish (e.g. Slobin and Aksu 1982: 199), Bulgarian (personal knowledge), Macedonian (Friedman 2003: 210–211), Albanian (Victor Friedman, p.c.) and marginally also in Lowland Lithuanian dialects (Axel Holvoet, p.c.). Witness the Turkish example in (18), adapted from Meydan (1996: 131):

- (18) *Bu gece çok güzel bir rüya gör-dü-m. Büyük bir bahçe-dey-miş-im.*
 garden-LOC-INDIR-1SG
 ‘This night I had a very beautiful dream. I was in a huge garden.’

Parameter 15: Evidential forms are used with sensory and mental state verbs as objectivizers of speaker’s feelings. This feature is attested in Albanian (Duchet and Përnaska 1996: 35–36) and its typological relevance is guaranteed by evidence from Tibetan languages (Central Tibetan and Amdo; Tournadre 1996: 206–207). Witness the following Albanian example, adapted from Duchet and Përnaska (1996: 36):

- (19) – *Po më ardhka keq për Franin, – tha*
 me come.PRS.ADM.3SG bad for Fran say.AOR.3SG
vëllai.
 brother.DEF
 ‘– I feel very sorry for Fran, the brother said.’

Parameter 16: The formally marked evidential forms display conventionalized uses in traditional narratives (fairy tales, myths, legends, folk songs, jokes etc.) as basic forms of the predicate. This feature is attested in all languages, but to different extents. While in Turkish (e.g. Slobin and Aksu 1982: 187), Bulgarian (e.g. Demina 1959: 375–376; GBE II: 364–365), Macedonian (e.g. Friedman 2005: 27–30), Latvian (e.g. Wälchli 2000: 192) and Estonian (e.g. Muižniece, Metslang and Pajusalu 1999) the feature seems to be present in most of the genres, in Albanian it is restricted to epic songs and to onsets of fairy tales (Fiedler 1966; Duchet and Pěrnaska 1996: 43–45; Norbert Boretzky, p.c) and in Lithuanian to legends of origin (Wälchli 2000: 192–193). Witness the Latvian example in (20) (Wälchli 2000: 186–187):

- (20) *Skrīver.u pagast.ā ne.tālu no Puriņu*
 Skrīveri.GEN:PL municipality.LOC not.far from Puriņi.GEN:PL
māj.as atrod.as kāda pļava. Š.o
 house.GEN be_located.3:RFL which meadow. This:ACC
pļav.u sauc par Uguns pļav.u. Jāņ.u
 meadow call:3 for Fire:GEN meadow:ACC John.GEN:PL
nakt.ī šī pļava spīd.ot vien.ās ugun.īs,
 night.LOC this meadow shine.EV:PR one.LOC:PL fire.LOC:PL
it kā kad viņa deg.tu. Tādēļ to iesauk.uši
 as if when she burn:SBJ. That_is_why this call.EV:PS:PL:M
par Uguns pļav.u.
 for fire:GEN meadow:ACC

‘In the municipality of Skrīveri not far from the Puriņi farm there is a meadow. This meadow is called Fire Meadow. At Midsummer Eve this meadows **is said to shine** in fire all over as if it would burn. That is why this meadow **has been called** Fire Meadow.

4.1.3. Structural availability parameters

Parameter 17: The formally marked term of the evidentiality system is used in direct non-echoic questions. With this restrictive definition I eliminate indirect questions, i.e. reports of direct questions (e.g. *Jim wonders whether Mary is coming to the exhibition*), and echo questions, i.e. questions repeating the utterance of interlocutor (*Where is John? – Probably in the hospital. – Probably where?*). This feature is attested in Turkish (e.g. Johanson 2003: 286; Friedman 2005: 33–34), Bulgarian (e.g. Friedman 2005: 33–34), Macedonian (e.g. Friedman 2003: 201–202)¹⁶, Albanian (Duchet and Pěrnaska 1996: 36; Friedman 2005: 33–34), and to a lesser extent in Lithuanian (see example (26) in article II which I owe to Axel Holvoet)¹⁷. Witness (21) from Macedonian. Under normal circumstances the sentence conveys that the speaker may have heard that the addressee has a new car, or may be expressing mirativity or dubitativity:

- (21) *Kade si ja kupi-l kola-ta?*
where be+PRES+2SG it+ACC buy+L+MASC car-DEF
'Where did you buy the car?' (Friedman 2003: 202)

Parameter 18: The formally marked evidential term is used to express reported commands or requests. This feature is attested in Turkish (Kononov 1956: 245 ff.; Ceyhan Temürcü, p.c.), Bulgarian (e.g. Demina 1959: 330–332), Macedonian (e.g. Friedman 2003: 202–203), Albanian (e.g. Friedman 2003: 202–203) and Latvian (e.g. Holvoet 2001b: 229, 242–249).¹⁸ Witness (22) from Bulgarian:

- (22) *Neka da razkažel za sebe si.*
let COMP tell_story.IMPF.PST.PTCP about him_self
'(Reportedly) he should explain about himself.'

¹⁶ The claim concerning obligatoriness of exclamation marker in wh-questions made in article II (pp. 826–827) needs to be weakened. As Friedman convincingly shows in a recent study (Friedman 2005: 33–34), in addition to Albanian, wh-questions also need not be marked with an exclamation marker in Turkish and Macedonian.

¹⁷ Holvoet (2007: 109) presents also an example with interrogative sentence containing evidential marker from Latvian, but the question involved seems to be indirect rather than direct.

¹⁸ It should be noted that Estonian has a special paradigm of 'jussive' forms, which is employed for the expression of reported commands or requests (see Erelt 2002). These forms are excluded from the following discussion because i) they are different from the evidential forms used in declaratives and the definition of parameter 18 captures only evidential markers, which are used in declaratives and are extended to commands and requests, and ii) because, as shown by Erelt (2002), the primary function of these forms is modal rather than evidential.

Parameter 19: The formally marked evidential forms co-occur with morphological mood. This feature serves as direct evidence against the claim that evidentials are mood markers. As different mood markers occupy one and the same morphotactic slot, they are expected to be mutually exclusive (see Holvoet 2007: 87 for a same remark). The restrictive qualification ‘morphological’ here brings clarity to the notion of mood as it expels various syntactic constructions with modal meanings which in some descriptive traditions are given mood status, in others are not. The feature is attested in Turkish (e.g. Kononov 1956: 251, Ceyhan Temürçü, p.c.), Latvian (e.g. Holvoet 2007: 87–89) and Lithuanian (e.g. Holvoet 2007: 87–88).¹⁹ Consider example (23) from Latvian:

- (23) *Cilvēki runā, ka, ja izņemot visu akmeni,*
 people say:PRS3 COMP if remove:IRR.OBL whole stone:ACC
būtot redzams ceļš uz eļli.
 be:IRR.OBL visible way to hell:NOM
 ‘People say that if the whole stone were removed the way to hell would become visible.’ (Holvoet 2007: 88)

4.1.4. A comparison with de Haan’s parameters

The principal difference between the method used in the present study for the comparison of evidentiality systems and the method used by de Haan (2005b, 2005c) is that I apply the classical feature analysis which uses the same values (in this case +, (+) and –) for all parameters, whereas de Haan applies a model in which each parameter is defined in terms of specific values and thus the values and their number differ across parameters. De Haan applies two broad parameters, the first of which is termed ‘semantic distinctions of evidentiality’ (De Haan 2005b), and the second ‘coding of evidentiality’ (De Haan 2005c). The values of the parameter of semantic distinctions of evidentiality attested in his language sample are:

- 1) No grammatical evidentials,
- 2) Indirect evidentiality only,
- 3) Direct and indirect evidentiality.

The values of the coding of evidentiality attested in his sample are:

- 1) No grammatical evidentials,
- 2) Verbal affix or clitic,
- 3) Part of the tense system,
- 4) Separate particle,

¹⁹ The archaic synthetic forms of the Bulgarian conditional mood may also have non-firsthand marking (see e.g. Gerdzikov 1984: 246), but owing the fact that they are obsolete by now, Bulgarian is not given a positive value for this parameter.

- 5) Modal morpheme,
- 6) Mixed.

This method presupposes that the values of a parameter are mutually exclusive. For the coding of evidentiality this condition is satisfied: a language may not have grammatical evidentiality coding, or it may code only indirect evidentiality, or it may code both direct and indirect evidentiality, but cannot have two of these possibilities at the same time. In the case of the semantic distinctions, however, the choice of values is not so satisfactory. In the description of values, de Haan does not explain for example what the relationship between the values ‘verbal affix or clitic’ and ‘modal morpheme’ is. Although it is apparent from the numbers in the relevant table (see De Haan 2005c) that he has kept their instantiations apart they seem to partially overlap. Therefore, when applying de Haan’s values to new data one may easily end up counting a single element twice, once as a verbal affix or clitic and the second time as a modal morpheme.

In general, the comparison of the parameters (and their values) in de Haan’s study with my parameters reveals the pros and cons of large sample typology as opposed to small sample typology. As can be seen from the above list, de Haan defines values with notions that are more abstract than the notions used in the definitions of my parameters. This is due to the fact that his notions were targeted to a much larger number of cases. Furthermore, de Haan’s view of evidentiality seems to be less restrictive than Aikhenvald’s view (Aikhenvald 2004: 1) as he subsumes various modal verbs (such as Dutch *moeten*; see De Haan 2005c) under the notion of (grammatical) evidentiality. This, in turn, leads to the inclusion of several Western European languages in the class of languages with grammatical evidentiality coding.

Applying a large number of narrowly defined parameters and a more restrictive view of evidentiality in this study, allows me, on the other hand, to be more specific about the particular morphosyntactic strategies and semantic distinctions of evidentiality. For example, while de Haan merely declares a relationship between evidentiality coding and tense system, my parameters allow identifying the specific time reference of the markers engaged in evidentiality coding. With respect to the semantic distinctions, de Haan is not explicit about the number of evidential terms in a language. He uses the general terms of ‘direct’ and ‘indirect’ evidentiality, whereas my parameters capture finer evidential distinctions. Furthermore, he does not pay any attention to the conventionalized uses of evidentials in discourse, or to their syntagmatic properties. Finally, although his values are meant to be all-inclusive, some characteristics of evidentiality coding can not be accounted for by any of his values. Such is for example the relationship of evidentiality with verbal categories other than tense, like voice for example (see parameter 7 in 4.1.1.).

De Haan’s contribution (2005b, 2005c) conforms to the purposes of the World Atlas of Language Structures in demonstrating the world-wide distribution of the coding and semantic types of evidentiality. As he takes some

418 languages into consideration, it is inevitable that his comparison is built on only the most salient properties of evidentiality. Less salient properties are usually not mentioned in the grammars, and considering them would have led to the exclusion not only of languages, but also of entire linguistic areas and genetic families from his sample.

4.2. The relationship between evidentiality and epistemic modality

Section 2.1. introduced the two opposite standpoints concerning the relationship between evidentiality and epistemic modality and the status of evidentiality as a functional category in general. The last article (IV) of this dissertation tests these standpoints against empirical data. The stipulations are drawn from an analysis of Bulgarian and Estonian sentences with two co-occurring sentential modifiers – one grammatical (evidential) and one lexical (epistemic and/or evidential). Of particular interest are cases in which the grammatical and the lexical item are combined in order to reinforce a common meaning component.

Since the early 1980s, many studies have attempted to order different evidential meanings on a scale according to the strength of evidence. As noted by Boye (2006: 112–113, cf. also Willet 1988: 85), this scale corresponds to a scale of reliability. The results of such studies are usually formulated in terms of preference hierarchies in which each evidential meaning is equated with certain amount of reliability; cf. the following hierarchies (all of them modified in order to make their comparison more transparent).

Table 1. Types of evidence along the reliability scale.

	more reliable/certain	>>>>>	less reliable/certain
Barnes (1984: 262–264) ²⁰	'visual' – 'nonvisual' – 'secondhand' – 'apparent' – 'assumed'		
Oswalt (1986: 43)	'performative' – 'factual-visual' – 'auditory' – 'inferential' – 'quotative'		
De Haan (1999: 88)	'visual' – 'auditory' – 'non-visual' – 'inference' – 'quotative'		

Boye (2006: 109) integrates the notion of reliability and the different types of evidence in a coherent and complex semantic domain labelled 'epistemic meaning domain'. The postulation of such a domain is evoked by a broad cross-linguistic comparison. Article IV presents empirical data providing further evidence for the conceptual affinity between evidentiality (type of evidence)

²⁰ Quoted in Aikhenvald (2004: 307).

and certainty (epistemic modality). In this article, items are arranged along an epistemic scale which is labelled ‘scale of certainty’. The notion of certainty applied there has to be briefly explained as it is a potential source of confusion: if an item is positioned low on the scale of certainty, it conveys that it is very uncertain (or doubtful) that the propositional facts obtain, and often implies that they are improbable. In other words, it implies relatively strong certainty that the propositional facts do not obtain. It should be stressed, however, that the given scale concerns the strength of assertion of the propositional content, and not the strength of its negation.

The data presented in Article IV consists of attested examples and their surrounding context. In addition to ‘certainty’ and ‘type of evidence’ I also introduce a third variable whose values can also be structured along a strength scale and on which the other two variables depend. This variable, named ‘referent identity’, concerns the degree to which the original referent of the information is retrievable from the context (e.g. ‘an individual’, ‘unspecified individual within a limited group’, ‘generally unspecified individual’).

4.3. The historical development of evidentiality markers

As any other grammatical category, evidentiality does not evolve by chance from a random array of linguistic elements. Certain grammatical or lexical elements are much more likely to develop evidential connotations than others. Aikhenvald (2004: 271–284) distinguishes the following major sources of evidential markers:

- 1) grammaticalized verbs:
 - a) verbs of speech: e.g. the West Greenlandic reportive evidentiality marker *-unni* derives from the verb stem *unnir-* ‘say (that)’;
 - b) verbs of perception: e.g. the Wintu non-visual sensory marker *nt^hEr* derives from a passive form based on the verb ‘hear’ and is followed by an inferential evidential marker;
- 2) deictic and locative markers: e.g. the Sissala hearsay and reported marker *ré* (or *é*) stems from the locative demonstrative *ré* ‘here’, ‘this’;
- 3) evidential extensions of non-evidential verbal categories:
 - a) modalities: e.g. the Hill Patwin marker of indirect evidentiality *-boti/-beti* derives from a combination of the auxiliary *bo/be* ‘be (locational)’ with the definite future suffix;
 - b) perfective aspect, past tenses, and other forms with resultative meaning: e.g. Cree-Montagnais-Naskapi non-firsthand marker *-shapan* derives from a Proto-Algonquian perfect;
 - c) participles or other nominalizations: see example (7) from Lithuanian and example (8) from Estonian;

- 4) speech complements (de-subordination of an erstwhile subordinate clause): the only well-described cases come from Estonian; example in Campbell (1991);
 - 5) copula constructions: e.g. Chinese Pidgin Russian non-firsthand term marked by a construction containing verb root and the infinitive of the copula verb *est*'.
- Finally, evidentials can be borrowed across linguistic communities (see Aikhenvald 2004: 288–299).

These paths are not mutually exclusive. The Balkan languages (Turkish, Bulgarian, Macedonian and Albanian) for example, are generally believed to instantiate path 3b (see e.g. Johanson (2003: 287–288) for Turkish, Lindstedt (2000a: 374–378) for Bulgarian and Macedonian, and Friedman (2003: 209–210) for Albanian). Article I introduces the major points on the cline from compound perfective past to evidentiality. On a more atomistic account, however, the situation in Balkan Slavic can be also accounted for in terms of type 3c. There are many other cases in which the sources listed above do not seem to represent different sources but different perspectives to a single source. In Estonian for instance, evidentiality is most commonly marked with ‘participles or other nominalizations’ (Path 3c) some of which are likely to have evolved through de-subordination of speech complements (Path 4; see e.g. Campbell 1991) others from evidential extensions of perfect tense (Path 3b; see e.g. Ikola 1953: 51–58; Künnap 1994). A similar impression of multiple paths arises for Latvian and Lithuanian (compare Aikhenvald (2004: 281) with Endzelīns (1951: 976) and Holvoet (2007: 92–96), for example).

In the remaining part of this section, I will discuss the historical development of the coding of evidentiality in Estonian, which, unlike its Balkan equivalents, is still a controversial topic in need of further investigation. The historical reconstruction of the Estonian evidentiality marking based on participles poses problems which can be expected in the reconstruction of any other evidentiality system employing multiple markers integrated in a single paradigm with rather specific meaning (such as reported). The following discussion is based on a study by Kehayov and Siegl (2007), which in addition to Estonian also discussed the situation in Eastern Finno-Ugric (Permic and Volgaic) and Baltic. I will start the exposition with the less controversial topic, namely, the question about the development of the Estonian present participle into an evidentiality marker.

In Standard Estonian, the so-called *vat*-form encoding reported evidentiality consists of the partitive case of the present participle. Certain dialects also use the genitive or nominative case forms of the present participle in this function (see Kask 1984: 254–261). It is commonly agreed that the reportive meaning of the present participle is due to ellipsis of the superordinate verb of a speech-act or mental state in participial subordinating construction. This idea was first formulated by Wiedemann (1875: 651–652) and has survived with minor modifications until today (cf. Kettunen 1924: 20; Saareste 1932: 18; Airila 1935: 48; Grünthal 1941: 259–260; Ikola 1953: 41–43; and Hakulinen and

Leino 1987). Synthesizing the ideas of his predecessors, Campbell (1991) presented an explicit scenario of the process of de-subordination. The bridging context for the reanalysis to occur is characterized by a contamination of a subordinating construction with the complementizer *et* and the participial subordinating construction; see (24) (Campbell 1991: 287):

- (24) a) *sai kuulda, et seal üks mees*
 got hear-INF that there one.NOM man.NOM
elab
 live-3PRES.INDICATIVE
 ‘She came to hear / she heard that a man lives there.’
- b) *sai kuulda seal ühe mehe elavat*
 got hear-INF there one.GEN man.GEN live-PRES.PARTCP
 ‘He came to hear / he heard of a man’s living there.’
- c) *sai kuulda, (et) seal üks mees*
 got hear-INF that there one.NOM man.NOM
elavat
 live-MODUS.OBLIQUUS
 ‘He came to hear / he heard that (they say) a man lives there.’
- d) *ta tegevat tööd*
 he.NOM do-PRES.INDIR work-PARTV
 ‘They say he is working.’

The contamination (c) of the complement clause (a) and the participial construction (b) resulted in the extension of the subordinated verb form (the partitive form of the present participle) into the main clause (d).

In contrast to the question about the development of the present participle to a marker of reported evidentiality, the question about the development of the past participle to that function has been a subject for long-lasting debate among linguists. In Kehayov and Siegl (2007), we distinguish between three main hypotheses. According to the first hypothesis, the past participle acquired reportive meaning following a grammaticalization path similar to the path presented above for the present participle (see for example Wiedemann 1875: 651–652; Saareste 1932: 18; Campbell 1991; Alvre 1993; Künnap 1994). Künnap (1994: 24) extended the model presented by Campbell to explain the reanalysis of the past participle; cf. (25):

- (25) a) *sai kuulda, et seal üks mees*
 got hear-INF that there one.NOM man.NOM
elas
 live-3IMPRF.INDICATIVE
 ‘She came to hear / she heard that a man lived there.’
- b) *sai kuulda seal ühe mehe elanud*
 got hear-INF there one.GEN man.GEN live-PST.PARTCP
(olevat)
 be-PST.PARTCP
 ‘He came to hear / he heard of a man’s who lived there.’
- c) *sai kuulda, (et) seal üks mees*
 got hear-INF that there one.NOM man.NOM
elanud
 live-MODUS.OBLIQUUS
 ‘He came to hear / he heard that (they say) a man lived there.’
- d) *seal elanud üks mees*
 there live-MODUS.OBLIQUUS one.NOM man.NOM
 ‘A man lived there.’

According to the second hypothesis, the evidentiality coding with a past participle originates from a compound past tense with omitted auxiliary. Here one can distinguish between two slightly different viewpoints. According to the first, the omitted element was the *vat*-form or the infinitive (which is also used as a marker of evidentiality: see Parameter 3, example (8)) of the auxiliary. This scenario is proposed by Kask (1984: 243). According to the second viewpoint, the omitted element is a finite form of the auxiliary and the source construction is a regular compound tense form of the indicative (see Ikola 1953: 51–58; Künnap 1992, 1994; Metslang 1994). These two scenarios are presented in (26a–b):

- (26) a) *ta olevat / olla tulnud* → *ta Ø tulnud.*
 s/he be.vat / be.INF come.PST.PTCP s/he Ø come.PST.PTCP
 ‘Reportedly, she has come.’ ‘Reportedly, she came.’
- b) *ta on / oli tulnud* → *ta Ø tulnud*
 s/he be.3SG / be.PST.3SG come.PST.PTCP s/he Ø come.PST.PTCP
 ‘She has / had come.’ ‘Reportedly, she came.’

According to the third hypothesis, the evidential application of the past participle derives from its use as an unmarked form of the predicate in ancient monologic narratives. The idea was first suggested by Grünthal (1941: 261–262), who referred to similar developments in some other Uralic languages. Grünthal’s idea was rejected in subsequent studies (e.g. Ikola 1953: 58) and recently revived by Muižniece, Metslang and Pajusalu (1999). This hypothesis is based on the assumption that the use of the past participle as a predicate is an ancient feature both in Finno-Ugric and Indo-European (see Muižniece, Metslang and Pajusalu 1999: 137). The aim of Kehayov and Siegl (2007) was to provide a critical assessment of this hypothesis.

The first two hypotheses state that the evidential use of the participle evolved from a certain type of construction (subordinating participial construction or compound tense forms), whereas the third sees the source of such uses in the ancient predicative function of the participle. In other words, the first two hypotheses build on the assumption that the past participle was extracted from a construction which as a whole expressed reported evidentiality. In contrast, the third hypothesis states that the participle was initially the unmarked form of the predicate, and that its sphere of use became gradually narrower, until it became eventually restricted in modern standard Estonian to the reportative and a few other peripheral meaning domains.

In what follows, I will briefly present internal evidence from Estonian for or against each of these hypotheses.

The strongest evidence in favour of the first hypothesis comes from the observation that the past participle in utterances like *ta tulnud* (cf. example (26b)) has reported and not inferential or other evidential meaning. If the evidential construction with the past participle is derived from a compound tense form (such as present or past perfect), one should have expected it to develop an inferential or a general non-firsthand meaning, which are cross-linguistically common meaning extensions of perfect tense (see Bybee, Perkins and Pagliuca 1994: 54, 97; Lindstedt 2000a). At the same time, the development of reported evidentiality conforms perfectly to the scenario involving the de-subordination of the speech-complement encoded by the past participle.

The second hypothesis (and its second scenario in particular) is endorsed by two pieces of linguistic evidence, both of which concern the distribution of verb phrases in connected narratives. The first concerns the negative forms of the past participle expressing reported evidentiality. One would expect that such negative forms would be formed following the regular pattern of verbal negation, i.e. preposing the non-finite negative verb *ei* to the past participle. Using the example with the verb ‘to come’, we would expect that negative form of the participle *tulnud* (s/he come-PST.PTCP) when used as an evidential predicate will be *ei tulnud* (NEG come-PST.PTCP). In Kehayov and Siegl (2007), we looked at Estonian folk narratives to find out whether this is the common negation pattern. Although this pattern was attested, we found many

cases in which the evidential past participle is negated with the pattern *ei ole tulnud* (NEG be come-PST.PTCP). This pattern coincides with the negative form of the (compound) perfect tense. Consider example (27):

- (27) *ühe koõra olnd jälle mîs ja naõne*
 one.GEN time.GEN be-PST.PTCP again man and woman
abi,elupâr. olnd neil üks veikke
 marriage_couple be-PST.PTCP they-ADE one small
tuba veì saun mâ külle sisse
 room or sauna hill.GEN side-GEN in-ILL
tehtud, tagumene külG olnd mâ
 do-PST.PASS.PTCP back side be-PST.PTCP hill-GEN
külle sîs, îs külles ol'ti kaitus piäl.
 side.GEN then front side-INE be-IMPF.3SG roof above.ADE
sis olnd naõtukke peldu kâ, õbest
 then be-PST.PTCP little field-PART also horse-PART
ei ole olnd, lehm olnd ja siga. ...

NEG be be-PST.PTCP cow **be-PST.PTCP** and pig

“Once there was (= was said to be) a married couple. They had (= were said to have) a small room or cottage built onto a hill, the back was (= was said to be) built onto the hill and there was a roof on the front side. Then, there **was (= was said to be)** also a little piece of land, there **was no (PRF)** horse, there **was (= was said to be)** a cow and a pig. ... (Must 1965: 374–375, quoted in Kehayov and Siegl 2007: 105–106)

Examples like (27) show that there is a close relationship between the evidential predicate in the form of past participle and the forms of the perfect tense. Unfortunately, we could not go through all folklore texts systematically, and therefore we refrained from speculating as to which of the alternative patterns (NEG V-PST.PTCP or NEG be V-PST.PTCP) is the more frequent negative counterpart of the evidential past participle. This is a topic for a corpus-based study.

