



HAL
open science

Verbless and zero-predicate sentences : an English and Russian contrastive corpus study

Antonina Bondarenko

► **To cite this version:**

Antonina Bondarenko. Verbless and zero-predicate sentences : an English and Russian contrastive corpus study. Linguistics. Université Paris Cité, 2021. English. NNT : 2021UNIP7110 . tel-03706573

HAL Id: tel-03706573

<https://theses.hal.science/tel-03706573>

Submitted on 27 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

École doctorale Sciences du Langage 622

Laboratoire CLILLAC-ARP et SEDYL

Verbless and Zero-Predicate Sentences: *An English and Russian Contrastive Corpus Study*

Par Antonina Bondarenko

Thèse de doctorat de Linguistique

Dirigée par Agnès Celle
Et par Christine Bonnot

Présentée et soutenue publiquement le 13 Décembre 2021

Devant un jury composé de :

Christine BONNOT	<i>(INALCO, Professeur des Universités)</i>	Directrice
Agnès CELLE	<i>(Université de Paris, Professeur des Universités)</i>	Directrice
Eric CORRE	<i>(Université Sorbonne Nouvelle, Professeur des Universités)</i>	Rapporteur
Dagmar DIVJAK	<i>(University of Birmingham, Professor)</i>	Rapporteur
Serge FLEURY	<i>(Université Sorbonne Nouvelle, Maître de Conférences)</i>	Examineur
Rudy LOOCK	<i>(Université de Lille, Professeur des Universités)</i>	Examineur
Graham RANGER	<i>(Université d'Avignon, Professeur des Universités)</i>	Examineur



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support and generosity of a great many people to whom I would like to express my sincere gratitude.

First and foremost, I am deeply grateful to my research supervisors, Professor Agnès Celle and Professor Christine Bonnot, for their absolutely invaluable guidance.

Words cannot do justice to all that I owe to Agnès Celle. Thank you so much for all of the stimulating discussions, for your openness and attentiveness, for your invaluable expertise and passion for discovery. You have inspired, positively challenged and securely supported me in every way during this climb, from the first day of my Masters through the present thesis. I cannot thank you enough for always being there for me, no matter the issue nor the time, for your investment into my academic and professional development and your confidence in me. I could not have dreamt of a better research director.

I am also sincerely grateful to my co-supervisor, Christine Bonnot, for sharing her expertise and passion for Russian linguistics and helping me to navigate in this domain. Thank you for your insights and introducing me to new perspectives, for helping me to sharpen my arguments and improve the quality of my writing. Thank you also for integrating me into the Russian formal linguistics seminars at INALCO and for warmly welcoming me into the SeDyL laboratory.

I would like to thank Université Paris Diderot, École Doctorale de Sciences du Langage (ED622) and the Clillac-Arp laboratory for financing the present thesis, offering me a PhD contract and an ATER position, and for generously contributing to the development of my research and making possible the many important international conferences, seminars and workshops. Thank you to Labex-EFL for providing the funding for the international summer-

schools which had a pivotal impact on the present work. I am also very grateful to Université Sorbonne Paris Cité and the selection committee for awarding me with the MIEM scholarship during my Masters studies and motivating this doctoral project.

I am enormously honoured that Professors Eric Corre, Dagmar Divjak, Serge Fleury, Rudy Loock and Graham Ranger have accepted to be a part of my defence jury. Thank you sincerely for making the time to evaluate my work and for all of your attention and precious feedback.

I am also incredibly lucky to have had the opportunity to work with Professors Sylvie Archambault, Serge Fleury and Mathieu Valette as part of my thesis advisory committee. Your expertise, generous feedback, and enthusiasm for this project provided me with powerful motivation and have been essential for the development of this thesis. I am very grateful to Sylvie Archambault for all of the very helpful insights and suggestions, attention and warm support. I am forever indebted to Serge Fleury for all of the intensive work to help turn this dream into a reality. Thank you for always taking the time to help, for believing in this project and in my capacities, and for providing me with a wealth of inspiration. A very special thank you to Mathieu Valette for a highly engaging Masters seminar in corpus semantics that was very influential on me, and for the consistent encouragement of this research project from its very conception.

I am forever grateful to Serge Fleury and Maria Zimina for the powerful Trameur, without which this study would have been impossible. I am also eternally indebted to Elena, Oleg and Anton Borrisov for their excitement and investment in this project, precious time, colossal help with the scripts and immense support. I also owe a very special thank you to Loïc Liégeois for the initial script that set the automatization process in motion.

I would also like to thank all of the members of my two labs, Clillac-Arp and SeDyL, for their support during these years. I am especially grateful to Isabelle Léglise whose help and feedback was invaluable for the realization of this thesis. A very special thank you to Sophie Vassilaki and Ioana Chitoran whose comments and enthusiasm for this project really meant a lot to me. I also owe a big thank you to Anne Jugnet, Rèbh Charifi, and everyone at the UFR d'Études Anglophones, as well as Duna Troiani and Laurence Guernalec, for all of their help with administrative and scheduling issues and their kindness and energy.

A big thank you to Ruslan Mitkov, Graham Ranger, Sebastian Hoffmann, Martin Weisser, Laurent Gosselin, Eric Corre, Nicolas Ballier, Laure Lansari, Lisa Brunetti and Lucia Tovenà for their feedback and encouragement. Thank you very much to Olga Shchetinkova,

Ljudmila Dorofeeva, Oksana Koroteeva, Irina Bahtijarova and Almira Tatkenova for useful and inspiring conversations about Russian literature and linguistics.

I would also like to address special thanks to everyone I met and had the chance to discuss this project with at seminars, conferences, at the Université de Louvain and University of Birmingham summer-schools and during the workshops at the University of Vienna, University of Tartu and University of Konstanz.

I would like to express my gratitude to all of my teachers through the years, including all of my linguistics professors at Université de Paris and at the University of Arizona online class. I am especially grateful to my philosophy professors Spiro Panagiotou, Alison Miculan, Brigitte Sassen, Sébastien Gandon and my French professor Corina Sandu from McMaster University for their inspiration and encouragement.

I am eternally grateful to my French family without whom this thesis would not have been possible – Gilles Charles Dockerty, Alain and Françoise Souillard, Sandrine Boussard-Nilly, Alain Lapert, Gwendal Demeslay, and all of my friends, for always being there for me.

Finally, I could not have completed this dissertation without the unconditional support of my family. The fearless sacrifices that my parents and grandparents have made for me deserve a separate book, and it is to them that I dedicate this thesis.

CONTENTS

ACKNOWLEDGEMENTS	i
CONTENTS	1
ABSTRACTS	9
ENGLISH	10
RUSSIAN	14
FRENCH	17
PRELIMINARIES	20
TRANSLITERATION	21
GLOSSING ABBREVIATIONS	22
EXAMPLE GUIDELINES	24

PART ZERO: INTRODUCTION

MOTIVATION AND ROAD MAP	26
--------------------------------------	-----------

PART ONE: THEORETICAL FRAMEWORK

CHAPTER 1: THE PHENOMENON	40
CHAPTER 2: PREDICATION AND SENTENTIAL STATUS	42
2.1 Modern Logic	49
2.2 Questioning Equality from a Linguistic Perspective	51
2.3 The Russian Linguistic Tradition	59
2.3.1 The Verb-centric Syntactic Model in Russia	60
2.3.2 Non-Verb-centric and Other Sentential Models in Russia	67

2.3.2.1	a. Lifting Morphological Restrictions	67
2.3.2.2	b. New Part of Speech	68
2.3.2.3	c. One-Part / Two-Part Sentence Typologies.....	74
CHAPTER 3: ELLIPTICAL SENTENCES AND NONSENTENTIALS		82
3.1	The Notion of Ellipsis: Its Importance and Versions.....	82
3.2	Syntactic Sentence Structure	84
3.3	Syntactic Ellipsis.....	85
3.4	Arguments Against Syntactic Ellipsis	87
3.4.1	Argument 1 – <i>Elliptical structures behave differently</i>	87
3.4.2	Argument 2 – <i>Various possibilities for reconstruction without antecedent</i>	88
3.4.3	Argument 3 – <i>Syntactic structure is what it appears: Antecedent mismatch</i>	89
3.4.4	Argument 4 – <i>Hidden structure is too costly</i>	90
3.5	Objections to Nonsententials.....	92
3.5.1	Objection 1 – <i>Underemphasizing the fall of hidden structure: Self-sufficiency</i>	92
3.5.2	Objection 2 – <i>Sentence as a top-down syntactic category for all phrases</i>	92
3.5.3	Objection 3 – <i>Agreement</i>	93
3.5.4	Objection 4 – <i>Distinguishing verbless embedded phrases</i>	95
3.5.5	Objection 5 – <i>Distinguishing verbal embedded phrases</i>	96
3.5.6	Objection 6 – <i>False dichotomy between Elliptical Sentence & Nonsentential hypotheses</i>	96
3.5.7	Objection 7 – <i>Nonsentential speech acts: Unresolved issues</i>	97
3.5.7.1	a. Potential semantic differences	98
3.5.7.2	b. Illocutionary force.....	99
3.5.7.2.1	Step 1: Uttered tokens and linguistic types	99
3.5.7.2.1.1	Objections to Step 1: Type-token divorce.....	100
3.5.7.2.2	Step 2: Context-free analysis	101
3.5.7.2.2.1	Objections to Step 2: Illocutionary force without context.....	102
3.5.7.2.3	Illocutionary force cross-linguistically: Stability considerations	103
3.5.7.2.3.1	Present tense copula in Russian is not an exception	105
3.5.7.3	c. Sentential status: Fight or flight.....	112
3.6	Semantic Ellipsis.....	114
3.6.1	Problem 1: Choosing the propositional meaning within a single context	116
3.6.2	Problem 2: Multiple propositional meanings in different contexts.....	122
3.6.3	Problem 3: Reintroduction changes the meaning	127
3.6.4	Problem 4: Inconvenience of encoded meanings.....	130
3.6.5	Problem 5: The proposition, its encoding, and sentential status	131
3.7	Constructionist Approaches.....	133
3.7.1	Advantages for verbless constructions	135
3.7.2	Disadvantages for verbless constructs.....	136
3.8	Summary	141

CHAPTER 4: VERBLESS SENTENCES.....	142
4.1 Preliminaries.....	143
4.1.1 Sentence vs. Clause	143
4.1.2 Sentence vs. Sub-Sentential Utterance Unit	143
4.2 Initial and Final Marking.....	144
4.3 Contextual Insertion.....	147
4.3.1 Information Structure	148
4.3.1.1 Information	148
4.3.1.2 Topic.....	150
4.3.1.2.1 Aboutness.....	150
4.3.1.2.2 Revealed by Discourse Context.....	151
4.3.1.2.3 Not Old Information	152
4.3.1.2.4 Precise Criteria	152
4.3.1.3 Focus	155
4.3.1.3.1 Fundamental to the Sentence.....	155
4.3.1.3.2 Not a Complement of the Topic.....	155
4.3.1.3.3 Not New Information	156
4.3.1.3.4 Focus Relation May Include Old Constituents	157
4.3.1.4 Focus Structure Types.....	158
4.3.1.4.1 Sentence Type 1: Topic-Comment	158
4.3.1.4.1.1 Criteria	158
4.3.1.4.1.2 Unmarked.....	159
4.3.1.4.1.3 Illustration	160
4.3.1.4.2 Sentence Type 2: Identificational.....	161
4.3.1.4.2.1 Criteria	162
4.3.1.4.2.2 Illustration	163
4.3.1.4.3 Sentence Type 3: Thetic.....	164
4.3.1.4.3.1 Criteria	165
4.3.1.4.3.2 Illustration	168
4.3.1.5 Summary of Information Structure	170
4.3.2 Predicational Support.....	172
4.3.2.1 The Explicitness Debate.....	172
4.3.2.2 Predication with Dummy Subjects & Extralinguistic Support	173
4.3.2.3 Predication in One-Word Utterances & Extralinguistic Support	176
4.3.2.3.1 In the Present Framework.....	177
4.3.2.3.2 Illustration	178
4.3.2.3.2.1 Vocatives	178
4.3.2.3.2.2 Insults	180
4.3.2.4 Predication Supported by Linguistic Context	182
4.3.3 The Sentence.....	184
4.4 Absence of Verb.....	185
4.4.1 Finite & Non-finite: Verbless as a Subtype of Predicate-less.....	185
4.4.2 Entirely Verbless: Compound.....	187

4.5	Two Fundamental Types of Verbless Sentences	188
4.5.1	Elliptical Verbless Sentences	189
4.5.1.1	Sluicing	189
4.5.1.2	Stripping.....	190
4.5.1.3	Gapping.....	191
4.5.1.4	Antecedent	192
4.5.2	Non-Elliptical Verbless Sentences	193
4.5.3	The Zero-Predicate	194
4.6	The Meaning of Verbless Sentences	199
4.6.1	Implicature	203
4.6.2	Assertion.....	206
4.6.3	Summary	211
4.7	Conclusion: Proposed Framework	212

PART TWO: METHODOLOGY

CHAPTER 1: AN INTERDISCIPLINARY APPROACH	216
CHAPTER 2: THIRST FOR CORPORA	217
2.1 The Challenge	217
2.1.1 Grammatical Structure	218
2.1.2 Indeterminate.....	218
2.1.3 Problems of Zero Markers.....	218
2.1.4 Problems of Parsed Corpora	219
2.1.5 Centered on Absence	224
2.2 Strategies of Previous Studies	224
2.2.1 Adjust Scope to the Searchable	224
2.2.2 Steer Away from the Corpus Approach	225
2.3 Data Limits	226
2.4 Technological Biases	228
2.5 Context of the <i>Verbless Sentence</i>	229
CHAPTER 3: CONTRASTIVE LINGUISTICS	231
3.1 Russian and English	232
3.1.1 Zero-Copula Constructions.....	232
3.1.2 More than Zero-Copula Constructions.....	235
3.1.3 Ellipsis and Morphological Case	238
3.2 Analysis of Translations	241
3.2.1 Source Language Interference	242

3.2.2	Traits of Translation Language	242
3.2.3	Translation Legitimacy	244
3.3	Multiple Directions.....	245
3.3.1	Directions of Translation	246
3.3.2	Directions of Correspondence.....	246
3.3.3	Third-Language Reciprocal Correspondences.....	247
3.3.4	Reference Features	248
3.3.5	Multiple Translations, Genre, Time, Number of Originals	248
3.4	Framing Translation Correspondences	252
3.5	Context of Verbless Sentence <i>Translations</i>	253
CHAPTER 4: AUTOMATIC PROCESSING		255
4.1	Pilot Studies and Accuracy Measures	256
4.2	Sentence Segmentation.....	264
4.3	Alignment.....	267
4.4	Structure & Trameur	270
4.4.1	Trameur.....	270
4.4.2	Base	272
4.5	Tagging.....	274
4.5.1	A Sensitive Issue for Verbless Sentences	274
4.5.2	Tagging of Russian and English.....	275
4.5.3	Post-Editing	275
4.6	Classification.....	276
4.6	Additional New Programs.....	277
4.7	Summary	278
4.7.1	Three Keys Issues of Verbless Sentence Retrieval	279
CHAPTER 5: QUANTITATIVE AND QUALITATIVE ANALYSIS.....		281
5.1	Monolingual Perspective	281
5.2	Parallel Perspective.....	285
5.3	Third-Language Perspective.....	288
CHAPTER 6: CONCLUSION: METHODOLOGY		289

PART THREE: CORPUS

CHAPTER 1: DESIGN AND SELECTION CRITERIA		291
1.1	The Target Population	293
1.2	The Kinds of Texts and Their Selection	294

CHAPTER 2: COMPOSITION	299
2.1 General Domain	299
2.2 Number of Texts	299
2.3 Number of Words	300
2.4 Description of Texts	301
CHAPTER 3: COMPARABILITY ANALYSIS	315
3.1 Russian and English in Number of Texts	315
3.2 Russian and English Source Type in Number of Words	317
3.3 Russian and English Genre in Number of Words	318
3.4 Russian and English Time Period in Number of Words	319
3.5 Parallel-Text Comparability Map	320

PART FOUR: RESULTS

CHAPTER 1: SYNTACTIC EXPLANATIONS	323
1.1 Frequency of Verbless Sentences	323
1.2 Ellipsis	328
1.2.1 Steps: Utterances and Annotation.....	329
1.2.2 Ellipsis overrepresented in <i>English</i>	330
1.2.3 Comparative Stripping, Sluicing, Gapping.....	333
1.2.4 Non-elliptical Dominate.....	335
1.2.5 Summary: The Limits of Syntax.....	336
CHAPTER 2: SEMANTIC FEATURES	338
2.1 Steps and Classification Schema	338
2.2 A Look Inside The Verbless Sentence	341
2.2.1 Agreement and Disagreement.....	342
2.2.2 Expressiveness.....	347
2.2.3 Quantification.....	350
2.2.4 Formality and Informality.....	355
2.2.5 Immediacy and Deixis.....	356
2.2.6 Questioning.....	358
2.2.7 Emphasis.....	362
2.2.8 Reference.....	367
2.3 Conventionalization and Repeated Segments	379
2.4 Length and Lexical Complexity	383
2.4.1 Quantifying 'Short'.....	383

2.4.2	Lexical Complexity	387
2.5	Statistical Portrait of the Verbless Sentence	389
CHAPTER 3: TRANSLATION LANGUAGE AND GENRE		390
3.1	Source-type	390
3.1.1	Monolingual Approach	390
3.1.2	Parallel Approach	393
3.2	Genre	395
3.2.1	Monolingual Approach	395
3.2.2	Parallel Approach	397
3.3	Direct Speech	398
CHAPTER 4: INFORMATION STRUCTURE		400
4.1	Information Structure Type	400
4.2	Topic	402
4.3	Focus	406
CHAPTER 5: TEMPORAL AND ASPECTUAL MEANING		412
5.1	Aspect	412
5.1.1	States and Events	418
5.1.2	Viewpoint	419
5.2	Tense	419
5.3	Verbal Lemmas	422
CHAPTER 6: INDIRECT SPEECH ACTS IN QUESTIONS		425
6.1	Sensitivity to the Verb	426
6.2	Speech Act of Verbless Questions	429
6.3	Speech Act of Verbal Correspondences	432

PART FIVE: CONCLUSION

VERBLESS SENTENCES	436
---------------------------------	------------

APPENDICES

APPENDIX 1: CORPUS LABELS	443
APPENDIX 2: TRAMEUR ILLUSTRATION	445
APPENDIX 3: MORPHOSYNTACTIC SPECIFICATIONS	454
APPENDIX 4: VERBAL TAGS	460

APPENDIX 5: CONTRASTIVE SEARCH & INTERFACE.....	461
APPENDIX 6: CHARACTERISTIC ELEMENTS	464
APPENDIX 7: CHARACTERISTIC N-GRAMS.....	473

RÉSUMÉ LONG EN FRANÇAIS

MOTIVATION DE L'ÉTUDE.....	482
FEUILLE DE ROUTE	488
Partie 1 : Le Cadre Théorique	488
Partie 2 : La Méthodologie.....	488
Partie 3 : Le Corpus.....	492
Partie 4 : Les Résultats.....	492
Chapitre 1 : Les Explications Syntaxiques	493
Chapitre 2 : Les Caractéristiques Sémantiques	494
Chapitre 3 : La Langue de Traduction et le Genre	494
Chapitre 4 : La Structure Informationnelle.....	495
Chapitre 5 : Le Sens Temporel et Aspectuel	496
Chapitre 6 : L'Acte de Langage	496
Partie 5 : Conclusion.....	497
LES IMPLICATIONS	500
1 Une Explication Phrastique.....	500
2 Une Approche Interdisciplinaire à la Description Sémantico-Pragmatique.....	501
3 La Recherche d'Absence	502
4 Un Nouveau Corpus Parallèle et Comparable	502
5 Dans quelle mesure l'ellipse syntaxique peut-elle expliquer le phénomène de la phrase averbale ?.....	503
6 Quelles caractéristiques sémantiques et pragmatiques distinguent les phrases averbales des phrases verbales dans les deux langues ?	504
7 Comment est-ce que la langue de la traduction et le genre influencent-ils le phénomène de la phrase averbale ?.....	505
8 Les phrases averbales omettent-elles des informations prévisibles et récupérables ?	505
9 Comment est-ce que les catégories sémantiques associées au verbe, y compris la signification temporelle et aspectuelle, se rapportent aux phrases averbales ?.....	506
10 Comment est-ce que les phrases averbales sont utilisées dans les actes de langage ?	507
11 Les Implications Plus Larges.....	507
LIST OF FIGURES.....	509
LIST OF TABLES	511
REFERENCES.....	512

ABSTRACTS

ENGLISH

ABSTRACT

Verbless and Zero Predicate Sentences: An English and Russian Contrastive Corpus Study

Key words:

contrastive linguistics, parallel and comparable corpora, NLP, semantics, pragmatics, syntax, translation, English, Russian, predication, verbless sentences, zero predicate, nonsententials, ellipsis

Focusing on structures in which the typical marker of sentential status – the verbal predicate – is absent, the present dissertation takes an interdisciplinary approach that combines contrastive linguistics, corpus methods and enunciative analysis in order to explore the semantic and pragmatic characteristics of verbless sentences in English and Russian.

Verbless sentences have principally received attention in the domain of syntax where the departure from canonical clause structure is typically analyzed as a nonsentential clause, an elliptical sentence, a fragment, a nominal utterance, but not a full sentence in its own right. Limited attention has been paid to the semantic and pragmatic domains, where the difficulty of

automatically processing absence in corpora has predominantly led the existing analyses to either concentrate on particular predefined types of verbless structures or rely on fragmented data.

This thesis aims to provide a semantico-pragmatic description of verbless sentences, of their characteristic features and of the possible restrictions on their use and meaning in two typologically profoundly different languages by studying them through the lens of contrastive corpus linguistics. A 1.4-million-word parallel, reciprocal and comparable corpus of 19th to 21st century literature is sentence-segmented, POS-tagged, and aligned with multiple translations at sentence-level. Retrieved automatically using a new method developed for accurate identification of the absence of the verb, verbless sentences are analyzed from monolingual, parallel-text and third-language perspectives. Annotation for syntactic, semantic and information structure features is carried out on sub-parts of the corpus. Quantitative and qualitative analysis is performed on verbless sentences in their original context. Contrastive analysis of systematically reoccurring translation patterns makes evident linguistic constraints that remain hidden from the perspective of a single language. Corpus analysis of the verbless sentences, their context and translations, is realized against a reference corpus. Enunciative analysis reveals the contextual features of the utterance situations in which the verbless sentences are found.

The results show that the statistically significant overrepresentation of verbless sentences in Russian as compared to English cannot be explained by appealing to typological differences in the productivity of syntactic verbal ellipsis. Contrary to expectations, antecedent-based ellipses are significantly overrepresented not in Russian, the language that stands out among the Indo-European language family for permitting the most liberal use of verbless sentences, but rather in English, the language that is typically associated with a dependence on the finite verb phrase. Furthermore, in both languages, the results reveal that verbless sentences are predominantly non-elliptical. Casting doubt on the extent to which syntactic analysis of ellipsis may explain the phenomenon of the verbless sentence, it is argued that verbless structures constitute a type of sentence in their own right and should not be treated as syntactic or semantic reductions of verbal sentences. Special attention in terms of semantic and pragmatic analysis is required in order to account for the verbless phenomenon in general and for the observed frequency differences cross-linguistically.

Among the results is a semantic classification of almost 20,000 verbless sentences in terms of the lexical constituents that statistically distinguish them from verbal sentences. The constituents that set verbless sentences semantically apart from verbal sentences in both

languages are statistically revealed to include indexical, negative quantification, expressive and informal lexical elements and segments; while indefinite reference and *wh*-words are key in English verbless sentences, elements marking deixis and emphasis of intensity are key for Russian. In spite of shorter length, verbless sentences are found to have greater lexical complexity compared to verbal sentences.

The verbless sentence is shown to be a linguistic phenomenon that is not significantly affected by translation, but rather by genre. Verbless sentences are found to be more affected by the difference between speech and narration than whether the source of the language is an original or a translation. A correlation with direct speech highlights the pragmatic requirements on the common ground between the interlocutors. Furthermore, the correlation suggests that the verbless sentence may be a linguistic feature of a functionally-based genre distinction between discourse-based fiction, scripted speech and narrative-based fiction.

Information structurally, the verbs that are potentially implicated by verbless sentences are found through contrastive analysis to be a part of focus. This finding challenges the extent to which the verbless sentence can be accounted for in terms of an omission of a predictable and reconstructable element. It provides further evidence that the constituents of a verbless sentence can be sufficient to express a complete thought and satisfy the requirements for constituting full instances of predication.

Furthermore, a difference in the instantiation of the informational topic between the two languages is observed to correlate with the instantiation of the verb. This finding suggests that some of the frequency differences between the two languages concerning verbless sentences may in part be due to typological differences in topic activation.

The results also explore the relation between verbless sentences and semantic properties that are typically associated with canonical verbal clauses: tense, situation and viewpoint aspect, and verbal lexical meaning. The present parallel data and contrastive methodology makes it possible to approach this question from a different perspective than the one that is typically taken. Instead of testing the semantic properties on non-verbal elements, in the present analysis it is verbal translations that are used to study the temporal, aspectual and lexical potential that is conversationally implicated in a verbless sentence. The results show that verbless sentences have the potential to implicate a wide range of temporal, aspectual and verbal meaning. Language specific correlations suggest typological differences. Defending a complete semantics for sentences without verbs, the potential existence of semantic properties typically associated with the verb are presently argued to be a matter of potential pragmatic implicature as opposed to semantic entailment.

Questions, as compared to other types of sentences, show particular sensitivity to the presence and absence of the verb in contrastive analysis. English is observed to use the absence of the verb as a grammatical marker of indirect speech acts in questioning. This finding, as well as the correlation between topic activation and the verb, contribute to explaining some of the frequency differences in the use of verbless sentences in the two languages.

The main contribution of this thesis is therefore a semantic and pragmatic corpus-based description of verbless sentences in English and Russian. The language-specific questions that it deals with include: What semantic characteristics distinguish verbless sentences from verbal sentences? What contextual factors influence their use? What happens with verbless sentences in translation from Russian to English and vice versa in terms of semantic content? How does textual genre and translation language influence the use of the verbless sentence? Are verbless sentences semantically, pragmatically or information structurally restricted compared to verbal sentences? How do semantic categories associated with the verb relate to verbless sentences? How are verbless sentences used in speech acts?

Wider implications pertain to the theoretical account of the verbless sentence. In addition to language-specific results, the present analysis contributes arguments toward the more general question of whether or not verbless sentences are reductions of verbal sentences. To what extent is their meaning a matter of assertion versus implicature? Where do verbless sentences stand in the debate between compositional versus conventional meaning? An account of the sentential status of verbless structures is proposed that does not rely on a hidden syntactic structure, but rather insists on the primacy of information structure features as necessary sentential criteria and treats the semantic content of a verbless sentence in terms of a linguistically-explicit assertion and contextually-revealed potential conversational implicature. A fundamental distinction between elliptical and non-elliptical verbless sentences is made without appealing to hidden elements.

Another wider implication concerns the automatic processing of absence. Verbless sentences correspond to an open class of grammatical structures centered on the absence of a formal marker. A series of specific problems associated with their automatic retrieval are identified and overcome with a reproducible semi-automated method. The results provide methodological suggestions and perspectives for further contrastive treatment of verbless sentences in corpora.

RUSSIAN

КРАТКОЕ СОДЕРЖАНИЕ

Безглагольные Предложения и Предложения с Нулевым Глагольным Сказуемым: Контрастивное Корпусное Исследование в Русском и Английском Языках

Ключевые слова:

контрастивная лингвистика, сопоставительная лингвистика, параллельный и сопоставимый корпус, обработка естественного языка, автоматическое извлечение, семантика, прагматика, синтаксис, перевод, английский язык, русский язык, предикация, безглагольные предложения, бессказуемые предложения, нулевое глагольное сказуемое, нулевой предикат, нулевая связка, неполное предложение, критерий неполноты, эллипсис

Диссертация посвящена структурам, в которых канонический критерий предложения – глагольное сказуемое отсутствует. Сочетая методы контрастивного,

корпусного и дискурсивного анализа, исследуются семантико-прагматические характеристики безглагольных предложений в английском и русском языках.

Предшествующие исследования в основном синтаксические и обычно рассматривают отклонение от канонической финитной структуры как ‘клауза несоответствующая предложению’, ‘неполное или эллиптическое предложение’, ‘фрагмент’ или ‘номинативное высказывание’. Ограниченное внимание уделяют семантико-прагматическим областям, где сложность автоматической обработки ‘отсутствия’ в корпусах вынуждает концентрироваться на заранее определенных типах или опираться на фрагментные данные.

В диссертации предлагается семантико-прагматическое описание безглагольных предложений и возможных ограничений на их употребление в двух глубоко различных языках через призму контрастивной корпусной лингвистики. Взаимно-параллельный сопоставимый корпус литературы 19 – 21 вв. объемом 1,4 млн. слов сегментирован по предложениям, снабжен морфосинтаксической-POS разметкой и полностью выравнен с многократными переводами на уровне предложений. Извлечённые автоматически с помощью нового метода, разработанного для точной идентификации отсутствия глагола, предложения анализированы с моноязычной и параллельной точек зрения, а также с позиции перевода с третьего языка. Синтаксические, семантические и информационно-структурные параметры аннотированны на субчастях корпуса. Количественный и качественный виды анализа реализованы в контексте. Контрастивный анализ переводов выявил лингвистические ограничения, скрытые в моноязычном подходе. Корпусный анализ безглагольных предложений, их переводов и контекста проводился в сопоставлении с референтным корпусом. Дискурсивный анализ позволяет изучить особенности контекста ситуации.

Результаты показывают, что безглагольность предложения не объясняется типологической продуктивностью синтаксического эллипсиса. Антецедентные эллипсы преобладают не в русском языке, являющимся индоевропейским лидером по безглагольным конструкциям, а в английском, который известен своей глагольной зависимостью. Кроме того, выявлено, что в обоих языках безглагольные предложения преимущественно неэллиптические. Эти результаты свидетельствуют в пользу семантико-прагматического объяснения. Безглагольная структура рассматривается как предложение без синтаксической или семантической редукции.

Семантическая классификация 20,000 безглагольных предложений по лексическим компонентам, статистически отличающим их от глагольных предложений,

показывает, что неопределенная референция и *wh*-слова являются ключевыми в английском языке, а дейксис и эмфаза интенсивности – в русском.

Установлено, что на безглагольные предложения существенно влияет не столько язык перевода, сколько жанр. Корреляция с прямой речью подчеркивает требование общего знания.

Контрастивный анализ указывает на принадлежность глаголов, потенциально подразумеваемых в безглагольных предложениях, к информационному фокусу. Такой результат ставит под сомнение возможность объяснения безглагольного предложения с точки зрения отсутствия предсказуемого реконструируемого элемента. Это свидетельствует о том, что составляющие безглагольного предложения могут быть достаточными для выражения законченной мысли и удовлетворять требованиям, предъявляемым к полной предикации.

Безглагольные предложения способны имплицировать широкий спектр временных и аспектуальных значений; потенциальные семантические свойства глагола рассматриваются как возможность прагматической импликации, а не семантического следования.

Вопросительные предложения демонстрируют особую чувствительность к глаголам. В английском языке наблюдается использование глагольного отсутствия в качестве грамматического маркера косвенных речевых актов в вопросах. Данный результат, а также корреляция между топиком и глаголом, способствуют частичному объяснению частотного различия между языками.

Более широкие последствия данного исследования относятся к теории предложения и обоснования безглагольных структур в качестве предложений, а также к автоматической обработке отсутствия глаголов. Безглагольное предложение анализируется как сочетание утверждения и потенциальной коммуникативной импликатуры. Различие между эллиптическим и неэллиптическим типом безглагольных предложений предлагается без обращения к скрытой структуре. Проблемы автоматического поиска конкретизированы и преодолены с помощью воспроизводимого полуавтоматизированного метода для дальнейших контрастивных исследований безглагольных предложений в корпусах.

FRENCH

RÉSUMÉ

Les Phrases Averbales et Les Phrases à Prédicat Zéro : Étude Contrastive Anglais-Russe Basée sur Corpus

Mots-clés :

linguistique contrastive, corpus parallèle et comparable, TAL, sémantique, pragmatique, syntaxe, traduction, anglais, russe, prédication, phrases averbales, prédicat zéro, nonsententials, énoncé non-phrastique, ellipse

Cette thèse porte sur les structures dans lesquelles le marqueur typique de la phrase – le prédicat verbal – est absent, et convoque des méthodes contrastives, énonciatives ainsi que la linguistique de corpus pour explorer les traits sémantico-pragmatiques des phrases averbales en anglais et en russe.

Dans les études antérieures, principalement syntaxiques, l'écart avec la structure canonique est typiquement analysé en tant que proposition non-phrastique, phrase elliptique, fragment ou énoncé nominal. Peu d'attention a été accordée aux domaines sémantico-

pragmatiques où la difficulté du traitement automatique de l'absence dans un corpus a conduit les analyses à se concentrer sur des types prédéfinis ou à utiliser des données fragmentées.

La thèse vise une description sémantico-pragmatique des phrases averbales et leurs éventuelles restrictions d'utilisation dans deux langues profondément différentes à travers le prisme de la linguistique contrastive de corpus. Un corpus parallèle réciproque et comparable de 1,4 million de mots de la littérature des 19 – 21^e siècles a été segmenté en phrases, étiqueté POS et aligné à la phrase avec plusieurs traductions. Récupérées automatiquement avec une nouvelle méthode développée pour l'identification fiable de l'absence verbale, les phrases averbales sont analysées dans une perspective monolingue, parallèle et du point de vue de la troisième langue. L'annotation syntaxique, sémantique et de structure informationnelle est effectuée sur les sous-parties. L'analyse quantitative et qualitative est réalisée en contexte. L'analyse contrastive des traductions récurrentes révèle des contraintes linguistiques sous-jacentes à une seule langue. L'analyse du corpus des phrases averbales, de leurs traductions et de leurs contextes est effectuée par rapport à un corpus de référence. L'analyse énonciative révèle les traits situationnels conditionnant l'énoncé.

Les résultats montrent que la phrase averbale ne s'explique pas par la productivité syntaxique et qu'elle est statistiquement nonelliptique. L'ellipse en antécédent est surreprésentée non pas en russe, langue où les structures averbales sont pourtant fortement présentes, mais en anglais, connu pour sa dépendance du groupe verbal. Ce résultat milite pour une explication sémantico-pragmatique. Les structures averbales sont considérées comme des phrases sans réduction syntaxique ou sémantique.

La classification sémantique de 20,000 phrases averbales en fonction des composants lexicaux qui les distinguent statistiquement des phrases verbales montre que la référence indéfinie et les mots-wh sont clés en anglais, tandis que la deixis et l'accentuation de l'intensité le sont en russe.

La phrase averbale n'est pas affectée significativement par la traduction, mais plutôt par le genre. La corrélation avec le discours direct met en évidence les exigences du savoir partagé.

L'analyse contrastive révèle que les verbes potentiellement impliqués pragmatiquement dans les phrases averbales font partie du focus. Cette découverte remet en question la façon dont la phrase averbale peut être expliquée en termes d'omission d'un élément reconstituable prévisible. Elle démontre que les constituants d'une phrase averbale peuvent être suffisants pour exprimer une pensée entière et satisfaire aux exigences pour constituer des instances complètes de prédication.

La phrase averbale peut impliquer un éventail de significations temporo-aspectuelles; les propriétés sémantiques verbales potentielles sont proposées comme relevant de l'implicature pragmatique plutôt que de l'encodage sémantique.

Les questions montrent une sensibilité particulière au verbe; l'anglais utilise l'absence de verbe comme un marqueur grammatical de l'acte de langage indirect dans les questions. Ce résultat, ainsi qu'une corrélation observée entre le topic informationnel et le verbe, contribuent à expliquer en partie la variation entre les langues.

Des implications plus larges portent sur la justification phrastique des structures averbales, ainsi que sur le traitement automatique de l'absence. La phrase averbale est justifiée comme une alliance d'assertion et d'implicature conversationnelle potentielle. Une distinction non/elliptique est maintenue sans faire appel à une structure cachée. Les problèmes de récupération automatique sont précisés et surmontés par une méthode semi-automatique reproductible pour des études contrastives futures des phrases averbales dans les corpus.

Un résumé long en français se trouve à la fin de la thèse (p. 481–508).

PRELIMINARIES

TRANSLITERATION

Cyrillic	Latin	Phonetic			
а	<i>a</i>	[a]	ъ	"	
б	<i>b</i>	[b]	ы	<i>y</i>	[ɨ]
в	<i>v</i>	[v]	ь	'	[j]
г	<i>g</i>	[g]	э	<i>è</i>	[ɛ]
д	<i>d</i>	[d]	ю	<i>ju</i>	[ɯ]
е	<i>e</i>	[e]	я	<i>ja</i>	[æ]
ё	<i>ë</i>	[ø]			
ж	<i>ž</i>	[ʒ]			
з	<i>z</i>	[z]			
и	<i>i</i>	[i]			
й	<i>j</i>	[j]			
к	<i>k</i>	[k]			
л	<i>l</i>	[l]			
м	<i>m</i>	[m]			
н	<i>n</i>	[n]			
о	<i>o</i>	[o]			
п	<i>p</i>	[p]			
р	<i>r</i>	[r]			
с	<i>s</i>	[s]			
т	<i>t</i>	[t]			
у	<i>u</i>	[u]			
ф	<i>f</i>	[f]			
х	<i>x</i>	[x]			
ц	<i>c</i>	[ts]			
ч	<i>č</i>	[tʃ]			
ш	<i>š</i>	[ʃ]			
щ	<i>šč</i>	[ʃtʃ]			

Romanization of the Cyrillic alphabet into Latin script corresponds to the ‘Linguistic’ system that serves as an adapted phonetic alphabet (e.g., Timberlake, 2004: 24–25).

GLOSSING ABBREVIATIONS

Abbreviation	Interpretation
M / F / N	masculine / feminine / neuter
1 / 2 / 3	first person / second person / third person
SG / PL	singular / plural
NOM	nominative case
GEN	genitive case
DAT	dative case
ACC	accusative case
INS	instrumental case
PRE	prepositional case
PS / PR / FT	past / present / future
PFV / IPFV	perfective / imperfective
GRND / IMV / INF / PTCP	gerund / imperative / infinitive / participle
ADJ / ADJS	adjective / short adjective
ADV	adverb
ANIM / INANIM	animate / inanimate
COMP	comparative
CONJ	conjunction
DEM	demonstrative
DET	determiner
INTJ	interjection
NEG	negation
NN	noun
NUM	numeral
PART / P	particle
PREP	preposition
PRO	pronoun
REF	reflexive
SUPR	superlative
V	verb
[e]	ellipsis

Glossing Notes

The annotation of examples for grammatical properties and meanings of the individual words corresponds to the Leipzig glossing conventions (Comrie et al., 2015), slightly modified for relevant categories. The above provides a full list of the glossing abbreviations used.

- The glosses indicate only the labels necessary for the analysis; whenever possible a literal translation of the word replaces the category label.
- A period is used to separate the category labels; however, gender, person and number are not separated by a period (e.g., M1SG indicates masculine first person singular).
- An underscore is used when a single word gloss is lacking.
- For in-text examples, the transliteration is followed by a literal gloss which includes any necessary category labels, e.g., ‘русский’ (*russkij*; lit. ‘*russian.ADJ*’).

EXAMPLE GUIDELINES

Examples are annotated according to the following guidelines:

— *Context:*

The context is provided in several types of brackets. Square braces indicate the [linguistic context]. Curly braces indicate a summary of the {situational context}. Round braces belong to the (original text).

— *Line 1 – Transliteration:*

Romanization of the Cyrillic alphabet into Latin script corresponds to the ‘Linguistic’ system. The transliteration correspondences are provided in the *Preliminaries Transliteration* section.

— *Line 2 – Gloss:*

Morpho-syntactic glosses correspond to the Leipzig Glossing Abbreviations described in the *Preliminaries Glossing Abbreviations* section.

— *Line 3 – Literal Translation:*

Line three provides a possible word-by-word literal translation.

— *Line 4 – Sentential Translation:*

Line four, where necessary, gives a sentential translation that comes from either (a) the corpus and is labelled according to the guidelines in *Appendix 1*, or (b) the author and not labelled. Such labelling of the source of examples holds for originals and translations.

PART ZERO:

INTRODUCTION

MOTIVATION AND ROAD MAP

The present dissertation focuses on structures in which the typical marker of sentential status – the verbal predicate – is absent. The semantic and pragmatic features of verbless sentences are explored in English and Russian using an interdisciplinary approach that combines contrastive linguistics, corpus methods and enunciative analysis.

At what point do words become sentences? The question of what precisely constitutes a sentence in natural language is central to any linguistic theory. A sentence requires something more than a random succession of words. This was first noted by Plato in 360 B.C. in the *Sophist*. The statement that “*the verbs are mingled with the nouns; then words fit, and their first combination is a sentence*” (Plato, 360 B.C./1921: *Sophist* 262b–d) is usually identified as the catalyst for the discussion of the defining activity that takes place in a sentence. Since Plato, this central activity, referred to as predication, has been associated with the presence of the verb. The traditional verbal requirement renders contradictory the notion of a verbless sentence and consequently often pushes verbless structures to the sidelines of linguistic discussion as incomplete, atypical and irrelevant for linguistic analysis.

Yet, the phenomenon of sentences without a verb has been noted to exist in many natural languages, most prominently by Emile Benveniste in *Problems in General Linguistics* (1966/1971). Among the Indo-European language family, the Russian language is generally known for permitting the most liberal use of verbless sentences (McShane, 2000; Kopotev, 2007b). It readily allows sentences such as (i) below, literally ‘*I Tonya*’, where the copula verb ‘*be*’ would typically be expected in English, as well as sentences without full lexical verbs, such as (ii), literally ‘*I to store*’. The typological characteristics of Russian include a morphological case system that is highly developed and contributes to the language’s extraordinary capacity for subject and verbal ellipsis; flexible word order and intonation; the nonexistence of articles; it typically suppresses the copula verb ‘*быть*’ (*byt*’; lit. ‘*be*’) in the

present tense, as illustrated in the “zero copula” construction (Stassen, 2013) in (i), and allows verbless structures such as (ii) without requiring an antecedent verb in the context.

- (i) Я Тоня.
ja tonja
I Tonya.NOM
‘I am Tonya.’
- (ii) Я в магазин.
ja v magazin
I to.PREP store.ACC
‘I am going to the store.’

Although the English language is known for a dependence on the finite verb phrase, verbless sentences are also found across various speech acts, as illustrated in the declaration, directive, exclamation and question in (iii).

- (iii) a. No sign of Dmitry Fyodorovich yet.
b. Just a moment.
c. How terrible!
d. What about my parental blessing?

The characteristics of English provide a sharp contrast to Russian and include a limited case system that is aided by the verb to reduce ambiguity; register restrictions on syntactic verbal ellipsis, particularly gapping (McShane, 2000); the typical necessity of an overt copula, articles, and strict word order. These profound typological differences between the two languages make their contrast particularly relevant for the study of verbless sentences.

Although verbless structures have fascinated many linguists, the difficulty of retrieving them automatically has meant that most analyses have relied on fragmented data and invented examples. The challenge for any automated search of the phenomenon is that the defining element of the query is a grammatical structure that is indeterminate and centered on absence. For natural language processing, verbless structures constitute open strings consisting of grammatical elements that cannot be specified a priori. In other words, they represent a quasi-absurd search for something indeterminate that contains something that doesn’t exist. The presence of a verb via a null element, and the eventual position of such an element in a linguistic structure, is a matter of theoretical dispute. Corpus annotation does not include such zero markers (Loock, 2016: 33). Parsed corpora, which is typically recommended for the treatment of grammatical structures, are mostly built around syntactic models that are verb-centric and

do not allow for the search of verbless clauses (Landolfi et al., 2010). Furthermore, little attention has been paid to the accuracy of automatic retrieval of the structures in existing corpora. As a result, most of the corpus work on the phenomenon has focused on particular verbless structures that are predetermined and include some searchable formal marker.

The verbless sentence has principally received attention in the domain of syntax. Its departure from canonical clause structure (i.e. a two-part structure consisting of a subject and a verbal predicate that is linked to the subject by means of inflection for agreement and tense) has earned it various names including ‘fragment’, ‘elliptical sentence’, ‘nonsentential clause’, ‘nominal utterance’, amongst others, but not a full-fledged sentence in its own right. Most of the syntactic discussion of verbless structures centers on recovering a verb. In order to give verbless structures sentential status, albeit elliptical, models are proposed of an underlying hidden structure that is verbal and either (a) transformed into a verbless one by means of surface level deletion or (b) includes special null elements that stand in for the verb (Elugardo & Stainton, 2005). Alternatively, verbless structures are proposed to be generated without any hidden elements in their structure but their sentential status is withdrawn due to lack of inflection for agreement and tense (Barton & Progovac, 2005). The latter analysis seems preferable since it does not assign any extra structure to the verbless construction beyond the words that are there. However, in both cases, the words of a verbless structure are treated as themselves not enough to be a full sentence: they are either incomplete left-overs from a full sentence and receive ‘elliptical sentence’ status, or they are self-sufficient ‘nonsententials’.

From a semantic perspective, the verbless sentence poses further riddles. At first glance, it is reasonable to assume that the elements that may be omitted from a sentence are those elements that are predictable and therefore their meaning may be easily recovered in order to complete the meaning of the sentence. After all, the meaning of a verbless sentence is typically clear for interlocutors. However, recovering the meaning of the verb turns out to be no less problematic than finding a place for it in the syntactic structure.

First, it is far from clear which predicate should be re-established; the same verbless sentence can be felicitous with a range of predicates that may be narrowed down but not fully resolved by the context. For instance, the sentence in (iv), literally ‘*He to hospital*’, may be used in the context of (a) motion of the agent toward a destination (voluntary, e.g. *walk, drive*, and involuntary, e.g. *arrive*), (b) violence at an illegal military target (actual, e.g. *shoot, bomb*, and prospective, e.g. *aim*), (c) flowers for the patients inside (acquisition, e.g. *buy, order*, and distribution, e.g. *give, deliver*), as well as in other contexts that narrow down a range of possibilities for the lexical meaning of a predicate.

- (iv) Он в больницу.
 on v bol'nicu
 he to.PREP hospital.ACC

Similarly, the context may restrict the range of possible temporal and aspectual meanings that a sentence can take, e.g. observing an agent who (a) has just departed, versus, (b) is getting ready for departure, would change the range of temporal and aspectual possibilities for the predicate. Nevertheless, the sentence structure alone is compatible with multiple possibilities, and the context, beyond delimiting a range, does not specify which particular predicate within the range is to be adopted for re-introduction.

Furthermore, even if a particular predicate could be recovered, re-establishing it appears to change the meaning of the sentence. Existing semantic analyses highlight the changes associated with making explicit a salient predicate for certain Russian structures. Differences concerning referential value and situational links have been noted by Ol'ga Selivërstova (1973) and Denis Paillard (1984). For instance, a famous example from Selivërstova (1973) illustrates different readings for the verbless sentence in (v) 'У него седые волосы' (*u nego sedye volosy*; lit. 'at him gray hair') and its verbal counterpart with the present tense 'быть' (*byt'*; lit. 'be').

- (v) a. У него седые волосы.
 u nego sedye volosy
 at him gray hair
 'He has gray hair.'
- b. У него есть седые волосы.
 u nego est' sedye volosy
 at him is gray hair
 'He has some gray hair.'

While the verbless sentence in (a) is typically used to attribute the quality 'gray' to the entire head of hair, the verbal alternative in (b) is a marked structure that blocks such a reading. The latter can only be used to assert the existence of a limited quantity of gray hair. Recovering a verbal predicate for a verbless structure, even in situations where the verb seems fully predictable, is not straightforward syntactically and even less so semantically.

The present dissertation aims to break new ground in the semantico-pragmatic description of verbless sentences by taking a contrastive corpus approach to the phenomenon in Russian and English. The characteristic features of verbless sentences and possible restrictions on their use and meaning are analyzed through the lens of a 1.4-million-word

parallel comparable corpus that was specially created to permit the development and utilisation of a new method of accurate verbless sentence retrieval, quantitative analysis of verbless sentences against a reference corpus, and context-dependent semantic and pragmatic analysis of the structures.

The **first part** outlines the theoretical background of the study. It begins by defining the target phenomenon from a perspective that makes minimal assumptions. Verbless sentences are formalized as strings of text that (a) are found in context, (b) do not include a finite verb, nor any other verb form, and (c) are delimited by initial and final punctuation marking and turn change. The advantages and limits of the current definition are explored. A fundamental binary distinction of verbless sentences in terms of elliptical and non-elliptical is established. The section discusses the essential terminology with regard to verbless structures, including various uses of the term ‘ellipsis’ and the notion of a zero predicate.

Part two describes the methodological framework. Three separate methods to the study of language – contrastive, corpus and enunciative approaches – are presented in terms of their individual strengths and vulnerabilities, together with the potential from their unison for the semantic study of verbless phenomena.

Guiding this study is the contrastive analysis method developed by Jacqueline Guillemin-Flescher (2003) and the principle that analysis of systematically re-occurring translation patterns can make evident linguistic constraints that remain hidden from the perspective of a single language. In order to mitigate the influence of source language interference and disentangle linguistic patterns from potential translation universals (Nádvořníková, 2017; Zanettin, 2013; Olohan, 2002; Baker, 1993), measures are taken to include multiple, reciprocal and third-language translations in the corpus composition. The corpus is fully aligned with translations at the level of the sentence which makes it possible to systematically retrieve not only verbless sentences, but also their translation correspondences across multiple texts.

The under-researched semantic and pragmatic aspects of verbless sentences are presently targeted through analysis of the contextual factors of the enunciative situation in which the verbless sentences are found. Existing studies in these domains, due to technological difficulties, have mostly steered away from corpora. As a result, they are vulnerable to the critiques that face data that has been eclectically selected from various sources by the researcher to illustrate a preconceived model (Kohnen, 2015; McEnery et al., 2006; McEnery & Wilson, 2001; Garside et al., 1997). However, in order to confront the gap in semantic and pragmatic analysis from a corpus perspective, it is necessary to make the context central at all steps of the corpus design and data retrieval. This adds another complication to the retrieval of

verbless sentences in corpora. Maintaining the vital link to context in order to permit semantic analysis is currently still not an easy feat since it pertains not to the typical concordance lines centered on a word or an expression, but to verbless sentences and simultaneously to their translation correspondences. The challenge is presently tackled through a corpus design and retrieval method that resolves all three – verbless sentences, translation correspondences, and context.

The present new method of automatic retrieval was developed with a particular concern for accuracy in an attempt to minimize some of the biases from the use of corpus tools (Anthony, 2012). The method aims to be replicable on any raw text and, at its essence, involves classifying appropriately delimited and part-of-speech-tagged sentence units while maintaining their place in the overall structure. First, the texts are segmented into sentences using a script that unites a custom series of algorithms that target graphical markers in order to distinguish direct speech, clitics and other issues that pose problems for recall and precision of verbless sentence retrieval. The texts are then manually aligned across multiple translations at the sentence level and united into a parallel-text structure. All of the words are then automatically sequenced to form a single data-thread, morpho-syntactically tagged, the sentence-segments are classified into verbal and non-verbal, and the alignment is visualized and explored using the multi-level text processing and statistical analysis software package Trameur (Fleury, 2019a; Fleury & Zimina, 2014). The latter processes custom segmented data, permits automatic correction of a large portion of tagging errors, permits the visualization of verbless sentences aligned with multiple translations in their original context, as well as statistical analysis against a reference corpus of verbal sentences.

Following extraction, verbless sentences and their translation correspondences in sub-parts of the corpus are further segmented manually into smaller utterance-units and manually analyzed for syntactic, semantic and pragmatic features. Amongst the manually annotated categories are antecedent-based verbal ellipsis, subject and predicate, verbal translation correspondence and its tense and aspect, discourse type, information structure, direct and indirect speech act, and other categories, the definitions of which are presented in this chapter. The chapter also defines the statistical analysis that was carried out on the verbless sentences and their context from a monolingual perspective, including characteristic elements and n-grams, and also provides a guide to the semantic classification for the statistically identified results.

The corpus is examined from three perspectives: monolingual, parallel and third-language translation. First, from a monolingual perspective, it is not only the original texts, but

also the translations, that are treated as genuine samples of language production in their own right. Concerning the use of translations in monolingual analysis, the present study sides with Zanettin (2014), Olohan (2002), Baker (1996), Biber (1993), amongst others, in arguing that translations constitute an important part of the production and reception of a language, and as a result must be included in any corpus design that aims to be representative of a language, whether in terms of linguistic variety (Biber, 1993) or proportional usage (Leech, 2007). Thus, Russian originals are compared not only with English originals, but also with Russian translations; and the same for English. The four language types – Russian originals, English originals, Russian translations, English translations – are compared in terms of normalized frequencies, semantic classification of the characteristic elements and n-gram results, and the manually annotated categories.

Secondly, combining the contrastive analysis principles of Guillemin-Flescher (2003) with criteria for reliable parallel corpora (Nádvorníková, 2017; Stolz, 2007a; McEnery & Xiao, 2008; Malmkjaer, 1998), the corpus is studied from a parallel perspective. The present analysis looks for correspondence patterns that reoccur across multiple translations of the same original, across several works by different authors, across different genres, and in reciprocal translation directions. In addition to the two typical translation directions, (a) from Russian originals to English translations, and (b) from English originals to Russian translations, the commitment to treating translations as a type of language in its own right and the nature of the phenomenon leads the present study to also include the analysis of correspondence patterns (c) from English translations to Russian originals and (d) from Russian translations to English originals. For the present parallel analysis, particular attention is given to verbal correspondences of non-antecedent based verbless sentences and their correlation with the manually annotated categories.

Finally, the corpus is examined from a third-language translation perspective. In order to identify patterns that indicate language specific constraints, it is necessary to disentangle them from patterns that result from the influence of the original language on the translation (for instance, a literal translation of an English verbal sentence may result in a verbal sentence in Russian, even though in the context of translation from another language, or in a non-translational context, a verbless sentence would typically be used), as well as those that may be due to the very act of translation (for instance, the tendency to simplify in the passage from one language to another and other potential universal features of translation). In an attempt to control for the interference of the original language and potential translation universals, a sub-corpus consisting of Russian and English translations from French is added. This sub-corpus

undergoes the same monolingual and parallel analyses. The third language perspective controls for Russian and English interference on one another and provides another type of translated data to compare with originals.

Part three presents the 1.4-million-word, fully sentence aligned and morpho-syntactically tagged parallel and comparable corpus that was built for the present study. Its general domain is realist fiction from the late 19th through early 21st century that has received high exposure and has been translated into the target languages. The total of 32 texts consists of 13 different works and includes 12 originals, 16 translations and 4 third-language translations, that are equally divided between Russian and English. Three function-based genres are represented, including discourse-based fiction, which represents the main targeted genre, as well as scripted-speech and narrative-based fiction, the latter two representing control groups in order to disentangle patterns that are due to genre interference. The section describes the composition of the corpus, outlines the selection criteria, and evaluates the texts against the criteria. The comparability of the corpus is analyzed and evaluated as loosely comparable, according to the sampling frame, including size, time period, genre, language (i.e., Russian and English) and language type (i.e., original, translation, third-language translation). The corpus aims to be representative of a wide variety and high frequency of verbless sentences. Representativeness in terms of proportion of speakers or receivers of the targeted languages (e.g. proportion of English speakers represented by the composition of the present corpus) is traded in favour of linguistic variety of the targeted phenomenon.

Part four combines the description of the results and the quantitative and qualitative analysis in several chapters that each target a different question.

