A Computational Explanatory Combinatorial Dictionary of Religious Terminology in Early Baltic Catechisms*

Silvia Piccini[1,*,†], Adriano Cerri[2,†], and Pietro U. Dini[2]

[1] Institute of Computational Linguistics "A. Zampolli", CNR, via G. Moruzzi 1, 56124 Pisa, Italy

[2] University of Pisa, via S. Maria 36, 56126 Pisa, Italy

Abstract

This study, part of the Italian PRIN project "Old Words for a New World: Translating Christianity to Baltic Pagans", aims to illustrate a computational Explanatory and Combinatorial Dictionary (ECD) of the religious lexicon extracted from the earliest Baltic catechisms (Old Prussian, Old Lithuanian and Old Latvian). The ECD will offer insights into the linguistic and cultural transformations spurred by Christianization of Europe's last pagans, illuminate the translational strategies of early Baltic translators, and enable systematic comparisons of translation choices across languages and confessional traditions.

Keywords

Explanatory Combinatorial Dictionary, semantic web, religious terminology, Old Lithuanian, Old Latvian, Old Prussian, Catechisms, Christianization

1. Introduction

The conversion to Christianity in the Baltic region was an epochal event of the Middle Ages, profoundly transforming the political, social, and cultural organization of these societies [1, 2, 3]. Lithuanians, Latvians, and Prussians, as the last pagans of Europe, underwent conquest and gradual Christianization by the Teutonic Order, a process formally completed in 1387 with the baptism of the Grand Duke of Lithuania Jogaila.

Christianization not only reshaped societal structures but also left a deep imprint on language and culture. The introduction of Christianity catalyzed the creation of written forms of Baltic languages and the development of early literary works, such as catechisms [4, 5]. This shift also necessitated the emergence of a specialized religious lexicon, as new theological concepts demanded corresponding terminology. Indeed, conversion to a new belief system represents an extraordinary cultural revolution, introducing novel notions, realia, and logics that demand an immediate linguistic response: new terms arise to express these new concepts, while obsolete terms fade with the concepts they once designated.

Despite the valuable insights offered by existing studies on the Baltic religious lexicon (e.g., [6]), there has been little systematic research focusing on how Christian terminology was shaped and integrated into the linguistic framework of these communities. The research we present here, conducted within the framework of the Italian PRIN project "Old Words for a New World: Translating Christianity to Baltic Pagans"[2], aims to systematically map these processes by

[2] https://www.ilc.cnr.it/progetti/ownw/.

constructing a computational Explanatory and Combinatorial Dictionary (ECD) dedicated to the religious lexicon of the earliest Baltic catechisms.

The dictionary seeks to illuminate the profound linguistic and cultural transformations triggered by Christianization and help scholars to understand how Baltic pagan communities reacted to the introduction of Christianity. In addition to its utility for documenting and analyzing the religious lexicon, the ECD will provide a crucial foundation for studying the translational strategies employed by early Baltic translators, for a systematic comparison of translation choices across languages and between confessional traditions.

2. Corpus for Terminological Analysis

The analysis of religious terminology in the Baltic languages is based on a selected corpus of five catechisms, dating to the second half of the 16th century (see Table 1). These texts were chosen based on several key criteria that align with the scope and objectives of this study.

Language	Translator(s), Title	Year	Confession	Nr. of Tokens
Old Prussian	A. Will – P. Megott, *Enchiridion*	1561	Lutheran	6.712
Old Lithuanian	B. Vilentas, *Enchiridion*	1579	Lutheran	8.987
	M. Daukša, *Kathechismas arba mokslas kiekwienam krikszczioni priwalvs*	1595	Catholic	13.052
Old Latvian	Anon., *Catechismvs Catholicorum*	1585	Catholic	4.869
	Rivius et al., *Enchiridion*	1586	Lutheran	7.513

Table 1. The Selected Corpus of Early Baltic Catechisms

The first criterion is chronological. All five catechisms are part of the very early phase of the Baltic written tradition, preceding the emergence of secular literature. As such, they provide some of the earliest attestations to the uptake of Christianity among the Balts, offering a unique glimpse into the initial stages of Christianization. While the first crusades in the Baltic region date back to the 13th century, these catechisms reflect the subsequent and pivotal moment when Christianity began to be integrated into Baltic culture.