The second piece of evidence is provided by contexts in which past participles are isolated from the forms of pluperfect and start to function as main predicates. Consider (28):

- (28) *t'sōru meizahn oļ ka heēr oļnu? siš*
 tsooru manor-INE be-IMPF.3SG also master be-PST.PTCP then
tū oļ meļsavahteḡa^p pahanḡanu? eḡ: "mis
 he be-IMPF.3SG ranger-PL-COM accuse-PST.PTCP that what
ti^l lašt varasta?!" siš oļ
 2PL let-PRS.2PL steal-INF then be-IMPF.3SG
paḡḡnu^h hobēze eḡte, oļ eži^k ka vaht^{ma}
 put-PST.PTCP horse.GEN before be-IMPF.3SG self also guard-INF
lāḡḡnū?. eḡ meļsavahi? eḡ jaḡiⁿ noḡḡ meļsavaḡḡiḡ,
 go-PST.PTCP that ranger-PL NEG hunt those wood_thief-PL-PART
eḡ ma^{lā} eži^v vaht^{ma}. siš nū?
 that I go-PRS.1SG self guard-INF then they
oļti^l tuḡḡḡmaḡa^m mehe? oļnu?
be-IMPF-3PL unknown man-PL be-PST.PTCP
nū^k kraḡanu^k kiḡḡni?. uma nōri
 they **seize-PST.PTCP off** own roop-GEN
veḡḡnu^r rḡ^l manḡ ja paḡḡnu^p paruni
 take-PST.PTCP sledge.GEN from and put-PST.PTCP baron.GEN
sāni peḡrā. taļvel oļnu? ...
 sledge.GEN behind winter-ADE be-PST.PTCP
 “There was a baron at Tsooru manor. He was angry with the rangers:
 “Why do you let them steal!” Then he hitched up the horse and had
 gone to guard himself, being angry with the rangers for not hunting
 the wood thieves. There **had been (PLPRF)** strangers, who **had**
seized him (= were said to have seized him). They had taken (= were
 said to have taken) a rope from the sledge and fastened (= were said to
 have fastened) the baron behind the sledge. It was (= was said to have
 been) winter. ... (Keem – Kāsi 2002: 334, quoted in Kehayov and
 Siegl 2007: 106–107)

Here the first part of the story is narrated using forms of pluperfect. At a certain moment, however, the finite form of the auxiliary verb becomes redundant in its function to mark specific time reference in discourse and the past participle is rendered as an independent predicate. While the bold and underlined verb form in the example is perceived as an elliptic form, from the next sentence the past participle becomes the regular form of the predicate. The fact that there is a noun phrase between the auxiliary and the participle in the foregoing bold form of pluperfect has apparently played a role in the mechanism which led to the isolation of the participle. Following the principles of iconicity, we would expect that in cases in which the auxiliary and the participle are separated by other constituents a switch of forms is more likely to occur. This kind of switch between compound tense and bare past participle is frequently attested, not only

in traditional narratives, but also in contemporary fiction. Consider example (29):

- (29) *Ta jutustas Lavranile oma isast, kes
 she tell-IMPF.3SG Lavran-ALL own father-ELA who
 olevat tüüpiline New Yorki juut. Kord oli
 be-PRS.REP typical New York-GEN Jew Once be-IMPF.3SG
 üks naine, kelle pool ta Tartus
 one woman who-GEN side she Tartu-INE
 oli elanud, hakanud temaga rääkima
 be-IMPF.3SG live-PST.PTCP start-PST.PTCP she-COM talk-INF
 marurikastest Ameerika juutidest. Naine
 madly_rich-PL-ELA America-GEN Jew-PL-ELA woman
 teinud seda mõnevõrra halvustavalt, ning
 do-PST.PTCP it to_some_degree contemptuously and
 Carol polnud öelnud oma võõrustajale,
 Carol NEG.be-PST.PTCP say-PST.PTCP her host-ALL
 et temagi on juut, kelle isa on ajaloolane ja
 that she-ENCL is Jew who-GEN father is historian and
 on New Yorgi Ülikoolis ajalugu õpetanud.
 is New York-GEN University-INE history-PART teach-PST.PTCP
 ‘She told Lavran about her father who was said to be a typical New
 York Jew. There was a woman once, at whose place she was living in
 Tartu, who **had started** to tell her about all those madly rich
 American Jews. The woman **had done** (= **was said to have done**) this
 with some contempt in her voice, so Carol had not told her that she
 was Jew herself and that her father was a historian who had been
 teaching history at the University of New York.’ (Pilter: 28, quoted in
 Kehayov and Siegl 2007: 77, 107–108)*

Here again the finite form of the auxiliary (the first bold form) is used to link the event to a specific moment of reference. The reason for the form switch to occur exactly in this place is that we do not only find the noun phrase *üks naine* ‘a woman’, but also its extension – a long relative clause, situated between the auxiliary and the participle, which together constitute the form of pluperfect. In the next sentence, the past participle alone functions as a predicate conveying reported evidentiality (see the bold and underlined form). This mechanism of omission of the current relevance expression (the finite auxiliary) and conventionalization of the remaining expression (the past participle) as a main plot-advancing device has already been acknowledged in the relevant literature. Koptjevskaja-Tamm and Wälchli (2001: 723) for example, point out the tendency for using perfect forms with an overt auxiliary at the beginning of tales. A possible explanation of such uses is that the overt coding of the current

relevance is always necessary there in order to link the following chain of events to the specific moment of reference. The scope of the current relevance expression extends over the first predication unit of the tale. As a result, the finite auxiliary is perceived as redundant in the following sentences.

The third hypothesis is novel not only in the literature on Finno-Ugric, but also in general. Unfortunately however, the existing empirical evidence does not seem to support it. All the evidence brought in this section defies it; for additional counter-evidence see Kehayov and Siegl (2007).

5. SOURCES REFERRED TO IN THE INTRODUCTORY PART (Sections 1–4)

- Aikhenvald, Alexandra Y. 2003: Evidentiality in typological perspective. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality. Typological Studies in Language 54*. Amsterdam – Philadelphia: John Benjamins, 1–31.
- Aikhenvald, Alexandra Y. 2004: *Evidentiality*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y., Robert M. W. Dixon (eds) 2003: *Studies in Evidentiality. Typological Studies in Language 54*. Amsterdam – Philadelphia: John Benjamins.
- Airila, Martti 1935: Viron kielen *oratio obliqua*'sta. – *Eripainos Suomi 5*, 17-sta. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Aksu-Koç, Ayhan A., Dan I. Slobin 1986: A Psychological Account of the Development and Use of Evidentials in Turkish. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 159–167.
- Alvre, Paul 1993: Anmerkungen zum obliquen Modus im Estnischen und Livischen. – *Linguistica Uralica 29*, 2, 95–104.
- Asenova = Асенова, П. 2002: Балканско езикознание. Основни проблеми на балканския езиков съюз. Велико Търново: Фабер.
- van der Auwera Johan 1998a: Introduction. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe. Empirical approaches to language typology 20*. Berlin: Mouton de Gruyter, 1–23.
- van der Auwera 1998b: Phasal adverbials in the languages of Europe. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe. Empirical approaches to language typology 20*. Berlin: Mouton de Gruyter, 25–145.
- van der Auwera Johan 1998c: Conclusion. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe. Empirical approaches to language typology 20*. Berlin: Mouton de Gruyter, 813–836.
- van der Auwera, Johan 1998d: Revisiting the Balkan and Meso-American Linguistic Areas. – *Language Sciences 20*, 3, 259–270.
- Barnes, Janet 1984: Evidentials in the Tuyuca verb. – *International Journal of American Linguistics 50*, 3, 255–271.
- Boye, Kasper 2006: *Epistemic meaning: a cross-linguistic study*. Ph.D. dissertation. Department of English, Germanic and Romance Studies, University of Copenhagen.
- Buchholz, Oda, Wilfried Fiedler 1987: *Albanische Grammatik*. Leipzig: WEB Verlag Enzyklopädie Leipzig.
- Bybee, Joan L., Revere Perkins and William Pagliuca 1994: *The evolution of grammar – tense, aspect, and modality in the languages of the world*. Chicago – London: University of Chicago Press.
- Campbell, Lyle 1991: Some Grammaticalization Changes in Estonian and their Implications. – E. C. Traugott, B. Heine (eds), *Approaches to Grammaticalization. Typological Studies in Language 19*. Amsterdam: Benjamins. 285–299.
- Chafe, Wallace, Johanna Nichols (eds) 1986: *Evidentiality: the Linguistic Coding of Epistemology. Advances in Discourse Processes 20*. Norwood, New Jersey: Ablex.
- Chirikba, Vjacheslav 2003: Evidential category and evidential strategy in Abkhaz. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality*, 243–272.

- Comrie, Bernard, Maria Polinsky 2007: Evidentials in Tsez. – Z. Guentchéva, J. Landaburu (eds), *L'énonciation médiatisée II. Le traitement épistémologique de l'information: illustrations amérindiennes et caucasiennes*, 335–350.
- Croft, William 2003: *Typology and Universals*. (Second edition) Cambridge: Cambridge University Press.
- Dahl, Östen 1985: *Tense and Aspect Systems*. Oxford: Blackwell.
- Dahl, Östen, Maria Koptjevskaja-Tamm 2001: The Circum-Baltic Languages. Introduction to this volume. – Ö. Dahl, M. Koptjevskaja-Tamm (eds), *Circum-Baltic Languages*. Vol. 1. Past and Present. Amsterdam – Philadelphia: John Benjamins, XV–xx.
- DeLancey, Scott 2001: The mirative and evidentiality. – *Journal of Pragmatics* 33, 3, 369–382.
- Demina = Демина, Е. И. 1959: Пересказывательные формы в современном болгарском литературном языке. – С. Б. Бернштейн, (ред.), *Вопросы грамматики болгарского литературного языка*, Москва: Издательство Академии наук СССР, 313–378.
- Dendale, Patrick, Liliane Tasmowski 2001: Introduction: Evidentiality and related notions. – *Journal of Pragmatics* 33, 3, 339–348.
- Dobrushina, Nina, Sergei G. Tatevosov 1996: Usage of verbal forms. – A. E. Kibrik, S. G. Tatevosov, A. Eulenberg (eds), *Godoberi*. München: Lincom Europa, 91–106.
- Duchet, Jean-Louis, Remzi Pernäska 1996: L'admiratif albanais: recherche d'un invariant sémantique. – Z. Guentchéva (éd.), *L'énonciation médiatisée*. Louvain – Paris: Peeters, 31–46.
- Eintrei = Эйнтрей, Т. И. 1982: *Албанский язык*. Ленинград: Издательство Ленинградского университета.
- Endzelin, Janis 1922: *Lettische Grammatik*. Riga: A. Gulbis.
- Endzelīns, Jānis 1951: *Latviešu valodas gramatika*. Rīga: Latvijas Valsts izdevniecība.
- Erelt, Mati 2002: Does Estonian Have the Jussive? – *Linguistica Uralica* 38, 2, 110–117.
- Erelt, Mati (ed) 2003: *Estonian Language*. *Linguistica Uralica*. Supplementary Series 1. Tallinn: Estonian Academy Publishers.
- Erelt, Mati, Helle Metslang, Karl Pajusalu 2006: Tense and Evidentiality in Estonian. – B. Cornillie, N. Delbecque (eds), *Topics in Subjectification and Modalization*. *Belgian Journal of Linguistics* 20, 125–136.
- Fiedler, Wilfried 1966: Zu einigen Problemen des albanischen Admirativs. – *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig* 15, 561–566.
- Friedman, Victor A. 1978: On the Semantic and Morphological Influence of Turkish on Balkan Slavic. – *Papers from the Fourteenth Regional Meeting*. Chicago Linguistic Society, April 13–14, 1978.
- Friedman, Victor A. 1986: Evidentiality in the Balkans. Bulgarian, Macedonian and Albanian. – W. Chafe, J. Nichols (eds). *Evidentiality: the Linguistic Coding of Epistemology*, 168–187.
- Friedman, Victor A. 1988b: Morphological Innovation and Semantic Shift in Macedonian. – *Zeitschrift für Balkanologie* 24, 1, 34–41.
- Friedman, Victor A. 1994: Surprise! Surprise! Arumanian Has Had an Admirative! – *Indiana Slavic Studies* 7, 1994, 79–89.

- Friedman, Victor A. 1997: A Contrastive View of Subordinate Aspect and the Opposition Confirmative/Non-confirmative in the Balkans with Particular Reference to Macedonian and Turkish. – *Balkanistica* 10. Columbus, Ohio, 185–201.
- Friedman, Victor A. 2003: Evidentiality in the Balkans with special attention to Macedonian and Albanian. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality. Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 189–218.
- Friedman, Victor 2005: Admirativity: between modality and evidentiality. – *Sprachtypologie und Universalienforschung* 58, 1, 26–35.
- ГБЕ II = Тилков, Д., Стоянов, С., Попов К. (ред.) 1983: Граматика на съвременния български книжовен език. Том 2, Морфология. София: Издателство на Българска Академия на Науките, 1983.
- Gerdžikov = Герджиков, Г. 1984: Преизказването на глаголното действие в българския език. София: Наука и изкуство.
- GLJ = Амбразас, В., Валецкене, А., Валюлите, Э. 1985: Грамматика литовского языка. Академия Наук Литовской ССР. Вильнюс: Мокслас.
- Gronemeyer, Claire 1997: Evidentiality in Lithuanian. – *Lund University, Department of Linguistics Working Papers* 46, 93–112.
- Grünthal, Willem 1941: Itämerensuomalaisten kielten yksikön nominatiivi objektin edustajana aktiivin yhteydessä. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Guentchéva, Zlatka (éd.) 1996a: L'énonciation médiatisée. Louvain – Paris: Peeters.
- Guentchéva, Zlatka 1996b: Le médiatif en Bulgare. – Z. Guentchéva (ed), *L'énonciation médiatisée*, 47–70.
- Guentchéva Zlatka, Jon Landaburu (eds) 2007: L'énonciation médiatisée II. Le traitement épistémologique de l'information: illustrations amérindiennes et caucasiennes. Louvain – Paris – Dudley, MA: Peeters (Bibliothèque de l'Information Grammaticale 63).
- De Haan, Ferdinand. 1999: Evidentiality and epistemic modality: Setting boundaries. – *Southwest Journal of Linguistics* 18, 1, 83–101.
- De Haan, Ferdinand 2005a: Encoding speaker perspective: Evidentials. – Z. Frajzyngier, A. Hodges, D. S. Rood (eds), *Linguistic Diversity and Language Theories*. Amsterdam: John Benjamins, 379–397.
- De Haan, Ferdinand 2005b: Semantic Distinctions of Evidentiality. – M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 314–317.
- De Haan, Ferdinand 2005c: Coding of Evidentiality. – M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 318–321.
- Haarmann, Harald 1970: Die indirekte Erlebnisform als grammatische Kategorie. Eine eurasische Isoglosse. Wiesbaden: Otto Harrassowitz.
- Haarmann, Harald 1976: Arealtypologie der finnisch-ugrischen Sprachen – Prinzipien einer interdisziplinär orientierten Finnougristik. – *Finnisch-Ugrische Mitteilungen* 1, 3–14.
- Hakulinen, Auli, Pentti Leino 1987: Finnish Participial Construction from a Discourse Point of View. – *Ural-Altäische Jahrbücher* 59, 35–43.

- Haspelmath, Martin 2004: Does linguistic explanation presuppose linguistic description? – *Studies in Language* 28, 3, Special issue on linguistic evidence, (guest eds. M. Penke, A. Rosenbach), 554–579.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, Bernard Comrie (eds) 2005: *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Helimski, Eugen 2003: Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. – *Nyelvtudományi Közlemények* 100, 156–167.
- Holvoet, Axel 2001a: On the Paradigm of the Oblique Mood in Lithuanian and Latvian. – *Linguistica Baltica* 9, 69–86.
- Holvoet, Axel 2001b: Mood and Modality in Latvian. – *Sprachtypologie und Universalienforschung* 54, 3, 226–252.
- Holvoet, Axel 2001c: *Studies in the Latvian Verb*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Holvoet, Axel 2007: *Mood and Modality in Baltic*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Ikola, Osmo 1953: Viron ja liivin modus obliquuksen historiaa. – *Suomi* 106, 4, Helsinki: Suomalaisen Kirjallisuuden Seura.
- Izvorski, Roumyana 1997: The Present Perfect as an Epistemic Modal. – A. Lawson, E. Cho (eds), *Proceedings of Semantics and Linguistic Theory (SALT) VII*. Ithaca, New York: CLS Publications, 222–239.
- Jakobson, Roman 1971: Shifters, Verbal Categories, and the Russian Verb. – R. Jakobson (ed), *Selected writings. Vol. 2, Word and Language*. The Hague: Mouton, 131–147.
- Johanson, Lars 1998: Zum Kontakteinfluß türkischer Indirektive. – N. Demir, E. Taube (eds), *Turkologie heute – Tradition und Perspektive. Materialien der dritten Deutschen Turkologen-Konferenz, Leipzig, 4.–7. Oktober 1994*, Wiesbaden: Otto Harrassowitz, 141–150.
- Johanson, Lars 2003: Evidentiality in Turkic. – *Studies in Evidentiality*. A. Y. Aikhenvald, R. M. W. Dixon (eds), *Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 272–290.
- Kask, Arnold 1984: *Eesti murded ja kirjakeel*. Tallinn: Valgus.
- Kehayov, Petar, Florian Siegl 2007: The evidential past participle in Estonian reconsidered. – *Études finno-ougriennes* 38, Paris, 75–117.
- Kettunen, Lauri 1924: *Lauseliikmed eesti keeles*. Tartu: Eesti kirjanduse selts.
- Kiefer, Ferenc 1994: Modality. – R. E. Asher (ed), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 2515–2520.
- Klaas, Birute 1997: The Quotative Mood in the Baltic Sea Areal. – M. Erelt (ed), *Estonian: Typological Studies II*. Tartu Ülikooli Eesti Keele Õppetooli Toimetised 7, 73–97.
- Koneski = Конески, Б. 1967: *Граматика на македонскиот литературен јазик*. Скопје: Култура.
- Копонов = Кононов, А. Н. 1956: *Граматика современного турецкого литературного языка*. Москва – Ленинград: Издательство академии наук СССР.
- Koptjevskaja-Tamm, Maria, Bernhard Wälchli 2001: The Circum-Baltic languages: An areal-typological approach – Ö. Dahl, M. Koptjevskaja-Tamm (eds), *The Circum-*

- Baltic Languages. Typology and Contact. Vol. 2, Grammar and Typology. Amsterdam – Philadelphia: John Benjamins, 615–750.
- Kozintseva = Козинцева, Н. А. 1994: Категория эвиденциальности (проблемы типологического анализа). – Вопросы языкознания, 1994/3, 92–104.
- Künnap, Ago 1992: On the analytical past tense forms in the Uralic languages. – *Linguistica Uralica* 28, 3, 173–179.
- Künnap, Ago 1994: On the Uralic Indirectal. – *Ural-Altische Jahrbücher* 13, 18–27.
- Lazard, Gilbert 1999: Mirativity, evidentiality, mediativity, or other? – *Linguistic Typology* 3, 1, 91–109
- Lazard, Gilbert 2001: On the Grammaticalization of Evidentiality. – *Journal of Pragmatics* 33, 3, 359–367.
- Leinonen, Maria 2000: Evidentiality in Komi Zyryan. – L. Johanson, B. Utas (eds), *Evidentials: Turkic, Iranian and Neighbouring Languages. Empirical Approaches to Language Typology* 24. Berlin – New York: Mouton de Gruyter, 419–440.
- Lewis, Geoffrey L. 1966: *Turkish. Teach Yourself Books*. London: The English Universities Press LTD.
- Lindstedt, Jouko 2000a: The Perfect – Aspectual, Temporal and Evidential. – Ö. Dahl (ed), *Tense and Aspect in the Languages in Europe. Empirical Approaches to Language Typology* 20, 6 = EURO TYP. Berlin – New York: Mouton de Gruyter, 365–383.
- Lindstedt, Jouko 2000b: Linguistic Balkanization: Contact-Induced Change by Mutual Reinforcement. – D. G. Gilberts, J. Nerbonne, J. Schaeken (eds), *Languages in Contact. Studies in Slavic and General Linguistics* 28. Amsterdam – Atlanta, GA: Rodopi, 231–246.
- Lyons, John 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.
- Maslov = Маслов, Ю. С. 1981: *Грамматика болгарского языка*. Москва: Высшая школа.
- Mathiassen Terje 1996: Some aspects of language contacts and convergency phenomena in the Baltic Area – Evidence from Lithuanian, Latvian, Livonian, Estonian, Russian and Polish. – J. Fernandez, R. Raag (eds), *Contacts des langues et de cultures dans l'aire baltique / Contacts of languages and cultures in the Baltic area: Mélanges offerts à Fanny de Sivers*. (= Uppsala Multiethnic Papers 39). Uppsala, 171–179.
- McCready, Eric, Norry Ogata 2007: Evidentiality, modality and probability. – *Linguistics and Philosophy* 30, 2, 147–206.
- Metslang, Helle 1994: Temporal relations in the predicate and the grammatical system of Estonian and Finnish. – *Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja* 39. Oulu.
- Meydan, Métiyé 1996: Les emplois médiatifs de *-miş* en Turc. – Z. Guentchéva (éd.), *L'énonciation médiatisée*. Louvain – Paris: Peeters, 125–143.
- Muižniece, Liena, Helle Metslang, Karl Pajusalu 1999: Past Participle Finitization in Estonian and Latvian. – M. Erelt (ed), *Estonian: Typological Studies III*. Tartu Ülikooli Eesti Keele Õppetooli Toimetised 11, 128–157.
- Nichols, Johanna 1986: The bottom line: Chinese Pidgin Russian. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 239–257.
- Nichols, Johanna 2007: What, if anything, is typology? – *Linguistic Typology* 11, 1, 231–238.

- Nitsolova = Ницолова, Р. 2006: Взаимодействие эвиденциальности и адмиративности с категориями времени и лица глагола в болгарском языке. – *Вопросы языкознания*, 2006/4, 27–45.
- Nuyts, Jan 2001: Epistemic modality, language and conceptualization. A cognitive-pragmatic perspective, Amsterdam: John Benjamins.
- Oswalt, Robert L. 1986: The evidential system of Kashaya. – W. Chafe, J. Nichols (eds). *Evidentiality: the Linguistic Coding of Epistemology*, 29–45.
- Palmer, Frank, R. 2001: *Mood and Modality*. Cambridge: Cambridge University Press.
- Rudzīte, Marta 1964: *Latviešu dialektoloģija*. Rīga: Latvijas Valsts Izdevniecība.
- Saareste, Andrus 1932: *Die estnische Sprache*. Tartu: Akadeemiline Kooperatiiv.
- Schaller, Helmut W. 1975: *Die Balkansprachen. Eine Einführung in die Balkanphilologie*. Heidelberg: Carl Winter.
- Sepper, Maria-Maren 2002: Evidentsiaalsuse väljendamise eesti kirjakeeles ühe sajandi jooksul. – Väikeisi kiili kokkoputmisõq / Väikeste keelte kontaktid. Võro Instituudi Toimõndusõq 14. Võru: Võro Instituut, 185–206.
- Serebrennikov = Серебrenников, Б. А. 1960: Категории времени и вида в финно-угорских языках пермской и волжской групп. Москва: Наука.
- Siegl, Florian 2004: The 2nd past in the Permic languages – form, function, and a comparative analysis from a typological perspective. Unpublished MA thesis. University of Tartu. Chair of Finnic languages. <http://www.utlib.ee/ekollekt/diss/mag/2004/b16700053/Siegl.pdf>
- Slobin, Dan, Ayhan A. Aksu 1982: Tense, Aspect, and Modality in the Use of the Turkish Evidential. – P. J. Hopper (ed), *Tense-Aspect: Between Semantics and Pragmatics*. Amsterdam – Philadelphia: John Benjamins, 185–200.
- Stolz, Thomas 1991: Sprachbund im Baltikum? Estnisch und Lettisch im Zentrum einer sprachlichen Konvergenzlandschaft (= Bochum-Essener Beiträge zur Sprachwandelforschung 13). Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Toomet, Piret 2000: Mõnest kaudsuse väljendamise võimalusest tänapäeva eesti keeles. – *Keel ja Kirjandus* 43, 4, 251–259.
- Tournadre, Nicolas 1996: Comparaison des systèmes médiatifs de quatre dialectes tibétains. (tibétain central, ladakhi, dzongkha et amdo). – Z. Guentchéva (éd.), *L'énonciation médiatisée*. Louvain – Paris: Peeters, 195–213.
- Wälchli, Bernhard 2000: Infinite predication as a marker of evidentiality and modality in the languages of the Baltic region. – *Sprachtypologie und Universalienforschung* 53, 2, 186–210.
- Whaley, Lindsay, J. 1997: *Introduction to typology: the unity and diversity of language*. Thousand Oaks – London – New Delhi: Sage Publications.
- Wiedemann, Ferdinand J. 1875: *Grammatik der Ehstnischen Sprache, zunächst wie sie in Mittelehstland gesprochen wird, mit Berücksichtigung der anderen Dialekte*. St. Pétersbourg.
- Willet, Thomas 1988: A Cross-Linguistic Survey of the Grammaticalization of Evidentiality. – *Studies in Language* 12, 1, 51–97.

ARTICLES

Kehayov, Petar 2002:
Typology of Grammaticalized Evidentiality in Bulgarian and Estonian. –
Linguistica Uralica 38, 2, 126–144.

Kehayov, Petar 2004:
Eesti keele evidentsiaalsussüsteem mõne teise keele taustal.
Morfosüntaks ja distributsioon. – Keel ja Kirjandus 47, 11, 812–829.

Kehayov, Petar 2004:
Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Semantika. –
Keel ja Kirjandus 47, 12, 895–914.

Kehayov, Petar 2008 (in print):
Interactions between grammatical evidentials and lexical markers of epistemicity
and evidentiality: a case-study of Bulgarian and Estonian. – V. A. Plungian, B.
Wiemer (eds), Lexikalische Evidenzialitäts-Marker in slavischen Sprachen. Wiener
Slawistischer Almanach, Sonderband 72, Wien.

Petar Kehayov

**INTERACTIONS BETWEEN GRAMMATICAL EVIDENTIALS
AND LEXICAL MARKERS OF EPISTEMICITY AND
EVIDENTIALITY: A CASE STUDY OF BULGARIAN
AND ESTONIAN¹**

1. Introduction

This study is concerned with the interface between evidentiality and epistemic modality in two genetically unrelated languages. Bulgarian is a strongly 'Balkanized' South Slavic language with grammatical marking of evidentiality, which is largely accepted to be a non-Slavic feature in its grammar. Estonian, which like Bulgarian has grammaticalized evidentiality as a coherent category, is a member of the Finnic branch of the Finno-Ugric language family and as such is not related to the Slavic languages. What these languages have in common, however, is that they are spoken close to the external borders of the Slavic linguistic area (where Bulgarian lies within this area and Estonian just outside it) and that the historical sources of their grammatical evidentiality systems have been generally assumed to be outside of rather than inside of what could be considered a common Slavic grammaticon.

By also taking a Finno-Ugric language into consideration the present article goes slightly beyond the restrictions set by the major aims of this volume. Moreover, it also transgresses in 1) taking into consideration not only the lexical but also the grammatical coding of evidentiality, and 2) taking into consideration not only the lexical coding of evidentiality, but also, and even more so, the lexical coding of epistemicity. The reason for adopting such a broad view is that I will not be concerned so much with the individual status of different functional notions or forms, but rather with their structural interplay. Example (1) and its possible interpretations provide an idea about this structural interplay in Bulgarian.

¹ This study was supported by the Estonian Science Foundation (grant no. 7006). I am also obliged to Mati Erelt and Björn Wiemer for their valuable comments on previous drafts of this paper. Needless to say, I am solely responsible for any remaining misconceptions or errors.

Interactions between grammatical and lexical markers

(1) *Стефан май бил от Бургас.*
Stefan probably≈it seems **be.PST.PTCP(=EV)** from Burgas

I will argue that, given the lack of additional context, there are four ways in which this sentence may be accessed by the person who hears it.

- (1a) The speaker has heard that Stefan is perhaps from Burgas.
- (1b) The speaker thinks/recalls that he has heard that Stefan is from Burgas.
- (1c) The speaker has heard that Stefan is from Burgas and thinks that Stefan is (perhaps) from Burgas.
- (1d) Stefan seems to be (according to the speaker) from Burgas.

(1a) concerns wide scope evidentiality, i.e. an interpretation in which the word *май* ‘probably, as it seems’ is in the scope of the report and therefore is not considered as a part of the speaker’s utterance, but as a part of the utterance of the person from whom the speaker has obtained the information about Stefan. (1b) relates to the opposite scope relation where the evidential form *бил* (the auxiliary-less past participle) is in the scope of *май*. Here the speaker thinks, but is not sure, that he has heard (somewhere) that Stefan is from Burgas. In (1c) the proposition is modified twice and neither of the two modifiers is superordinate with respect to the other. This means that both the word *май* and the evidential form *бил* independently modify the proposition. In (1d) the items *май* and *бил* are comprehended as parts of a single entity, which as such conveys both reference to the source of information (in this case the unspecified third part) and reference to the speaker’s epistemic judgement. This multiplicity of approaches is due to three possible ambiguities:

- (i) ambiguity as to whether the items are in scopal dependency or not, cf. (1a–b) vs. (1c–d),
- (ii) scope ambiguity, cf. (1a) vs. (1b),
- (iii) ambiguity based on differences in the analysis of the sequential structure of the sentence, i.e. ambiguity between free-collocation and more idiomatic reading of the sequence of grammatical and lexical modifier, cf. (1c) vs. (1d).

I will henceforth refer to the reading (1c) as ‘analytic reading’ and to reading (1d) as ‘concord (or holistic) reading’ of the sequence of grammatical evidential and an epistemic or evidential word.

The major claim of this study is that the concord account for such sequences is in many cases more adequate than an analytic or a scopal one. I will argue that the concord readings are triggered by an overlap in the functions of the

grammatical evidential marker and the epistemic or evidential word. The second important claim is that such concord readings provide a good analytical tool for diagnosing the functions of the grammatical evidentials of any language.