The **first chapter** of this section tackles syntactic explanations of the typological differences between the languages. It analyzes results concerning the frequency of verbless sentences in the two languages, as well as the fundamental elliptical and non-elliptical verbless sentence distinction. Based on the present data, this chapter first reveals the frequency with which verbless sentences are used in the two languages and establishes that the expected overrepresentation of verbless sentences in Russian compared to English is statistically significant. It then shows that the more frequent use of verbless sentences in Russian as compared to English cannot be explained by appealing to typological differences in the productivity of syntactic verbal ellipsis. Contrary to expectations, antecedent-based ellipses are significantly overrepresented not in Russian, the language that stands out among the Indo-European language family for permitting the most liberal use of verbless sentences, but rather in English, the language that is typically associated with a dependence on the finite verb phrase.

Furthermore, the results reveal that verbless utterances are predominantly non-elliptical in both languages. These results cast doubt on the extent to which syntactic ellipsis may explain the phenomenon of the verbless sentence. In order to account for the verbless phenomenon in general (i.e., the high proportion of the non-elliptical type) and for the observed cross-linguistic frequency differences, special attention is required in terms of semantic and pragmatic analysis. This attention is given in the results of the following chapters which focus on determining semantic and pragmatic meaning and possible restrictions on verbless sentences.

The first chapter also lays the foundation for the argument that runs throughout the thesis which is that verbless structures constitute a type of sentence in their own right and should not be treated as syntactic or semantic reductions of verbal sentences. This chapter provides empirical evidence showing that the phenomenon of the verbless sentences is not explained by a null element, nor the deletion of a predicate, that exists in the verbless sentence and is reconstructable from the linguistic context. It is proposed that the zero predicate may help to explain the elliptical type of sentence, i.e. the type that does potentially involve deletion of an underlying predicate. However, based on the present results most verbless sentences appear to be independent, at least syntactically, from an existing predicate.

Chapter two provides empirical evidence in support of the argument that they are also independent sentences semantically. In spite of shorter length, verbless sentences are found to have greater lexical complexity compared to the average sentence in English. Such a result is not expected for mere reductions and fragments of verbal sentences.

This chapter also analyzes the distinguishing semantic features of verbless sentences. It presents the results of a semantic classification of almost 20,000 verbless sentences in terms of the lexical constituents that statistically distinguish them from verbal sentences. The constituents that set verbless sentences apart from verbal sentences in both languages are statistically revealed to include indexical, negative quantification, expressive and informal lexical elements and segments. While indefinite reference and *wh*-words are key in English verbless sentences, elements marking deixis and emphasis of intensity are key for Russian.

Chapter three explores the question of translation language and textual genre. What influence does translated language have on verbless sentences? How does textual genre affect verbless sentence use? The verbless sentence is shown to be a linguistic phenomenon that is language-specific and not significantly affected by translation from both a monolingual and a parallel perspective. Verbless sentences are found to be more affected by the difference between speech and narration than whether the source of the language is an original or a translation.

These results are supportive for contrastive analysis of the phenomenon. A linguistic analysis based on translations typically faces the need to defend against the argument that translations represent a type of language in which some particular phenomenon may disappear, drowned under potential universal features of translated language. The present results suggest that this argument can safely be rejected for the study of verbless sentences. From a contrastive perspective, verbless sentences behave similarly in originals and translations in terms of verbal translation correspondences; from a monolingual perspective, only a slight difference is observed in the frequency of verbless sentences.

This chapter also analyzes the correlation of verbless sentences with direct speech. This correlation is exposed through (a) manual annotation of verbless sentences for direct speech versus narration, (b) statistical analysis of the constituents of verbless sentences that reveal the saturation of elements that require contextual resolution (e.g. indexicals) and interaction (e.g. interjections, question words), and (c) variation in the frequency of verbless sentences across the examined genres. The correlation highlights the pragmatic requirements on the common ground between the interlocutors. Furthermore, it is proposed that the verbless sentence may be a linguistic feature helping to establish a functionally-based genre distinction between discourse-based fiction, scripted speech and narrative-based fiction.

Chapter four analyzes the information structure (Lambrecht, 1994) of the verbless sentences from a monolingual and a contrastive perspective. Considered monolingually, verbless sentences are found across all information structure types, including topic-comment, identificational andthetic sentences.

The results in this chapter also show that the verbs that are implicated by verbless sentences are found through contrastive analysis to be a part of focus. This finding challenges the extent to which the verbless sentence can be accounted for in terms of the omission of a predictable and reconstructable element. It provides further evidence to support the argument that the constituents of a verbless sentence can be sufficient to express a complete thought and satisfy the requirements for constituting full instances of predication.

The chapter also reveals a difference in the instantiation of the informational topic in the verbless sentences of the two languages. The instantiation of the topic is observed to correlate with the instantiation of the verb. This finding suggests that some of the frequency differences between the two languages concerning verbless sentences may in part be due to typological differences in topic activation. The chapter thus contributes a pragmatic explanation that may account for some of the variation in the verbless sentences between the languages.

Chapter five targets the question of potential semantic, temporal and aspectual restrictions on verbless sentences. What is temporal and aspectual meaning without their typical marker – the verb ‘constellation’ (Smith, 1997: 10)? This chapter explores the relation between verbless sentences and the semantic properties that are typically associated with canonical verbal clauses – tense, situation and viewpoint aspect, as well as their relation with verbal lexical meaning. The present parallel data and contrastive methodology makes it possible to approach this question from a different perspective than the one that is typically taken in existing analyses of verbless sentences. Instead of testing the semantic properties on non-verbal elements, in the present analysis verbal translations are used to study the temporal, aspectual and lexical potential that is conversationally implicated in a verbless sentence.

The results show that verbless sentences have the potential to implicate a wide range of temporal, aspectual and verbal lexical meaning. Language specific correlations are used to reveal typological differences. Defending a complete semantics for sentences without verbs, the semantic properties typically associated with the verb are presently argued to be a matter of pragmatic implicature as opposed to semantic entailment.

Chapter six reveals another pragmatic explanation for the non-elliptical verbless sentence variation between the languages. The chapter presents evidence that the absence of the verb may be used as a grammatical marker in English to signal that a question is an indirect speech act.

It starts by showing that the analysis of translation correspondences identifies questions, as compared to other types of sentences, as being particularly sensitive to the presence and absence of the verb: regardless of language type (source, translation or third-language translation), verbs vary more frequently in correspondences of verbless questions than in all of the other sentence types combined.

Secondly, from a monolingual perspective, in both languages, verbless questions are found to correlate with indirect speech acts. In other words, the absence of the verb in questions tends to be associated with situations in which the speaker is not requesting information from the addressee (e.g. rhetorical questions, surprise questions) in both languages.

However, a contrastive look exposes typological differences in the use of the verb in questions. English verbal correspondences of Russian verbless questions correlate with direct speech acts across all language types. In other words, English is observed to use the absence of the verb as a grammatical marker of indirect speech acts in questioning. The finding suggests that in English the verb allows to distinguish between questions as direct versus indirect speech acts, whereas in Russian questions the direct-indirect distinction is not related to the verb.

Rather than syntax, these results suggest that contributing to the explanation of verbless sentence frequency differences between the languages is a typological difference in the pragmatic use of the verb in questions.

The **conclude**, the main contributions of the present dissertation are:

- a sentential account that justifies verbless sentences as constituting a legitimate linguistic phenomenon;
- the development of semantic and pragmatic analysis of verbless sentences in Russian and English through a contrastive corpus approach;
- a semi-automatic method of retrieval of verbless sentences for semantic analysis;
- a new parallel and comparable corpus of thirteen works of 19th – 21th century Russian and English realist fiction.

Several questions that pertain to typological differences between Russian and English are explored, including empirical evidence concerning:

- How successful are syntactic differences in the productivity of ellipsis at explaining the variation between the languages?
- What semantic and pragmatic features distinguish verbless sentences from verbal sentences in the languages? What contextual factors influence their use?
- How does textual genre and translation language affect verbless sentences?
- What happens with verbless sentences in translation from Russian to English and vice versa? How do semantic categories associated with the verb relate to verbless sentences? Are verbless sentences semantically, pragmatically or information structurally restricted?
- How are verbless sentences used in speech acts?

The wider implications of the dissertation pertain to the theoretical account of the verbless sentence. In addition to language specific results, the analysis contributes arguments toward more general questions, including:

- Are verbless sentences reductions of verbal sentences?
- To what extent is their meaning a matter of assertion versus implicature?
- Where do verbless sentences stand in the debate between compositional versus conventional meaning?

An account of the sentential status of verbless structures is proposed that does not rely on a hidden syntactic structure, but rather insists on the primacy of information structural focus as the necessary sentential criteria. It is proposed that the semantic content of a verbless sentence

should be treated in terms of linguistically explicit asserted elements and contextually-revealed conversation implicatures.

A final contribution of the present thesis concerns corpus treatment of the absence of the verb. Development of the necessary corpus design is typically a multi-step process, as stressed by Biber (1993), amongst others. The present corpus design and method of retrieval are achieved through several smaller pilot studies. Attention is drawn to the specific challenges facing the automatic retrieval of the open class of grammatical structures centered on the absence of the verb. Methodological suggestions for further treatment of verbless sentences in corpora are put forward, including the necessary developments and next steps in the contrastive study of the phenomenon.

PART ONE:

THEORETICAL FRAMEWORK

CHAPTER 1

THE PHENOMENON

What is a sentence without a verb? A verbal predicate is typically considered a necessary feature of sentential status. Verbless structures, such as (1) – (4) below, depart from the canonical model of a syntactic clause, i.e. a two-part structure consisting of a subject and a verbal predicate that is linked to the subject through inflection for agreement and tense.

(1) {Speaker opens the window:}

- a. English verbless: Fresh air.
- b. Russian verbless: Свежий воздух.

<i>svežij</i>	<i>vozduh</i>
ADJ.NOM	NN.NOM
<i>fresh</i>	<i>air</i>

(2) {Speaker pulls an envelope out of the mailbox and says:}

- a. English verbless: From France.
- b. Russian verbless: Из Франции.

<i>iz</i>	<i>francii</i>
PREP	NN.GEN
<i>from</i>	<i>France</i>

(3) {Speaker encourages a friend:}

- a. English verbless: You genius!
- b. Russian verbless: Ты гений!
ty genij
2SG NN.NOM
you genius

(4) {Announcement in a grocery store:}

- a. English verbless: Manager to the flower department.
- b. Russian verbless: Начальник в цветочный отдел.
načal'nik v cvetočnyj otdel
NN.NOM PREP ADJ.ACC NN.ACC
manager to flower department

As a result of this syntactic difference from the norm, verbless structures are typically referred to as ‘fragments’, ‘elliptical sentences’, ‘nonsentential clauses’, ‘nominal utterances’, but not full-fledged sentences in their own right. The present section starts by exploring the origins of the notion of predication and sentential status. It then proceeds to examine several existing accounts of verbless structures and the divergent uses of key terminology, including the term ‘ellipsis’ and the notion of a zero predicate. After analyzing the strengths and weaknesses of existing accounts, a working definition of the verbless sentence is presently proposed from a pre-theoretical perspective that makes minimal assumptions. For the current study, verbless sentences are formalized as strings of text that:

- a. are found in context
- b. do not include a finite verb, nor any other verb form
- c. are delimited by initial and final marking

This definition is explored for its advantages and potential limits. Finally, two fundamental types of verbless sentences are defended: elliptical and non-elliptical.

CHAPTER 2

PREDICATION AND SENTENTIAL STATUS

The possibility of a sentential structure without a verb stirs up timeless questions about the relationship between language and thought. A sentence is informally defined as the expression of a complete thought. A key task for any linguistic theory is to provide a precise definition of what constitutes this required **completeness**.

The first known account of **sentential requirements** dates over two thousand years. In the *Sophist*, Plato notes that the sentence, as opposed to ordinary words, necessarily involves combination:

[V]erbs and nouns do not make discourse if spoken successively [...] For instance, *walks, runs, sleeps* and the other verbs which denote actions [...] And again, when *lion, stag, horse*, and all other names of those who perform these actions are uttered, such a succession of words does not yet make discourse; for in neither case do the words uttered indicate action or inaction or the existence of anything that exists or does not exist, until the verbs are mingled with the nouns; then the words fit, and their first combination is a sentence, about the first and shortest form of discourse. (Plato, 360 B.C./1921: *Sophist* 262b–d)¹

Thus, the defining feature of a sentence is identified by Plato as the activity of combining two essential elements.

¹ Quotations:

- Changes to quotations are indicated in square brackets: [].
- **Boldface** has been used to add emphasis inside quotations; *italics* belong to the original quote.
- Translations from Russian quotations are provided in the footnotes; they have been translated by the present author, unless stated otherwise.

The necessity of the verb as one of the essential sentential elements is often traced to originate in these lines. Yet the requirement that is made by Plato is not necessarily the presence of a noun and a verb. The linguistic distinctions are made in the context of solving a paradox about using language to tell lies. The puzzle is that making a false statement, such as ‘*Theaetetus flies*’, requires using language, which refers to things that exist, in this case ‘*Theaetetus*’ and ‘*flying*’, in order to talk about things that don’t exist, i.e. ‘*a flying Theaetetus*’. To solve the puzzle, Plato insists on the need to treat the sentence as something that results from the combination of two essential parts, instead of something that names one homogenous whole:

Plato’s solution consists in [...] clarifying that making a “statement” (*logos*) is something complex and thus different from naming something [...] Even the simplest statement [...] is a combination of “a noun/subject” (*onoma*) and a “verb/predicate” (*rhema*), in which the latter states something of the named object. Therefore, to make a false statement (“Theaetetus flies”) is not to name a nonexistent object (flying Theaetetus), but to say of something that exists (Theaetetus) something which also exists (flying) but does not “combine” with it. [...] Plato’s pioneering distinction between the two essential components of a statement is a fundamental contribution. It is not clear, however, whether it amounts to a grammatical distinction (noun/verb), a syntactical one (subject/predicate), or an ontological one (thing/property). (Castagnoli & Di Lascio, 2012: 814)

Plato’s linguistic insight into sentences thus arises instrumentally from the need to explain how it is that language allows both true and false statements. His concern is not a linguistic definition of the elements of a sentence, and, as emphasized above by Luca Castagnoli and Ermelinda Di Lascio (2012), it is not at all clear **to what extent Plato’s writings associate the sentence with a verbal predicate**: the elements the interaction of which he is calling for may be a noun and a verb, a subject and a predicate, or a thing and a property. It should also be noted that there is no mention of whether or not the elements required must be explicitly stated.

The force and focus of Plato’s writings is not a description of the elements involved, but rather the idea that a sentence involves combination. His contribution to linguistics is the idea that a **sentence consists of saying something about something else**; and that these two elements – (a) the thing about which something is said and (b) the thing that is said – can be separated from one another. That is what allows him to explain true and false statements, i.e. a match between (a) and (b) results in truth, mismatch in falsity.

In linguistics, the act of combining these two essential elements – saying something about something else – has since come to be known as **predication**. The origin of the concept is

attributed to the above lines from Plato's *Sophist* (360 B.C./1921: 262b–d), whereas the first use of the **terminology** 'predication', 'predicate', 'to predicate' is subsequently traced to his student's, Aristotle's, *Analytica Priora* (Aristotle, 350 B.C./1989: Prior Analytics 24b18), e.g. in (Ildefonse, 1994: 18).

The notion of a sentence and the notion of predication are thus directly linked. A **sentence**, in linguistics, is typically identified as the place where predication occurs:

[L]a prédication, selon sa définition la plus courte et la plus concise, consiste à dire quelque chose de quelque chose, et cette façon même de décrire le phénomène implique un acte d'ajout. Ce qui nous intéressera essentiellement dans la prédication [...] c'est la division de nombreuses phrases en deux parties, (1) la partie faisant référence à la chose dont il est parlé, et qu'on appelle le sujet, et (2) ce qui est dit du sujet [...], à savoir le prédicat. (Gardiner, 1932/1989: 223–224, in Touratier, 2009: 17).

Plato's essential elements (a) and (b) have become the casual definitions for **subject and predicate**. The completeness of thought that is required for a sentence is found in the act of combining these two essential elements. However, more specific questions concerning the precise nature of these essential sentential elements – What constitutes a subject and a predicate? To what extent must they be explicit for sentential status to be granted to a particular string of text? – were not addressed by Plato, and, as witnessed in the special issues of the journals *Faits de Langues 31–32: La Prédication* (ed. Merle, 2009a; Merle, 2009b), *Revue de Linguistique et de Didactique des Langues 37: Syntaxe et Sémantique des Prédicats* (ed. Novakova & Guentchéva, 2008b; Novakova & Guentchéva, 2008a), *Syntaxe & Sémantique 6: Aux Marges de la Prédication* (ed. Behr et al., 2005b; Behr et al., 2005a), they remain a source of debate throughout history and to this day.

Predication, the defining requirement for sentential status, has gone through several developments since the ancient Greek philosophers. Aristotle is credited with inspiring the **classical logic** conception of the notion. Working with Plato's definition of the sentence as necessarily involving combination, Aristotle further develops the two essential elements involved. He characterizes (a), the element that is being talked about, as a necessary underlying basis that supports the predication and gives it the name 'hupokeímenon' (literally translated from Greek as 'subject'). Element (b), the thing that is said about the support, receives the name 'katégorêma' (literally 'predicate'). Predication itself is analyzed as the activity of combining these two elements which results in a judgement (also called 'proposition') that is either true or false. Analyzing Aristotle's contributions from a linguistic perspective, it is

important to note that, like Plato, Aristotle's goal in providing these definitions is also to defend the existence of false statements (Ildefonse, 1994: 18); this means that his focus, like Plato's, is not on language in general but on those aspects of language that make false speech possible. Furthermore, Aristotle makes clear that the essential elements discussed are not the words, but the actual physical **referents** of the words:

On n'insistera jamais trop à mon sens sur le fait que la théorie aristotélicienne de la prédication s'ancre dans le contexte philosophique précis d'une logique indissociable d'une physique. (Ildefonse, 1994: 19)

The classical logic conception of predication continues Aristotle's focus on the referents of the linguistic expressions, as opposed to the linguistic expressions themselves. Thus, in classical logic, predication is a relation between a subject *referent* and a predicate, which says something *about that referent*, and this relation can be evaluated as true or false. For instance, the statement '*Theaetetus flies*' from the classical logic perspective would be analyzed in terms of the subject referent of the word '*Theaetetus*', i.e. the man named Theaetetus, combining with the predicate referent of the word '*flies*', i.e. the physical activity of flying, and yielding a proposition which is false.

In characterizing the definitions from classical logic, it is notable that classical logicians follow the ancient Greek philosophers not only in the focus on the referent, but also in their primary concern with evaluating the **truth or falsity** of the link that is made between the subject and the predicate:

L'important pour le logicien est de savoir si la relation sujet-prédicat est vraie ou fausse et quelles sont les conditions de sa validité. (Maillard, 2009: 23)

Furthermore, the two elements of which predication is composed, i.e. subject and predicate, are of equal importance for the classical logic conception: the subject is just as valuable and necessary as the predicate (Feuillet, 2009: 133). For this reason, the conception is often called '**binary**'. Together, these three key characteristics are usually identified as posing problems for a direct transfer of the classical logic explanation to the field of grammar and language in general.

This transfer occurred during the middle ages. The classical logic conception of sentences was extended to the linguistic domain. The definitions of the logical subject and predicate, which were originally intended only for logical – i.e. binary, truth-evaluable, referential –

propositions, were applied to all types of sentences. As noted by Christian Touratier (2009), the **over-extension from classical logic** has led to much confusion in linguistics:

Les grammairiens vont d'abord emprunter à Aristote sa terminologie logique de sujet et prédicat, pour décrire non plus les propositions, au sens logique du mot, mais les phrases simples ou complexes qu'ils trouvent dans les textes. Puis, avec le temps ils vont faire passer dans le domaine proprement grammatical ces termes qui avaient, à l'origine, une portée exclusivement logique, ce qui infléchira sérieusement leur sens et risquera d'entraîner quelque confusion. (Touratier, 2009: 14)

As a consequence of using terminology that originated from a different domain, the linguistic definitions for subject, predicate and predication, must be clearly separated from the homonyms in the logic domain which have a much narrower scope than all of language.

Illuminating the ties that linguistic terminology has with logic, and tracing the developments of the definitions through history, Touratier (2009) insists that modern linguists define these terms with regard to the syntax above all else:

Si donc le linguiste, entend éviter tout risque de confusion ou de dérapage verbal, il donnera une portée exclusivement grammaticale aux deux termes de sujet et prédicat, en désignant par-là la fonction syntaxique que remplissent, l'un par rapport à l'autre, les deux constituants immédiats de la phrase exocentrique, la phrase exocentrique étant en effet ce qui correspond, pour le linguiste à la proposition des logiciens. Et il désignera les fonctions logiques que les logiciens classiques appellent sujet et prédicat à l'aide des deux termes de thème et rhème, ce qui permettra de dire, sans la moindre contradiction, qu'un sujet, au sens syntaxique, peut être le thème d'un énoncé, mais que le thème d'un énoncé peut être autre chose qu'un sujet au sens syntaxique. (Touratier, 2009: 16)

His suggestion is that that the **logical elements** (a), the element talked about, and (b), what is said about the element, should be renamed; in linguistic discussion, these key logical elements are to be referred to as 'theme' and 'rhema', as opposed to their original classical logic names 'subject' and 'predicate'.

In turn, linguistic, as opposed to logical, use of the homonymous terminology is typically said to carry at least three characteristic features. In the following, John Lyons (1980) clearly distinguishes the first key particularity of the **linguistic use** of the terms subject and predicate:

Il faut, [...], premièrement établir une distinction entre les expressions et leurs référents. Selon cette distinction on pourrait dire que dans l'énoncé

(1) Jean est sorti

le sujet est soit l'expression 'Jean' soit le référent de 'Jean' (c'est-à-dire Jean). En général, le terme de sujet est appliqué par les linguistes aux expressions et non à leurs référents. Dans ce que nous avons dit jusqu'ici sur les sujets et les prédicats, nous avons adopté cette convention terminologique [...] (C'est le discours des linguistes). Si l'on se reporte à une autre convention [...], selon laquelle les propriétés sont attribuées à des entités en prédisant des expressions de ces dernières, il s'ensuit que le prédicat n'est pas prédiqué du sujet, mais du référent du sujet. (C'est le discours du logicien). (Lyons, 1980/1978: 132–133, in Touratier, 2009: 15)

Thus, the first feature of the linguistic use of the terminology is a **break with referentiality**. This means that it is the expression, as opposed to the referent of the expression, with which the linguist must be concerned. For example, in '*John has left*', the subject is the linguistic expression '*John*', not the person who is the referent of the expression. The predicate '*has left*' also corresponds to the words, not the referenced activity. Moreover, the predicate expression says something about the linguistic subject – the expression '*John*', not about the logical referent.

The second particularity concerns the linguistic subject, which is to be identified with **only the head**. The linguistic subject is stripped of the words that accompany it. For instance, in '*The chocolate covered birthday cake is all mine*', it is only the noun '*cake*' that is considered the subject for grammatical and linguistic purposes. The complete noun phrase '*the chocolate covered birthday cake*' corresponds to the logical subject and its referent.

Finally, it is above all the **syntactic function** that counts in identifying the subject of the sentence for linguistic purposes. For example, in '*It is the syntactic function that counts*', the grammatical subject of the sentence would be the dummy-subject '*it*', as opposed to the real or logical subject '*syntactic function*'. Non-referential elements, such as the empty '*it*', are perfectly coherent as subjects from the perspective of syntactic function.

Another issue is to what extent the linguistic notion of predication is **decomposable**. In logic, predication is the union of the logical subject and predicate. The dichotomy in the logical conception is necessary for the goals set out: it is only the union of two referential elements that can be evaluated for truth or falsity. In linguistics, the focus is shifted from the logical physical referents to the syntactic expressions. One may thus wonder about the purpose of extending the binary requirement, since in any case syntactic expressions cannot be truth-evaluable independently from their logical referents. For instance, the linguistic expressions '*Theaetetus flies*' and '*Theaetetus doesn't fly*' will receive truth values depending on the correspondence between the physical man Theaetetus and the physical activity of flying. The correspondence between the word '*Theaetetus*' and the word '*flies*' cannot have truth or falsity

without the physical world: from a purely syntactic perspective ‘*Theaetetus flies*’ and ‘*Theaetetus doesn’t fly*’ are identical in terms of truth value. If physical world reference is removed from the linguistic definitions, it is not clear what purpose the binary requirement serves. Since the linguistic definitions of subject, predicate and predication are focused on the syntactic expressions, truth-evaluability is no longer the aim, nor is it possible without consideration of the referents. Thus, the original logical purpose of the subject and predicate dichotomy, i.e. truth-evaluability, does not transfer to linguistics.

Nonetheless, the binary definition proposed by classical logicians for truth-evaluable propositions was transferred to the linguistic definition of a sentence and, as Novakova and Guentchéva (2008a) point out below, it has influenced many linguists and had a great impact on modern syntax:

De très nombreuses grammaires traditionnelles et, en grande partie, les travaux de syntaxe modernes partagent encore ce principe fondamental de décomposition tout en admettant qu’à chaque phrase déclarative simple, constituée d’un sujet nominal et d’un prédicat verbal, peuvent s’adjoindre en plus des constituants facultatifs (de lieu, de temps, de cause, etc.). (Novakova & Guentchéva, 2008a: 6)

Upon this ‘decompositional’ definition, the main concern of predication becomes dividing the sentence into a subject and a predicate, and most typically, a nominal subject and a verbal predicate. This binary decomposition was set up as a powerful precedent for linguists by the influential *Grammaire de Port-Royal* (Arnauld & Lancelot, 1660). As seen in the definitions below, the transposition from classical logic is direct:

[L]e jugement s’appelle aussi proposition, [et] il est aisé de voir qu’elle doit avoir deux termes : l’un de qui l’on affirme, ou de qui l’on nie, lequel on appelle sujet; [et] l’autre que l’on affirme, ou que l’on nie, lequel s’appelle attribut ou praedicatum. (Arnauld & Nicole, 1662: 156, in Touratier, 2009: 13)

The Port Royal grammar treats the subject and predicate as carrying equal weight in the constitution of the grammatical sentence, just like they do in classical logic propositions. As Jack Feuillet (2009: 133) points out, the influence of this ‘egalitarian dichotomy’ on more recent work can be seen in the early generative grammar where the combination of a nominal subject and verbal predicate is said to create a brand-new structure that is not headed by any single constituent, i.e. the model ‘*Sentence* → *Noun Phrase* + *Verb Phrase*’.

2.1 MODERN LOGIC

Following nineteenth century advances in mathematics, **modern logic** has fine-tuned the ‘principle of decomposition’ by replacing the subject-predicate egalitarian dichotomy with a mathematical function. In the field of logic, Gottlob Frege, in the *Begriffsschrift* (1879), develops a ‘formula language’ that strives to eliminate the ambiguity of ordinary language in order to facilitate logical analysis:

In studying the concept of number, Frege was confronted with difficulties when he attempted to give a logical analysis of the notion of sequence. The impression and ambiguity of ordinary language led him to look for a more appropriate tool; he devised a new mode of expression, a language that deals with the “conceptual content” and that he came to call “Begriffsschrift”. This ideography is a “formula language”, that is, a *lingua characterica*, a language written with special symbols, “for pure thought”, that is, free from rhetorical embellishments, “modeled upon that of arithmetic”, that is, constructed from specific symbols that are manipulated according to definite rules. (Van Heijenoort, 1967: 1)

In this work, Frege re-defines the logical predicate as a function that operates over a specific number of necessary arguments, one of which is the logical subject. The logical predicate thus becomes the organizing element of the logical structure. The logical subject becomes just one of the arguments over which the predicate operates. The classical egalitarian balance between two elements is thus replaced with a hierarchical model in which the logical predicate is primary.

It must be emphasized that Frege is very critical of the link between logic and language and develops his new model precisely to distance his chosen subject matter from the subject matter of linguists:

I hope that logicians, if they do not allow themselves to be frightened off by an initial impression of strangeness, will not withhold their assent from the innovations that, by a necessity inherent in the subject matter itself, I was driven to make. These deviations from what is traditional find their justification in the fact that logic has hitherto always followed ordinary language and grammar too closely. In particular, I believe that the replacement of the concepts *subject* and *predicate* by *argument* and *function*, respectively, will stand the test of time. (Frege, 1879/1967: 7)

Putting aside the question of whether it is the logical domain or the linguistic domain that overextends its influence, or perhaps even a coincidence of similar thoughts, it is clear in the

above that Frege proposes the innovations in logic precisely to divorce logic from ordinary natural language. Nonetheless, a break between the two disciplines does not seem to be in the cards. Many linguists see advantages in updating linguistic models to the new models proposed in modern logic.

The modernization in the domain of logic has been mirrored in the **linguistic domain** and extended to sentences. Novakova and Guentchéva (2008a) summarize the transfer to linguistics and note the flexibility that the new conception provides, particularly in accounting for quantifiers and conjunctions:

La prédication est ainsi analysée comme une *relation* associant un *prédicat* à un nombre n d'arguments [$P(x)$ pour les prédicats à une place, $P(x, y)$ – pour les prédicats à deux places, $P(x, y, z)$ – pour ceux à trois places]. La possibilité de combiner la conceptualisation d'un prédicat sous forme d'une fonction à plusieurs arguments avec la quantification universelle ou existentielle pour analyser par exemple les quantificateurs comme *tout, tous, certains* dans *tout homme, tous les hommes, certains hommes*, ou encore les connecteurs *et, ou, ne...pas*, est considérée comme l'un des apports majeurs de G. Frege dans ce domaine. (Novakova & Guentchéva, 2008a: 7)

The change in models has been concisely represented as ' $S / P \rightarrow f(x, y)$ ' in (Sériot, 2000: 249). The linguistic model of the sentence is replaced from '*Sentence* \rightarrow *Subject + Predicate*' to '*Sentence* \rightarrow *Predicate (Arguments)*'. For instance, the linguistic analysis of '*I love you*', modeled on the conception stemming from classical logic as the subject '*I*' combining with predicate '*love you*', is re-modeled following modern logic developments as the predicate function '*love*' operating on and relating the arguments '*I*' and '*you*': *love (I, you)*.

In terms of the definition of predication overall, for linguistics the change from classical logic is not as drastic as it may seem, at least not for the binary factor. In discussing the new model, which is also referred to as 'nuclear predication', Touratier (2009) insists that:

cela ne change pas grand-chose à l'idée générale que le linguiste peut se faire de la prédication (Touratier, 2009: 16)

As noted by Touratier (2009), the principle of decomposition withstands modernization. The division of sentences in terms of subject and predicate is replaced by the division in terms of argument and function, but the idea that predication is created from explicit parts of the sentence being put together, persists. Concerning the linguistic definition of predicate, Claude Muller (2013) explains that it also keeps its fundamental trait, i.e. the predicate remains that which combines with the subject:

Au total, *prédicat* gardera son ancienne signification, celle de la logique classique : tout ce qui complète un sujet pour faire une proposition. Mais dans la logique rénovée, comme dans la linguistique naissante, il désigne aussi le centre fonctionnel régissant des variables qui sont les arguments (aussi bien le sujet que les compléments), dans le *calcul des prédicats*. (Muller, 2013: 52)

What changes is the repartition of the value between the subject and the predicate. A hierarchy is created. The predicate takes on the role of being the organizing element that creates the relation between the other parts of the sentence, one of which is the subject. Treated as a relational element, the predicate becomes primary. Conversely, the role of the subject in predication is deemphasized; the subject becomes just one of the elements that undergoes the relation and is thus relegated to secondary status.

2.2 QUESTIONING EQUALITY FROM A LINGUISTIC PERSPECTIVE

The egalitarianism between the predicate and the subject in predication has also been put into question for purely linguistic reasons, not related to the field of logic. The work of Lucien Tesnière (1959) is strongly critical of the subject's role in linguistic definitions of the sentence:

Dans aucune langue, aucun fait proprement linguistique n'invite à opposer le sujet au prédicat [...] D'autre part, il est difficile de mettre sur un pied d'égalité le sujet qui ne contient qu'un seul mot, et qui peut même n'être pas pleinement exprimé, avec le prédicat, dont l'énonciation est obligatoire [...] L'opposition du sujet et du prédicat [...] conduit à isoler comme sujet un des actants, à l'exclusion des autres, lesquels se trouvent rejetés dans le prédicat pêle-mêle avec le verbe et tous les circonstants. C'est là accorder à l'un des éléments de la phrase une importance disproportionnée, qu'aucun fait strictement linguistique ne justifie. (Tesnière, 1959: 103–105, in Feuillet, 2009: 134)

Tesnière (1959) argues that the importance given to the one-word subject is purely logical, not based on any linguistic observations, and disproportionate since, unlike the predicate, the subject does not need to be fully expressed. Though he distances himself from the logical terminology, it is the predicate that is central in Tesnière's dependency grammar model of sentence structure. This hierarchical model, consisting of a governing 'root' and its dependent

‘actants’, is, in its basic essence, very similar to the hierarchical model from modern logic, as explains Muller (2013).

Le prédicat est alors proche de la fonction mathématique du terme qui nécessite l’attribution de valeurs à des arguments qu’il introduit dans sa dépendance. [...] à la nuance importante près que Tesnière n’utilise pas ce terme et emploie une terminologie purement grammaticale : le verbe et ses actants. Les actants sont distingués des circonstants (facultatifs) : il s’agit bien de l’équivalent grammatical des arguments. Le centre d’analyse grammaticale est le verbe, dont le sujet n’est qu’un actant parmi d’autres. (Muller, 2013: 53)

In Tesnière’s linguistic model, like in modern logic, a central organizing element relates the rest of the sentence. This organizing role is given to the grammatical category of the verb.

While others also rejected the subject-predicate division, Tesnière was one of the first in modern linguistics to clearly replace this division with verb centrality. (Kahane & Osborne, 2015: 11)

Sentential structure on this account is defined by the presence of the finite verb together with the arguments that depend on it.

In the same vein, Denis Creissels conceives of the predicate as a relational function carried out by the verb. Driven by the desire to emphasize the dependent status of the subject, Creissels (2004) elevates the verb phrase by defining it as an unsaturated sentence:

Il est notamment difficile de soutenir l’existence d’un groupe verbal réunissant le verbe et les termes de sa construction autre que le sujet dans les langues qui placent systématiquement le verbe en tête de phrase et le sujet immédiatement après le verbe. [...] Les constituants nominaux ont en principe pour tête un nom, les phrases et constituants phrastiques ont en principe pour tête un verbe; en effet, dans cette optique (qui est celle qu’adopte ce cours), le groupe verbal n’a pas à être défini de manière indépendante, car ce n’est ni plus ni moins qu’une phrase non saturée (c’est-à-dire à laquelle manque un constituant nominal en fonction de sujet pour être une unité phrastique complète). Ce qui sur un plan théorique crée des difficultés, c’est la conception selon laquelle le sujet est extérieur à la construction maximale ayant pour tête le verbe. (Creissels, 2004: 11)

Creissels (2004) argues that the verb phrase is much more than a structure with a verb as the head. The verb is rather the head of the entire sentential structure. While the noun is the maximal value of a *noun phrase*, i.e. a structure consisting of the noun and its dependent elements, the verb is the maximal value of the *sentence*, i.e. a structure consisting of the verb and any dependent arguments. The latter include the subject and any other complements

(Creissels, 1995: 45). The subject on this conception is a noun phrase that enters the structure as required by the verb, together with other necessary arguments determined by the verb, and, once the necessary requirements of the verb are filled, the structure acquires sentential status. Since it is the finite verb that unites the arguments of the sentence, the finite verb is always necessarily the predicate in the model of predication proposed by Creissels (Novakova & Guentchéva, 2008a: 4).

Furthermore, Creissels makes a strong claim against the ability of nominal constituents to carry out the predicate function. The predicate is that which structures the nominal constituents into a sentential unit; it is that which remains once the nominal constituents are removed:

[O]n reconnaîtra comme expression prédicative ce qui reste une fois dégagés les constituants nominaux ou quasi nominaux. (Creissels, 1995: 42)

As a result, the idea that a predicate may be nominal is contradictory to the proposed model:

Une conséquence méthodologique importante de cette approche est que, si on développe de façon tout à fait cohérente la notion de prédicat comme élément qui structure en unité phrastique un ensemble de constituants nominaux, on doit rejeter comme contradictoire dans ses termes mêmes la notion de « prédicat nominal » : dans la mesure où on maintient distinctes l'une de l'autre la notion discursive de propos et la notion syntaxique de centre organisateur de l'unité phrastique, et où on réserve le terme de prédicat pour cette dernière notion, on doit conclure que les notions de constituant nominal et de prédicat sont complémentaires, et qu'un même fragment d'énoncé ne saurait simultanément être reconnu comme constituant nominal et comme prédicat. [... II] peut certes y avoir là des problèmes d'analyse délicats dans la description de telle ou telle langue, mais cela ne remet pas en cause le principe de l'incompatibilité entre le statut de constituant nominal et celui de prédicat. (Creissels, 1995: 48)

Creissels claims that to treat a nominal element as a predicate is to confuse the syntactic predicate, i.e. organisational center of the sentence, with the basic discursive notion of something said about a theme:

La notion de prédicat nominal ne peut en réalité être justifiée qu'au prix d'une confusion entre élément organisateur de la phrase et propos développant un thème. (Creissels, 1993: 76, in Novakova & Guentchéva, 2008a: 11)

Even a semantically empty copula verb syntactically carries out the predicative function. For instance, '*Michel is the son of Jean and Marie*' would be analyzed in terms of two nominal constituents – the arguments '*Michel*' and '*the son of Jean and Marie*' – linked together by the

verb ‘is’. For Creissels, nominal constituents and the predicate are by definition incompatible (Creissels, 1995: 49).

In spite of such strong claims concerning verb centrality, Creissels does attempt to accommodate the phenomenon of verbless structures within the verb-centered model of the sentence. He does this in several ways. The first may be generally characterized as the ‘**verbal paraphrase approach**’ or ‘**variant one of the hidden verbal structure approach**’. This is when a verbless sentence is paraphrased into some verbal form and then, based on this paraphrase, assigned an underlying abstract structure which is verbal. Thereby, the dangerous contradiction posed by verbless structures for verb-centered sentential models is neutralized. The absence of the verb is treated as a surface-level syntactic mirage and does not interfere with the model: the verbal requirement for sentential status is satisfied by the verbless structure via the verbal paraphrase which is said to reveal its real abstract structure. The verb-centered model is thus saved – it can be extended to all language; even non-verbal phenomenon that at first does not appear to be within its jurisdiction is only superficially so. However, the cost of this approach is that the actual verbless structures found in natural language receive only marginal status as sentences: at their core they are attached to a verbal structure which is what actually gives them the predicative status.

Creissels (1995: 36–38) uses this paraphrase approach when analyzing a common French verbless structure illustrated in ‘*intéressant, ce livre*’ (lit. ‘*interesting, this book*’). He treats the structure as a semantically complete assertion that is associated with the conceptual event of ‘the fact of a book to be interesting’. He then explains that this event can be reworded in another way that is just as semantically complete as the verbless structure – the verbal ‘*ce livre est intéressant*’ (lit. ‘*this book is interesting*’). The verbal form is more flexible syntactically since it can be embedded as a subordinate clause, i.e. ‘*je crois que ce livre est intéressant*’ (lit. ‘*I believe that this book is interesting*’), while the verbless one cannot, i.e. *‘*je crois qu’intéressant, ce livre*’ (lit. *‘*I believe that interesting, this book*’). Creissels explains that the marginal syntactic status of the verbless structure is due to the limitations with embedding and syntactically manipulating it. In essence, for Creissels there is one conceptual event, several syntactic forms to capture it, and a competition between these forms for the “maximum manipulations of the constituents and maximum mechanisms of integration into complex structures”: the one that is most syntactically flexible is recognized as the basic underlying abstract form (i.e. “*unité phrastique de base*”).

De telles limitations aux possibilités syntaxiques de (72) [*intéressant, ce livre*] suggèrent comme explication qu'il s'agit d'un énoncé construit selon un schème qui prévoit la présence du verbe *être*, mais que dans certaines conditions, il n'est pas obligatoire que le verbe *être* soit effectivement présent. C'est notamment le cas lorsque le sujet est représenté par l'indice correspondant et reporté en fin d'énoncé, dans une position où l'énonciateur le fait apparaître comme le rappel d'un élément thématique – ex. (76) [(*il est*) *intéressant, ce livre*].

Autrement dit, (72) [*intéressant, ce livre*] appartient à une famille d'unités phrastiques à l'intérieur de laquelle il est justifié de reconnaître dans (73) [*ce livre est intéressant*] l'unité phrastique de base. On peut voir ainsi de manière plus générale qu'il n'y a **pas véritablement de contradiction** entre le fait évident que le français abonde en type d'énoncés dépourvus de formes verbales et la position traditionnelle en syntaxe française selon laquelle **le verbe est nécessaire à la bonne formation** des unités phrastiques. Cette position demande seulement à être précisée et nuancée : à partir d'observations comme les précédentes, on peut accepter que, si la présence effective d'un verbe dans l'unité phrastique réalisée ne constitue pas en français une nécessité absolue, par contre **la structure abstraite** rendant compte d'une famille d'unités phrastiques du français **représentant un même événement conceptualisé** comporte nécessairement une position qui est généralement occupée par une **forme verbale** et qui ne peut être laissée vide que sous certaines conditions. (Creissels, 1995: 37–38, emphasis added)

Creissels thus gives the verbless structure '*intéressant, ce livre*' the underlying verbal structure '*il est intéressant, ce livre*' (lit. '*it is interesting, this book*'). The latter structure is itself a manipulation of the basic underlying abstract form '*ce livre est intéressant*'. It has been allowed to omit the verb and subject ('*il est*') as a result of the movement of the logical subject to final position, but the basic verbal sentential structure remains intact on an abstract level. Despite the absence of the verb, the syntactic structure is still verbal and semantically complete. The contradiction is neutralized and the verb-centered model is saved. The verbless structure receives marginal sentential status, i.e. indirectly through its verbal paraphrase. In summary, sentential status is given to the verbless structure based on syntactic manipulation, including movement and deletion, of a hidden verbal structure.

The second approach may be called the '**verbal placeholder approach**'. This approach is a variant of the hidden verbal structure approach. Using this approach Creissels (1995: 50–51) targets those verbless sentences that are capable of being embedded without requiring the introduction of a verb. He gives the example of a verbless structure in Hungarian, '*érdekes ez a könyv*' (lit. '*interesting, this book*'), that may be embedded as a subordinate clause without needing to be altered by re-establishing a verb.

Un type d'énoncé qui diffère de l'exemple français (72) [*intéressant, ce livre*] au sens où il n'y aurait aucune justification à voir là la version « tronquée » d'une unité phrastique comportant une forme verbale dont le rétablissement serait nécessaire pour que puissent fonctionner certains mécanismes syntaxiques. [...] En particulier, aucune adjonction d'une quelconque forme verbale n'est nécessaire pour insérer une telle séquence à un contexte comme *nem tdja hogy ...* « il ne sait pas que ... » (Creissels, 1995: 50)

Though such a structure does not have an explicit predicate that surfaces in syntactic manipulation, it is also given sentential status in Creissels' model. Sentential status is granted to even such 'embeddable and still verbless' structures in what can be condensed to three essential steps.

First, it is necessary to establish that none of the explicit elements of the structure is itself a predicate. In other words, it must be proved that neither the noun phrase '*this book*' nor the adjective '*interesting*' function as a verb. To do this, Creissels proposes the 'Question Test' according to which the central difference between nominal constituents and verbs is that verbs may not be questioned (Creissels, 1995: 47). The fact that each of the elements of the structure may be replaced with a question word, i.e. '*ce livre*' with '*quoi?*' (lit. '*what*') and '*intéressant*' with '*comment?*' (lit. '*how*'), shows that they are nominal – not verbal – constituents. In this way it is determined that none of the explicit elements are themselves a predicate.

The second step requires accepting that a syntactic structure is made not only of explicit but also of inexplicit 'hidden' elements. Creissels insists that the status of 'syntactic completion' may be granted to a structure even when the structure does not contain an explicit element that is recognisable as the predicate:

[I]l n'y a pas lieu de poser, comme on le fait souvent, que toute phrase doit comporter un terme **reconnaisable comme** prédicat. On peut certes accepter l'idée que tout énoncé comporte normalement un propos, mais il n'y a aucune raison d'en déduire que tout énoncé **syntactiquement achevé** doit présenter **explicitement** un terme répondant à la notion de prédicat telle qu'elle est définie ici. On peut concevoir (et bien des langues exploitent cette possibilité) qu'un énoncé vérifiant la propriété de complétude syntaxique **se présente dans sa réalisation** comme une simple juxtaposition de constituants nominaux. (Creissels, 1995: 50, emphasis added)

Thus, syntactic completion may be achieved not only through explicit recognizable elements; something as central as the predicate may be inexplicit. This important step allows inexplicit 'hidden' elements to create the predicative expression and come to the rescue of structures that do not fit the model.

The third crucial step is to justify the actual presence of the predicate on this hidden level. That is, it must be proved that the nominal constituents, though they appear as simple juxtaposed elements, are actually related to one another through a predicate that indeed exists in the particular structure but has just been left unexpressed. This placeholder for the predicate – i.e. the element that relates the nominal elements together and allows integrating into the model even those verbless structures that remain verbless through embedding – Creissels refers to as the ‘ \emptyset ’.

The justification for the existence of this placeholder is also based on a paraphrase of the verbless structure. However, the paraphrase this time involves, instead of embedding, a change of temporal meaning.

[C]ette possibilité de questionnement sur chacun des deux constituants d’un tel énoncé prouve qu’on a bien affaire à une expression prédicative à deux variables, mais à une expression prédicative qui a la particularité de **présenter comme** la simple juxtaposition des deux variables, la place que **pourrait occuper un élément prédicatif explicite étant laissée vide**. L’expression prédicative sous-jacente à cet énoncé peut être figurée comme *milyen \emptyset ki/mi* « *qui/quoi est comment* », la **justification du \emptyset** qui apparaît dans cette formule résidant dans le fait que **l’apparition d’un élément prédicatif explicite accompagne nécessairement la variation de sens** qui ordinairement se traduit en hongrois par le remplacement d’une forme verbale de « présent » par une forme verbale de « passé » – ex. (101) [*érdekes volt ez a könyv*, « *ce livre était intéressant* »]. (Creissels, 1995: 51, emphasis added)

To prove that the predicate placeholder actually exists, Creissels paraphrases the structure so that it references a time that requires an explicit tensed verb. For instance, changing the meaning of the verbless structure to the past is accompanied in Hungarian by the necessary introduction of an explicit past tense verb. On the basis of this change of meaning, an unarticulated verb form is said to exist inside the verbless structure. The \emptyset predicate is positioned in the place of the verb in an explicitly tensed structure.

Thus, a verbless structure may present itself as a simple juxtaposition of elements, it may remain verbless even though embedded, but the juxtaposed elements are nevertheless linked together into one syntactically complete predicative expression – not by an explicit verb, and not thanks to the characteristics of the explicit elements, but by means of a hidden structure which contains a hidden verbal predicate. This \emptyset placeholder substitutes for the verb, relates the nominal constituents together into an instance of predication, and saves the verb-centered model.

Finally, for those verbless structures for which the above strategies do not work, Creissels proposes a third approach. This approach does not aim to integrate the structures into the verb-centered model. It may be referred to as the ‘**non-verbal predicate approach**’ since it admits predicative status to explicit elements that are not verbs. Cases that withstand all of the variants of paraphrasing, that is (a) the verb doesn’t surface during embedding and (b) altering the temporal reference also doesn’t result in the introduction of a verb, are treated as instances of “non-verbal predicates”.

[I]l n’est pas impossible que l’on soit amené à reconnaître dans une langue l’existence d’un nombre limité d’expressions prédicatives trop particulières pour pouvoir se rattacher à un schème productif de formation d’expressions prédicatives. On pourra si c’est le cas les désigner comme « prédicats non verbaux ». Mais il est raisonnable de s’attendre à ce que **l’immense majorité des expressions prédicatives d’une langue, sinon toutes, se laissent ramener à un nombre limité de schèmes de formation (éventuellement même, à un schème unique de formation)** mettant en jeu un choix lexical. En effet, une langue dans laquelle ce ne serait pas le cas serait une langue dans laquelle les expressions prédicatives ne pourraient être qu’en nombre très limité. La chose n’est pas tout à fait inconcevable, mais il est toutefois permis de douter qu’une telle langue existe. (Creissels, 1995: 56, emphasis added)

In the above lines, Creissels hesitantly allows the possibility that an element other than a verb may carry out the predicative relational function. However, he emphasizes that such structures based on a non-verbal predicate are highly unlikely. He expects that the majority of the sentences in a language can be recognized as belonging to a small number of structures, perhaps even a single one, in which either explicitly or by means of a hidden structure there exists a verb. Structures based on a non-verbal predicate must withstand syntactic variation to show that a verb does not exist on a hidden level; and if they make it through, the potential number of such non-verbal structures is expected to be so limited that their existence does not pose any danger for the verb-centered model.

It is clear that a verb-centric sentential model must resolve the existence of verbless structures in natural language and simultaneously maintain the necessity of the verb. The replacement of logical definitions with grammatical categories is a balancing act not only for classical egalitarian subject / predicate models, but also for the hierarchical models where the predicate is central.

2.3 THE RUSSIAN LINGUISTIC TRADITION

In Russia, the linguistic discussion faces the same enemy when it comes to defining the sentence – the **conflict between logic and grammar**. Furthermore, certain features of the Russian language – principally, the commonplace use of ‘impersonal structures’ – have led to the fact that the debate about the hierarchy between the subject and the predicate surfaced in Russian linguistic literature even earlier than Frege’s (1879) modernization in logic.

In the historical outline provided by Patrick Sériot (2000: 240), the conflict between natural language and logic flared up in Russia in the 1860’s as a consequence of the dominant and richly illustrated grammar of the time – that of Fëdor Ivanovič Buslaev (1858). The grammar, in theory, used the subject / predicate classical logic conception, but, in reality, drew attention to the problems of applying the model to actual language data. The indiscreet culprits propelling the conflict to arise so early in Russia are **impersonal sentences** which are structures that do not have what is typically defined as the subject.

The structures at issue do not have a noun in the nominative case. For this reason, they are typically treated as structures without a subject. Example (5) illustrates the contrast between (a) a canonical structure with the nominative case ‘я’ (*ja*; lit. ‘I’), and (b) an impersonal structure with the dative case ‘мне’ (*mne*; lit. ‘me’).

- (5) a. Я не сплю
 ja *ne* *splju*
 I.NOM *not* *sleep*
- b. Мне не спится
 mne *ne* *spitsja*
 me.DAT *not* *sleep.REF*

While the nominative case morphologically marks a noun as having the subject function, the dative indicates that it functions as an indirect object. Impersonal structures such as (5b) ‘Мне не спится’ (*mne ne spitsja*; lit. ‘*me.DAT not sleep.REF*’) are sometimes analyzed as presenting the events from an objective, as opposed to a subjective, perspective and are said to have an ‘experiencer’ instead of a ‘subject’ (Sériot, 2000: 237–242).

Impersonal sentences are so common in Russian that the contradiction with a classical logic model emerges forcefully and almost immediately upon comparison with natural

language. The clash between the theory and the data showing the absence of one of the essential elements – the nominal subject – could not be ignored:

Il y avait en particulier une chose étrange, qui intriguait beaucoup les grammairiens : c'est la coexistence de deux types de propositions, très différents. Il s'agit, bien sûr, des propositions « personnelles » à deux éléments, et des propositions « impersonnelles », qui, en termes traditionnels, semblaient comme privées d'un des deux éléments obligatoires et nécessaires de toute proposition : le sujet. C'est autour de cette anomalie logique que tournait l'essentiel de la discussion. On voit qu'un fait propre au russe faisait surgir une question qui à l'époque se posait dans tous les pays d'Europe : que devait-on considérer comme primaire dans la langue : le nom (et la fonction de nomination) ou le verbe (et la fonction de prédication)? (Sériot, 2000: 240)

As explained by Sériot (2000), the wide use of such structures led the Russian linguistic discussion to question a conception based on equality between the subject and the predicate in advance of the developments in modern logic and do so for purely linguistic reasons.

2.3.1 The Verb-centric Syntactic Model in Russia

The pressing need to account for impersonal structures contributed to bringing about a **verb-centric syntactic model** in Russia. At the time, the Russian verb, due to its unique and rich aspectual features, was already in the process of being distinguished as central to the analysis of Russian, as argued most notably by Konstantin Sergeevič Aksakov (1855). As shown by Sériot (2000), the dissonance between the subject / predicate model and the prevalence of impersonal propositions resulted in a rejection of equality in favour of a predicate dominant model.

Les linguistes du courant slavophile [...] proposaient une analyse différente de la structure de la proposition, selon laquelle c'était le verbe (=prédicat) qui était le centre absolu de la proposition, alors que les noms (y compris le sujet) occupaient une position de *dépendance*. Le sujet était donc traité comme une *variété de complément*, étant considéré comme un membre secondaire de la proposition. (Sériot, 2000: 243)

The nominative subject is deemphasized and treated as one of the complements, while the verbal predicate is given primacy as the central organizing element of the sentence. This revolutionary syntactic conception is put forward in 1877 by Aleksandr Alekseevič Dmitrevskij (Sériot, 2000: 243; Kornilov, 2017). Rejecting the classical logical binary model

for its inability to account for impersonal structures, Dmitrevskij insists on the secondary role of the subject and argues for the autonomous sufficiency of the verbal predicate:

Появляется взгляд, что подлежащее не равноценно со сказуемым, что «подлежащее не может считаться одним из главных членов предложения, а должно быть низведено в разряд второстепенных, и именно дополнений» (8). [Note (8) references A. A. Dmitrevskij, “*Практические Заметки о Русском Синтаксисе*” (*Practical Notes on Russian Syntax*) which was published as a series of articles in several issues of the journal “*Филологические Записки*” (*Philological Notes*), including: 1877(3): 1–15, 1877(4): 15–37, 1878(1): 37–61, 1878(2): 61–76, 1878(4): 79–89, 1878(6): 15–27, 1880] (Vinogradov, 1958: 290)²

Сказуемое есть неограниченный властитель, царь предложения : если есть в предложении, кроме него, другие члены, они строго ему подчинены и от него только получают свой смысл и значение; если нет их, даже подлежащего, сказуемое само собой достаточно выражает мысль и составляет целое предложение. Иначе сказать: и само предложение есть не что иное, как сказуемое или одно, или с приданными ему другими членами. (Dmitrevskij, 1877b: 23)³

Dmitrevskij’s conception is strikingly similar to that which will later be proposed by Tesnière in 1959 (Храковский, 1983; Gasparov, 1995: 134; Sériot, 2004; Kornilov, 2017).

A fascinating fact emerges from these historical considerations. It must be noted that Russian is known not only for impersonal sentences, which *challenge the status of the subject*, but also for verbless sentences, which *challenge the status of the predicate*. Yet, the effort to account for the widespread use of impersonal sentences in Russian led most of the linguistic community to **focus on the inconsistencies with only one of the essential elements** – the nominal subject.

The **reasons** for the attention to impersonal sentences and the status of the subject, as opposed to verbless sentences and the status of the predicate, could be multiple.

Conceivably, it would have been drastic to attack the natural language inconsistencies of both the subject and the predicate at once. After all, at the time, Dmitrevskij’s proposal to

² Translation: « [L]e sujet n’est pas à mettre sur le même plan que le prédicat, mais doit être relégué (*nizvedeno*) au rang des membres secondaires de la proposition, c’est-à-dire des compléments. » (Translated in Sériot, 2000: 244)

³ Translation: « Le prédicat est le maître illimité de la proposition : s’il y a en dehors de lui d’autres membres de la proposition, ils lui sont strictement subordonnés et ce n’est que de lui qu’ils reçoivent leur sens; s’il n’y en a pas d’autres, même pas de sujet, le prédicat à lui tout seul suffit à exprimer la pensée et constitue une proposition entière. En d’autres termes, la proposition n’est autre chose que le prédicat, tout seul ou accompagné des autres membres qui lui sont rattachés. » (Translated in Sériot, 2000: 244)

demote the subject received very mixed **reactions** (so mixed in fact that, as shown by Nikolaj Kornilov, 2015, the pioneering work of what is now a widely regarded model is to this day often attributed to a scholar named A. A. Dmitrievskij).