The second criterion is typological. Although other religious texts from the same period, such as postils, translations of psalms, and hymns, exist, catechisms stand out as the most comparable and suitable for the present study. Three of the selected catechisms — Abel Will's Enchiridion (1561) in Old Prussian, Vilentas' Catechism (1579) in Old Lithuanian, and the Rivius-Enchiridion (1586) in Old Latvian — represent a unicum in Baltic philology: these are the only parallel texts we possess in the three languages, as they are all translations of the same source-text, namely Luther's Small Catechism. This makes them invaluable for a comparative study of religious terminology across different linguistic traditions.

Out of the five selected catechisms, three belong to the Lutheran tradition, reflecting the historical circumstances of the region's Christianization. However, to ensure broader confessional representation, we have also included two Catholic catechisms: the Old Latvian anonymous *Catechismvs Catholicorum* (1585), often referred to as Canisius' Catechism, and Daukša's Catechism (1595) in Old Lithuanian. As for Old Prussian, no Catholic catechisms were ever published.

This diverse collection of texts is representative of the three major Baltic languages — Old Prussian, Old Lithuanian, and Old Latvian — and provides a balanced representation of both the Lutheran and Catholic confessions in the region.

3. The Explanatory Combinatorial Dictionary for Baltic Catechisms

The extraction of the terminological repertoire from the Catechisms was performed manually. This approach, made feasible by the relatively limited length of the documents, was primarily driven by the significant orthographic variability characterizing these texts. Following extraction, the terminological repertoire is being formalized in the form of an ECD, adhering to the principles of Explanatory and Combinatorial Lexicology (ECL). This model, developed by Mel'čuk and his colleagues as an integral component of the Meaning-Text Theory [7], was chosen for several compelling reasons.

Firstly, the ECD allows for a rigorous, exhaustive, and precise description of both the semantics and combinatorics of lexical units, making it uniquely suited for capturing the nuances of religious terminology. Secondly, this model has been successfully applied to describe the vocabulary of specialized domains [8, 9, 10, 11][3], demonstrating its flexibility and methodological robustness. Thirdly, the ECD is particularly appropriate for this project, as it sheds light on semantic, syntactic, and morphological mismatches between languages. This is crucial for understanding the translational dynamics and motivations underlying the choices made by Baltic translators when adapting Christian texts, often from German originals.

According to the principles of ECL, terms in our dictionary are provided with:

i) Rigorous lexicographic definitions: Each term is defined using a pre-established formal metalanguage. It is important to emphasize that this definition differs from the conceptual definition, which is formalized in a separate ontology using OWL. This distinction underscores the foundational assumption of this work: terminology operates as a dual discipline [12, 13, 14], encompassing both linguistic and conceptual dimensions. As a result, the semantic features that constitute the sense of a term do not necessarily align with the characteristics of the corresponding concept [15].

ii) Morphosyntactic information: For predicative lexical units, the Government Pattern is specified, detailing the surface realizations of the semantic actants, if any, introduced in the definition.

iii) Lexical Functions: A wide range of paradigmatic and syntagmatic relationships is described through Lexical Functions. These include collocations, idiomatic expressions, and lexical derivations, which facilitate the study of translational aspects, as will be shown in the following section.

3.1. Technological Implementation of the Baltic ECD

To ensure that the dictionary is machine-readable and adheres to FAIR (Findable, Accessible, Interoperable, and Reusable) principles [16], semantic web technologies [17] and Linked Data paradigm [18] were employed. Specifically, the dictionary is developed using the OntoLex-Lemon model [19], which is widely recognized as the de facto standard in computational lexicography. Within this framework, the Core module (also referred to as OntoLex), as well as the Lexicog and Lexfom modules have been applied. Both Lexicog and Lexfom integrate with the Core: the former is used to formalize structures and annotations typical of dictionaries and lexicographic resources, while the latter formalizes lexical functions. The SynSem module will also be incorporated in the future to address syntactic aspects.

---

[3] It should also be noted that at the *Observatoire de linguistique Sens-Texte* (OLST) of the University of Montréal, several terminological dictionaries have been developed, including DiCoEnviro (focused on environmental terminology), DiCoInfo (focused on IT terminology), DiCoLexitrans (focused on legal terminology), among others.