Section 2 introduces the major theoretical problems relevant to this study. Section 3 discusses the specific goals and the method used in the study. The types of interactions between the grammatical evidentials and the epistemic-or-evidential words are presented in Section 4. Section 5 looks for the triggers of what we call the ‘concord (or holistic) reading’, while Section 6 presents further evidence in support of the concord-hypothesis. Section 7 discusses in detail some general consequences instigated by the results of the previous sections. Section 8 summarizes the overall results of the study.

2. Key issues

Since the early 1980s there has been an upswing in interest towards delimiting the conceptual boundaries of the notion of ‘evidentiality’. More specifically, the main puzzle has been (and still is) the relationship between the notions of ‘evidentiality’ and ‘modality’. There is agreement among scholars that what is often referred to as ‘evidentiality’ is a somewhat Janus-faced category residing partly in the epistemic sector of the modal domain and partly outside it. One can distinguish between three major viewpoints as regards the relationship between these two notions: 1) (epistemic) modality and evidentiality are distinct functional categories with no overlap between them (see e.g. Nuyts 2001, 27-28; Aikhenvald 2003a; 2004, 7); 2) there is a subsumption relation between them, i.e. one of these notions includes the other one (see e.g. Chafe 1986; Kiefer 1994; Ifantidou 2001; Boye 2006: 21), and 3) there is an area of overlap, but no full subsumption (see e.g. Kozintseva 1994; Plungian 2001); the main candidate for such an area is provided by the notion of ‘inferentiality’ (see e.g. Palmer 2001, 8-9, 24; Dendale & Tasmowski 2001).²

For the purposes of this study, I will not presuppose any of the above possibilities. The only preliminaries applied henceforth are first, that these notions are conceptually graspable (if not necessarily linguistically discernable), and second, that for any form which expresses both meanings – epistemic judgement and reference to the source of information – one of these meanings is in a given context supposed to be pragmatically foregrounded or more prominent than the other.

The possibility of combining grammatical evidentials with epistemic or evidential lexemes within a single sentence is recognized and seen as a proof of

² In some studies the second and the third possibility are not differentiated.

Interactions between grammatical and lexical markers

the conceptual sovereignty of the notion of evidentiality (Aikhenvald 2004, 257). Unfortunately, there are virtually no studies exclusively devoted to such combinations. One may ask why such combinations should at all be of interest. In essence it boils down to the more general problem about the restrictions on combinability of two or more sentential modifiers, and lately there have been several attempts to formulate the principles underlying such restrictions (see e.g. Hengeveld 1989; Van Valin & LaPolla 1997, 40-52; Alexiadou 1997; Van Valin 2005, 19-21; Ernst 2006, 92-148). We will take a closer look at two studies which are of particular interest to us: Cinque (1999) and Nuyts (forthcoming). These studies are instructive not only because of their explanatory power, but also because they represent opposite theoretical traditions. Cinque (1999), who approaches the problem from the perspective of the Chomskyan generative paradigm, is interested in the relative order of adverbials and functional heads as a possible sign of universal phrase-structure constraint. Nuyts, on the other hand, working within the framework of Cognitive Linguistics, is interested in the combinability of such modifiers as an indication of the hierarchical nature of the qualifications of states of affairs. Both studies arrive at rather strict hierarchies.

To begin with, Cinque's hierarchy is advanced as a universal hierarchy of clausal functional projections. The hierarchy derives from the observation that various classes of adverbs enter a strictly ordered sequence, and this ordered sequence coincides with the order of the dependent morphemes encoding various functional notions (such as mood, modality, tense, aspect and voice). It is then stipulated that these different classes of adverbs enter into a transparent Spec/head relation with the different functional heads³ of the clause. Each specific class of adverbs (e.g. tense adverbs) is an overt manifestation of a distinct functional projection, which in certain languages may also be overtly expressed in the corresponding functional head position (e.g. as a tense affix). The restrictive sequential order of the adverbs or affixes is therefore nothing but a reflection of the hierarchical relations of the corresponding functional projections. Cinque's list of projections is very comprehensive; Figure 1 presents only the first part of it (starting from left), which contains the projections 'evidential' and 'epistemic'.

³ Cinque's notion of 'functional head' corresponds in this case to a dependent (mood, tense, aspect, or other) morpheme.

Figure 1. **The universal hierarchy of clausal functional projections (modified version of Cinque 1999, 106)**

[speech act [evaluative⁴ [evidential [epistemic [tense [irrealis [deontic [habitual [repetitive [frequentative [volitional [celerative [anterior [terminative [continuative ...

As can be seen from this figure, evidentials (lexical or grammatical) precede (lexical or grammatical) epistemic expressions. A direct indication for this relative order is the acceptability of (2a), where the evidential adverb *evidently* precedes the epistemic adverb *probably*, compared to the unacceptability of (2b) where the opposite order is present (see Cinque 1999, 135).

(2) (a) *Evidently John has probably left.*

(b) **Probably John has evidently left.*

Cinque (1999, 141) claims that although many of the relative orders among such functional elements may eventually reduce to scope relations (as maintained in the functionalist literature; see e.g. Bybee (1985), not all orders are so explicable. If the relative order is determined by the fact that the notion of evidentiality is semantically superordinate with respect to the notion of epistemicity, then one should not expect (3) to be acceptable:

(3) *It is probable that it is evident that he is the guilty one.* Cinque (1999, 135)

He assumes therefore that the hierarchy above should be considered a property of the Universal Grammar rather than only a reflection of the semantic structure (see also Cinque 2006, 119-144 for discussion).⁵

The hierarchy of Nuyts (forthcoming) (see Figure 2) is postulated in two-dimensional format, but due to space limitations we will present it here as one-dimensional, thus also simplifying the comparison with Cinque's hierarchy. While in Cinque's hierarchy the relation 'A hierarchically higher than B' was

⁴ As the figure shows, Cinque draws a distinction between evaluative and epistemic modalities. Evaluative modalities do not refer to the degree of certainty in the truth of the proposition, "but rather express the speaker's (positive, negative, or other) evaluation of the state of affairs described in it" Cinque (1999, 84). The following English adverbs could be considered evaluative: *(un)fortunately*, *luckily*, *regrettably*, *surprisingly*, *strangely/oddly (enough)*, *(un)expectedly*.

⁵ See Svenonius (2001, 211) for an explanation of examples like (3), which rescues the semantic scope account.

indicated with 'A [B', in Nuyts' hierarchy the relevant indexation is 'A > B'. A basic primitive in this hierarchy is the cognitive-functional notion of 'qualification'.

Figure 2. **Hierarchy of qualifications of states of affairs [Nuyts, forthcoming]**

evidentiality > epistemic modality > deontic modality > time > quantificational aspect (frequency) > phasal aspect > (parts of the) STATE OF AFFAIRS

Nuyts believes that although this hierarchy is part of syntactic and/or lexical semantic representation in grammar, it does not belong in grammar at all, but rather constitutes a primary dimension of human conceptualization (see also Nuyts 2001, 353-357). This means that although the hierarchy postulated is based on solely linguistic facts about scope relations between (and the grammatical behaviour of) qualificational expressions in language, it obviously reflects basic principles to a much greater extent, which are characterized by Nuyts as principles of human perception and conceptualization.

Both Cinque (1999) and Nuyts (forthcoming) arrive at their hierarchies using material which is strictly constrained with respect to the level of linguistic expression. As many other scholars, they explore the combinability of items, which are either lexical or grammatical, but not the combinability of lexical *and* grammatical markers. Therefore, by addressing combinations of grammatical and lexical markers of evidentiality and epistemicity respectively, we enter into an unexplored area in the research paradigm (see also [Makarcev, this volume] for another contribution on this topic). The sequences of grammatical evidentials and lexical markers of epistemicity or evidentiality are significant because they involve more conventionalized meaning-to-form mapping than the sequences of two lexical markers. According to Mushin (2001, 170), we should expect 'a much higher degree of conventional mapping between actual source of information and adoption of epistemological stance in languages with grammatical evidentiality than in languages which lack such systems.' As a rule, the degree of conventionalization is mirrored in the frequency of the given pattern (see Hopper & Traugott 2003, 126-130; Brinton & Traugott 2005, 100). This means that we should expect in Bulgarian and Estonian, which have grammatical patterns of evidentiality, such combinations to be more frequent than for example in Russian, where no grammatical patterns of evidentiality exist.

Interestingly, Bulgarian and Estonian seem to allow all possible orders and combinations of epistemic words and grammatical evidentials, and therefore we are left with the possibility of drawing generalizations based on the semantic

and pragmatic interpretations of such sequences, but not on their word order patterns. Furthermore, looking at the degree to which such sequences correspond to the above hierarchies, one has to admit that in Bulgarian and Estonian we are dealing with tendencies rather than with rules. Provided that in example (1) we have a co-occurrence of the arguably epistemic lexical marker *май* and the arguably evidential grammeme encoded as *l*-form (auxiliary-less past participle), we can immediately identify a certain discrepancy with the above hierarchies. It is embodied by reading (1b), in which the evidential is in the scope of the epistemic – a pattern which does not conform to the above hierarchies. Nevertheless, these hierarchies appear to be strong structural tendencies as readings like (1b) are extremely rare in Bulgarian and Estonian.

3. Description of the items under consideration

Before analyzing such sequences, we need to specify which particular Bulgarian and Estonian items appertain to the notions of ‘grammatical evidential’ and ‘lexical marker of epistemicity and/or evidentiality’. We can easily delimit the universe in which we operate by referring to notions used by the traditional descriptive grammars.

Let a sequence of a grammatical evidential and an epistemic or evidential lexical marker be a sequence in which a grammatical marker classified as ‘evidential’ by the descriptive grammars of Bulgarian and Estonian co-occurs with a lexical marker classified as ‘a word (adverb or particle) expressing epistemic assessment’. The latter includes both epistemic and evidential lexemes and this is due to the fact that descriptive grammars do not postulate a separate class of evidential adverbs or particles, but include such items in the class of modal (epistemic) words (see GBE II, 405-406, 494-495 for Bulgarian and EKG II, 187-190 for Estonian). With the help of this definition, we considerably restrict the array of sequential types factored in the study. First, we retract those potential grammatical markers of evidentiality which are not descriptively promoted to a categorial status, and second, we retract those lexical markers of epistemicity or evidentiality that belong to inflectional classes, such as verbs for example. The reason for this latter delimitation is that with non-inflectional classes it is easier and faster to look for natural examples on the Internet.

In Bulgarian, the grammaticalized evidential category *преизказно наклонение* ‘renarrative mood’ (or – for those who consider its mood status problematic – the class of *преизказни форми* ‘forms of renarration’) is encoded by past active participles (ending in *-l*), which in third person singular and plural are not accompanied by the auxiliary verb *съм* ‘to be’. Compare the minimal pair in (4); (4a) conveys direct and (4b) indirect evidentiality:

Interactions between grammatical and lexical markers

(4) a. *Стефан замина за Бургас.*

Stefan **leave**.AOR.3SG for Burgas

‘Stefan left for Burgas.’

b. *Стефан заминал за Бургас.*

Stefan **leave**.PST.PTCP for Burgas

‘Allegedly / as it seems, Stefan has left for Burgas.’

From a typological perspective, the classificatory notions of ‘firsthand’ and ‘non-firsthand’ capture the difference between (4a) and (4b) and the Bulgarian evidentiality system corresponds to type A1 in Aikhenvald’s classification (see Aikhenvald 2004, 29-31). The ‘non-firsthand’ term comprises the meanings of reported evidentiality, inferentiality and mirativity, while the ‘firsthand’ term takes default reading in which the information is acquired through vision.

In contrast, the grammatical evidentiality system of Estonian represents the type A3 in Aikhenvald’s classification (see Aikhenvald 2004, 33). Type A3 encodes the distinction between ‘reported’ and ‘everything else’. The ‘reported’ term of the Estonian evidentiality system is labelled *kaudne kõneviis* ‘oblique mood’ and is manifested by the dedicated marker *-vat* suffixed to the first verbal form of the predicate. Compare the difference between the direct evidence in (5a) and the reported evidence in (5b):

(5) a. *Jaan sõit-is Pärnu-sse.*

Jaan **leave**-PST.3SG Pärnu-ILL

‘Jaan left for Pärnu.’

b. *Jaan ole-vat sõit-nud Pärnu-sse.*

Jaan **be**-EV **leave**-PST.PTCP Pärnu-ILL

‘Reportedly, Jaan has left for Pärnu.’

As for the particular items characterized as ‘a word (adverb or particle) expressing epistemic assessment’ we will use the notion of epistemicity as a cover term even though some of these items have evidential meanings. It is well known that markers of evidentiality imply different degrees of certainty about the state of affairs under consideration. In other words, at this stage we will not distinguish between items with focal epistemic meanings and items (such as evidential words) with only implicational epistemic meanings. Instead, we will catalogue lexical items according to the degree of certainty they express (or imply), i.e. we will assign each Bulgarian and Estonian item a rough position on an ‘epistemic scale’ (see Givón 1982; Akatsuka 1985; Nuyts 2001, 22 about this notion). This has been done in Table 1. Such characterization is necessary,

because in Section 5 we will look for a correlation between the degree of certainty that these words induce and the way they interact with grammatical evidentials.

Table 1. **Epistemic and evidential words according to the degree of certainty**

	Bulgarian	Estonian
FULL CERTAINTY	<i>безспорно</i> ‘indisputably’, <i>без съмнение</i> ‘beyond doubt’, <i>несъмнено</i> ‘undoubtedly’, <i>със сигурност</i> ‘certainly’,	<i>kahtlemata</i> ‘undoubtedly’, <i>kindlasti</i> ‘certainly’,
STRONG CERTAINTY	<i>вероятно</i> ‘most likely’, <i>навярно</i> ‘most likely, presumably’, <i>очевидно</i> ‘apparently, evidently’, <i>сигурно</i> ‘probably’,	<i>ilmselt</i> ‘obviously, apparently’, <i>nähtavasti</i> ‘apparently, evidently’, <i>tõenäoliselt</i> ‘probably’,
MEDIUM CERTAINTY	<i>евентуално</i> ‘possibly’, <i>май</i> ‘it seems (that), probably’, <i>може би</i> ‘perhaps, maybe’,	<i>arvatavasti</i> ‘possibly, presumably’, <i>ehk</i> ‘maybe, perhaps’, <i>vahest</i> ‘perhaps, possibly’, <i>vist</i> ‘perhaps, maybe’, <i>võib-olla</i> ‘maybe, perhaps’,
WEAK CERTAINTY	<i>едва ли</i> ‘hardly, scarcely’, <i>надали</i> ‘hardly, scarcely’	<i>vaevalt</i> ‘hardly, scarcely’,

We have 13 Bulgarian and 11 Estonian items whose co-occurrences with the relevant grammatical evidentials will be studied in the remaining part of this paper. It is worth noting that we will only be concerned with those cases where these words function as sentential modifiers, i.e. with cases where their scope is equivalent to the scope of the grammatical evidentials, which always operate at sentential level. This means that co-occurrences of grammatical evidentials and epistemic or evidential words where the latter have constituent scope are not factored in the present study.

An important caveat is that the borders between the four degrees of certainty are drawn intuitively, and it is possible that exclusive testing of the degrees of certainty would slightly modify the above classification, especially with regard to the middle area in the table. Another intuition is that the invariant meanings of the majority of the items in the table are epistemic. Only Bulgarian *очевидно* ‘apparently, evidently’ and Estonian *ilmselt* ‘obviously, apparently’ and *nähtavasti* ‘apparently, evidently’ always convey evidential meanings. With

Interactions between grammatical and lexical markers

regard to Bulgarian *навярно* ‘most likely, presumably’ and *май* ‘probably, it seems (that)’ as well as to Estonian *tõenäoliselt* ‘probably’, they seem vague with respect to the distinction between epistemicity and evidentiality. As it is irrelevant for the purposes of this study whether these intuitions are fully sound or not, they will not be further tested.

The sentences where a grammatical evidential co-occurs with an epistemic or evidential word were collected from the Internet. To make the search as simple as possible, I looked only for co-occurrences where the epistemic or evidential lexeme immediately precedes or follows a verbal form grammatically marked as evidential.

4. Types of interactions

We already saw that there are four ways in which the sentence in (1) can be accessed. Accordingly, we will draw distinctions between four types of interaction between evidential grammemes and epistemic lexemes. In this section, I will discuss these types in greater detail, illustrating them with attested Bulgarian examples. Each type is introduced with a Roman number, cf.:

I. An epistemic lexeme is in the scope of an evidential grammeme. This type is exemplified in (6), an example from an online forum. The evidential form (the auxiliary-less past participle *имал* ‘have.PST.PTCP’ in bold) indicates that the speaker refers to someone else’s words. Although the referent is non-specific – what is referred to is rather the attitude of the ruling political class in the US –, the adverb *несъмнено* ‘undoubtedly’ is perceived as a part of the reported statement and is thus within the scope of the report. If EV stands for a grammatical marker of evidentiality, EP for an epistemic word, *p* for a proposition and square brackets indicate scope relations, this type can be formally represented as [EV [EP [*p*]]].

- (6) *Днес всички интересуваци се знаят, че това е била*
today all interested know.3PL that this be.3SG be.PST.PTCP
чиста лъжа. Чиста лъжа беше и повод-ът за
obvious lie obvious lie be.IMP.3SG also occasion-DEF for
интервенция-та в Ирак – Садам несъмнено имал
intervention-DEF in Iraq Saddam **undoubtedly** **have.PST.PTCP**
ОМП. *Оказа се, че не е*
weapons_of_mass_destruction it_turn.AOR.3SG_out that NEG be.3SG

имал, ама какво от това.
have.PST.PTCP but what of this

‘Now all people who are interested know that this was an obvious lie. An obvious lie was also the excuse for the military intervention in Iraq – **reportedly, there wasn’t any doubt** that Saddam **has** weapons of mass destruction. Well, he didn’t have them as it turns out, but so what.’
(<http://www.pro-anti.net/show.php?article=1&issue=735>)

II. An evidential grammeme is in the scope of an epistemic lexeme. Consider the background of the story in (7). For a month or so Nasko has had some strange wounds on his legs. As he is not able to recover, he visits his Turkish friend whose grandmother says that Nasko will die if he does not follow her advice. She gives him the address of a certain imam and Nasko goes to see him. Despite the fact that the whole story is grammatically marked for indirect evidentiality (every predicate is in the form of auxiliary-less past participle), the evidential form *дал* ‘give.PST.PTCP’ (in bold) is within the scope of the word *май*. As already noted, this word lends itself both to the domain of epistemicity and to the domain of evidentiality, and can be roughly translated as ‘probably’ or ‘it seems’. The speaker in (7) no longer remembers the details reported in the sentence containing the forms in bold. By embedding the reference to the source of information in the scope of *май*, he indicates that he does not trust his memory. This type can be formally stated as [EP [EV [p]]].

(7) Дала му координати-те на някакъв ходжа и той
give.PST.PTCP him coordinates-DEF of some imam and he

отишъл при него. Ходжа-та му **дал** **май**
go.PST.PTCP to him imam-DEF him **give.PST.PTCP** **probably**≈as it seems

няква кърпа / не си спомням точно вече / да спи върху нея
some piece_of_cloth NEG remember.1SG exactly anymore to sleep on it

или нещо такова и след това да му я занесе.
or something like_this and after this to him it bring.3SG

‘(Reportedly), she had given him the address of some imam and he had gone to him. And then (**I think the story went like this**), the imam had **given** him a piece of cloth, – well, I don’t remember exactly anymore – to sleep on it, or something like this, and to bring it back to him after that.’
(<http://forum.rozali.com/viewtopic.php?p=32156&sid=94472861ac1c962874a819c64d2620b9>)

Interactions between grammatical and lexical markers

Examples like (6) and (7) show that in Bulgarian scope relations are not coded by word order. In both examples the direction of scope dependency is opposite to the relevant order of the grammatical evidential and the epistemic word with respect to the propositional core. The word order is also irrelevant for the description of the remaining two types of interaction. In these two types neither the grammatical evidential nor the epistemic word takes scope over the other one, which means that these items share exactly the same semantic scope. Although often disregarded in the study of evidentiality (and related categories), such cases are not exceptional cross-linguistically (see [Aikhenvald 2004, 87-95] and [Boye 2006, 191-194] for examples).

III. The two items modify independently the proposition; witness the example in (8). The sequence in bold indicates that although the speaker does not have direct evidence to show how educated Abraham was, he is confident (based on common knowledge) that Abraham was a very educated man. Here the sequence of the grammatical evidential and the epistemic word can be seen as a free collocation of two forms, which both maintain their functional identity – the first expressing reported evidentiality and the second certainty. This type can be formally stated as [EV [*p*]] & [EP [*p*]], i.e. as a conjunction of two qualifications of the status of *p*. As already noted in Section 1, I will conventionally speak in this case about ‘analytic reading’ of the sequence of evidential grammeme and epistemic word.

- (8) Днес обаче знаем, че Авраам в никакъв случай не може да
today however know.1PL that Abraham in NEG.PRON case NEG can.3SG to
бъде поставен на едно равнище и сравняван с
be.3SG place.PASS.PTCP at one level and compare.PASS.PTCP with
примитивни-те, суеверни бедуини. Обратно, той бил
primitive-DEF superstitious Bedouins on_the_contrary he be.PST.PTCP
със сигурност високопросветен човек, потомък на културно и
with confidence highly_educated man descendant of cultivated and
високоцивилизовано общество.
highly_civilized society

‘Yet, we know by now that there is no way in which Abraham can be put on the same level as the primitive and superstitious Bedouins. On the contrary, he is **supposed to be**, and we are **confident** about this, a well-educated man, a descendant of highly civilized society with great cultural achievements.’
(<http://www.bgbible.sdabg.net/arheolog/a-6.htm>)

IV. The two items are understood as a single entity, which as such includes both reference to the source of information and reference to the speaker's epistemic judgement. Which one of these referential meanings prevails in this unified form, depends first on discourse-pragmatic factors and second on the genuine semantics of the items classified in the grammars as 'evidential' and 'epistemic'. Using a hyphen as a mark of structural blending, this type can be stated as [EV-EP [p]]. Consider (9), in which the speaker addresses his forum mates, asking them for a favour. The string in bold cannot be disassembled into an expression of epistemicity and evidentiality. Rather, it is perceived as one unit, which as a whole conveys uncertainty based on indirect evidence. This effect is due to the fact that neither of the forms involved can be identified as only evidential or only epistemic; instead, both can express both meanings. We saw that the word *май* can be translated as 'probably' or 'as it seems'. In addition, the auxiliary-less past participle *бил* ('be-PST.PTCP') conveys here inferentiality (an inference based on remote evidence) and has a scent of doubt in its presupposition. As a result, the two formal elements merge into one unit expressing hesitation and doubt over the state of affairs expressed in the proposition. As noted in Section 1, I will speak in such cases about 'concord (or holistic) reading' of the sequence of grammatical marker of evidentiality and epistemic word.

- (9) *Ако някой скоро ще проверява в архив-а в Търново [...] нека*
 if somebody soon FUT check.3SG inarchive-DEF in Tŕrnovo HOR
да види информация за име-то Цеко Иванов (или) Иванчов
 to see.3SG information about name-DEF Tseko Ivanov or Ivančov
Драшански от гр. Бяла Слатина бил се по
 Drašanski from town Byala Slatina fight.PST.PTCP on
фронтовe-те на Добруджа през ПСВ, не знам в
 fronts-DEF of Dobrudja duringFirst_World_War NEG know.1SG in
коя частe бил или какъв чин e имал.
 which unit be.3SG be.PST.PTCP or what rank be.3SG have.PST.PTCP
*Мисля че **май бил** убит от свой другар по*
 think.1SG that **maybe**≈as it seems **be.PST.PTCP** killed by his friend by
погрешка докато оня нещо си оправял пушка-та, но това
 mistake while that something fix.PST.PTCP rifle-DEF but these
са само догадки.
 be.3PL only guesses

Interactions between grammatical and lexical markers

‘If anybody is going to the archives in Tärnovo in the near future, please let him check for information about the name Tseko Ivanov (or) Ivančov Drašanski from the town of Byala Slatina, who fought on Dobrudja front during the First World War; I don’t know in which unit or what rank he would have had. I think he **may have been** killed by mistake while his friend was fixing his rifle, but these are only guesses.’

(<http://forum.boinaslava.net/archive/index.php/t-6691.html>)

Note that when we mentioned ambiguity between different interpretations of (1), we did not refer to any actual ambiguity in the specific speech situation. Rather, we were concerned with the possibility of different interpretations of a particular sentence in different contexts. As can be seen from examples (6)–(9), the co-occurrences of grammatical evidential and epistemic word are usually given specific interpretation by the context, or in other words, they are disambiguated by the contexts.

In the next section we will focus on the properties of type IV, which is of main interest in this study, and we will try to identify the conditions which license such concord interpretations.

5. The triggers of the concord (or holistic) interpretation

The opposition ‘analytic vs. holistic’ access to a linguistic element was introduced into the study of grammaticalization and lexicalization by Lehmann (2002b). In philosophy of language, the doctrine of semantic holism defended by Quine (see Quine 1953) refers to the effect that a certain part of language can only be understood through its relations to a (already understood) larger segment of language. In our specific case, the concept of ‘concord (or holistic) reading’ refers to a very early phase of semantic coalescence of two formally distinguishable propositional modifiers. I will maintain that this effect is triggered by an overlap in the functions of what we referred to as ‘grammatical evidential’ and ‘epistemic word’. Previous acknowledgments of this effect on co-occurring modals include Halliday (1970, 331), Lyons (1977, 807-808) and Coates (1983, 46, 138). In a recent study, Kasper Boye classified cases with two co-occurring epistemic markers⁶, which share the same scope into four types (see Boye 2006, 189-196). His second type coincides with our notion of concord interpretation, cf.:

‘Two or more epistemic items or constructions which have overlapping meanings (...) may co-occur non-obligatorily in a unified expression of

⁶ In Boye’s system evidentials belong to the epistemic meaning domain.

justificatory support for a proposition. [T]he co-occurring epistemic items or constructions cooperate in what may be thought of as the specification of one epistemic meaning.’ (Boye 2006, 191)

5.1. Kinds of functional overlap

Let us first elaborate on the different ways in which the meanings of the linguistic elements may overlap. In both Bulgarian and Estonian the element characterized as ‘grammatical evidential’ is a constant and the element characterized as ‘epistemic word’ is a variable. This is obvious as in both languages there is only one grammaticalized category of evidentiality, but many lexical words expressing epistemic and/or evidential meanings. Therefore we may say that the variable ‘epistemic word’ takes different values in a single language. We already saw that one of the major distinctions that can be drawn between the items in Table 1 is that some of them have predominantly epistemic functions, whereas other predominantly evidential functions. Based on the distinction between words with primary (or foregrounded) epistemic meaning and such with primary (or foregrounded) evidential meaning we can distinguish between two subtypes of our type IV:

IV(a): [EV-EP_{≈EP} [p]];

IV(b): [EV-EP_{≈EV} [p]].

Once more, the index EP should not be misleading – it signifies a word which is classified in the descriptive grammars of Bulgarian and Estonian as expressing epistemic assessment. The indices EP and EV on the other hand stand for the actual functional value of the signatum of EP. The first subtype is exemplified in (10); the story is about the security surveillance cameras:

- (10) *Видеомасивите на Лондон и Кайро не са интегрирани*
video_surveillance_areas of London and Cairo NEG be.3PL integrated
на определено ниво. После се оказа, че експлозивът
at certain level then turn_out.AOR.3SG that explosive-DEF
може би имал балкански произход – но дали
maybe have.PST.PTCP Balkan origin but Q(yes/no)
складове-те и въобще работа-та с взривни материали у
storehouses-DEF and in_general work-DEF with explosive material at
нас например се контролира по подходящ начин?
us for_example monitor.IMPS in adequate way

Interactions between grammatical and lexical markers

‘The areas with video surveillance in London and Cairo are not integrated at a certain level. It turned out then that the explosives **were maybe** of Balkan origin – but who knows whether here (in Bulgaria, P.K.) for example the storage of the explosive materials and the work with them in general is monitored in an adequate way.’