A summary of the reactions to Dmitrevskij's proposal can be found in (Kornilov, 2017), but one heated exchange, concerning *impersonal sentences*, draws particular attention for its **peculiar resemblance** to a multi-century echo of present-day arguments concerning the use of another type of problematic structure that is found in natural language – the *verbless sentence*.

At the heart of both issues is a challenge inadvertently thrown to logic-based syntactic sentential models by **real language use**. It starts with Dmitrevskij (1877b) criticizing the accepted sentence model of the time for avoiding inconvenient data:

[Л]ишь дело коснётся рассмотрения состава предложения и начнётся приведение примеров, между коими попадётся на грех безличное предложение, тогда наши составители учебников грамматики поднимут такую разногласицу, что невольно при этом вспомнишь басню Крылова про лебедя с братьей. (Dmitrevskij, 1877b: 17)⁴

In the above, he predicts that as soon as the binary model of the sentence is compared to actual language data, which will inevitably include impersonal sentences, those defending the model will engage in never-ending unproductive disputes resembling those of a famous Russian fable. The referenced classic is Ivan Andreevič Krylov's *The Swan, The Pike and The Crab* (1816), which is prefaced with a warning and goes roughly as follows:

THE SWAN, THE PIKE AND THE CRAB

When comrades lack agreement,
their project will not be productive
and will result in nothing, except trouble.

Once the Swan, the Crab and the Pike
set out to pull a loaded cart,
together all three harnessed themselves in;
they pull with all their might,
but the cart won't budge!

The load, for them, seems like it should be light:
but the Swan is raring for the skies,

⁴ Translation: "As soon as the discussion turns toward examining the constituents of the sentence and examples start to be put forth, amongst which an impersonal sentence will surface maliciously, then our writers of grammar textbooks will start so many disputes that it will be impossible not to think of Krylov's fable about the swan and its brothers."

the Crab is scrambling backwards, and the Pike
is headed for the water.
Who is the guilty party, who is right, – that judgement
is not for us;
It's just that the cart is to this day in the same place.

The cart represents the definition of the sentence and the different directions that it is pulled are the different ingenious attempts to maintain in place the classical logic model despite its evident incompatibility with actual language data, i.e. in this case, the attempts to either find a nominative subject in data that does not appear to have one or, alternatively, to discredit the data and make it inconsequential. In essence, in the above lines Dmitrevskij makes a **prophecy** that nothing productive will come from defending a theoretical model that pushes incongruous structures that do not fit the model, in this case impersonal sentences, to the periphery. It has been over two-hundred years, and Dmitrevskij's criticisms concerning the status of the subject have been taken seriously – the hierarchical verbal-predicate model that he was arguing for has become a widely regarded alternative to the binary model, following the developments in modern logic and the work of Tesnière. However, although the traditional subject's incongruence with language data has since been recognized, most sentential models today are still pushing to the periphery the incongruous use of language of yet another type: the verbless sentence.

What's more, certain arguments that were raised to *minimize the impact* of impersonal sentences on the accepted model are very similar, in their general traits, to those that are currently often raised against verbless sentences. The first of these is the **hidden presence** argument. One of the ways that Dmitrevskij's contemporaries tried to defend the status of the subject and with it, the binary classical logic model, against contradictory data is by appealing to the idea of a hidden element. Dmitrevskij (1877a) severely criticizes the conception of the hidden subject, including that maintained by Buslaev, which integrates the subject into the impersonal verb, as well as the alternative versions of the conception used in grammar textbooks. He insists that the hidden subject conceived as a 'formless form', i.e. a form that is not grammatically marked in any way and is indeterminate and unpronounceable by any language speaker, is contradictory to what can constitute an essential grammatical element (Dmitrevskij, 1877a: 3–7). By definition it is neither grammatical, since it has no form – 'What grammatical form points to all of these indeterminate and hidden subjects?' (Dmitrevskij, 1877a: 6), nor essential, since positing it as necessary for the sentence amounts

to saying ‘*Here is a rule without exceptions, which (i.e. this rule without exceptions) has exceptions*’ (Dmitrevskij, 1877a: 4).

Another strategy, also familiar to verbless sentences, was to **discredit** the inconvenient data in some way in order to neutralize its consequences for the adopted model. This included completely **ignoring** the structures in grammatical descriptions, which avoided contradictions and maintained the coherence of the model. Others relegated the data to the status of anomalies that require **correction**. For instance, Grigorij Alekseevič Milovidov referred to impersonal structures as ‘archeological ruminants of ancient times’ (Milovidov, 1878: 15), a historically restricted phenomenon that he urged will soon be replaced by modern binary usage, and, notably, insisted that impersonal sentences are just as inconsequential for the binary model as verbless sentences:

[То, что и в современном языке встречаются безличные предложения] не уменьшает и не изменяет значения подлежащего в предложении, как не уменьшается значение сказуемого от того, что есть предложения и без сказуемых. Например, дитя видит жука и кричит: «Жук!» (Milovidov, 1878: 16)⁵

The argument is that a sentence with only one element (impersonal or verbless) has the same meaning as the binary one, and will with time come to be used in the correct binary form. In contrast, Dmitrevskij insists that language data is not a ‘heresy to be corrected in order to fit dogmatic definitions’ (1877a: 8). Defending real language use, he points to the short-sightedness of syntactic models that refuse to explain the existence of impersonal structures (1877a: 8) and challenges his contemporaries’ prescriptive approach:

[Б]езличное предложение [...] не погребено [...] под развалинами бесчисленных переворотов языка, а, являясь живым, неумиряющим свидетелем всей истории языка, и поныне живет себе [...] и оно умрет разве только с языком. (Dmitrevskij, 1878a: 17)⁶

It thus appears to be **performance** – language data concerning impersonal sentences – that drove the criticism of the classical logic model and incited the new hierarchical syntactic model

⁵ Translation: “The fact that impersonal sentences are also found in modern language does not diminish or change the status of the subject in the sentence, just as the status of the predicate does not diminish from the fact that there are sentences without a predicate. For example, a child sees a bug and yells: *Bug!*”

⁶ Translation: “The impersonal sentence is not buried under the ruins of innumerable language revolutions, but, constitutes a living, persisting witness of the entire history of language, and thrives to this day and will disappear only simultaneously with language.”

which strove to better represent the competence necessary to account for observed language, i.e. a predicate alone was deemed a sufficient representation of a complete thought. The tension is evident in the above exchanges, but, as history has shown, Dmitrevskij's opponents have not succeeded in writing off the observed structures without a subject as inconsequential performative anomalies. The subject data eventually shook the status of one of the essential syntactic elements, and it is not the only skeleton in the closet. It seems reasonable to wonder how long it will take before the parallel between impersonal and another type of incongruous performance – the verbless sentence – is acknowledged.

The evident fight to insist on the *theoretical implications of the subject's absence* is potentially one of the reasons that, despite all of his resolve for a performance driven model, Dmitrevskij's treatment of *verbless* sentences was entirely eclipsed by the drive to establish the verbal predicate at the top of the hierarchy. In spite of the observable absence of the verb, verbless sentences were analyzed as consisting of 'secondary elements into which the verbal predicate has been integrated'. To argue for the verbal predicate-centered conception, he attributed verbal properties to elements such as nouns, adverbs, particles and interjections:

[С]казуемое, для своего выражения, не только заимствует формы именные и наречные, но даже не брезгует и частицами, когда они способны выполнять роль глагола. Таковы в русском языке: ну, да, чтоб (чтоб его!), кроме звукоподражательных хлоп, стук, и др. Даже ну и на принимают глагольные флексии множ. числа 2 лица: ну-те, на-те (также часто и наречие прочь: прочь-те). (Dmitrevskij, 1878b: 49 in Vinogradov, 1958: 291–292)⁷

His justification is that the elements that do not belong to the grammatical category of the verb, nevertheless appear to be capable of carrying out the functions of the verb since they show agreement and some are even observed to conjugate for person (as for instance the particle 'на' (*na*; lit. 'here') when used in the sentence 'На-те' (*na-te*; lit. 'here-te.2PL') to, for example, offer something to someone). For sentences consisting of two nouns, such as 'Земля – планета' (*zemlja – planeta*; lit. 'Earth.F.NOM – planet.F.NOM') in which the Earth is identified as a planet, the noun is treated as a predicate which receives the verbal properties of conjugation and present tense:

⁷ Translation: "The predicate, for its expression, does not only borrow nominal and adverbial forms, but doesn't shy away even from particles, when they are capable of performing the role of the verb. Such are in the Russian language: *nu, da, čtob (čtob ego!)*, apart from onomatopoeia *xlop, stuk*, amongst others. Even *nu* and *na* accept verbal conjugation of the second person plural: *nu-te, na-te* (as does, just as frequently, the adverb *proč: proč-te*)."

Само собой разумеется, что, придавая основное, господствующее значение сказуемому и сказуемости, А. А. Дмитриевский склонен расширять объем понятия «глагольности» или «спрягаемости», которое рассматривается им как синоним «сказуемости». Так, по его мнению, под влиянием «метафоризма языка» в тех случаях, когда сказуемым служит имя без глагола, это имя получает «вербальную форму или спрягаемость и само в себе уже включает признак настоящего времени» (ср. *земля – планета*). (Vinogradov, 1958: 291)⁸

Similarly, a single word verbless sentence, such as the nominative-case noun ‘Зима’ (*zima*; lit. ‘*Winter*’), is analysed as itself constituting a verbal predicate integrated inside what initially appears to be a nominative subject (Vinogradov, 1958: 291, 293). In essence, what has happened is that the hidden presence of the subject, that is so convincingly refuted by means of appeal to language data, is, under the pressure of predicate-centrality, replaced by a determined hunt for the hidden presence of the verb through giving verbal qualities to non-verbal elements.

The conflicts, raised by Dmitrevskij drawing attention to data that was incongruent with the classical logic subject requirements, may in part explain how it is that Russian impersonal sentences overshadowed verbless sentence data, and led to the model in which the verbal predicate is dominant originating particularly in Russia. Another potential explanation is the rich aspectual system of the Russian verb and, as shown in (Sériot, 2000; Gasparov, 1995), the concentration of studies at that time on comparing and emphasizing the properties of Russian verbs to the verbs of other languages. The focus on ‘subject-less’ sentences could also be because positing the existence of a hidden nominative noun inside the impersonal structure, where a noun in a different case is already present, intuitively seems more difficult than it is to posit a hidden verb inside a structure without a verb. It may also be that the data concerning sentences without a verb was not as amply described at the time in linguistics literature as it was for structures without a nominative noun. Perhaps socio-linguistic and other reasons also contributed. The reasons deserve attention, as the historical chain points to a curious fact: the language that is known for having the most productive use of verbless sentences in the Indo-

⁸ Translation: “It is self-explanatory, that, upon giving the main, dominant role to the predicate and predicative function, A. A. Dmitrievskij is inclined to expand the scope of meaning of ‘verbality’ and ‘conjugation’, which he considers synonymous to the ‘predicative function’. In this way, in his opinion, under the influence of the ‘metaphor of language’, in those cases where a noun without a verb serves as the predicate, this noun receives ‘a verbal form or conjugation and in itself already includes signs of the present tense’ (for instance, ‘*Earth – planet*’).”

European family appears to have been the *first to arrive at the verb-centric syntactic model*, and to have done so nearly one hundred years ahead of Tesnière.

2.3.2 Non–Verb-centric and Other Sentential Models in Russia

2.3.2.1 a. Lifting Morphological Restrictions

The dominance of the verb was not the only solution that was put forth in the Russian linguistics literature to deal with the inconsistencies of applying the classical logic model to natural language sentences. Vladimir Ignat’evič Klassovskij (1870) argued to maintain the classical subject / predicate division but disassociate these essential syntactic elements from their traditional morphological restrictions. Striving to resolve the logical structure of a judgment, which is universal, with the existing grammatical structures of the sentence, of which there is a profusion, he insisted that the only way the binary model can be maintained is if the role of the subject and the predicate can be carried out by any grammatical category (Vinogradov, 1958: 284). Thus, a subject, if it is a noun, can be of any case, and the predicate is not to be limited to the verb:

В суждении, с логической точки зрения, все может быть и подлежащим и сказуемым, смотря по данному случаю, так сказать, по ударению на той или другой мысли. (Klassovskij, 1870: 13 in Vinogradov, 1958: 284)⁹

For Klassovskij, the sentence is identified as a logical judgement, of which the requirements are a logical predicate and a logical subject, i.e. something is said about something else, without any additional restrictions (Sériot, 2000: 241). One-word sentences, such as ‘Рассветает’ (*rassvetaet*; lit. the impersonal verb ‘*dawning*’), where the subject (i.e. the day that is dawning) is not explicit, are treated as truncated judgements (Sériot, 2000: 242). Anomalies to the nominative noun subject and finite verb predicate sentential model are many and by removing the grammatical category restrictions Klassovskij strives to account for them.

⁹ Translation: “In a judgment, from a logical point of view, anything can be a subject or a predicate, depending on the particular case, so to speak, on the emphasis of the particular thought.”

It is clear that the absence of the verb from a sentence does not a priori pose a problem for such a conception and it does not even necessarily qualify the sentence as truncated or marginal: if there is explicit reference to both a logical subject and a logical predicate, then the judgment, and thus the sentence, is complete. At the same time, one may wonder whether this sentential model is still in the domain of syntax, since it is no longer the words, but the referents of ‘something that is said’ and ‘the thing about which it is said’, that constitute the sentence.

Furthermore, and arguably more importantly, it is not clear that even this conception, free from grammatical category restrictions, can deal with a verbless sentence such as ‘*Coffee!*’ when it is said about the referent coffee, as for instance in the context of a coffee pot forgotten on the stove or in the context of recognizing the contents inside a cup. In this case, the logical subject and the logical predicate appear to coincide: the utterance ‘*Coffee!*’ appears to simply be naming the referent, as opposed to saying something about it. The logical subject exists (i.e. the referent coffee serves as the support for the predication) but the thing that is said about this referent, i.e. ‘*Coffee!*’, seems to require attributing some additional information that is not explicitly uttered for it to constitute a logical predicate, for instance that we forgot about it or even simply that we are identifying it as being coffee. The positing of this additional inexplicit information seems to slip the structure into a truncated judgment, and thus, again, a marginal sentence form. If the additional inexplicit information is not posited, the structure is left without one of the essential elements.

2.3.2.2 b. New Part of Speech

In 1928, a new part of speech is introduced into Russian linguistics analysis. It is called the ‘category of state’ (‘категория состояния’, *kategorija sostojanija*), and later comes to be known as ‘predicative words’ or simply ‘predicatives’ (‘предикативы’, *predikativy*). The original proposal for the new class of words is made by Lev Vladimirovič Ščerba (1928) and concerns words that, in the context of a particular sentence, are difficult to assign to a grammatical category:

Есть ряд слов, как *нельзя, можно, надо, пора, жаль* и т. п., подведение которых под какую-либо категорию затруднительно. Чаще всего их, по формальному признаку неизменяемости, зачисляют в наречия, что в конце концов не вызывает практических неудобств в словарном отношении, если оговорить, что они употребляются со связкой и функционируют как

сказуемое безличных предложений. Однако при ближайшем рассмотрении оказывается, что указанные слова не подпадают под категорию наречий, так как не относятся ни к глаголу, ни к прилагательному, ни к другому наречию. Далее, оказывается, что они составляют одну группу с такими формами, как *холодно, светло, весело*, и т. д. во фразах: *на дворе становилось холодно; в комнате было светло; нам было очень весело* и т. п. Подобные слова тоже не могут считаться наречиями, так как эти последние относятся к глаголам (или прилагательным), здесь же мы имеем дело со связками (см. ниже). Под форму среднего рода единственного числа прилагательных они тоже не подходят, так как прилагательные относятся к существительным, а здесь этих последних нет, ни явных, ни подразумеваемых. Может быть, мы имеем здесь дело с особой *категорией состояния* [...] в отличие от такого же состояния, но представляемого как действие: *нельзя [...] / запрещается; можно / позволяется; [...] становится темно / темнеет; [...]* и т.д. (Ščerba, 1928: §VII)¹⁰

The **motivation** for the proposal is the ambiguity of a series of words within a particular type of copula structure. Appeal is made to the fact that a word may find itself without an unambiguous morphological class, with only the syntactic function as the sole distinction between identical forms. For instance, morphological inflection is not always able to distinguish an adjective from an adverb. Such a distinction must often be made on the basis of whether the word modifies a noun or another part of speech, which is a syntactic consideration. It is thus generally recognized that for words that have the same form and lack inflection, i.e. “indeclinables, syntax is decisive for determining part of speech” (Wertz, 1994: 308). However, the problem is that even syntactic structure is not always able to resolve an uninflected word, particularly when the word that the latter modifies is also ambiguous for part of speech. It is here that the attention falls particularly on the ambiguity created by impersonal

¹⁰ Translation: “There exists a series of words, such as *nel’zja* [lit. ‘impossible’], *možno* [lit. ‘permissible’], *nado* [lit. ‘necessary’], *pora* [lit. ‘timely’], *žal’* [lit. ‘regrettably’], amongst others, which are difficult to place into any grammatical category. Most often these words, due to the fact that their formal restriction is the absence of inflection, are classified as adverbs, which ultimately does not raise any practical inconveniences in terms of word relations, if we set aside the fact that they are used with the copula and that they function as the predicate in impersonal sentences. However, upon closer examination, it turns out that these words do not fall under the category of adverbs, as they do not relate neither to a verb, nor to an adjective, nor to another adverb. Next, it turns out that they constitute a single group with words such as *xolodno* [lit. ‘cold’], *svetlo* [lit. ‘luminous’], *veselo* [lit. ‘fun’], etc., in phrases: *na dvore stanovilos’ xolodno* [lit. ‘in yard.PRE was_becoming cold’]; *v komnate bylo svetlo* [lit. ‘in room.PRE was luminous’]; *nam bylo očen’ veselo* [lit. ‘for_us.DAT was very fun’], amongst others. Such words also cannot be considered adverbs, since the latter relate to verbs (or adjectives), but here we are dealing with copulas (see below). They also do not fit as neuter-gender singular adjectives, since adjectives relate to nouns, but here the latter do not exist, neither explicitly, nor implicitly. It is possible that we are dealing here with a special *category of state* [...] in contrast to a similar state, but presented as an action: *nel’zja* [lit. ‘impossible’] [...] / *zapreščajetsja* [lit. ‘prohibited’]; *možno* [lit. ‘possible’] [...] / *pozvoljaetsja* [lit. ‘allowed’]; [...] *stanovitsja temno* [lit. ‘becoming dark’] / *temneet* [lit. ‘darkening’]; [...] etc.”

structures (i.e. without a nominative subject) that use a copula. For instance, ‘Мне было холодно’ (*mne bylo xolodno*; lit. ‘*me.DAT was cold*’), where ‘холодно’ (*xolodno*; lit. ‘*cold*’) is ambiguous. Ščerba rejects its analysis both as an adverb and as an adjective. The adverb analysis is rejected on the basis that what the target word modifies in this structure is usually not considered a verb, but a copula, the verbal status of which is widely debated.¹¹ Treating the form as an adjective is also rejected since ‘холодно’ (*xolodno*; lit. ‘*cold*’), which is identical in form to a gender-neutral singular short adjective, in this structure does not have a corresponding gender-neutral noun to modify and also does not inflect like a short adjective for gender and number. The modified element is ambiguous and the targeted uninflected word in the impersonal copula structure is left wanting a morphological category. It is not mentioned by Ščerba (1928), but the morphological and syntactic ambiguity is particularly accentuated when using such a structure to refer to the present time in an unmarked way, since the copula is omitted in such cases.

It is thus that the new morphological class was originally motivated by Ščerba and called the ‘category of state’ – the name reflecting a discernable semantic similarity between the morphologically ambiguous forms. He explains the similarity in meaning between the words that fall into the proposed category, e.g. ‘нельзя’ (*nel’zja*; lit. ‘*impossible*’) and ‘становится темно’ (*stanovits’ja temno*; lit. ‘*becoming dark*’), as referring to a state, and makes a contrast between this type of state and a state presented as an action, e.g. ‘запрещается’ (*zapreščaetsja*; lit. ‘*forbidden*’) and ‘темнеет’ (*temneet*; lit. ‘*darkening*’), respectively.

Formally, Ščerba’s proposed morphological class consisted of ambiguous words or phrases that (a) are combined with the copula, and, simultaneously, (b) are either: uninflected forms (e.g. ‘нельзя’; *nel’zja*; lit. ‘*impossible*’), or nouns preceded by a preposition (e.g. ‘в сюртуке’; *v sjurtuke*; lit. ‘*in frock.PRE*’), or forms with the masculine suffix ‘–ø’, feminine ‘–а’, neuter ‘–о’ or ‘–ə’ (e.g. ‘красен’; *krasen*; lit. ‘*red*’), or nouns in the instrumental-case (e.g. ‘солдатом’; *soldatom*; lit. ‘*by_soldier.INS*’) (Ščerba, 1928: §VII).

The grouping of such words into a new part of speech raised heated **debates** that continue to this day. In his original proposal, Ščerba was cautious, expressing worries that the ambiguous words found in this syntactic structure may be too diverse to constitute a morphological class:

¹¹ Andrea Moro (1997) historically traces the idea that a copula itself cannot be considered neither a verb, nor a predicate, to Aristotle (Moro, 1997: 250, 253) and provides an overview of the debates around its status, which remains an open question: “copular sentences have always constituted and still constitute a challenging field for all grammatical models” (Moro, 2006: 2).

Однако мне самому не кажется, чтобы это была яркая и убедительная категория в русском языке. [...] Все эти параллели едва ли укрепили мою новую категорию, так как слишком разнообразны средства ее выражения, однако несомненным для меня являются попытки русского языка иметь особую категорию состояния, которая и вырабатывается на разных путях, но не получила еще, а может и никогда не получит, общей марки. (Ščerba, 1928: §VII)¹²

Summaries of the divergent reactions from that time may be found in (Tixonov, 1960) and (Galkina-Fedoruk, 1958). **Critics** of the proposed word class (e.g. Šapiro, 1955; Travniček, 1956; Apresjan, 1985) emphasized the absence of any special morphological features, the lack of historical studies to justify its existence, as well as the inconsistencies of the proposed semantic classifications of the words (Benson, 1954: 285, 289; Wertz, 1994: 313). For instance, for Jurij Derenikovič Apresjan, the issue is a matter of syntax rather than morphological class: syntactic features, as opposed to part of speech, characterize the ability or inability of a lexeme to occupy certain syntactic positions in particular structures (Apresjan, 1985: 291 in Wertz, 1994: 314). **Support** for the proposed part of speech came from Nikolaj Semenovič Pospelov (1955), Aleksandr Vasil’evič Isačenko (1955), as well as Viktor Vladimirovič Vinogradov (1947/1986). The latter, working in the perspective of a different school of linguistics from that of Ščerba, initially defends the new part of speech in *Russkij Jazyk* (Vinogradov, 1947/1986: 42–44, 399–421), but then ignores the category in his syntactic discussion of the Russian language in *Grammatika Russkogo Jazyka* (Vinogradov & Istrina, 1954) only a few years later (Benson, 1954: 284). Christopher Wertz argues for the recognition of the part of speech due to the pedagogical advantages of a separate label for the words (Wertz, 1994: 315). As he points out, the proposed ‘category of state’ does not actually fit under the traditional definition of ‘part of speech’ (Wertz, 1994: 308) and the question of whether the category of state “is a separate part of speech, and if so, which words constitute it, is not yet settled” (Wertz, 1994: 306). Some of the still unanswered questions that block consensus concerning the existence of the part of speech are summarized in (Zimmerling, 2018: 46).

Furthermore, even those who concur with the existence of the category often disagree in terms of what should and should not be included in it, and the category continues to take on

¹² Translation: “However, it doesn’t seem to me that this category would be a distinctive and convincing category in the Russian language. [...] All of these parallel examples [i.e. such as the contrast between *nel’zja* (lit. ‘impossible’) / *zapreščajetsja* (lit. ‘forbidden’)] scarcely strengthen my new category, since the means of its expression are far too diverse, however the attempts of the Russian language to have a particular category of state appear undeniable to me, a category which is being developed in different ways, but has not yet received, and perhaps will never receive, a unified representation.”

various formalizations. For instance, forms resembling short adjectives are included in the proposed part of speech by Ščerba (1928) and Vinogradov (1947/1986) when they are a part of an impersonal structure (e.g. ‘ему было холодно’; *emu bylo xolodno*; lit. ‘*him.DAT was cold.ADJS.NSG*’); but, short adjectives in structures with a *nominative subject* are included only when the short adjective at issue does not have a corresponding full form or the full form has a different meaning from the short form (Wertz, 1994: 306). Wertz (1994: 308–310) argues that all short adjectives found in any copula structure should be included, regardless of whether the structure is *impersonal or personal*; thus, extending the new part of speech. In contrast, Benson (1954), Pospelov (1955), Belošapkova (1989), as well as more recently Antonova (2004), argue that short adjectives, in all structures, should be *excluded* from the category. Morton Benson (1954: 286) rejects all short adjectives from the class on the grounds that they may be used with a nominative subject (and thus have a noun to modify) and inflect for gender and number (and thus are not indeclinable), e.g. ‘Он был холоден’ (*on byl xoloden*; lit. ‘*he.NOM was cold.ADJS.MSG*’). Whether in the wider or the narrower version, those who recognize the proposed part of speech at minimum agree that the grouping of words concerns strictly a subgroup of sentences that use the copula.

It is important to note that the existence of the predicate in the ambiguous structures at issue is presupposed. Preliminary predicative status is key to defining the new part of speech. The words are grouped into a single part of speech according to their syntactic position. That this position carries a predicative function is not determined by any prior morphological restrictions. For instance, the following definition from Benson shows that uninflected words in one-word utterances are assumed to function as the predicate, which is then used to justify their predicative status in the ambiguous copula utterances:

A predicative in Russian is an uninflected word that regularly constitutes a **complete utterance when standing alone**, i.e. when preceded and followed by silence. [...] Whether an independent word-class (part of speech) or a subclass is recognized, is not the essential question here. What should be established is the special syntactic role of predicatives. (Benson, 1954: 285, emphasis added)

As shown by Wertz, the new category is defined by a conception of ‘predicative use’ that is not restricted to the verb:

Category of State: [...] words which, although they **can be used only as predicatives**, for morphological and surface-syntactic reasons are clearly not verbs. (Wertz, 1994: 303, emphasis added)

Vinogradov also groups the words that belong to the new part of speech based on their function as predicate, and in addition, citing Aleksandr Afanas'evič Potebnja (1888), attributes *tense* to the new part of speech based on its combination with the copula:

Под категорию состояния подводятся несклоняемо-именные и наречные слова, которые имеют формы времени (для прошедшего и будущего времени аналитические, образованные посредством присоединения соответствующих форм связки 'быть') и **употребляются только в функции сказуемого.**⁵ [5: Ср. замечание А. А. Потебни: «'Он прав', несмотря на опущении глагола, имеет настоящее время, так как место его в языке в этом отношении определено оборотами 'он был, будет прав'» (Потебня, 1888: 394–395).] (Vinogradov, 1947/1986): §2, emphasis added)¹³

Irina Antonova (2004) emphasizes that the predicative function must be determined before identifying the part of speech:

Вот к какому выводу, который разделяет и автор статьи, приходит В.А. Белошапкова: «Категорию состояния как часть речи составляют неизменяемые полнозначные слова, единственная синтаксическая функция которых – функция сказуемого» (Белошапкова [1989: 521]). Автор отмечает, что в каждом конкретном случае необходимо определить синтаксическую функцию неизменяемого полнозначного слова: *Мне весело* (КС) // *Он весело* (наречие) *пел*. (Antonova, 2004: 129–130)¹⁴

The absence of any prior morphological restrictions on the identification of the predicate is evident; it is precisely this morphological freedom of the predicate that allows to group the words into a separate morphological class. The proposal for the new class thus concerns structures in which the existence of the predicate and sentential status is presupposed.

This presupposition is not unreasonable, but important. Since Klassovskij (1870) opened the door for the lifting of morphological restrictions on the logical subject and predicate (i.e. conceiving of the sentence as something, of any part of speech, that is said about something

¹³ Translation: “The category of state is constituted by uninflected nominal and adverbial words, which have forms of tense (for the past and future tense these are analytical, created through the attachment of appropriate forms of the copula *byt'*) and used only in the function of the predicate. (5) [Footnote 5: Cf. the comment of A. A. Potebnja: “ ‘*On prav'* [lit. ‘*he.NOM correct'*], despite the absence of the verb, has the present tense, since its position in language in this respect is determined by the expressions ‘*on byl, budet prav'* [lit. ‘*he.NOM was, will_be correct'*]” (Potebnja, 1888: 394-395).]”

¹⁴ Translation: “V. A. Belošapkova arrives at the following conclusion, which is shared by the author: “The category of state as a part of speech consists of indeclinable full words, whose only syntactic function is – the function of the predicate” (Belošapkova [1989: 521]). The author notes that in each particular case it is necessary to determine the syntactic function of the indeclinable full word: *mne veselo* (KS) [lit. ‘*me.DAT fun.CAT-OF-STATE'*] // *on veselo* (adverb) *pel* [lit. ‘*he.NOM fun.ADV sung.V'*].”

else), it does not seem strange that the sentential status of the structures at hand, which lack a nominative subject and sometimes simultaneously a verbal predicate, was assumed. However, this assumption must be underlined so as not to fall into a circular trap of using the new morphological category of ‘predicative words’ as an argument to justify the existence of a (non-logical) predicate in these structures. There is a risk of running into the following fallacy: if there exists a predicate then there must be a ‘predicate word’; there is a ‘predicate word’ so there must exist a predicate. The existence of the predicate and sentential status is assumed in the definition of the class of ‘predicative words’ and, as tempting as it may be, it would be circular reasoning to use the presence of ‘predicative words’ to justify the sentential status of a structure. This also means that Russian structures with ‘predicative words’ are not immune to the problematic issues, outlined in the previous section, that are faced by Klassovskij’s proposal to remove a priori part of speech restrictions on the subject and the predicate.

2.3.2.3 c. One-Part / Two-Part Sentence Typologies

Russian has several sentence typologies which include sentences without verbs. One of these centers on Aleksej Aleksandrovič Šaxmatov’s (1925) famous **one-part / two-part sentence distinction**. Two-part sentences explicitly involve a subject and a predicate, in contrast to one-part sentences which make explicit only the subject or only the predicate. The distinction is explained as follows:

[П]редложения русского языка распадаются по форме на следующие две основные разновидности: *предложения односоставные, не представляющие словесного обнаружения тех двух членов, на которые распадается каждая психологическая коммуникация*, и на предложения *двусоставные*, один состав из которых является господствующим и соответствует **психологическому субъекту**, а с другой состав – зависимым и соответствует **психологическому предикату**. (Šaxmatov, 1941/2015: 53, emphasis added)¹⁵

¹⁵ Translation: “Sentences in the Russian language consist of two main types according to their form: *one-part sentences*, which **do not include the lexical occurrence of the two elements that make up each communicative act**, and *two-part sentences*, one part of which dominates and corresponds to the **psychological subject**, while the other part is dependent and corresponds to the **psychological predicate**.”

The above makes clear that the elements of subject and predicate, at the basis of this sentential model, involve a communicative psychological dimension. Emphasis is made on the psychological communicative act, as opposed to the logical proposition, since the latter may only account for one particular type of communication, i.e. a judgment in which an assertion or a negation is made (Šaxmatov, 1941/2015: 13). Each *communicative act* is said to consist of two elements – a psychological subject and a psychological predicate.

The *sentence*, defined as the grammatical expression of this communicative act, need not explicitly display both psychological elements in words: the grammatical expression need not be an exact reflection of the communicative act. In Šaxmatov's words:

[П]редложение – это словесное, облеченное в грамматическое целое (посредством согласования составных его частей или соответствующей интонации) выражение психологической коммуникации. [...] Не следует, однако, чтобы предложение было сколько-нибудь точным отображением коммуникации. (Šaxmatov, 1941/2015: 27–28)¹⁶

A two-part sentence consists of the expression of a psychological subject, which is said to be dominant in the sentence, as well as the expression of a psychological predicate, which is said to depend on the subject. The *one-part sentence* also contains the two psychological elements that constitute the communicative act, but it lacks their direct expression as two separate elements; it is said to express both the psychological subject and the psychological predicate in a single sentential element (typically, a single word):

[П]редложения, в которых указанное сочетание субъекта и предиката находит себе соответствие в одном члене предложения (выраженном большей частью одним словом) – это предложения односоставные; [...] как 'вчера морозило', [...] где сочетание субъекта и предиката подлежащей коммуникации находит себе соответствие в слов[е] 'морозило' [...] ('морозило' соответствует сочетанию конкретного признака с отвлеченным признаком в прошедш. времени [...]). (Šaxmatov, 1941/2015: 28)¹⁷

¹⁶ Translation: "The sentence is the lexical expression, which has been framed as a grammatical whole (through the agreement of its parts or appropriate intonation), of psychological communication. [...] It doesn't follow, however, that a sentence is to any extent an exact reflection of communication."

¹⁷ Translation: "Sentences in which the indicated combination of subject and predicate finds correspondence in a single part of the sentence (expressed typically as a single word) – these are one-part sentences; [...] for instance, *včera morozilo* [lit. 'yesterday.ADV freeze.V.PS'], [...] where the combination of the subject and the predicate of the underlying communicative act finds itself a correspondence in the wor[d] *morozilo* [lit.

As a result of this communicative dimension, the one-part sentence is treated as a sentence, just like the two-part variant. In other words, a psychological subject and predicate are necessary, but their grammatical expression is flexible.

The **psychological subject and predicate** are differentiated from the *logical subject and predicate* on the basis that the latter are said to apply only in a judgment, i.e. the logical subject is that about which an assertion or negation is made, and the logical predicate is that which is asserted or negated; whereas the psychological elements, in addition to judgments, also apply in communicative acts that do not assert or negate (Šaxmatov, 1941/2015: 16). Determined sometimes by the speaker and sometimes by inherent nature, the psychological elements are defined as mental conceptions, with the *psychological subject*, i.e. the conception of some entity, dominating over the psychological predicate, i.e. the conception of some attribute about that entity (Šaxmatov, 1941/2015: 21–22).

These are also distinguished from the **grammatical subject and predicate**, which are tied to *lexical expression* and, more specifically, receive the following definition:

[Г]рамматическим подлежащим будет название конкретного, индивидуального предмета, а сказуемым название общего, родового понятия, название предмета, с которым сочетаются представления о признаках, о совокупности признаков; об этом было сказано выше по поводу такого предложения, как *‘Иванов портной’* [подлежащее *‘Иванов’* и сказуемое *‘портной’*; утверждение что субъектом является *‘портной’* имеет смысл, если слово *‘портной’* означает здесь того определенного портного, о котором была только что речь], а так же таких предложений, как *‘неклен – дерево’*, *‘шведы – германцы’*. Только там, где утверждается тождество одного предмета представления с другим, можно безразлично употреблять то или другое название в качестве субъекта или в качестве предиката [...]. (Šaxmatov, 1941/2015: 22–23)¹⁸

‘freeze.V.PS’ [...] (*morozilo* [lit. *‘freeze.V.PS’*] corresponds to the combination of a concrete feature with an abstract feature in the past tense [...]).”

¹⁸ Translation: “Grammatical subject will be the name for a specific, individual entity, and predicate will be the name for a global generic notion, the name of an entity which corresponds to the conception of features and of the combination of features; this point was discussed above in relation to a sentence such as *‘Ivanov portnoj’* [lit. *‘Ivanov.NN.MSG.NOM tailor.NN.MSG.NOM’*] [i.e. grammatical-subject *‘Ivanov’* and grammatical-predicate *‘tailor’*; the argument that the psychological-subject is *‘tailor’* makes sense only if the word *‘tailor’* here means that specific tailor that has just been mentioned], as well as sentences such as *‘neklen – derevo’* [lit. *‘maple.NN.NSG.NOM – tree.NN.NSG.NOM’*] and *‘švedy – germancy’* [lit. *‘swiss.NN.PL.NOM – germanics.NN.PL.NOM’*]. Only in those case where identity is asserted between the conception of the one entity with the other, is it possible to indifferently use the one or the other name as the psychological-subject or as the psychological-predicate.”

The grammatical subject is the lexical evocation of the mental conception of some specific entity, whereas the grammatical predicate is the lexical evocation of the mental conception of an attribute. For instance, if ‘*maple – tree*’ is analyzed as the entity ‘*maple*’, conceived of as a specific object, and the entity ‘*tree*’, conceived of as an attribute of this object, then the former is the grammatical subject and the latter is the grammatical predicate. However, if both ‘*maple*’ and ‘*tree*’ are conceived of as equal entities that are being identified with one another, then either of them may serve as the psychological subject or the psychological predicate, and thus either may correspond to the grammatical subject or the grammatical predicate. It is emphasized that any potential ‘*grammatical*’ restrictions on identifying a word as a grammatical-subject is due to a ‘*psychological inability*’ to conceive of it as psychological-subject:

[Г]рамматическая невозможность сделать ‘портной’ подлежащим при слове ‘Иванов’ [...] коренится, конечно, в невозможности психологической. (Šaxmatov, 1941/2015: 21–22)¹⁹

Therefore, it is ultimately the speaker’s conceptions that determine which lexical element is identified as the grammatical subject and which as the grammatical predicate; formal features, such as agreement, intonation and word order, are a posteriori reflections of the psychological conceptions. In other words, these grammatical definitions do not depend on part-of-speech or grammatical restrictions, but are rather determined by speakers.

It is also important to mention that the existence of one-part sentences, does not prevent Šaxmatov from identifying some sentences as ‘**incomplete**’. To this category he assigns what appear to be semantic ellipses and fragments.²⁰ In particular, he includes sentences in which case marks the presence of the grammatical subject (e.g. ‘входит’; *vxodit*; lit. ‘*enters.V.PR.3SG*’, which is analyzed as omitting a third person noun due to the conjugation of the verb) or the presence of another element dominating over the lexically explicit item (e.g. ‘бочку’; *bočku*; lit. ‘*barrel.NN.ACC*’, is analyzed as omitting a finite verb due to the case of the noun), as well as those sentences that mention the dominating element in the linguistic context (Šaxmatov, 1941/2015: 52). For Šaxmatov, the distinguishing feature between an

¹⁹ Translation: “The grammatical impossibility of making the word ‘*portnoj*’ [lit. ‘*tailor.NN.MSG.NOM*’] the grammatical subject, given the presence of the word ‘*Ivanov*’ [lit. ‘*Ivanov.NN.MSG.NOM*’], is rooted, of course, in a psychological impossibility.”

²⁰ The different conceptions of ‘ellipses’, as well as the debate about whether or not elliptical structures constitute sentences, will be discussed in the sections that follow.

‘incomplete’ and a ‘complete’ sentence is whether or not the sentence is used in the same way as the version in which the element is explicitly evoked:

‘входит’, ‘сахару’ признаем неполными предложениями, но предложения как ‘виноват’, ‘рады стараться’ считаем **полными, ибо ‘я виноват’, ‘мы рады стараться’ в своем употреблении не совпадают с предложениями без ‘я’, ‘мы’.** Предложения с опущенными главными членами называем *неполными*, причем предложения с опущенным подлежащим назовем *недостаточными*, а с опущенными главным членом или сказуемым – *нарушенными*. (Šaxmatov, 1941/2015: 53)²¹

An *incomplete sentence* is said to be one that has the same use as the version which lexically evokes the subject (e.g. ‘enters.V.PR.3SG’ and ‘she enters’ are analyzed to be synonymous in use and thus the former is said to be incomplete). A sentence that is *complete* is one that does not have the same use if an element is evoked. (It is notable that this analysis thus assumes that a word may be added without changing the meaning of the sentence.) Furthermore, intonation is also said to contribute to the complete/incomplete distinction: exclamative sentences have complete status due to emphatic intonation when they are pronounced. For instance, in a sentence such as ‘Ах, как мил!’ (*ax, kak mil*; ‘ah, how cute.ADJS.MSG’), emphatic pronunciation is identified as replacing the need for an overt subject, i.e. the need for an explicit masculine singular noun to correspond to the masculine singular adjective (Šaxmatov, 1941/2015: 73).

Based on these definitions, Šaxmatov proposes an extremely rich typology of sentences. The following, *Figure 1*, illustrates selected **principal divisions** with examples from Šaxmatov (1941/2015: 54–316).

²¹ Translation: “We recognize ‘vходит’ (lit. ‘enters.V.PR.3SG’) and ‘сахару’ (lit. ‘sugar.NN.GEN’) as incomplete sentences, but sentences such as ‘виноват’ (lit. ‘guilty.ADJS.M1SG’) and ‘рады стараться’ (lit. ‘happy.ADJS.PL to_try.V.INF’) we treat as full because ‘я виноват’ (lit. ‘I.PRO.1SG guilty.ADJS.M1SG’) and ‘мы рады стараться’ (lit. ‘we.PRO.3PL happy.ADJS.PL to_try.V.INF’) in their use do not correspond to sentences without ‘я’ (lit. ‘I.PRO.1SG’) and ‘мы’ (lit. ‘we.PRO.3PL’). Sentences with omitted main elements we call *incomplete*, moreover sentences with an omitted grammatical-subject we call *insufficient*, and sentences with an omitted dominating-element [i.e. the element, in one part-sentences, in which the psychological subject and the psychological predicate combine] or grammatical-predicate we call *broken*.”

ŠAXMATOV'S TYPOLOGY OF SENTENCES (REDUCED TO SELECTED MAJOR TYPES)

ONE-PART SENTENCES: combine a psychological subject with a psychological predicate that involves the conception of being, having, or appearing, in the same expressed part

PREDICATE-LESS SENTENCES: psychological subject cannot be made explicit as a separate lexical entity

- 'Зима.' (*zima*; 'winter.NOM')
- 'Было морозно.' (*bylo morozno*; 'was frosty.ADV')
- 'Яблок-то, яблок!' (*jablok-to, jablok*; 'apples.GEN-that.PART, apples.GEN')

SUBJECT-LESS SENTENCES: psychological subject could be made explicit as a separate word without changing the meaning

- 'Ах, как мил!' (*ax kak mil*; 'ah, how cute.ADJS.MSG')
- 'Вот злодейка!' (*vot zlodejka*; 'here villain.NOM.FSG')
- 'Перед всеми каюсь!' (*pered vsemi kajus*; 'in_front_of everyone confess.V.1SG')
- 'На улицах стреляют.' (*na ulicax streljajut*; 'on streets shoot.V.3PL')
- 'Там поставить кровать.' (*tam postavit' krovat*; 'there put.V.INF bed')
- 'Брысь!' (*brys*; 'shoo.INTJ.2SG'), 'Брысьте!' (*brys'te*; 'shoo.INTJ.2PL')

TWO-PART SENTENCES: combine a psychological subject and psychological predicate expressed separately

WITHOUT GRAMMATICAL AGREEMENT

- Уехать не удалось. (*uexat' ne udalos*; 'leave.V.INF not succeeded.V.PS')
- Кататься весело. (*katat'sja veselo*; 'skate.V.INF fun.ADV')
- В кошельке осталось пустяки. (*v košel'ke ostalos' pustjaki*; 'in wallet remain.V.NSG crumbs.NN.MPL')
- Ум – хорошо, а два еще лучше. (*um – xorošo, a dva ešče lučše*; 'brain.SG – good.ADV, but two even better')
- Что за бумаги? (*čto za bumagi*; 'what.Q this.PART papers.NN.PL?')

WITH GRAMMATICAL AGREEMENT

- Дети вернулись из школы. (*deti vernulis' iz školy*; 'children.3PL.NOM returned.V.3PL.PS from school')
- Промилуйте вы меня. (*promilujte vy menja*; 'forgive.V.2PL.IMP you.2PL.NOM me')
- Вы хороший. (*vy xorošij*; 'you.2PL good.ADJ.MSG')

Figure 1. Šaxmatov's Typology of One-Part and Two-Part Sentences

There are many more sub-types in this classification which is based on a combination of the overall meaning of the sentence, the grammatical category of the main element, as well as the anticipated emphasis during pronunciation. It is important to note that communicative function is fundamentally integrated into the divisions: a psychological subject and psychological predicate are present regardless of sentence type, what varies is whether or not, and the way that, they are evoked.

Since Šaxmatov's influential work, many different versions of the one-part / two-part sentence typology have been proposed, a comparison of which may be found in (Potapova, 2017). They include: Peškovskij (1928), Vinogradov and Istrina (1954), Galkina-Fedoruk (1958), Gvozdev (1968), Babajceva and Maksimov (1981), Lekant (1986), Belošapkova et al. (1989), Valgina (2000), Skoblikova (2006), Dolin (2008). Simultaneously, the partition has also been criticized. The famous grammar by Natal'ja Švedova et al. (1980) rejects the one-part / two-part distinction and replaces it with categorization by the morphological form of the main element, making the principal division in terms of finite-verb sentences and non-finite verb sentences. However the partition is drawn, Šaxmatov makes a coherent proposal to explain the variety of sentence types that are very difficult to account for with a sentential model that is based on a verbal predicate. Many of the key aspects of Šaxmatov's detailed and explanatory analysis that are summarized in the present section went on to become very influential on Russian linguistics.

Nonetheless, one may still wonder about the way that the model compares with the analyses of sentential status discussed thus far. Šaxmatov's proposal, in a way, represents a step away from logic – the criticism of the insufficiency of a logical subject and logical predicate is clear, i.e. something said about something else does not represent all of language as it excludes many communicative acts, such as questions, imperatives, suggestions, where a judgment (assertion or negation) is not made (Šaxmatov, 1941/2015: 14). However, at the same time the proposed model of the sentence is rooted in a physics that is beyond tangible linguistic expression – it is the referents beyond the words that form the backbone of the sentence, and as such the question concerning the role of syntax in this psychological sentential model seems to remain just as relevant as for the logical models. In other words, one-part sentences are said to be sentences because they have the psychological subject and predicate necessary for a communicative act, but they still lack a separate grammatical expression of one of these psychological elements and, consequently, can still be objected to if the sentence is treated as a grammatical entity requiring both a grammatical subject and a grammatical predicate.

The above review proves that in the Russian linguistics tradition significant efforts were made in order to rise to the challenge of language data and account for the sentential status of structures that are incongruous with the classical subject/predicate logical model, including the important proposals to prioritize the verbal predicate, to lift the grammatical restrictions on the subject and the predicate, to introduce a new part of speech, and to draw a one-part/two-part distinction. Yet, a closer analysis of the non-verb centric proposals confirms the firm grip that conceptions related to logic and psychology have on the definition of the grammatical sentence. The precisely grammatical grounds which would allow a verbless structure to be called a sentence in the same respects as a verbal structure remain contentious across the various historical accounts of the sentence reviewed in the previous two sections. The focus of the next section now shifts to accounts that aim particularly at verbless structures.

CHAPTER 3

ELLIPTICAL SENTENCES AND NONSENTENTIALS

3.1 THE NOTION OF ELLIPSIS: ITS IMPORTANCE AND VERSIONS

Central to the discussion of the verbless sentence is the **notion of ellipsis**. The term ‘ellipsis’ takes on various definitions in the literature.

It is sometimes used very loosely as a synonym for any general difference between what is thought and what is explicitly said. For instance, when a speaker utters ‘*It started*’, in order to reference a loaf of bread that has just now started rising in the oven, more is meant than what is literally said. This is an example of the omission of contextually available information which the speaker expects the hearer to fill in: what started (*the particular loaf in the oven*), when it started (*just now*), and started to do what (*the process of rising*) – this information is not made linguistically explicit. The omission is possible thanks to the pragmatic context, and in this sense, it may be called ‘pragmatic ellipsis’.

However, such **pragmatic** omission of extra-linguistic contextually available information is not what is usually meant by the term ‘ellipsis’, or, rather, it should not be in linguistic discussion. Robert Stainton calls this the ‘extremely weak sense’ of the term and strongly warns against such use (Stainton, 2004: 272–273).

In support of the argument that the term ‘ellipsis’ should not be used with regard to pragmatic context, it may be added that pragmatic omission is part of any and all utterances, to various degrees, by virtue of the fact that any string of words necessarily has a context (even if

that context is as informationally-bare as a blanc page or an empty room). Even an unquestionably canonical structure, such as ‘*Socrates runs*’, said upon looking out the window and seeing Socrates run, omits extra-linguistic elements of the context (for instance his speed, the direction, the reason, what he is wearing, where the running takes place, etc.) and as a result can never be pragmatically complete. The immediately evident facilitated interpretability of (a) ‘*Socrates runs.*’ as compared to (b) ‘*It started.*’, if both were found written on a blanc page, is due to the fact that the linguistic content of (a) allows to reduce the *range* of possible interpretations more than (b); but not to the resolution or satisfaction of pragmatic factors since in the context of a blanc page these remain just as opaque for both sentences. The options for the pragmatic context of (a) are easier to imagine than that of (b), but in neither case is the pragmatic content of the utterance fully resolved or complete.

Further support for this perspective comes from the fact that language itself is necessarily a recursive mechanism with continual linguistically-explicit additions and precisions possible to infinity. This consequently means that no sentence can ever be expected to capture the full pragmatic context in and of itself; if it did then language can no longer be recursive. Pragmatic completeness thus runs counter to language itself. As a result, to give pragmatic omission the title of ‘ellipsis’ would subsume under this title all linguistic structures. Using the term may be useful as a reminder of the pragmatic dependency of language, but it does not make for a very good distinction between structures.

Reference to ‘ellipsis’ is usually reserved for various types of omission of specific linguistic material that is presupposed to exist in the *syntactic* structure of the string of words. Alternatively, it is also sometimes used to reference a particular *semantic* encoding, as will be explained below.

The reason that the notion of ellipsis is key to the discussion of verbless sentences is that it is typically used to defend the sentential status of a verbless structure. The **ultimate goal** of the various accounts of verbal ellipsis is to integrate what appears to be incongruent data into sentential models in which a finite verb is a necessary condition for the sentence. The following provides an overview of the way that this is done and analyzes the strengths and limits of such accounts. Are verbless structures (syntactically and/or semantically) elliptical sentences? Are they nonsententials? Both possibilities will be disputed and an alternative, sentential, account will be proposed for the structures in *Part 1: Chapter 4* that aims to overcome the presented criticisms.

3.2 SYNTACTIC SENTENCE STRUCTURE

One of the most discussed conceptions of the syntactic structure of the sentence is that proposed by Noam Chomsky:

The maximal projection of INFL" consists of INFL' and its specifier, the NP subject of INFL"; this maximal projection is what we have called S. (Chomsky, 1986: 161, in Stainton, 2000: 453)

This means that the dominant and necessary element of any sentential structure is inflection, i.e. the morphological marking of the verb for subject-verb agreement and for tense. It is around this inflection, abbreviated as INFL (or I), that the rest of the sentence is built. The inflection necessarily combines with a complement, i.e. the verb on which the inflection is marked (VP). Once this is achieved, this inflected verb phrase additionally necessarily requires a specifier, i.e. the nominal subject (NP). This can be summarized in the following diagram, based on (Stainton, 1995: 283), which illustrates the structure that organizes the string of words 'The letter is from France' into a sentence (Figure 2).

Canonical Sentence Structure: Inflection, Its Complement and Its Specifier

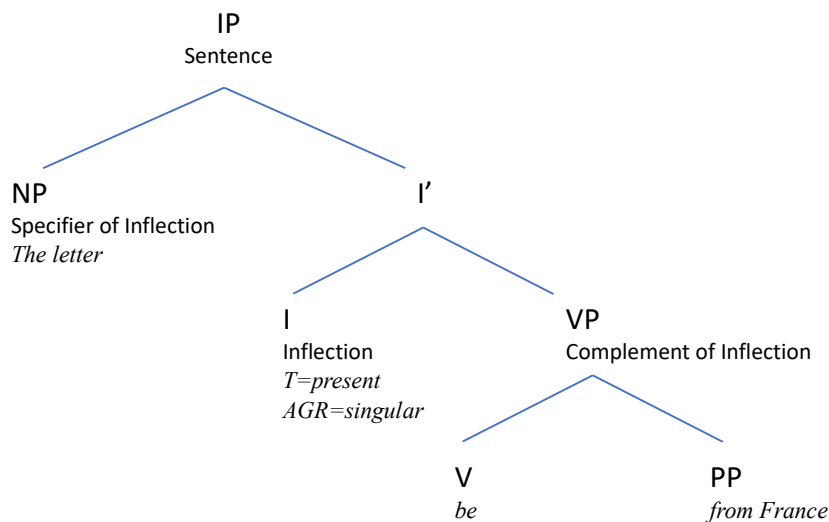


Figure 2. The syntactic structure of a canonical sentence

This syntactic model of the canonical sentence emphasizes the importance of inflection which heads the sentential structure. Details aside, the model is in line with the basic traditional syntactic characterizations of the necessary requirements of a sentence: “A sentence syntactic has a subject, verb and (optional) object” (Stainton, 2000: 445).

3.3 SYNTACTIC ELLIPSIS

The complete structure of a canonical syntactic sentence described above is also attributed to cases where syntactic ellipsis is said to occur. For instance, in the context of receiving a letter, the utterance of the words ‘*from France*’ is sometimes analyzed as omitting much more than extra-linguistic contextually available information: on some accounts, it represents an instance of the omission of linguistic material.

- (6) {Speaker pulls a letter out of the mailbox and says:}
- a. The letter is from France.
 - b. From France.

In (a), the embedded segment ‘*from France*’ constitutes a prepositional phrase within the sentence ‘*The letter is from France*’ that has the canonical structure illustrated in *Figure 2* above. In (b), the words ‘*from France*’ occurring independently, in the same extra-linguistic context, are said to constitute a syntactically elliptical version of sentence (a). On such an account, (b) differs from (a) in the surface expression, i.e. the sound pattern or written text realization, where it appears as an abbreviated version of (a). However, syntactically, (b) ‘*From France*’ is analyzed as having the same complete canonical structure as ‘*The letter is from France*’, i.e. on the syntactic level it is said to include inflection, the verb and the nominal subject, and is therefore considered a syntactic sentence, though called ‘elliptical’ in order to differentiate with cases where the same structure not only exists but is also expressed.

At the basis of such accounts is the following definition of ‘syntactic ellipsis’ and ‘syntactically elliptical sentence’:

[Syntactic ellipsis:] abbreviation occurring between level one (sound pattern) and level two (syntactic structure). (Stainton, 2004: 273)

[Syntactically elliptical sentences:] expressions which were tokened have the structure/form of a sentence_{syntactic}, even though what was pronounced sounds exactly like a sub-sentence_{syntactic}. (Stainton, 2000: 447)

In other words, when found in the particular extra-linguistic context, (b) is not analyzed as syntactically constituting a prepositional phrase, but rather as constituting a syntactic sentence that is headed by inflection. The perceived departure from the model is said to concern an abbreviation of the phonetic or written realization by the speaker, and not the structure of the sentence.

There are two general mechanisms by way of which the syntactic structure of an elliptical sentence, such as (b), gets to stay intact with the canonical syntactic structure of the sentence illustrated in *Figure 2* above. As outlined by Reinaldo Elugardo and Robert Stainton (2005: 2), the difference between the verbless surface realization and the canonical sentence model is explained in theoretical syntax in terms of an underlying hidden structure that is verbal and either:

- (i) transformed into the verbless one by means of surface level *deletion*; or,
- (ii) includes special *null elements* that stand in for the verb.

The first alternative, called the ‘deleted ordinary material’ account, treats the sentence not as a single entity, but rather as a set that simultaneously consists of the hidden abstract version and the explicit version of the sentence. For the previous example (6b), this corresponds to the following pair:

- (i) <[_s the letter is from France], *from France*>

The initial version of the sentence gives the underlying syntactic structure, [_s], that is input into the semantics. The second version of the sentence is the transformed surface realization: this is the version that is pronounced or written, and from which certain parts that exist in the hidden deep structure have been deleted.