Below, a user-friendly representation of an entry from the Old Lithuanian dictionary, *venčiavoti* 'to marry' (Figure 1):
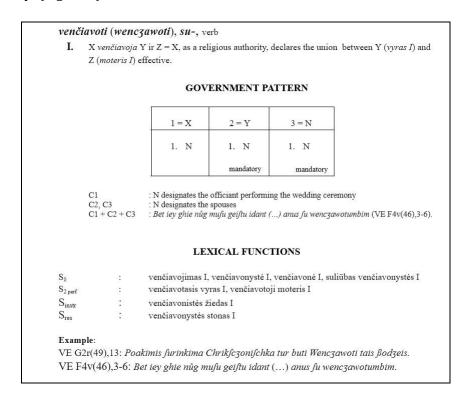

Figure 1: The Old Lithuanian dictionary entry *venčiavoti* 'to marry'

The entry begins with a lexicographical definition in propositional form, following the principles of Explanatory and Combinatorial Lexicology: "*X venčiavoja Y ir Z* = X, as a religious authority, declares the union between Y (*vyras I*) and Z (*moteris I*) effective." The morphosyntactic realization of the three semantic actants, X, Y and Z, is outlined in the Government Pattern section. All three are expressed as nouns: X refers to the officiant, while Y and Z designate the spouses, with additional semantic constraints specified in the same section.. An excerpt from Vilentas' catechism illustrates all the actants in context.

The Lexical Functions section describes paradigmatic and syntagmatic relationships, focusing on collocations and lexical derivation. Four nominal derivatives (S_0) are identified: *venčiavojimas I, venčiavonystė I, venčiavonė I, suliūbas venčiavonystės I* ('marriage').

Lexical function S_{2perf} indicates the patient (e.g., *venčiavotasis vyras I* 'groom' and *venčiavotoji moteris I* 'bride') introduced by the verb *venčiavoti*. Additionally, the lexical functions S_{instr} and S_{res} denote nominal circumstantial derivatives: S_{instr} represents the instrument associated with the lexical unit (*venčiavonistės žiedas I* 'the wedding ring'), while S_{res} refers to the resulting condition after the marital ritual (*venčiavonystės stonas I* 'marital status').

Below, in Figure 2, is the RDF formalization of the Old Lithuanian dictionary entry *venčiavoti* 'to marry' according to the aforementioned vocabularies.

```
:venčiavoti_lex_entry a ontolex:LexicalEntry ;
    lexinfo:partOfSpeech lexinfo:verb ;
    ontolex:sense :venčiavoti_sense_I ;
    rdfs:label "venčiavoti"@lit .

:venčiavoti_dict_entry a lexicog:Entry ;
    rdf:_1 :venčiavoti_entry_comp.

:venčiavoti_entry_comp a lexicog:LexicographicComponent ;
    lexicog:describes :venčiavoti_lex_entry ;
    rdf:_1 :venčiavoti_sense_I_comp .

:venčiavoti_sense_I_comp a lexicog:LexicographicComponent ;
    rdfs:label "I" ;
    lexicog:describes :venčiavoti_sense_I .

:venčiavoti_sense_I a ontolex:LexicalSense ;
    skos:definition "X venčiavoja Y ir Z
     = X, as a religious authority, declares the union
    between Y (vyras I) and Y (moteris I)effective.] " ;
    rdfs:seeAlso  :vyras_sense_I ,
                  :moteris_sense_I.

:venčiavoti_sense_I lf:S₀      :venčiavojimas_senseI,
                               :venčiavonystė_senseI,
                               :venčiavonė_senseI,
                               :suliūbas venčiavonystės_senseI.

:venčiavoti_sense_I lf:S₂perf  : venčiavotasis_vyras_senseI,
                               : venčiavotoji_moteris_senseI.

:venčiavoti_sense_I lf:Sinstr  : venčiavonystės_žiedas_senseI.

:venčiavoti_sense_I lf:Sres    : venčiavonystės stonas I_senseI.
```

Figure 2: The RDF formalization of the Old Lithuanian dictionary entry *venčiavoti* 'to marry'

4. ECD for Ancient Languages: Issues and Benefits

Constructing an ECD for ancient languages is an innovative and experimental endeavor that poses unique challenges while offering significant advantages. One of the primary difficulties lies in the impossibility of relying on native speaker intuition or linguistic competence. This limitation affects all stages of the process, especially the formulation of definitions. Unlike modern ECDs, where definitions are typically written in the language of the dictionary, we have opted to use English, supplementing it with corresponding Baltic terms in parentheses wherever they are attested in the source material.