(http://newteck.orbitel.bg/computer/articles.php?issue_id=56&unit_id=495&article_id=1651)

We are dealing here with a semantically and pragmatically unified expression, although it can be formally deconstructed into an expression of inferential evidentiality inducing a certain amount of doubt, and into a corroborating adverb of middle certainty (*може би* ‘perhaps, maybe’).

Comparing the examples of concord reading (9) and (10), we can see that in both cases the function of the Bulgarian grammeme expressing indirect evidentiality is contextually specified as inferential. This observation is consonant with the view according to which the inferential meaning is located at the borderline of evidentiality and modality (see van der Auwera & Plungian (1998, 85); Palmer (2001, 8-9, 24-26); Dendale & Tasmowski 2001). It is crucial however, that the concord reading is licensed also in cases where the function of the evidential form is not specified as inferential. Witness (11), in which the writer sends a query to an online forum. The evidential form *бил* ‘be.PST.PTCP’ conveys here a hearsay from unspecified source(s) and together with *май* ‘probably, it seems’, which is vague between epistemicity and evidentiality, expresses uncertainty in the truth of the proposition. Here also, I believe, a concord (or holistic) analysis is more adequate than an analytic one.

- (11) *Знае ли някой как стои въпросът с цена-та на*
know.3SG Q(yes/no) anybodyhow stands question-DEF with price-DEF of
хостинг-а за Java и PHP, защото искам да си правя
hosting-DEF for Java and PHP since want-1SG to REFL do.1SG
нещата на Java, а не на PHP (не го знам и не
things-DEF on Java but NEG on PHP NEG it know-1SG and NEG
искам да го уча ако може). Май бил
want-1SG to it learn.1SG if possible **probably**≈as it seems **be.PST.PTCP**
в пъти по-скъп.
many times more_expensive

‘Does anybody know how the question stands with regard to the hosting price for Java and PHP since I want to do my things with Java, not with

PHP (which I'm not familiar with and don't want to learn if at all possible). It is **supposed to be** many times more expensive.'
 (<http://clubs.dir.bg/showflat.php?Board=java&Number=1937935356&page=0&view=collapsed&sb=3&vc=1>)

Further evidence for the claim that the concord reading is not triggered only in contexts marked for inferential evidentiality is provided by Estonian. As already noted in Section 3, the Estonian grammatical evidential only encodes the meaning of reported evidentiality. If the necessary condition for the concord interpretation were the inferential reading of the grammatical evidential, then one would not expect to find instantiations of this interpretation in Estonian. Nevertheless, such instantiations are attested. Consider (12) in a short commentary on black economy and prostitution, published in the Estonian newspaper 'Sakala' on 12 March 2004.

- (12) *Olen mitme tuttava käest päri-nud, kas Viljandi-s*
 be.1SG several acquaintance from ask-PST.PTCP Q(yes/no) Viljandi-INE
on võimalik prostituuti tellida. Kõik nad on vasta-nud
 be.3SG possible prostitute order all they be.3SG respond-PST.PTCP
õlakehituse-ga või öel-nud kõhklevalt, et mõne aasta eest
 shrug-COM or say-PST.PTCP hesitantly that some year ago
vist ole-vat saa-nud küll.
perhaps be-EV get-PST.PTCP indeed

'I have asked many of my acquaintances if it is possible to order a prostitute in Viljandi (a town in South Estonia; P.K.). All of them have answered by shrugging their shoulders, or hesitantly saying that some years ago it **had (perhaps) been** possible.'
 (<http://vana.www.sakala.ajaleht.ee/rubriigid.html?number=576>)

Here the refusal to give a univocal answer to the question asked by the narrator, i.e. the avoidance of responsibility, is explicitly marked by the word *kõhklevalt* 'hesitantly'. Therefore the sequence in bold is best accessed holistically. None of the other readings seems adequate. Reading this sentence one would hardly assume that the word *vist* 'perhaps' is in the scope of the report (cf. reading I). Even more unlikely is the reading with narrow scope evidentiality (cf. reading II), i.e. where the people answering the question are not sure whether they have heard the facts about prostitution or not. The analytic reading (cf. reading III) is discarded because the word *vist* 'perhaps, maybe' and the hearsay form *ole-vat* 'be-EV', which induces doubt, are felt too similar and somehow functionally

redundant, which means that we are not dealing with two independent qualifications of the status of the proposition.

The discussion so far brings us to the necessary conclusion that the functional overlap of the grammatical evidential and the epistemic words is larger than the area covered by the notion of inferentiality.

Consider now the second group of cases with concord reading – the subtype [EV-EP_{≈EV} [*p*]]. Here the grammatical evidential co-occurs with an evidential word and the resultant effect is different from the effect triggered by the first subtype. The difference is due to the fact that the functional domain covered by the grammatical evidential does not overlap with, but rather contains the functional domain covered by the evidential word. This is a natural consequence of the fact that lexical expressions are more specific in meaning than grammatical ones. Such sequences are therefore best accounted for in terms of specification and disambiguation (terms used by Boye 2006, 133, 191). Witness (13), where the grammatical evidential encodes a more general evidential term and the evidential adverb specifies/disambiguates the meaning of this term. The evidential form *имал* ‘have.PST.PTCP’ expresses general indirect evidentiality and the adverb *очевидно* ‘apparently’ specifies it into inferential evidentiality and more specifically into an inference based on visual perception.

- (13) *Обсъдихме* *впечатления-та* *от състезание-то* *и от други*
 discuss.AOR.1PL impressions-DEF from contest-DEF and from other
неща *случващи се* *наоколо* *в един фургон,* *на студени* *безалкохолни*
 things taking_place around in one van at cold non-alcoholic
напитки. *Charlie* *(който* *очевидно* *имал* *тежка* *вечер)*
 drinks Charlie who **apparently** **have.PST.PTCP** heavy evening
прецени, *че* *просто* *нещата* *няма* *така* *лесно* *да*
 decide.AOR.3SG that simply things NEG.FUT so easily INF.COMP
се избистрят *в глава-та* *му* *и* *бе* *принуден* *да* *си*
 become_clear.3PL in head-DEF his and be.AOR.3SG forced to REFL
вземе *бира.*
 take.3SG beer

‘We discussed our impressions of the contest and from other things going on while sitting in the van with our soft drinks. Charlie (who **obviously had had** a heavy night) decided that the things in his head wouldn’t become clear that simply and was forced to get himself a beer.’

(<http://clubf1.net/forums/index.php?showtopic=3797&pid=83133&mode=threaded&start=>)

A native speaker of Bulgarian may have doubts in the acceptability of the sentence containing the grammatical evidential form and the evidential adverb in (13). It may seem elliptic, in which case the omitted element would be the form *беше* ‘be.IMPF.3SG’ occurring between the two forms in bold. Adding this form, we would have a free collocation of the word *очевидно* ‘apparently’ and the regular form of pluperfect indicative *беше имал* ‘be.IMPF.3SG have.PST.PTCP’. One may speculate that the example in (13) is nothing but a case of auxiliary ellipse caused by carelessness. Nevertheless, similar occurrences are also attested in edited texts. An example is found in a historical survey of the uprising of the Bulgarian Paulicans in the eleventh century; cf. (14).

- (14) *Както разказва византийска-та историчка Анна Комнина, през*
as narrate.3SG Byzantine-DEF historian Anna Komnene during
1084 г. (начало-то на въстание-то) Травъл бил вече
1084 year beginning-DEF of uprising-DEF Traväl be.PST.PTCP already
шеста година личен и доверен ‘слуга’ (т.е. един от близки-те
sixth year personal and confidential servant i.e. one of close-DEF
адютанти или оръженосци) на нейния баща, прославени-ят
adjutants or armour-bearers of her father celebrated-DEF
пълководец и император Алексий I Комнин (1081–1118).
military_commander and emperor Alexios I Komnenos 1081–1118
Като пресметнем време-то на тази служба, то пловдивският
as calculate-1PL time-DEF of this position then PLOVID.ATTR
*павликянин **очевидно бил** сред най-верни-те хора на*
paulican **apparently** be.PST.PTCP among most_faithful-DEF people of
Алексий I Комнин още от 1078–1079 г. и е
Alexios I Komnenos ever since 1078–1079 year and be.3SG
преживял заедно със своя ‘патрон’ редица победи и
experience.PST.PTCP together with his patron series_of victories and
триумфи, както и немалко тежки изпитания.
triumphs as_well_as and not_a_few hard probations

‘As is documented by the Byzantine historian Anna Komnene, in the year 1084 (in the beginning of the uprising) Traväl had already for six years been a trusted personal ‘servant’ (i.e. one of the close adjutants and armour-bearers) of her father, the celebrated military commander and emperor Alexios I Komnenos (1081–1118). If we add up the time of his

Interactions between grammatical and lexical markers

service, then the Plovdiv-born Paulican **had apparently been** one of the most trusted people of Alexios I Komnenos ever since the years 1078–1079 and had experienced together with his patron a number of victories as well as many hardships.’
(http://litenet.bg/publish13/p_pavlov/buntari/travyl.htm)

As already noted, in such occurrences the evidential word specifies/disambiguates the more general evidential term encoded by the grammatical evidential. This effect is possible only in Bulgarian, because the Bulgarian indirect evidential is such that it can be specified into narrower terms by the lexical items listed in Table 1. The Estonian evidential term is narrowly reported, and therefore co-occurrences with inferential adverbs like those in (13) and (14) cannot have concord readings, but only readings where the items in the sequence are in scopal dependency, cf. (15).

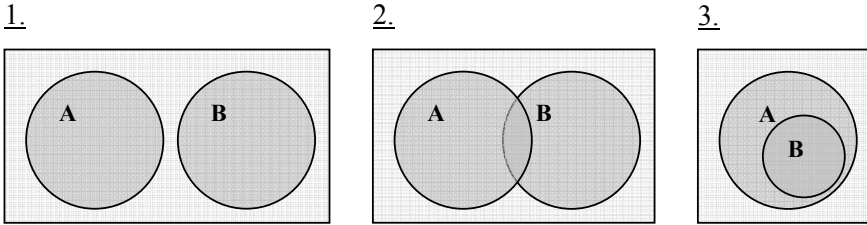
(15) *Ta* **ole-vat nähtavasti** *töö-l*.
s/he **be-EV** **apparently** work-ADE

‘It is said that s/he **is apparently** at work.’

This granted, we can now try to identify the licensing conditions of what we called ‘concord (or holistic) reading’. As we are dealing with overlapping domains, this can be done with the help of the tool from the set theory known as Venn diagrams. Before doing that, we will take for granted, that evidentiality and epistemic modality imply each other. Any explicit marking of the source of information correlates with the degree of the speaker’s commitment to the truth of the proposition, and vice versa, the degree of commitment correlates with the temporal or spatial distance between the state-of-affairs referred in the proposition and the participant which functions as a deictic center of the utterance.⁷ Consider now the types of intersection illustrated in Figure 1. We have only two sets, indicated respectively with A for the meaning domain covered by the grammatical evidential, and B for the meaning domain covered by the epistemic or evidential word. We will be interested here only in cases where the grammatical evidential and the epistemic (or evidential) lexeme co-occur and share the same semantic scope. Bearing this in mind we can identify the following types of intersections.

⁷ This entails that evidentiality is a deictic notion, a position recently defended by de Haan (2005) that I agree with.

Figure 1. **Intersections of the meanings of the grammatical evidential and the epistemic (or evidential) word**



In the first diagram, there is no overlap in the meanings of the grammatical and the lexical item. In such cases we expect that their combination is accessed always analytically as a free collocation of two independent qualifications of the propositional content. This corresponds to our type III.

In the second diagram, the intersection is not an empty set, i.e. there is an overlap between the meanings of the two elements. In this case, their combination can be accessed holistically as subtype IV(a). The intersection comprises at least the notional domain of inferentiality, but does not seem to be restricted to it. As a result, the two elements are comprehended as somewhat redundant. In certain functionalist studies of modality (see [Lyons 1977, 807-808; Coates 1983, 45-46, 137-138; Palmer 2001, 35]), similar cases are labelled ‘harmonic combinations’ of two modals while in the formally orientated studies this phenomenon has been recently labelled ‘modal concord’ [Geurts & Huitink 2006] in parallel with the phenomenon called ‘negative concord’ where two overt negators yield a single operator. Halliday noted already in 1970 that in such cases the co-occurring modals reinforce each other [Halliday 1970, 331]. In our case, the ‘reinforced’ meanings are those meanings that are confined to the overlapping area in the diagram.

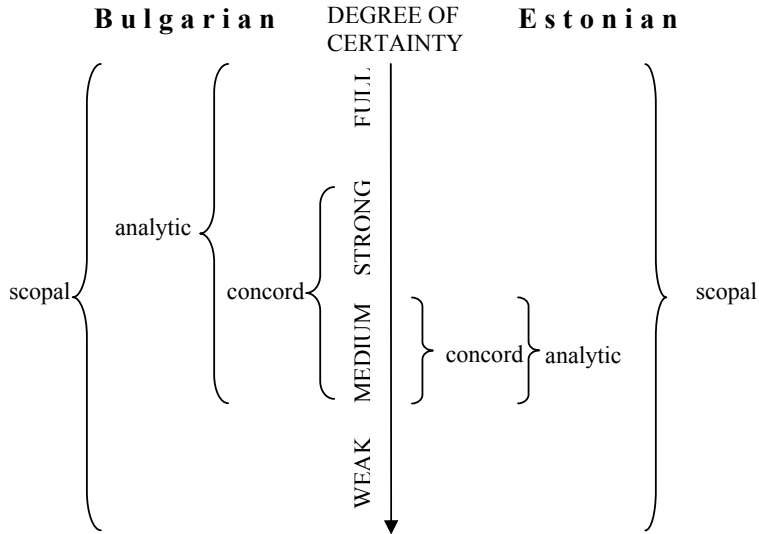
In the third diagram, one of the sets is confined within the other one. As already said, such cases can be accounted for in terms of specification or disambiguation. The item whose functional range is designated by the set A can be said to be underspecified, and the item whose functional range corresponds to the set B specifies it. This is our subtype IV (b), which as we saw, is attested only in Bulgarian. As the Estonian evidential grammeme encodes reported evidence, combining it with an inferential word would yield either scopal reading, i.e. reading in which either the inference is in scope of the report or the report is in the scope of the inference, or same-scope, but ‘analytic’ reading. If, on the other hand, the Estonian evidential is combined with a particle or an adverb expressing reported evidentiality, the meaning areas covered by these items would just coincide and bring about redundancy.

5.2. Medium certainty as a licenser of the concord reading

Let us now elaborate on the question of how the semantics of the particular epistemic and/or evidential words correlate with the different readings. In particular, we are interested to discover which items of those listed in Table 1 engage with the evidentials in combinations that can be accessed holistically. It seems that such readings are licensed only in combinations where the grammatical marker of evidentiality co-occurs with a lexeme expressing or implying a middle degree of certainty (or commitment) that the facts referred to by the proposition obtain. An apparent explanation for this tendency is that the evidential grammemes of Bulgarian and Estonian imply middle rather than full or low certainty. Using an expression of indirect evidentiality usually implies that the speaker does not want to commit himself with an opinion as to whether the reported state of affairs obtains or not. Therefore in those combinations where the grammatical evidential and the lexical expression of middle certainty share the same scope, the latter are perceived as somehow superfluous. Due to the meaning intersection, the analytic reading is usually not selected and the only remaining way is that of convergence of the two items into a single unit, where the epistemic stance which can be paraphrased as ‘I am not absolutely sure in p ’ or ‘I do not fully commit myself to p ’ is reinforced.

With regard to the combinations with words expressing full or weak certainty, our evidence is that the first can receive either scopal (types I and II) or – only in Bulgarian – analytic interpretation (type III), whereas the latter seem to be restricted to only scopal interpretation. The reason for this is that words expressing low certainty (or commitment) can be considered as implicit negators: qualifications like ‘hardly p ’ normally imply ‘not p ’. As a result such combinations yield clearly two distinct operators, one of which (evidentiality) is not truth-functional whereas the other one (negation) is truth-functional, and hence the combination cannot be accessed holistically. These observations are illustrated in Figure 2 which follows the conventional distinction between four degrees of certainty used in Table 1.

Figure 2. **Correlation between the type of interaction and the degree of certainty adduced by the epistemic (or evidential) word in Bulgarian and Estonian.**



The distribution in the figure leads to a number of repercussions for the licensing conditions of the different readings. First, it is clear that the scopal interpretation is possible regardless of the degree of certainty adduced by the lexical word. In both Bulgarian and Estonian there were instantiations of scopal dependency between the grammatical evidential and the epistemic (or evidential) word all along the spectrum of certainty adduced by the latter.

With regard to the difference between the analytic and concord reading, it is not easy to decide between them without a thorough examination of the relevant context. The area on the scale where the analytic reading is available is larger or equivalent to the area permitting concord interpretation. This is a data-driven generalization, which can also be reached deductively. If we assume that any free collocation of two independent qualificational expressions is prior to cases where these expressions are felt as parts of a single semantic unit, it follows that the analytic reading is licensed at least in the same environments where the holistic reading is licensed. In Bulgarian, cases where the grammatical evidential co-occurs with an epistemic or evidential word inducing full degree of certainty could only have scopal or analytic reading, whereas the concord reading is reserved for co-occurrences with words inducing medium-to-strong certainty. The instantiations of concord reading with words inducing strong

Interactions between grammatical and lexical markers

certainty are rarer and more specific than those with words expressing medium certainty. Strong certainty items can only yield the type of concord reading that we characterized as evidential specification (recall examples (13) and (14)).

Compared to Bulgarian, Estonian seems to be more restrictive with regard to the availability of the analytic and the concord interpretations. Combinations with words inducing strong certainty always receive scopal interpretation. For example, the Estonian sentence in (16) where the grammatical evidential is combined with an adverb expressing strong certainty can only receive scopal interpretation:

- (16) *Ta ole-vat tõenäoliselt töö-l.*
s/he **be-EV** **probably** work-ADE

‘It is said that s/he **is probably** at work.’

*‘It is said and I think it **is probable** that s/he is at work.’

The difference between Bulgarian and Estonian in the availability of the different readings mirrors the functional range of the respective grammatical markers of evidentiality. The Bulgarian evidential grammeme has broader semantics and so the array of epistemic words with which it can form a ‘harmonic combination’ is larger than in Estonian. In Section 7, we will return to this regularity and look for an operational device that can be used to detect the precise semantic range of any evidential grammeme in any language.

5.3. Non-specific referent as a licenser of the concord (or holistic) reading

We noted in Section 5.1. that the concord interpretation is not restricted to cases where the grammatical evidential receives an inferential specification, but also to cases where it has hearsay interpretation. The analysis of the Bulgarian and Estonian data pinpoints a specific condition that triggers the concord reading in such cases of reported evidentiality. It relates to the identity of the referent of the report. The grammatical evidential may be used in a context which specifies the individual from whom the speaker has acquired the information concerning *p* or in a context which does not specify the source of the report, but rather indicates that the speaker has acquired the information about *p* from different sources at different times or that determining the referent of the report is irrelevant in the given speech situation. In the first case we can talk about a specific referent of the evidential expression, in the second about non-specific referent of the evidential expression.

In Bulgarian and Estonian, the concord reading is found in contexts with non-specific referent. Consider the following example from Bulgarian. It comes from

an Internet discussion concerning an earthquake which took place the day before. The earthquake was light; it was felt only by some people in certain districts of Eastern Sofia, where the protagonist of the story lives.

- (17) *Аз пък си помислих, че съсед-а Тошо се е
 I but REFL.DAT think.AOR.1SG that neighbour-DEF Tošo REFL be.3SG
 изтърсил по гъз. И нищо чудно, нали епицентър-а
 tumble_down.PST.PTCP on ass and nothing surprising isn't epicentre-DEF
 май бил в Младост ...
 probably~as it seems be.PST.PTCP in Mladost*

‘And I thought that my neighbour Tošo fell on his backside. It doesn’t surprise me, wasn’t the epicentre **supposed to be** in Mladost (a residential area in Sofia; P.K.).

(<http://muro.biz/old/?p=253>)

It seems that the author of (17) has gathered the information about the epicentre of the earthquake from one or several sources, none of which is fully reliable. The referent of the evidential qualification is not contextually specified, which in turn triggers an interpretation in which the hearer cannot identify any scopal relation between the two elements in bold. Considering that these elements have partly overlapping meanings, the speaker is left with the possibility for interpreting them in terms of redundancy and reinforcement, in which case the concord interpretation is activated. These stipulations are easy to check. If we insert into the second sentence of (17) a clause referring to a specific source of information about the epicentre then the scopal interpretation [EV [EP [p]]] arises leaving no space for any other interpretation. In other words, if we added a clause like ‘according to my uncle’ to (17) the word *май* automatically receives a narrow scope interpretation with respect to the expression of evidentiality.

This observation has important consequences. It seems that utterances with a non-specific referent of the report involve an inferential step by the speaker. This means that the evidential qualification in sentences like (17) can be paraphrased as ‘From what I have heard, I infer that *p*’ or ‘I guess from hearsay that *p*’. In fact, it is this inferential step which binds the reported evidentiality and the degree of certainty in a unified expression. It is interesting that Estonian – a language with a grammaticalized term of reported evidentiality, also allows such inferential interpretations in contexts with non-specific referent. Witness (18), where the story is about the Eurovision song contest.

Interactions between grammatical and lexical markers

- (18) *Aga sel aastal oli siiski üllatavalt hea. Tavapärase*
 but this year be.PST.3SG however surprisingly good usual
saastalaadung loomulikult ka – lood, mis mitte millegagi
 scum_load naturally too songs which NEG.ADV anything.ENCL
silma ei paista ja mis pane-vad mõtlema, et kui sellised
 eye.INE NEG shine and which put-3PL think.SUP that if such
lood on saa-nud 10 parema hulka, siis milline see
 numbers be.3SG get-PST.PTCP 10 best.GEN set.INE then what_kind_of this
üldine tase veel oli, mis ka üllatavalt kõrge vist
 common level again be.PST.3SG which also surprisingly high **perhaps**

ole-vat ol-nud.

be-EV be-PST.PTCP

‘But it was surprisingly good this year. The ordinary crap, too, of course – faceless songs that make you think that if such songs are among the ten best, then what might the general level have been, which is also **supposed to have been** surprisingly high.’

(<http://www.allstarz.ee/foorum/viewthread.php?fid=12&tid=629&action=printable>)

The source of information about the quality of this year’s Eurovision is undetermined and probably of dubious reliability. The underspecification of the information source together with the adverb expressing less than full certainty leads to an inferential reading of the sequence in bold. If the information about a particular event is obtained through sources of different reliability, the speaker synthesizes it in a similar way as in a typical case of inferentiality where he gathers pieces of physical evidence for *p*.

These facts lead to two important repercussions. First, they indicate that the licensing conditions we outlined presuppose each other. In other words, it seems that the syntagm of the grammatical evidential and the word expressing medium certainty is assigned concord reading only in case the referent of the report is not specified, and conversely, a non-specific referent triggers concord interpretation only in case the grammatical marker is combined with a word expressing medium certainty.

The second and more important repercussion is that contrary to the stipulation made in Section 5.1, the instantiations of concord reading may still be reduced to the functional notion of inferentiality. Even if second-hand evidence is involved, this evidence is filtered out through what we called an inferential step.

6. Further evidence for the ‘concord’ hypothesis

The reader may have noticed that I have so far not presented any empirical evidence showing that what was called the concord (or holistic) reading really exists. Speaking about the holistic reading presupposes at least some degree of formal bondedness, i.e. an increase of the intimacy with which the two collocating elements are connected to each other (see [Lehmann 2002b, 131-139]). In our case, however, there are no direct signs whatsoever of increased bondedness, and therefore we have to admit that the ‘holistic’-claim belongs to the realm of native linguistic intuitions that cannot be easily verified. The ‘concord’-claim, on the other hand, is a weaker version of the ‘holistic’-claim and therefore seems more suitable for the description of cases where a certain amount of semantic but not necessarily formal coalescence is at play.

One sign for the increase of the intimacy between the two elements is provided by their collocational frequencies. There is evidence indicating that the Bulgarian and Estonian grammatical evidentials show preferences with respect to particular lexical items. Table 2 presents some (preliminary) statistical data for the existence of such preferences. The first column of the table indicates the number of the most frequent collocations of the grammatical marker of evidentiality and epistemic or evidential word. In Bulgarian, the most frequent collocation is ‘evidential grammeme + *май*’. This means that the number in the first cell indicates the frequency of the cases where the word *май* ‘probably, it seems’ immediately follows or precedes the auxiliary-less past participle. In Estonian, the most frequent collocation is ‘evidential grammeme + *vist*’. Thus, the number in the second cell of the first column indicates cases where the word *vist* ‘perhaps, possibly’ immediately follows or precedes the evidential *vat*-form. The second column contains the respective numbers for the second most frequent collocations; in Bulgarian this is *вероятно* ‘most likely’ preceding or following the auxiliary-less past participle, and in Estonian *ehk* ‘maybe, perhaps’ preceding or following the evidential *vat*-form.

Table 2. **Comparison of the frequencies of the two most common collocations of evidential grammeme and epistemic (or evidential) word in Bulgarian and Estonian**

	the most frequent collocation	the second most frequent collocation
Bulgarian	(ev. grammeme + <i>май</i>) 183	(ev. grammeme + <i>вероятно</i>) 82
Estonian	(ev. grammeme + <i>vist</i>) 161	(ev. grammeme + <i>ehk</i>) 14

Interactions between grammatical and lexical markers

The distribution in the table is significant (chi-square=31.19425, p=.000)⁸ showing that the lexical items *май* ‘probably, it seems’ in Bulgarian and *vist* ‘perhaps, possibly’ in Estonian are much more preferable in collocation with the relevant grammatical markers of evidentiality than any of the remaining lexical items. This, of course, applies only in case we assume that the overall frequency of these lexical items is similar. A Google search, which shows only approximate numbers, confirms that this assumption is more or less correct. The search for *май* yields approximately 1 680 000 Google hits and the search for *вероятно* approximately 1 420 000 hits. In Estonian, *vist* shows approximately 1 740 000 hits, and *ehk* approximately 1 500 000 hits. Thus, despite the similarity in the token frequencies of the Bulgarian *май* and *вероятно*, the first occurs over twice more often than the second in collocation with the grammatical evidential. In Estonian, this tendency is even more striking. Despite the similarity in token frequency of *vist* and *ehk*, the first occurs over eleven times more frequently in collocation with the grammatical gram than the second. The only conclusion that forces itself upon us is that the grammatical evidential is sensitive to a particular word, which expresses medium certainty and thus generates redundancy.

Is it interesting that native speakers of Bulgarian and Estonian often consider sentences containing combinations of the grammatical marker of evidentiality and a word expressing middle certainty somewhat overburdened and redundant if such sentences are out of the context, but if they are surrounded by the original context their acceptability to the speakers increases substantially, and the characterization of ‘redundancy’ is replaced with something like ‘motivated reinforcement’. This granted we can draw a parallel with the reinforcement occurring within the paradigms of grammatical evidentiality. The paradigm of the Bulgarian evidential is composed by aorist or imperfect past participles which can occur with or without the past participle of the auxiliary verb. The past participle of the auxiliary usually denotes criticism and distrust on behalf of the speaker (see [Demina 1959, 323; GBE II, 360]). Therefore some studies (such as [Nitsolova 2006]) postulate a separate paradigm of dubitative forms which is to be distinguished from the paradigm of the ‘renarrative’ forms. Compare the evidential renarrative form in (19a) with the evidential dubitative in (19b):

⁸ The calculation tool uses Yates’ correction for continuity, which reduces the magnitude of the difference between expected and observed frequencies by 0.5.