The second alternative, called the ‘empty element’ account, inserts ‘extraordinary linguistic material’ which is never expressed (i.e. null elements) into the syntactic structure. As a result, the example ‘*From France*’ would be represented as a single entity containing null elements (represented by Δ), and be attributed the following structure:

- (ii) [s [NP Δ *the letter*] [I' Δ [I Δ *present, singular, 3rd person*] [VP Δ *be*]] from France]

The null elements give this account two advantages in comparison to the former. First, there is no need for surface deletion since the structure can be pronounced as it is, i.e. null elements are not pronounced. Second, because the null elements are considered present, they step in to complete the syntactic structure of the elliptical sentence, and they contribute to the semantic content of the proposition; they do this by invisibly replacing the noun phrase '*the letter*', the inflection specifications for present tense and 3rd person singular agreement, and the verb '*be*', with which they are said to be co-indexed.

Early versions of the 'deleted material' alternative have been traced to Ivan Sag (1976), whereas the early 'empty elements' account to Edwin Williams (1977). It is clear that both of the strategies rely on a hidden syntactic structure to turn the produced non-canonical expression into a sentence. In other words, structures that lack a finite verb are treated as sentential through the attribution of a syntactic structure that goes beyond what is pronounced.

3.4 ARGUMENTS AGAINST SYNTACTIC ELLIPSIS

Several arguments in the domain of syntax have been raised against the ability of the above-described syntactic ellipsis hypotheses to successfully account for the sentential status of a structure.

3.4.1 Argument 1 – *Elliptical structures behave differently*

One of the arguments is that elliptical sentences often do not license VP ellipsis in another sentence; yet, they are expected to if they covertly have the identical structure of a canonical sentence. Stainton (1997: 65) gives the following example:

- (7) a. John: The boat is going very fast.
 Mary: That car is too. [VP ellipsis]
- b. John: Very fast.
 Mary: *That car is too. [VP ellipsis]

While (a) ‘*The boat is going very fast*’ licenses the deletion of the verb phrase ‘*going very fast*’ in the reply ‘*That car is too*’; (b) ‘*Very fast*’ does not, since it cannot be followed by ‘*That car is too*’. This indicates that (a) and (b) do not have the same syntactic structures. The elliptical sentence does not behave in a way that would suggest it has a verb at its core; though it is said to fully meet requirements of the sentential model, its structure does not stimulate the same reactions and does not have the same capacities as a canonical structure. As a result of such behaviour, an elliptical structure does not appear to contain a verb in its syntax, nor to meet the sentential requirements in *Figure 2*. If there is any unpronounced additional syntactic structure, it does not appear to give syntactic signs of its existence. The inevitable question then arises: What, other than the desire to fit language data to a preconceived theory, leads one to maintain the existence of a hidden unexpressed structure?

3.4.2 Argument 2 – *Various possibilities for reconstruction without antecedent*

A key distinction between the above examples (6b) ‘*From France*’ and (7a) ‘*That car is too*’ concerns the presence of a linguistically explicit antecedent in the context. In the terminology of George Hankamer and Ivan Sag (1976) the ellipsis in (7a) constitutes syntactically controlled surface anaphora (which in addition to the illustrated VP deletion, also includes sluicing, stripping, and gapping types of ellipsis), whereas (6b) would be said to have a non-linguistic antecedent and correspond to pragmatically controlled deep anaphora.

In (7a), the occurrence of ‘*going*’ in John’s utterance ‘*The boat is going very fast*’ serves as the antecedent for the verb phrase ellipsis in Mary’s utterance ‘*That car is too*’. The latter may potentially be reconstructed, on a syntactically-elliptical sentence account, as (7c):

- (7) c. That car is [going very fast] too’.

In contrast, (6b) ‘*From France*’ does not contain a linguistically explicit antecedent in the context. The pragmatic context (i.e. in which the speaker pulls a letter out of the mailbox), allows multiple possibilities for potential reconstruction, including:

- (8) a. [The letter is] from France.
- b. [The letter came] from France.
- c. [The letter must have been sent] from France.
- d. [The envelope that Bob is sending me this letter in is the only one in his treasured collection] from France.
- e. From France [was written on the envelope].

It is not clear which verb, and in this case also which subject, is to be reconstructed from the pragmatic context alone. This leaves if not an infinite then at least a wide range of possibilities for the hidden syntactic structure. If reconstruction of linguistic antecedent-based (7a) is imaginable, the vast possibilities for (6b) make it much more difficult to imagine its covert structure.

3.4.3 Argument 3 – *Syntactic structure is what it appears: Antecedent mismatch*

Arguments against the existence of any hidden syntactic material, linguistic-antecedent-based or not, are also put forth by Mary Dalrymple (2005). She argues that even in cases where there appears to exist a syntactic antecedent (i.e. the syntactic source from which to reconstruct), the “syntactic structure of a sentence containing ellipsis is exactly what appears on the surface” (Dalrymple, 2005: 31).

Dalrymple’s arguments draw attention to the frequent mismatch between the constituent that appears as the syntactic antecedent and the syntactic form of the constituent that is required to reconstruct the elliptical sentence. For instance, in the following cases (from Dalrymple, 2005: 35), the element that is supposed to serve as the antecedent cannot simply be copied into the elliptical structure in order to reconstruct it:

- (9) a. Avoid getting shampoo in eyes – if it does, flush thoroughly with water.
- b. This letter deserves a response, but before you do ...

In (a), in order to reconstruct the VP ellipsis ‘*if it does*’, the antecedent ‘*getting in eyes*’ is not suitable (i.e. it predicts the resolution ‘*if it [getting in eyes]*’) and requires syntactic

transformation of its aspect and agreement (i.e. ‘*if it [gets in eyes]*’). Similarly, in (b), the syntactic antecedent found is the noun ‘*response*’, whereas the supposed reconstructed version of the elliptical sentence ‘*before you [respond]*’ requires the syntactic antecedent ‘*respond*’.

These and other syntactic arguments lead Dalrymple to propose that the reconstruction of syntactic ellipsis is not itself syntactic. Ellipsis leads to semantic, rather than syntactic, resolution:

[E]llipsis resolution depends on a semantic equality between the source and target clause: some relation P is obtainable given the meaning of the source clause and is then used in the interpretation of the target. (Dalrymple, 2005: 44)

For instance, in (9a), the elliptical ‘*if it does*’ is interpreted as a statement that some *property* ‘P’ holds of the subject ‘*it*’, symbolized as P (*it*). The resolution requires, not a syntactic antecedent, but to determine semantically – through the interpretation of ‘*avoid getting shampoo in eyes*’ – what this property P is. Formally, it requires solving for P in the parallel:

P (*it*) = getting shampoo in eyes (*you*).

Such an argument means that a syntactic structure that lacks the necessary sentential features of *Figure 2*, may be reconstructed into a semantic proposition that is similar to syntactically canonical sentences, but this reconstruction is not syntactic.

In this way, it is argued that the presence of an antecedent does not necessarily indicate any additional hidden syntax: even cases with an explicit antecedent would receive a semantic explanation on Dalrymple’s account, which, consequently, removes the need to introduce extra syntactic structure.

3.4.4 Argument 4 – *Hidden structure is too costly*

Arguing against any deletion of hidden structure even for cases of antecedent-based ellipsis, Ellen Barton and Ljiljana Progovac (2005) provide a syntactic explanation of verbless structures and the way that they could be built from the perspective of the minimalist generative grammar framework. They point out that the need for deletion rules has become difficult to motivate keeping with considerations of economy.

The strength of their account is that they insist that all structures, including both phrases and sentences, are built bottom-up:

[T]he distinction between sentences and phrases is not a significant theoretical problem in the framework we are adopting in this paper, Chomsky's (1989, 1995) Minimalist Program. Given Minimalism, a clause is simply a phrase whose head is I (short for Inflection). Moreover, the derivation of a structure is built bottom-up, by merging words and phrases in a binary fashion, but only as long as there is evidence for further merger. [...] Superfluous structure is prohibited by the Economy Principle [...] A constituent is considered to be a maximal projection if it projects no further, allowing a head, which projects no further, to be at the same time a maximal projection. Such a head can be a N projecting to NP, a V projecting to VP, a P projecting to PP, an I projecting to IP, and so on. [...] there is nothing special about sentence/clause in this framework. (Barton & Progovac, 2005: 74)

This means that the way that a verbless structure is built would not be any different from that of a structure with a finite verb, in the sense that that both will stop merging as soon as the requirements of the words are satisfied. For a structure to be elliptical, any hidden elements would need to first be generated and then deleted, which is a much costlier alternative to the simple merging of the elements that are present. For instance, the merging of the words '*from*' and '*France*' into the prepositional phrase '*from France*', and stopping there, is more economical than the alternative of additionally continuing to combine '*from France*' with '*the letter is*' and subsequently deleting '*the letter is*' in order to get the same prepositional phrase '*from France*' but now with the hidden inflection that allows it to be a sentence. (This need for economy holds regardless of the presence or the absence of a linguistic antecedent.)

On this account, the difference between a verbless and a verbal structure would not be in their generation, but in that the former will have all of its requirements satisfied and stop merging without reaching inflection. Thus, verbless, and other structures that do not fit the canonical model, are freed from any additional hidden structure. However, simultaneously, they become what Barton and Progovac call 'nonsententials':

' X^{\max} nonsententials' [are] generated not as sentences, but as maximal projections of nonsentential phrases. [They are ...] phrases below sentential level [...] there is no evidence for their sentential status [...] no evidence of any other category on top of these [...] no evidence of any higher projection [...]. (Barton & Progovac, 2005: 74–75)

Verbless structures are thus proposed to be generated without any hidden elements, however, their sentential status is withdrawn due to lack of inflection for agreement and tense.

3.5 OBJECTIONS TO NONSENTENTIALS

3.5.1 Objection 1 – *Underemphasizing the fall of hidden structure: Self-sufficiency*

While the above arguments together with the proposed syntactic strategy from Barton and Progovac (2005) successfully eliminate hidden structure, objections arise with regard to the sentential versus nonsentential distinction that is made following this important syntactic theoretical update. It is clear that the update puts verbless and verbal structures on the same footing in terms of the way that they are built: the creation of structures is based on the bottom-up merger of words in both cases. The implication of this is that there is no longer anything hidden or missing from a verbless structure: the verbless structure is no longer a version of a verbal sentence but is independently constructed. The importance of this move is of cardinal significance. Although it is not presented as such, this elimination of hidden elements and independence of verbless structures should be considered a great step forward toward their sentential status. Syntactically they have, in the above argument, been disentangled from verbal structures; their syntactic dependence on the verb – at some hidden level – has been one of the most frequent arguments for their lack of sentential status. Eliminating the hidden structure eliminates the hidden verb, and with it the insufficiency and dependency of the verbless structure on the verb. The syntax of verbless structures therefore becomes self-sufficient.

3.5.2 Objection 2 – *Sentence as a top-down syntactic category for all phrases*

Next, it should be noted that the only thing that now makes the verbless ‘nonsentential’ and the verbal ‘sentence’ syntactically different is the point at which the merging stops: in the latter this point is the satisfaction of verbal inflection which heads the structure, whereas the former lacks this element. However, endowing the inflectional phrase with such powers seems contradictory to the bottom-up strategy. While a structure that satisfies inflectional tense and agreement is proposed to be generated bottom-up, the requirement of inflection for sentential

status is a top-down requirement. Top-down requirements are deemed superfluous, unmotivated and arbitrary:

Projecting a sentential category on top of these NPs would not only be arbitrary, but also contrary to the minimalist principles of structure building. [... A] top-down strategy [starts] with an arbitrary top category, such as sentence [... T]he general requirement [...] prohibits any superfluous and unmotivated pieces of structure [such as the arbitrary top category]. (Barton & Progovac, 2005: 75)

In (Barton & Progovac, 2005), the economy principle is applied to verbless structures and all phrases which do not carry inflectional requirements: no sentential category is applied on top of these structures. However, if the sentential-versus-nonsentential distinction that they simultaneously propose is a syntactic one, then the economy principle is put into question with regard to inflectional phrases. In this case, the structures that do satisfy inflectional requirements defy the economy principle: they are assigned a top category – the sentence – as a result of their inflection (while the non-inflectional phrases are said to remain merely phrases). From a bottom-up perspective, the inflected phrase is just an inflected phrase (i.e. a phrase with satisfied inflection thanks to agreement between the verbal complement and the nominal specifier, but still a phrase). Attributing sentential status to a structure that is built bottom-up conflicts not only with noninflectional phrases (as noted by Barton & Progovac, 2005), but also with inflected ones. The expectation of a certain form for the structure cannot enter into *bottom-up* construction. Therefore, endowing inflectional phrases with sentential status and noninflectional phrases the status of nonsententials cannot be part of bottom-up syntax. If this distinction is kept, it must be recognized that it does not come from bottom-up syntactic considerations and is merely rhetorical. Consequently, this means that syntax does not determine the borderline between sentential and nonsentential structures, and can only speak of different types of phrases – some inflectional and others not.

3.5.3 Objection 3 – *Agreement*

It is worth noting that not only verbal, but verbless phrases as well, carry inflectional requirements that they must satisfy. The discussion so far has been about verbal inflectional requirements, i.e. tense, aspect and agreement, which need a verbal complement and a nominal specifier in order to be satisfied. However, inflection itself is an autonomous entity:

“inflectional morphemes do not belong to the verb phrase nor to any other specific element of the clause” (Moro, 1997: 258).

The inflectional feature of ‘agreement’ is also extremely relevant to verbless structures, as may be observed in the following:

(10) Inflection in verbless structures

English

- | | | |
|----|--|--|
| a. | <i>Articles</i>
*an problem / *a egg
*a milk
*a problems | [initial letter]
[countability]
[number] |
| b. | <i>Number</i>
*three dog / *one cats
*three milks / *three milk
*piece of cakes | [quantifier]
[countability]
[compound] |
| c. | <i>Case</i>
*from she
*we room | [preposition]
[pronoun] |
| d. | <i>Preposition</i>
*ticket on the opera | [preposition] |

Russian

- | | | |
|----|---|--|
| e. | <i>Number</i>
*три собака (<i>tri sobaka; three dog.SG</i>)
*три молоко (<i>tri moloko; three milk.SG</i>)
*города-герой (<i>goroda-geroj; cities.PL-hero.SG</i>)
*один родина (<i>odin rodina; one.M.SG home.F.SG</i>)
*мой друзьям (<i>moi druz'jam; my.SG.NOM friends.PL.DAT</i>) | [quantifier]
[countability]
[compound]
[gender]
[case] |
| f. | <i>Case</i>
*от она (<i>ot ona; from she.NOM</i>)
*рыжий коту (<i>ryžij kotu; orange.NOM cat.DAT</i>) | [preposition]
[adjective] |
| g. | <i>Preposition</i>
*билет на театр (<i>bilet na teatr; ticket on theater</i>) | [preposition] |
| h. | <i>Gender</i>
*рыжий кошка (<i>ryžij koška; orange.M cat.F</i>)
*рыжая кот (<i>ryžaja kot; orange.F cat.M</i>) | [adjective]
[adjective] |

The ungrammaticality illustrated in the above NPs and PPs shows that there exists agreement between the non-verbal lexical elements, particularly with regard to articles – (a), number – (b) and (e), case – (c) and (f), preposition – (d) and (g), and gender – (h). The fact that inflectional morphemes also occur in phrases other than the verb phrase suggests that, even though these structures do not have tense, nor agreement with a specifier subject, their lexical elements have their own particular syntactic inflectional requirements to satisfy. It seems coherent that inflection, as an autonomous entity, in phrases built from the bottom-up, be allowed different requirements that vary with the phrase type. Verbless structures are not without inflection, but their inflection is of a different type than inflected verbal structures – i.e. not requiring a specifier, nor tense. Phrases are syntactically satisfied from the point that the requirements of the lexical elements in the structure are fulfilled, and automatically so if they consist of only one word.

3.5.4 Objection 4 – *Distinguishing verbless embedded phrases*

If the reasoning presented so far is correct, then syntax appears to be in a serious bind with regard to the sentence, which has traditionally been in its domain of ‘jurisdiction’, i.e. the syntactic structure of a string of words is often treated as the ultimate judge of what does and does not constitute a sentence. The problem is that, keeping in mind the above updates, syntactically, it is not possible to tell the difference between (a) and (b) in:

- (11) a. car problem
b. Car problem!

When a hidden structure was supposed, it was the hidden structure that did this work. That is, ‘*Car problem!*’, treated as an elliptical sentence, was given hidden verbal structure that turned it into a sentence, whereas ‘*car problem*’ was assumed to be embedded in another structure and, not having been assigned any hidden elements, it remained just a phrase. On the bottom-up conception, both structures are noun phrases without any additional hidden elements. It is syntactically no longer possible to tell the difference between them. The problem must be overcome by some means other than syntax. A potential means will be presently proposed in *Part I: Chapter 4* which will provide a sentential account of the structures from an information-structural perspective.

3.5.5 Objection 5 – *Distinguishing verbal embedded phrases*

The above argument applies not only to verbless structures such as (11), but also extends to inflectional phrases, such as:

- (12) a. she has a car problem
b. She has a car problem.

Both (a) and (b) syntactically have equal status. On the grounds that both are built bottom-up, and considering objection 2 above, endowing (a) with sentential status would amount to a top-down requirement. Both (a) and (b) meet the requirements of inflection for tense, aspect and agreement, and, from a syntactic bottom-up perspective remain inflectional phrases. The difference between them is that (b) cannot be embedded into another structure. However, this, also, cannot be determined from a bottom-up syntactic perspective. The solution proposed in *Part 1: Chapter 4* targets this problem also; this is addressed with the requirement of linguistic and extra-linguistic context insertion.

3.5.6 Objection 6 – *False dichotomy between Elliptical Sentence & Nonsentential hypotheses*

The rejection of a covert structure lead Elugardo and Stainton (2005) as well as Barton and Progovac (2005) to conclude that what appear to be elliptical sentences (i.e. whether antecedent-based or not) do not actually have the syntactic structure of a sentence. However, the defence of the sentential status of the structures is only considered in terms of ellipsis.

A dichotomy becomes apparent between: (a) the ‘elliptical sentence hypothesis’ which appeals to hidden elements (i.e. treats the explicit words of a verbless string as visible left-overs from a full sentence, insufficient on their own to constitute a full syntactic structure and thus named ‘elliptical sentence’), and (b), the ‘nonsentential hypothesis’ in which verbless strings are treated as syntactically self-sufficient structures, but these self-sufficient syntactic structures, without the hidden elements, are deemed incapable of constituting sentences. The dichotomy is obviously false. The arguments against the existence of hidden elements are

devastating to the ‘elliptical sentence hypothesis’. However, their strength does not prove the ‘nonsentential hypothesis’ true.

What the strength of the arguments against the ellipsis defense reveals is that language uses independent (i.e. unembedded) structures that do not fit, explicitly or invisibly, the syntactic definition of the sentence. That these structures are not sentential holds true only as long as the syntactic definition of the sentence, that has now been confirmed to have ineradicable dissonance with language data, is accepted. Elugardo and Stainton (2005) provide a rigorous analysis of the implications of the fall of the elliptical sentence defense. The most important implication however appears to be overlooked: the question ‘*Is the syntactic model of the sentence still a credible one for language?*’, keeping in mind the admitted restricted application of a model free from hidden structure, should now receive a negative response.

Thus, nonsentential analysis is not a necessary conclusion of the rejection of hidden syntactic structure: other options include a search for alternative sentential models. Furthermore, like the ‘*elliptical sentence hypothesis*’, the ‘*nonsentential hypothesis*’ turns out to be another way that the traditional verbal models of the sentence are kept in place.

3.5.7 Objection 7 – *Nonsentential speech acts: Unresolved issues*

The most obvious implication of the rejection of the ellipsis defense – i.e. the need to revise the definition of sentential status – does not arise, because it is maintained that the structures are able to constitute speech acts without needing sentential status.

Stainton (1993, 1995, 1997, 2004) shows that the so-called nonsentential status of many of our speech acts is not merely an appearance: they actually do not fit either syntactic or semantic models of a sentence in several important respects – including not only the absence of sentential syntactic structure, but also the absence of semantic proposition and illocutionary force. Nevertheless, the fact that such structures are used to assert, ask, command and perform other speech acts, a fact which is taken to be self-evident from observation, leads Stainton to the conclusion that sentential status is not a requirement for felicitous speech acts.

The following reviews another key way that independent verbless structures are deemed to be different from canonical structures, analyzes the extent of the difference, and argues that the recognition of their use as speech-acts should motivate rather than neutralize updates to existing sentential models.

3.5.7.1 a. Potential semantic differences

The discussion so far has compared verbless and canonical structures in terms of syntax. However, independent phrases without inflection are also typically said to be *semantically* different from independent finite verb phrases.

The first semantic issue concerns whether or not the structure expresses a ‘proposition’. The notion of a proposition is itself full of nuances, as evidenced in an overview of its different conceptions by Scott Soames (2012). In short, it is minimally defined as: an abstract cognitive structure by means of which an agent predicates something of something else (Soames, 2012: 218). Typically, only syntactic sentences are said to encode propositions; whereas words and sub-sentential structures are said to have the semantic content of concepts, properties or quantifiers, but not propositions (Stainton, 1995: 284).

The second issue concerns whether or not the expression has ‘illocutionary force’. Illocutionary force refers to the use of language to perform a social act (Siemund, 2018: 33). For instance, a declarative sentence is often used in order to assert, an interrogative to question, an imperative to request, and an exclamative to exclaim (Stainton, 1995: 285). The particular force may be both linguistically encoded and contextually supplied (Siemund, 2018: 33). For example, an important social act is assertion. To ‘assert’ something (or to make an ‘assertion’) means to commit oneself to the truth of a proposition, i.e. the speaker treats the proposition as something that he or she knows, believes and is justified in believing, and thereby becomes liable for being correct or incorrect, so that if it turns out that the assertion does not reflect the reality of things, e.g. if the speaker asserts that the letter is from France when in reality it is not, then the speaker is taken to be mistaken or lying (Green, 2017: 3; Soames, 2012: 217). Only syntactically sentential structures, but not sub-sentential independent words or phrases, are claimed to exhibit an illocutionary force property (Stainton, 1995: 285, 288).

These two concerns constitute very strong additional semantic motivators for defending the sentential status of independent verbless strings (which, as discussed above, has typically been done by means of attributing hidden verbal syntactic structure). After all, when someone pulls out a letter from a mailbox and utters ‘*From France*’, it is reasonable to desire that upon hearing the speaker’s utterance it is possible to (a) attribute to it the semantic content of a proposition in which arrival from said country is predicated of the object pulled out from the mailbox, and, no less important, (b) to treat the speaker’s utterance as having the illocutionary

force of an assertion, rely on its truth and hold the speaker liable (i.e. if it is later claimed that the letter is not from France, cognitive dissonance will no doubt arise).

To get noncanonical structures out of this predicament, some have argued that the independent strings should be attributed canonical sentence syntax (i.e. the hidden structure syntactically elliptical sentence arguments); this would immediately neutralize the above concerns but it requires proof that additional hidden structure actually exists, which, as has been shown above, is problematic. Others have argued that they are semantic propositions (i.e. the semantic ellipsis argument that will be discussed below). Finally, yet others have argued that they have none of these, but that this does *not* pose as serious of a problem as it first seems. Stainton's work belongs to this latter perspective. He argues that expressions which do not have a syntactically sentential structure, do not encode the content of a proposition, and are not force-bearing, are found used to make an assertion and thus carry out a speech act (Stainton, 1995: 285). In other words, on such an account, neither syntactic sentential structure, nor propositional content, nor illocutionary force, would be deemed present in an independent verbless string, but the resolution of the predicament would be that these are not necessary for all language since speech acts can be performed anyway.

3.5.7.2 b. Illocutionary force

The argument against the presence of illocutionary force deserves particular attention and goes as follows.

3.5.7.2.1 Step 1: Uttered tokens and linguistic types

To start, Stainton (1995: 288) draws a distinction between two uses of the term 'illocutionary force'. The fact that the act of uttering (i.e. whenever someone asserts, asks, commands, exclaims, etc.) necessarily attributes to the uttered expression an illocutionary force is not questioned. What is at issue concerns the inherent properties of the expressions themselves, and more particularly of the expression types.

The claim is that illocutionary force is *not* a property of the noncanonical type of expression – i.e. the type that does not have the syntactic structure of a canonical sentence, but *is* a property of the canonical type of structure that is headed by verbal inflection. In Stainton's

words, the illocutionary force under question concerns the types of expressions, not their uttered instances:

It is a platitude that whenever someone asserts, commands, or asks their utterance has illocutionary force. Since it is part of our claim that these expressions – whether they turn out to be ordinary words and phrases or semantically elliptical sentences – are commonly used to make assertions, we of course agree that utterances of them have illocutionary force; in particular some have assertoric force. The question at hand concerns the *expressions*, not utterances of them. That is, to employ some standard vocabulary: we are inquiring about the properties had by certain linguistic types, not their tokens. Our conclusion shall be that these linguistic types do not have illocutionary force. (Stainton, 1995: 288)

Thus, a token (or uttered instance), of for example ‘*From France.*’, is admitted to have illocutionary force (in this case assertive), but it is rather the type to which this uttered instance belongs – the so-called ‘nonsentential’ – whose illocutionary force is questioned.

3.5.7.2.1.1 *Objections to Step 1: Type-token divorce*

This distinction between uttered tokens and the properties of types is not a minor one. It would normally not be controversial to say that *types* are abstractions that are derived from data, i.e. from *tokens*: one looks at instances and divides them into types.

Since all of the tokens at issue are said to be necessarily force-bearing (by virtue of their utterance), how is it possible that any of the types that describe these tokens are not also force-bearing? If the types are derived from tokens, then the types must also be force-bearing; or, alternatively, it cannot be that all tokens, i.e. all uttered expressions, are force-bearing (but this latter option is explicitly ruled out).

Alternatively, if the types are not derived from the tokens – i.e. if the fact that the tokens are said to have been uttered, whereas the types have not, is taken to represent a clean division between them – then it is no longer clear what the types represent. A complete divorce between the data tokens and the abstract types destroys the value of the types. Abstractions are only as relevant as they attempt to reflect what is actually going on, otherwise any internally coherent abstract model, i.e. any types that don’t contradict with one another, would suffice. The division between types and tokens cannot be a complete one, otherwise the entire purpose of types is lost and their verification becomes impossible. In this particular case, when all uttered tokens are said to necessarily carry illocutionary force, it is hard to see how any abstract type that does not have illocutionary force would pass verification with the data. Thus, even if the

rest of the argument (for the absence of illocutionary force in so-called nonsententials) proves to be valid, this foundational distinction means that the linguistic data that the argument applies to does not actually exist. (In other words, even if proved that some type of sentences does not have illocutionary force, it is not a type that can ever be uttered.)

3.5.7.2.2 Step 2: Context-free analysis

The fact of making the distinction between uttered tokens and their linguistic types, allows one to take a crucial step: it makes it possible to consider linguistic types without context (after all, it is the tokens that are actually found in context, whereas the types abstract from the individual occurrences).

The illocutionary force of abstract types is thus considered in terms that are said to be separate from any utterance context. The syntactically canonical type is illustrated with paradigm cases, including the following from Stainton (1995: 289).

- (13) a. Context-free of the type: Snow is white
 b. Context-free of the type: Is John wearing a hat?

Such sentences are said to have the inherent property of illocutionary force thanks to their recognized standard use to assert and to question without needing any appeal to context:

Here is our hypothesis: taken apart from any context, someone who knows English can make an educated guess about what the speaker of each expression would be doing. This, we think, is the property which all expressions with illocutionary force share. (Stainton, 1995: 289)

The so-called nonsentential type, for instance (14), is denied the illocutionary force property because it is not possible to recognize the standard use of such expressions from their words alone; some context is needed in order to form even a guess about what the speaker is doing by using this expression.²²

- (14) a. Context-free of the type: White
 b. Context-free of the type: In a hat?

²² The idea that abstract schemas of clause types are associated with potential illocutionary forces is also traced to François Recanati (1987: 127) and William Alston (2000). See discussion in (Siemund, 2018: 36).

Thus, a canonical sentence type is said to have illocutionary force, while departures from this type do not.

3.5.7.2.2.1 *Objections to Step 2: Illocutionary force without context*

No doubt a perceptible difference in the ease of interpretation between (13) and (14) exists. However, this is not necessarily due to a context-independency of one and context dependency of the other. To what extent can the interpretation of types be really considered context-independent?

a) Besides the fact that context-free analysis requires accepting the type-token divorce described above (and thus, to come to terms with the unverifiability of language models that necessarily results from definitive separation of ‘context-free types’ from ‘context-dependent tokens’), the following arguments give additional reason to reject the possibility of context-free analysis of illocutionary force.

b) In the canonical structure type, such as ‘*Snow is white*’, it is the potential attribution of the structure to a context – i.e. to an imagined context – that makes it possible to assign to this structure the illocutionary force of an assertion prior to any actual use. Thus, it is not just the structure alone and context-free that carries an illocutionary force. Rather, it is the expectation of the insertion of this structure into a context that is actually responsible for the perceived force.

c) Moreover, it is worth noting that the imagined assertive force here is only a potential interpretation – a likely use based on most expected contexts, not a necessary one. For example, uttered three times in a row by a theater director upon seeing the wrong coloured snow brought in, ‘*Snow is white*’ would take on the force of an insistent request to change the snow. This signals again that the force attributed to the abstract schema is an expectation, based on an imagined ‘regular’ context of use, not an inherent property of a context-free structure.

d) The perceived difference of force between the canonical and the noncanonical structure type when analyzed from the perspective of a blank page, could be due to the differences in the content of the structures. Unlike (14) ‘*White*’, (13) ‘*Snow is white*’ rules out subjects other than snow, which consequently narrows down the range of contexts imagined for the use of (13) and makes it easier to interpret than (14). This ease of interpretation gives the illusion of a context-free structure with an inherent force property. In reality, the less precise structure of (14) means that it can be used in more contexts, thus have more potential force interpretations not less.

Thus, it seems reasonable that the more precise linguistic elements are evoked in a structure, the easier it becomes to imagine the context and the potential force that would be gained upon use of the structure. Illocutionary force arises from use in a real or imagined context; its analysis as a property of an abstract schema is only conceivable through statistical association with a particular use. In other words, the contention presently made is that the existence of an inherent force property cannot be a distinction between canonical and non-canonical sentences: force is absent not only from non-canonical context-free abstract schemas, but also from canonical ones.

3.5.7.2.3 Illocutionary force cross-linguistically: Stability considerations

The contradictory nature of treating force as a property of only canonical structures becomes even more obvious when considered from a cross-linguistic perspective. Particularly, attributing force as an inherent property of a certain abstract schema (i.e. syntactic structure type) means different abstract schemas correspond to differences in force: as illustrated above, it is said to be present in (13) but absent in (14). In other words, tying illocutionary force to syntax means that it will naturally vary with the syntax. Such variation leads to the question of the domain to which illocutionary force belongs. Translation aims to keep semantic and pragmatic properties stable between the source and the target, despite typological variation in syntactic structure between languages. If illocutionary force is tied to syntax, the translation of a canonical verbal structure in one language as a noncanonical verbless structure in another language, would mean that illocutionary force is found in the former but not the latter. The direct link of illocutionary force to syntactic structure means that speaking of it in terms of a semantic or pragmatic property, which is expected to remain stable in translation, is no longer possible.

For instance, it is reasonable to expect that the translation of an assertion will keep the assertive illocutionary force (i.e. if a speaker makes a truth commitment in English, that commitment will be maintained in translations). The following provides unmarked Russian translations of the paradigm English canonical structures from (13):

(15) a. English: Snow is white.

Russian: Снег белый.
sneg belyj
NN ADJ
snow white

b. English: Is John wearing a hat?

Russian: Джон в шапке?
Džon v šapke
NN PREP NN.PRE
John in hat

The Russian structures in (15) are not headed by inflection. Any attribution of a verb to their structure would face the hidden structure arguments above. (This includes the Russian verbless translation of the English copula sentence in (15a), a claim which may at first glance seem controversial but should not, as will be discussed next in 3.5.7.2.3.1, since, in the context of the current arguments, such is the coherent reading of the rejection of hidden structure.)

On the account that illocutionary force is an inherent property of the syntactic structure types, this means that the English canonical sentence in (15a) ‘*Snow is white*’ is an assertion, and thus carries a truth commitment, whereas its Russian equivalent, a noncanonical syntactic structure cannot be treated as such. Similarly, the canonical sentence (15b), an English interrogative clause, has a questioning force, while its Russian translation is not headed by verbal inflection and thus would lack force.

This provides another reason as to why it would be inappropriate to directly link illocutionary force to syntax; a reason that applies even in discussions that despite the arguments above still admit context-free types. It seems reasonable to speak of certain linguistic and syntactic elements as being *indirectly* linked with highlighting illocutionary force through frequent association; however, it would be contradictory to make the link a direct one and posit an inherent property of syntactic structure types.

Such a separation of syntactic structure type and illocutionary force also has another important positive consequence. The question of whether or not verbless sentences can be assertions is naturally resolved. Such a question simply does not arise if the link between syntax and force is not treated as a direct one. There is no direct link between syntactic structure and illocutionary force; only an indirect one based on frequency of use. Illocutionary force is realized upon the insertion of the structure into a context. As a result, (i) force remains stable

through translation and (ii) a priori any syntactic structure may have illocutionary force and it remains to frequency of use to see if it does. Unless a direct link holds between illocutionary force and syntax, and unless all of the problems of such a link can be resolved – i.e. the resulting cross-linguistic instability of force and the question of the domain to which illocutionary force belongs – then the question becomes why any bias against particularly the illocutionary force of verbless sentences should arise. Verbless sentences would have the illocutionary force of assertions if they are used to assert, and the same for all of the other types of illocutionary force. In other words, illocutionary force, which arises from the utterance of a structure in order to perform some social act, does not pose any problems for verbless sentences.

The question concerning the semantic content, e.g. what it is that is actually asserted when an independent verbless structure is used to assert, will be explored below as part of the discussion on whether or not semantic, rather than syntactic, ellipsis is involved, and in the sentential account proposed in *Part 1: Chapter 4*.

3.5.7.2.3.1 *Present tense copula in Russian is not an exception*

Before continuing, it is important to emphasize that the Russian structure in (15a) is not an exception to the hidden structure arguments. In other words, analyzing (15a) as a structure that contains a present tense copula would need to respond to the syntactic problems with hidden structure detailed above.

The reason that this tangent is important is that in Russian, the present tense form of the copula verb ‘быть’ (*byt’*; lit. ‘*be*’), i.e. ‘есть’ (*est’*; lit. ‘*is*’), is not used in unmarked contexts. As a result, a paradigm is created for Russian unmarked copula structures which treats the absence of the copula verb as being akin to the morphological use of absence that is evident in, for instance, the English expression of number, where an ‘*s*’ is added only for plural nouns and the absence of the ‘*s*’ typically carries a singular reading.²³ The copula paradigm is often

²³ The Concept of the ‘Zero’:

The analogy between a zero-form of the verb in a verbless sentence and a zero-morpheme is famously made by Igor’ Mel’čuk (1974; 1979) and the definition of a zero as an existing form (as opposed to a fact of absence) is actively maintained in e.g. Letuchiy (2018), Frolova (2012), Kopotev & Gurin (2007), amongst others.

The concept of a zero-form is found discussed in the context of arguments for dispensing with (or limiting) Šhamatov’s one-part / two-part distinction, in favour of a two-part sentence model in which the presence of a zero is said to complete the missing part (Dolin, 2001). Such a perspective is found in E. A. Sedel’nikov (1961: 67, 73), who is traced as being the first to apply the concept of the ‘zero’ to all Russian sentences that lack an explicit expression of the subject or the verbal predicate; with such a foundation, the present tense is attributed to verbless sentences syntactically through the presence of a zero-copula in a past/present/future paradigm

presented as: past tense ‘БЫЛ’ (*byl*; lit. ‘was’), present tense ‘Ø’ (Ø; lit. ‘is’), future tense ‘будет’ (*budet*; lit. ‘will_be’).

While it is certainly true that when one desires to speak of the present time in an unmarked way the present tense form of the copula verb is not used, this paradigm must not be overgeneralized by attributing the present tense to verbless syntactic structures. The following illustrates that a zero present tense copula also faces the syntactic problems that come with attributing hidden structure that were discussed above.

(Tixanova, 2019; Potapova, 2017: 12). This concept of a deep level zero-form of the verb is also used in studies analyzing the meaning of particular verbless structures that are said to contain the zero-forms (e.g. Dymarskij, 2018; Mel’čuk, 2019).

Systems have been developed for identifying whether the verbless structure contains an ellipsis versus a zero-verb form versus a zero-copula form (e.g. Letuchiy, 2015; 2013). The key distinction between ellipsis and a zero often centers on whether or not there is a change in meaning from adding the verb. More specifically, a change in meaning from the re-introduction of the verb signals a zero, whereas the absence of a change in meaning from re-introduction is identified as an ellipsis, as summarized in (Dymarskij, 2018: 11).

The term ‘zero’ tends to take on a particular use in Russian literature, studies within Igor’ Mel’čuk’s highly influential Meaning-Text framework, and not only, corresponding to the warning that ‘zero’ should be reserved for occurrences of particular types of zero-lexemes:

[С]ловосочетания типа «нулевое сказуемое» (подлежащее, дополнение), «нулевой глагол», «нулевое имя», «нулевой артикль», и т.п., а также «нулевой синтаксический элемент» и «нулевой вариант слова» (термы Л. С. Бархударова – Бархударов, 1966: 180) должны употребляться только для обозначения нулевых словоформ или лексем: нулевое сказуемое = сказуемое, выраженное нулевой словоформой; нулевой глагол = нулевая глагольная лексема; нулевой вариант слова = нулевая словоформа лексема; и т.п. (Mel’čuk, 1974 : 355)

Translation: “Terms such as ‘zero predicate’ (subject, complement), ‘zero verb’, ‘zero noun’, ‘zero article’, etc., as well as ‘zero syntactic element’ and ‘zero variant of a word’ (terminology of L. S. Barxudarov, 1966: 180) must be used only for the description of zero word-forms or lexemes: zero-predicate = predicate expressed using a zero lexeme; zero-verb = zero verb lexeme; zero variant of a word = zero word-form for a lexeme; etc.”

The definition in emphasizes most importantly that the zero is to be treated as an actual form, not simply absence.

The results of the studies which appeal to zero-verbs often attribute specific meanings to the zero-forms of various lexical verbs that may be introduced in specific structures and contexts. The framework involved in identifying and describing lexical zeros is extremely intricate. Nevertheless, due to the necessity of appealing to hidden structure, the studies of lexical zeros seem to ultimately face the issues discussed above that come with such an appeal. Furthermore, the attribution of meaning to zero-forms of verbs would mean that a verb would thus carry not only its standard lexical definitions for when it is present but also for when it is absent. Lexical definitions would be attributed to zero-verbs from structures where a verb may have been introduced but was not. As a consequence, the number of dictionary entries for a single zero-verb would extend to the various structures that it may be a part of, plus, due to word order flexibility, potentially also to the various positions in which it may be found within the same structure.

i) First, it must be stressed that a present tense form of the copula verb exists.

The form, ‘есть’ (*est’*; lit. ‘*is*’), not only exists, but it is used. Its use may be evidenced in (16):

- (16) a. Снег был, есть и будет белый.
sneg byl est’ i budet belyj
snow was, is and will_be white
- b. Я и есть тот человек.
ja i est’ tot čelovek
I PART am that person

The above structures are of course marked, that is to say they suggest contrastive readings. For instance, (16a) contrasts the past, present and future time of the association of ‘*snow*’ with the property ‘*white*’, and (16b) uses an emphatic particle prior to the present tense copula in order to suggest that the identification of ‘*I*’ with ‘*that person*’ should not be doubted. Despite the uses being marked, the fact that there exists an expressed form of the copula verb in the present tense is not a minor one.

The existence of the present-tense form reveals nuances in the paradigm. The copula paradigm becomes: past tense ‘*был*’ (*byl*; lit. ‘*was*’), present tense ‘ \emptyset ’ (\emptyset ; lit. ‘*is*’) or ‘*есть*’ (*est’*; lit. ‘*is*’), future tense ‘*будет*’ (*budet*; lit. ‘*will_be*’). This precision results in questions concerning the choice between the ‘ \emptyset ’ (\emptyset ; lit. ‘*is*’) and ‘*есть*’ (*est’*; lit. ‘*is*’) surface forms. It becomes necessary to explain the syntactic way that a zero present tense form of the copula appears and disappears, as well as the reasons for the variation. Present-tense inflection is said to be involved in both the explicit and the zero-form, which means that the syntactic mechanism would need to condition the variation on something other than tense. In other words, positing a zero-form on the syntactic level to stand in for what appears to be a missing present tense copula would require hidden syntactic structure and a deletion mechanism that, as discussed above, is a problematic costly endeavour.

ii) The second point important to emphasize is that, even for well-studied patterns, the paradigm is difficult to generalize.

Upon encountering a verbless structure such as (15a), ‘Снег белый’ (*sneg belyj*; lit. ‘*snow white*’), the fact that the explicit past tense or future tense verbal copula may be introduced in a paraphrase does not indicate that the structure necessarily does carry an

inexplicit present tense copula. After all, paraphrases with other verbs are also possible, e.g. ‘падает’ (*padaet*; lit. ‘falls.V.PR’), ‘лежал’ (*ležal*; lit. ‘laid.V.PS’), ‘привезли’ (*privezli*; lit. ‘was_brought.V.PS’), which creates competition both for the copula and the supposed present tense inflection. Yet, the alternation with the past and future forms is often presented as the central principle for identifying the zero copula in the paradigm:

Понятие «нулевая связка» используется прежде всего тогда, когда нулевая форма сказуемого в одних временах и наклонениях чередуется с ненулевой в других. В русском языке нулевая связка заменяет формы настоящего времени индикатива глагола *быть*. (Letuchiy, 2018: 1) ²⁴

The crucial problem with the principle of alternating tense is that it works only when one is already aware that the verbless structure is part of the paradigm in which an unmarked way of indicating the present is desired, but not as a rule for determining that the structure actually is a part of that paradigm.

This point is also made by Mikhail Dymarskij (2018), who points out the practical difficulty in distinguishing between the zero-copula and a zero-verb:

Трудность, однако, состоит в том, что, [...] в некоторых случаях полноценное сказуемое регулярно опускается, в результате чего предложение становится практически неотличимым от предложений с нулевой связкой. (Dymarskij, 2018: 6) ²⁵

Чередование с нулем отнюдь не может служить достаточным основанием для вывода о связочной природе глагола. (Dymarskij, 2018: 11) ²⁶

Dymarskij (2018) shows that the difficulty exists even for the typical possessive structure in Russian.

This structure ‘У *x-a y*’ (*u x-a y*; lit. ‘at *x*.GEN *y*.NOM’), which has been famously studied in (Selivërstova, 1973; Paillard, 1984; Apresjan, 2017; Mel’čuk, 2019), is usually considered from the perspective of the paradigm in which a choice is made between the explicit present

²⁴ Translation: “The concept ‘zero-copula’ is used above all when a zero form of the predicate in certain tenses and moods alternates with a non-zero form in others. In the Russian language, the zero copula replaces the present tense indicative forms of the verb ‘быть’ [*byt*]; lit. ‘be’.”

²⁵ Translation: “A difficulty, however, consists in the fact that, in some cases a full predicate is regularly dropped, and as a result the sentence becomes practically indistinguishable from a ‘zero-copula’ sentence.”

²⁶ Translation: “The alternation with a zero cannot in anyway be a sufficient basis for conclusions regarding the copula status of the verb.”

tense form of the copula verb and the zero copula alternative, e.g. between ‘У него есть седые волосы’ (*u nego est’ sedye volosy*; lit. ‘at him is gray hair’) and ‘У него седые волосы’ (*u nego sedye volosy*; lit. ‘at him gray hair’). However, even this verbless structure can work not only with a present tense copula, but also with full lexical verbs, and not necessarily in the present tense, as for instance in (17):

(17) a.	У	него	∅	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>∅</i>	<i>sedye</i>	<i>volosy</i>
	<i>at</i>	<i>him</i>	<i>∅</i>	<i>gray</i>	<i>hair</i>
b.	У	него	есть	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>est’</i>	<i>sedye</i>	<i>volosy</i>
	<i>at</i>	<i>him</i>	<i>is</i>	<i>gray</i>	<i>hair</i>
c.	У	него	растут	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>rastut</i>	<i>sedye</i>	<i>volosy</i>
	<i>at</i>	<i>him</i>	<i>is_growing</i>	<i>gray</i>	<i>hair</i>
d.	У	него	выросли	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>vyrosli</i>	<i>sedye</i>	<i>volosy</i>
	<i>at</i>	<i>him</i>	<i>grew</i>	<i>gray</i>	<i>hair</i>

Though the verbless structure in (a) is said to correspond to the past/present/future paradigm, it is not at all clear that the zero in the structure has the meaning of the present tense zero copula in (b), as opposed to a full lexical verb, such as (c), and another tense such as (d). Furthermore, all three options for the zero, as well as the verbless structure itself, are a priori possible in the same context, for instance upon observing a close friend’s hair and discovering that some of them are gray. Thus, even individual consideration of each verbless structure within a specific given context, as insisted on for instance by Leonid Iomdin (2003: 1) and Jekaterina Mažara (2011: 3), does not resolve the competition for the meaning that the zero actually signifies,²⁷

²⁷ There exists another perspective regarding the implication of the impossibility of the resolution of the verb. Mažara (2010) is said to follow Apresjan (1986: 113) in the argument that:

Every zero has a certain meaning, but it is not possible to reconstruct that meaning, since it does not match the meaning of any existing Russian lexeme [...] it is impossible to complete the sentence with the missing word without changing the meaning. (Mažara, 2010: 232-233)

Such a conception of the ‘zero’ is an interesting one: it maintains the existence of a zero form that carries its own meaning while simultaneously admitting the impossibility of reconstruction or even identification of that meaning. The zero form thus becomes something that in addition to being surface-invisible carries an unknown – and unknowable – semantics. If the existence of such a form – a form that is both syntactically and semantically inaccessible – is granted, it must be noted that such a form cannot by this very definition be identified as a verb or as carrying any specific inflection, nor their meanings. As a result, the convenience of the ‘zero’ is lost from

and this is true even for structures that fall within the copula paradigm (i.e. into which a past or future explicit copula may be introduced).

Similarly, but focusing on a version of the possessive which simultaneously specifies location, Dymarskij (2018) provides the following examples:

- (18) a. У меня в кармане трояк завалился.
u menja v karmane trojak zavaljalsja
at me.GEN in pocket.PRE money laid_around.PS
- b. А ну-ка, глянь, что у меня в кармане водится.
a nu-ka gljan' čto u menja v karmane voditsja
P P look what at me.GEN in pocket.PRE lives.PR

The above illustrate the use of verbs other than the present tense copula in structures that, were they to be verbless, would typically be treated as part of the copula paradigm and attributed present tense.

If the verbless structure is treated as carrying a verb on a hidden level, the competition is thus not only between a zero present tense copula and its explicit version ‘есть’ (*est*’; lit. ‘*is*’), but also between full lexical verbs and other tense inflection possibilities. Furthermore, the introduction of a non-copula verb in an inflection other than the present should in principle be ruled out if the verbless structure fits the past/present/future paradigm and inherently carries the zero element, which as the above examples show is evidently not the case.

iii) Alternative reading: unspecified for tense or other inflection

Thus, a hidden structure cannot resolve either a copula verb, or a lexical verb. In other words, the appeal to a zero copula form that represents a deep level verb and tense inflection is problematic even when the anticipated verb is a copula and the verbless structure is part of

the perspective of its utility for granting sentential status (i.e. fulfilling the need for a verb and inflection). In other words, the concept of a zero form that was introduced in order to save the canonical sentence model is given another blow if, adding to its syntactic issues, it cannot be semantically identified either. It seems important that if any zero form is posited it must have some existing lexeme with which it matches; otherwise, what actually exists is ‘a potential to introduce some form’ but not something that is itself a form. It thus would be more reasonable to consider the impossibility of reconstruction of the meaning as support for the idea of a form-less absence, as opposed to a meaningful syntactic zero form.

the copula paradigm.²⁸ Therefore, tying illocutionary force to syntactic structure means that the notion of illocutionary force would lose cross-linguistic stability; and this is also true when copula structures, such as ‘*Snow is white*’, are compared to the typical unmarked verbless translations, such as ‘Снег белый’ (*sneg belyj*; lit. ‘*snow white*’).

Concerning inflection, it seems preferable to treat verbless sentences, even sentences that fit the copula paradigm, as structures that are a priori unmarked for any tense. That is to say that in an independent verbless structure there is no additional hidden verbal structure and thus no tense marking, nor other inflection tied to verbal categories.

This obviously does not exclude the use of these structures with a temporal meaning that is not encoded in them by way of verbal inflection (i.e. temporal reference may be encoded by other elements such as adverbs or arise as implicature from the context); though, the temporal meaning may also remain, both syntactically and semantically, unspecified. Thus, while it is not necessarily a problem for temporal meaning, the absence of the hidden verb and its inflection does however pose a serious problem for sentential status if the latter is considered in accordance with the canonical syntactic requirements.

²⁸ There exists another loose use of the term ‘zero copula’:

The term ‘zero copula’ is sometimes used in a way that is neutral to a hidden level and does not carry the connotation with an existing form. The term would simply be used to state that in unmarked situations only the past and future tenses are encoded using the copula; the structure in this case is attributed the present tense by a default, hypothetically and without any claims of encoding. Such a neutral use does not permit the structures to be called syntactic sentences, which, though it is the main reason for the development of the concept of the zero, is not concerned with sentential status. Leon Stassen (2013) draws attention to the difference, and as he does not take a stand on the sentential status of verbless structures, makes the following comment to emphasize this distant neutral use:

To avoid misunderstandings, a couple of remarks on the use of the term *zero copula* may be in order. This term has been used traditionally in the literature on predicate nominal sentences (see, for example, Benveniste, 1966). It will therefore be familiar to many readers, which is why I have decided to employ it in this chapter. I realize, however, that the term may give rise to Euro-centric interpretations, in that it might suggest that languages which do not have an overt copula are somehow “defective”. On a more theoretical level, the term might suggest that, in languages with zero copula encoding of predicate nominal sentences, there is some phonologically null copula present in the syntax of these sentences, or, alternatively, that there is something corresponding to a copula in the conceptual/semantic structure underlying predicate nominal sentences in these languages. It should be stressed that none of these inferences is warranted. The term *zero copula* is used here as a strictly neutral technical label, in that it refers purely to a construction in which the relation between a subject and a nominal predicate is not marked by an overt item. (Stassen, 2013: 1)

3.5.7.3 c. Sentential status: Fight or flight

This section began with the claim that one of the key implications of the rejection of hidden structure or ellipsis defense – i.e. the resulting need to revise the definition of sentential status – is overlooked, because independent verbless structures are able to constitute speech acts without syntactically constituting sentences. For Stainton, they are able to constitute speech acts even without illocutionary force. While the idea that there may be a speech act without illocutionary force is an interesting one, it faces problems. In the present section it has been argued that the foundation for the idea of a force-free speech act – i.e. the type/token divorce and the context-free analysis of illocutionary force – is unstable. Furthermore, it has presently been argued that such an idea destroys the cross-linguistic applicability of the notion of illocutionary force. Thus, the claim that speech acts may exist without illocutionary force is presently rejected. Instead it is presently maintained that independent verbless structures are capable of carrying illocutionary force and need to if they are to be considered speech acts. In other words, the reasons as to why independent verbless structures can constitute speech acts differ, but the conclusion that they can is upheld. It is now possible to get back to the point with which this section began, i.e. the objection to using the status of a speech act in order to neutralize the need to update existing sentential models.

It is clear that the concept of the sentence is in danger from the arguments against hidden structure. However, settling for non-sentential status for independent verbless structures is also problematic, and not least for the reason that speech acts are indeed performed without meeting the canonical syntactic requirements. Given that syntactically nonsentential structures can be speech acts points to the limits of syntactic sentential models in accounting for language use.

Historically, various attempts have been proposed to address this problem with alternative versions of hidden elements. The efforts put toward the development of a hidden structure, despite ultimately not arriving at satisfactory results, were made in order to adjust the canonical model of the sentence so as to account for actual language and the inerasable existence and use of non-canonical structures.

The non-sentential speech act argument now makes a call to give up those efforts. Simultaneously, it does not provide any new sentential model or even call for the need to update sentential criteria. The question ‘What should we do with non-canonical structures?’, in essence, instead of traditional attempts to make them fit, receives the new answer: let them be as they are and keep going on without them, i.e. without integrating them into sentential

models. More precisely, independent verbless structures become bottom-up constructed syntactic entities, their use to make a speech act is accepted as a given reality of language, which however is not seen as worthy of a sentential explanation.

The target scope of the syntactic sentence is thus reduced to only canonical structures. *Figure 3* illustrates the reduced goals of the nonsentential hypothesis compared to the ellipsis or hidden-structure-based hypotheses.

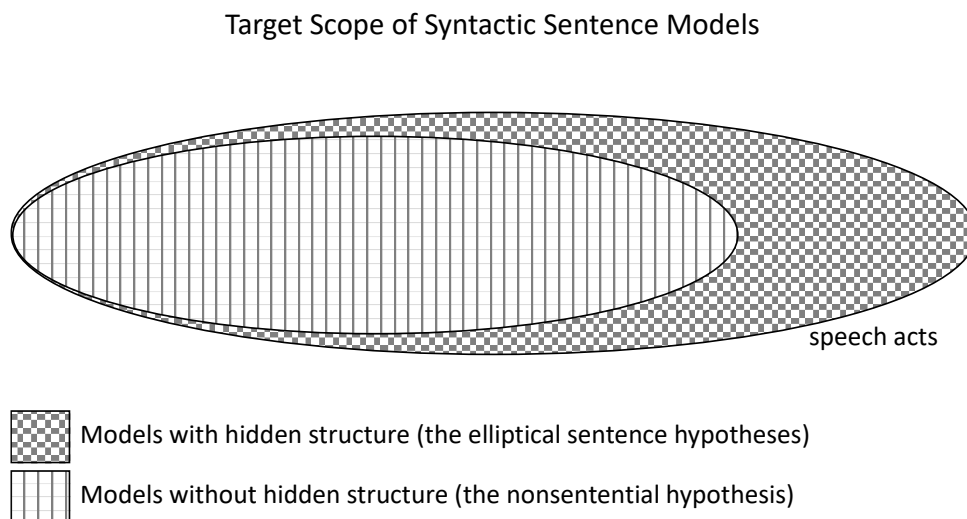


Figure 3. Target scope of syntactic sentence models

The figure represents the aims of the sentence models with regard to independent, i.e. unembedded, structures used as speech acts; the actual scope of course depends on the success of the models (which as the discussion in the present section attempts to show is dubious not only for models involving hidden structure but also for the models that argue against them). The nonsentential hypothesis is less ambitious in the sense that it accepts that independent verbless and other non-canonical structures are a challenge that syntactic models of the sentence are not able to meet. The use of such structures unembedded to perform speech acts only emphasizes the importance of that challenge.

3.6 SEMANTIC ELLIPSIS

Another way that one may attempt to account for independent verbless structures, without shaking the canonical syntactic models of the sentence, is to argue for their sentential status not in terms of their syntactic structure but in terms of their semantic encoding.

When one utters (6b) '*from France*' upon taking a letter out of a mailbox, it is possible to analyze what has occurred not as a syntactic ellipsis, i.e. there is no zero verb nor any hidden syntactic elements in the structure of '*from France*', but instead a semantic one. That is to say that there are hidden semantic elements which make it possible for the expression '*from France*' to encode a semantic proposition.