Another key innovation lies in redefining the purpose of the ECD. While traditional ECDs are designed primarily to support speech production, the main objective of our ECD is to provide scholars with tools to analyze and deeply understand the translational techniques employed by the original translators of the early Baltic catechisms. By identifying potential losses, additions, or semantic transformations, lexical functions reveal the translational dynamics at play. This analysis is particularly valuable in determining whether translators adhered to or deviated from the collocational norms of the target language, uncovering instances of interlinguistic influence or deliberate stylistic and doctrinal choices. Additionally, lexical functions provide insights into whether translators prioritized literal fidelity (form) or semantic fidelity (content), when adapting theological concepts across different linguistic and cultural contexts.

We propose here a brief example to illustrate this. In Luther's *Small Catechism*, the entry of spouses into the new marital state is expressed as follows: *Hans. N. und Greta. N. wollen (…) zum heyligen stande der ehe greiffen* ('Hans N. and Greta N. desire (…) to enter [lit. 'seize, grab'] the holy estate of marriage'). Baltic translators, however, appear to have adopted slightly different

strategies, as demonstrated below (Table 2). The three corresponding entries for 'marriage' (OLith. *suliūbas venčiavonystės,* OPr. *salaūban,* and OLatv. *laulība*) exhibit a complex lexical function (*IncepOper*)[4], highlighting how the transition into the marital state was rendered by the three Baltic translators.

Old Lithuanian	IncepOper$_1$(*suliūbas venčiavonystės*): *įeiti į [~ą venčiavonystės], nusiduoti į [~ą venčiavonystės]*
Old Latvian	IncepOper$_1$ (*laulība*): *iekšan [DEM ~<e>] doties*
Old Prussian	IncepOper$_1$(*salaūban*): *preiſtan Swintan buſennien ſteiſe [~] kackint*

Table 2: The lexical function *IncepOper$_1$* in the three Baltic ECD entries.

Specifically, Lithuanian and Latvian employ motion verbs combined with a directional preposition (*į* in Lithuanian and *iekš/iekšan* in Latvian). Unlike, Old Prussian aligns more closely with the German construction, employing a telic verb (*kackint* 'to reach') that conveys the attainment of a new state, implying either a physical or metaphorical movement. Furthermore, in Old Prussian, the attainment of the marital state is emphasized by the noun *buſennien* (corresponding to the German *Stand* 'condition'), a feature absent in the Lithuanian and Latvian renditions.

As mentioned earlier, it is also important to emphasize another innovative aspect of this work: the linguistic descriptions provided in the ECD are complemented by a formal computational ontology. This dual-level approach distinguishes between linguistic and conceptual representations, shedding light on how translators addressed lexical or conceptual gaps. When a language lacks a direct equivalent for a theological term or concept, the ECD reveals alternative strategies, such as explanatory descriptions, metaphors, or the creation of new terms. For example, the concept of <sacrament> is linked in the ontology through the *ontolex:reference* property to the entries in the three dictionaries. This immediately highlights how, in relation to the same concept, the Prussian translator opted to adopt a loanword from German (*sacramentan*), while the Lithuanian and Latvian translators chose indigenous terms. However, if the Lithuanian term *paslaptinė*[5] refers to the concept of *mysterion*, in contrast, the Latvian translator used *lielā mīlestība* ('great loving-kindness'), a term that appears to be fully in line with Luther's writings. This term symbolizes the union of Christ with his Church as an act of divine love, allegorically represented by the marriage between a man and a woman [4].

5. Conclusions

The present paper has outlined the initial steps in the construction of a computational ECD dedicated to the religious terminology of the early Baltic catechisms. The ECD approach is especially relevant, as it enables a detailed analysis of how lexical items were adapted, reinterpreted, or innovated to accommodate new Christian concepts in languages deeply rooted in their pagan heritage. In the future, the computational formalization of the entries will be further developed to include aspects of the Government Pattern, which have not yet been considered in the current stage.

---

[4] This complex lexical function consists of the lexical function *Incep*, which indicates the initial phase of the process, and the lexical function *Oper*, which primarily acts as a support verb.

[5] This word is an *o*-grade derivative from the Lithuanian root *slep-*, cf. vb. *slėpti* 'cover up, hide, conceal', cf. [20].