- (19) a. *мой казал*
he say.PST.PTCP
‘Reportedly, he **said** ...’
- b. *мой бил казал*
he be.PST.PTCP say.PST.PTCP
‘Reportedly (but I doubt it), he **said** ...’

Due to its function to increase the distance between the speaker and the event referred to by the proposition, the participle of the auxiliary can be characterized here as a ‘distance particle’, a term originally used by Johanson (see Johanson 1998, 146). Such distance particles enhance the epistemic component in meaning of the compound.

Consider now an analogous case in Estonian. Besides the dedicated marker of evidentiality *-vat*, Estonian has the multifunctional modal verb *pidama* (with premodal meaning ‘to hold’) which in addition to dynamic, deontic and epistemic necessity is used as a marker of evidentiality. The dedicated marker *-vat* is often suffixed to this verb, which enhances the sense of doubt; witness (20) adopted from Kehayov (2002, 136). In (20a), where the suffix *-vat* is the only marker of evidentiality, we are dealing with a typical case of reported evidentiality, which may, but need not, imply a sense of doubt. In (20b) *-vat* is suffixed to the verb *pidama* and the cooperative meaning of the whole verb form is that of report accompanied by a stronger sense of doubt.

- (20) a. *Ta ole-vat Tallinna-s.*
s/he be-EV Tallinn-INE
‘Reportedly, s/he is in Tallinn’
- b. *Ta pida-vat Tallinna-s olema.*
s/he must-EV Tallinn-INE be-SUP
‘Reportedly (but I do not subscribe to this view), s/he is in Tallinn.’

Now, what is common between the cases in (19) and (20) and the combinations of grammatical and lexical items with concord interpretation is the semantic effect of reinforcement. The difference, on the other hand, between these cases is that in (19) and (20) we are dealing with a grammaticalized means of reinforcement whereas in the combinations of grammatical and lexical items with concord reading the relationship between the two elements is not grammaticalized.

Interactions between grammatical and lexical markers

It was stressed on several occasions that there is an implicational relation between the meanings of evidentiality and epistemicity as the cognitive remoteness of the source of information correlates with the degree of certainty. Despite this, only one of these meanings is considered a focal meaning (cf. Wiemer 2006, who draws a distinction between focal and associated meanings in his analysis of Polish lexical evidentials). Therefore, we can claim that in a complex sequence of two items with similar meanings, a certain meaning (evidential or epistemic) can be focalized both by grammatical or non-grammatical structural means. The only reason why the concord interpretation does not seem so obvious in cases where grammatical and lexical items are combined is that it is not overtly marked in the morphosyntax of the language.

7. Some consequences of general relevance

The effect of reinforcement can be compared with cases of reduplication where the property denoted by the repeated word is enhanced. If someone says *good good dog* it normally means that the dog is very good, i.e. the concept of 'goodness' is reinforced. We saw that certain combinations of 'evidential' and 'epistemic' yield an increase of doubt in the truth of the proposition, in which case we may say that the concept of 'doubt' is reinforced. Although this parallel might seem speculative, both examples involve reinforcement of a term. The very existence of such effect leads us again to the question whether we should look for an umbrella term for evidentiality and epistemicity. Such an umbrella term would be a narrowly defined functional category which encompasses only two subcategories: the one of evidentiality and the one of epistemicity. In such a way, the concord analysis provides another piece of evidence for the conceptual affinity of evidentiality and epistemic modality. Bulgarian and Estonian grammatical evidentiality systems are among the first discovered and best described evidentiality systems in the world. Moreover, due to the work of Jakobson (Jakobson 1971), the Bulgarian evidentiality system has played an important role in establishing the cross-linguistic category of evidentiality. It is thus beyond doubt that the cognitive basis that warranted an establishment of a new grammatical category distinct from modality in these languages is firm enough. Nevertheless, the surprising number of cases where the grammatical evidential and an epistemic word 'reinforce' a common meaning component urge us to reconsider whether even these languages do not warrant an umbrella term for the notions of evidentiality and epistemicity from which this meaning component can be abstracted.

The second important consequence emerges as we look at the size of the area of overlap of the linguistic elements which are said to be evidential and

Interactions between grammatical and lexical markers

expressions than a smaller region or by the same number of linguistic expressions with more general meanings. Both possibilities have frequential effect. In our case, the chance that the overlapping area is encoded linguistically is greater in Bulgarian than in Estonian.

Unfortunately, we cannot test these deductively achieved claims as we do not have comparable corpora for Estonian and Bulgarian. Nonetheless, these generalizations are significant as they seem to hold for any two languages with grammatical evidentials.

The evidence from Bulgarian and Estonian that we have looked at so far can be helpful if we want to elaborate a waterproof method with which we can identify the exact functional range of any grammatical evidential in any language. We will take for granted that any two markers with cognitively adjacent and/or partly overlapping meanings are subjected to specific restrictions as for their co-occurrence in the same sentence. Furthermore, we will adopt the assumption that if the co-occurrence of these two markers in a single sentence is accepted, there is still another set of restrictions which govern the semantic effects induced by such co-occurrence. Suppose we study the functions of a certain marker in certain language. If we find out what the relevant restrictions are we would be able to determine precisely the functional boundaries of this marker and to better locate it in semantic space. Say we study the functions of certain grammatical marker of evidentiality. Combining this marker with different items expressing epistemic modality helps to determine its functional range. In light of the above evidence, the following aspects should be taken into consideration:

- 1) Does the combination of the grammatical evidential with epistemic items generate concord readings?
- 2) If it does, with which particular epistemic items does this happen?
- 3) What is the position of these epistemic items on the scale of certainty?
- 4) Are there any epistemic items which particularly often enter into a concord relation with the grammatical evidential?

The fieldworker's guides advanced in Kozintseva (1994) and Aikhenvald (2004, 385-390) consider the compatibility of evidential and modal markers a relevant criterion for determining the type of evidential coding. The questions above could be considered as a supplement to these guides. By answering these four questions we could test the functional boundaries of the grammatical evidential of any random language. The last question is crucial as for whether a particular combination of evidential and epistemic item is in process of becoming conventionalized. The concord readings of the collocations of Bulgarian and Estonian grammatical evidentials with the items *май* 'probably, it seems' and *vist* 'perhaps, possibly', respectively, form the majority of cases where the collocation of grammatical evidential and epistemic or evidential word has

concord reading in these languages.⁹ Moreover, as we saw in Table 2, these collocations make up a considerable share of the total amount of attested co-occurrences of grammatical evidential and epistemic or evidential word. It seems therefore that we are dealing with sensitivity between the grammatical evidential and a certain epistemic word, which might reflect an early stage of conventionalization of such complex expressions. An advanced stage of such conventionalization, on the other hand, would be a situation where their co-occurrence has become obligatory. Boye reports for a good number of cases from different languages where two epistemic items or constructions with overlapping meaning co-occur obligatorily in a unified qualificational expression (Boye 2006, 78-80, 189-191) and, following the postulates of the grammaticalization theory, we could assume that such expressions originate in non-obligatory syntagmatic patterns.

As a final point, it should be noted that the co-occurrences of grammatical and lexical expressions that we studied are surprisingly common compared to the co-occurrences of two grammatical or two lexical expressions of the relevant categories. We saw in Section 2 that there are a number of studies concerned with combinations of lexical or grammatical markers of epistemicity and/or evidentiality. We focused on the third possibility, namely on co-occurrences of grammatical and lexical marker. This choice turned out to be successful as we easily managed to gather a considerable body of examples. In contrast, both in Bulgarian and Estonian, combinations of grammatical evidentials with ‘epistemic’ moods, such as the conditional mood, are ungrammatical.¹⁰ The compatibility of two lexical means of evidentiality and epistemicity in Bulgarian and Estonian has not yet been studied, but the intuition is that many of the possible combinations are not acceptable. It seems therefore that if evidential and epistemic modifiers are combined at different levels of linguistic expression (lexicon and morphology, for example), they are more acceptable than if they are combined at the same level of linguistic expression. This pattern might be due to some general principle which blocks redundancies at the same level of grammar, but allows them at different levels.

⁹ I will refrain from presenting exact frequencies, because many of the examples of concord reading might be considered ambiguous, also allowing for other readings and which may therefore be considered controversial among speakers. It is clear, however, that they form the majority of all cases of concord reading.

¹⁰ Except from the short conditional forms in Bulgarian, which permit evidential marking (see GBE II, 370 for examples), but which are obsolete by now. These evidential conditional forms were in many cases homonymic with the relevant imperfective past participles and therefore do not qualify as adequate examples.

8. Conclusions

In this contribution, I outlined the types of interaction between grammatical markers of evidentiality and lexical markers of epistemicity and evidentiality. These types were stipulated based only on Bulgarian and Estonian data, but were assumed to be cross-linguistically pertinent. The following four types of interactions were distinguished:

- 1) The grammatical marker of evidentiality outscopes the epistemic (or evidential) word.
- 2) The epistemic (or evidential) word outscopes the grammatical marker of evidentiality.
- 3) The two items are not in scopal dependency, but represent two independent qualifications of the status of the proposition;
- 4) The two items are understood as parts of a single entity, which ‘reinforces’ a common meaning component.

This fourth type was referred to as ‘concord (or holistic) reading’ of the sequence of the grammatical marker of evidentiality and an epistemic (or evidential) word. The domain of medium certainty as well as the absence of specific referent of the report were shown to trigger the concord reading.

The central claim of this study was that the possibility for a concord reading of such sequences should be seriously considered in the description of the evidential and/or modal system of any language. It was argued that if one wants to determine the array of meanings expressed by a given grammatical evidential, the possibility for concord readings, the regularity of these readings and the range of the specific semantic effects should be explanatory as for the exact functions of this grammeme.

Abbreviations

ADE – adessive case, ADV – adverb, AOR – aorist, ATTR – attributive adjective, COM – comitative case, DAT – dative case, DEF – definite article, ENCL – enclitic, EP – epistemic: a conventional term in the descriptive grammar, _{EP} – notionally epistemic, EV – evidential: a conventional term in the descriptive grammar, _{EV} – notionally evidential, FUT – future tense, GEN – genitive case, HOR – hortative, ILL – illative case, IMP – imperative, IMPF – imperfect (tense), IMPS – impersonal form, INE – inessive case, INF.COMP – infinitival complement, INTERJ – interjection, NEG – negative, *p* – proposition, PASS – passive, PL – plural, PRON – pronoun, PST – past tense, PTCP – participle, Q – question marker, REFL – reflexive, SG – singular, SUP – supine, 1 – first person, 2 – second person, 3 – third person

References

- Aikhenvald, A. Y. 2003a. "Evidentiality in typological perspective", Aikhenvald, A. Y., Dixon, R. M. W. (eds.), *Studies in Evidentiality*, Amsterdam: John Benjamins, 1-31.
- Aikhenvald, A.Y. 2004. *Evidentiality*, Oxford: Oxford University Press.
- Akatsuka, N. 1985. "Conditionals and the epistemic scale", *Language* 61, 1985/3, 625-639.
- Alexiadou, A. 1997. *Adverb Placement. A Case Study in Anty-Symmetric Syntax*, Amsterdam, Philadelphia: John Benjamins.
- van der Auwera, J., Plungian, V. 1998. "Modality's semantic map", *Linguistic Typology* 2, 1998/1, 79-124.
- Boye, K. 2006. *Epistemic meaning: a cross-linguistic study*. Ph.D. dissertation. Department of English, Germanic and Romance Studies, University of Copenhagen. [unpublished manuscript]
- Brinton, L.J., Traugott, E.C. 2005. *Lexicalization and Language Change*, Cambridge: Cambridge University Press.
- Bybee, J.L. 1985. *Morphology. A Study of the Relation between Meaning and Form*, Amsterdam: John Benjamins.
- Chafe, W. 1986. "Evidentiality in English Conversation and Academic Writing", Chafe, W., Nichols, J. (eds.), *Evidentiality: The Linguistic Coding of Epistemology*, Norwood: Ablex, 261-271.
- Cinque, G. 1999. *Adverbs and Functional Heads. A Cross-Linguistic Perspective*, New York, Oxford: Oxford University Press.
- Cinque, G. 2006. *Restructuring and Functional Heads: The Cartography of Syntactic Structures*, 4, New York: Oxford University Press.
- Dendale, P., Tasmowski, L. 2001. "Introduction: Evidentiality and related notions", *Journal of Pragmatics* 33, 2001/3, Amsterdam: Elsevier, 339-348.
- EKG II = *Eesti keele grammatika II. Süntaks. Lisa: kiri*, Erelt, M., Kasik, R., Metslang, H., Rajandi, H., Ross, K., Saari, H., Tael, K., Vare, S. (eds.), Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut, Tallinn, 1993.
- Ernst, Th. 2006. *The Syntax of Adjuncts*, Cambridge: Cambridge University Press.
- Geurts, B., Huitink, J. 2006. "Modal Concord", Dekker, P., Zeijlstra, H. (eds.), *Proceedings of the ESSLLI 06 Workshop Concord Phenomena at the Syntax-Semantics Interface*, Málaga: University of Málaga, 15-20.
- Givón, T. 1982. "Evidentiality and epistemic space", *Studies in Language* 6, 1982/1, 23-49.
- De Haan, F. 2005. "Encoding speaker perspective: Evidentials", Frajzyngier, Z., Hodges, A., Rood, D. S. (eds.), *Linguistic Diversity and Language Theories*. Amsterdam: John Benjamins, 379-397.
- Halliday, M.A.K. 1970. "Functional diversity in language as seen from a consideration of mood and modality in English", *Foundations of Language* 6, 1970/3, 322-361.
- Hengeveld, K. 1989. "Layers and operators in functional grammar." *Journal of Linguistics* 25, 1989/1, 127-157.
- Hopper, P.J., Traugott, E. C. 2003. *Grammaticalization*, Cambridge: Cambridge University Press.
- Ifantidou, E. 2001. *Evidentials and Relevance*, Amsterdam: John Benjamins.
- Jakobson, R. 1971: "Shifters, Verbal Categories, and the Russian Verb", Jakobson, R. (ed.), *Selected writings, Vol. 2: Word and Language*, The Hague: Mouton, 131-147.

Interactions between grammatical and lexical markers

- Johanson, L. 1998. "Zum Kontakteinfluß türkischer Indirektive", Demir, N., Taube, E. (eds.), *Turkologie heute – Tradition und Perspektive. Materialien der dritten Deutschen Turkologen-Konferenz, Leipzig, 4.-7. Oktober 1994*, Wiesbaden: Harrassowitz, 141–150.
- Kehayov, P. 2002. "Typology of Grammaticalized Evidentiality in Bulgarian and Estonian", *Linguistica Uralica* 2002/2, 126-144.
- Kiefer, F. 1994. "Modality", Asher, R. E. (ed.), *The Encyclopaedia of language and linguistics*, Oxford: Pergamon Press, 2515-2520.
- Lehmann, Chr. 2002a. *Thoughts on grammaticalization*, Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt 9.
- Lehmann, Chr. 2002b. "New reflections on grammaticalization and lexicalization", Wischer, I., Diewald, G. (eds.), *New reflections on grammaticalization*, Amsterdam, Philadelphia: John Benjamins, 1-18.
- Lyons, J. 1977. *Semantics. Vol. 2*, Cambridge: Cambridge University Press.
- Makarsev = Макарецев, М. М. "Связь лексических и грамматических показателей эвиденциальности в болгарском языке", this volume.
- Mushin, I. 2001. *Evidentiality and Epistemological Stance. Narrative Retelling*, Amsterdam: John Benjamins.
- Nuyts, J. 2001. *Epistemic modality, language and conceptualization. A cognitive-pragmatic perspective*, Amsterdam: John Benjamins.
- Nuyts, J. (forthcoming). "The 'one-commitment-per-clause' principle and the cognitive status of qualificational categories", *Linguistics*.
- Palmer, F. 2001. *Mood and Modality*. Cambridge: Cambridge University Press.
- Plungian, V. 2001. "The Place of Evidentiality within the Universal Grammatical Space", *Journal of Pragmatics* 33, 2001/3, 349-357.
- Quine, W.V.O. 1953. *From a Logical Point of View*, Harvard: Harvard University Press.
- Svenonius, P. 2001. "Subject Positions and the Placement of Adverbials", Svenonius, P. (ed.), *Subject, Expletives and the EPP*, New York: Oxford University Press, 201-242.
- Van Valin, R.D., LaPolla R.J. 1997. *Syntax: Structure, meaning and function*, Cambridge: Cambridge University Press.
- Van Valin, R.D. 2005. *Exploring the Syntax-Semantics Interface*, Cambridge: Cambridge University Press.
- Wiemer, B. 2006. "Particles, parentheticals, conjunctions and prepositions as evidentiality markers in contemporary Polish. (A first exploratory study.)", *Studies in Polish Linguistics* 3, 1-52.
- Демина, Е.И. 1959. "Пересказывательные формы в современном болгарском литературном языке", Бернштейн, С. Б., (ред.), *Вопросы грамматики болгарского литературного языка*, Москва, 313-378.
- ГВЕ II = *Грамматика на съвременния български книжовен език. Том 2/ Морфология*, Тилков, Д., Стоянов, С., Попов, К. (ред.), Българска Академия на Науките, София 1983.
- Козинцева, Н.А. 1994. "Категория эвиденциальности (проблемы типологического анализа)", *Вопросы языкознания* 1994/3, 92-104.
- Ницолова, Р. 2006. "Взаимодействие эвиденциальности и адмиративности с категориями времени и лица глагола в болгарском языке", *Вопросы языкознания* 2006/4, 27-45.

Petar Kehayov

Резюме

Взаимодействие между граматически показатели за евиденциалност и лексически средства за изразяване на епистемична модалност и евиденциалност в български и естонски

Статията разглежда изречения като *Той май бил от Бургас*, които се характеризират със съчетаване на лексема маркираща епистемична модалност (или евиденциалност) от една страна и граматически показател за евиденциалност от друга. Задачата на статията е да очертае основните видове взаимодействие между двата типа показатели в подобни съчетания. Въпреки че въпросните видове са формулирани въз основа на данни от български и естонски, може да се предположи, че те са валидни в универсален план. Предложената типология обхваща четири вида взаимодействие:

1. Лексемата предаваща епистемична модалност или евиденциалност е в обсега на действие на граматическия показател за евиденциалност.
2. Граматическият показател за евиденциалност е в обсега на действие на лексемата предаваща епистемична модалност или евиденциалност.
3. Двата показателя не се намират в отношение на подчиненост по признак семантичен обсега, а се тълкуват като два независими оператора санкциониращи отделно статуса на пропозицията.
4. Двата показателя се тълкуват като части от едно цяло, чиято цел е да подсили общ семантичен компонент във функциите им.

Последният вид е означен като 'съгласувано (или холистично) тълкуване' на комбинацията от граматически показател за евиденциалност и лексема маркираща епистемична модалност (или евиденциалност). Основният извод на статията е, че възможността за подобно холистично тълкуване на въпросните комбинации в даден език е индикатор за обхвата на граматическата евиденциална система на този език. Главните предпоставки за подобни холистични тълкувания са: а) лексемата да изразява резервираност (или непълна сигурност) от страна на говорещия по отношение на истинността на предадената информация; б) крайният източник на информацията да не е контекстуално маркиран.

CONCLUSIONS AND PERSPECTIVES FOR FURTHER RESEARCH

The development of grammatical evidentiality systems is a feature shared by both linguistic areas considered in this dissertation. There is a difference, however, in the linguistic significance that evidentiality has had for the establishment of these *Sprachbünde*. As noted in Section 3.1, the Balkan hotbed of evidentiality is just a piece of much larger Eurasian pattern, and therefore evidentiality cannot be taken as an emblematic feature of the Balkan linguistic area (see Appendix 2). The Baltic evidentiality hotbed, on the other hand, is cut off from the large Eurasian evidential area and therefore seems to be more closely related to the convergence processes specific to the restricted area where it is found (the eastern coast of the Baltic Sea). It was stimulating, however, to study the relationship of this separate area with the larger Eurasian pattern and to verify whether the geographical gap corresponds to a typological gap, or, in other words, whether the fact that these areas are not contiguous geographically is reflected in them being typologically non-contiguous. No previous study has addressed this question, which makes the present study highly relevant.

Of course, the scope of the research programme was considerably narrowed. Two articles (II and III) compared seven evidentiality systems, three from languages of the Baltic area and four from the Balkan area. The other two articles (I and IV) dealt with a comparison of the Estonian system, which is found at the geographical extreme of Baltic evidentiality zone, and the Bulgarian system, which can be viewed as a prototypical member of the Balkan area of evidentiality. In Section 4.3., I also discussed the possible connections between the coding of evidentiality with the past participle in Estonian and tense system. This is an important question because the historical connection between evidentiality coding and the tense system is still fairly transparent in many languages in the Balkans, Asia Minor and Caucasus (this is visible in de Haan's Map 78; see Haspelmath, Dryer, Gil and Comrie 2005) and could give us an important clue with regard to the relationship of the Estonian evidentiality system with the Eurasian evidential area.

The following concluding remarks do not only comprise a summary of articles, but also partial revision of the results of articles II and III. I will start with the results of the comparison of the evidentiality systems under concern, and then proceed with the relationship between evidentiality and epistemic modality, arriving finally at the conclusions concerning the historical development of evidential markers.

Results of the comparison of evidentiality systems

In articles II and III, the Turkish, Bulgarian, Macedonian, Albanian, Lithuanian, Latvian and Estonian evidentiality systems were compared in relation to 32 features. In Section 4.1., these 32 features were changed to a restricted set of 19 better defined and typologically relevant features. Some of the old features were left unchanged, some were redefined, some were collapsed into a single feature and some excluded.

Tables 1–3 in Appendix 3 show the value assignments for each evidentiality system under consideration. Comparing the values from Tables 1–2 in article II and Tables 1–2 in article III with the values from the tables in Appendix 3, one can see that there are no significant reassignments. Most of the reassignments are due to recapitulation of the value (–) into either (+) or –, which actually means that these two values were expanded to encompass a larger choice of cases. In addition, in some cases a (+) value is changed to +. Unfortunately, there is no strict procedure which would assign the right value to occurrences of various significance (in terms of structural weight, frequency or other), and a certain degree of arbitrariness in choosing between values is inevitable. In some cases there seem to be considerable changes, but this is only due to the narrowing of the definition of the relevant parameter; compare for example the values concerning the co-occurrences of evidential and mood markers in the old and in the new parameter set (cf. Table 1 in the end of article III with Table 3 in Appendix 3). All these reassignments are due to new evidence and additional deliberation. Crucially though, except for the cases with changed definitions, there are no disruptive reassignments on the scale of values +, (+), (–), –. In other words, there are no cases in which + or (+) are changed to – or vice versa. This fact alone shows that no significant relocations in typological distances between the initial and the revised parameter sample could have been expected.

Table 1 in Appendix 4 presents the typological distances between each two languages in the sample. The calculation is based on number of shared values. As already noted in Section 1.2., I draw the boundary between considering a feature present or not present between values (+) and – and the count in Table 1 (Appendix 4) is based on such a binary value-system. The highest number in the table is 17. This number of shared features is reached by the pairs Turkish–Bulgarian, Turkish–Macedonian and Bulgarian–Macedonian. This result is not surprising. Bulgarian and Macedonian are close relatives, which only 100 years ago were considered to be dialects of a single language, and which have both gone through intense contact with Turkish during the Ottoman rule in the Balkans. The lowest attested figure (7), on the other hand, is instantiated by the pairs Estonian–Turkish, Estonian–Macedonian and Latvian–Albanian. This result is more interesting, because it seems to outline some kind of typological radicals in the sample. Nevertheless, it is trivial as it roughly correlates with the geographical distances between these languages.

Consider now Figures 1–7 in Appendix 4. In these figures each total of shared features represented in Table 1 is illustrated as a specific line. As noted in Section 1.2., these lines are called isopleths and the relevant figure is called an isopleth map. The position of a language on this map roughly corresponds to its relative geographical position on the map of Europe (where Estonian, Latvian, Lithuanian and Bulgarian, for example, line up along the 24th meridian east of Greenwich). Each map illustrates isopleth clustering around a certain language. The distribution of isopleths shows the degree of similarity of the evidentiality systems of other languages to the system of this certain language. The first map shows the attested levels of similarity to the Turkish system. The innermost isopleth line enclosing Turkish, Bulgarian and Macedonian shows that the latter two languages share 17 values with Turkish. The next level of similarity is reached by Lithuanian sharing 15 values with Turkish. Further on, at the level of 12 shared values, Albanian enters the picture. The outermost isopleth lines add Latvian with 8, and Estonian with 7 values common with Turkish. Figure 2 shows respective clustering based on Bulgarian, Figure 3 on Macedonian, Figure 4 on Albanian, Figure 5 on Lithuanian, Figure 6 on Latvian and Figure 7 on Estonian.

The most interesting findings in these two-dimensional representations, are first, that the Lithuanian evidentiality system turns out to be more similar to Bulgarian, Macedonian and Turkish systems than to the system of its neighbour and close relative Latvian, and second, that the evidentiality system of Albanian stands clearly apart from the group Turkish–Bulgarian–Macedonian. Going back to the tables in Appendix 3, we see that the unexpected similarity of Lithuanian to Balkan Slavic and Turkish on the one hand, and the unexpected remoteness of Albanian from the latter two is not due to a mismatch of values only in a certain domain of parameters. The tables show clearly that both types of parameters, the formal and the semantic ones, contribute to this result. For example, Lithuanian shares with Bulgarian 5 formal values, but only 4 with Latvian. With regard to semantics (and pragmatics), Lithuanian shares with Bulgarian 9 values, but only 5 with Latvian.

Other, minor discoveries may seem interesting, but they are not significant enough to warrant sweeping generalizations. The scores in Table 1 (Appendix 4) and the resultant isopleths show for example that the Estonian system is closer by one point to the Albanian and Bulgarian systems than Latvian, despite the fact that on genetic and geographical grounds Latvian should be expected to show more common features with these two languages. Along similar lines, Albanian has two more values in common with Bulgarian than with its neighbour Macedonian. These differences, however, are so small that may well be observational artefacts, or in other words distortions due to unbalanced feature selection or feature definition.

All in all, on the basis of Figures 1–7 we may outline two distinct groups in which languages share at least 15 features with each other. These groups, shown

in Figure 8 (Appendix 4) are 1) Turkish–Bulgarian–Macedonian–Lithuanian, and 2) Latvian–Estonian. The distribution of isopleths is so clear-cut that these clusters cannot be incidental. Despite the great distance in space, Turkish, Bulgarian and Macedonian share with Lithuanian 15 values, while Macedonian shares only 12 values with its neighbour Albanian (which therefore remains outside the group). In a similar way, Latvian and Estonian share 16 values, but only 10 and 9 values respectively with the adjacent Lithuanian.

Although the cluster comprising Turkish, Bulgarian, Macedonian and Lithuanian is an interesting discovery, one should keep in mind that different pairs of languages in this group might not share exactly the same features. The pairs Bulgarian–Turkish and Bulgarian–Macedonian, for example, both score 17 shared values, but as can be seen from tables 1–3 in Appendix 3, not exactly for the same parameters. Therefore the question arises of which parameters have the same values in all languages of the group Turkish–Bulgarian–Macedonian–Lithuanian, or in other words, what is the standard of this similarity cluster. Judging from Tables 1–3 in Appendix 3, we can list the following characteristics of this standard:

- 1) evidentiality coding based on past participles;
- 2) infinitives or other nominalizations (except for participles) are not employed as markers of evidentiality;
- 3) auxiliary ellipsis creates contrast with the tense system;
- 4) weak dedication of evidentiality markers;
- 5) the participial form of the existential copula verb is employed as a distance particle;
- 6) reported, inferential and mirative meanings are integrated into a broad non-firsthand term;
- 7) the existence of strategies to specify reported subtype of evidentiality;
- 8) the non-firsthand term does not encode generic statements;
- 9) the non-firsthand term can be used in reports of dreams;
- 10) the non-firsthand term is not used as objectivizer of sensory perceptions and feelings;
- 11) the non-firsthand term is conventionalized in traditional narratives;
- 12) the non-firsthand term can be used in direct non-echoic questions.