As mentioned, a 'proposition' is an abstract cognitive structure by means of which an agent predicates something of something else (Soames, 2012: 218). Typically, only syntactic canonical sentences are said to encode propositions (Stainton, 1995: 284). For instance, the phrase '*from France*' would typically be said to encode the semantic content of a *property*, but not of a *proposition*. The propositional meaning would require that the property be explicitly attributed to something, but in '*from France*' this attribution is not linguistically explicit and as a result, a proposition would not be encoded into the expressed bare phrase. A propositional meaning may be understood from the context, but would not be encoded.

In contrast, the semantic ellipsis hypothesis would argue that '*from France*' may encode not only a property, but also a proposition. In cases when '*from France*' is embedded, e.g. '*This letter is from France*', it would have the semantic type of a property. However, when it is used independently, i.e. syntactically unembedded, as in '*From France.*' uttered in the context of receiving a letter, it would belong to the propositional semantic type. In the latter case, it is not only the concept of the property '*from France*', but also the support to which the property is attributed (i.e. the concept of the letter), as well as the attribution (i.e. the concept of the relation between the letter and the property), that gets encoded as part of a complete proposition. The ellipsis is not syntactic (a syntactic subject and syntactic inflection for agreement and tense does not exist) but semantic (the support and the relation of the proposition, without which there would only be a property, are invisibly encoded).

The notion of semantic ellipsis is developed by Dalrymple et al. (1991) and Dalrymple (2005) and explored in detail in (Stainton, 1995; 1997; 2000; 2004; Elugardo & Stainton, 2005). It is summarized as follows:

When speakers (appear to) utter words or phrases in isolation, what they produce are not syntactically sentences – the expressions uttered have no subject, verb, etc. But they are semantically sentences, in the sense that they express propositions. [...] It is tempting to think that this expression [*fire*], which is used in theaters and such, has a proposition as its meaning: the same proposition expressed by [*There is a fire*']. (Stainton, 1997: 66)

Whenever a speaker makes an assertion by uttering an unembedded word or phrase, what that speaker really utters is an elliptical sentence in the sense that the semantic type of the expression uttered is propositional. (Stainton, 1995: 289–290)

Using an abbreviated sound pattern that corresponds to an equally abbreviated syntactic structure, but where that syntactic structure somehow linguistically encodes the complete message recovered. (Stainton, 2004: 275)

In other words, the semantic ellipsis hypothesis posits that in an independent non-canonical syntactic structure more is semantically encoded than what is said. This ‘more’ is retrieved from the linguistic and extra-linguistic context and encoded into the non-canonical structure. Its semantic proposition is reconstructed and complete even though the syntactic structure is not.

Thus, with regard to an independent verbless structure, the semantic ellipsis account would encode a proposition and attribute sentential status, albeit not syntactically but semantically. Syntactic models of the sentence thus do not need to introduce hidden syntactic elements to explain the independent use of incongruous structures: structures that do not meet syntactic requirements are said to be semantically resolved through propositional encoding. The insufficiency of syntactic models is thus made up for by an encoded proposition whose meaning has been reconstructed from the context.

While this solution may at first seem like a reasonable compromise between treating the structures as completely nonsentential and attributing them hidden syntax, several difficulties suggest that it only moves the problem into the semantic domain. The success of this account hinges on the success of propositional encoding. However, it appears that the proposition to be encoded is rarely recoverable from the context. Recovering additional hidden encoded meaning turns out to be no less problematic than additional hidden syntactic structure.

3.6.1 Problem 1: Choosing the propositional meaning within a single context

There are several problems that arise with the semantic ellipsis hypothesis. The first is that semantic reconstruction turns out to be just as ambiguous as syntactic reconstruction. The difficulty of identifying a specific verb, specific inflection, specific subject, and specific positions for the hidden elements in order to recreate from the various possibilities a single syntactic hidden structure, is also true when considering the semantic elements and choosing amongst the possible propositions that the string may encode. In other words, the possible interpretations of (6b) ‘*From France.*’, that were mentioned in (8) and are reproduced below, pose problems not only for the syntactic structure but also for the semantic proposition. The propositional meaning of ‘*From France.*’, may be the same as any of the propositions expressed by the following:

- (19) a. [The letter is] from France.
- b. [The letter came] from France.
- c. [The letter must have been sent] from France.
- d. [The envelope that Bob is sending me this letter in is the only one in his treasured collection] from France.
- e. From France [was written on the envelope].

Notably, all of the above may be used in the same context of taking a letter out of a mailbox. As a result, each has an equal chance of constituting the encoded semantic proposition and the choice cannot be resolved by the context.

Similarly, a stable context does not lead to felicitous resolution of the predicate and proposition to be encoded in the following seemingly predictable structure. Illustrated in (20) is an independent verbless structure and a context in which the resolution of the meaning of the predicate appears to be salient and likely to be considered as relatively easy to reconstruct.

(20) {Bob has just walked out of the room. }

Он в	больницу.
<i>on v</i>	<i>bol'nicu</i>
<i>he to</i>	<i>hospital.ACC</i>

The context is someone who has just walked out of the room; the subject and the destination are identified in the structure. A possible analysis of this structure is that a motion verb in the present tense completes its propositional meaning. The reason for its omission

would be similar to the reason for replacing ‘Bob’ with the indexical ‘he’: salience and predictability, i.e. why *repeat* something that is already crystal clear for the *hearer*. Instead of an explicit pronoun and its antecedent ‘Bob’, it is absence that would serve as the anaphor for the contextually salient meaning of the predicate antecedent. A predicate and propositional meaning would thus be semantically encoded into the structure. Such is the reasoning that leads to the semantic ellipsis account that more is semantically encoded than that which is syntactically expressed.

However, recovering the meaning of the verb and the proposition, even when it appears to be contextually salient, is a problem for encoding. Example (20) is felicitous with a range of predicates which, though they have been narrowed down by the context, are not fully resolved within this range. Example (21) shows the minimum range of potential meanings of the predicate and proposition that would compete for a semantic ellipsis reconstruction of (20) uttered in the context of Bob walking out of the room.

(21) *The potential meaning to be encoded into (20) in the context of Bob walking out of the room may correspond to that expressed by:*

a. ‘walk’/ ‘идти’, ‘уйти’, ‘пойти’

i. ‘идти’ (*idti*; lit. ‘walk.IPFV’)

Он	идет	в	больницу.
<i>on</i>	<i>idet</i>	<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>walking.PR.IPFV</i>	<i>to</i>	<i>hospital</i>

ii. ‘уйти’ (*ujti*; lit. ‘walk_away.PFV’)

Он	ушел	в	больницу.
<i>on</i>	<i>ušel</i>	<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>walked_away.PS.PFV</i>	<i>to</i>	<i>hospital</i>

iii. ‘пойти’ (*pojti*; lit. ‘walk.PFV’)

Он	пойдет	в	больницу.
<i>on</i>	<i>pojdet</i>	<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>will_walk.FT.PFV</i>	<i>to</i>	<i>hospital</i>

b. 'run' / 'бежать', 'убежать', 'побежать'

i. 'бежать' (*bežat*); lit. 'run.IPFV')

Он	бежит	в	больницу.
<i>on</i>	<i>bežit</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>running.PR.IPFV</i>	<i>to</i>	<i>hospital</i>

ii. 'убежать' (*ubežat*); lit. 'run_off.PFV')

Он	убежал в больницу.		
<i>on</i>	<i>ubežal</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>ran_off.PS.PFV</i>	<i>to</i>	<i>hospital</i>

iii. 'побежать' (*pobežat*); lit. 'run.PFV')

Он	побежит в больницу.		
<i>on</i>	<i>pobežit</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>will_run.FT.PFV</i>	<i>to</i>	<i>hospital</i>

c. 'drive' / 'ехать', 'уехать', 'поехать'

i. 'ехать' (*exat*); lit. 'drive.IPFV')

Он	едет	в	больницу.
<i>on</i>	<i>edet</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>driving.PR.IPFV</i>	<i>to</i>	<i>hospital</i>

ii. 'уехать' (*exat*); lit. 'drive_off.PFV')

Он	уехал	в	больницу.
<i>on</i>	<i>uechal</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>drove_off.PS.PFV</i>	<i>to</i>	<i>hospital</i>

iii. 'поехать' (*poexat*); lit. 'drive.PFV')

Он	поедет	в	больницу.
<i>on</i>	<i>poedet</i>	<i>v</i>	<i>bol'nicu</i>
<i>he</i>	<i>will_drive.FT.PFV</i>	<i>to</i>	<i>hospital</i>

d. 'plan to drive' / 'собираться ехать', 'собраться ехать'

i. 'собираться ехать' (*sobirat'sja exat*); lit. 'plan.IPFV drive.IPFV')

Он	собирается	ехать	в больницу.
<i>on</i>	<i>sobiraetsja</i>	<i>exat'</i>	<i>v bol'nicu</i>
<i>he</i>	<i>planning.PR.IPFV</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

ii. ‘собраться ехать’ (*sobrat’sja exat’*; lit. ‘*plan.PFV drive.IPFV*’)

Он	собрался	ехать	в больницу.
<i>on</i>	<i>sobral’sja</i>	<i>exat’</i>	<i>v bol’nicu</i>
<i>he</i>	<i>planned.PS.PFV</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

iii. ‘собирается ехать’ (*sobirat’sja exat’*; lit. ‘*plan.IPFV drive.IPFV*’)

Он	будет	собирается	ехать	в больницу.
<i>on</i>	<i>budet</i>	<i>sobirat’sja</i>	<i>exat’</i>	<i>v bol’nicu</i>
<i>he</i>	<i>be.FT</i>	<i>plan.IPFV.INF</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

e. ‘*plan to drive fast*’ / ‘собирается быстро ехать’, ‘собраться быстро ехать’

i. ‘собирается быстро ехать’ (*sobirat’sja bystro exat’*; lit. ‘*plan.IPFV fast drive.IPFV*’)

Он	собирается	быстро	ехать	в больницу.
<i>on</i>	<i>sobiraetsja</i>	<i>bystro</i>	<i>exat’</i>	<i>v bol’nicu</i>
<i>he</i>	<i>planning.PR.IPFV</i>	<i>fast</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

ii. ‘собраться быстро ехать’ (*sobrat’sja bystro exat’*; lit. ‘*plan.PFV fast drive.IPFV*’)

Он	собрался	быстро	ехать	в больницу.
<i>on</i>	<i>sobral’sja</i>	<i>bystro</i>	<i>exat’</i>	<i>v bol’nicu</i>
<i>he</i>	<i>planned.PS.PFV</i>	<i>fast</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

iii. ‘собирается быстро ехать’ (*sobirat’sja bystro exat’*; lit. ‘*plan.IPFV fast drive.IPFV*’)

Он	будет	собирается	быстро	ехать	в больницу.
<i>on</i>	<i>budet</i>	<i>sobirat’sja</i>	<i>bystro</i>	<i>exat’</i>	<i>v bol’nicu</i>
<i>he</i>	<i>be.FT</i>	<i>plan.IPFV.INF</i>	<i>fast</i>	<i>drive.IPFV.INF</i>	<i>to hospital</i>

f. *verb final*

i. Он в больницу идет.
on v bol’nicu idet
he to hospital walking.PR.IPFV

ii. Он в больницу ушел.
on v bol’nicu ušel
he to hospital walked_away.PS.PFV

iii. Он в больницу пойдет.
on v bol’nicu pojdet
he to hospital will_walk.FT.PFV

g. *verb initial*

i.	Идет <i>idet</i> <i>walking.PR.IPFV</i>	он <i>on</i> <i>he</i>	в <i>v</i> <i>to</i>	больницу. <i>bol'nicu</i> <i>hospital</i>
ii.	Ушел <i>ušel</i> <i>walked_away.PS.PFV</i>	он <i>on</i> <i>he</i>	в <i>v</i> <i>to</i>	больницу. <i>bol'nicu</i> <i>hospital</i>
iii.	Пойдет <i>pojdet</i> <i>will_walk.FT.PFV</i>	он <i>on</i> <i>he</i>	в <i>v</i> <i>to</i>	больницу. <i>bol'nicu</i> <i>hospital</i>

The propositional meaning of (20), uttered in the context of Bob leaving the room, may be the same as that expressed by, at minimum, any of the above structures. This includes differences not only in the *type of motion* (e.g. (a) ‘walk’ vs. (b) ‘run’ vs. (c) ‘drive’), but also in whether it is *realized or potential* (e.g. (c) ‘drive’ vs. (d) ‘plan to drive’), as well as the *manner* (e.g. (e) ‘plan to drive fast’). Differences in *tense and aspect* are also not resolved by the context (e.g. (i) present tense imperfective aspect vs. (ii) past tense perfective aspect vs. (iii) future tense perfective aspect). Furthermore, word order flexibility multiplies the potential structures, and, assuming that a difference in form leads to a difference in meaning, the propositional meanings of the possible *alternative word orders* (e.g. (f) verb final and (g) verb initial) should also be considered as competing for a reconstructed propositional meaning of the verbless structure uttered in the context of (20).

All of these show that it is far from clear which predicate should be re-established and which one of the potential propositional meanings is to be encoded. Allowing the encoding of more than that which is expressed opens a Pandora’s box of competing propositional meanings even within a single and very clear context. Schematically this problem may be illustrated as in *Figure 4*.

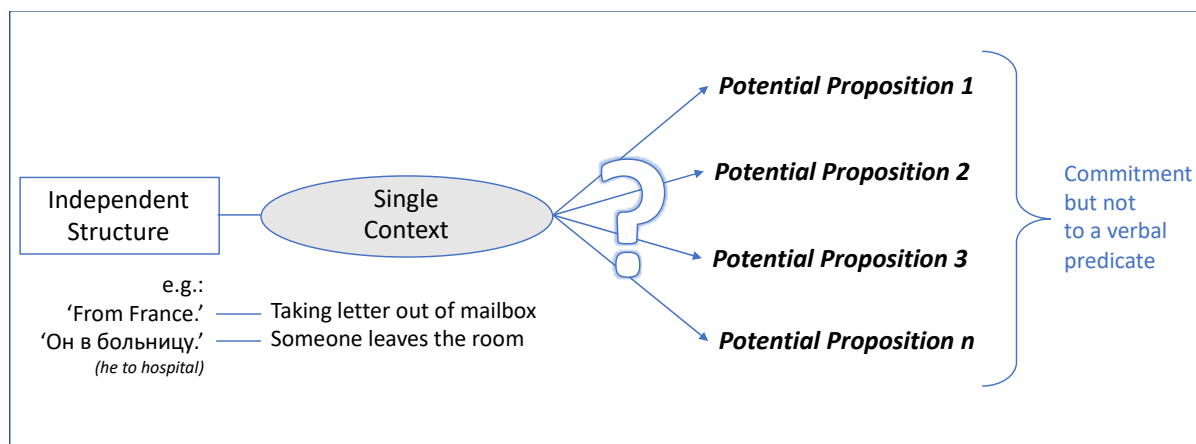


Figure 4. Multiple potential propositional meanings in single context

Notably, the competition illustrated in the above examples, highlights that two key issues – (a) ease of interpretation by the hearer and (b) speaker commitment – work against the hypothesis of semantic encoding.

Ease of interpretation: While it is clear that the meaning of (20) and (6b) do not pose interpretation problems for the interlocutors, the potential propositions in (21) and (19) show that it is a rush to say that this is due to encoding. The propositional meaning – a cognitive notion – is not necessarily semantically encoded; removing the assumption that it is, would remove the necessity to select a proposition with more meaning than is actually expressed with a verbless structure. The ease of use and interpretation of the structures suggest that speaker relies on the hearer to be able to arrive at a propositional meaning, but semantic encoding is not the only way that propositional meaning may arise for the hearer.

Speaker commitment: Furthermore, these examples also make it possible to note the limits of the commitment that is made by the speaker. Having asserted (20), it seems unreasonable to commit the speaker to the truth of all of the potential meanings in (21). Yet, committing the speaker to only one, for instance (21a – i), i.e. that Bob is walking to the hospital, is not possible without ruling out the other potential propositions: if it turns out that Bob drove, it will not make the speaker's utterance of (20) false. The alternative possible propositional meanings show that the extent of the speaker commitment stops at the explicit elements, i.e. 'he' and 'to hospital', plus one key non-encoded pragmatic feature: a relation of *relevance* between the explicit elements – a pragmatic relation that arises from the fact that they are found as part of an independent syntactic structure in a particular context.

As will be proposed in *Part I: Chapter 4*, both (a) the relevance of the explicit elements to one another (i.e. the predicative relation), and (b) the propositional meaning of the structure, appears to be a matter of pragmatic implicature, not semantic encoding.

3.6.2 Problem 2: Multiple propositional meanings in different contexts

Another argument against semantic ellipsis is provided by Stainton (1995). The problem concerns the various forms of the semantic propositions that arise when using the same structure in various contexts. He argues that the various contexts of use should not change the encoded semantic proposition, and shows that they clearly do when the structure at issue is a non-canonical one.

The one-word utterance in (22) is used by Stainton (1995) to illustrate the problem.

(22) Red.

The word ‘*red*’ is typically said to have the meaning of a property, but treated as a semantic ellipsis, when found in use as an independent unembedded structure, it would be said to encode the meaning of a proposition. At minimum, the following four contexts could be felicitous with its independent use; as indicated, each one leads to a different semantic type identified by Stainton (1995: 290–294).

(23) Potential contexts for the independent structure in (22).

a. *Context A:*

When asked to identify a paint sample, the speaker utters ‘*red*’.

- The meaning of the one-word structure would be that ‘*the contextually specified item is red*’.
- The semantic type of the proposition to be encoded would have the *argument-predicate form* (i.e. the contextually specified item is the argument; the property ‘red’ is the predicate that is attributed to it).

b. *Context B:*

When talking about favorite things, the speaker utters ‘*red*’.

- The meaning in this case would be that ‘*red is the speaker’s favorite colour*’.
- The semantic type of the proposition to be encoded would have the form of an *identity between properties* (i.e. the property ‘red’ is identified with the contextually specified property ‘favorite colour’).

c. *Context C:*

When looking at a series of paintings each of which is a different shade of red, the speaker utters ‘*red*’.

- The meaning expressed would be that ‘*every painting is red*’.
- The semantic type of the proposition to be encoded would have the form of a *universal quantification* (i.e. the property ‘red’ is identified as being numerically identical to all instances of the contextually specified property ‘painting’).

d. *Context D:*

An interior decorator walks into an unpainted room and utters ‘*red*’.

- The meaning would be that ‘*red is the colour that should be used to paint the room*’.
- The semantic type of the proposition to be encoded would have an *argument-predicate form*, but compared to the argument-predicate form of Context A, the word ‘*red*’ in Context D no longer said to express the predicate (i.e. the property ‘red’ would now be the argument and the predicate would be the contextually specified need to paint the room in some colour).

Stanton argues that such contextual variance of the propositional form indicates that there is no one particular semantic type that is encoded into the string:

propositional form – i.e., the kind of proposition exhibited by an expression – is not the sort of thing that is context dependent [...] a univocal expression must have single semantic type. [...] expression E cannot, in context A, express a proposition of (say) argument-predicate form, and, in context B, express a proposition of quantificational form – even if the expression E does contain indexicals. (Stanton, 1995: 291)

As a result, the semantic ellipsis hypothesis that a structure, despite non-canonical syntax, may have the meaning of a full semantic proposition, is implausible when considered in terms of semantic encoding. The appeal to an encoded semantic ellipsis requires an encoded semantic

proposition but the variance of this latter across contexts rules out such encoding. A propositional meaning thus appears outside of the possible *semantic encoding* of non-canonical structures.

The importance of the problem of multiple contexts, and their consequence on potential propositional meaning, is emphasized in further analysis of the Russian structure in (24).

(24) Он в больницу.
on v bol'nicu
he to hospital.ACC

In (20) and (21), this structure was analyzed for the multiple potential propositional meanings arising from the perspective of a single context (i.e. someone leaving the room). The meaning of the structure in this type of context can be referred to in general as that of *physical motion of the agent toward a destination*. However, it is notable that the same independent verbless structure is also compatible with contexts of different types, with each type leading to a range of different potential propositional meanings. The structure in (24) is, as a minimum, compatible with the contexts in (25), each of which would lead to a different range of potential propositional meanings that would multiply the competition for encoding.

(25) Potential contexts for the independent structure in (24).

a. *Context A:*

Speaker observes Bob moving toward the hospital and utters ‘он в больницу’ (*on v bol'nicu*; lit. ‘*he to hospital*’).

Meaning: Physical motion of agent toward destination.

Similar to the meaning expressed by:

Он идет в больницу.
on idet v bol'nicu
he walking.PR.IPFV to hospital.ACC

b. *Context B:*

Speaker utters ‘он в больницу’ (*on v bol’nicu*; lit. ‘*he to hospital*’) in order to explain why Bob is speaking very loudly to someone on the telephone. The hearer knows that the telephone is not working very well and that Bob has been trying to call the hospital all day.

Meaning: Communication of agent with interlocutor.

Similar to the meaning expressed by:

Он	дозвонился		в	больницу.
<i>on</i>	<i>dozvonilsja</i>		<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>telephoned_through.PS.PFV</i>		<i>to</i>	<i>hospital.ACC</i>

c. *Context C:*

Speaker assigns plans of military crimes and utters ‘он в больницу’ (*on v bol’nicu*; lit. ‘*he to hospital*’).

Meaning: Violence of agent toward an illegal target.

Similar to the meaning expressed by:

Он	будет	стрелять	в	больницу.
<i>on</i>	<i>budet</i>	<i>streljat’</i>	<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>be.FT</i>	<i>shoot.IPFV.INF</i>	<i>to</i>	<i>hospital.ACC</i>

d. *Context D:*

Speaker sees a child playing with toy stethoscopes and describes the situation as ‘он в больницу’ (*on v bol’nicu*; lit. ‘*he to hospital*’).

Meaning: Describing the contextual items (i.e. toys) as an abstraction that must be occurring in the mind of the agent.

Similar to the meaning expressed by:

Он	играет		в	больницу.
<i>on</i>	<i>igraet</i>		<i>v</i>	<i>bol’nicu</i>
<i>he</i>	<i>playing.PR.IPFV</i>		<i>to</i>	<i>hospital.ACC</i>

In addition to the potential meanings competing for encoding when (24), lit. ‘*he to hospital*’, is used in the context of (a) to indicate physical motion of an explicitly specified agent toward an explicitly specified destination (the breadth of the competition is illustrated in (21)), the potential meanings arising from the contexts in (25), which can each be expanded with a range similar to (21), further complicate the problem of reconstruction. Example (25) shows that the

same structure could be felicitously used in contexts where it would take on a completely different range of meanings. In a context such as (b), the structure would be used to reference the communication of the agent, not physical motion; furthermore, the object ‘*hospital*’ is used in this context to reference an interlocutor at the location, not the location itself. The structure may also be used in contexts of violence, such as (c), where the hospital becomes a criminal target. In a context such as (d), the object ‘*hospital*’ is used as an abstract notion (i.e. a game), not as any physical destination.

Thus, the potential meaning to be encoded becomes even more difficult to determine when the structure may be used in multiple contexts with different meanings. This problem can be illustrated schematically as in *Figure 5*.

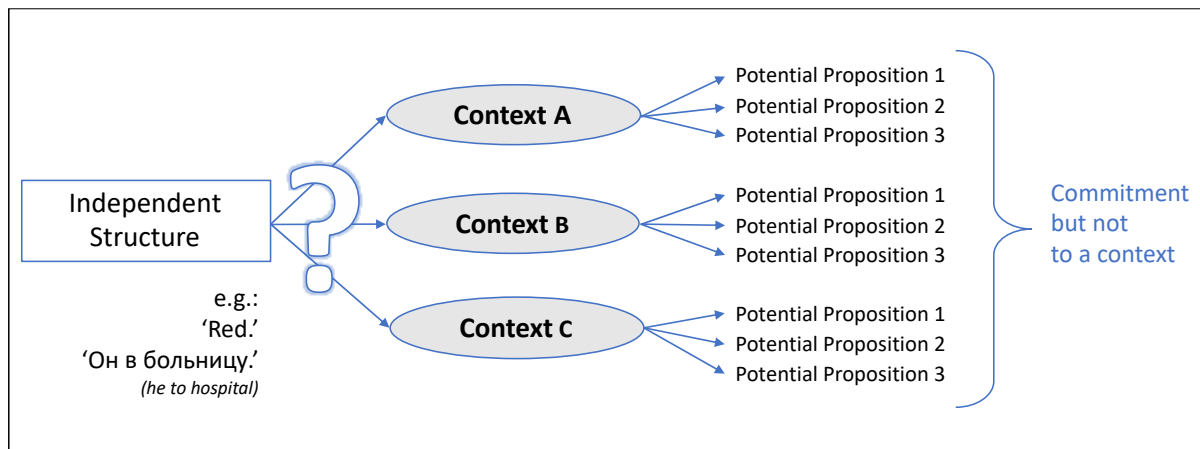


Figure 5. Multiple propositional meanings in different contexts

As shown on the diagram, the problem of multiple contexts further multiplies the competition for propositional meanings if the issue is to be considered in terms of encoding. If more is encoded than what is said, the question of what particularly the speaker is committing to in uttering the structure must now include not only the meaning of the predicate, and the propositional meaning (that were put into question in Problem 1), but also the general context of use. The change of propositional meaning according on the context appears to contradict the idea of its encoding into an independent structure.

3.6.3 Problem 3: Reintroduction changes the meaning

Even if a single meaning could be determined within a context (i.e. setting aside for a minute Problem 1), and if the propositional form were to remain stable across the various contexts into which the structure may be inserted (i.e. neutralizing Problem 2), there is yet another problem that would prevent treating the verbless sentence as semantically encoding a canonical proposition. This problem is that reconstruction of the proposition has been shown to change the meaning of that very same proposition.

Existing semantic analyses show that making explicit a salient predicate leads to changes in the meaning of the structure. Semantic differences have been explored with regard to the Russian structure ‘У *x-a y*’ (*u x-a y*; lit. ‘at *x*.GEN *y*.NOM’) when it is treated within the copula paradigm as a possessive structure which alternates the expression and non-expression of the present tense copula verb ‘БЫТЬ’ (*byt*’; lit. ‘*be*’). The semantic differences from explicitly instantiating the predicate include changes in referential value, as well as changes in situational links. These changes have been noted in detailed semantic analyses of the structure in (Selivërstova, 1973; Paillard, 1984). The following famous examples from Selivërstova (1973) illustrate these two types of changes.

Differences in *referential value* from re-establishing the predicate are made obvious in (26), which compares ‘У него седые волосы’ (*u nego sedye volosy*; lit. ‘at *him gray hair*’) with the verbal counterpart using the present tense ‘быть’ (*byt*’; lit. ‘*be*’).

- (26) a. У него седые волосы.
u nego sedye volosy
at him.GEN gray hair
‘*He has gray hair.*’
- b. У него есть седые волосы.
u nego est’ sedye volosy
at him.GEN is gray hair
‘*He has some gray hair.*’

The verbless sentence in (a) would typically be used to attribute the property ‘gray’ to the entire head of hair (i.e. to identify the colour). However, this reading is not available for the verbal alternative. The verbal structure (b) can only be used to assert the existence of a limited quantity of gray hair. This means that re-establishing the predicted predicate would change the meaning of the verbless structure.

This change in meaning is a problem for the semantic ellipsis hypothesis. The above shows that the verbless structure (a), which is supposed to encode the meaning of its verbal counterpart (b), would be encoding a meaning that is not actually available to (b). Furthermore, in a context of surprise, the verbless structure in (a) allows the meaning of (b), e.g. the speaker may utter (a) upon noticing the existence of a single gray hair on the head of someone who is not supposed to have any. Considered in terms of encoding, this would mean that (a) can semantically encode more than (b).

It must be noted that the absence or presence of the verb, with regard to the same ‘У x-a y’ (*u x-a y*; lit. ‘at x.GEN y.NOM’) structure, does not always entail the same difference in meaning. In (26), the variation concerns referential value, however, other differences have also been noted that do not concern referential value.

Making explicit a salient predicate has also been shown to lead to semantic differences in terms of *situational links*. This factor is illustrated in (27) through the comparison of the verbless ‘У меня дети’ (*u menja deti*; lit. ‘at me children’) with its verbal counterpart using the present tense ‘быть’ (*byt*’; lit. ‘be’).

(27) a. Ана: Бежим.
bežim
run.IMV
 ‘Let’s run.’

Bob:	Я	не	могу	бежать.	У	меня	дети.
	<i>ja</i>	<i>ne</i>	<i>možu</i>	<i>bežat’</i>	<i>u</i>	<i>menja</i>	<i>deti</i>
	<i>I</i>	<i>not</i>	<i>can</i>	<i>run</i>	<i>at</i>	<i>me.GEN</i>	<i>children</i>
	<i>‘I can’t run.’</i>				<i>‘I have children.’</i>		

b. Ана: Бежим.
bežim
run.IMV
 ‘Let’s run.’

Bob:	Я	не	могу	бежать.	У	меня	есть	дети.
	<i>ja</i>	<i>ne</i>	<i>možu</i>	<i>bežat’</i>	<i>u</i>	<i>menja</i>	<i>est’</i>	<i>deti</i>
	<i>I</i>	<i>not</i>	<i>can</i>	<i>run</i>	<i>at</i>	<i>me.GEN</i>	<i>be.PR</i>	<i>children</i>
	<i>‘I can’t run.’</i>				<i>‘I have children.’</i>			

The verbless sentence in (a) is used to make a link between the existence of children and the role that they have in the ongoing situation: the children are presented as an explanation, in this case an obstacle that prevents Bob from performing the action that is requested of him. Notably, the children in (a) do not necessarily belong to the speaker; they are linked to the situation.

The verbal structure in (b) blocks such a situational reading. The structure expressing the predicate can only be used to establish the existence of children belonging to the speaker. Notably, it links the existence of the children to the speaker, not to the situation. However, in the context above, where the speaker is unable to run, it is not the link of the children with the speaker that poses a problem, but the attention that they require or something else that is associated to their existence in the situation; a situational reading is thus not available to (b). The verbal sentence in (b) does not present the children’s existence as an explanation, but simply asserts their existence.

These examples further highlight that the difficulty of recovering the predicate exists not only syntactically, but also semantically. The use of the verbal predicate in (b) gives the structure a different meaning from its verbless counterpart in (a). As a result, the semantic ellipsis hypothesis according to which a verbless structure encodes the propositional meaning of a verbal structure is clearly in trouble: the propositional meaning available to verbless (a) – i.e. the situational obstacle reading – is not available to the verbal structure (b).

These changes in meaning further emphasize the problems with the idea of encoding a semantically ellipsed element into a verbless structure. Schematically, this third problem may be illustrated as in *Figure 6*.

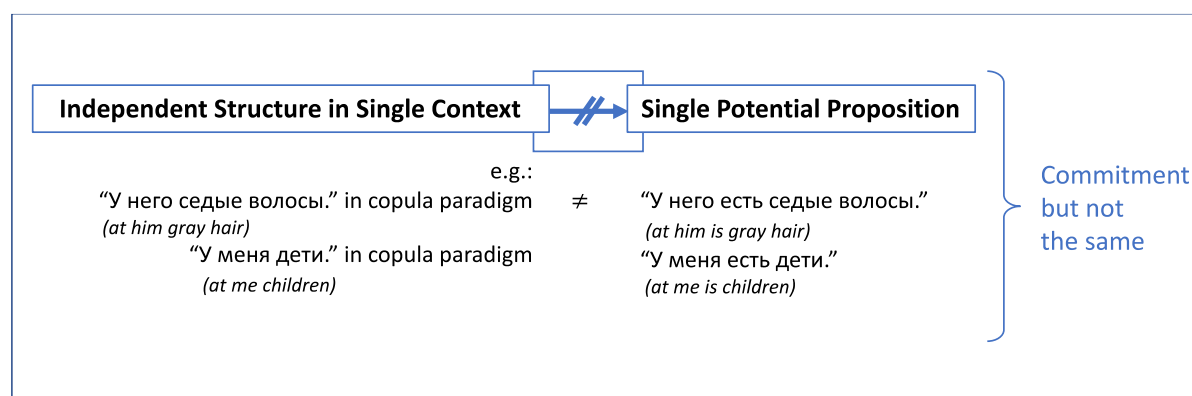


Figure 6. Reintroduction of the predicate changes meaning

As shown on the diagram, the link between the independent structure and the potential propositional meaning of a verbal counterpart, even when the latter concerns a single potential proposition and a single context, is broken. The propositional meaning implied by the verbless structure is not the same as the propositional meaning implied by its verbal counterpart.

Considered within the same context, the speaker would commit to different propositions depending on whether or not the verb is made explicit. In (a), it is the children's link to the situation that is presented as a problem and a quality that is attributed to hair, whereas in (b) it is the existence of children belonging to the speaker and some limited quantity of gray hair that is being asserted.

3.6.4 Problem 4: Inconvenience of encoded meanings

The problems above make it difficult to see how a propositional meaning would be *semantically encoded* into an independent structure that is not already explicitly of the particular propositional form:

1. What specifically would that proposition look like? (i.e. *Problem 1*)
2. How can it be encoded and yet change its encoding with the context? (i.e. *Problem 2*)
3. If the propositional meaning is encoded how can that encoded meaning change when the implicit encoding is made explicit? (i.e. *Problem 3*)

There is an additional issue. Even if the above problems are set aside and it is imagined that such propositional encoding were possible, it would not be a very convenient solution in terms of the cognitive load that such encoding implies. It would mean that each word, each phrase, each structure would have encoded into it not only the meaning that it is already typically said to have encoded into it, but also all of the possible propositional meanings that it could have in all of the contexts in which it can be used.

For instance, dictionary entries for the word '*red*' include: a colour, a first name, and a political direction. To what extent even these non-propositional meanings are actually *semantically encoded* into the word, as opposed to associated with the word by some other means, e.g. potentially through insertion of the word into a context that implies a certain meaning which comes to be habitually attributed to the word with a strength that is proportional to the salience of the word's use, is also a source of debate. If the proposed propositional

encoding was also possible, in addition to the non-propositional definitions, the word ‘*red*’ would require dictionary entries for all of the possibilities of its occurrence as an independent unembedded structure, such as the propositions associated with the contexts in (22) above, as well as other potential uses. Propositional encoding for non-canonical structures, would thus result in a significant augmentation of the meanings encoded in each word, each phrase and each possible structure.

This issue was raised by Stainton (1995) who makes the following comments concerning the use of propositional semantic encoding to justify sentential status for independent non-canonical structures, i.e. the semantic ellipsis hypothesis:

How many semantically elliptical sentences would there be? [... A] minor point, but one worth making – if the semantic ellipsis hypothesis were true, there would be a *very large* class of one-word and one-phrase sentences, in addition to the infinitely large class of syntactic sentences and the infinitely large class of ordinary words and phrases. [...] It may be easy enough to suppose that there are a scattered few one-word and one-phrase sentences, just like idioms and such. Indeed, if there were just a few, one could give their meaning by providing a short list. But if semantically elliptical sentences are to do the work demanded of them, there cannot be just a few of them: if the proponent of the semantic ellipsis hypothesis is to handle *all* possible assertoric utterances of (apparent) words or phrases, then he must postulate a very large class of extra formatives. (Stainton, 1995: 292–293)

It is clear that, even *if* a particular proposition could be reconstructed (i.e. all three problems above successfully passed), the claim that more is semantically encoded than what is expressed would need to be used extremely sparingly.

3.6.5 Problem 5: The proposition, its encoding, and sentential status

The final problem concerns the way that the arguments against elliptical semantic encoding of the proposition in non-canonical structures are used. As the above quote illustrates, Stainton takes the line of reasoning that the absence of propositional encoding in non-canonical structures indicates nonsentential status. However, abandoning *sentential status* is not a necessary consequence of accepting the arguments against *propositional encoding*.

Question (i), *whether or not semantic encoding of a proposition of the type expressed by a typical canonical syntactic structure is necessary for sentential status*, is a separate question from (ii) *whether or not such a proposition is semantically encoded*.

It appears that question (ii) must receive a negative answer. The above problems have suggested that semantic encoding does indeed require explicitness: to say that more is encoded than that which is made explicit is problematic from the point of view of determining what that more is. The claim that a proposition *that is encoded* into a syntactic structure which consists of a subject, a predicate, and inflection, is the same as the proposition *encoded* into a structure that does not have these elements, appears to defeat the purpose of encoding. It may be the same proposition behind both structures, but it is not equally encoded.

However, this does not necessitate non-sentential status. First, a proposition may still exist. A proposition is an abstract cognitive notion and, as such, it may exist without being linguistically encoded. For instance, it may remain an unexpressed mental image.²⁹ Secondly, this proposition may be related to a structure in a different way than encoding. For instance, it may be pragmatically implicated, by way of its combination with the context, rather than semantically encoded into the expressed elements of the structure.

The next chapter (*Part 1: Chapter 4*) will explore how to reconcile the existence of a potential propositional meaning (or several potentials) with verbless structures. (The solution proposed will involve pragmatic implicature and a re-consideration of the essential elements of sentential status in terms of embedding and information structure). Presently, it is important to emphasize that the question of encoding and that of sentential status are separate. Although independent verbless structures do not *semantically encode* either the meaning of a verb, or the semantic categories associated with the verb (i.e. tense and aspect), or other potential meaning that goes beyond that which the structure makes explicit – this does not mean that such meanings cannot arise from the use of a verbless structure. What the present section aimed to show is that it is the *encoding* that is problematic – notably, it is problematic whether from the *compositional* perspective, i.e. the combination of the explicit words of the structure does not appear to encode a predicate, but, it is similarly problematic if the encoding is said to be *non-compositionally* derived, i.e. as a single encoded propositional meaning that expresses more than the combination of the explicit words: both accounts face all of the challenges illustrated above with regard to *encoding* a recovered potential proposition. The propositional meaning associated with a canonical verbal syntactic structure appears outside of the possible semantic

²⁹ A discussion of propositions and their relation to mental images can be found in John Perry's (1986) analysis of thoughts that do not have any linguistic representation (e.g. '*It is raining*' which assumes a place that is not articulated), as well as in Elugardo and Stainton (2003), Stainton (2004: 282–285), Elugardo and Stainton (2005: 13, 17)'s references to a 'mentalese' language and mental representations of sentences which are "not in any natural language" (Stainton, 2004: 285).

encoding of non-canonical structures, but this does not make non-sentential status a necessary consequence. Instead, it throws out a challenge to look for other ways that *un*-encoded propositional meaning may arise.

3.7 CONSTRUCTIONIST APPROACHES

As shown in the previous sections, independent verbless structures shake not only the traditional conceptions of syntax but also raise issues concerning the source of meaning in a sentence. Hidden elements have been used as an argument to justify their sentential status not only in the syntactic domain, but also in the semantic domain. Though the quest to justify their sentential status must be maintained in order to account for actual language use, there are serious reasons to doubt the success of positing either a hidden syntax or a hidden, compositional or non-compositional, semantic encoding.

The traditional view of sentence meaning, i.e. that verbs project an argument structure and thereby specify the meaning of a sentence, is unable to account for the meaning of an independent verbless structure: the absence of a verb, given the inability to reconstruct it syntactically or semantically, pushes such structures to the periphery where, at best, they may be analyzed as abnormal exceptions to the way that language actually works.

Recently, an alternative conception of language meaning has been established through the Construction Grammar (CxG) linguistic framework (Goldberg, 1995; Goldberg, 2003). Construction grammar approaches treat language as a set of surface forms (of various levels, including morphemes, words, phrases, abstract phrasal patterns, and any combination thereof), that are each paired with a particular function (an identifiable meaning which arises through frequent use, i.e. ‘conventionalization’, of the particular form with the particular interpretation); these form-function pairings are cognitively stored, and freely combine into larger units of language as long as they do not conflict with one another. The word ‘construction’ is used to refer to any linguistic pattern in which form and function have been paired through frequency use. For instance, Adele Goldberg and Devin Casenhiser (2006: 348–349) identify a basic ordinary expression, such as (28) ‘*What did Chris buy her mother?*’, as a ‘construct’ that involves the ‘constructions’ in (a) through (f):

(28) **Construct:** *What did Chris buy her mother?*

Constructions invoked:

- | | |
|---------------------------------|--|
| a. Words: | <i>Chris, buy, her, mother, what, did</i> |
| b. Ditransitive: | <i>Chris buy her mother + what</i> |
| c. Interrogative: | <i>initial wh-word + subject-auxiliary construction + 'missing' argument</i> |
| d. Subject-Auxiliary Inversion: | <i>did Chris</i> |
| e. Verb Phrase: | <i>buy her mother</i> |
| f. Noun Phrases: | <i>What, Chris, her mother</i> |

Constructions make it possible to account not only for basic patterns of language that involve compositional meaning (i.e. through the combination of lexical items, each with a unique form-function pairing) but also for patterns of language that involve more problematic non-compositional meaning, such as idioms (i.e. through the ability of the 'form' in the form-function pairing to involve a level beyond the individual words). Furthermore, if a regular pattern of language, such as '*What did Chris buy her mother?*' which has been analyzed as a construct that invokes constructions in (28), became a frequent expression with its own definable and generalizable function (i.e. a 'conventional pairing of form and function'), say, for instance, through a heavy marketing campaign it became regularly used as an informal way to say that 'buying a gift is better than making one', it would itself become a 'construction':

[W]hile most linguists agree that constructions are required for unusual patterns, constructionists invoke constructions for the basic, regular patterns of language as well. [...] any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, many researchers observe that there exists linguistic and psycholinguistic evidence that patterns are stored even if they are fully predictable as long as they occur with sufficient frequency [...] Thus, these highly frequent expressions, even if fully compositional, are sometimes labeled "constructions" as well. (Goldberg & Casenhiser, 2006: 348–349)

The difference between a construct and a construction (Cx) is thus established based on conventionalized meaning (i.e. frequency of the form-function correspondence). This may be illustrated schematically as in (29).

- (29) CONSTRUCT = [C_{X1} + C_{X2} + C_{XN}] + Not conventionalized
(e.g. ‘What did Chris buy her mother?’ without conventionalized meaning)
- CONSTRUCTION = [C_{X1} + C_{X2} + C_{XN}] + Conventionalized through frequency
(e.g. ‘What did Chris buy her mother?’ if regularly used as an informal way to express the importance of buying a gift)

3.7.1 Advantages for verbless constructions

From the perspective of independent verbless structures, the constructionist approach represents several major advantages compared to the traditional account.

First, a construction is not in any way a verb-centric notion: a structure with or without a verb may constitute a construction. Thus, to start with, the constructionist approaches do not ignore noncanonical structures. This aim and asset of the approach is powerfully summarized in the following:

Most generative grammarians thus conclude, with Chomsky that “the search for explanatory adequacy requires that language structure must be invariant, except at the margins” (Chomsky, 2000). These researchers accordingly attempt to predict the properties of patterns such as [*‘Sam slept the whole trip away’*; *‘Him, a trapeze artist?!’* ...] on the basis of general, universal principles. If such attempts are unsuccessful in this endeavor, the pattern is relegated to the “periphery” or “residue” of language. As such, it is determined to be an uninteresting bit of a language that is not subject to the same cognitive principles at work in the ‘core’ grammar of a language. [...] a theory is only explanatorily adequate if we can ultimately account for how languages can be learned from the initial state on the basis of the input. The approaches differ, however, both in what each theory believes it is necessary to account for, and in each theory’s view of the richness of the initial state. As mentioned, generative linguists often relegate constructions such as the *incredulity construction* [e.g. *‘Him, a trapeze artist?!’*] to the periphery of the theory. As such, they have no reason to account for the way in which they are learned. Moreover, more prolific constructions such as the *passive* are considered to exist in many languages and as such may be universal and part of the genetic language component. Constructionists hold neither of these views and therefore believe that a theory of language learning must necessarily account for how all constructions are learned. (Goldberg & Casenhiser, 2006: 345–347)

Notably, constructionist approaches do not require the surface form to be altered in any way by hidden syntactic elements: the form is treated as it is.

The second major advantage is that the context appears to be systematically involved in all form-function pairings, since they are formed through the frequency of use in which the particular form is found with the particular interpretation. In other words, the requirement of conventionalized use in order to identify a construction means that any construction necessarily involves context in its very formation (i.e. in establishing the form-function pattern). This represents another powerful asset from the perspective of independent verbless structures. It makes it possible not only for verbal, but also for verbless structures to have propositional meaning: the propositional meaning of the construction arises through frequency, regardless of whether that construction is verbal or verbless. For instance, the verbless structure ‘*Him, a trapeze artist?!*’ is analyzed as an incredulity construction with full propositional meaning that indicates that the speaker takes a skeptical attitude toward the proposition expressed, i.e. that ‘*he is a trapeze artist*’:

Incredulity construction (e.g., *Him, a trapeze artist?!*). This construction is used to express an attitude towards a proposition, one of incredulity. The speaker in the example above expresses incredulity that the person in question is a trapeze artist. The form of the construction does not obey general rules of English. For one thing, there is no verb and yet the expression stands alone as a full utterance and conveys an entire proposition. In addition, the accusative case marking is normally used for objects, and yet the initial NP would seem to act as a subject or topic argument (cf. *He’s a trapeze artist?!*) (Goldberg & Casenhiser, 2006: 344)

Thus, it appears that the constructionist approach *does not a priori rule out* that verbless conventionalized form-meaning pairings may have a propositional meaning (e.g. in addition to the property meaning of ‘*red*’, the propositional meaning ‘*this object is red*’, is not a priori ruled out from the potential meaning of the lexical surface form ‘*red*’), and the approach even has an account of how such a propositional meaning comes about in verbless *constructions* (i.e. through frequent use).

3.7.2 Disadvantages for verbless constructs

While the constructionist framework presents major advantages for the treatment of verbless *constructions*, it does, however, appear that the approach also faces difficulties when it comes to the problems raised in the previous sections concerning the meaning of independent verbless structures.

A key feature of the constructionist approach is that the ‘argument structure construction’ is treated as *separate* from the ‘verb construction’: the argument structure is said to have *its own semantics* that interacts with the semantics of the verb (Goldberg, 1997; Bencini & Goldberg, 2000). This separation is crucial as it de-centralizes the verb; it means that the argument structure is just as important, it has a conventional meaning of its own which interacts with the conventional meaning of the verb, and together they form the propositional meaning:

argument structure patterns contribute directly to the overall meaning of a sentence, and a division of labour can be posited between the meaning of the construction and the meaning of the verb in a sentence. While the constructional meaning may, perhaps prototypically, be redundant with that of the main verb, the verb and construction may contribute distinct aspects of meaning to the overall interpretation. [...] In many cases, however, the meaning of the construction contributes an aspect of meaning to the overall interpretation that is not evident in the verb in isolation. (Bencini & Goldberg, 2000: 642).

Therefore, keeping in mind the difference between construction and construct above, illustrated in (29), there are two different *ways that a propositional meaning may arise for a structure involving a finite verb*: (a) through the combination of the argument structure construction and the verb construction, and (b) through conventionalization of the combination of the argument structure construction and the verb construction. These two possibilities for the propositional meaning of a finite-verb structure are represented schematically in (30).

- (30) a. *Verbal Construct* = [Argument structure Cx + Verb Cx] + Not conventionalized
- b. *Verbal Construction* = [Argument structure Cx + Verb Cx] + Conventionalized through frequency

When it comes to the *way that propositional meaning arises in a verbless structure*, the options are not the same: its possibilities include only conventionalization. This is represented schematically in (31).

- (31) a. *Verbless Construction* = [Cx] + Conventionalized through frequency
- b. *Verbless Construct* ≠ [Cx] + Not conventionalized

The propositional meaning of a verbless structure would need to be compositional for it to be considered a ‘construct’. However, independent verbless structures, from a compositional

perspective, by definition, do not have access to conventionalization in order to acquire propositional meaning. As a result, a verbless structure whose meaning is not conventionalized is still left without an account of propositional meaning (for instance, an account of how ‘*red*’ may go beyond the property meaning of ‘*red*’ to a propositional meaning without being conventionalized).

The discussion in the previous sections showed that the feat of recovering the particular propositional meaning for an independent verbless structure (which is without doubt interpreted as implying some proposition when it is used independently) is not as evident as it seems. It is of course always possible to paraphrase a specific structure aiming for a maximally finely nuanced explicit propositional meaning, yet the question remains – what is the relation of that paraphrase to the actual structure which has stimulated this paraphrase? Treating the propositional meaning of the paraphrase as semantically encoded into the verbless proposition (whether compositionally or non-compositionally) is a problem: upon closer examination it often becomes obvious that several slightly different propositional meanings are possible within the same context (Problem 1 above), that different contexts exponentially multiply the potential propositional meanings (Problem 2 above), and, what is even more damaging to the idea of reconstruction and recovery of the proposition, the fact of making explicit the propositional meaning that (having successfully passed Problem 1 and Problem 2) is finally attributed to the structure, turns out to transform the meaning of that very structure (Problem 3 above). These problems do exist even for treating verbless structures as constructions (i.e. involving a propositional meaning that is not predictable from the combination of its parts, but is generalized from form-function correspondences); however, while for constructions the attribution of non-compositional meaning can at least be envisaged (i.e. if a conventionalized form-function pairing is found, then that specific verbless structure would carry that particular propositional meaning), a compositional account requires not only a search for a propositional meaning but, also, the attribution of that meaning to specific elements of the structure (i.e. not as a conventionalized function attributed to the entire surface form, but as a compositional meaning derived from the individual parts of the surface structure) which is a task that is difficult to envisage since the structure at issue is precisely one whose meaning goes beyond its explicit surface elements (e.g. it would require decomposing the propositional meaning ‘*this object is red*’ among the lexical elements of the surface form ‘*red*’).

Simply put: there exist serious challenges to the generalization of a form-function correspondence when it comes to the propositional meaning of a verbless sentence, and this challenge is even greater from the perspective of compositional meaning. While independent

verbless structures are interpreted by speakers and hearers as implying some sort of proposition, it seems to be that an individual context is needed for each individual case in order to be able to estimate a *potential* proposition that may be *implicated* by the use of the structure in that context. Conventionalized meaning is possible for *some* verbless structures, just like it is possible for *some* verbal structures (e.g. idioms exist both as finite verb structures and non-finite verb structures); however, when a conventionalized meaning cannot be determined (i.e. the form-function correspondence remains at the individual level of each structure and its context), individual consideration is necessary in order to determine the precise propositional meaning. However, in such case-by-case non-conventionalized instances, it is only the verbal structures whose propositional meaning is accounted for. Unlike verbless constructions, independent verbless structures that *do not* have conventionalized meaning are still left without an explanation as to how it is that their propositional meaning is able to go beyond that expressed by their parts.

To summarize, it appears that on a constructionist approach, the problem of meaning is resolved for a verbless structure when it is a *construction*, i.e. when its meaning arises from systematic correspondence of its form with a particular function as determined through frequency of association that suggests that the same propositional meaning can be generalized to an instance of the particular form. However, if the verbless structure does not constitute a construction, i.e. a consistent form-function correspondence cannot be determined, then the verbless structure is again in trouble even from a constructionist perspective. The propositional meaning of a verbless structure poses a problem for construction grammar accounts from the moment that it does not arise through form-function conventionalized correspondence, i.e. that the verbless structure is not a construction. As a ‘construct’, the verbless structure is either:

- (i) left without a propositional meaning, or
- (ii) attributed a propositional meaning but without an explanation of how it got there (i.e. how its meaning was able to go beyond the combination of its explicit elements when conventionalization is blocked).

Thus, it appears that the constructionist perspective, while providing a powerful explanation of verbless structures in terms of constructions, appears to overlook their potential to constitute constructs.

The discussion of verbless structures within the construction grammar framework typically concerns the search to find conventionalized meaning for the structures. A recent

example is a study by Laura Janda et al. (2020) of the Russian verbless ‘NOM~INS’ construction, lexically instantiated in ‘дурак дураком’ (*durak durakom*; lit. ‘*fool.NOM fool.INS*’; translated as ‘*a fool times two*’). By manually annotating a corpus for semantic types, the NOM~INS form was statistically correlated with three particular meanings (one conveying a maximum degree of the meaning of the main noun, e.g. an extreme fool; the second, a milder ordinary fool; and a third that minimizes the emphasis on the given folly of the character by shifting the discourse away to a different aspect of the person’s character), and the construction is also analyzed in terms of how it interacts with other constructions.

A theoretical account of ellipsis from a construction grammar perspective has been proposed by Adele Goldberg and Florent Perek (2019). Structures that are commonly discussed in the syntactic domain as involving surface level ellipsis of the verb, e.g. gapping, sluicing, stripping, as well as verbless structures that are deemed elliptical in a very loose way of “elliptical in the sense of not providing an overt main verb”, e.g. ‘*Elise, Casey*’, ‘*Down with etiquette*’, ‘*Well, I never*’ (Goldberg & Perek, 2019: 11), are characterized as constituting constructions, i.e. conventional form-function correspondences, that do not depend on a hidden syntactic structure for reconstruction. A psychological pointer mechanism is introduced in order to link a linguistic antecedent and an elliptical construction:

a general pointing function [...] allows some constructions to point to a quite specific overt linguistic string, while others only require that a semantic entity or proposition be evoked (Goldberg & Perek, 2019: 4)

Thanks to this mechanism the antecedent need not be analyzed syntactically; this allows elliptical constructions to be “constructions in their own right” and not “simply shorter variants of full-fledged sentence patterns” (Goldberg & Perek, 2019: 10). The analysis thus spotlights the strong commitment of construction grammar to surface structure. Consequently, from the perspective of sentential status, elliptical constructions appear to be explicitly distanced from ‘full-fledged’ syntactic sentences.

To conclude, it may be noted that even from the perspective of the constructionist account, independent verbless structures still struggle when it comes to sentential status. Construction grammar provides a framework in which any verbless structure may potentially be analyzed as a construction – i.e. a conventionalized pairing of form and function – without, however, neither requiring, nor granting, sentential status. Furthermore, in terms of propositional meaning, it seems that taking a constructionist perspective would present

advantages when it comes to those structures that have been conventionalized; however, the propositional meaning of a verbless structure appears to be very rebellious to recovery.

3.8 SUMMARY

The previous sections have attempted to outline the key syntactic and semantic difficulties that are faced by verbless structures when it comes to their sentential status. Despite these apparent shortcomings, the ultimate motivation for the introduction of hidden syntactic structure, like the motivation for the introduction of a hidden semantic encoding, seems to be a legitimate quest to find some way to adjust sentential models to incongruent data.

The following section is driven by a similar motivation – the desire to explain how it is that in the language that humans use structures that do not have a finite verb are commonly used and understood by interlocutors as full sentences, carrying a propositional meaning and expressing a complete thought. The aim will be to propose a *sentential* definition of verbless structures that strives to account for the above-mentioned concerns. Following the proposal for sentential status, verbless sentences will then be explored in terms of their language-specific semantic characteristics and pragmatic uses in English and Russian as revealed by corpus analysis.

CHAPTER 4

VERBLESS SENTENCES

On what grounds can verbless structures be considered sentences? The present section focuses on the defining elements of sentential status and further delimits the target phenomenon.

It is presently proposed that in order to be considered a ‘**verbless sentence**’, a string of text must meet the three conditions in (32).

(32) NECESSARY AND SUFFICIENT CONDITIONS OF A VERBLESS SENTENCE

A string of text is a verbless sentence, if and only if:

1. It contains initial and final marking.
2. It is found in a context.
3. It does not include a finite verb, nor any other verb form.

To emphasize, all three conditions must hold simultaneously, and if they do, this is sufficient for the structure to be a verbless sentence. Notably, the first two conditions concern sentential status and the third narrows down the verbless structures addressed in the present thesis. They will each be discussed in turn below.

Before doing so, two important terminological points are in order with regard to the notion of a sentence. The first concerns the preservation of the typical finite-verb structure in the notion of a ‘clause’, and the second highlights the compound nature of a sentence.

4.1 PRELIMINARIES

4.1.1 Sentence vs. Clause

Of note is that a ‘sentence’ is a different notion from that of a ‘clause’. The term ‘**clause**’ is set aside for the combination of a syntactic subject, i.e. a noun in a nominative case, and a syntactic finite-verb predicate, what Renaat Declerck et al. (2006) call the ‘prototypical clause’:

A clause, then is a linguistic unit made up of, minimally, a noun phrase and a verb phrase. [...] However, the predicate constituent may contain other elements in addition to the VP. [...] The prototypical sentence is made up of one or more clauses, which means it prototypically contains one or more verb phrases. (Declerck et al., 2006: 13)

In this way, the traditional finite-verb definition is preserved for referring to a canonical syntactic clause.