Acknowledgments

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] N. Blomkvist, The Discovery of the Baltic. The Reception of a Catholic World-System in the European North (AD 1075-1225), Brill, Leiden / Boston, 2005.

[2] F. Young (Ed.), Pagans in the Early Modern Baltic: Sixteenth-Century Ethnographic Accounts of Baltic Paganism, Arc Humanities Press, Leeds, 2022.

[3] E. Christiansen, The Northern Crusades, Macmillan, London, 1980.

[4] E. Castelli, A. Cerri, P.U. Dini, M. Monaca, M. Mormino, S. Piccini, Early Baltic Catechisms in the Transition from Paganism to Christianity: Showcasing an Interdisciplinary Research Project, in: D. Bronner, E. Parina (Eds.), Contents of Faith in Transfer, J.B. Metzler, forthcoming.

[5] P. U. Dini, The Earliest Catechisms in the Baltics: Cultural Influences and Translation Issues in the 16th Century, in: D. Bronner, E. Parina (Eds.), Contents of Faith in Transfer, J.B. Metzler, forthcoming.

[6] N. Vėlius, Baltų religijos ir mitologijos šaltiniai, 4 volls., Mokslo ir enciklopedijų leidybos institutas, Vilnius, 1996–2005.

[7] I. Mel'čuk, A. Clas, A. Polguère. Introduction à la lexicologie explicative et combinatoire, Duculot, Bruxelles, 1995.

[8] J. Dancette J., Les relations lexico-sémantiques dans un dictionnaire spécialisé, in: Th. Szende (Ed.), Le français dans les dictionnaires bilingues, Champion/Slatkine, Paris et Genève, 2006, pp. 144-156.

[9] M.C. L'Homme, Fonctions lexicales pour représenter les relations sémantiques entre termes, Traitement automatique de la langue (TAL) (2002), 43/1: 19-41.

[10] M.C. L'Homme, Using Explanatory and Combinatorial Lexicology to Describe Terms, in: L. Wanner (Ed.), Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In honour of Igor Mel'čuk, Benjamin's Academic Publishers, Amsterdam, 2007, pp. 167-202.

[11] M.C. L'Homme, Lexical Semantics for Terminology: An introduction, John Benjamins, Amsterdam 2020.

[12] S. Piccini, Eppur si muove! Per una rappresentazione formale della variazione diacronica in terminologia, in: C. Burgassi, E. Guadagnini et G. Vaccaro (Eds.), Storie di idee nell'Europa mediterranea. Trasmissione delle conoscenze nel Medioevo e nella prima Età moderna, CNR-Edizioni, Roma, 2024, pp. 569-602.

[13] S. Piccini Silvia, E. Giovannetti, E. Corsi, Une ressource termino-ontologique bilingue chinois-italien : le cas de la traduction de la Mappemonde de Matteo Ricci par Pasquale D'Elia, in C. Roche (Ed.), Terminologie & Ontologie : Théories et Applications, Série : Terminologica, Actes de la conférence Toth 2017, Éditions de l'Université Savoie Mont Blanc, Chambéry, 2018, pp. 33-49.

[14] C. Roche, Ontoterminology : how to unify terminology and ontology into a single paradigm, in: N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J.

Odijk et S. Piperidis (Eds.), Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), Istanbul, European Language Resources Association (ELRA), 2002, pp. 2626-2630.

[15] S. Piccini, D. Saponaro, E. G. Vilela Ruiz, La connotation logique en terminologie. Le mariage dans l'Israël antique comme étude de cas, Cahiers de lexicologie (2024), 125 - 2: 103-129.

[16] M. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, vol. 3, 2016.

[17] T. Berners-Lee, J. Hendler, O. Lassila, O., The semantic web, Scientific American (2001), 284(5): 34-43.

[18] Ch. Bizer, T. Heath, T. Berners-Lee, Linked data: The story so far, in: A. Sheth (Ed.), Semantic services, interoperability and web applications: emerging concepts, 1st ed., IGI Global, Hershey, Pennsylvania, 2011, pp. 205-227.

[19] Ph. Cimiano, J. P. McCrae, P. Buitelaar, Lexicon model for ontologies: Community report. W3C Ontology-Lexicon Community Group, 2016.

[20] W. Smoczyński, Lithuanian Etymological Dictionary, Peter Lang, Berlin / Bern / Bruxelles / New York / Oxford / Warszawa / Wien, 2018.