It is beyond doubt that many of these characteristics also constitute the standard of the extended Eurasian evidentiality area (cf. for example Friedman 1988a; DeLancey 2001; Aikhenvald 2004: 27–31). The above results, supplemented by what is known from the previous studies, show that the eastern European evidentiality area comprising Turkish, Bulgarian, Macedonian and Lithuanian adhere to the type of system that has been broadly labelled ‘evidential perfect’ (see Nichols 1986: 253), whereas the Latvian, Estonian and Albanian evidentiality systems should rather be accounted for in terms of geographically restricted innovation.

These stipulations lead to a rather interesting consequence for areal linguistics. Along with other categories, grammatical evidentiality may play an important role for delimiting the Balkan and Baltic *Sprachbünde*, but this holds only as long as we do not delve deeper into the properties of the category in the different languages. Such a deeper look reveals underlying areas of affinity (see again Figure 8 in Appendix 4) which are rather different from the areas postulated on the basis of superficial comparison of grammatical systems. The type of evidentiality system instantiated by Lithuanian, Bulgarian, Macedonian and Turkish seems to drift smoothly into Asia Minor, Caucasus and Central Asia wherein an increase in geographical distance leads to gradual type-alternation with no abrupt changes. The outlined similarity cluster need not be due to language contact, but to more general processing principles. What is striking about Lithuanian, Bulgarian and Macedonian is that Lithuanian has the most complex tense-aspect system among the existent Baltic languages whilst Bulgarian and Macedonian have the most complex tense-aspect systems in Slavic. It seems thus, that there is a straightforward correlation between having a complex tense-aspect system and developing the type of evidentiality system outlined in the standard above.

The above results also provide indirect evidence against the claim that the Bulgarian and Macedonian evidentiality systems are structural borrowings from Turkish. The Lithuanian evidentiality system is also very similar to the Turkish system despite the lack of direct contact between Turkish and Lithuanian.

The parameter values in Tables 1–3 (Appendix 3) elicit also several implications between features. Here are the most striking of them:

- a) If a language uses nominalizations other than participles as markers of evidentiality, it also uses present participles as markers of evidentiality, and if it uses present participles as markers of evidentiality, it also uses past resultative participles as markers of evidentiality. In other words, there is a hierarchy ‘past participle’ > ‘present participle’ > ‘other nominalization’.
- b) If a term covers the expression of inferentiality and mirativity in a two-term system, it always covers also the expression of reported evidentiality.
- c) A language may have devices for encoding a tripartite distinction with a functionally unmarked member only if it has an evidentiality system with a broad non-firsthand (or indirective) term.
- d) If a language has an evidentiality system with a broad non-firsthand term, it will probably have the formal means to further distinguish reported evidentiality from it.
- e) A formally marked evidential term can be used for reports of dreams only if the language has a broad non-firsthand system.

- f) The necessary condition for conventionalization of a formally marked evidential term in a system with two choices in traditional narratives is its use to mark reports, not its inferential or mirative uses.
- g) A formally marked term in a system with two choices can be used in direct non-echoic questions only if the language has a broad non-firsthand system.
- h) Distance particles derived from copula verbs are used only in languages with broad two-term systems (i.e. in languages with non-firsthand systems).

Based exclusively on the data in Tables 1–3 (Appendix 3) we could state that implications (a)–(d) are unidirectional, i.e. they are not valid the other way round, whereas implications (e)–(h) seem to be bidirectional, i.e. they do hold even if we switch the contents of the antecedent and the consequence. Let us now elaborate on the soundness of each implication against a wider linguistic background. Aikhenvald’s book (2004) is an excellent source for such an assessment.

The first part of the first implication (a) is violated by some Estonian dialects in which past participles and infinitives are used as evidentiality markers, but not present participles (see Kask 1984: 254–270). At the same time, I am not aware of any evidence against the implication ‘from past resultative participles to present participles’. This implication conforms to the view according to which the past time reference is the locus of the cross-linguistic category of evidentiality (see for example Wälchli 2000). In more abstract terms, this implication says that evidentiality marking with present (or imperfective) nominalizations presupposes evidentiality marking with past (or perfective) nominalizations. Future research should investigate whether this implication is an absolute universal or not. The second implication (b) is violated by examples from languages like Nepali and Jarawara (see Aikhenvald 2004: 156). Although I do not know of any apparent cases against (c), this implication is probably too specific to be typologically contributive. Implication (d) is in turn very interesting and deserves further assessment; so far I am not aware of any evidence going against it (see Aikhenvald 2004: 155–156 for supportive evidence). Implication (e), which applies only to two-term systems, is interesting and cross-linguistically contributive, although it should be weakened into a unidirectional (cf. Aikhenvald 2004: 344–347), and even in this case it might be violated by the situation in Jarawara (Aikhenvald 2004: 23–24, 345, referring to Dixon 2004). Similarly, implication (f) seems to survive cross-linguistic comparison (cf. Aikhenvald 2004: 310–315), but only in its unidirectional version. Implication (g) does not withstand cross-linguistic comparison due to evidence from Sochiapan Chinantec (Aikhenvald 2004: 249, 291) and Eucree (Aikhenvald 2004: 249), and the last implication (h) seem valid (i.e. no counter-examples can be found), but is probably too specific to be typologically challenging.

Summing up, the only implications that deserve further typological scrutiny are the second half of (a) ('past participle' > 'present participle') and implications (d), (e) and (f). This is a topic for future research.

Finally, Tables 1–2 (Appendix 3) reveal an intriguing regularity. The tables show that languages (such as Latvian and Estonian) that are rich in grammatical material encoding evidentiality (employing several participles and nominalizations) are poor in evidential functions, i.e. their marked evidential terms have rather narrow meaning ('reported'). On the other hand, the languages that are poor in grammatical material encoding evidentiality distinctions (such as the Balkan languages) are rich in evidential functions, i.e. their marked evidential terms have broader meaning ('non-firsthand' or 'indirect'). This pattern remains unexplained in this dissertation. Future investigation taking into account also other structural characteristics of the language at hand may be able to explain it.

The relationship between evidentiality and epistemic modality

In the last article (IV) I sketched out the types of interaction between grammatical markers of evidentiality and lexical markers of epistemicity and evidentiality. These types were based only on Bulgarian and Estonian data, but were assumed to be cross-linguistically valid.

My basic claim was that there is a conceptual mapping between the point of middle certainty on the epistemic scale and the non-firsthand evidentiality. This is reflected in the languages by the possibility for concord readings of sequences of evidential and epistemic markers. In some cases such readings seem to conventionalize which is a straightforward evidence for an overlap between the evidential and epistemic space. In Bulgarian, for example, the concord interpretations of the sequence of the word *maü* 'probably, as it seems' and the evidential grammeme (auxiliary-less past participle) are very common. This means that even in a language which has played a definitional role in the formulation of the cross-linguistic category of evidentiality, epistemic modality and evidentiality interfere in a way which cannot be expected for unrelated categories.

These conclusions corroborate in general the results of Boye's recent study, which subsumes evidentiality and epistemic modality under a unified descriptive category of epistemicity (Boye 2006). The meaning domain of epistemicity is complex, as it comprises two separate subdomains: 1) 'epistemic-support meaning domain', and 2) 'evidential meaning domain'. On the other hand, the meaning domain of epistemicity is inherently coherent, because all specific meanings it comprises can be narrowly defined in terms of 'justificatory support'.

A promising topic for further research is to look into the scope restrictions of evidentiality in conditionals. Analysing the behaviour of certain evidentials in Japanese, McCready and Ogata (2007) show that they can be embedded in conditionals and under modals. These facts are problematic for those theories of evidentials which assign evidentials widest possible scope, and indicate that such expressions indicating source of evidence are part of the propositional content and should be best analyzed as a special kind of epistemic modal.

The historical development of evidentiality markers

Section 4.3. discussed the origins of Estonian evidentiality coding based on participles. Whereas the problem about the sources of the present participles marking evidentiality can be considered solved by the earlier research, and thus leaves little space for further investigation, the problem about the origin of the past evidential participles remains open. In Section 4.3., I distinguished between three scenarios for the development of the past participles into markers of evidentiality: i) de-subordination of speech (or mental state) complements, ii) auxiliary ellipsis in compound past tense, and iii) retention of the ancient use of participles as predicates. I also presented results from an independent study by Kehayov and Siegl (2007), which discarded the third scenario. With respect to the first two scenarios, I presented positive evidence for both of them. Supportive evidence for the first scenario comes from the fact that past participles express reported evidentiality, and thus can be easily conceived as remnants of speech complements. The evidence in support of the second scenario is twofold. First, the negated form of the evidential past participle is often identical with the negated form of compound tense forms, and second, the past participle is often extracted from the forms of compound past tenses. Taking into account the available evidence, the present study does not favour any of these two scenarios, considering them equally probable. Therefore, it must be admitted that the contribution made here to the research of the historical development of evidentiality coding is modest compared to the contribution made in relation to the first two goals set forth by the thesis.

Prospects for further research include an exclusive corpus study that would examine whether the negative forms of the evidential past participle in Estonian are more often identical with the negative forms of (indicative) simple past or with the negative forms of compound past. Another research area that might bear fruit is to reconsider (following Kask 1984: 282–285) the possibility of cumulative influence from German. As Breitbarth has shown in a series of recent papers (e.g. Breitbarth 2006), the ellipsis of the finite auxiliary in Early Modern German (ca. 1350–1700) correlates with non-assertive contexts, and with non-firsthand evidentiality in particular. If not more, such uses in German

may have had reinforcing effect in written Estonian on an already existing tendency to mark reported statements with auxiliary-less past participles.

Finally, some of the categorizations present in articles I–III concerning the status of past participles and infinitives in Estonian as markers of evidentiality are in need of revision. In accordance with the situation with the past participles in the Balkan languages, these markers are accounted for in terms of evidentiality strategy (see article I: pp. 129, 140; article II: pp. 818, 819 and article III: 910). Unlike the Balkan languages, however, the construction in which they are used in Estonian has exclusively evidential meaning, and therefore they should perhaps be characterized as instantiations of evidentiality proper. This confusion is created by the fact that Estonian has also a special morpheme (the so called *vat*-form) which can be listed in morphemicon as a dedicated evidential marker, whereas past participles and infinitives may have various functions and therefore do not qualify as morphologically dedicated markers. If, however, we expand the necessary criteria for dedication to the syntax, the relevant constructions with past participles and infinitives certainly satisfy the dedication requirement, and accordingly, if they are dedicated to express evidential meanings, they certainly should be accounted for in terms of evidentiality proper.

SOURCES REFERRED TO IN THE ENTIRE DISSERTATION

- Aikhenvald, Alexandra Y. 2000: Evidentials. La Trobe University. Melbourne.
- Aikhenvald, Alexandra Y. 2003: Evidentiality in typological perspective. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality*. *Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 1–31.
- Aikhenvald, Alexandra Y. 2004: Evidentiality. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y., R. M. W. Dixon (eds) 2003: *Studies in Evidentiality*. *Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins.
- Airila, Martti 1935: Viron kielen *oratio obliqua*'sta. – *Eripainos Suomi* 5, 17-sta. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Akatsuka, Noriko 1985: Conditionals and the epistemic scale. – *Language* 61, 3, 625–639.
- Aksu-Koç, Ayhan A., Dan I. Slobin 1986: A Psychological Account of the Development and Use of Evidentials in Turkish. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 159–167.
- Alexiadou, Artemis 1997: Adverb Placement. A Case Study in Anti-Symmetric Syntax. Amsterdam – Philadelphia: John Benjamins.
- Alvre, Paul 1993: Anmerkungen zum obliquen Modus im Estnischen und Livischen. – *Linguistica Uralica* 29, 2, 95–104.
- Anderson, Lloyd B. 1986: Evidentials, Paths of Change, and Mental Maps. *Typologically Regular Asymmetries*. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 273–312.
- ARS = Кочи, Р. Д., Скенди, Д. И., Серебренников, Б. А. (сост.) 1950: Албанско-русский словарь. Москва: Государственное издательство иностранных и национальных словарей.
- Asenova = Асенова, П. 2002: Балканско езикознание. Основни проблеми на балканския езиков съюз. Велико Търново: Фабер.
- van der Auwera, Johan 1998a: Introduction. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe*. *Empirical approaches to language typology* 20. Berlin: Mouton de Gruyter, 1–23.
- van der Auwera 1998b: Phasal adverbials in the languages of Europe. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe*. *Empirical approaches to language typology* 20. Berlin: Mouton de Gruyter, 25–145.
- van der Auwera Johan 1998c: Conclusion. – J. van der Auwera, D. P. O. Baoill (eds), *Adverbial constructions in the languages of Europe*. *Empirical approaches to language typology* 20. Berlin: Mouton de Gruyter, 813–836.
- van der Auwera, Johan 1998d: Revisiting the Balkan and Meso-American Linguistic Areas. – *Language Sciences* 20, 3, 259–270.
- van der Auwera, Johan, Vladimir Plungian 1998: Modality's semantic map. – *Linguistic Typology* 2, 1, 79–124.
- Barnes, Janet 1984: Evidentials in the Tuyuca verb. – *International Journal of American Linguistics* 50, 3, 255–271.
- Boye, Kasper 2006: Epistemic meaning: a cross-linguistic study. Ph.D. dissertation. Department of English, Germanic and Romance Studies, University of Copenhagen.

- Breitbarth, Anne 2006: Pragmatic aspects of Early Modern German auxiliary drop. – *Leiden Papers in Linguistics* 3.1. (eds N. Elouazizi, F. Landsbergen, M. Poss, M. Salzmann), 1–15.
- Brinton, Laurel J., Elizabeth Closs Traugott 2005: *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Buchholz, Oda, Wilfried Fiedler 1987: *Albanische Grammatik*. Leipzig: WEB Verlag Enzyklopädie Leipzig.
- Bybee, Joan L. 1985: *Morphology. A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.
- Bybee, Joan L., Revere Perkins and William Pagliuca 1994: *The evolution of grammar – tense, aspect, and modality in the languages of the world*. Chicago – London: University of Chicago Press.
- Campbell, Lyle 1991: Some Grammaticalization Changes in Estonian and their Implications. – E. C. Traugott, B. Heine (eds), *Approaches to Grammaticalization. Typological Studies in Language* 19. Amsterdam: Benjamins. 285–299.
- Ceplīte, Brigita, Laimdota Ceplītis 1991: *Latviešu valodas praktiskā gramatika*. Rīga. Zvaigzne.
- Chafe, Wallace, Johanna Nichols (eds) 1986: *Evidentiality: the Linguistic Coding of Epistemology. Advances in Discourse Processes* 20. Norwood, New Jersey: Ablex.
- Chafe, Wallace 1986: *Evidentiality in English Conversation and Academic Writing*. – W. Chafe, J. Nichols (eds), *Evidentiality: The Linguistic Coding of Epistemology*, 261–272.
- Chirikba, Vjacheslav 2003: Evidential category and evidential strategy in Abkhaz. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality*, 243–272.
- Cinque, Guglielmo 1999: *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. New York – Oxford: Oxford University Press.
- Cinque, Guglielmo 2006: *Restructuring and Functional Heads: The Cartography of Syntactic Structures*, 4. New York: Oxford University Press.
- Comrie, Bernard 1995: *Aspect: an Introduction to the Study of Verbal Aspect and Related Problems*. Fifth edition. Cambridge: Cambridge University Press.
- Comrie, Bernard, Maria Polinsky 2007: Evidentials in Tsez. – Z. Guentchéva, J. Landaburu (eds), *L'énunciation médiatisée II. Le traitement épistémologique de l'information: illustrations amérindiennes et caucasiennes*, 335–350.
- Croft, William 2003: *Typology and Universals*. (Second edition) Cambridge: Cambridge University Press.
- Dahl, Östen 1985: *Tense and Aspect Systems*. Oxford: Blackwell.
- Dahl, Östen, Maria Koptjevskaja-Tamm 2001: The Circum-Baltic Languages. Introduction to this volume. – Ö. Dahl, M. Koptjevskaja-Tamm (eds), *Circum-Baltic Languages*. Vol. 1. Past and Present. Amsterdam – Philadelphia: John Benjamins, xv–xx.
- Darden, Bill 1977: On the Admirative in Bulgarian. – *Folia Slavica*, 1, 1. Columbus, Ohio, 59–63.
- DeLancey, Scott 1997: Mirativity: The grammatical marking of unexpected information. – *Linguistic Typology* 1, 1, 33–52.
- DeLancey, Scott 2001: The mirative and evidentiality. – *Journal of Pragmatics* 33, 3, 369–382.

- Demina = Демина, Е. И. 1959: Пересказывательные формы в современном болгарском литературном языке. – С. Б. Бернштейн, (ред.), Вопросы грамматики болгарского литературного языка, Москва: Издательство Академии наук СССР, 313–378.
- Dendale, Patrick 1993: Le conditionnel de l'information incertaine: marqueur modal ou marqueur evidential? – G. Hilty (ed), Proceedings of the XXe Congrès International de Linguistique et Philologie Romanes. Tome 1, Section I. La phrase. Tübingen: Francke, 165–176.
- Dendale, Patrick, Liliane Tasmowski 2001: Introduction: Evidentiality and related notions. – Journal of Pragmatics 33, 3, 339–348.
- Dobrushina, Nina, Sergei G. Tatevosov 1996: Usage of verbal forms. – A. E. Kibrik, S. G. Tatevosov, A. Eulenberg (eds), Godoberi. München: Lincom Europa, 91–106.
- Duchet, Jean-Louis, Remzi Pernäska 1996: L'admiratif albanais: recherche d'un invariant sémantique. – Z. Guentchéva (ed), L'énonciation médiatisée. Louvain – Paris: Peeters, 31–46.
- Dudina = Дудина, Л. Н. 1993: Турецкий язык. Москва – Баку: Издательство “Азербайджан”.
- Eintrei = Эйнтрей, Т. И. 1982: Албанский язык. Ленинград: Издательство Ленинградского университета.
- EKG II = Erelt, M., Kasik, R., Metslang, M., Rajandi, H., Ross, K., Saari, H., Tael, K., Vare, S., 1993: Eesti keele grammatika II. Süntaks. Lisa: kiri. Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn.
- Endzelin, Janis 1922: Lettische Grammatik. Riga: A. Gulbis.
- Endzelīns, Jānis 1951: Latviešu valodas gramatika. Rīga: Latvijas Valsts izdevniecība.
- Erelt, Mati 2002: Does Estonian Have the Jussive? – Linguistica Uralica 38, 2, 110–117.
- Erelt, Mati (ed) 2003: Estonian Language. Linguistica Uralica Supplementary Series 1. Tallinn: Estonian Academy Publishers.
- Erelt, Mati, Helle Metslang, Karl Pajusalu 2006: Tense and Evidentiality in Estonian. – B. Cornillie, N. Delbecque (eds), Topics in Subjectification and Modalization. Belgian Journal of Linguistics 20, 125–136.
- Ernst, Thomas 2006: The Syntax of Adjuncts. Cambridge: Cambridge University Press.
- Fiedler, Wilfried 1966: Zu einigen Problemen des albanischen Admirativs. – Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig 15, 561–566.
- Fitneva, Stanka 2001: Epistemic Marking and Reliability Judgments: Evidence from Bulgarian. – Journal of Pragmatics 33, 3, 401–420.
- Friedman, Victor A. 1978: On the Semantic and Morphological Influence of Turkish on Balkan Slavic. – Papers from the Fourteenth Regional Meeting. Chicago Linguistic Society, April 13–14, 1978.
- Friedman, Victor A. 1985: Balkan Romani Modality and Other Balkan Languages. – Folia Slavica 9, 7. Columbus, Ohio, 381–389.
- Friedman, Victor A. 1986: Evidentiality in the Balkans. Bulgarian, Macedonian and Albanian. – W. Chafe, J. Nichols (eds), Evidentiality: the Linguistic Coding of Epistemology, 168–187.
- Friedman, Victor A. 1988a: The Category of Evidentiality in the Balkans and the Caucasus. – A. M. Schenker (ed), American Contributions to the Tenth International Congress of Slavists: Linguistics. Columbus: Slavica, 121–139.

- Friedman, Victor A. 1988b: Morphological Innovation and Semantic Shift in Macedonian. – *Zeitschrift für Balkanologie* 24, 1, 34–41.
- Friedman, Victor A. 1994: Surprise! Surprise! Arumanian Has Had an Admirative! – *Indiana Slavic Studies* 7, 1994, 79–89.
- Friedman, Victor A. 1997: A Contrastive View of Subordinate Aspect and the Opposition Confirmative/Non-confirmative in the Balkans with Particular Reference to Macedonian and Turkish. – *Balkanistica* 10. Columbus, Ohio, 185–201.
- Friedman, Victor A. 2001: Interrogation and Nonconfirmativity: A Pearceian Approach to the Intersection between Expressive Past and Present. – *Pearceian Semiotics: The State of the Art. An International Colloquium Canet-Plage*, June 27–30 2001. <http://www.univ-perp.fr/csp2001/friedman.htm>.
- Friedman, Victor A. 2002: Hunting the Elusive Evidential: The Third-Person Auxiliary as a Boojum in Bulgarian. – D. Dyer, V. Friedman (eds), *Of All the Slavs My Favorites: Studies in Honor of Howard I. Aronson on the Occasion of his 66th Birthday*. *Indiana Slavic Studies*, 12. Bloomington, Indiana: Slavica, 203–230.
- Friedman, Victor A. 2003: Evidentiality in the Balkans with special attention to Macedonian and Albanian. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality. Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 189–218.
- Friedman, Victor 2005: Admirativity: between modality and evidentiality. – *Sprachtypologie und Universalienforschung* 58, 1, 26–35.
- ГБЕ II = Тилков, Д., Стоянов, С., Попов К. (ред.) 1983: Граматика на съвременния български книжовен език. Том 2, Морфология. София: Издателство на Българска Академия на Науките, 1983.
- Gerdžikov = Герджиков, Г. 1984: Преизказването на глаголното действие в българския език. София: Наука и изкуство.
- Geurts, Bart, Janneke Huiting 2006: Modal Concord. – P. Dekker, H. Zeijlstra (eds), *Proceedings of the ESSLI 06 Workshop Concord Phenomena at the Syntax-Semantics Interface*, Málaga: ESSLI 06, 15–20.
- Givón, Talmy 1982: Evidentiality and epistemic space. – *Studies in Language* 6, 1, 23–49.
- GLJ = Амбразас, В., Валецкене, А., Валюлите, Э. 1985: Грамматика литовского языка. Академия Наук Литовской ССР. Вильнюс: Моклас.
- Graves, Nina 2000: Macedonian – a Language with Three Perfects? – Ö. Dahl (ed), *Tense and Aspect in the Languages in Europe. Empirical Approaches to Language Typology* 20, 6 = EUROTYP. Berlin – New York: Mouton de Gruyter, 479–494.
- Gronemeyer, Claire 1997: Evidentiality in Lithuanian. – *Lund University, Department of Linguistics Working Papers* 46, 93–112.
- Grünthal, Willem 1941: *Itämerensuomalaisen kielten yksikön nominatiivi objektin edustajana aktiivin yhteydessä*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Guentchéva, Zlatka (ed) 1996a: *L'énonciation médiatisée*. Louvain – Paris: Peeters.
- Guentchéva, Zlatka 1996b: *Le médiatif en Bulgare*. – Z. Guentchéva (ed), *L'énonciation médiatisée*, 47–70.
- Guentchéva Zlatka, Jon Landaburu (eds) 2007: *L'énonciation médiatisée II. Le traitement épistémologique de l'information: illustrations amérindiennes et caucasiennes*. Louvain – Paris – Dudley, MA: Peeters (Bibliothèque de l'Information Grammaticale 63).

- De Haan, Ferdinand. 1999: Evidentiality and epistemic modality: Setting boundaries. – *Southwest Journal of Linguistics* 18, 1, 83–101.
- De Haan, Ferdinand 2005a: Encoding speaker perspective: Evidentials. – Z. Frajzyngier, A. Hodges, D. S. Rood (eds), *Linguistic Diversity and Language Theories*. Amsterdam: John Benjamins, 379–397.
- De Haan, Ferdinand 2005b: Semantic Distinctions of Evidentiality. – M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 314–317.
- De Haan, Ferdinand 2005c: Coding of Evidentiality. – M. Haspelmath, M. S. Dryer, D. Gil, B. Comrie (eds), *The World Atlas of Language Structures*. Oxford: Oxford University Press, 318–321.
- Haarmann, Harald 1970: Die indirekte Erlebnisform als grammatische Kategorie. Eine eurasische Isoglosse. Wiesbaden: Otto Harrassowitz.
- Haarmann, Harald 1976: Arealtypologie der finnisch-ugrischen Sprachen – Prinzipien einer interdisziplinär orientierten Finnougristik. – *Finnisch-Ugrische Mitteilungen* 1, 3–14.
- Hakulinen, Auli, Pentti Leino 1987: Finnish Participial Construction from a Discourse Point of View. – *Ural-Altäische Jahrbücher* 59, 35–43.
- Halliday, Michael A. K. 1970: Functional diversity in language as seen from a consideration of mood and modality in English. – *Foundations of Language* 6, 3, 322–361.
- Haspelmath, Martin 2004: Does linguistic explanation presuppose linguistic description? – *Studies in Language* 28, 3, Special issue on linguistic evidence, (guest eds. M. Penke, A. Rosenbach), 554–579.
- Haspelmath, Martin, Matthew S. Dryer, David Gil, Bernard Comrie (eds) 2005: *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Helimski, Eugen 2003: Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. – *Nyelvtudományi Közlemények* 100, 156–167.
- Hengeveld, Kees 1989: Layers and operators in functional grammar. – *Journal of Linguistics* 25, 1, 127–157.
- Holvoet, Axel 2001a: On the Paradigm of the Oblique Mood in Lithuanian and Latvian. – *Linguistica Baltica* 9, 69–86.
- Holvoet, Axel 2001b: Mood and Modality in Latvian. – *Sprachtypologie und Universalienforschung* 54, 3, 226–252.
- Holvoet, Axel 2001c: *Studies in the Latvian Verb*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Holvoet, Axel 2001d: Impersonals and Passives in Finnic and Baltic. – Ö. Dahl, M. Koptjevskaja-Tamm (eds), *The Circum-Baltic Languages. Typology and Contact. Vol. 2, Grammar and Typology*. Amsterdam – Philadelphia: John Benjamins, 363–389.
- Holvoet, Axel 2007: *Mood and Modality in Baltic*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Hopper, Paul J., Elizabeth C. Traugott 1994: *Grammaticalization*. Cambridge: Cambridge University Press.
- Hopper, Paul J., Elizabeth C. Traugott 2003: *Grammaticalization. Second edition*. Cambridge: Cambridge University Press.