4.1.2 Sentence vs. Sub-Sentential Utterance Unit

Furthermore, the term ‘sentence’ is an entity that may itself constitute a **compound**. Following Frank Palmer (1974: 11), the ‘sentence’ is used to refer to the matrix of constituents, as opposed to the constituents themselves. Thus, a single sentence, which is marked with initial and final punctuation marking, may consist of several smaller units.

The basic sub-sentence level ‘**utterance**’ unit has been defined in various ways in the literature and a detailed overview of its definitions is provided by David Traum and Peter Heeman (1997). Of note is that the meaning of the term ‘utterance’ goes beyond that of the result of an act of uttering; it refers to a basic sub-sentence unit. For the present contrastive analysis, the boundaries of these smaller utterance units are defined based on syntactic criteria. For instance, the sentence in (33) ‘*Not just one, but both of them.*’ is analyzed as consisting of two utterances ‘*not just one*’ and ‘*but both of them*’.

(33) One sentence, two utterances:

- a. [Go and hurry. Be near your brothers.] Not just one, but both of them.

The above example also highlights the importance of the division into sub-sentence level utterances. In (33a), there are two instances of the stripping ellipsis of the antecedent ‘*be near*’, as illustrated in (33b).

(33) Two occurrences of stripping:

- b. [Be near] not just one, but [be near] both of them.

The focus on the smaller sub-parts of the sentence is of prime necessity for the coherent analysis of antecedent-based ellipsis. Without such a division the example would be treated as a single instance of ellipsis, when in fact it represents two occurrences. Thus, analysis of ellipsis must occur at the level of the sub-sentence level utterances, and *not* at the level of the sentence.

The discussion of ellipsis will be further developed shortly, but it is important for the following description of the three necessary and sufficient conditions of verbless sentences to highlight the compound nature of the notion of the sentence.

4.2 INITIAL AND FINAL MARKING

The goal of the first condition is to delimit the boundaries of the sentence. The string must contain initial and final marking. This requirement establishes that the string of text is independent as opposed to embedded.

In writing, the initial marking is generally the capitalization of the first letter in the string; the final marking usually occurs by means of major punctuation marks including the period, exclamation mark and question mark.³⁰ In addition, turn change has an important role in delimiting the sentence and overrides the typical final punctuation and initial capitalization

³⁰ As automatizing the process of sentence delimiting for the present study has shown, these are by far not the only markings that delimit the sentence unit in writing. Other important markings include combinations of the major punctuation marks, and in many cases also include the continuation symbol, the quotation symbol, long dashes, and the interaction of various markers with the quotation symbol during turn change.

marking. For instance, this may be a change from one speaker to the next in a written dialogue, as illustrated in (34a), or it could be a change from narration to direct speech, as in (34b). A change of turn is thus also included as a marker of sentence boundaries.

(34) Turn change:

- a. Speaker 1: Go and hurry.
Speaker 2: Be near your brothers.
Speaker 3: Not just one, but both of them.
- b. “Go and hurry,” he said.

In spoken language, it is supposed that the delimiting of the sentence may carry special intonation to mark its start and end; the length of the pauses may also constitute a relevant marker. Like in writing, turn change (from one speaker to another, as well as direct quotation), would override the other typical intonation marking. In Russian syntactic discussion, the subject of intonation is often intertwined with sentential status. As noted by Potapova (2017: 26), the importance of intonation as an indicator of predication is emphasized in the syntax of Nina Valgina (2000), who relies on intonation particularly as a marker of Russian nominative sentences, as well as in the Russian syntax of Aleksej Rudnev (1968).³¹

Initial and final marking is required in order to establish that the structure at issue is not itself embedded into another structure. This formal marking is *part* of what distinguishes a structure as a word or a phrase (e.g. ‘red’ in the ‘*The red painting on the wall.*’) from a structure that is itself sentential (e.g. ‘*Red.*’), without needing to evoke any hidden structure.

The importance of the latter part, i.e. without any hidden structure, is key. As discussed in *Part 1: Chapter 3* above, the syntactic ellipsis hypothesis makes the distinction between the noun phrase ‘*car problem*’ in (35a) and the sentential use of the same noun phrase in (35b) ‘*Car problem!*’ on the basis of hidden structure.

- (35) a. She has a car problem.
- b. Car problem!

On the syntactic ellipsis hypothesis, the noun phrase in (b) is attributed a hidden verbal predicate and inflection which give it sentential status as an ‘elliptical sentence’. For several

³¹ A thorough discussion of intonation patterns in Russian is found in the work of Irina Fougeron (1984; 1986; 1988).

reasons that were outlined above, hidden syntactic structure is a very difficult argument to maintain. Upon rejecting hidden syntactic structure, it may be tempting to take a nonsentential perspective with regard to the structure in (b), i.e. the perspective which states that although (b) is syntactically built using the same mechanism as (a), i.e. bottom-up merging without any hidden structure, (b) lacks an inflected verb which makes it a nonsentential structure. However, the previous section also raised several objections against accepting the nonsentential conclusion. One of the issues there raised is that the elimination of hidden structure, which seems correct, requires reconsideration of the difference between the independent (35b) ‘*Car problem!*’ versus the embedded noun phrase ‘*car problem*’ in (35a), reproduced as a plain noun phrase in (35c); as well as the verbal independent structure (35a) ‘*She has a car problem*’ versus the version that is ready for embedding in (35d).

- (35) c. car problem
- d. she has a car problem

It appears that these differences can no longer be treated as being of the syntactic order; syntactically (b) and (c), just like (a) and (d), have the same structure. Each instance is built using the same bottom up method as the other, and, without a top-down requirement (which would be contrary to the bottom-up economy principle), each instance remains a phrase, i.e. (a) and (d) are inflected verb phrases, whereas (b) and (c) are noun phrases.

Furthermore, basing sentential status on syntactic structure would not only represent a top-down contradiction to bottom-up models, but it would also be, in principle, undesirable from the perspective of recursion. For instance, it is necessary that syntactically it is possible to embed ‘*car problem*’ into ‘*she has a car problem*’, and then further expand the latter by ‘*someone said that she has a car problem*’, and so on to a potential infinity. Positing the absence of top-down sentential requirements in a minimalist bottom-up model appears to allow such productive syntactic construction. Thus, it seems to be a positive development that sentential status is not determined by syntax.

Initial and final marking is a formal restriction on sentential status. Its satisfaction requires analysis of the *linguistic context* of the structure in order to determine whether or not it is embedded into another structure. It is notable that the un-embedded requirement is necessary for all types of sentences, both verbal and verbless.

4.3 CONTEXTUAL INSERTION

Apart from the initial and final marking, no other restrictions on the form of the sentence are posited. However, the satisfaction of this requirement does not render a syntactic string into a sentence. It is most importantly contextual insertion that assures the necessary element for sentential status, i.e. the ‘focus’ in the sense of Knud Lambrecht (1994).

Crucially, it must be stressed that what is being proposed is not the replacement of the syntactic predicate and syntactic subject with a theme and rhema model, which, as discussed in *Part 1: Chapter 2*, has historically been attempted several times with much criticism stemming from the fact that it mixes up logic and linguistics. As mentioned above in *Section 4.1.1*, the traditional linguistic definitions apply with regard to the syntactic ‘clause’, i.e. the syntactic clause is a combination of the syntactic subject, defined as a noun in the nominative case, and the syntactic predicate, defined as a finite verb. What the present discussion stresses is that the notion of the syntactic clause is not sufficient for sentential status since (1) it does not on its own rule out embedding (the linguistic context is necessary for this, as was discussed in *Section 4.2*), and (2) it does not carry the essential element for predication (i.e. an information structural ‘focus’ which can be acquired only through the insertion of the structure into a context).

The present proposal is that a structure, that is syntactically built bottom-up, constitutes a **sentence** in so far as it is found in a context where it is used *unembedded* and carries an information structural *focus*.

For instance, for either ‘*car problem*’ or ‘*she has a car problem*’ to constitute a sentence it must be found used un-embedded in a context that gives the entire structure, or a part of it, a focus. From the perspective of predication, it is the focus that starts things off; the presence of the focus makes the word, phrase, or combination of syntactic phrases into a sentence. Without contextual insertion, the syntactic structure remains a series of words that lacks informational value and may itself be part of another structure. By a loose analogy, the syntactic structure alone may be compared to bricks and cement that constitute the mechanism of building a house from the ground up: just like the solid attachment of well-fitted bricks to one another does not constitute a house, the syntactic requirements of each element must necessarily be satisfied, but syntactic satisfaction alone does not constitute a sentence. It is information structure that provides the blue print for the sentence.

At this point it is necessary to be precise about what particularly is meant by information structure and the ‘focus’ requirement which is attained through contextual insertion of an unembedded syntactic structure.

4.3.1 Information Structure

It is widely acknowledged that when using words to convey information about something to someone, the structure of the sentence that a speaker produces will necessarily be influenced by the assumptions that the speaker makes about the hearer’s state of mind. For example, when talking about a particular person, mentioning the person’s full name at the beginning of every sentence would be redundant: from the point that the person’s name is mentioned once, its referent typically becomes predictable and, if necessary to evoke again, the speaker will do so by means of a pronoun. Thus, assuming that the hearer is familiar with certain elements and not familiar with others, changes the sentence that is produced. In other words, pragmatic concerns about the utility of certain information influences the grammatical structure of the sentence.

There exist various definitions of information structure and its key components. The present study follows Knud Lambrecht’s (1994) analysis of the essentials of information structure. In what follows, the central aspects of Lambrecht’s definitions, which are presently adopted, are summarized and illustrated using examples from the current corpus.³²

4.3.1.1 Information

One of the distinguishing and powerful features of particularly Lambrecht’s conception of information structure is the notion of ‘information’ itself. For Lambrecht, information only arises when something new is related to something that has been taken for granted (Lambrecht, 1994: 48). In other words, information is something in which new and old elements are

³² In particular, unless otherwise stated, the illustrative examples in *Part 1: Section 4.3.1* and its sub-parts are from the corpus parts ВК_Р0 (Russian original ‘Братья Карамазовы’) and ВК_Е1 (English translation ‘The Brothers Karamazov’), which are respectively presented in part (a) and part (b) of the examples.

combined (Lambrecht, 1994: 51). Thus, the establishment of a *relation* is central to the concept of information.

Notably, Lambrecht issues the warning that an old or new element of the sentence cannot itself be identified as information:

Information is not conveyed by lexical items or individual sentence constituents but only by establishing RELATIONS between denotata and propositions. (Lambrecht, 1994: 209)

He emphasizes that neglecting the relational role of information ultimately leads to misleading definitions of Topic and Focus as themselves types of information.

Practically, Lambrecht's definition of information means that in (36), it would be incorrect to say that information is carried by the new element in the sentence, i.e. 'как философ Дидерот' (*kak filosof Diderot*; lit. 'like philosopher Diderot') in Russian or 'am like the philosopher Diderot' in English.

(36) a. Russian

[Я только в последнее время усомнился, но зато теперь сижу и жду великих словес.]

Я,	ваше	преподобие,	как	философ	Дидерот.
<i>ja</i>	<i>vaše</i>	<i>prepodobie</i>	<i>kak</i>	<i>filosof</i>	<i>Diderot</i>
<i>I</i>	<i>your.2PL</i>	<i>reverence</i>	<i>like</i>	<i>philosopher</i>	<i>Diderot</i>

b. English

[It's only lately that I've begun to have doubts, but to make up for it I'm sitting and waiting to hear lofty words.]

I am, reverend Father, like the philosopher Diderot.

Rather, information arises when the new element which is unpredictable (in this case, this is being like Diderot), is related to some presupposed element (in this case, the presupposed element happens to be the linguistically explicit syntactic subject and simultaneous Topic under discussion, i.e. the speaker himself). Thus, a new piece of *information* is created from the *relation* of a non-recoverable element to *something* recoverable.

4.3.1.2 Topic

The concepts of Topic and Focus are fundamentally different from that of the subject and the predicate. The former are *pragmatic relations* to the sentence, whereas the latter are syntactic categories. As mentioned at the start of this section in 4.1.1, the term subject and predicate are presently used according to their traditional syntactic definitions; to be clear, the notions of Topic and Focus do not replace those of the subject and the predicate, nor is the combination of the former considered to constitute a syntactic clause.

As with the concept of information, the terms Topic and Focus are found used in several ways in the literature. The current section outlines the particularities of the presently adopted definition of Topic which follows that of Lambrecht (1994).³³

4.3.1.2.1 Aboutness

To be extremely concise, the Topic of the sentence is what the sentence is *about*. More precisely:

The topic of the sentence is the thing which the proposition expressed by the sentence is ABOUT. (Lambrecht, 1994: 118)

This notion of ‘aboutness’ is clarified in Strawson’s (1964) definition of the Topic as:

the matter of current interest which a statement is about and with respect to which a proposition is to be interpreted as relevant. (Lambrecht, 1994: 119).

³³ Notably, the pragmatic notions of Topic and Focus are different not only from the syntactic notions of Subject and Predicate, but also from the notions of Theme and Rhema. As discussed in *Part 1: Chapter 2* above, the latter notions are typically purely logical; however, their defining characteristics also sometimes involve word order and intonation requirements. For instance, the account of Theme proposed in (Bonnot, 1999) appeals to word order as one of the essential formal criteria, i.e. the Theme is defined as a constituent that can only occur in sentence initial position:

In order to escape the ambiguity that results from the traditional semantic definitions of the theme, we start from a formal definition, based on word order and intonation: the theme is a constituent always in initial position which can be separated from the rest of the utterance by a potential pause. (Bonnot, 1999: 15)

As will be shown, the Theme, whether it is defined from a purely logical perspective or from a formal one, is a different notion from the notion of Topic proposed by Lambrecht.

This concise notion reveals particular nuances when it comes to recognizing the Topic in practice.

4.3.1.2.2 Revealed by Discourse Context

A distinguishing feature of Lambrecht’s definitions is that the notion of Topic does not a priori integrate a particular word order. Although there are tendencies for the Topic to be expressed in the beginning of the sentence in many languages, Lambrecht stresses that it is not the syntactic structure that reveals the Topic of the sentence, but that it is necessary to look to the discourse context of the sentence (Lambrecht, 1994: 120). In other words, the position of the Topic is a matter of synthetic analysis, not analytic definition.

This particular aspect of the notion of Topic is supported by examples such as (37), which illustrates a case where the subject is *not* the Topic.

(37) a. Russian:

[Раз много лет тому назад, говорю одному влиятельному даже лицу: «Ваша супруга щекотливая женщина-с», – в смысле то-есть чести, так сказать, нравственных качеств, а он мне вдруг на то: «А вы её щекотали?»]

Ваша	супруга	щекотливая	женщина-с
<i>vaša</i>	<i>supruga</i>	<i>ščekotlivaja</i>	<i>ženščina-s</i>
<i>your.2PL</i>	<i>wife.NOM</i>	<i>ticklish</i>	<i>woman-sir</i>

b. English:

[Once, this was many years ago now, I said to an influential person, “Your wife, sir, is a ticklish woman,” referring to her honor, her moral qualities, so to speak. And he suddenly retorted, “Did you tickle her?”]

Your wife, sir, is a ticklish woman

In the above, the subject, i.e. ‘wife’ in English and ‘супруга’ (*supruga*; lit. ‘wife’) in Russian, is being introduced into the conversation. It cannot be the Topic (or more accurately, this constituent does not have a Topic relation to this sentence) since it cannot be presupposed to be in the mind of the hearer at the time of its utterance.

4.3.1.2.3 Not Old Information

Contrary to other definitions, Lambrecht insists that Topic must not be identified with ‘old information’, if only because ‘old information’ is often in Focus:

Given the fact that expressions with ‘old’ referents can be either topics or foci in a sentence, there can be no one-to-one correspondence between pragmatic relations and pragmatic properties of referents. Therefore, to assert, as is often done in discussions of topic, that the topic of a sentence is ‘the old information’ is, to say the least, misleading. However, it is equally misleading to assert that there is no necessary relationship at all between the two parameters. (Lambrecht, 1994: 164)

Lambrecht explicitly rejects Wallace Chafe’s (1976) definition of Topic as the “hitching post for new knowledge” and Focus as the new information “hitched” to it (Lambrecht, 1994: 206). For Chafe, the Topic is:

not so much what the sentence is about as the frame within which the sentence holds. (Chafe, 1976: 50)

Lambrecht argues that Chafe’s definition is motivated by a desire to distinguish two different kinds of topics, a distinction that Lambrecht finds unnecessary (1994: 118).

4.3.1.2.4 Precise Criteria

Instead of defining Topic as old information, Lambrecht urges that Topic must be thought of as a “pragmatically construed relation to a proposition” which is predictable and recoverable (Lambrecht, 1994: 218).

In practice, in order to determine whether an expression stands in a Topic relation to the proposition, a series of questions must be asked. *Figure 7* narrows down five criteria that an expression must necessarily meet in order to qualify as the Topic of the sentence, according to Lambrecht (1994).³⁴

³⁴ On criterion four, a relevant discussion of accessibility theory is found in Mira Ariel (2001).

1 The sentence must be expressing information *about* the referent of the expression.

2 A Topic must be *referential*.

- A dummy subject cannot be a Topic since it is not referential (1994: 228); neither can a referent that is not presupposed to exist (e.g. 'car' in 'Bill doesn't have a car').
 - Existential 'there' cannot be a Topic expression; however, deictic 'there' can (provided it meets the other criteria). (1994: 155–156)
 - It is important to note that both entities and propositions may be discourse referents. (1994: 74)
-

3 A Topic must be *identifiable*.

- An identifiable referent is a referent "for which a representation exists in the addressee's mind". (1994: 77)
 - Unique referents (e.g. 'the sun'), classes of entities salient for interlocutors (e.g. 'the kids'), deictic (e.g. 'those', 'there', 'your leg') and anaphoric referents, are automatically assumed identifiable referents. (1994: 87–89)
 - There exists a correlation between *identifiability* (i.e. a cognitive category which exists in all languages) and *definiteness* (i.e. a grammatical category, which exists in English but not in Russian); however, it is imperfect (1994: 80). Identifiability is a matter of degree (1994: 84).
 - Brand new referents are often anchored so as to be more identifiable (e.g. 'a guy I work with' is anchored to the speaker). This may affect the acceptability of an expression as Topic. (1994: 86)
 - Indefinite pronouns and quantified expressions (e.g. 'nobody', 'everybody', 'many people') cannot be Topics, while universals can under certain conditions. (1994: 156)
-

4 A Topic must be *accessible*.

- The referent must have "pragmatic salience" in discourse. (1994: 262)

- Specifically, a Topic referent must be either *active* (i.e. “currently lit up” for the speaker and the hearer) or *accessible* (i.e. in the peripheral consciousness but not being focused on). (1994: 94)
- Unused referents (i.e. inactive but identifiable) make the *borderline* case for acceptability as a Topic: new information expressed about unused referents requires a lot of mental effort to process. (1994: 166)
- A brand-new referent (i.e. unidentifiable for the hearer) cannot be a Topic.

[It] forces a hearer to put the predication on hold, so to speak, until she finds what she is receiving information about. (1994: 166)

One cannot ‘add’ information about a referent unless this referent is in some important sense already available in the discourse as a starting point. (1994: 164)

From a certain degree of inactiveness, the Topic becomes a Focus (1994: 164).

- Pronouns are necessarily active, with the exception of deictic pronouns. Some deictic are inactive, e.g. ‘*that*’ in ‘*I want that*’ said while pointing, is activated by the utterance (1994: 95–96). Those deictic that have a “salient presence in the text-external world” are active (1994: 110).

5 A Topic must *not be the focus*.

- A Topical Expression may be a part of the Focus Domain, but in such cases this expression cannot be the Topic of the sentence. In other words:

Focus domains must be allowed to contain non-focal elements.
[However] focus elements cannot be part of topical domains. (1994: 216)

This means that when an expression, even a pronoun, that meets the previous four conditions, is in Focus, it cannot have a Topic relation to the sentence. This is frequently the case in contrastive sentences (1994: 288). The crucial point is that “the topic must be taken for granted” (1994: 153). Sentences where the element is in Focus signal that a new relation, one that is not taken for granted, is being established.

Example (37) above illustrates the case of a subject in Focus, i.e. having an unpredictable relation to the proposition.

Figure 7. The five necessary criteria of a Topic

4.3.1.3 Focus

4.3.1.3.1 Fundamental to the Sentence

The crucial reason for the present emphasis on context insertion as a necessary requirement for sentential status is because it is the context that assures the foundational element of any sentence: the information structural Focus.

Regarding the definition of a sentence, Lambrecht (1994) holds the following perspective:

For a structure to qualify as an independent sentence it must express an assertion, i.e. the proposition expressed by it must contain a focus. There are no independent sentences expressing only pragmatically presupposed propositions. (This is true even for sentences like ‘*I love you*’, which may have been said hundreds of times to the same addressee.) (Lambrecht, 1994: 236)

This perspective perfectly corresponds to that which is presently being proposed. A few points must be made with regard to the definition of Focus.

4.3.1.3.2 Not a Complement of the Topic

The first is that defining the Focus as a constituent of a sentence which complements the Topic is merely a “convenient shorthand”, as explained by Lambrecht (1994). In reality:

All sentences must have a focus. However, not all sentences have a topic. (Lambrecht, 1994: 206)

The Topic may be omitted from the sentence, since it is by definition something that is taken for granted. However, a constituent in Focus cannot “be omitted without depriving the utterance of some or all of its informational value” (Lambrecht, 1994: 224). The fact that the Topic need not be explicit makes it unequal to the Focus. Thus, it is misleading to define the Focus as a complement of the Topic.

4.3.1.3.3 Not New Information

Neither is the Focus to be defined as new information. First, Focus, like Topic, for Lambrecht (1994), is a *relation* and *not* a property. As such, the Focus must *not* be identified with any type of *information* – old or new:

An expression [...] can have information value only as an element of the proposition expressed by the entire sentence. (Lambrecht, 1994: 209)

Secondly, Lambrecht argues that the property of ‘newness’ is *not what matters* in a Focus constituent:

The function of grammatical focus marking must be to express such relations rather than to attribute the property ‘new’ to the denotata of individual sentence constituents. (Lambrecht, 1994: 209)

Here again, Chafe’s (1976) definition differs from Lambrecht; for the latter, Topic and Focus are pragmatic relations, not categories of the sentence. Recent work by Zsuzsanna Gécseg (2011) also appears to depart from Lambrecht’s view. Gécseg treats the Focus as both an element that complements the Topic, and as itself constituting information:

[L]es deux constituants de base de la phrase ne sont pas le sujet grammatical et le prédicat grammatical, mais le Topique et le Commentaire. La position du Topique accueille généralement le ou les constituants dénotant les référents à propos desquels le Commentaire fournit une nouvelle information. (Gécseg, 2011: 33)

Thus, the Focus, like the Topic, also has various definitions in the literature. The essence of Lambrecht’s is that the Focus is (i) a pragmatic relation (not information; not a property) and (ii) does not complement the Topic, but is primary in the sentence.

The Focus “has to do with conveying new information” only in the sense that conveying new information without a Focus is impossible:

The focus relation relates the pragmatically non-recoverable to the recoverable component of a proposition and thereby creates a new state of information in the mind of the addressee. (Lambrecht, 1994: 218)

Thus, it is the Focus which turns the structure into a piece of information, i.e. an assertion (Lambrecht, 1994: 217). In other words, it is the presence of a pragmatic Focus relation that makes a structure into an instance of predication.

4.3.1.3.4 Focus Relation May Include Old Constituents

A Focus often contains ‘old’ or presupposed propositions, and, crucially, it is sometimes even “coextensive with such constituents” (Lambrecht, 1994: 217). This is because such ‘old’ constituents may in fact have a relation with the proposition that is unpredictable and unrecoverable.

One such instance is illustrated in (38). The pronoun ‘you’ in English and ‘ты’ (*ty*; lit. ‘you.2SG’) in Russian is assumed to be active in the mind of the hearer, yet it is part of the Focus in this example.

(38) a. Russian

[Максимов: «Да ведь и я не фон-Зон, я Максимов.»
Федор Павлович: «Нет, ты фон-Зон.»]

Нет,	ты	фон-Зон.
<i>net</i>	<i>ty</i>	<i>fon-Zon</i>
<i>no</i>	<i>you.2SG.NOM</i>	<i>von-Sohn</i>

b. English

[Maximov: “But I’m not von Sohn either, I am Maximov.”
Fyodor Pavlovich: “No, you’re von Sohn.”]

No, you’re von Sohn.

This sentence is identifying a particular person as being von Sohn, thus establishing a new unpredictable relation between the active element ‘you’, or ‘ты’ (*ty*; lit. ‘you’), and the rest of the proposition.

Keeping in mind all of the above points, a better way to identify the Focus in practice is made precise in (39):

(39) IDENTIFYING THE FOCUS:

That unpredictable or non-recoverable element of a proposition which is left if we take away the elements which are predictable or recoverable. (Lambrecht, 1994: 217)

In other words, instead of old and new information it is more appropriate to think of Topic and Focus in terms of predictability and recoverability:

Like the topic, the focus is an element which stands in a pragmatically construed relation to a proposition. But while the pragmatic relation between the topic and the proposition must be predictable or recoverable, the relation between the focus element and the proposition is assumed to be unpredictable or non-recoverable for the addressee at the time of utterance. (Lambrecht, 1994: 218)

It is precisely this definition of Focus which is presently adopted.

4.3.1.4 Focus Structure Types

The above notions of Topic and Focus allow Lambrecht to identify three basic sentence structures that exist cross-linguistically: Topic-Comment, Identificational and Thetic. Each structure corresponds to a different communicative function.

4.3.1.4.1 Sentence Type 1: Topic-Comment

The first type of sentence structure corresponds to the communicative situation where the speaker wants to make some comment about a topic that he or she believes to be the matter of current interest in the discussion (Lambrecht, 1994: 222). For example, if in a discussion about borsch, I want to comment that it is delicious, I could structure my sentence in a wide variety of ways such as '*The borsch is delicious*', or '*It's delicious*', or just '*Delicious*'. In all of the different versions the common element is that some property is being predicated of an established discourse referent (Lambrecht, 1994: 126). This type is called 'Topic-Comment' or alternatively 'Predicate-Focus'.

4.3.1.4.1.1 Criteria

The Topic-Comment structure is identifiable by the following two necessary criteria.

i) Subject and Topic Coincide

First, the subject and the Topic of the sentence must coincide. Most natural for this type of structure is for the constituent in which the subject-Topic coincide to be either a pronoun or null (i.e. as in absent); however, it could also be a lexical noun phrase (Lambrecht, 1994: 223).

The key is that the lexical noun phrase or pronoun is optional: its omission would not result in any loss of propositional information. The only communicative reason for the full lexical noun phrase version is to make sure that the Topic is indeed active in the hearer's mind. If there does exist an explicit subject in the sentence, that subject must be the Topic for the sentence to have the Topic-Comment structure.

ii) Predicate is in Focus

Secondly, the predicate must necessarily be in Focus. If the predicate is a noun phrase, it has an additional requirement – it must be *non-referential*. If it is referential, the sentence is Identificational.

Consider the example in (40) from Lambrecht (1994: 123) and the possible readings of the noun phrase '*my friends*':

(40) The ones who did that are my friends.

Lambrecht explains that in the Topic-Comment reading of the sentence, '*my friends*' is non-referential:

the subject 'the ones who did that' refers to an identifiable set of individuals [...] in which '*my friends*' is a non-referring predicate nominal. (Lambrecht, 1994: 123)

Alternatively, if '*my friends*' is identifying a group of people as the missing argument in the open proposition '*The ones who did that are X*' or '*It's X that did that*' or '*X did that*', the communicative purpose is no longer to simply comment on the ones who did that but to identify them.

4.3.1.4.1.2 Unmarked

Formally, the Topic-Comment structure is said to be the most common. Lambrecht calls "unmarked" this most frequently used structure precisely because, considered out of context, it provides no specific marking of what type it actually is:

Topic-comment sentences [...] are syntactically and prosodically UNMARKED with respect to their information structure, i.e. their formal

structure is compatible with other pragmatic construals, in which the subject is not a topic. (Lambrecht, 1994: 122)

Considered in isolation, a Topic-Comment sentence often looks exactly like one of the other types. It is the context that reveals what type it actually is. The context must provide evidence to suggest that the structure is *not* a Topic-Comment. If there is no such evidence in the context, then the structure is “unmarked” and thus, the default, Topic-Comment reading can be accepted.

4.3.1.4.1.3 *Illustration*

The practical consequences of this fine nuance are best illustrated by Lambrecht in a series of examples provided near the very end of *Information Structure and Sentence Form*. Consider the sentence ‘*Her husband is sick*’ as part of the following example analyzed by Lambrecht (1994: 308).

- (41) Speaker A: Why didn’t she come to work today?
 Speaker B: Her husband is sick.

This sentence would be considered a Topic-Comment sentence unless there is evidence that Speaker B wants to explicitly stress that it is the husband who is the someone that is sick. If the speaker is merely providing a reason for why the woman didn’t come to work – i.e. that someone is sick – and has no intention of emphasizing the husband, the statement is Topic-Comment. This is true even if Speaker A has not even the slightest suspicion that the woman is married (i.e. even if the husband is *not* active in the conversation). What matters is that the predicate ‘*is sick*’ is in Focus, and, that there is *no* reason to believe that the subject is also in Focus. If there is no evidence to believe that the subject is also in Focus, the Topic-Comment reading is accepted.

The following example illustrates a prototypical case of a Topic-Comment structure:

(42) a. Russian

[Он не мог слышать известных слов и известных разговоров про женщин.]

Эти	«известные»	слова	и	разговоры,
<i>èti</i>	<i>izvestnye</i>	<i>slova</i>	<i>i</i>	<i>razgovory</i>
DEM.NOM	ADJ.NOM	NN.NOM	CONJ	NN.NOM
<i>these</i>	<i>famous</i>	<i>words</i>	<i>and</i>	<i>conversations</i>
к	несчастью,	неискоренимы	в	школах.
<i>k</i>	<i>nesčast'ju</i>	<i>neiskorenimy</i>	<i>v</i>	<i>školax</i>
PREP	NN.DAT	ADJS.PL	PREP	NN.PREP
<i>to</i>	<i>misfortune</i>	<i>ineradicable</i>	<i>in</i>	<i>schools</i>

b. English

[He could not bear to hear certain words and certain conversations about women.]

These “certain” words and conversations, unfortunately, are ineradicable in schools.

In (42), the Topic, i.e. in English, ‘*these certain words and conversations*’ and in Russian, ‘эти известные слова и разговоры’ (*èti izvestnye slova i razgovory*; lit. ‘*these famous words and conversations*’), has been established in the previous sentence. In the current sentence, something unpredictable is said about these certain famous words. The focus is on ‘*are ineradicable in schools*’ in English and on ‘неискоренимы в школах’ (*ne iskorenimy v školax*; lit. ‘*ineradicable in schools*’) in Russian.

4.3.1.4.2 Sentence Type 2: Identificational

The communicative function of the second type of sentence structure is to identify a referent (Lambrecht, 1994: 222). For example, if when asked ‘*What is the most delicious kind of soup?*’ the speaker wants to identify that out of all the possible soups the most delicious is the borsch, there are various ways that the speaker could syntactically structure the sentence, but they would all be similar in terms of information structure. I could say for instance ‘*The most delicious kind of soup is the borsch*’, or ‘*The borsch is the most delicious*’, or just ‘*Borsch*’. In all of the different versions the common element is that a relation is established between the argument and a previously evoked open proposition (Lambrecht, 1994: 126). In other words, the argument ‘*borsch*’ provides the missing referent in the proposition ‘*X is the*

most delicious kind of soup’, a proposition that can be presupposed based on the context. This type of information structure is called ‘Identificational’, or alternatively, ‘Argument-Focus’.

4.3.1.4.2.1 Criteria

i) Presupposed Open Proposition

The Identificational structure is distinguished by the fact that a Focus relation is established between only one referential element and the rest of the sentence. The key is that an open proposition into which this element is added must be presupposed.

Notably, it is not a question of the constituent provided being ‘new information’. For one, because the constituent can be already active (or ‘old’), but also, because the constituent alone can have no informational value without the rest of the proposition:

The expression can have informational value only as an element of the proposition expressed by the entire sentence. What is ‘new’ is not the constituent nor its designatum, but its role as the [...] argument of the predicate [...] in the pragmatically presupposed open proposition. (Lambrecht, 1994: 209–210)

The defining feature of this structure is that the constituent provided must be the one and only Focus of the sentence: “its addition to the sentence makes the sentence a new piece of information” (Lambrecht, 1994: 211).

ii) Subject as Topic or Focus

Regarding the subject of Identificational sentences it is worth stating that it may be the Topic or the Focus. For this type of sentence, if the subject is in Focus, the predicate cannot also be in Focus, since the latter must be part of the presupposed proposition. For instance, the subject ‘*borsch*’ is the Focus in (43):

(43) What’s so delicious? The borsch is delicious.

Alternatively, there exist Identificational sentences where, like in Topic-Comment sentences, the subject is the Topic. The subject ‘*I*’ is the Topic in the following Identificational sentence:

(44) What would you like to eat? I would like to eat borsch.

In an Identificational structure where the subject is the Topic, the Focus must be on the predicate; this predicate must include a referential element; finally, everything in the sentence, other than the referential element, must constitute an open-proposition which can be reasonably assumed as presupposed at the time of utterance. Such are the essential elements for distinguishing Identificational sentences from the other types of structures.

4.3.1.4.2.2 *Illustration*

A prototypical example of an Identificational sentence is example (45), discussed above in the section on Focus in (38). It is reproduced here to illustrate that the open proposition is not always created by a question, as it was in (43) and (44).

(45) a. Russian

[Максимов: «Да ведь и я не фон-Зон, я Максимов.»
Федор Павлович: «Нет, ты фон-Зон.»]

Нет,	ты	фон-Зон.
<i>net</i>	<i>ty</i>	<i>fon-Zon</i>
<i>no</i>	<i>you.2SG.NOM</i>	<i>von-Sohn</i>

b. English

[Maximov: “But I’m not von Sohn either, I am Maximov.”
Fyodor Pavlovich: “No, you’re von Sohn.”]

No, you’re von Sohn.

In example (46), it is a question which is the Identificational sentence.

(46) a. Russian

[«Я был, был, я уже был... Un chevalier parfait!» и помещик пустил на воздух щелчок пальцем. «Кто это chevalier?» спросил Миусов.]

Кто	это	chevalier?
<i>kto</i>	<i>èto</i>	<i>chevalier</i>
<i>who</i>	<i>this</i>	<i>chevalier</i>

b. English

[“I’ve been, I’ve been already... Un chevalier parfait!” And the landowner loosed a snap of his fingers into the air. “Who is a chevalier?” asked Miusov.]

Who is a chevalier?

The open proposition that ‘*someone is a chevalier*’ in English, or ‘кто-то chevalier’ (*kto-to chevalier*; lit. ‘*someone chevalier*’) in Russian, is presupposed, and the particular referent of that someone is being inquired about.

4.3.1.4.3 Sentence Type 3: Thetic

The final type of sentence structure corresponds to the communicative situation where the speaker either wants to explicitly introduce³⁵ a new referent into the discourse, or, they want to report some event (Lambrecht, 1994: 222). For example, a speaker could introduce ‘*borsch*’ into the conversation by saying ‘*There exists a really delicious Russian soup called borsch*’ or ‘*One of the most famous Russian dishes is borsch*’; or, if she wanted to introduce the event of ‘*cooking borsch*’ she could say ‘*Once upon a time I cooked borsch*’ or ‘*Today we’re cooking borsch*’ or ‘*It is time to cook borsch*’. In all of these, ‘*borsch*’ is *not* a Topic, and an open proposition (e.g. ‘*It is time to cook something*’) *cannot* be presupposed.

³⁵ ‘*Explicitly introduce*’ stresses the important difference between the *Topic-Comment* construction (e.g. (a), where the inactive but accessible referent ‘*his lover*’ plays the role of Topic), and the *Presentational* construction, (e.g. (b), where the same inactive but accessible referent ‘*his lover*’ is part of the Focus). This example is analyzed in Lambrecht in detail (1994: 110-114). The example illustrates that the distinction is *not* merely a matter of activation status.

- (a) Topic-Comment:
Remember Mark? His lover just died of aids.
(predicates something of ‘his lover’)
- (b) Presentational:
Remember Mark? I ran into his lover yesterday, and he told me he had aids.
(introduces ‘his lover’ into the discourse)

It is the *communicative purpose* that makes the Presentation construction fundamentally different. In the example, the noun phrase ‘*his lover*’ is *equally accessible* in both sentences. The difference is that that the purpose of the Topic-Comment sentence is to predicate something about the subject ‘*his lover*’, whereas the purpose of the Presentational sentence is to introduce ‘*his lover*’ into the discourse.

The common feature among these sentences is that they introduce:

a new element into the discourse without linking this element to an already established topic or to some presupposed open proposition. (Lambrecht, 1994: 144)

These sentences have what Lambrecht calls an “all-new character”. This type of structure is called ‘Thetic-Presentational’ (or ‘Presentational’) when a discourse referent is being introduced, and ‘Thetic-Event-Reporting’ (or just ‘Event-Reporting’) when it is an event that is being introduced, or alternatively, the two types are referred to under the single title of ‘Sentence-Focus’ (or simply ‘Thetic’).

4.3.1.4.3.1 *Criteria*

i) Entire Sentence is in Focus

Like the other types, it is impossible to determine if the information structure of the sentence is of the Thetic type without considering its context: a structure which is Thetic in one context may be Topic-Comment or Identificational in another. Lambrecht (1994: 299) demonstrates this with the sentence ‘*Bill went straight home*’:

- | | | | |
|---------|-------------------|---------------------------|---------------------------------|
| (47) a. | Identificational: | <i>Where did Bill go?</i> | <i>Bill went straight home.</i> |
| b. | Topic-Comment: | <i>What about Bill?</i> | <i>Bill went straight home.</i> |
| c. | Thetic: | <i>What happened?</i> | <i>Bill went straight home.</i> |

The sentence ‘*Bill went straight home*’ is *Identificational* in the context of ‘*Where did Bill go?*’. The same sentence becomes *Topic-Comment* if Bill is seen as topical (even if inactive), and the predicate ‘*went straight home*’ is being attributed to it. Finally, it would be *Thetic* if the fact that someone went straight home cannot be presupposed and the prior conversation had nothing to do with Bill, such as if someone asked ‘*What happened?*’.

In the Thetic case, the Focus “extends over the entire proposition” (Lambrecht, 1994: 233). The key to identifying a Thetic structure is that the entire sentence is in Focus. For this to happen, the *subject* (if it exists) *cannot be a Topic* and the *proposition cannot be partly presupposed*.

ii) Important Nuance: Presupposed Relation vs. Support of Predication

The difference between the Topic, as a pragmatic relation, and the support of the predication must be emphasized. It may be tempting to altogether reject the existence of Thetic sentences when Topic is not treated as a relation.

For instance, Francis Cornish (2009) argues that Thetic statements do not constitute instances of predication due to the fact that by definition they do not have a Topic, and, therefore, should be analyzed as missing the support of the predication:

Pour qu'il y ait prédication (sémantique), il faut que l'entité dont on prédique quelque chose ait le statut de topique potentiel (donc de « support » d'une prédication). Or, dans les propositions thétiques, le sujet n'est pas un topique potentiel, il fait partie du focus – toute la proposition étant focalisée dans ce cas. (Cornish, 2009: 121)

Lambrecht's notion of Topic as a pragmatic *relation* makes it possible to overcome such an objection to Thetic sentences. Notably, the absence of such a Topic does not constitute the absence of a support for the predication. Rather, the absence of Topic constitutes the absence of a predictable relation between the discourse context and any referent of the syntactic elements of the sentence structure. For instance, when the speaker utters (47c) '*Bill went straight home*' in the context of '*What happened?*', neither the referent of the subject '*Bill*', nor an open proposition, can be presupposed: nothing in the structure '*Bill went straight home*' is predictable from this particular context.

Consideration of the important question raised by Cornish (2009), as to whether there really exist Thetic sentences, emphasizes the importance of treating the information structure categories of Topic and Focus as pragmatic relations, which are, notably, not equal in value (i.e. the Focus is primary). Treating the Focus and Topic as categories, appears to slip the sentence back to the classical logic-based approaches of segmenting its explicit elements.

The support of the predication is a separate notion from that of an element of the sentence which is presupposed. In *Part 1: Section 4.3.2*, predicational support will be explored in more detail and it will be argued that it may arise from the linguistic context, or may even remain in the extra-linguistic context: it does not need to be explicitly evoked. In the same vein, the predicational support for Thetic sentences, despite the entire sentence being in Focus, exists: it comes from the linguistic context or the extra-linguistic situation. To illustrate on (47c), this means that the predicational support of the Thetic sentence '*Bill went straight home*', in which the entire sentence is in Focus, is provided by the linguistic context: the support is the reference

to something important happening, but it is not explicitly evoked in ‘*Bill went straight home*’. The subject and predicate make (47c) a clause; but it is the context that makes it possible for (47c) to have a Focus, which in turn allows the independent (syntactically-unembedded) structure to be a sentence. Whether it is the entire structure that is in Focus, or a part of it, does not pose a problem for sentential status.

The potential objection that it does, i.e. that Thetic sentences, due to their all-in-focus structure, are not predicational, is proposed the present reply that consists of (a) emphasizing that the Topic is a relation, and (b) linking their predicational support with the context. The temptation to view the absence of a Topic as a barrier to predicational status is thus overcome.

iii) Another Nuance: Subject in Focus, Regardless of Accessibility

It merits mentioning that there is a general tendency for Thetic sentences to commonly have subjects with “unidentifiable or otherwise highly inaccessible referents” (Lambrecht, 1994: 168). However, this general trend must not be mistaken for a defining feature.

It is not enough for a subject to simply be *inaccessible* for it to be considered in Focus. This is made obvious in the following case from Lambrecht (1994: 308), which, as discussed in (41) above, may have a Topic-Comment or a Thetic reading, even though the subject appears inaccessible in both.

- (48) Speaker A: Why didn’t she come to work today?
 Speaker B: Her husband is sick.

A Thetic reading of ‘*Her husband is sick*’ is acceptable if and only if there is reason to believe that Speaker B wants to explicitly stress that it is particularly the husband’s sickness, as opposed to someone else’s, that is keeping the woman from coming to work. Such evidence would be provided if it were established in the context that the woman works as a doctor, and the fact that some unspecified person is sick would not be perceived as an explanation for her absence. Thus, the entire sentence is shown to be in Focus, not just the predicate ‘*is sick*’. Otherwise, as explained above, the sentence would be considered Topic-Comment, in spite of the inaccessibility of the subject.

Neither is an *active* subject enough to rule out that the entire sentence is *not* in Focus, i.e. to rule out the Thetic interpretation. Nothing stops an active referent from being inside a Focus domain. For example, Lambrecht (1994: 39) analyzes the sentence in (49) ‘*Here he*

comes' where '*he*' refers to a cat that has just entered into the room while the two people are speaking about this cat.

(49) {During a conversation about the referent of '*he*'.}

Here he comes.

The pronominal '*he*' reveals that the cat is an established Topic in the conversation. Nevertheless, the sentence is analyzed as Thetic. This is because the sentence is not merely predicating something about the established Topic. The sentence is indicating 'the passage of the referent from the text-internal to the text-external world', i.e. the relation between the cat's text-external presence and the conversation is not predictable. The Focus is not on the predicate '*comes*' but on the entire sentence '*here he comes*', since the event of his, or anyone's, entrance into the room could not be predicted or presupposed. As a result, the sentence is considered Thetic-Presentational.

The above examples show that, in spite of the trend of Thetic subjects to have certain properties, whether something is or is not in Focus is not determined by its identifiability and activation status:

Focus is free with respect to identifiability and activation. (Lambrecht, 1994: 263)

What matters for Thetic sentences is not so much the accessibility of the individual elements. Rather, what matters is that the Focus includes the entire sentence, as opposed to merely a part of it.

4.3.1.4.3.2 *Illustration*

A typical example of a Thetic sentence is illustrated in (50). The example was used in (37) to show how the subject differs from a Topic. The subject i.e. '*wife*' in English and '*супруга*' (*suprugá*; lit. '*wife*') in Russian, does not have a Topic relation to this sentence, i.e. it cannot be presupposed to be in the mind of the hearer at the time of the utterance, nor can the rest of this sentence. The influential person's wife is as much in focus as the quality that is predicated of her.

(50) a. Russian:

[Раз много лет тому назад, говорю одному влиятельному даже лицу: «Ваша супруга щекотливая женщина-с», – в смысле то-есть чести, так сказать, нравственных качеств, а он мне вдруг на то: «А вы её щекотали?»]

Ваша	супруга	щекотливая	женщина-с
<i>vaša</i>	<i>supruga</i>	<i>ščekotlivaja</i>	<i>ženščina-s</i>
PRO.2PL	NN.NOM	ADJ.NOM	NN.NOM-PART
<i>your</i>	<i>wife</i>	<i>ticklish</i>	<i>woman</i>

b. English:

[Once, this was many years ago now, I said to an influential person, “Your wife, sir, is a ticklish woman,” referring to her honor, her moral qualities, so to speak. And he suddenly retorted, “Did you tickle her?”]

Your wife, sir, is a ticklish woman

Example (51) illustrates another Thetic-Presentational structure. The sentence is used to introduce into the conversation the female pointed out by the speaker, i.e. the referent that comes from far away and is marked by the pronoun ‘*she*’ in English, and, in Russian, the contextually accessible female explicitly characterized as ‘далекая’ (*dalekaja*; lit. *distant_one*.ADJ.F.NOM).

(51) a. Russian:

[Многие из теснившихся к нему женщин заливались слезами умиления и восторга, вызванного эффектом минуты {...} «А вот далекая!» указал он на одну еще вовсе не старую женщину, но очень худую и испитую, не то что загоревшую, а как бы всю почерневшую лицом.]

А	ВОТ	далекая!
<i>a</i>	<i>vot</i>	<i>dalekaja</i>
CONJ	PART	ADJ.F.NOM
<i>and</i>	<i>here</i>	<i>distant_one</i>

b. English:

[Many of the women who pressed towards him were shedding tears of tenderness and rapture, called up by the effect of the moment {...} “But she comes from far away!” He pointed at a woman who was not at all old yet but very thin and haggard, with a face not tanned but, as it were, blackened.]

But she comes from far away!

Similarly, the Focus also extends over the entire sentence in (52). The example shows the topical referent of the element ‘*your faith*’ in English, ‘*вашей веры*’ (*vašej very*; lit. ‘*your faith*’) in Russian, being included in the Focus domain.

(52) a. Russian

[«Да, вот вы тогда обедали, а я вот веру-то и потерял!» поддразнивал Федор Павлович.]

Какое	мне	дело	до	вашей	веры!
<i>kakoe</i>	<i>mne</i>	<i>delo</i>	<i>do</i>	<i>vašej</i>	<i>very</i>
PRO.NOM	PRO.DAT	NN.NOM	PREP	PRO.GEN	NN.GEN
<i>what</i>	<i>me</i>	<i>business</i>	<i>for</i>	<i>your</i>	<i>faith</i>

b. English

[“So you were having dinner then, and I just lost my faith!” Fyodor Pavlovich went on teasing him.]

What do I care about your faith!

4.3.1.5 Summary of Information Structure

The present section has attempted to summarize the key details of Lambrecht’s (1994) definitions of information structure which have been adopted in the present study. It is significant that Lambrecht’s sentence types are defined based on a correspondence between the sentence structure and communicative function. This link makes it possible to analyze the syntactic form of a sentence in connection with its meaning. Lambrecht summarizes the important advantages of such an analysis in the following:

One advantage of my approach is that it offers a way out of the ‘segmentation’ problem [...] by identifying focus domains with major syntactic and semantic categories.

Another advantage is that it makes it possible to capture semantic correspondences between formally divergent but functionally identical sentences across or within languages. (Lambrecht, 1994: 221)

Indeed, Lambrecht’s information structural framework presents several major advantages, amongst which is its emphasis on treating the Topic and the Focus as relations,

and the consequent freedom to break with binary segmentation models of the sentence which are based on the presence of certain categories (i.e. the necessity of explicit subject and predicate, or, alternatively, explicit theme and rhema). The sentence becomes a piece of information made possible through the presence of a contextually gained pragmatic Focus relation.

Another important advantage is in terms of what the framework brings for cross-linguistic analysis. As mentioned by Lambrecht in the second part of the above quote, the approach maintains *cross-linguistic stability in terms of communicative function*, in order to study the ways that formally divergent languages express these functions language-specifically.

This advantage may be evidenced in the illustrative Russian and English examples in the present section. In addition to exposing the way that the major information structure concepts apply in both Russian, in part (a) of the example, and English, in part (b), the sentences selected for the examples in *Section 4.3.1* are purposefully those that are verbless in part (a) of the example and have a verbal translation correspondence in part (b). The examples are thus syntactically divergent not only in terms of their formal language-specific structure, but also in the fact that the verbless structures in (a), on traditional definitions, would not be recognized as sentences, while their functionally and semantically (near) equivalent translation correspondence in (b) would be recognized as expressions of a complete thought by any standard of any sentential model.

As may be noted in these examples, both the structures in (a) and their translation correspondences in (b) maintain the context and, consequently, the communicative function which is driven by that context. This is true despite the Russian structures in (a) being verbless and the English structures in (b) being verbal. The pragmatic context, which reveals whether a referent is accessible or unpredictable to the interlocutors, does not change depending on whether they are speaking Russian or English. This cross-linguistic stability of communicative function is integrated as a foundational element of Lambrecht's information structure model. In other words, it becomes an open research question how the cross-linguistically stable communicative functions, including contextually revealed Focus and Topic relations, are expressed in syntactically divergent languages.

This question, as well as some of the examples from the present section, will be returned to with regard to the information structure of verbless sentences, their translation correspondences, and the communicative function of the verb that is pragmatically implicated in translation. Though in a verbless sentence, it may be tempting to consider the omission of

the verb as corresponding to its predictability, *Part 1: Chapter 3* aimed to show that its identification and reconstruction, whether syntactic or semantic, is subject to serious doubt; furthermore, in the results (*Part 4: Section 4.3*), evidence will be presented that suggests that a pragmatically implicated verb is also not predictable from an information structural perspective.

It appears that for the notion of the ‘sentence’ to remain relevant with regard to independent verbless structures, neither hidden structure, nor semantic reconstruction, nor pragmatic predictability will do.

4.3.2 Predicational Support

In addition to providing the focus, contextual insertion also resolves another important element: the predicational support. This was mentioned in the discussion of Thetic sentences above (*Part 1: Section 4.3.1.4.3.1.ii*). In the present section, the extra-linguistic and linguistic support of predication will be given more detailed attention and illustrated in both Russian and English using examples from a sub-part of the present corpus.³⁶

4.3.2.1 The Explicitness Debate

The actual level of linguistic explicitness required for predication is the subject of a heated debate. An utterance such as ‘*This borsch is delicious*’ is an uncontroversial case of predication. Here, the predicate ‘*is delicious*’ attributes a quality to the explicit subject ‘*borsch*’. The referent of the word ‘*borsch*’ is the **support** for the attribute ‘*delicious*’; it is what the sentence is about in the logical sense. However, utterances with non-referential *dummy subjects* (e.g. ‘*it*’ in ‘*It’s borsch that I love*’), *one-word utterances* (e.g. ‘*Delicious*’ said while pointing at the borsch), as well as utterances which *depend on other utterances* in the dialogue (e.g. ‘*Borsch*’ said in response to ‘*What would you like to eat?*’) are controversial. It

³⁶Unless otherwise stated, the illustrative examples in *Part 1: Section 4.3.2* and its sub-parts are from the corpus parts BK_RO (Russian original ‘Братья Карамазовы’) and BK_E1 (English translation ‘The Brothers Karamazov’), which are presented respectively in part (a) and part (b) of the examples.

is sometimes insisted that all parts of the predication must be linguistically present to constitute a sentence; this perspective puts into question the existence of support for such utterances and thus their predicational status. This perspective is important to address as many verbless sentences belong to these types.

The controversy is resolved once it is acknowledged that the notion of predication involves not simply the structure in isolation, but the structure in combination with its utterance context. In their discussion of children's utterances, Laurent Danon-Boileau and Aliyah Morgenstern (2009) concentrate precisely on the above-mentioned controversial types of utterances. They respond to the question of whether or not the utterances involve predication by stating that:

La réponse ne dépend que de l'acceptation qu'on aura de la notion de prédication et si on veut que la prédication soit explicite dans la production verbale ou si l'on considère que **l'énoncé, le contexte linguistique, le contexte extra-linguistique, sans oublier le posturo-mimo-gestuel et l'intonation** forment un tout. (Danon-Boileau & Morgenstern, 2009: 59, emphasis added)

Indeed, it is not clear why it should be insisted that structures be analyzed for predicational status in isolation from their linguistic and extra-linguistic context (including gestures and intonation). As emphasized by Danon-Boileau and Morgenstern (2009: 60), the explicit linguistically produced elements may receive the support for predication from the situational context.

Danon-Boileau and Morgenstern (2009) discuss examples from children; however, the types of structures they discuss may also be used by adults. Although adults, unlike young children, are definitely capable of being more linguistically explicit, there is no reason to believe that adults limit themselves to making linguistically explicit the support of their predication. The following examples illustrate extra-linguistic predicational support, as well as cases in which the support comes from the linguistic context, on sentences with dummy-subjects, one-word sentences, and other types of structures.

4.3.2.2 Predication with Dummy Subjects & Extralinguistic Support

To start, consider the English utterance '*It's precisely the time*', in (53). The word '*it*' is a *non-referential dummy subject*, which, being empty, cannot serve as the support of the

predication: it cannot be what the utterance is about. Instead, the support is the *situation* in which a clock has struck the appropriate hour.

(53) a. Russian

[Пробившие часы помогли начать разговор. Ударило скорым боем на дешевых маленьких настенных часах с гирями ровно двенадцать. «Ровнешенько настоящий час,» вскричал Федор Павлович]

Ровнешенько	настоящий	час
<i>rovnešen'ko</i>	<i>nastojaščij</i>	<i>čas</i>
ADV	ADJ.MS	NN.MS
<i>precisely</i>	<i>real</i>	<i>time</i>

b. English

[The chiming of the clock helped to start conversation. A cheap little wall clock with weights rapidly struck twelve. "It's precisely the time," cried Fyodor Pavlovich]

It's precisely the time.

The predicate '*is precisely the time*' is supported not linguistically, but implicitly by the situation which is in the extra-linguistic context. In this sense, this sentence is similar to '*It's raining*', the French version of which is analyzed in a similar fashion by Danon-Boileau and Morgenstern (2009).

In the Russian sentence (53a), the absence of a dummy-subject is not surprising. An equivalent for the semantically-empty dummy subject, such as the English '*it*' or existential '*there*', or the French '*il*' or '*ça*', does not exist in Russian (Guiraud-Weber, 2009: 85); nor is there a structural need for it. The presence of the dummy '*it*' is structurally, not semantically required in English.³⁷

Margarite Guiraud-Weber stresses that a Russian sentence does not require an explicit subject and can function perfectly well without one (2009: 81). She argues that it is time to

³⁷ Lambrecht (1994) says the following regarding the primarily syntactic function of the English subject:

Typologically, English presents itself as an example of extreme 'subject prominence' (Li & Thompson, 1976), i.e. as a language in which a great variety of semantic and pragmatic functions may be associated with the invariant syntactic function of subject and in which order is to a large extent grammatically and not pragmatically controlled. (Lambrecht, 1994: 24)

stop describing Russian as a language where the subject is mandatory ('à sujet obligatoire'), and points out that doing so would facilitate understanding of why Russian sentences often omit the subject and why the language does not need its semantically empty form:

permettrait de comprendre pourquoi tout sujet désémantisé ou récupérable par ailleurs s'efface si facilement et pourquoi le russe n'a besoin d'aucun sujet 'vide' ou 'apparent' [...] Cependant l'absence d'un sujet ne signifie pas l'absence de support à la prédication. (Guiraud-Weber, 2009: 85)

The insistence that 'the absence of a subject does not signify the absence of predicative support' is echoed in the present example. The predicative support for both the subject-less Russian sentence (53a), and for the English sentence carrying a semantically empty dummy-subject (53b), is found in the extra-linguistic situation. Both sentences are thus justified instances of predication.

There is another aspect that is worth drawing attention to in the English example in (53b). Although the English sentence contains the verb '*be*' this verb is as semantically empty, in this case, as the dummy-subject '*it*'. Furthermore, on some accounts, the main syntactic predicate is not '*is precisely the time*', but rather, the main predicate is what follows the two English semantically empty but syntactically necessary elements, i.e. '*precisely the time*'. This is the perspective of Kees Hengeveld (1992), for whom this English sentence would constitute an instance of 'non-verbal predication' due to the finite verb being a copula.

Without appealing to the extra-linguistic context, the predicational status of even the English canonical finite-verb clause when it uses a dummy subject is clearly put into question: in addition to lacking a lexical verb (which means that the meaning of the verb in the syntactic predicate is limited to only the inflectionally marked tense and aspect categories), it would lack the fundamental support necessary for predication (i.e. it would be saying something about a semantically empty element, basically about nothing). Notably, the appeal to context is necessary not only for the predicational support of verbless structures, but also for canonical clauses such as (53b) if we wish to maintain their predicational status. The example illustrates that in utterances where the English subject is semantically empty, the predication takes place thanks to the implicit support provided by the extra-linguistic situation.

4.3.2.3 Predication in One-Word Utterances & Extralinguistic Support

The predication in independent one-word utterances is similarly justified by support from the extra-linguistic context. Danon-Boileau and Morgenstern (2009) discuss an example of a fifteen-month-old child uttering (54) ‘*Baba*’ upon hearing the footsteps of her caretaker Barbara walking in the corridor.

(54) {Upon hearing the footsteps of her caretaker, Barbara, a child utters:}

Baba.

As Danon-Boileau and Morgenstern (2009) convincingly argue, the one-word structure uttered by the child is in fact an instance of predication, not to be brushed off as an instance of simple ‘labelling’.

In their argument, Danon-Boileau and Morgenstern (2009) reject the proposal that one-word utterances are merely labels, i.e. that they merely name some element in the overall context and do not involve any predicative relation. From the latter perspective, these utterances are not considered instances of predication or, consequently, sentences:

Il s’agit dès alors d’une simple « activité d’étiquetage » qui ne peut être considérée comme une production de phrase. À ce stade, il n’y a ni prédication, ni grammaire, ni phrase. ‘A label is not a sentence.’ (Danon-Boileau & Morgenstern, 2009: 60)

On such accounts, an independent one-word structure would be rejected sentential status on the grounds that it is merely a label for some element in the context; i.e. identifies a stable predication-supporting element without attributing anything to it.

To the contrary, Danon-Boileau and Morgenstern argue that the utterance ‘*Baba*’ does not “qualify a stable property” (2009: 60, translation). They argue that ‘*Baba*’ is not the support of the predication. Rather, the utterance ‘*Baba*’ is to be treated as the predicate, in the classical logic sense of the term (i.e. the something said), whose support (i.e. the thing that it is said about) is the situation in which the event of the noise of the steps takes place.

More specifically, they analyze the predication in ‘*Baba*’ as including two stages: first, a stage to qualify the event of the footsteps, and a second to attribute Barbara to this footsteps-event:

Quand Lauren dit ‘*baba*’ en entendant des pas, ce qu’elle qualifie n’est pas une propriété stable du réel, mais un événement : la survenue de bruit de pas. Elle fait une prédication comparable à ‘*il pleut*’. Elle dit en somme : ‘*Tiens, on marche*’. Mais elle fait également une seconde opération, puisqu’elle dit aussi ‘*c’est Barbara*’. Elle caractérise un événement et le rapporte à un pôle de stabilité (sa nounou Barbara). (Danon-Boileau & Morgenstern, 2009: 60)

Like with the previous dummy-subject example, the support of the predication in this one-word sentence is the extra-linguistic situation.

4.3.2.3.1 In the Present Framework

In some respects, the ‘A label is not a sentence’ argument echoes the rejection of the semantic ellipsis hypothesis discussed in *Part 1: Chapter 3*, i.e. the inability of single words to semantically encode propositional meanings. However, as was argued in the same section, the non-sentential status of independent one-word utterances is not a necessary consequence of the rejection of such encoding; rather, as it was there suggested, the relation between propositional meaning and explicit elements of an independent structure must be reconsidered (i.e. to be considered in terms of pragmatic implicature as opposed to semantic encoding).

Similarly, the ‘A label is not a sentence’ argument also relates to the problem raised in *Part 1: Chapter 2* concerning the analysis of the sentence in terms of classical logic categories of subject and predicate. For example, when ‘*Coffee!*’ is uttered in the context of a coffee pot forgotten on the stove, or in the context of recognizing that the contents inside a cup is the desired beverage, the classical logic account of a sentence faces the problem that ‘*Coffee*’ is the only explicit linguistic element and hence, regardless of whether it is analyzed as the support or as something that is said about the support, some additional inexplicit information is needed in order to make up for the other ‘missing’ half of the dichotomy, i.e. the respective logical subject or logical predicate. The appeal to extra-linguistic context for this element slips the structure into a truncated version of the binary sentence model; alternatively, if extra-linguistic context is not appealed to, the structure is left without one of its essential elements and would then be not simply truncated but non-sentential. Danon-Boileau and Morgenstern’s (2009) arguments provide an important clarification in this puzzle: the support of the predication is the situational context and the single-word structure ‘*Baba*’ or ‘*Coffee*’ is to be treated as something that is said about that support. However, the problem of the binary categories of the logical model remains even with such a clarification. This is one of the reasons

that the framework of the present thesis insists on the presence of a Focus relation which is gained through contextual insertion of the independent structure.

4.3.2.3.2 Illustration

In the same vein, the support of the predication in the following independent structures is analyzed as coming from the extra-linguistic situation.

4.3.2.3.2.1 *Vocatives*

The uttered independent one-word vocative in (55), ‘*Teacher!*’ in English, ‘Учитель!’ (*učitel’*; lit. ‘*teacher*’) in Russian, is an instance of predication in which the speaker attributes the role of a teacher to the addressee in the extra-linguistic context.

(55) {This utterance is part of a multi-party conversation where the speaker, Fyodor Pavlovich, is responding to something the elder has just said.}

a. Russian

[Вот потому я и шут, от стыда шут, старец великий, от стыда. От мнительности одной и буяню. Ведь если б я только был уверен, когда вхожу, что все меня за милейшего и умнейшего человека сейчас же примут, - Господи! какой бы я тогда был добрый человек! Учитель! – повергся он вдруг на колени, - что мне делать, чтобы наследовать жизнь вечную? – Трудно было и теперь решить: шутит он или в самом деле в таком умилении?]

Учитель!
učitel’
NN.MS.NOM
teacher

b. English

[“That’s why I am a buffoon, I’m a buffoon out of shame, great elder, out of shame. I act up just because I’m insecure. If only I were sure, when I came in, that everyone would take me at once for the most pleasant and intelligent of men – oh, Lord! what a good man I’d be! Teacher!” he suddenly threw himself on his knees, “what should I do to inherit eternal life?” It was hard to even now tell whether he was joking or was indeed greatly moved.]

Teacher!

The predication is supported by the extra-linguistic situation in which a particular person is being addressed. Notably, it does more than evoke the addressee's name, and more than draw the attention of the interlocutor. By means of this utterance the speaker attributes a property (i.e. the role of a teacher) to the referent in the extra-linguistic context.

Similarly, the independent vocative in example (56) shows that what appears to be a simple utterance of an addressee's name is also an instance of predication.

(56) a. Russian

[– Катерина Ивановна присылает вам через меня вот это, – подала она ему маленькое письмецо. [...]

– Хорошо, я приду, – решил Алеша, пробежав коротенькую и загадочную записочку, в которой, кроме убедительной просьбы прийти, не было никаких пояснений.

– Ах, как это с вашей стороны мило и великолепно будет, – вдруг, вся одушевясь, вскричала Lise. – А я ведь маме говорю: ни за что он не пойдет, он спасается. Экой, экой вы прекрасный! Ведь я всегда думала, что вы прекрасный, вот что мне приятно вам теперь сказать!

– Lise! – внушительно проговорила мамаша, впрочем тотчас же улыбнулась.]

Lise!

b. English

[“Katerina Ivanovna sends you this by me.” She handed him a small letter. [...]

“Very well, I’ll go,” Alyosha decided, glancing through the short and mysterious note, which, apart from an urgent request to come, contained no explanations.

“Ah, how nice and splendid it will be of you,” Lise cried with sudden animation. “And I just said to mother: he won’t go for anything, he is saving his soul. You’re so wonderful, so wonderful! I always did think you were wonderful, and it’s so nice to say it to you now!”

“Lise!” her mama said imposingly, though she immediately smiled.]

Lise!

Taking context into account reveals that this is not an instance of simply stating predicational support. The utterance of ‘Lise’ here is not the support for predication, though in other circumstances it may be used as such; here, the utterance is the predication. The predication ‘Lise!’ is supported by the situation in which Lise has just said something which Lise’s mother believes merits scolding. Though no particular propositional meaning, such as for instance ‘*you shouldn’t say that*’ or ‘*you’re behaving badly*’ or some other linguistically propositional form

of scolding, is semantically encoded into this predication, an abstract ‘mentalese’ proposition of scolding is pragmatically implicated by means of the predication.

4.3.2.3.2.2 *Insults*

Although the following examples consist of more than one-word, they are similar to the one-word sentences just examined in that the predication relies on a support that is found in the extra-linguistic context.

(57) {Fyodor Pavlovich and Dmitry Fyodorovich are arguing and exchanging insults. }

a. Russian

[– Митя! Митя! – слабонервно и выдавливая из себя слезы, вскричал Федор Павлович, – а родительское-то благословление на что? А ну проклянц, что тогда будет?
– Бесстыдник и притворщик! – неистово рявкнул Дмитрий Федорович.]

Бесстыдник	и	притворщик!
<i>besstydnik</i>	<i>i</i>	<i>prityvorščik</i>
NN.MS.NOM	CONJ	NN.MS.NOM
<i>shameless_man</i>	<i>and</i>	<i>pretender</i>

b. English

[“Mitya! Mitya!” Fyodor Pavlovich cried tremulously, trying to squeeze out a tear. Don’t you care about a father’s blessing? And what if I should curse you?”
“Shameless imposter!” Dmitry Fyodorovich roared in a fury.]

Shameless imposter!

(58) {Miusov is criticizing Ivan for accepting the dinner invitation. }

a. Russian

[Миусов с ненавистью посмотрел на Ивана Федоровича.
«А ведь идет на обед как ни в чем не бывало! – подумал он. – Медный лоб и карамазовская совесть.»]

Медный	лоб	и	карамазовская	совесть!
<i>mednyj</i>	<i>lob</i>	<i>i</i>	<i>karamazovskaja</i>	<i>sovest’</i>
ADJ.MS.NOM	NN.MS.NOM	CONJ	ADJ.FS.NOM	NN.FS.NOM
<i>brazen</i>	<i>forehead</i>	<i>and</i>	<i>Karamazov</i>	<i>conscience</i>

b. English

[Miusov looked at Ivan Fyodorovich with hatred.

“He goes off to dinner as if nothing had happened!” he thought. “A brazen face and a Karamazov conscience.”]

A brazen face and a Karamazov conscience.

(59) {The Superior bowed deeply to Fyodor Pavlovich.}

a. Russian

[– Те-те-те! Ханжество и старые фразы! Старые фразы и старые жесты! Старая ложь и казенщина земных поклонов! Знаем мы эти поклоны! «Поцелуй в губы и кинжал в сердце», как в «Разбойниках» Шиллера. Не люблю, отцы, фальши, а хочу истины!]

«Поцелуй	в	губы	и	кинжал
<i>poceluj</i>	<i>v</i>	<i>guby</i>	<i>i</i>	<i>kinžal</i>
NN.MS.NOM	PREP	NN.PL.ACC	CONJ	NN.MS.NOM
<i>kiss</i>	<i>to</i>	<i>lips</i>	<i>and</i>	<i>dagger</i>

в	сердце»,	как	в	«Разбойниках»	Шиллера.
<i>v</i>	<i>serdce,</i>	<i>kak</i>	<i>v</i>	<i>Razbojnikax</i>	<i>Šillera</i>
PREP	NN.NS.ACC	CONJ	PREP	NN.MPL.PREP	NN.MS.GEN
<i>to</i>	<i>heart</i>	<i>as</i>	<i>in</i>	<i>Robbers</i>	<i>Shiller's</i>

b. English

[“Tut, tut, tut! Humbug and old phrases! Old phrases and old sentences! Old lies and conventional bows. We know these bows! ‘A kiss on the lips and a dagger in the heart,’ as in Schiller’s *Robbers*. I don’t like falseness, fathers, I want the truth!”]

‘A kiss on the lips and a dagger in the heart,’ as in Schiller’s *Robbers*.

The function of insults has been identified by Catherine Chauvin (2009) as being particularly relevant for English verbless sentences.

The English and Russian insults in (57) ‘*Shameless imposter!*’ and ‘Бесстыдник и притворщик!’ (*besstydnik i pritvorščik*; lit. ‘*shameless_man and pretender*’) are expressed as verbless sentences for which support is found in the situational context. Here, the predication consists of a certain quality being attributed to an individual, i.e. the addressee, in the extra-linguistic context.

Similarly, example (58) characterizes an individual in the extra-linguistic context using a more elaborate combination of noun phrases ‘*a brazen face and a Karamazov conscience*’ in

English and ‘медный лоб и карамазовская совесть’ (*mednyj lob i karamazovskaja sovest’*; lit. ‘*brazen forehead and Karamazov conscience*’) in Russian.

Example (59) characterizes not an individual, but an extra-linguistic situation in which the addressee bows to the speaker. Notably, the Russian utterance ‘поцелуй в губы и кинжал в сердце’ (*poceluj v guby i kinžal v serdce*; lit. ‘*kiss to lips and dagger to heart*’) is similar in its syntactic structure to the above-discussed example (20) ‘он в больницу’ (*on v bol’nicu*; lit. ‘*he to hospital.ACC*’) and receives a verbless translation in the English version of the insult.

Despite the differing syntactic complexity, the illustrated structures acquire support and receive their predicational status in the same way, i.e. through the extra-linguistic context.

4.3.2.4 Predication Supported by Linguistic Context

The examples in the previous section happened to be supported by the extra-linguistic context. The linguistic context also plays an important role in justifying predication in independent one-word utterances and other utterances in which, on first glance, the essential support of the predication appears to be missing.

This may be evidenced in the one-word sentence and its linguistic context in (60):

(60) a. Russian

[– Un chevalier parfait! – и помещик пустил на воздух щелчок пальцем.
– Кто это chevalier? – спросил Миусов.
– Старец, великолепный старец, старец... Честь и слава монастырю.
Зосима.]

Зосима.
Zosima
NN.NOM
Zosima

b. English

[“Un chevalier parfait!” And the landowner loosed a snap of his fingers into the air.

“Who is a chevalier?” asked Miusov.

“The elder, the splendid elder, the elder... The honour and glory of the monastery. Zosima.]

Zosima.

The one-word independent utterance ‘*Zosima*’ is an instance of predication which depends on the previous question ‘*Who is a chevalier?*’ for the supporting element; similarly, the Russian one-word predication ‘*Зосима*’ (*Zosima*) depends on the prior linguistic context, ‘*Кто это chevalier?*’ (*kto èto chevalier*; lit. ‘*who this chevalier*’), for its support. The support is the identity of the chevalier. In other words, the answer ‘*Zosima*’ provides the referent inquired about without making explicit the support of the predication, i.e. ‘*chevalier*’. Notably, the identity relation between ‘*chevalier*’ and ‘*Zosima*’ is also taken for granted in the answer. (The presence of the copula in the English question makes the English answer a case of syntactic stripping ellipsis, as will be discussed shortly.) Thus, the one-word sentence ‘*Zosima*’, in English, and ‘*Зосима*’ (*Zosima*), in Russian, is a verbless sentence whose predicational support is found in the linguistic context.

While in (60), the support of the predication for both the Russian and the English sentence is in the linguistic context, it is worth noting that this is not always the case for translation correspondences. Example (61) illustrates an instance where the predicational support is in the extra-linguistic context for the Russian sentence in (a), whereas in the corresponding English sentence in (b), it is made explicit in the sentence itself.

(61) {A woman tells a story about herself to the elder. The elder has listened to her and now asks her the following question. }

a. Russian

Издалека?

izdaleka

ADV

from_far_away

b. English

Have you come from far away?

The predication in the Russian sentence (a) ‘Издалека?’ (*izdaleka*; lit. ‘*from_far_away*.ADV’) is supported by the extra-linguistic context (i.e. the woman’s presence). In the English example in (b) ‘*Have you come from far away?*’, the support of the predication (i.e. the referent woman) is made linguistically explicit with the subject ‘*you*’.

Furthermore, in Russian, the extra-linguistic contextual support is carried over into the answer, (62a) ‘За пятьсот верст отселева’ (*za pjat’sot verst otseleva*; lit. ‘*over five_hundred versts from here*’), which immediately follows the question in (61a). The Russian question in (61a) is a one-word predication expressed as an adverb; as such, it provides minimal linguistic support for the answer that follows.

(62) a. Russian

[– Издалека?
– За пятьсот верст отселева.]

За	пятьсот	верст	отселева.
<i>za</i>	<i>pjat’sot</i>	<i>verst</i>	<i>otseleva</i>
PREP	NUM	NN.PL	ADV
<i>over</i>	<i>five_hundred</i>	<i>versts</i>	<i>from_here</i>

b. English

[“*Have you come from far away?*”
“*Over three hundred miles from here.*”]

Over three hundred miles from here.

In contrast, the English answer in (62b), though verbless, receives the support of its predication from the linguistic context, i.e. the question in (61b). Notably, the example suggests that English is perhaps less comfortable with the extra-linguistic context being used in predication over distances.

4.3.3 The Sentence

The discussion in *Part 1: Section 4.3* has emphasized the fundamental importance of contextual insertion for the sentential status of an independent syntactic structure, both in terms

of receiving an information structural Focus (*Section 4.3.1*) and in terms of predicational support (*Section 4.3.2*).

The present requirement of contextual insertion, in addition to that of the initial and final marking (*Part 1: Section 4.2*), constitute the necessary and sufficient conditions for the notion of the ‘sentence’. These conditions strive to resolve the challenges to the notion of a sentence that were presented in *Part 1: Chapters 2 and 3*, which are so serious that they have led some to abandon the notion of a ‘sentence’ as something that is relevant to more than canonical syntactic clauses.

Notably, the present conditions on the sentence hold for both verbal and verbless structures. The following section will focus on the absence of the verb from such a sentence, further define the notion of a verbless sentence, and propose two fundamental types: ‘elliptical’ and ‘non-elliptical’.

4.4 ABSENCE OF VERB

It is now possible to be more precise about the intentions behind the term ‘verbless sentence’, which is the object of the present investigation. In addition to meeting sentential conditions, i.e. that a syntactic structure is uttered in a context and that it has initial and final marking, the particularly *verbless* sentence is characterized by the *absence of any finite and non-finite verb form*. This restriction has two nuances.

4.4.1 Finite & Non-finite: Verbless as a Subtype of Predicate-less

This last requirement means that sentences which include a verb with a syntactically non-predicative function, such as an infinitive, a participle or a gerund, are treated as *verbal* sentences. Although all verbless sentences are necessarily ‘predicate-less’ (i.e. they necessarily do not include a predicate in the traditional sense of the term, which would require a finite verb), the term ‘verbless’ does not encompass all of the different types of sentences without a syntactic predicate. In other words, the verbless sentence is a subtype of the predicate-less

sentence. For instance, the term *predicate-less* would apply not only to *verbless* sentences such as those in (63a), but also to *verbal* sentences such as (63b).

(63) a. *Verbless and Predicate-less Sentence*

- i. {Speaker pulls an envelope out of the mailbox and says in Russian:}

Из	Франции.
<i>iz</i>	<i>francii</i>
PREP	NN.GEN
<i>from</i>	<i>France</i>

- ii. {Speaker pulls an envelope out of the mailbox and says in English:}

From France.

b. *Verbal and Predicate-less Sentence*

- i. {The speaker is leaving a party because he has to get up early the next morning. He interrupts the ongoing conversation in order to explain his departure and says in Russian:}

Мне	спать.
<i>mne</i>	<i>spat'</i>
DAT	INF
<i>me</i>	<i>sleep</i>

- ii. {The speaker is leaving a party because he has to get up early the next morning. He interrupts the ongoing conversation in order to explain his departure and says in English:}

Going	to	bed.
PTCP	PREP	NN

Though the sentences in (63b), ‘*Going to bed*’ in English and ‘Мне спать’ (*mne spat'*; lit. ‘*me sleep*’) in Russian³⁸, do not include a predicate, they are nevertheless *verbal sentences* since they contain a verb form, i.e. the participle ‘*going*’ in English and the infinitive ‘спать’ (*spat'*;

³⁸ In the Russian structure ‘Мне спать’ (*mne spat'*; lit. ‘*me.DAT sleep.INF*’) there is neither a syntactic predicate (i.e. no finite verb), nor a syntactic subject (i.e. no noun in the nominative case). Thus, it would not be considered a canonical syntactic clause. On the present account, this example is nevertheless *sentential* because it meets the two necessary criteria: (a) initial and final marking (which indicates that it is a syntactically independent structure), and (b) it is uttered in a context (thus, it has predicational support, i.e. the situation in which the speaker is leaving, and a Focus, i.e. the infinitive). In terms of meaning, this sentence, being an impersonal one (due to the dative subject), suggests an objective presentation of the fact that speaker must go to sleep. The meaning of the Dative-Infinitive type of structure in Russian, and particularly its link to obligation, is thoroughly explored in a recent thesis dissertation by Bastien Poreau (2020).

lit. ‘*sleep*’) in Russian, and meet the necessary sentential conditions. (Notably, having neither a syntactic predicate, nor a syntactic subject, both structures in (63b) would not be considered syntactic clauses; yet, they may be analyzed for propositional meaning since they meet the above-proposed sentential criteria.)

4.4.2 Entirely Verbless: Compound

The second nuance concerns the compound nature of the notion of the sentence. In particular, as discussed in *Part 1: Section 4.1.2*, the sentence is treated as what Palmer (1974: 11) calls a ‘matrix of constituents’, in the sense that it may encompass multiple syntactic constituents, including multiple clauses. In other words, it may be made up of several sub-sentential ‘utterance units’ (Traum & Heeman, 1997). For a ‘verbless sentence’ it is necessary that the verb restriction apply to the entire sentential structure. This means that sentences consisting of multiple utterances, of which only one contains a finite-verb or a non-finite verb form, are considered verbal.

To illustrate, the following example would constitute a single verbal sentence with two utterance units.

(64) {The speaker describes plans for the day: }

a. Russian

Она	идет	в	музей,	а	я	в	магазин.
<i>ona</i>	<i>idet</i>	<i>v</i>	<i>muzej</i>	<i>a</i>	<i>ja</i>	<i>v</i>	<i>magazin</i>
NOM	PR	PREP	ACC	CONJ	NOM	PREP	ACC
<i>she</i>	<i>go</i>	<i>to</i>	<i>museum</i>	<i>and</i>	<i>I</i>	<i>to</i>	<i>store</i>

b. English

She’s going to the museum, and I to the store.

In particular, the single *verbal* sentence consists of two utterance units: a *verbal* utterance, i.e. ‘*She’s going to the museum*’ in English and ‘Она идет в музей’ (*ona idet v muzej*; lit. ‘*she go to museum*’) in Russian, and a *verbless* utterance, i.e. ‘*and I to the store*’ in English and ‘а я в магазин’ (*a ja v magazin*; lit. ‘*and I to store*’) in Russian.

As will be discussed shortly, the latter utterance is analyzed as an instance of the syntactic ellipsis of the verb. This example further develops the point made in *Part 1: Section 4.1.2*, concerning the importance of the distinction between the utterance unit and the sentence. While for the analysis of ellipsis it is the utterance unit that is crucial, the verbless sentence requires initial and final marking and contextual insertion.

4.5 TWO FUNDAMENTAL TYPES OF VERBLESS SENTENCES

The previous sections have defined a *verbless sentence* as a string of text that is found in a context, carries initial and final marking, and does not include a finite verb, or a non-finite verb form (i.e. participle, gerund, infinitive), in any of its parts. Simultaneously to maintaining what appear to be convincing arguments against the existence of hidden structure (arguments presented in *Part 1: Chapter 3*), it seems crucial to appeal to specific syntactic criteria for the analysis of verbless sentences. The present section will show that there appears to be an important contrast within the verbless sentence category which consistently divides verbless sentences into two fundamentally different types.

What is presently proposed is that within verbless sentences there exists a basic dichotomy in terms of *elliptical and non-elliptical*. As shown in *Part 1: Chapter 3*, the term ‘ellipsis’ is wrapped up in controversy and has been used in various ways, including sometimes very loosely to refer to the general phenomenon of the absence of any contextually salient element. Presently, the term is reserved strictly for cases of syntactically defined ellipsis that are based on the existence of a verbal antecedent in the linguistic context.³⁹

In short, the proposed fundamental binary distinction of verbless sentences concerns the presence or the absence of a linguistically explicit verbal antecedent. The distinction is context dependent: it requires the analysis of the linguistic context for the explicit presence of a verbal

³⁹ The current distinction is in some respects in line with, for instance, the work of Mikhail Kopotev (1999; 2007a; 2007b; 2015) and Marjorie McShane (2000). For the former, verbless sentences that are verbless due to ellipsis are distinct from ‘verbless sentences’ (e.g. Kopotev 2007b: 117). The essence of this distinction is presently maintained, although currently we propose that elliptical and non-elliptical *both constitute types of verbless sentences*. McShane (2000) discusses different licensing conditions associated with the syntactic omission of the verb, including those that are antecedent-based (and which we also currently call ellipses), as well as other types of structures that are non-antecedent based but commonly analyzed as ellipses.

antecedent prior to, or following, the verbless sentence. Furthermore, as discussed in *Part 1: Sections 4.4.2 and 4.1.2*, since a sentence may consist of multiple utterances, a single sentence may include several ellipses.

The following will illustrate more precisely what is meant by non-elliptical and elliptical verbless sentences, and the importance of the distinction.⁴⁰

4.5.1 Elliptical Verbless Sentences

More precisely, *elliptical verbless sentences* are those for which there exists a verbal antecedent in the linguistic context. In accordance with this antecedent-based requirement and the common syntactic definitions of ellipses summarized in (McShane, 2000), three types of elliptical verbless sentences are presently distinguished: *sluicing*, *stripping* and *gapping*. The following illustrates and explains the licensing conditions of the three types.

4.5.1.1 Sluicing

The sluicing type of ellipsis commonly tends to be associated with questions. In addition to the presence of a linguistically explicit verbal antecedent in the context, this ellipsis is syntactically “licensed by the question word that introduces [a] question” (McShane, 2000: 205). An elliptical verbless sentence of the sluicing type is illustrated in English and Russian in (65).

⁴⁰ The illustrative examples (65), (66), (68) and (69) are from the corpus parts BK_R0 (Russian original ‘Братья Карамазовы’) and BK_E1 (English translation ‘The Brothers Karamazov’), which are presented respectively in part (a) and part (b) of the examples.

(65) *Elliptical Verbless Sentence of the Sluicing Type*

a. Russian

{Ракитин:} – По-моему, старик действительно прозорлив: уголовщину
пронюхал. Смердит у вас.

{Алёша:} – Какую уголовщину?

уголовщину	пронюхал
<i>ugolovščinu</i>	<i>pronjuxal</i>
NN.FS.ACC	V.3MS.PS
<i>crime</i>	<i>smell</i>

Какую	уголовщину?
<i>kakuju</i>	<i>ugolovščinu</i>
PRO.FS.ACC	NN.FS.ACC
<i>what</i>	<i>crime</i>

b. English

{Rakitin:} “The old man is really astute, if you ask me: he smelled crime. It stinks in your family.”

{Alyosha:} “What crime?”

For the verbless sentences in (65), ‘*What crime?*’ in English and ‘*Какую уголовщину?*’ (*kakuju ugovovščinu*; lit. ‘*what crime?*’) in Russian, there exists a verbal antecedent in the prior linguistic context. The question word, i.e. ‘*what?*’ in English and ‘*какую?*’ (*kakuju*; lit. ‘*what?*’) in Russian, allows for the ellipsis of the embedded ‘*he smelled?*’ and ‘*пронюхал?*’ (*pronjuxal*; lit. ‘*smell.V.3MS.PS?*’), which is recoverable from the previous utterance.

4.5.1.2 Stripping

The stripping type of ellipsis centers on parallelism between two syntactic structures. McShane explains that when there exists a parallelism between structures, it is often possible to strip “all but one major lexical category in the second clause of a coordinate structure” (2000: 205). This parallelism hinges on the repetition of a single constituent. An elliptical verbless sentence of the stripping type is illustrated in (66).

(66) *Elliptical Verbless Sentences of the Stripping Type*

a. Russian

[Ступай и поспеши. Около братьев будь.
Да не около одного, а около обоих.]

Около	братьев	будь.
<i>okolo</i>	<i>brat'ev</i>	<i>bud'</i>
PREP	NN.MPL.ACC	V.2S.IMP
<i>near</i>	<i>brothers</i>	<i>be</i>

Да	не	около	одного,	а	около	обоих.
<i>da</i>	<i>ne</i>	<i>okolo</i>	<i>odnogo</i>	<i>a</i>	<i>okolo</i>	<i>oboix</i>
CONJ	NEG	PREP	MS.ACC	CONJ	PREP	MPL.ACC
<i>and</i>	<i>not</i>	<i>near</i>	<i>one</i>	<i>but</i>	<i>near</i>	<i>both</i>

b. English

[Go and hurry. Be near your brothers.
Not just one, but both of them.]

In the above, the ellipsis is syntactically licensed by the existence of a parallel verbal structure, i.e. the imperative ‘*be near x*’ in English and ‘около *x* будь’ (*okolo x bud'*; lit. ‘*near x be*’) in Russian. The verbal antecedents are said to be stripped from the verbless structures; the latter are tied to their parallel structures by means of a single element, i.e. the reference to brothers. Notably, as discussed in (33) above, in this example the stripping ellipsis occurs twice within the same sentence.

4.5.1.3 Gapping

The parallelism between two structures also allows to syntactically ‘gap’ a verb, i.e. omit a verbal antecedent leaving the rest of the sentence, or *at least two* main constituents, intact (McShane, 2000). An elliptical verbless sentence of the gapping type is illustrated in (67).

(67) *Elliptical Verbless Sentence of the Gapping Type*

a. Russian

Speaker A: Она идёт в музей.

Speaker B: А я в магазин.

Она	идёт	в	музей.
<i>ona</i>	<i>idët</i>	<i>v</i>	<i>muzej</i>
NOM	2S.PR.IPFV	PREP	ACC
<i>she</i>	<i>going</i>	<i>to</i>	<i>museum</i>

А	я	в	магазин.
<i>a</i>	<i>ja</i>	<i>v</i>	<i>magazin</i>
CONJ	NOM	PREP	ACC
<i>and</i>	<i>I</i>	<i>to</i>	<i>store</i>

b. English

Speaker A: “She is going to the museum.”

Speaker B: “And I to the store.”

The linguistically explicit verbal antecedent, i.e. ‘*is going*’ and ‘идёт’ (*idët*; lit. ‘*is going*’), together with the parallel structure in the linguistic context, license the syntactic ellipsis of the present continuous form of the verb ‘go’ in the English verbless sentence and ‘идти’ (*idti*; lit. ‘go’) in the Russian. In contrast with stripping, in the gapping ellipses in (67) more than one linguistic constituent is left intact and the verb’s absence corresponds to a gap in the parallel structure.

4.5.1.4 Antecedent

Example (68) illustrates stripping and gapping in English verbless questions, but not in Russian. For the latter, a verbal antecedent does not exist in the linguistic context which means that similar syntactic ellipses in (68) are not licensed. Consequently, the Russian verbless sentences in (68) are of the *non-elliptical type* according to the present definitions.

(68) a. Russian

{Алёша:} – Ах, да, я забыл, ведь она тебе родственница...
 {Ракитин:} – Родственница? Это Грушенька-то мне родственница?

ведь	она	тебе	родственница
<i>ved'</i>	<i>ona</i>	<i>tebe</i>	<i>rodstvennica</i>
PART	PRO.FS.NOM	PRO.2S.DAT	NN.FS.NOM
<i>well</i>	<i>she</i>	<i>to_you</i>	<i>relative</i>

Родственница?	Это	Грушенька-то	мне	родственница?
<i>rodstvennica</i>	<i>eto</i>	<i>Grušen'ka-to</i>	<i>mne</i>	<i>rodstvennica</i>
NN.FS.NOM	DEM	NOM-PART	2S.DAT	NN.FS.NOM
<i>relative</i>	<i>this</i>	<i>Grushenka-PART</i>	<i>to_me</i>	<i>relative</i>

b. English

{Alyosha:} “Ah, yes, I forgot, she’s your relative...”
 {Rakitin:} “My relative? Grushenka, my relative?”

The antecedent present-tense verb ‘*be*’ and the parallel verbal structure ‘*she’s your relative*’, licence a stripping ellipsis (of ‘*she*’ and the present tense ‘*be*’) in the English question ‘*My relative?*’ and a gapping ellipsis (of the present tense ‘*be*’) in the English question ‘*Grushenka, my relative?*’.

4.5.2 Non-Elliptical Verbless Sentences

A *non-elliptical type of verbless sentence* differs from an *elliptical* one by its relationship to the *linguistic context*: there does not exist a verbal antecedent in the linguistic context of a non-elliptical verbless sentence. For instance, the linguistic context of the verbless exclamative in (69) does not include a verbal antecedent.

(69) *Non-Elliptical Verbless Sentence*

a. Russian

Так точно было и с ним: он запомнил один вечер, летний, тихий, отворенное окно, косые лучи заходящего солнца (косые-то лучи и запомнились всего более) [...] и вдруг вбегает нянька и вырывает его у нее в испуге. Вот картина! Алеша запомнил в тот миг и лицо своей матери: он говорил, что оно было иступленное, но прекрасное, судя по тому, сколько мог он припомнить.

Вот	картина!
<i>vot</i>	<i>kartina</i>
PART	NN.NOM
<i>here</i>	<i>picture</i>

b. English

That is exactly how it was with him: he remembered a quite summer evening, an open window, the slanting rays of the setting sun [...] and suddenly a nurse rushes in and snatches him from her in fear. What a picture! Alyosha remembered his mother's face, too, at that moment: he used to say that it was frenzied, but beautiful, as far as he could remember.

What a picture!

In verbless sentences such as (69) there is no indication that an omission has taken place.

4.5.3 The Zero-Predicate

Where do antecedent-based elliptical verbless sentences stand with regard to additional syntactic structure? The discussion in *Part 1: Chapter 3* showed that reconstruction of a verbal predicate is a difficult argument to maintain. The present section has proposed a sentential definition that diminishes the traditional leading role of syntactic structure in the notion of the sentence (i.e. without context and initial-final marking a clause cannot be sentential). As a result, on the present account, the linguistic antecedent in the context of a verbless sentence does not alter its sentential or non-sentential status. The necessity to appeal to a hidden *verbal* structure in order to justify the verbless structure as being sentential is neutralized. In other words, elliptical verbless structures **do not require** additional hidden syntactic structure, or to be considered as omitting a predicate, in order to be sentential.

However, there does appear to be a considerable **difference** between verbless sentences for which an antecedent exists and those for which it does not. The explicit presence of a verbal antecedent (i.e. precisely one that fits the syntactic licensing conditions of sluicing, stripping or gapping), is the linguistic marker that indicates that a particular verbless sentence involves omission (i.e. is missing a particular predicate and is dependent on the context more than another sentence without an antecedent). An elliptical verbless structure may thus be analyzed as having a verb that may be recovered from its linguistic context. Notably, tense, aspect and other properties that are typically associated with the verb are also marked on the antecedent and recovered in the same way.

The necessity to acknowledge some form of recovery in antecedent-based elliptical structures is illustrated in comparing the elliptical verbless sentence of Speaker B in (70) with the non-elliptical verbless sentence of Speaker C in (71).

(70) {Bob has just walked out of the room. }

a. Russian

<i>Speaker A:</i>	Он	едет	в	город?
	<i>on</i>	<i>edet</i>	<i>v</i>	<i>gorod</i>
	PRO	V.PR.IPFV	PREP	NN.ACC
	<i>he</i>	<i>driving</i>	<i>to</i>	<i>city</i>

<i>Speaker B:</i>	В	больницу.
	<i>v</i>	<i>bol'nicu</i>
	PREP	NN.ACC
	<i>to</i>	<i>hospital</i>

b. English

Speaker A: Is he driving to the city?

Speaker B: To the hospital.

(71) {Bob has just walked out of the room.}

a. Russian

<i>Speaker C:</i>	В	больницу.
	<i>v</i>	<i>bol'nicu</i>
	PREP	NN.ACC
	<i>to</i>	<i>hospital</i>

b. English

Speaker C: To the hospital.

The commitment made by Speaker B in (70) is not the same as that made by Speaker C in (71). Although driving is not explicitly mentioned in the elliptical verbless sentence in (70), the existence of the antecedent tacitly commits Speaker B to this method of transport, as well as to the temporal and aspectual reference made by the present-tense continuous form of the verb. This is the case because the verbless structure and its linguistic context together meet the licensing conditions for an ellipsis. This suggests that the verbal predicate ‘*is driving*’ and ‘едет’ (*edet*) is, in some respects, part of, or *closely intertwined* with, the verbless elliptical sentence in (70). In this linguistic context, denying that ‘*Bob is driving*’ is presupposed by Speaker B would be difficult given the explicit linguistic presence of the antecedent.

However, if the verbless sentence ‘*To the hospital*’ or ‘В больницу’ (*v bol'nicu*) was uttered without the antecedent in the linguistic context, as illustrated by Speaker C in (71), i.e. if it did not meet the licensing conditions for antecedent-based syntactic ellipsis, maintaining speaker commitment to any particular predicate would not be justified on linguistic grounds. Even if for both interlocutors the event of ‘Bob driving’ appears to be particularly salient, for instance it may even seem to be the only contextually available option, the linguistic structures produced by Speaker B and Speaker C nevertheless carry, at minimum, a different level of speaker commitment; moreover, they may, in addition to differences in commitment, be used for entirely different pragmatic purposes and have different implicatures. Thus, while Speaker B tacitly allows the presupposition of the particular predicate antecedent in (70), in Speaker C’s non-elliptical sentence in (71), a predicate is neither explicit nor can a particular predicate be presupposed.

That said, the difference between Speaker B’s utterance of the elliptical verbless sentence in (70) and Speaker D’s explicit verbal predicate sentence in (72) must also be acknowledged.

(72) {Bob has just walked out of the room.}

a. Russian

<i>Speaker D:</i>	Он	едет	в	больницу.
	<i>on</i>	<i>edet</i>	<i>v</i>	<i>bol'nicu</i>
	PRO	V.PR.IPFV	PREP	NN.ACC
	<i>he</i>	<i>driving</i>	<i>to</i>	<i>hospital</i>

b. English

Speaker D: He is driving to the hospital.

Speaker D's verbal sentence, in (72), linguistically makes a different commitment to that of Speaker B's elliptical verbless sentence: Speaker's D commitment to the event of 'Bob driving' is explicit, Speaker B's is linguistically recovered and presupposed, while Speaker C's is a pragmatically implied potentiality. In addition to differences in commitment, each sentence may have its own semantic and pragmatic uses that are subject to investigation.

The acknowledgement of a difference between elliptical (70) and non-elliptical (71), does *not* necessitate that the elliptical verbless sentence in (70) is a verbal sentence that has the syntactic structure of (72) and had been reduced through syntactic deletion of the verb. In other words, an elliptical verbless sentence need not be considered a syntactically reduced version of a verbal sentence, since, as was shown in *Part I: Chapter 3*, it is not necessary that syntactic reconstruction take place. The **recovery** of the antecedent verb, in the case of syntactic antecedent-based ellipsis, may occur by means of a purely semantic mechanism, as suggested in the model developed by Dalrymple (2005); alternatively, it could occur via a psychological pointing mechanism, as suggested by Goldberg and Perek (2019).

Alternatively, the difference between the verbal (72) and the antecedent-based elliptical (70) could also be accounted for if it was proposed that the syntactic structure of the antecedent-based elliptical sentence includes a **Zero-Predicate**. The latter would take the place of the verbal predicate suggested by the linguistically-explicit antecedent in (70). The problem of the zero's particular position would remain (i.e. full syntactic reconstruction would require a decision on whether it should be '*ϕ_{Bob} ϕ_{is driving to the hospital}*' or '*to the hospital ϕ_{Bob} ϕ_{is driving}*' or any other alternative placements possible in English, and respectively concerning the possibilities of the placement of the antecedent in Russian; however, such precision would potentially lead to as many different meanings as there are different word-order possibilities for the same explicit sentence).

As a result, it presently seems more reasonable to consider that, if syntactic reconstruction is desired, the zero-predicate would exist as *external to the syntactic structure of the antecedent-based elliptical sentence*. This proposal is symbolically summarized in (73), using the examples from (70) and (71):

(73) a. Elliptical Verbless Sentence:

[SENTENTIAL STRUCTURE] + EXTERNAL: ZERO PREDICATE ANTECEDENT

Example:

- | | | | | |
|-----|--------------------|---|-----------|-------------------|
| i. | [To the hospital] | + | External: | ∅ he is driving |
| ii. | [В больницу] | + | External: | ∅ он едет |
| | <i>v bol'nicu</i> | | | <i>on edet</i> |
| | PREP NN.ACC | | | PRO V.PR.IPFV |
| | <i>to hospital</i> | | | <i>he driving</i> |

b. Non-Elliptical Verbless Sentence:

[SENTENTIAL STRUCTURE]

Example:

- | | |
|-----|--------------------|
| i. | [To the hospital] |
| ii. | [В больницу] |
| | <i>v bol'nicu</i> |
| | PREP NN.ACC |
| | <i>to hospital</i> |

Considered as an *external addition to the syntactic structure of the sentence*, as opposed to a part of it, the zero-predicate would not actually alter the syntactic structure of the sentence (i.e. as an external appendix it would not need to be integrated into the structure; in other words, the presence of the antecedent would not make the elliptical structure into a verbal one). However, its presence would make it possible to syntactically distinguish antecedent-based (70), non-elliptical (71) and the verbal (72).

As such, it seems appropriate that the title of ‘**Zero-Predicate Sentence**’ be used as an alternative for particularly elliptical antecedent-based verbless sentences. The ‘**Zero-**

Predicate' would be an external element that distinguishes an elliptical verbless sentence from a non-elliptical one, and, also, an elliptical-verbless sentence from a canonical clause.

The proposed elliptical (or zero-predicate) versus non-elliptical distinction makes it possible to maintain the former verbless structures as actual instances of **omission**, and simultaneously to free the latter verbless structures from overextended speaker commitments. On this account, the notion of omission requires a precise linguistic antecedent that is recoverable. In other words, the marker of the omission of the verb is *not* merely its absence, but rather the contextual *presence* of a verbal antecedent. In this way, it is aimed to clearly distinguish something that is missing, from something that simply does not exist. The distinction is made possible by the fact that the presently proposed definition of the sentence integrates verbless structures without requiring that they are treated as instances of verbal omission. The precision of this distinction seems important due to its consequences for speaker commitment.

4.6 THE MEANING OF VERBLESS SENTENCES

If for the elliptical type of verbless sentence, the existence of the antecedent resolves the predicate and with it the propositional meaning, the same cannot be said for the non-elliptical verbless sentence for which an antecedent does not exist, and, hence, speaker commitment to a specific predicate cannot even be tacitly presupposed. As a result of the problems discussed in *Part 1: Chapter 3*, rather than treating verbless sentences as carrying semantically encoded propositional meaning, it is the present perspective that the notion of conversational implicature provides a better account of verbless sentences in term of (a) their propositional meaning, and (b) the predicative relation between the elements of the sentence.

The notion of conversational implicature was famously put forth as a systematized theory by Herbert Paul Grice (1967/1975; 1989). As emphasized by Scott Soames (2008), the guiding idea behind the notion is that there is more to the meaning of a sentence than its literal semantically encoded meaning:

constraints on the use of sentences, and information conveyed by utterances of them, arise not only from their conventional meaning (the information they

semantically encode) but also from the communicative uses to which they are put. (Soames, 2008: 440)

Grice makes a fundamental contrast between something that is *said* (i.e. the semantically encoded conventional meaning of the words that were uttered) and something that is *implicated* (i.e. the meaning that goes beyond the conventional meaning of the words that were uttered). An example of this implicated ‘beyond’ may be gleaned in the following famous recommendation letter, from Grice (1989: 33).

A is writing a recommendation letter for a student who is a candidate for a philosophy job, and his letter reads as follows:

“Dear Sir,
Mr. X’s command of English is excellent, and his attendance in class has been regular.
Yours, truly.”

A implicates that his student is no good at philosophy. Why? A knows that a more informative letter is desired. Since he is also in a position to provide the needed information, there must be a reason he hasn’t put it in the letter. There is no reason to think he is being uncooperative, since if he were, he wouldn’t have written. Given that he is being cooperative, he would surely give us a positive evaluation if he had one to give. (People are more reluctant to state negative evaluations than positive ones, especially in writing.) Thus, his evaluation must be negative, and he simply doesn’t want to explicitly say so. Hence, he must think the student is no good. (Soames, 2008: 442)

More formally, conversational implicature is defined as something that arises as a result of general features of discourse; specifically, it arises as a result of what Grice calls the Cooperative Principle that participants of any conversation are expected to observe:

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. One might label this the Cooperative Principle [... and] distinguish four categories [...] the following of which will, in general, yield results in accordance with the Cooperative Principle. Echoing Kant, I call these categories Quantity, Quality, Relation and Manner. (Grice, 1967/1975: 45)

When participants cooperate in discourse, Grice posits four maxims that guide both the sentences that a speaker produces and the hearer’s interpretation of what the speaker has made linguistically explicit. The maxim of *Quantity* expects that what the speaker says will be informative, but not overly so. *Quality* predicts that it will not be something that the speaker believes to be false or something for which the speaker lacks evidence. The *Relation* maxim

presupposes that what the speaker says is relevant to the context at hand. As a general category of cooperativeness in a conversation, it can also be assumed that the speaker's *Manner* avoids obscurity and ambiguity, and is brief and orderly, as far as possible. Thus, in contrast to conventional meaning which is semantically encoded into the words of the sentence, conversationally implicated meaning arises as a result of the above general features of discourse being preserved in the specific utterance context.

In other words, it is through the combination of the asserted proposition '*Mr. X's command of English is excellent and his attendance has been regular*' with the utterance context and Grice's basic conversational maxims, that the recommendation letter conversationally implicates the proposition that '*Mr. X is a bad student*' (or was bad at philosophy, or will be bad for this position, or something along these lines), even though the latter conversationally implicated proposition is neither specified in an exact formulation, nor explicitly asserted.

Furthermore, Soames (2008) makes two important precisions regarding the distinction between conversational implicature and linguistic meaning, and explicitly links the latter to assertion:

First, on the standard picture, a conversational implicature is a piece of information conveyed, over and above what is said, or **asserted** – which is itself closely related to the meaning of the sentence uttered, and perhaps identical with what would now be called *the semantic content of the sentence in the context*.³

[Footnote 3: Like Grice, I use the indirect-discourse sense of 'say' in such a way that '*A says that S*' is essentially equivalent to '*A asserts that S*'. For both Grice and me, saying/asserting that S is a way of **committing** oneself to the truth of the claim that S, distinct from merely implicating that S. In discussing implicature he typically contrasts "what is said by S" with what an utterance of S merely implicates – making it sound as if the sentence itself says or asserts something. What he means, I take it, is (i) that what a speaker says/asserts by uttering S is to be contrasted with what the utterance merely implicates, and (ii) that what the speaker says/asserts is (in standard cases) to be identified with what is now called the semantic content of S in the context – a statement determined by disambiguating S, and settling the reference of any indexicals, time references, or other context-sensitive elements.]

Second, in order for a proposition q to count as conversationally implicated, the conclusion – that the speaker believes or accepts q, and is inviting his hearers to do the same – must, in principle, be derivable by an argument of the specified type from information available to speaker-hearers about the meaning of the sentence uttered, the context of the utterance, the conversational maxims, and other background information in the context. (Soames, 2008: 441–442, emphasis added)

Crucially, a meaning that is conversationally implicated does not carry the same commitment as one that is semantically encoded into the words. A conversational implicature, unlike semantically encoded meaning or an implicature that has been conventionalized, may be cancelled:

implication is not part of the meaning of a sentence since the speaker can cancel the implication by adding '*I do not meant to imply that...*' (Grandy & Warner, 2020: §1)

An implicature that *p* is *cancelable* if it is permissible to conjoin (to the sentence that allegedly implicates *p*) "but not *p*" or "I do not mean to imply that *p*". Example: if I say "Either *p* or *q*" I conversationally implicate that I don't know which of the two is true, just that at least one is. But I can explicitly *cancel* this by adding "*but I don't mean to imply that I don't know which is true.*" [...] On the other hand, consider our earlier example: "*She is poor, but she is honest.*" The speaker has implicated that there is a clash or contrast between (her) poverty and (her) honesty. But this implicature cannot be cancelled. For I cannot coherently say: "*She is poor but she is honest, and I don't mean to imply that there is a clash or contrast between (her) poverty and (her) honesty.*" The reason is that the implicature here is *conventional*, not conversational. The implicature is carried by the *very words the speaker uses*. In a conversational case, the implicature is carried not by the words, but by the speaker's *saying* them (or the manner in which he says them). [...] But what starts out as conversational implicature may become conventionalized. (Cohen, 2008: 7–8)

Conversational implicature explains propositional meaning that is neither semantically encoded compositionally into the individual elements, nor conventionalized to the point where a non-compositional meaning is non-cancellable.

From the present perspective, the appeal to Gricean maxims and the notion of conversational implicature appears to be applicable not only to the meaning of verbal sentences, but its explanatory power may also be extended to key questions concerning verbless sentences; it appears to provide the meaning necessary for sentential status (i.e. a propositional meaning) without introducing additional syntax or additional semantic encoding. The following will illustrate the present proposal that: (a) the propositional meaning in a verbless sentence involves conversational implicature, and (b) a predicative relation, despite the absence of a predicate, results from the relevance maxim.

4.6.1 Implicature

Слово не воробей, вылетит не поймаешь.
(*slovo ne vorobej, vyletit ne pojmaeš'*)

A word is not a sparrow, once it flies out you can't catch it again.

– *Russian proverb*

Once a word is uttered it cannot be unsaid. From the perspective that every word carries some sort of commitment, conversational implicature presents itself as a powerful tool for the speaker to mitigate and control that commitment.

The arguments that a verbless sentence neither (a) contains a verb in its syntactic structure, nor (b) semantically encodes verbal meaning (whether lexical, temporal or aspectual meaning), are presently expanded with the suggestion that (c) the addition of any element that is not already explicitly present in the sentence, including not only optional adjuncts (e.g. the addition of an adjective to give a more precise description) but also a predicate (i.e. a finite verb in order to express the relation between non-verbal elements), however salient it may be, remains a potentiality to which the speaker does not make the same level of commitment as if it were explicitly stated.

For instance, in (74), '*From France*' and 'Из Франции' (*iz Francii*; lit. '*from France*') said in the context of taking an envelope out of a mailbox, a number of propositions may be attributed to the speaker.

(74) {Speaker pulls an envelope out of the mailbox and says:}

a. Russian

Из	Франции.
<i>iz</i>	<i>francii</i>
PREP	NN.GEN
<i>from</i>	<i>France</i>

b. English

From France.

The speaker's sentence could be interpreted as for instance the following propositional meanings: '*This letter came from France*', '*The contents inside this envelope were written in*

France, ‘*The words ‘from France’ are written on the envelope*’, ‘*I am expecting this letter to be from France*’, ‘*This envelope was made in France*’, amongst others; and, of course, a series of relevant potential propositions in Russian.

Notably, the speaker does not make an explicit linguistic commitment to any particular subject, nor any particular predicate: neither a subject nor a predicate is asserted. Such an analysis is supported by the fact that the speaker is able to deny the precisions that are made by any of the potential propositions attributed to the explicit sentence. For instance, in the case of ‘*From France*’, the potential implicated proposition that ‘*This was sent from France*’ may be **cancelled** by adding (75), and the same for Russian ‘Из Франции’ (*iz Francii*; lit. ‘*from France*’) and ‘Это было послано из Франции’ (*èto bylo poslano iz Francii*; lit. ‘*this.DEM was.V.NSG.PS sent.V.NSG.PS.PTCP from France*’):

(75) a. English cancellation

I don’t mean to imply that this was sent from France; what I mean is that John, who sent this, always gives inside information about the latest news coming from France.

b. Russian cancellation

Я не имею в виду что это было послано из Франции; а то, что это письмо послал Джон, а он всегда шлет эксклюзивные инсайды и новости из Франции.

(*ja ne imeju v vidu čto èto bylo poslano iz Francii; a to, čto èto pis'mo poslal Džon, a on vsegda šlet èkskljuzivnye insajdy i novosti iz Francii*)

The speaker does, however, make an **assertion**; though not of any of the specified propositions. The extent of the speaker’s assertion is the commitment that ‘*from France*’, or ‘Из Франции’ (*iz Francii*; lit. ‘*from France*’), **is relevant** to the context. The speaker may be held accountable for the words having some sort of relevance – as illustrated in the need for the additional explanation in (75). However, anything beyond the assertion of this basic relevance is cancellable. The predicate and subject are unspecified in the speaker’s utterance, and they remain undetermined for the hearer: they are *not part of the sentence*, neither syntactically nor semantically. The commitment made by the speaker is that the linguistically explicit elements are relevant.

Thus, the propositional meanings listed above are *potential*, or what Soames (2008: 456) calls ‘*possible pragmatic enrichment*’. These potential pragmatically enriched propositions are generated through the combination of the *explicit linguistic elements* of the sentence (which

constrain the semantic content) and the *context* (in which *relevance* and the shared expectation of the other *conversational maxims* constrain the potential propositions).

Together with the potential propositional meaning, the potential *predicative relation* between the linguistically explicit elements is generated and similarly conversationally implicated using the verbless sentence. The verbless sentence does not contain a verb either as an element of its syntactical structure, nor as a semantically encoded meaning, nor as a predictable pragmatic presupposition; any predicate, and with it, temporal and aspectual meaning, remains a potentiality which may be conversationally implicated but not resolved. The context and explicit linguistic elements make it possible for a verbless sentence to implicate a propositional meaning, and to constrain that meaning, without specifying it.

It is notable that conversational implicature is not a particularity of *verbless* structures, but is also common to *verbal* structures which also rely on the context, relevance and the other conversational maxims, for the generation of a propositional meaning beyond the linguistic assertion. For a verbless sentence, such as '*From France*', the assertion, to which the speaker makes a commitment, is that the linguistic elements '*From France*' are relevant to the context. The specific predicative relation is not part of the semantically encoded assertion, but rather a part of the conversational implicature of a verbless sentence. In contrast, when a verbal sentence, such as '*Mr. X's command of English is excellent*', is used with the conversational implicature that '*Mr. X shouldn't be hired for the philosophy position*', the predicative relation specified in the uttered sentence *is* part of the semantic content that restricts the implicated proposition through its interaction with conversational maxims. The difference between a verbless and a verbal sentence thus concerns the extent of the *semantic content that restricts any conversationally implicated proposition*. Thus, it appears safe to say that a conversationally implicated propositional meaning, and a conversationally implicated predicative relation, is not a deficiency of independent verbless structures. Rather, the appeal to conversational implicature provides a further argument in support of their sentential status.