- Ifantidou, Elly 2001: *Evidentials and Relevance*. Amsterdam – Philadelphia: John Benjamins.
- Ikola, Osmo 1953: Viron ja liivin modus obliquuksen historiaa. – *Suomi* 106, 4, Helsinki: Suomalaisen Kirjallisuuden Seura.
- Izvorski, Roumyana 1997: The Present Perfect as an Epistemic Modal. – A. Lawson, E. Cho (eds), *Proceedings of Semantics and Linguistic Theory (SALT) VII*. Ithaca, New York: CLS Publications, 222–239.
- Jakobson, Roman 1971: Shifters, Verbal Categories, and the Russian Verb. – R. Jakobson (ed), *Selected writings*. Vol. 2, *Word and Language*. The Hague: Mouton, 131–147.
- Johanson, Lars 1998: Zum Kontakteinfluß türkischer Indirektive. – N. Demir, E. Taube (eds), *Turkologie heute – Tradition und Perspektive*. Materialien der dritten Deutschen Turkologen-Konferenz, Leipzig, 4.–7. Oktober 1994, Wiesbaden: Otto Harrassowitz, 141–150.
- Johanson, Lars, Bo Utas (eds) 2000a: *Evidentials: Turkic, Iranian and Neighbouring Languages*. *Empirical Approaches to Language Typology* 24. Berlin – New York: Mouton de Gruyter.
- Johanson, Lars 2000b: Turkic indirectives. – L. Johanson, B. Utas (eds), *Evidentials: Turkic, Iranian and Neighbouring Languages*. *Empirical Approaches to Language Typology* 24. Berlin – New York: Mouton de Gruyter, 61–87.
- Johanson, Lars 2003: Evidentiality in Turkic. – *Studies in Evidentiality*. A. Y. Aikhenvald, R. M. W. Dixon (eds), *Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 272–290.
- Joseph, Brian. D. 2003: *Evidentials: Summation, Questions, Prospects*. – A. Y. Aikhenvald, R. M. W. Dixon (eds), *Studies in Evidentiality*. *Typological Studies in Language* 54. Amsterdam – Philadelphia: John Benjamins, 307–327.
- Kask, Arnold 1984: *Eesti murded ja kirjakeel*. Tallinn: Valgus.
- Kehayov, Petar 2002: *Typology of Grammaticalized Evidentiality in Bulgarian and Estonian*. – *Linguistica Uralica* 38, 2, 126–144.
- Kehayov, Petar 2004: *Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Morfosüntaks ja distributsioon*. – *Keel ja Kirjandus* 47, 11, 812–829.
- Kehayov, Petar 2004: *Eesti keele evidentsiaalsussüsteem mõne teise keele taustal. Semantika*. – *Keel ja Kirjandus* 47, 12, 895–914.
- Kehayov, Petar, Florian Siegl 2007: *The evidential past participle in Estonian reconsidered*. – *Études finno-ougriennes* 38, Paris, 75–117.
- Kehayov, Petar 2008 (in print): *Interactions between grammatical evidentials and lexical markers of epistemicity and evidentiality: a case-study of Bulgarian and Estonian*. – V. A. Plungian, B. Wiemer (eds), *Lexikalische Evidenzialitäts-Marker in slavischen Sprachen*. *Wiener Slawistischer Almanach, Sonderband 72*, Wien.
- Kettunen, Lauri 1924: *Lauseliikmed eesti keeles*. Tartu: Eesti kirjanduse selts.
- Kiefer, Ferenc 1994: *Modality*. – R. E. Asher (ed), *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press, 2515–2520.
- Klaas, Birute 1997: *The Quotative Mood in the Baltic Sea Areal*. – M. Erelt (ed), *Estonian: Typological Studies II*. Tartu Ülikooli Eesti Keele Õppetooli Toimetised 7, 73–97.
- Klaas, Birute 2002: *Reported Command in Lithuanian Compared to Estonian*. – *Linguistica Uralica* 38, 2, 118–125.

- Koneski = Конески, Б. 1967: Граматика на македонскиот литературен јазик. Скопје: Култура.
- Kononov = Кононов, А. Н. 1956: Грамматика современного турецкого литературного языка. Москва – Ленинград: Издательство академии наук СССР.
- Koptjevskaja-Tamm, Maria, Bernhard Wälchli 2001: The Circum Baltic languages: An areal-typological approach – Ö. Dahl, M. Koptjevskaja-Tamm (eds), *The Circum-Baltic Languages. Typology and Contact. Vol. 2, Grammar and Typology.* Amsterdam – Philadelphia: John Benjamins, 615–750.
- Kozintseva = Козинцева, Н. А. 1994: Категория эвиденциальности (проблемы типологического анализа. – *Вопросы языкознания*, 1994/3, 92–104.
- Künnap, Ago 1992: On the analytical past tense forms in the Uralic languages. – *Linguistica Uralica* 28, 3, 173–179.
- Künnap, Ago 1994: On the Uralic Indirectal. – *Ural-Altische Jahrbücher* 13, 18–27.
- Lazard, Gilbert 1999: Mirativity, evidentiality, mediativity, or other? – *Linguistic Typology* 3, 1, 91–109.
- Lazard, Gilbert 2001: On the Grammaticalization of Evidentiality. – *Journal of Pragmatics* 33, 3, 359–367.
- Leetberg, Kaarel 1925: Eesti lauseõpetus: keskkoolidele. Rakvere.
- Lehmann, Christian 2002a: Thoughts on grammaticalization. Second, revised edition. *Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt* 9. Erfurt.
- Lehmann, Christian 2002b: New reflections on grammaticalization and lexicalization – I. Wischer, G. Diewald (eds), *New reflections on grammaticalization.* Amsterdam – Philadelphia: John Benjamins, 1–18.
- Leinonen, Maria 2000: Evidentiality in Komi Zyryan. – L. Johanson, B. Utas (eds), *Evidentials: Turcic, Iranian and Neighbouring Languages. Empirical Approaches to Language Typology* 24. Berlin – New York: Mouton de Gruyter, 419–440.
- Lewis, Geoffrey L. 1966: *Turkish. Teach Yourself Books.* London: The English Universities Press LTD.
- Lindstedt, Jouko 1998: On the Balkan Linguistic Type. – *Studia Slavica Finlandensia* 15, Helsinki: Institute for Russian and East European Studies, 91–103.
- Lindstedt, Jouko 2000a: The Perfect – Aspectual, Temporal and Evidential. – Ö. Dahl (ed), *Tense and Aspect in the Languages in Europe. Empirical Approaches to Language Typology* 20, 6 = EURO TYP. Berlin – New York: Mouton de Gruyter, 365–383.
- Lindstedt, Jouko 2000b: Linguistic Balkanization: Contact-Induced Change by Mutual Reinforcement. – D. G. Gilberts, J. Nerbonne, J. Schaeken (eds), *Languages in Contact. Studies in Slavic and General Linguistics* 28. Amsterdam – Atlanta, GA: Rodopi, 231–246.
- Loorits, Oskar 1923: *Eesti keele grammatika.* Tartu: Odamees.
- Löttsch = Леч, Р. 1971: О специфическом характере грамматической интерференции связанной с происхождением так называемых пересказывательных форм болгарского языка. – *Исследования по славянскому языкознанию. Сборник в честь шестидесятилетия профессора С. Б. Бернштейна.* Москва: Наука, 173–181.
- Lyons, John 1977. *Semantics. Vol. 2.* Cambridge: Cambridge University Press.
- Макареџ = Макареџев, М. М. 2008: Связь лексических и грамматических показателей эвиденциальности в болгарском языке. – V. A. Plungian, B. Wiemer (eds),

- Lexikalische Evidenzialitäts-Marker in slavischen Sprachen. Wiener Slawistischer Almanach, Sonderband 72, Wien.
- Maslov = Маслов, Ю. С. 1981: Грамматика болгарского языка. Москва: Высшая школа.
- Mathiassen Terje 1996: Some aspects of language contacts and convergency phenomena in the Baltic Area – Evidence from Lithuanian, Latvian, Livonian, Estonian, Russian and Polish. – J. Fernandez, R. Raag (eds), *Contacts des langues et de cultures dans l'aire baltique / Contacts of languages and cultures in the Baltic area: Mélanges offerts à Fanny de Sivers.* (= Uppsala Multiethnic Papers 39). Uppsala, 171–179.
- McCready, Eric, Norry Ogata 2007: Evidentiality, modality and probability. – *Linguistics and Philosophy* 30, 2, 147–206.
- Metslang, Helle 1994: Temporal relations in the predicate and the grammatical system of Estonian and Finnish. – *Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja* 39. Oulu.
- Metslang, Helle, Karl Pajusalu 2002: Evidentiality in South Estonian. – *Linguistica Uralica* 38, 2, 98–109.
- Meydan, Métiyé 1996: Les emplois médiatifs de *-miş* en Turc. – Z. Guentchéva (ed), *L'énonciation médiatisée.* Louvain – Paris: Peeters, 125–143.
- Muižniece, Liena, Helle Metslang, Karl Pajusalu 1999: Past Participle Finitization in Estonian and Latvian. – M. Erelt (ed), *Estonian: Typological Studies III.* Tartu Ülikooli Eesti Keele Õppetooli Toimetised 11, 128–157.
- Mushin, Ilana 2001: Evidentiality and Epistemological Stance. *Narrative Retelling.* Amsterdam: John Benjamins.
- Nichols, Johanna 1986: The bottom line: Chinese Pidgin Russian. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 239–257.
- Nichols, Johanna 2007: What, if anything, is typology? – *Linguistic Typology* 11, 1, 231–238.
- Nitsolova = Ницолова, Р. 2006: Взаимодействие эвиденциальности и адмиративности с категориями времени и лица глагола в болгарском языке. – *Вопросы языкознания*, 2006/4, 27–45.
- Nuyts, Jan 2001: Epistemic modality, language and conceptualization. *A cognitive-pragmatic perspective*, Amsterdam: John Benjamins.
- Nuyts, Jan (forthcoming): The 'one-commitment-per-clause' principle and the cognitive status of qualificational categories. – *Linguistics*.
- Oswalt, Robert L. 1986: The evidential system of Kashaya. – W. Chafe, J. Nichols (eds), *Evidentiality: the Linguistic Coding of Epistemology*, 29–45.
- Palmer, Frank, R. 2001: *Mood and Modality.* Cambridge: Cambridge University Press.
- Plungian, Vladimir 2001: The Place of Evidentiality within the Universal Grammatical Space. – *Journal of Pragmatics* 33, 3, 349–357.
- Põld, Harald 1922: *Eesti keeleõpetus. II jagu. Lauseõpetus.* Tallinn: G. Pihlakas.
- Quine, Willard V. O. 1953: *From a Logical Point of View.* Harvard: Harvard University Press.
- Rätsep, Huno 1971: Kas kaudne kõneviis on kõneviis? – *Keel ja struktuur* 5. Tartu: Tartu Riiklik Ülikool, 45–69.
- Rudzīte, Marta 1964: *Latviešu dialektoloģija.* Rīga: Latvijas Valsts izdevniecība.
- Saareste, Andrus 1932: *Die estnische Sprache.* Tartu: Akadeemiline Kooperatiiv.

- Schaller, Helmut W. 1975: Die Balkansprachen. Eine Einführung in die Balkanphilologie. Heidelberg: Carl Winter.
- Sepper, Maria-Maren 2002: Evidentsiaalsuse väljendamise eesti kirjakeeles ühe sajandi jooksul. – Väikeisi kiili kokkoptumisõq / Väikeste keelte kontaktid. Võro Instituudi Toimõndusõq 14. Võru: Võro Instituut, 185–206.
- Serebrennikov = Серебренников, Б. А. 1960: Категории времени и вида в финно-угорских языках пермской и волжской групп. Москва: Наука.
- Siegl, Florian 2004: The 2nd past in the Permic languages – form, function, and a comparative analysis from a typological perspective. Unpublished MA thesis. University of Tartu. Chair of Finnic languages. <http://www.utlib.ee/ekollekt/diss/mag/2004/b16700053/Siegl.pdf>
- Slobin, Dan, Ayhan A. Aksu 1982: Tense, Aspect, and Modality in the Use of the Turkish Evidential. – P. J. Hopper (ed), Tense-Aspect: Between Semantics and Pragmatics. Amsterdam – Philadelphia: John Benjamins, 185–200.
- Stolz, Thomas 1991: Sprachbund im Balticum? Estnisch und Lettisch im Zentrum einer sprachlichen Konvergenzlandschaft (= Bochum-Essener Beiträge zur Sprachwandelforschung 13). Bochum: Universitätsverlag Dr. N. Brockmeyer.
- Svenonius, Peter 2001: Subject Positions and the Placement of Adverbials. – P. Svenonius (ed), Subject, Expletives and the EPP. New York: Oxford University Press, 201–242.
- Sytov = Сытов, А. П. 1979: Категория адмиратива в албанском языке и ее балканские соответствия. – А. В. Десницкая (ред.), Проблемы синтаксиса языков балканского ареала. Ленинград: Наука, 90–124.
- Toomet, Piret 2000: Mõnest kaudsuse väljendamise võimalusest tänapäeva eesti keeles. – Keel ja Kirjandus 43, 4, 251–259.
- Tournadre, Nicolas 1996: Comparaison des systèmes médiatifs de quatre dialectes tibétains. (tibétain central, ladakhi, dzongkha et amdo). – Z. Guentchéva (ed), L'énonciation médiatisée. Louvain – Paris: Peeters, 195–213.
- TÜKK = Tartu ülikooli kirjakeele korpus: <http://www.cl.ut.ee/korpused/baaskorpus/>.
- Valeckienė = Валецкиене, А. 1966: Литовский язык. – Языки народов СССР. Т. 1. Индоевропейские языки. Москва: Наука.
- Van Valin, Robert D., Randy J. LaPolla 1997: Syntax: Structure, Meaning and Function. Cambridge: Cambridge University Press.
- Van Valin, Robert D. 2005: Exploring the Syntax-Semantics Interface. Cambridge: Cambridge University Press.
- Wälchli, Bernhard 2000: Infinite predication as a marker of evidentiality and modality in the languages of the Baltic region. – Sprachtypologie und Universalienforschung 53, 2, 186–210.
- Weigand, Gustav 1923/4: The Admirative in Bulgarian. – The Slavonic Review 2, London, 567–568.
- Whaley, Lindsay, J. 1997: Introduction to typology: the unity and diversity of language. Thousand Oaks – London – New Delhi: Sage Publications.
- Wiedemann, Ferdinand J. 1875: Grammatik der Ehstnischen Sprache, zunächst wie sie in Mittelehstland gesprochen wird, mit Berücksichtigung der anderen Dialekte. St. Pétersbourg.

- Wiemer, Björn 2006: Particles, parentheticals, conjunctions and prepositions as evidentiality markers in contemporary Polish. (A first exploratory study.) – *Studies in Polish Linguistics* 3, 1–52.
- Willet, Thomas 1988: A Cross-Linguistic Survey of the Grammaticalization of Evidentiality. – *Studies in Language* 12, 1, 51–97.

SUMMARY IN ESTONIAN

Balkani ja Baltikumi keelte evidentsiaalsussüsteemid areaal-tüpoloogilisest perspektiivist

Käesolev doktoriväitekiri uuris evidentsiaalsust areaal-tüpoloogilisest perspektiivist seitsmes keeles: türgi, bulgaaria, makedoonia, albaania, leedu, läti ja eesti. Uurimus seadis endale kolm eesmärki:

- a) võrrelda erinevate keelte evidentsiaalsussüsteeme ja mõõta nendevahelisi tüpoloogilisi kaugusi,
- b) uurida evidentsiaalsust teiste verbikategooriate hulgas, eeskätt evidentsiaalsuse suhet episteemilise modaalsusega,
- c) selgitada partitsiibikujulise evidentsiaalsuse kõige tõenäolisemaid arenguradu.

Evidentsiaalsussüsteemide võrdluse tulemusi

Evidentsiaalsuse kujunemine koherentseks grammatiliseks kategooriaks on omane nii Balkani kui ka Läänemere, täpsemalt selle idakalda, keeleliidule. Samas on evidentsiaalsuse tähtsus nende keeleliitude defineerimisel olnud erinev. Kuna geograafiliselt moodustab Balkani evidentsiaalsusala ainult murdosa suuremast evidentsiaalsusvööndist, mida Nichols (1986: 253) iseloomustab sõnadega “Paneuraasia evidentsiaalne perfekt”, DeLancey (2001: 370) aga “Balkani ja Lääne-Aasia evidentsiaalsusvöönd” (vt lisa 2), ei saa evidentsiaalsust pidada Balkani keeleliidu defineerivaks tunnuseks. Läänemere evidentsiaalsusala jääb aga suurest Euraasia areaalist kõrvale, mistõttu võib evidentsiaalsust pidada Baltikumi keeleliidu defineerivaks tunnuseks. Üks käesoleva töö eesmärke oli võrrelda Baltikumis esinevaid evidentsiaalsussüsteeme Euraasia süsteemidega (Balkani evidentsiaalsussüsteemide näol) ning selgitada, kas kahe areaali geograafiline lahusolu kajastub ka nende evidentsiaalsussüsteemide tüpoloogilises lahknevuses. Kuna ükski varasem uurimus ei olnud sellist küsimust esitanud, oli käesolev uurimus asjakohane.

Artiklites II ja III võrdlesin türgi, bulgaaria, makedoonia, albaania, leedu, läti ja eesti evidentsiaalsussüsteeme 32 tunnuse (või parameetri) järgi. 4.1. peatükis vahetasin selle tunnuste valimi 19 paremini määratletud ja tüpoloogiliselt relevantset tunnust sisaldava valimi vastu. Mõned vana valimi tunnused jäid samaks, teised defineerisin uuesti, kolmandad taandasin ühele tunnusele ja neljandad jäid uuest valimist välja. Saadud tulemusi töötlesin keeletüpoloogias käibiva isopleetmeetodiga (vt van der Auwera 1998a, 1998b, 1998c, 1998d), mis võimaldab kindlaks määrata keelte (või nende struktuuriliste elementide) omavahelisi sarnasusi nende ühiste tunnuste hulga arvutamise kaudu.

Tabelid 1–3 lisas 3 näitavad iga evidentsiaalsussüsteemi väärtusi iga võrdlusparameetri kohta. Võimalikud väärtused 19 parameetri valimis on +, (+) ja –. Kui võrrelda tabeleid 1 ja 2 (artiklite II ja III lõpus) ja väärtusi lisas 3 omavahel, võib täheldada, et uue valimi väärtused ei erine oluliselt vana valimi väärtustest. Enamik erinevustest tuleneb vanas valimis esineva väärtuse (–) kaotamisest, mis sisuliselt tähendas väärtuste (+) ja (–) mahu laiendamist, selleks et haarata enam esinemisjuhte. Mõnel juhul on väärtus (+) vahetatud + vastu. Kuna puudub range protseduur, mille abil omistatakse õige väärtus mitte-diskreetselt jaotuvatele tunnustele, siis teatud suva väärtuste valimisel oli vältimatu. Mõnel juhul võib täheldada olulisi erinevusi väärtuste vahel vanas ja uues valimis, kuid see on üksnes antud parameetri definitsiooni ahendamise tagajärg. Kui välja arvata need parameetrid, mille definitsioon on oluliselt teisenenud, puudutavad muutused ainult piirnevaid väärtusi skaalal +, (+), (–), –. See tähendab, et puuduvad juhud, mille puhul on + või (+) väärtus muudetud – väärtuseks või vastupidi. Juba see näitab, et revideeritud valim ja uued andmed ei muuda oluliselt erinevate keelte parameetriväärtustest tuletatavaid tüpoloogilisi kaugusi.

Lisa 4 tabelis 1 on toodud tüpoloogilised kaugused keelte evidentsiaalsussüsteemide vahel, mille arvutus põhineb ühiste väärtuste arvul. Kaks keelt jagavad väärtust, kui antud tunnus esineb mõlemas keeles või ei esine mitte kummaski. Tunnus esineb juhul, kui ta saab väärtuse + või (+), ega esine juhul, kui ta saab väärtuse –. Lisa 4 tabelis 1 on kõige suurem arv 17. Nii palju ühiseid väärtusi on türgi ja bulgaaria evidentsiaalsussüsteemil, türgi ja makedoonia evidentsiaalsussüsteemil ning bulgaaria ja makedoonia evidentsiaalsussüsteemil. See tulemus ei ole üllatav, kuna bulgaaria ja makedoonia keel on väga lähedased sugulaskeeled, mida veel 100 aastat tagasi peeti ühe keele murreteks ning mõlemad keeled on kogenud tugevat türgi keele mõju. Kõige madalam arv tabelis on 7, mis kuulub süsteemipaaridele eesti-türgi, eesti-makedoonia ja läti-albaania. See tulemus on natuke huvitavam, kuna näitab Balkani ja Baltikumi keelte evidentsiaalsussüsteemide tüpoloogilisi äärmusi.

Lisas 4 toodud joonised 1–7 kujutavad iga tabelis 1 esinevat numbrit teatud joonena, mida nimetatakse isopleediks. Selliseid jooniseid nimetatakse aga isopleetkaartideks. Neil kaartidel vastab keelte asukoht umbkaudu keelte geograafilisele asukohale Euroopas. Iga kaart näitab keelte sarnasuse astet ühe teatud keelega, mis ilmneb isopleetide sees suletud alade võrdluses. Esimene kaart näitab sarnasuse astet türgi keelega. Kõige sisemine isopleetjoon sellel kaardil, mille sees on türgi, bulgaaria ja makedoonia keel, näitab, et viimased kaks keelt jagavad türgi keelega 17 ühist väärtust. Järgmine sarnasuse aste hõlmab leedu keelt, mis jagab türgi keelega 15 ühist väärtust. Albaania keel jagab aga türgi keelega 12 väärtust ning läti ja eesti keel vastavalt 8 ja 7 väärtust. Joonis 2 näitab, milline näeb välja isopleetide kobar bulgaaria keele ümber, joonised 3, 4, 5, 6 ja 7 vastavalt makedoonia, albaania, leedu, läti ja eesti keele ümber.

Kõige huvitavamad leiud nende isopleetkaartide peal on esiteks see, et leedu keele evidentsiaalsussüsteem sarnaneb enam bulgaaria, makedoonia ja türgi keele evidentsiaalsussüsteemiga kui lähedase sugulase läti keele evidentsiaalsussüsteemiga, ning teiseks see, et albaania keele süsteem seisab selgelt eraldi türgi-bulgaaria-makedoonia rühmast.

Minnes tagasi lisan 3 esitatud tabelite juurde, võib näha, et leedu süsteemi ootamatu sarnasus bulgaaria, makedoonia ja türgi süsteemiga ning albaania süsteemi ootamatu erinevus viimastest ilmneb nii formaalsete kui ka semantiliste (ja pragmaatiliste) tunnuste võrdluses. Näiteks leedu keele evidentsiaalsussüsteemil on bulgaaria keele süsteemiga viis ühist vormitunnust, läti keele süsteemiga aga ainult neli. Semantika ja pragmaatika vallas on leedu ja bulgaaria süsteemil üheksa ühist tunnust, leedu ja läti süsteemil aga ainult viis.

Ka mõned teised leiud tunduvad huvitavana, kuid ei ole kahjuks piisavalt ilmekad, et saada ulatuslike üldistuste aluseks. Näiteks näitavad numbrid tabelis 1 (lisa 4) ja vastavad isopleetjooned, et eesti keele evidentsiaalsussüsteem on läti süsteemiga võrreldes ühe punkti jagu lähemal albaania ja bulgaaria süsteemile, samas kui geneetilise keelesuguluse ja geograafia alusel oleks oodatav, et läti süsteemil oleks nende süsteemidega rohkem ühiseid tunnuseid kui eesti keelele. Samamoodi on üllatav, et albaania keele süsteem on kahe punkti võrra sarnasem bulgaaria süsteemiga kui naabruses oleva makedoonia keele süsteemiga. Need erinevused on aga nii tühised, et võivad vabalt olla artefaktid, mis on tuletatavad tasakaalustamatust parameetri-valimist või moonutatud parameetri-definitsioonist.

Isopleetkaardid joonistel 1–7 annavad kokkuvõttes võimaluse eristada kaks keelerühma, milles erinevad evidentsiaalsussüsteemid jagavad vähemalt 15 väärtust (vt joonist 8 lisa 4). Need rühmad on 1) türgi-bulgaaria-makedoonia-leedu ja 2) läti-eesti. Isopleetide jaotus on siin niivõrd ühemõtteline, et need rühmad ei saa olla ebaõnnestunud parameetri-valimi tulemus. Suurele vahele vaatamata jagab leedu keele evidentsiaalsussüsteem türgi, bulgaaria ja makedoonia süsteemiga 15 ühist joont, samas kui üle tuhande aasta kestnud naabrusele vaatamata jagavad makedoonia ja albaania süsteem ainult 12 ühist joont. Samamoodi kõnekas on asjaolu, et läti ja eesti süsteem jagavad omavahel 16 joont, kuid läti süsteemil on ainult 10 ja eesti süsteemil ainult 9 ühist joont leedu süsteemiga.

Kuigi rühm türgi-bulgaaria-makedoonia-leedu on kahtlemata huvitav leid, tuleb meele pidada seda, et isopleetmeetodist ei järeldu, et keelepaarid, mis jagavad sama arvu väärtuseid, jagavad väärtuseid täpselt samade parameetrite kohta. Paarid türgi-bulgaaria ja bulgaaria-makedoonia näiteks saavad mõlemad kokku 17 punkti, kuid nagu tabelid 1–3 lisa 3 näitavad, mitte täpselt samade parameetrite kohta. Sellepärast tuleb küsida, millised parameetrid saavad täpselt sama väärtuse rühma türgi-bulgaaria-makedoonia-leedu kõikides keeltes, ehk teisisõnu, mis on selle rühma evidentsiaalsussüsteemi standard. Lisa 3 tabelites

1–3 esitatud tulemuste põhjal võib seda standardit iseloomustada järgmiste tunnustega:

- 1) evidentsiaalsust vormistatakse mineviku partitsiipidega;
- 2) infinitiive ja muid mitte-partitsiibilisi verbi käändelisi vorme ei kasutata evidentsiaalsuse markeritena;
- 3) abiverbi ellips loob vormilise kontrasti ajasüsteemiga;
- 4) evidentsiaalsust vormistatakse vahenditega, millel on ka muid grammatilisi funktsioone;
- 5) eksistentsiaalse koopula partitsiibivormi kasutatakse distantspartiklina;
- 6) refereeritud, järelaluslikku ja miratiivset evidentsiaalsust kodeeritakse ühe grammatilise markeriga, mis markeerib kaudse evidentsiaalsuse üldmõistet;
- 7) on olemas strateegiaid refereeritud evidentsiaalsuse täpsustamiseks;
- 8) kaudse evidentsiaalsuse markeriga ei vormistata geneerilisi väiteid;
- 9) kõneleja võib kasutada kaudse evidentsiaalsuse markerit enda üne-nägude edasiandmiseks;
- 10) kaudse evidentsiaalsuse markerit kasutatakse enda füüsiliste ja vaimsete seisundite objektiveerijana;
- 11) kaudse evidentsiaalsuse marker esineb rahvanarratiivides predikaadi tavakohase vormina;
- 12) kaudse evidentsiaalsuse markerit võib kasutada otsestes mitte-kajaküsimustes.

Kahtlemata on paljud nendest tunnustest omased ka kogu Euraasia evidentsiaalsusareali standardile (vt nt Friedman 1988a; DeLancey 2001; Aikhenvald 2004: 27–31). Toodud tulemuste ning muude eelteadmiste põhjal võib järeldada, et türgi, bulgaaria, makedoonia ja leedu keel kuuluvad evidentsiaalsussüsteemi tüübi alla, mida Nichols nimetas Paneuraasia evidentsiaalseks perfekti. Läti, eesti ja albaania evidentsiaalsussüsteemi tuleb aga arvesse võtta grammatiliste uuendustena geograafiliselt piiratud alal.

Nendel järeldustel on omakorda üldisemad tagajärjed areaalilingvistika jaoks. Teiste grammatiliste nähtuste kõrval on evidentsiaalsus aidanud Balkani ja Baltikumi areaali tuvastada ja (osaliselt) naaberaladest eristada, kuid pilt muutub, kui uurida ja võrrelda evidentsiaalsuse tunnuseid erinevates keeltes. Selline sissevaade näitab sügavamaid sarnasuse alasid (vt jälle joonist 8 lisas 4), mis on üsna erinevad nendest, mida on tuvastatud tervete grammatiliste süsteemide võrdlemisel. Evidentsiaalsussüsteemi tüüp, mida me täheldasime leedu, bulgaaria, makedoonia ja türgi keeles, teiseneb Väike-Aasia, Kaukaasia ja Kesk-Aasia suunas, nii et vahemaa suurenemine viib astmelise tüübi vahetuseni ilma järsude muutusteta. Täheldatud Ida-Euroopa keelte rühm ei pruugi tingimata olla keelekontakti tulemus, vaid võib olla ka universaalsete keeletöötlemehhanismide ilming. Leedu, bulgaaria ja makedoonia keele puhul on silmatorkav see, et leedu keelel on kõige keerulisem aja-aspekti süsteem olemasolevate balti keelte hulgas, bulgaaria ja makedoonia keelel on aga kõige

keerulisem aja-aspekti süsteem slaavi keelte hulgas. Seega võib täheldada otsest korrelatsiooni keerulise aja-aspekti süsteemi ja kirjeldatud evidentsiaalsussüsteemi tüübi vahel.

Toodud tulemused räägivad kaudselt vastu väitele, et bulgaaria ja makeedoonia evidentsiaalsussüsteem on struktuurilised laenud türgi keelest. Ka leedu keele süsteem on türgi süsteemiga väga sarnane, kuigi leedu ja türgi keele vahel ei ole kunagi otsest kontakti olnud.

Parameetriväärtused tabelites 1–3 (lisa 3) tõstavad küsimuse implikatsioonisuhetest eri tunnuste vahel. Järgnevas nimekirjas on toodud kõige silmahakkavamad implikatsioonid.

- a) Kui keel kasutab verbi muid käändelisi vorme evidentsiaalsuse markeritena, siis ta kasutab ka oleviku partitsiipe evidentsiaalsuse markeritena; kui ta kasutab oleviku partitsiipe evidentsiaalsuse markeritena, siis ta kasutab ka mineviku partitsiipe evidentsiaalsuse markeritena. Niisiis tegemist on hierarhiaga 'mineviku partitsiip' > 'oleviku partitsiip' > 'verbi muu käändeline vorm'.
- b) Kui evidentsiaalsuse marker väljendab järelduslikku evidentsiaalsust ja miratiivsust, siis väljendab ta alati ka refereeritud evidentsiaalsust.
- c) Keelel võib olla vahendeid kolmeliikmelise opositsiooni kodeerimiseks, mille üks liige on funktsionaalselt markeerimata ainult juhul, kui selles keeles on laiema semantikaga indirektaali (ehk kaudse evidentsiaalsuse) tüüpi süsteem.
- d) Kui keeles on evidentsiaalsussüsteem, mis kodeerib kaudsuse üldmõistet, siis on selles tõenäoliselt ka vormivahendeid refereeritud evidentsiaalsuse eristamiseks.
- e) Vormiliselt markeeritud evidentsiaalsusopositsiooni liiget võib kasutada unenägude edasiandmisel ainult juhul, kui keeles on laiem kaudse evidentsiaalsuse süsteem.
- f) Kaheliikmelise evidentsiaalsussüsteemi vormiliselt markeeritud liige võib konventsionaliseeruda rahvanarratiivi predikaadi põhivormiks ainult tingimisel, et seda kasutatakse ka refereeritud evidentsiaalsuse vormistamiseks.
- g) Kaheliikmelise evidentsiaalsussüsteemi vormiliselt markeeritud liige võib olla kasutusel otsestes mitte-kajaküsimustes ainult juhul, kui tegemist on laia kaudse evidentsiaalsuse süsteemiga.
- h) Koopulaverbidest tuletatud distantsipartiklid esinevad ainult keeltes, milles on kaheliikmeline üldise kaudse evidentsiaalsuse süsteem.

Tabelites 1–3 (lisa 3) toodud andmete põhjal võib väita, et implikatsioonid a)–d) on ühesuunalised, teisisõnu nad ei ole kehtivad, kui me vahetame tingiva ja tingitava osalause sisu. Implikatsioonid e)–h) on seevastu kahesuunalised, ehk nad kehtivad isegi siis, kui me vahetame tingiva ja tingitava osalause sisu. Järgnevalt kontrollin iga implikatsiooni kehtivust laiemal tüpoloogilisel taustal. Aikhenvaldi uurimus (2004) on suurepärase allikas selliste katsetuste jaoks.

Implikatsiooni a) esimese poole võib tunnistada kehtetuks, kuna mõnes eesti murdes on evidentsiaalsuse markeritena kasutusel mineviku partitsiip ja infinitiiv, kuid mitte oleviku partitsiip (vt Kask 1984: 254–270). Samal ajal ei ole mul selliseid keelelisi andmeid, mis räägiks vastu implikatsioonile, mis seab oleviku partitsiibi evidentsiaalsuse kasutuse tingimuseks mineviku partitsiibi kasutuse selles funktsioonis. See implikatsioon sobib üldise arusaamaga, et evidentsiaalsuse kategooria on loomupäraselt seotud minevikuga (vt nt Wälchli 2000). Seda implikatsiooni abstraktsemalt sõnastades võib öelda, et evidentsiaalsuse vormistamine olevikuliste (või imperfektiivsete) nominalisatsioonidega eeldab evidentsiaalsuse vormistamist minevikuliste (või perfektiivsete) nominalisatsioonidega. Edaspidine uurimistöo peaks kontrollima, kas tegemist on absoluutse universaaliga või mitte. Teise implikatsiooni b) vastu räägivad andmed nepali ja jaruara keelest (vt Aikhenvald 2004: 156). Kuigi ei teata ühtegi kindlat juhtu, mis tühistaks c), on see implikatsioon natuke liiga spetsiifiline, et olla tüpoloogiliselt oluline. Implikatsioon d) on seevastu väga huvitav ja väärivad edasist katsetamist; vasturääkivaid andmeid minu teada ei ole, poolträäkivate andmete kohta vt Aikhenvald (2004: 155–156). Implikatsioon e), mis kehtib ainult kaheliikmeliste evidentsiaalsussüsteemide kohta, on huvitav ja universaalselt oluline, kuid see tuleb nõrgendada ühesuunaliseks (vt Aikhenvald 2004: 344–345). Kuid isegi sellisel juhul tundub see natuke liiga tugevana. Üks võimalik vastunäide tuleb jaruara keelest (vt Dixon 2004, viidatud Aikhenvaldi 2004: 23–24, 345 kaudu). Implikatsioon f) tundub tüpoloogilise sõela läbivat (vt Aikhenvald 2004: 310–315), kuid ainult ühesuunalisena. Implikatsioon g) vastu aga räägivad andmed tšinanteeği (Aikhenvald 2004: 249, 291) ja jutši keelest (Aikhenvald 2004: 249). Viimane implikatsioon h) tundub olemasolevate andmete põhjal kehtivat, kuid on ilmselt liiga spetsiifiline, et pakkuda tüpoloogiale huvi.

Kokkuvõttes võib nentida, et ainsad implikatsioonid, mis väärivad edasist tüpoloogilist tähelepanu, on implikatsiooni a) teine osa ('mineviku partitsiip' > 'oleviku partitsiip') ning implikatsioonid d), e) ja f). See jääb aga tulevase uurimuse teemaks.

Lõpuks tuleb märkida ka ühte huvitavat reeglipärasust, mis ilmneb evidentsiaalsussüsteemide võrdlusel. Tabelites 1–2 (lisa 3) tuleb välja, et keeled (läti ja eesti), mis on rikkad evidentsiaalsust kodeeriva grammatilise materjali poolest (st kasutavad peale partitsiipide ka muid nominalisatsioone), on funktsioonivaesed, st nende evidentsiaalsuse markerid on üsna kitsa tähendusega (refereeritud evidentsiaalsus). Samas keeled, mis on vaesed evidentsiaalsust kodeeriva grammatilise materjali poolest (Balkani areaali keeled), on funktsioonirikkad: see ilmneb nende evidentsiaalsussüsteemide markeeritud liikme laiemas tähenduses (kaudne evidentsiaalsus ehk indirektaal). Täheldatud reeglipärasus jääb käesolevas töös seletamata. Tulevased uurimused, mis võtavad arvesse vaadeldavate keelte üleüldisi struktuurilisi omadusi, võivad suuta seda seletada.

Evidentsiaalsuse ja episteemilise modaalsuse suhtest

Viimane artikkel (IV) tõi välja grammatiliste evidentsiaalsuse markeerimise vahendite ning leksikaalsete episteemilise modaalsuse ja evidentsiaalsuse markeerimise vahendite vastastikuse mõjustuse peamised tüübid. Need tüübid formuleerisin ainult bulgaaria ja eesti keele põhjal, kuid olen eeldanud, et need on universaalsema kehtivusega.

Järgmist eestikeelset lihtlauset võib kontekstist välja tõstetuna tõlgendada mitmeti:

- (1) *Lembit olevat vist Pärnus.*

Artikkel IV eristab kuni neli võimalikku tõlgendust:

- 1) Kõneleja on kuulnud kellegi käest, et Lembit on vist Pärnus.
- 2) Kõneleja arvab (kuid pole täiesti kindel), et ta on kuulnud, et Lembit on Pärnus.
- 3) Kõneleja on kuulnud, et Lembit on Pärnus, ja ta arvab, et Lembit on Pärnus.
- 4) Kõnelejale tundub, et Lembit on Pärnus.

Esimese tõlgenduse puhul on määrsõna *vist* refereeringu mõjualas, mida vormistatakse evidentsiaalse *vat*-vormiga. Seega sisaldub *vist* juba refereeritud väites. Teise tõlgenduse puhul on *vat*-vormiga vormistatud refereering vastupidi *vist* mõjualas: kõneleja on vist kuulnud Lembitu kohta. Kolmas tõlgendus sisaldab kahte hinnangut propositsiooni sisule, millest kumbki ei ole teise mõjualas. Siin modifitseerivad operaatorid *vist* ja *-vat* propositsiooni kumbki iseseisvalt. Viimases tõlgenduses on *-vat* ja *vist* tajutud ühe tervikuna, mis sellisena viitab informatsiooni kaudsele allikale ning sellega korreleeruvale episteemilisele hinnangule, mis väljendab seda, et väide ei ole täielikult usaldatav. Artiklis IV on tõlgendused 1)–2) sildistatud kui 'hõlmav tõlgendus', tõlgendus 3) kui 'analüütiline tõlgendus' ja tõlgendus 4) kui 'ühend tõlgendus'.

Põhirõhk artiklis IV on neil teguritel, mis võimaldavad selliste kollokatsioonide ühend tõlgendust. Nii eesti kui ka bulgaaria keeles on mõnede kollokatsioonide ühend tõlgendamine nii sagedane nähtus, et selles võib näha püsiühendi tekkimise algfaasi. Ühend tõlgendus viitab iseenesest evidentsiaalsuse ja episteemilise modaalsuse ontoloogilisele seosele. Kuna ühend tõlgendusi esineb ka sellistes keeltes nagu bulgaaria keel, mis on mänginud otsustavat rolli evidentsiaalsuse määratlemisel verbikateooriana, siis tuleks uuesti kaaluda võimalust evidentsiaalsuse ja episteemilise modaalsuse käsitlemiseks ühise kategooria raames. See järeldus langeb kokku Boye hiljutise uurimusega, mis põhjendab evidentsiaalsuse käsitlemist 'episteemilisuse' katusdomeeni sees (vt Boye 2006).

Partitsiibikujulise evidentsiaalsuse päritolust

Peatükis 4.3 vaatlesin üksikasjalikult partitsiibipõhise evidentsiaalsuse võimalikke arenguteid. Kui oleviku partitsiibi arenemise küsimust refereeritud evidentsiaalsuse markeriks võib pidada lahendatuks, siis mineviku partitsiibi kujunemine samasuguseks markeriks jääb lahtiseks.

Samas peatükis eristasin kolme mineviku partitsiibi evidentsiaalsuse markeriks arenemise stsenaariumi: i) kõnelemist või vaimset seisundit tähistavaid verbe laiendava partitsiiparandi desubordineerimine, vt (2a)²¹, ii) abiverbi väljajätt verbi liitvormist, vt (2b), iii) jäänuk verbi käändeliste vormide kasutusest predikaadi põhivormidena soome-ugri algkeeles, vt (2c).

- (2) a) *Räägitakse, (et) Lembit tulnud koju.* → *Lembit tulnud* [ev.] *koju.*
b) *Lembit olevat/olla/on/oli tulnud koju.* → *Lembit tulnud* [ev.] *koju.*
c) *Lembit tulnud* [≈ predikaadi markeerimata vorm] *koju.* → *Lembit tulnud* [ev.] *koju.*

Nende kolme arenguraja vaatlusel tuginesin Kehayovi ja Siegli uurimusele (2007), mis pidas kolmandat stsenaariumi kõige ebatõenäolisemaks. Valik esimese kahe stsenaariumi vahel jääb lahtiseks, sest pooltargumente on mõlema kasuks. Esimese kasuks kõneleb asjaolu, et mineviku partitsiip markeerib refereeritud evidentsiaalsust ja seega on kergesti taandatav kõnelemist märkiva verbi laiendiks. Tõendusmaterjal teise hüpoteesi kasuks on kahetine. Esiteks näitavad andmed seotud narratiividest, et evidentsiaalse mineviku partitsiibi eitusvorm langeb tihti kokku indikatiivi liitaegade eitusvormidega, teiseks eraldatakse mineviku partitsiip tihti narratiivis indikatiivse liitaja vormist. Edasises uurimistöös tuleb arvestada ka võimaliku kumulatiivse mõjuga saksa keelest. Breitbarth on hiljuti väitnud, et abiverbi sagedane väljajätt ülemsaksa kirja-keeles 14.–18. sajandil korreleerub kontekstis väljendatud kaudse evidentsiaalsusega (vt Breitbarth 2006).

Lõpuks tuleb revideerida ka üht artiklites esinevat väidet, mis puudutab partitsiipide ja infinitiivide staatust eesti keeles evidentsiaalsuse markeritena. Juhindudes olukorrast Balkani keeltes, on need evidentsiaalsuse markerid arvatud evidentsiaalsusstrateegiate hulka (vt artikkel II: lk 129, 140; artikkel II: lk 818, 819, ja artikkel III: 910). Erinevalt Balkani keeltest on aga konstruktsioonil, milles nad esinevad, üksnes evidentsiaalne tähendus. See tähendab, et nende süntaktilisi omadusi arvestades tuleks neid lugeda puht-evidentsiaalimarkeriteks.

²¹ Lühend *ev.* 'evidentsiaalne'.

Appendix I. Aikhenvald's (2004: xxiv) classification of evidentiality systems (including only attested types)

Systems with two choices:

- A₁**. Firsthand and Non-firsthand
- A₂**. Non-firsthand versus 'everything else'
- A₃**. Reported (or 'hearsay') versus 'everything else'
- A₄**. Sensory evidence and Reported (or 'hearsay')
- A₅**. Auditory (acquired through hearing) versus 'everything else'

Systems with three choices:

- B₁**. Direct (or Visual), Inferred, Reported
- B₂**. Visual, Non-visual sensory, Inferred
- B₃**. Visual, Non-visual sensory, Reported
- B₄**. Non-visual sensory, Inferred, Reported
- B₅**. Reported, Quotative, and 'everything else'

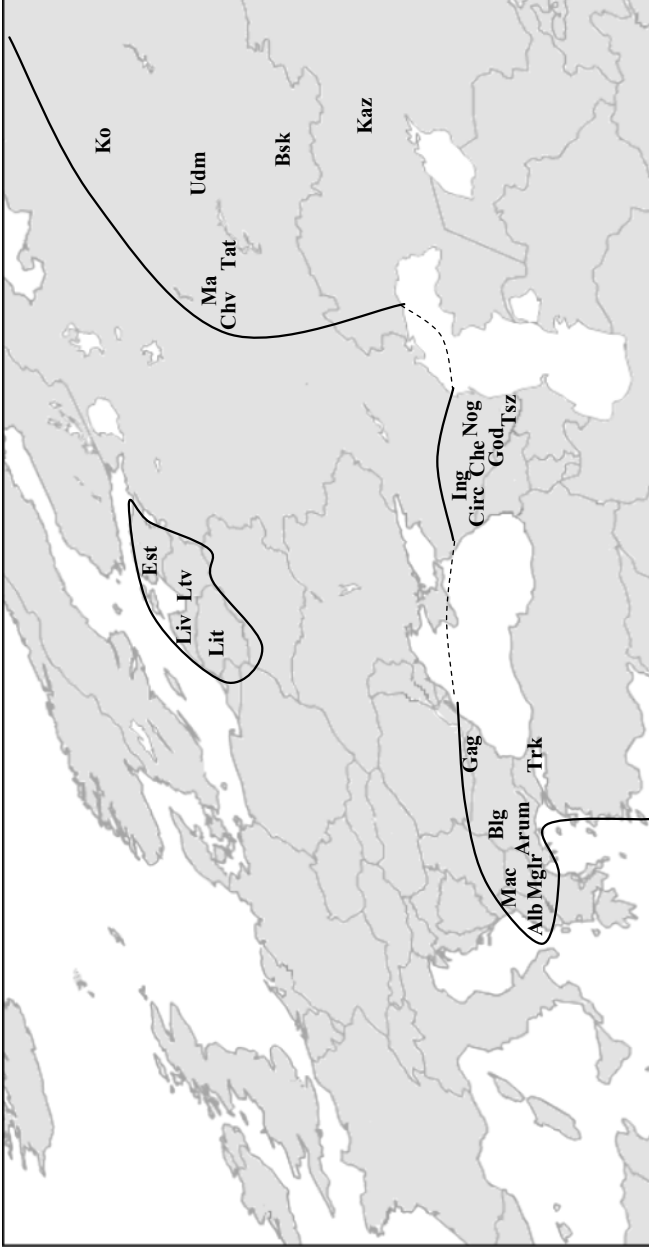
Systems with four choices:

- C₁**. Visual, Non-visual sensory, Inferred, Reported
- C₂**. Direct (or Visual), Inferred, Assumed, Reported
- C₃**. Direct, Inferred, Reported, Quotative

Systems with five choices:

- D₁**. Visual, Non-visual sensory, Inferred, Assumed, Reported

Appendix 2. Hotbeds of evidentiality in Europe



Appendix 3. Tabular data for parameter values

Legend: + the feature is present, (+) the feature is documented, but it is not central to the category, is very unfrequent or is restricted only to some dialects, – the feature is not present.

Table 1. Formal parameters

<i>EVIDENTIAL GRAMMEME</i>	Trk	Blg	Mac	Alb	Lit	Ltv	Est
PAST PARTICIPLE	+	+	+	+	+	+	+
PRESENT PARTICIPLE	–	–	–	–	+	+	+
OTHER NOMINALIZATION	–	–	–	–	–	(+)	+
COPULA ELLIPSIS	+	+	+	–	+	+	+
DEDICATED MORPHEME	–	–	–	+	–	+	+
DISTANCE PARTICLE	+	+	(+)	+	(+)	–	–
VOICE DISTINCTION	–	–	+	–	+	(+)	–

Table 2. Semantic and pragmatic parameters

<i>FUNCTION</i>	Trk	Blg	Mac	Alb	Lit	Ltv	Est
COVERS REPORTED	+	+	+	+	+	+	+
COVERS INFERENTIAL	+	+	+	+	+	–	–
COVERS MIRATIVE	+	+	+	+	+	–	–
TRIPATRITE DISTINCTION WITH UNMARKED MEMBER	+	–	+	–	–	–	–
SUBTYPE OF EVIDENTIALITY SPECIFIED	(+)	(+)	(+)	–	(+)	–	–
MARKS GENERIC STATEMENTS	–	–	–	+	–	–	–
USED IN REPORTS OF DREAMS	+	+	+	+	(+)	–	–
USED AS OBJECTIVIZER OF SENSORY PERCEPTIONS AND FEELINGS	–	–	–	+	–	–	–
CONVENTIONALIZED IN TRADITIONAL NARRATIVES	+	+	+	(+)	(+)	+	+

Table 3. Structural availability

<i>SYNTAGM</i>	Trk	Blg	Mac	Alb	Lit	Ltv	Est
IN DIRECT NON-ECHOIC QUESTIONS	+	+	+	+	(+)	-	-
IN REPORTED COMMANDS	(+)	+	+	+	-	+	-
WITH MORPHOLOGICAL MOOD	+	-	-	-	+	+	-

Appendix 4. Typological distances

Table 1. Number of shared values

	Trk					
Blg	17	Blg				
Mac	17	17	Mac			
Alb	12	14	12	Alb		
Lit	15	15	15	10	Lith	
Ltv	8	8	8	7	10	Ltv
Est	7	9	7	8	9	16

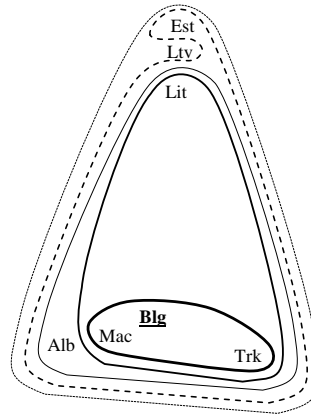
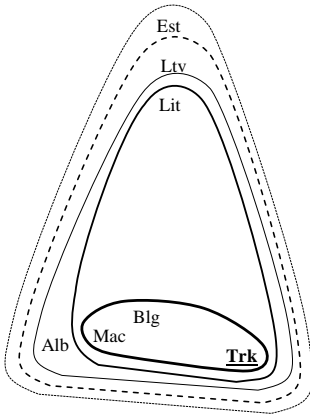


Figure 1. Degree of proximity to Turkish

- 17 features (shared values)
- 15 features
- 12 features
- - - 8 features
- 7 features

Figure 2. Degree of proximity to Bulgarian

- 17 features
- 15 features
- 14 features
- - - 9 features
- 8 features

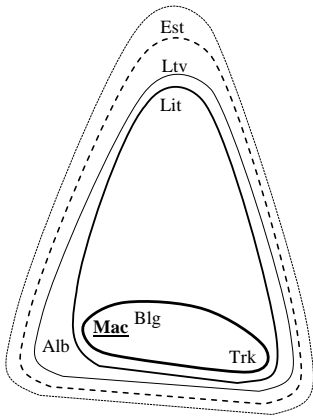


Figure 3. Degree of proximity to Macedonian

- 17 features
- 15 features
- 12 features
- - - 8 features
- 7 features

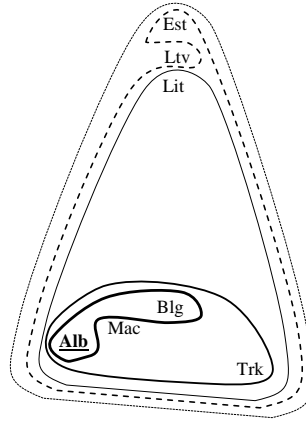


Figure 4. Degree of proximity to Albanian

- 14 features
- 12 features
- 10 features
- - - 8 features
- 7 features

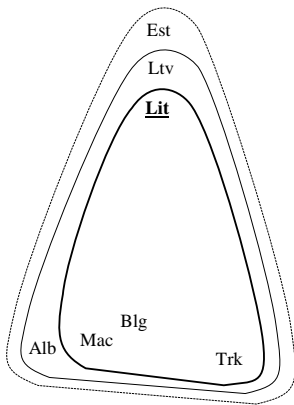


Figure 5. Degree of proximity to Lithuanian

- 15 features
- 10 features
- 9 features

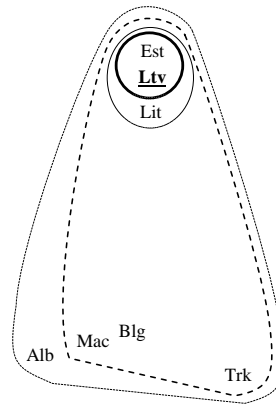


Figure 6. Degree of proximity to Latvian

- 16 features
- 10 features
- - - 8 features
- 7 features

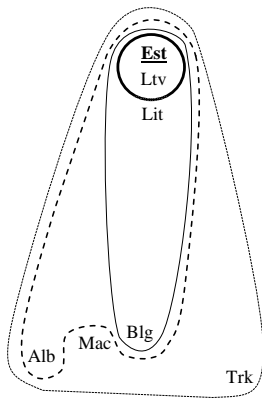


Figure 7. Degree of proximity to Estonian
 — 16 features
 - - - 9 features
 . . . 8 features
 - . - . 7 features

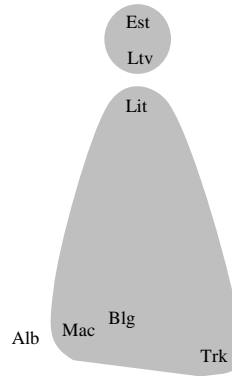


Figure 8. Similarity clusters sharing at least 15 features

CURRICULUM VITAE

PETAR KEHAYOV

Nationality: Bulgarian
Date of birth: 22.05.1972
Married
Address: University of Tartu, Department of Estonian Language,
Ülikooli 18–124, Tartu 50090, Estonia
Telephone: 737 5224
e-mail: petar.kehayov@ut.ee

Education

1990 Vasil Levski Secondary School, Sofia
2000 University of Tartu, Estonian as a Foreign Language
2003 University of Tartu, *magister artium*
from 2004 PhD student at the University of Tartu

Professional experience

2002–2003 University of Tartu, Chair of Uralic Languages, lecturer
2003–2004 University of Tartu, Chair of Estonian Language, specialist in
study
from 2004 University of Tartu, Institute of Estonian and General
Linguistics, research fellow

ELULOOKIRJELDUS

PETAR KEHAYOV

Kodakondsus: Bulgaaria
Sündinud: 22. mail 1972
abielus
Aadress: Tartu Ülikool, eesti keele osakond, Ülikooli 18–124, Tartu
50090, Eesti
Telefon: 737 5224
e-post: petar.kehayov@ut.ee

Haridus

1990 Vasil Levski Keskkool, Sofia
2000 Tartu Ülikool, eesti keel (võõrkeelena)
2003 Tartu Ülikool, *magister artium*
alates 2004 Tartu Ülikool, doktoriõpe

Teenistuskäik

2002–2003 Tartu Ülikool, uurali keelte õppetool, lektor
2003–2004 Tartu Ülikool, eesti keele õppetool, spetsialist
alates 2004 Tartu Ülikool, eesti ja üldkeeleteaduse instituut, teadur

DISSERTATIONES LINGUISTICAE UNIVERSITATIS TARTUENSIS

1. **Anna Verschik.** Estonian yiddish and its contacts with coterritorial languages. Tartu, 2000, 196 p.
2. **Silvi Tenjes.** Nonverbal means as regulators in communication: socio-cultural perspectives. Tartu, 2001, 214 p.
3. **Iiona Tragel.** Eesti keele tuumverbid. Tartu, 2003, 196 p.
4. **Einar Meister.** Promoting Estonian speech technology: from resources to prototypes. Tartu, 2003, 217 p.
5. **Ene Vainik.** Lexical knowledge of emotions: the structure, variability and semantics of the Estonian emotion vocabulary. Tartu, 2004, 166 p.
6. **Heili Orav.** Isiksuseomaduste sõnavara semantika eesti keeles. Tartu, 2006, 175 p.
7. **Larissa Degel.** Intellektuaalsfäär intellektuaalseid võimeid tähistavate sõnade kasutuse põhjal eesti ja vene keeles. Tartu, 2007, 225 p.
8. **Meelis Mihkla.** Kõne ajalise struktuuri modelleerimine eestikeelsele tekst-kõne sünteesile. Modelling the temporal structure of speech for the Estonian text-to-speech synthesis. Tartu, 2007, 176 p.
9. **Mari Uusküla.** Basic colour terms in Finno-Ugric and Slavonic languages: myths and facts. Tartu, 2008, 207 p.