Furthermore, conversational implicature makes it possible for the speaker to control the commitment that is made with a linguistic sentence and, simultaneously, forces the hearer to participate in the derivation of the implicated proposition. For instance, the writer of the reference letter has not committed to the assertion that '*Mr. X should not be hired*', but, through linguistic expression and conversational maxims, has involved the interlocutor in the generation of a conversationally implicated proposition. For a verbless sentence, the conversational implicature includes the very predicative relation between the linguistically explicit elements. As a result, by using a verbless sentence the speaker is able to control the

commitment to a specific predicative relation (not only concerning its lexical but also its temporal and aspectual properties, the specification of which might be, for instance, unknown, unnecessary, irrelevant, or distracting from the focus of the sentence), and, furthermore, involve the hearer in the generation of the predicative relation. Thus, rather than detract from sentential status, conversational implicatures associate verbless sentences with a powerful method of the creation of propositional meaning.

4.6.2 Assertion

It is the present perspective that verbless sentences should be considered in terms of conversational implicature (i.e. that any particular propositional meaning *beyond the assertion of relevance* is conversationally implicated); however, it is important to address the perspective that what occurs in (74) actually does constitute a more precise assertion.

The perspective that a more precise assertion occurs may be gleaned from Elugardo and Stainton (2003), for whom a similar example, ‘*From France*’, is analyzed as an assertion of a specific propositional meaning:

Imagine Andrew walks into a room, holds up a cigarette, and says ‘*From France*’ to Sylvia. We think that **it is obvious that**, in this case, Andrew communicates, about the cigarette, that it is from France. **A singular proposition**. He can do this because, as will be obvious to both Andrew and Sylvia, what Andrew means clearly is not the property $\lambda x . \text{from-France}(x)$. How could he mean *that*? Thus, to treat him as co-operating, Sylvia must find a proposition meant, and the **most obvious one** is **precisely this** singular proposition. (Elugardo & Stainton, 2003: 257, emphasis added)

While Elugardo and Stainton convincingly argue that the utterance of ‘*From France*’ does not constitute merely a property, but does involve propositional meaning, and that propositional meaning is made available through the co-operative principle, they additionally make a claim that presently appears to be a controversial one: they maintain that there is a single precise proposition that is most obvious, i.e. the proposition that *the cigarette is from France*. It is the present perspective that this latter claim is not as obvious as presented.

One may reasonably wonder what controls the choice between several propositions that may be *equally* available for the same particular context in order to arrive at “precisely this singular proposition”. While, obviously, a proposition using the present-tense singular copula

and a standard word order, i.e. *'the cigarette is from France'*, would seem to provide the least enrichment of the propositional meaning, it is still not clear on what grounds this least enriched option should be adopted as *the one* that the speaker asserts: perhaps a past-tense lexical predicative relation, such as *'the cigarette came from France'*, or alternative propositions, such as *'it is written that this cigarette is from France'* or *'this and other cigarettes like it have been smuggled in from France'*, would be just as salient for both participants as the present-tense copula reading but have the advantage of being more precise.

Moreover, even if a single precise proposition is successfully chosen amongst other potential propositional meanings, it is reasonable to wonder to what extent the speaker is actually committed to it. Elugardo and Stainton maintain that the speaker is fully accountable to that one chosen meaning:

In fact, there are lots of cases in which what the speaker communicates goes well beyond the meaning of her words. This surely occurs in conversational implicature, and many would maintain that it equally occurs in metaphor, indirect speech acts, irony, etc. [...] So, we think it is plausible that the same happens here: the hearer understands the proposition which the speaker meant, even though the speaker's words do not, even in context, mean that proposition. Thus, because of pragmatics, a person (e.g. Andrew) can successfully communicate a complete proposition by uttering something subsentential, with neither the syntax nor the semantics of a sentence. [...] That said, though a pragmatic process plays a part in determining the proposition, it does not seem that Andrew **merely implicated a proposition. He asserted one.** Certainly, he could not later say, accused of lying about the origin of the cigarette:

'Actually, I made no statement at all. Neither about the cigarette, nor about anything else. Sylvia just drew inappropriate conclusions.'

This would radically misdescribe the case. Given that an assertion was made, it is very **tempting**, upon first encountering such examples, **to dismiss them as 'elliptical'**. (Elugardo & Stainton, 2003: 258–259, emphasis added)

It is notable that in the above cancellation, i.e. *'Actually, I made no statement at all...'*, Elugardo and Stainton entirely cancel any and all propositional meaning. Indeed, such a sweeping cancellation of all possible asserted propositions appears to be impossible: the speaker has clearly made an assertion. Elugardo and Stainton's cancellation provides vital proof that *an* assertion has indeed been made.

However, the cancellation above also clearly extends far beyond the proposition that has been claimed to be the single precise one chosen by the speaker and hearer; the latter actually *is* cancellable. Having uttered *'From France'*, the speaker can cancel his commitment to the

precise proposition that the origin of the cigarette is France by saying for instance (76a) or (76b):

- (76) a. I don't mean to imply that the cigarette is from France; what I mean is that I picked up *the smoking habit* from France.
- b. I don't mean to imply that the origin of the cigarette is France; what I mean is that I *bought* it there.
- c. * I don't mean to imply that I am making a statement in which '*From France*' constitutes a proposition.

While having uttered '*From France*' the speaker *cannot* cancel that *a propositional meaning* was intended, the speaker *can* cancel the *precise* proposition. In other words, while it appears true that having uttered '*From France*' the speaker cannot claim that no assertion was made, the proposition that he asserted is not necessarily about the origin of the cigarette.

As a result, while agreeing with Elugardo and Stainton that a proposition has been asserted, it is the present perspective that even a minimally enriched proposition such as the one they suggest is *overdetermined*. Rather, it is presently proposed that the proposition that is asserted with '*From France*', and to which the speaker cannot cancel commitment, is simply that: '*from France*' is *propositionally relevant to the particular context*. This is supported by the impossibility of such a cancellation, as shown in (76c). From the present perspective, the proposition suggested by Elugardo and Stainton concerning the origin, as well as other possible propositions, constitute possible pragmatic enrichments which may potentially be conversationally implicated above and beyond the basic assertion that the verbless sentence constitutes a proposition that is relevant in the particular context. As such, the speaker is committed to having asserted some unspecified relevant proposition whose only semantically encoded content consists of '*From France*' and whose propositional meaning is conversationally implicated.

Furthermore, the fact that it is conversational implicature, as opposed to assertion, that is involved in the propositional meaning, seems to neutralize the temptation talked about by Elugardo and Stainton in the above quote to dismiss independent verbless structures as 'elliptical': neither additional hidden syntactic structure (i.e. syntactic ellipsis), nor the semantic encoding of multiple meanings (i.e. semantic ellipsis), is required by a conversationally implicated pragmatically enriched proposition or by the basic assertion that the independent verbless structure implicates a relevant proposition.

It should also be mentioned that pragmatic enrichment is involved not only in implicature, but also in the meaning of assertions. The way this happens is explained by Soames (2008). Soames makes a contrast between two types of pragmatic enrichment: one that is ‘required’ in order for a proposition to be truth-evaluable, and another that is ‘optional’. For instance, in ‘*I have five Fs*’ the pragmatic enrichment is of the required type: the sentence is analyzed as not truth-evaluable without pragmatic enrichment since its semantic content does not encode whether what is meant refers to ‘*at least five*’, ‘*exactly five*’, ‘*up to five*’, ‘*at most five*’, ‘*up to five but more than two*’ (Soames, 2008: 456). For the structure ‘*I have five Fs*’ to constitute an assertion, one of these possible pragmatic enrichments must stand out as obvious and relevant and as being the speaker’s primary intention to assert. This occurs through the application of the conversational maxims:

In addition to generating standard conversational implicatures, over and above what is asserted, the maxims help determine what an utterance asserts by narrowing the class of possible enrichments to those that most effectively advance the conversation. When several enrichments are otherwise feasible, the maxims dictate that one select the strongest, most informative, and relevant propositions among them for which one has adequate evidence. In this way, the maxims of quantity, quality, and relevance play a role in determining what is asserted, and thereby contribute to the truth conditions of utterances. (Soames, 2008: 456)

Soames explains that in the case that a pragmatic enrichment is not of the required type, i.e. if the literal proposition semantically encodes everything necessary for truth-evaluability, then this proposition is counted as one of its own possible pragmatic enrichments; it becomes an assertion only when it is the one obvious and relevant proposition that is intended by the speaker:

The general constraint governing assertion is that what is asserted by a normal, literal use of a sentence *S* must be an obvious and relevant pragmatic enrichment of the semantic content of *S*. When this content is a complete, truth-evaluable proposition, we let the proposition itself count as one of its own possible enrichments. [...] When [the semantic content of a sentence determines] a complete proposition, enrichment is optional. If the option is taken, the proposition semantically expressed by the sentence is itself asserted only when it is an obvious and relevant consequence of the enriched proposition that it is the speaker’s primary intention to assert. (Soames, 2008: 456)

This means that even an assertion that appears to be fully semantically encoded involves pragmatic enrichment of its semantic content.

Such analysis of assertion (and particularly, the emphasis on its overlap with pragmatic enrichment) provides additional support for the sentential status of a verbless structure. Reliance on pragmatic enrichment is *not a deficit*, but is common to all structures – even the most basic canonically-structured assertions.

Soames's distinction of required and optional pragmatic enrichment makes it possible to refine the present proposal for the meaning of verbless sentences. A *required* pragmatic enrichment would be involved in generating the basic initial assertion that the uttered verbless structure constitutes in and of itself a relevant proposition, i.e. by uttering the independent structure '*From France*' the speaker commits to the truth-evaluable assertion that '*From France*' is propositionally relevant to the context. This basic assertion is then used as the reference point; the conversationally implicated meaning extends over and above the assertion of this basic propositional relevance. Optional pragmatic enrichments are involved in generating the conversational implicature.

However, it is the present perspective that *optional* programmatic enrichment, which goes beyond the semantically encoded meaning, should be considered in terms of conversational implicature, and *not* assertion. As long as the meaning is cancellable, the speaker commitment is not the same. Even if a particular propositional meaning may be identified, e.g. that '*From France*' is certainly in that particular context referring to the acquisition of a smoking habit sometime in the past, as opposed to any of the other possible pragmatic enrichments, the speaker commitment is not the same as if the speaker were to explicitly utter the enriched proposition.

It may be important to emphasize that what is presently meant by the 'difference in commitment' between asserting and conversationally implicating, need not necessarily be explained as an explicit avoidance of responsibility on the part of the speaker to commit to a particular proposition; it could be an issue of style, a desire to involve the hearer, or the best way to bring some information into focus, or it could have another pragmatic purpose for the speaker. (As implicature involves the hearer in the generation of the proposition to a greater extent than a fully explicit assertion, the differences in speaker commitment, either to a predicative relation or an implicated proposition, need not be interpreted as being equivalent to a weaker proposition.) The difference in commitment is ultimately defined by the possibility or non-possibility of cancellation.

Given the same context, the utterance of '*From France*' should not be treated as asserting the pragmatically enriched proposition '*This cigarette is from France*', but this is *not* because assertions are free of pragmatic enrichment (they are not: pragmatic enrichment is

necessary both for assertions and for conversational implicature, as illustrated by Soames, 2008). Rather, they should not be treated as assertions because they involve a type of pragmatic enrichment to which the speaker has not overtly committed and may cancel. In other words, they involve differences in speaker commitment.

To amalgamate this difference, i.e. make even an extremely likely potential optional enrichment equal to an assertion, would seem to throw the baby out with the bath water. It would take away from the speaker the powerful ability to control the extent of the assertion and commitment that is made. Simultaneously, if optional pragmatic enrichment constitutes an assertion, some of the purpose for using an overt (and less-economical) proposition would seem to be lost. Thus, while acknowledging that assertion necessarily involves pragmatic enrichment, it seems entirely undesirable for the necessity to be reciprocal, i.e. it seems unnecessary for a pragmatically enriched proposition to constitute an assertion as opposed to conversational implicature. Associating optional enrichment with assertion would mute the pragmatic advantages of both implicature and assertion concerning controlling speaker commitment.

4.6.3 Summary

To summarize, it has been presently proposed that the propositional meaning and the predicative relation are semantically undetermined in a verbless sentence. By means of pragmatic enrichment, the verbless sentence (i.e. a verbless structure used in a context with initial and final marking), makes the basic assertion that its semantic contents are propositionally relevant, i.e. that the words of the structure constitute a proposition that is relevant to the particular context. This assertion is non-cancellable and marks the extent of speaker commitment. A more precise proposition which would involve a predicative relation, is generated through the interaction of the semantic content with the context and the conversational maxims. This more precise proposition is a potential pragmatic enrichment: it is conversationally implicated, but not asserted. As obvious as a particular propositional meaning may be to speaker-hearers, it must remain conversational implicature since it does not carry the same speaker commitment as an assertion. Crucially, such generation of meaning is not specific to verbless sentences: (a) pragmatic enrichment is appealed to for the generation of the propositional meaning of even basic assertions, and (b) conversational implicature is

involved even in sentences with a canonical subject-predicate structure. Thus, such an explanation of the propositional meaning of independent verbless structures provides the final necessary argument for their sentential status; and it does this without necessitating any additional syntactic structure or additional semantic encoding. In short: the meaning of a verbless sentence involves an asserted undetermined proposition and a conversationally implicated predicative relation.

4.7 CONCLUSION: PROPOSED FRAMEWORK

From the present perspective there is no longer any reason to refuse sentential status to independent verbless structures. They are accounted for in the same way as verbal sentences: built bottom-up syntactically, semantically encode only their explicit contents, assert a basic propositional meaning of relevance and conversationally implicate the predicative relation. Furthermore, the proof of their sentential status intertwines with the general notion of the sentence; it is hoped that the notion of a sentence may thus extend beyond the marginal application to a narrow subset of language – that of canonical syntactic structures – to which it has often been reduced to over the years.

Greatly motivated by Elugardo and Stainton (2005) and Soames (2008), amongst others, the present account differs in a few important respects. Unlike Elugardo and Stainton, for whom verbless structures are nonsentential assertions, it is presently proposed that they are sentential conversational implicatures. The shared perspective that they do not involve additional hidden syntax (i.e. syntactic ellipsis) and additional semantic encoding (i.e. semantic ellipsis), does not, on the present account, force verbless structures to surrender their sentential status. It is hoped that in *Part I*, it was possible to show that the concession of their sentential status is unjustified. Concerning assertion, Soames provides a vital account of pragmatic enrichment that, if it was applied to verbless sentences, would seem to make it possible to explain their use as fully pragmatically enriched assertions. However, it is presently proposed that such a route must be taken with caution and assertion limited to an absolute required non-cancellable minimum. Utilising Soames' vital distinction between required and optional pragmatic enrichment, the presently proposed account appeals to required pragmatic enrichment for the initial minimal assertion of propositional relevance, but insists that any

additional pragmatic enrichment, including the predicative relation, should be considered in terms of conversational implicature.

The present discussion leads to the more general conclusion that implicature is inevitably and profoundly ingrained in language. Even the most explicit and canonically structured sentences, such as those that may be found in legal texts, still do not escape pragmatic enrichment, since the proposition that they semantically encode constitutes one of multiple possible enrichments and needs to be selected as the most obvious and most relevant one for the particular context or the sentence will take on a different meaning. Thus, it appears that an entirely ‘context-free’ assertion is impossible. The idea of sentences having meaning in isolation also appears wrong-headed: (a) a structure cannot be identified as a sentence without a linguistic context to mark it as being unembedded and as carrying an information structural focus, and, (b) its meaning needs pragmatic enrichment which, even when the structure is said to be considered in isolation, is necessarily imagined. It seems that in each sentence, even for what may appear to be the most explicit of assertions, the hearer must take a ‘pragmatic leap’ to get from the semantically encoded content restrictions to the propositional meaning: in the case of conversational implicature this leap is greater and, therefore, it appears that verbless sentences (through their reliance on conversational implicature for the predicative relation) maximally exploit this pragmatic leap, which is, however, present in all of language and without which one could not deduce the most obvious, most relevant, intended meaning.

Combining such analysis of the meaning of verbless sentences with cross-linguistic differences in their use suggests that some languages rely more systematically on implicature than others. Russian and English differences in this regard are case in point: Russian appears to systematically conversationally implicate the predicative relation more than this is done in English. (As will be discussed further on, a verbless sentence is, for the purpose of contrastive analysis, presently considered to conversationally implicate a predicative relation that a corresponding verbal sentence, that arises in similar pragmatic conditions but from the perspective of the requirements of a different language, semantically encodes.) It must be emphasized that even though verbless and verbal sentences differ in terms of the pragmatic status of the predicative relation (i.e. conversationally implicated in the former versus asserted in the latter), they both minimally assert that they constitute a relevant proposition. Crucially, they are both instances of predication – though with a different syntactic structure, different semantic encoding and different type of evocation of the predicative relation. Such differences should not be considered from the perspective of compromising the predicational status of either structure, but be treated as typological constraints systematically ingrained into the way

that predicative relations are expressed: after all, typological differences may be not only syntactic and semantic, but also pragmatic.

The idea that language heavily relies on implicature, and that this is no reason for fear, is an echo of Peter Siemund (2018) who suggests that the evident under-specification of language should be taken as a call to enrich the scope of what we observe when we study language:

If sentences and utterances can be used to perform social acts and if certain linguistic units possess the force to perform such acts, the question is how much illocutionary force is linguistically encoded and how much of it is supplied contextually and computed by the interlocutors. This is a fundamental problem in the analysis of meaning, there being constant tension in language between descriptive or encoded meaning and pragmatic or contextual meaning. I think that we can take it for granted that **language is highly unspecified with only a tiny fraction of our world knowledge actually being encoded**. It is then interesting to observe which components of meaning get encoded and which do not. (Siemund, 2018: 33, emphasis added)

Thus, what gets encoded in one language and what gets implicated in another is typologically relevant. The appeal to pragmatic enrichment and conversational implicature in verbless sentences is, therefore, no reason to surrender sentential status.

Part I has attempted to show that criticisms of additional hidden syntax and additional semantic encoding are justified and must not be ignored; however, it is hoped that this part has also convincingly argued that the nonsentential white flag that is sometimes drawn does not provide an adequate solution. To address this challenge, presently, the sentential status of verbless sentences has been defended without resorting to additional syntax or semantic encoding; a proposal has been made with regard to the definition of the notion of a ‘sentence’ and the explanation of the *propositional* meaning of a verbless sentence; pragmatic differences have been appealed to in order to explain the differences in the predicative relation.

The focus so far has been on the following question: *Is a verbless structure a sentence, and if so, on what grounds?* Having defended their sentential status, and described what precisely constitutes a verbless sentence, the present study will now proceed to the analysis of particular semantic and pragmatic factors influencing the use of verbless sentences. A contrastive corpus methodology, described in the following section, will be used for this purpose. The methodology and corpus have been developed with the aim of providing insight into the subsequent question, which will be the focus of the rest of the thesis: *What are the main semantic and pragmatic features of Russian and English verbless sentences?*

PART TWO:

METHODOLOGY

CHAPTER 1

AN INTERDISCIPLINARY APPROACH

Having defined the object of study and defended it as constituting a legitimate linguistic phenomenon, the present section focuses on the methodology for which it was necessary to invent a new computational method of processing verbless sentences and to create a brand-new corpus that vitally differs in design and annotation from existing corpora.

Aiming to break new ground in the semantico-pragmatic description of verbless sentences, the present approach combines methods from contrastive, corpus, computational and enunciative linguistics. The reasons for taking such a perspective, as well as the precise steps taken, are made transparent.

CHAPTER 2

THIRST FOR CORPORA

2.1 THE CHALLENGE

Although verbless structures have fascinated many linguists, the difficulty of retrieving them automatically has meant that most existing analyses have been forced to rely on fragmented data. The challenge for any automated search of a verbless sentence is that the defining element of the search query is a grammatical structure that is indeterminate and is centered on absence. The three aspects of this challenge are summarized in (77):

(77) *The Challenge of Verbless Sentence from the Perspective of a Search Query*

- a. Grammatical structure
- b. Indeterminate
- c. Centered on absence

Notably, the query does not include any form, either lexical or grammatical.

2.1.1 Grammatical Structure

Point (77a) pertains to the difference between the search for a lexical string, e.g. ‘*superman to the rescue*’, and the search for its **grammatical structure**, e.g. ‘NN PREP DET NN’ (which would also concern, for instance, ‘*manager to the flower department*’ and other similar structures, when appropriately queried). The computational issues raised by such structures are outlined by Gaëtanelle Gilquin (2002). In recent years, significant progress has been made in morpho-syntactic (part-of-speech) annotation and the automatic search for precise grammatical structures.

2.1.2 Indeterminate

Yet, point (77b) highlights that verbless sentences concern not just a search for grammatical structures, but a search for **open** grammatical structures. More specifically, verbless sentences, as a general category, have no pre-determined words, grammatical categories, or order, which can serve as an anchor for the search query. Attempting to list all possible verbless structures a priori (e.g. (i) NN; (ii) ADJ NN; (iii) NN ADJ; (iv) DET ADJ NN, etc.) would be futile: it requires listing an infinite number of combinations of grammatical elements.

2.1.3 Problems of Zero Markers

It is also worth noting that approaching the issue from the perspective of **zero markers** would be controversial. At present, corpus annotation does not usually include zero markers (Loock, 2016: 33; Gilquin, 2002: 34; Kopotev & Gurin, 2007). As a potential solution for retrieving verbless phenomena, one might propose to add a zero marker to structures that are judged to be missing something in order to allow the subsequent automatic search for this marker.

However, (i) the eventual position of a zero element inside the structure, (ii) what it would represent, and (iii) how its presence would be theoretically justified, are serious

questions. As shown in *Part I*, the presence of a verb via a null element is a matter of heated theoretical dispute (that appears to be lost by the null element). It is the present perspective that such zero-annotation representing a verb, even if it were available, would risk yielding very questionable data. Although, all corpus annotation is a matter of analysis (e.g. morpho-syntactic annotation minimally requires assumptions concerning the definitions of the parts of speech), the annotation of a zero marker for the verb requires assumptions that do not seem justified. As discussed, even for well-defined antecedent-based syntactic ellipsis, it is not at all clear that a syntactic zero exists within the verbless structure. While classifying ‘syntactically elliptical’ structures together with the antecedents could be fruitful, this would be a different question from positioning zero markers in precise spots inside a structure that does not contain something that might possibly be added.

2.1.4 Problems of Parsed Corpora

Verbless sentences may be considered a type of ‘form-based grammatical phenomenon’. This term is used by Gilquin to refer to a phenomenon that is “centered on the **sentence** rather than on **morphemes or words**” (2002: 7, emphasis added); it corresponds to an open class of structures which cannot be listed a priori. Tagged corpora (i.e. corpora that are only morpho-syntactically annotated for part-of-speech) is typically not recommended for a form-based phenomenon that involves any null elements precisely because morphosyntactic annotation does not include a searchable zero marker. In general, automatic retrieval of this type of phenomenon is said to be best carried out on **parsed corpora** because it is not only tagged for part-of-speech, but also contains information concerning syntactic structure. This makes it possible to retrieve the desired grammatical forms even when they do not contain all of the elements that are characteristic of that grammatical sequence. For instance, Gilquin explains that relative clauses with a null relative pronoun require parsed corpora for automatic retrieval:

It should be noted that such an algorithm [i.e. for querying a tagged corpus] could not retrieve instances of the zero relative pronoun, since null elements are normally not encoded in tagged corpora. By contrast, a search for relative clauses on a parsed corpus would also retrieve relative clauses with a null relative pronoun. (Gilquin, 2002: 7/34)

Along the same lines of reasoning, it is parsed corpora that would tend to be recommended for the treatment of verbless structures. The parsed corpus could, in theory, be queried to retrieve a clause even when the latter does not contain a particular element.

That is the theory, but in **practice** the difficulties for the search for verbless sentences persist. Annamaria Landolfi et al. (2010) surveyed several existing parsed corpora, including the *Penn Treebank* (Taylor et al., 2003), the *Italian Syntactic-Semantic Treebank* (Montemagni et al., 2003), the *Prague Dependency Treebank* (Böhmová et al., 2003), as well as their own *AN.ANA.S Multilingual Treebank* (Landolfi et al., 2010), a 21,300-word corpus of Italian, English and Spanish elicited task-oriented dialogues and spontaneous speech, for their ability to retrieve verbless sentences. Their results show that in practice very few parsed corpora allow for the successful retrieval of the phenomenon. They found that automatic extraction of the structures is usually hampered by what can presently be narrowed down to two key issues.

The first problem that Landolfi et al. (2010) identify is that at the basis of the parsing algorithms are usually **VERB-CENTRIC** syntactic models (i.e. the ‘clause’ in automatic treatment, taking direction from theoretical syntax, is defined on the basis of a verbal predicate):

Surprisingly enough, despite the considerable number of existing treebanks, scant attention is paid to such structures. This is due to the difficulty of inserting such structures in the current syntactic models, **which are all more or less verb-centric**; i.e. they consider the presence of a verb a basic requirement to identify a sentence or a clause. The result is that in the spoken language, above all in dialogues or conversation, many linguistic structures that completely fit with whole utterances are **automatically considered as non-sentential material**. (Landolfi et al., 2010: 1188, emphasis added)

The second problem is that annotation in parsed corpora is usually fixed, with no possibility for adjustment by the user. Though this problem is not explicitly stated by Landolfi et al. (2010), its gravity is made evident through their study. **FIXED ANNOTATION** of existing parsed corpora is essentially what leads them to develop their own *AN.ANA.S* corpus with an annotation scheme that includes a ‘verbless clause’ tag, the definition of which they, as the developers of the corpus, control (Landolfi et al., 2010: 1190).

It is worth mentioning that the frequency of verbless ‘clauses’ found for the *AN.ANA.S* corpus – approximately one third of total clauses in Italian (28%), in English (24%), and in Spanish (33%) – is said to confirm the frequency of verbless clauses in *The Longman Grammar of Spoken and Written English* corpus (Biber et al., 1999) and that of the *C-ORAL-ROM* corpus (Cresti & Moneglia, 2005), as noted by Landolfi et al. (2010: 1192). The *C-ORAL-ROM* is a

corpus of Italian, French, Portuguese and Spanish spontaneous speech, consisting of 300,000 words per language (Cresti & Moneglia, 2005; 2007). Notably, these studies all note a similar frequency of verbless clauses in spoken parsed corpora for Italian, English, Spanish, French and Portuguese, i.e. around 30%. They are, furthermore, **unanimous** in their insistence that such a high frequency of verbless clauses indicates that parsed corpora based on verb-centric syntactic models do not allow to account for basically one third of clauses in spoken language. The inadequacy of such models is emphasized by Emanuela Cresti and Massimo Moneglia (2005) as follows:

More specifically, the statistical measurements made on the C-ORAL-ROM corpus show that verbless utterances constitute 38.1% in the Italian corpus, 24.1% in French, 37.23% in Spanish and 36.57% in Portuguese. Since on average 30% of utterances are verbless, all the definitions based on clause structure and verbal predication appear to be inadequate for spoken corpus analysis purposes. (Cresti & Moneglia, 2005: 14)

This call for the need to overcome the practical difficulties of extraction is also joined by Miriam Voghera (2010):

[V]erbless utterances [i.e. any uttered autonomous sequence] are so frequent in both formal and informal speech that they cannot be considered episodic of disfluencies phenomena. (Voghera, 2010: 95)

Voghera insists that the frequency of verbless structures does not allow them to be brushed off as disfluencies and disregarded. Interestingly, Voghera notes that work on the **prosody** of verbless structures in spoken Italian corpora shows that there is considerable similarity in the prosodic form of verbal and verbless sentences; thus, proposing an additional argument in defence of a complete and autonomous status for verbless sentences from a prosodic perspective (Voghera, 2010: 100). Furthermore, Voghera (2010: 95) draws attention to the fact that quantitative data regarding the **written** register is limited when compared to that of oral speech, despite studies such as those of Florence Lefevre (1999; 2001; 2007), Lefevre and Behr (2011), Lefevre and Nicolas (2004) showing that verbless sentences are also attested in literary and non-literary texts.

To resolve the problem of verbless sentence retrieval, Landolfi et al. (2010) suggest that parsed corpora should be built on models that treat verbless clauses as syntactic objects:

This shows that speakers, according to different communicative situations and exigencies, produce a variety of structures that cannot be easily subsumed under canonical sentencehood representation [...] This is a necessary step to produce

the **theoretical basis** for a **richer treebank annotation tagset**, which could consider **[verbless clauses] as proper syntactic objects** and not only as reduced forms of canonical sentences. (Landolfi, 2010: 1193)

It should be emphasized that their solution is essentially a call for changes to the theoretical syntax coming *from the computational domain*.

To summarize, what appears to be occurring with regard to (the stalling of) automatic retrieval of verbless sentences in parsed corpora is a well-meaning exchange of finger pointing. In essence, linguists working in non-computational domains draw attention to the technological barriers, while the automatic processing of the structures is hampered by controversy surrounding the linguistic definition of a clause. Thus, it seems that behind the problem of verbless sentence retrieval hides a more general problem of a too distant divide between computational and theoretical linguistics. The issue seems to be at a stalemate with each side waiting for the other in order to make progress: the development of necessary models for parsing requires linguists to agree on a theoretical basis for treating verbless structures as syntactic objects, while the linguists require more data in order to get a better picture of the implications of attested verbless structures on theoretical models.

In addition to the problem of (1) *verb-centric syntactic models*, and (2) *fixed-annotation*, the present experience of working with verbless sentences reveals a third important problem that has not received published attention to the best of present knowledge. This crucial issue is (3) **ACCURACY** of verbless sentence retrieval. Although, there has been some success with a few existing corpora (as mentioned, studies have succeeded in the task of retrieving indeterminate verbless clauses from *AN.ANA.S*, *C-ORAL-ROM*, and *LSWE* corpora), discussion of recall or precision rates does not arise.

A notable exception is the accuracy analysis in the work of Fernández et al. (2007), Fernández (2006), Fernández and Ginzburg (2002), who focus on automatically retrieving 15 pre-determined types of what they call ‘non-sentential utterances’. Working with data from the *BNC: British National Corpus* (Burnard, 2007), Raquel Fernández, Jonathan Ginzburg and Shalom Lappin (2007) ran a machine learning experiment on 200 speaker-turns which required manually annotated data; they provide accuracy, as well as frequency results with regard to the 15 analyzed types of their proposed taxonomy.

The general absence of an accuracy discussion may be due to the fact that access to the full parsed corpus is usually limited, i.e. it is often not possible to download it; a lack of access to the full text of the corpora would not permit to carry out the analysis required in order to measure accuracy. Moreover, such analysis is time-consuming as it requires the comparison of

automatic results against a manual standard. Furthermore, the annotation of existing parsed corpora is sometimes a ‘black box’, i.e. carried out automatically using statistical measures, as opposed to being based on rules that are set up ahead of time by the researcher; the researcher has no control over such statistically-based annotation and it would be impossible to alter it in any case.

What the absence of accuracy analysis means is that it is not clear what proportion of total verbless structures that really exist in the corpus are actually automatically retrieved in such studies (i.e. perhaps only 1% of existing verbless structures were retrieved). Similarly, it is also not clear what proportion of those verbless structures that were automatically retrieved are actually verbless (i.e. it could be that this retrieved 1% turns out to be almost entirely verbal).

As will be demonstrated shortly, the present experience with verbless sentence retrieval shows that such accuracy considerations are a fundamental issue for verbless phenomena. The crucial gravity of the accuracy problem revealed itself as a result of the unique nature of the present study in which the development of the computational method went hand in hand with specific linguistic aims. More precisely, the task of automatic extraction was not treated as an end goal in itself, but rather as an inevitable intervention in order to access necessary data to address specific linguistic goals. As a result of this focus, the present study began with a small corpus of manually retrieved verbless sentences, and the automatic method of retrieval (which will be proposed in the following section) was developed through several pilot studies that took accuracy against manual data into account at each additional automatic step, so as not to compromise the goals of semantico-pragmatic analysis. It is thus first and foremost linguistic concerns that led the elaboration of the present method of retrieval. As a result of this step by step approach, it was presently revealed that accuracy constitutes one of the most important pitfalls in the automatic retrieval of verbless structures. Specific tagging and delimiting issues were found to systematically interfere with the task and carry serious consequences for recall and precision of the structures. (These results are provided below in *Part 2: Chapter 4*).

Although the present method of retrieval is developed for verbless sentences, whereas the studies on parsed corpora usually focus on ‘clauses’, the present experiences with accuracy lead one to wonder about the accuracy of the retrieval in the existing studies. For instance, one may wonder to what extent the similarity in the frequency of verbless clauses in English, French, Italian, Spanish and Portuguese parsed corpora may be linked to the key pitfalls in tagging and delimiting that have been observed in the present study of verbless sentences.

Thus, in addition to the problems that have been noted by previous studies with regard to the practical difficulties of extracting verbless structures from parsed corpora, i.e. verb-centric models and fixed annotation, accuracy makes for another important obstacle that requires resolution.

It should also be mentioned that regarding Russian, despite the language generally being known for having the most productive use of verbless sentences out of all Indo-European languages (Kopotev, 2007b; McShane, 2000), corpus studies exploring the verbless phenomenon as a whole (i.e. as an indeterminate grammatical structure as opposed to a particular individual verbless sentence type or several predefined types) do not yet exist to the best of present knowledge. An overview of corpora resources for Russian to date may be found in (Kopotev et al., 2021) and details of their development in (Zakharov, 2013).

2.1.5 Centered on Absence

In addition to constituting indeterminate grammatical structures, point (77c) of the challenge highlights that the only formal element which characterizes the verbless sentence is an absent one. For natural language processing, verbless structures constitute open strings consisting of grammatical elements that cannot be specified a priori. Their only characteristic is the absence of the grammatical category of the verb. In essence, putting the points of the challenge together, automatic verbless sentence retrieval represents a quasi-absurd task of looking for something indeterminate that contains something that does not exist.

2.2 STRATEGIES OF PREVIOUS STUDIES

2.2.1 Adjust Scope to the Searchable

As a result of these difficulties, most of the corpus work on the phenomenon has focused on particular verbless structures that are predetermined and include some searchable formal

marker (e.g. the recent work by Vilinbakhova, 2016; Apresjan, 2017; Vishenkova & Zevakhina, 2019; Kim & Abeillé, 2019; Janda et al., 2020; Kim et al., forthcoming).

For instance, Elena Vilinbakhova and Mikhail Kopotev (2017) take a corpus approach to study the semantic aspects of the Russian verbless tautology ‘X это X’ (*X èto X*; lit. ‘*X this X*’) and the verbal variant ‘X есть X’ (*X est’ X*; lit. ‘*X is X*’). While ‘X’ is an open lexical element, the deictic ‘это’ (*èto*; lit. ‘*this*’) is a constant and, therefore, works as a lexical anchor for automatic retrieval of this particular verbless structure. Working with the *RNC: Russian National Corpus* (ruscorpora.ru; Sičinava, 2005; Sharoff, 2006), Vilinbakhova and Kopotev searched for various types of phrases within a certain distance of the constant.

Another example is Jong-Bok Kim and Anne Abeillé (2019) who searched the *COCA: Corpus of Contemporary American English* (Davies, 2008) with string queries, e.g. ‘*why det/art * [y*]*’ (2019: 373), in order to retrieve occurrences such as ‘*Why a wrestler?*’ for their study of the uses and the most appropriate theoretical account of verbless why-stripping elliptical constructions in English.

Thus, taking such a route means to limit the scope of the verbless sentences examined and focus on a specific type of verbless structures that can be searched for electronically.

2.2.2 Steer Away from the Corpus Approach

Alternatively, the technological challenge has led many studies of the semantic and pragmatic aspects of verbless sentences to steer away from corpora and concentrate exclusively on exploring the structures from a qualitative perspective.

For instance, this is the approach taken by Daniel Weiss (2011a; 2011b; 2013). Proposing a fine-grained typology of Russian verbless sentences, Weiss (2011a) makes explicit that he is forced to work with data that is fragmentary, disconnected and incomplete, due to the difficulty of the search for verbless constructions and expressly leaves the search in corpora for future studies:

База данных, на которую опираются мои наблюдения, состоит из разных записей спонтанной речи и фрагментов из художественных произведений; поиск безглагольных конструкций по электронным корпусам по

очевидной причине затруднителен и должен быть оставлен для дальнейших исследований в будущем. (Weiss, 2011a: 139) ⁴¹

Weiss (2013) reiterates the desire for a corpus investigation in a study of syntactic gaps in colloquial Russian:

It goes without saying that these observations have to be corroborated by empirical, corpus-based work. (Weiss, 2013: 97)

Although it is usually not as clearly acknowledged as is done by Weiss, manual selection of examples from eclectic sources is the most common approach in the existing semantico-pragmatic studies of verbless structures. Other qualitatively-rich studies of the semantic, pragmatic and enunciative aspects of verbless sentences include for instance the work of Selivërstova (1973) and Paillard (1984), who explore the semantic impact of re-introducing the verb into certain verbless structures. Using methods other than corpus analysis (i.e. in an exhaustive ‘all-inclusive’ fashion where all of the examples as well as the entire corpus is studied) may of course be due to researcher preferences; the intention of the present discussion is not at all to state that the corpus approach is the only valuable one or to undermine alternative complementary approaches which look at the phenomenon from a different perspective. Nevertheless, the inability to access full and accurate attested data through a corpus sets several important limits on verbless sentence research.

2.3 DATA LIMITS

The difficulty of automatic retrieval has meant that the majority of studies on verbless sentences have relied on data consisting of constructed examples and examples that were selected by the researcher from various sources in a non-exhaustive fashion, fragmented and abstracted from their original context. Although much qualitative progress has been made, the perspectives for verbless sentence research are limited as long as it relies solely on such data.

⁴¹ Translation: “The database on which my current observations are based consists of various fragments from spontaneous speech and fragments from literary works; the search for verbless constructions in electronic corpora is difficult for obvious reasons and should be left for other studies in the future.”

The overall use of corpora in most studies of verbless sentences can be called “illustrative eclecticism” (Kohnen, 2015: 56) or of the “corpus-illustrated” type, as opposed to “corpus-based” or “corpus-driven” (Divjak et al., 2017: 4). When data is eclectically selected, it can only be used for the purpose of illustration and statistical analysis must be sacrificed. The results of many qualitatively rich studies of verbless sentences are thus left vulnerable to critiques that face data that has been selected from various sources in order to illustrate a preconceived model.

The first issue is that a subjective retrieval process heightens the potential for researcher bias and circular reasoning. It is possible that the data might unintentionally be selected in order to prove a preconceived hypothesis, and so inconvenient or ambiguous examples might be accidentally overlooked. A corpus approach encourages linguists to treat inconvenient examples and puts the data ahead of the hypothesis.

Another serious problem is that the examples are abstracted from their original full-text context. The use of decontextualized examples carries significant qualitative and quantitative limits. Qualitatively, the systematic treatment of context is a fundamental necessity for semantic and pragmatic analysis. From a quantitative perspective, using a series of examples permits only raw frequency counts. This is due to the absence of a reference corpus to serve as point of comparison.

For instance, a high frequency of interjections in a series of verbless sentences does not tell us whether the element statistically characterizes these sentences. For the latter conclusion, it would be necessary to study the entire corpus since it may happen that interjections are found just as frequently in sentences containing verbs. Working with a series of verbless sentences, one may count the quantity of some marker, but one cannot say to what extent the observations are significant, i.e. they may turn out to be true for the entire corpus and not just for the examined series. Consequently, it is highly desirable for verbless sentence studies to look to automatic retrieval and treat verbless sentences in their *full* original context (i.e. here, the emphasis is that it is not just the immediate context of a specific verbless occurrence that is necessary, but the entire corpus in which the verbless sentences are found). The use of corpus methods creates the potential for the descriptive treatment of more complete data and allows for statistical inferences.

Such arguments, together with the strong desire to verify intuitions by more objective means, are usually the deriving factor behind the use of the corpus method and the development of natural language processing techniques (Zakharov & Bogdanova, 2020; Aijmer &

Rühlemann, 2015; Anthony, 2012; McEnery & Xiao, 2008; McEnery et al., 2006; Mitkov, 2003; McEnery & Wilson, 2001; Garside et al., 1997).

2.4 TECHNOLOGICAL BIASES

At the same time, the desire for attaining empirical conclusions in linguistics through the use of modern technology is sometimes criticized for sacrificing too much in terms of quality of analysis. Despite advances in automatic processing, a comfortable level of accuracy still requires the attention to detail often made possible only through human involvement. Powerful software makes it possible to instantly synthesize large amounts of data, however, knowledge of the corpus, of the annotation system, as well as of the technical nuances of the software is fundamentally necessary in order to make appropriate analysis of the results.

It must be acknowledged that the use of electronic corpora and language processing software often imposes its own limits on the data. For example, in a study of subject omission in Russian, Tatiana Zdorenko (2010) uses data drawn from different genres of the *Russian National Corpus*. Technological limitations impose restrictions on the type of utterances that may be identified:

Since the corpus was not marked up for syntactic relations, it was impossible to search for subject positions in clauses automatically. [...] In order to somewhat facilitate the analysis, I searched the selected subcorpora for all instances containing a verb. (Zdorenko, 2010: 125)

As a result, the analysis excludes an important type of subjectless sentence, notably, the sentence that is both subjectless *and* verbless. Attentive to the technological restriction, Zdorenko (2010) uses the retrieved data in order to draw conclusions about the verbs involved in subjectless structures. However, were this to be an analysis of subjectless utterances in general, such a technological limitation would be a serious concern. An important risk resides in the fact that restrictions on the data brought about in automatic processing are not always obvious.

Thus, while a corpus approach reduces researcher bias, encourages a more complete and descriptive treatment of data, allows for systematic analysis against a reference corpus and consequently statistical predictions, it also brings new problems into the picture. With the use

of natural language processing software and electronic corpora, one must not only pay attention to researcher bias (e.g. in the analysis of the automatic results), but also simultaneously to the bias introduced by the technical specifications of the software or of the electronic corpus used. Moreover, as will be shown shortly, issues such as sentence delimiting and morphosyntactic tagging, which the researcher often has little control over, influence the accuracy of a corpus analysis in sometimes unpredictable ways. Notably for verbless sentence studies, the important question of what qualifies as a sentence or a verb must be answered not only from a theoretical perspective but also from the computational perspective of what precisely is accurately extracted; therefore, potentially setting new limits on the data and influencing the results.

2.5 CONTEXT OF THE *VERBLESS SENTENCE*

A prior section mentioned the importance of context in the sense of the **overall corpus** from which the data is drawn, i.e. it is necessary to compare verbless sentences against a full reference corpus, one in which verbal sentences are present, in order to allow statistical analysis. This is one type of context that must be overcome by a verbless sentence corpus study, but there exists another. Additionally, it is crucially important to maintain the link to the context of the individual occurrences of the phenomenon.

The semantic and pragmatic aspects of verbless sentences are presently targeted through the analysis of contextual factors of the enunciative situation in which the verbless sentences are found. It is thus necessary that the verbless sentence be retrieved together with its individual **context of utterance** in order to allow for this subsequent analysis. The individual context is also fundamental for the identification of a potential antecedent in order to distinguish the antecedent-based elliptical versus the non-elliptical type of verbless sentence.

While this type of context is a simple task for most types of linguistic phenomenon, the individual utterance context of a verbless sentence hides another challenge from the perspective of automatic processing. Notably, the retrieval of the context in this case pertains not to the typical concordance lines which are centered on a word or on a particular lexical or grammatical expression, but for verbless sentences the concordance lines must be centered on a sentence – or, rather, on an indeterminate grammatical sentence unit that is itself defined by absence. Maintaining this vital link to the context of utterance for the verbless sentence is still

computationally not an easy feat. This adds another complication to the retrieval of verbless sentences in corpora and, if left unaddressed, crucially limits the possibility of semantico-pragmatic analysis of its utterance situation.

In order for the present study to attempt to confront the existing gap in semantic and pragmatic analysis of verbless sentences from a corpus perspective, it was necessary to make the context central at each step of the corpus design and the verbless sentence retrieval process, as well as to develop specific adjustments to existing software and new processing tools. The unique way this challenge was presently tackled will be described shortly in *Part 2: Chapter 4*.

CHAPTER 3

CONTRASTIVE LINGUISTICS

In order to gain a better understanding of the semantic and pragmatic factors associated with the absence of the verb, the present study adopts a methodological framework that combines corpus methods with contrastive linguistics.

Guiding this combination of methods is the fundamental conviction that cross-linguistic comparison makes evident linguistic constraints that would otherwise remain hidden. This conviction is at the basis of the contrastive analysis method founded by Jacqueline Guillemin-Flescher (1981; 1999; 2003; 2004; 2011; Gournay & Merle, 2004). In particular, Guillemin-Flescher shows the way that the study of translation patterns, i.e. systematically reoccurring choices made by translators, can reveal semantico-pragmatic linguistic constraints that are hidden from the perspective of a single language. The method of contrastive translation analysis was also simultaneously developed in the work of the Russian linguist Vladimir Grigor'evič Gak (1975; 1994).

In the present study, the contrastive translation-based methodology, in which a particular linguistic phenomenon is traditionally studied from the perspective of a systematically manually-analyzed all-inclusive series of translation correspondences, is combined with the opportunities made possible by the progress in parallel corpora. The present method also considers the possible pitfalls of using parallel-text studies (e.g. Gellerstam, 1986; Baker, 1993; Baker, 1996; Malmkjaer, 1998; Olohan, 2002; Gellerstam, 2005; Stolz, 2007a; Stolz, 2007b; McEnery & Xiao, 2008; Zanettin, 2013; Nádvorníková, 2017; Henkel, 2018) and makes appropriate adjustments in order to address these important concerns. Thus, Russian and English are presently studied with regard to verbless sentences and verbless sentence

translation correspondences. A reciprocal parallel corpus is specially created and used in order to uncover systematic translation patterns. The patterns are analyzed, both statistically and manually, in terms of what they reveal about the semantico-pragmatic constraints on the use of verbless sentences in the two languages.

The following sub-sections will explain the importance of studying particularly these two languages with regard to the verbless phenomenon, reveal the way that the potential pitfalls of contrastive analysis and parallel corpora were presently addressed, theoretically frame the relation between a verbless sentence and a verbal translation correspondence, and draw attention to particular issues of verbless sentences in parallel corpora.

3.1 RUSSIAN AND ENGLISH

Profound typological differences make it particularly relevant to compare English and Russian. A recent comparison of the general characteristics of the two languages may be found in (Kozlova, 2019; Kirillova, 2015). The present section illustrates the key differences that motivate the study of particularly the verbless phenomenon in these languages.

In sharp contrast to English, **Russian** is characterized by a highly developed morphological case system, flexibility of word order and intonation, the nonexistence of articles, and an extraordinary capacity for verbal and subject ellipsis. Russian typically does not use a copula verb in the present tense in unmarked contexts, and is highly productive with regard to non-elliptical verbless structures (i.e. that do not require a verbal antecedent).

3.1.1 Zero-Copula Constructions

The absence of present tense copula use is a phenomenon that is typically referred to as the ‘**zero-copula construction**’. Leon Stassen (2013) emphasizes that this name need not necessarily presuppose any hidden syntactic or semantic structure (cf. the discussion of the copula past–present–future paradigm in *Part 1: Section 3.5.7.2.3.1*):

The term *zero copula* is used here as a strictly neutral technical label, in that it refers purely to a construction in which the relation between a subject and a nominal predicate is not marked by an overt item. (Stassen, 2013: 1)

An example of such construction is the typical introduction in (78):

(78) ‘Zero Copula Construction’

{Introducing oneself}

Я Владимир.
ja Vladimir
I Vladimir.NOM
‘I am Vladimir’

As mentioned, the Russian verb ‘БЫТЬ’ (*byt’*; lit. ‘*be*’) is regularly employed in the past (‘БЫЛ’; *byl*; lit. ‘*was*’) and future (‘будет’; *budet*; lit. ‘*will_be*’) tenses, but its use in the present tense (‘есть’; *est’*; lit. ‘*is*’) is highly marked and generally emphasizes existence. For instance, one may use the present tense to stress that something really does exist in response to someone who has questioned such a predicative relation, e.g. (79).

(79) {Someone doesn’t believe that you are Vladimir}

Я и есть тот самый Владимир.
ja i est’ tot samyj Vladimir
I PART am that same Vladimir
‘I really am that very same Vladimir’

Thus, in the present tense, the verb ‘БЫТЬ’ (*byt’*; lit. ‘*be*’) tends to be used in a lexical sense that goes beyond a simple equivalence relation. The distinction between **the copula and the lexical** use of the verb ‘*be*’, and its correspondences across languages, goes back to Emile Benveniste (1966/1971) who explains that sometimes ‘*be*’ expresses merely the grammatical notion of the Copula, i.e. a link to say that the subject equals the predicate, while at other times ‘*be*’ is a full-fledged Lexical Verb. The first ‘*be*’ is a semantically empty link; the latter has meaning:

Two distinct terms that are confused in speaking of ‘to be’: one is the ‘copula’, the grammatical mark of equivalence; the other, a full-fledged verb. The two have co-existed and will always be able to co-exist since they are completely different. But in many languages, they have merged. The problem of ‘to be’ thus comes down to a development not of chronological succession but of the

dialectical coexistence of two terms, two functions, and two constructions.
(Benveniste, 1966/1971: 163)

Since Benveniste separated the two functions, it is usually recognized that the function of marking equivalence (i.e. the function that is typically associated with the verb ‘*be*’ and its cross-linguistic correspondences) need not only be carried out by the verb ‘*be*’. Thus, in (78) the absence of a verb in the Russian sentence and its presence in the English correspondence would suggest that in English the verb ‘*be*’ merges the lexical and the copula function, while in Russian the two functions are separated.

Kees Hengeveld (1992) develops the idea that the copula function may be grammatically **marked by different means** in different languages, i.e. it may be marked not only verbally (e.g. in addition to ‘*be*’ and ‘быть’ (*byt*; lit. ‘*be*’), similar verbs such as ‘*seem*’ and ‘является’ (*javljaetsja*; lit. ‘*appear*’), etc., may be used), but also via, for instance, a pronoun or a particle. He argues that, being a semantically empty link, the copula cannot constitute the main predicate of a sentence (Hengeveld, 1992: 26) and proposes a binary model of verbal and non-verbal predication in which not only verbless sentences, but also sentences with an explicit verbal copula would constitute the non-verbal predication type. Cross-linguistically, the model results in the transformation of predication type, which we investigated in (Bondarenko, 2018).

Further discussion of the distinction between the copula and lexical verbs for Russian may be found in the work of, e.g., Letuchiy (2015), Testelec (2008), Chvany (1975), and above in *Part 1: Section 3.5.7.2.3.1. Concerning English*, Rodney Huddleston and Geoffrey Pullum (2002: 113, 266) rework Benveniste’s distinction and distinguish six different uses of the **English verb ‘*be*’**, as in (80):

- (80) a. *Copula ‘be’*: The chief culprit was Kim. (*specifying*)
 His daughter is very bright. (*ascriptive*)
 b. *Progressive ‘be’*: She was sleeping peacefully.
 c. *Passive ‘be’*: They were seen by the security guard.
 d. *Quasi-modal ‘be’*: You are not to tell anyone.
 e. *Motional ‘be’*: She has been to Paris twice already.
 f. *Lexical ‘be’*: Why don’t you be more tolerant? If you don’t be quick
 you’ll lose.

For Huddleston and Pullum, the English copula use of ‘*be*’ corresponds to a syntactic link that relates the subject to the predicative complement, generally has little semantic content, and serves a merely syntactic function of carrying tense inflection (2002: 218).

Notably, in Russian, following the analysis proposed in *Part I*, structures such as the introduction in (78) are not grammatically marked for tense; they do not contain a hidden verb in their syntactic structure, nor encode one semantically. However, they may, potentially, conversationally implicate a temporal meaning through interaction with the context. Their predicative status is not gained based on a verbal meaning, nor based on simple juxtaposition of the words; rather, the verbless structure must meet the sentential criteria that were proposed above in *Part I* (i.e. like all sentences, it must not be embedded into any other syntactic structure and requires contextual insertion). As such, the sentence in (78), and other ‘zero-copula constructions’ defined according to the copula paradigm, do not carry tense. They are syntactically **tense-less** and semantically **time-less**, i.e. tense and time is undetermined and unspecified by the speaker who may, or may not, conversationally implicate an additional temporal meaning.

A final point about the copula verb in the present tense. The historical linguistic analysis of Mikhail Kopotev on verbless sentences in Russian shows that the present tense form of the copula verb ‘быть’ (*byt*; lit. ‘*be*’) started to **drop out of use** progressively to finally disappear around the 11th century (Kopotev, 2007b: 120). A potential question for future research is the diachronic fate of what seem to be syntactically similar sentences in English, e.g. ‘*You genius!*’, ‘*You crazy!*’, ‘*Me hungry.*’, though they are not usually treated as pertaining to the ‘zero copula’ type and are much less productive.

3.1.2 More than Zero-Copula Constructions

While verbless ‘zero copula constructions’ are common to several Indo-European languages, Russian is also known for a very productive use of other additional types of **non-antecedent based** verbless structures (e.g. McShane, 2000). One illustration is the verbless sentence in (81) which, in the particular context, implicates motion:

(81) *Verbless Sentence without Verbal Antecedent*

{Notifying someone about leaving}

Я в магазин.
ja v magazin
I to store.ACC
'I'm going to the store'

In English, the use of the verbless structure '*I to the store*' would typically require a verbal antecedent, e.g. (82a). It also, notably, takes on a formal register (as discussed in (88) below).

(82) a. *Antecedent-based*

Speaker A: She is going to the office.
Speaker B: I to the store.

b. *Marked*

{Announcing a plan of action}

Superman to the rescue! I to the store! You out.

The non-antecedent based use of '*I to the store*' is also possible in English, as is illustrated in (82b) which also shows two other similarly structured non-elliptical verbless sentences. However, such non-elliptical sentences would be marked, in the sense that they appear to be restricted to specific contexts. (It is one of the present aims to gain insight into what characterizes these contexts.)

The comparison of English and Russian is particularly relevant for the study of the verbless phenomena due to the fact that Russian is distinguished for permitting the **most productive and liberal use** of verbless sentences among the family of Indo-European languages (McShane, 2000; Kopotev, 2007b). In contrast with other languages, Russian is generally described as having an 'anti-structuring tendency' (Kopotev, 2007b: 131), largely as a result of its use of verbless and subject-less sentences and free word order:

Russian goes to extremes in ellipsis, in destroying syntactic constituents in favour of expressive-informal-rhythmic salience, and in leaving it to the hearer to guess the logical connections between predications, even to put together the predication from a scrambled sequence. (Leinonen, 1985: 138, in Kopotev, 2007b: 131)

In comparison, the typological **characteristics of English** include dependence on the finite verb phrase, the absence of a ‘zero-copula construction’ and register restrictions on verbal ellipsis (e.g. Stassen, 2013; Leech, 2004; McShane, 2000).

Nevertheless, verbless sentences are more **widespread** in English than is usually assumed about the language. Corpus data shows that in addition to sentences such as the exclamatives in (83), one also typically finds imperatives, questions, and assertions, such as illustrated, respectively, in (84) through (86).⁴²

(83) *Exclamatives*

- a. What brave and lofty words!
- b. How terrible!
- c. Oh, you gentry!
- d. Oh, miserable me!
- e. Shameless imposter!

(84) *Imperatives*

- a. And now to business.
- b. Just a moment.
- c. To a duel!

(85) *Questions*

- a. What about my parental blessing?
- b. Grushenka, my relative?
- c. What, again?

(86) *Assertions*

- a. So much, then, for the introduction.
- b. By the way, about Fyodor Pavlovich.
- c. No sign of Dmitry Fyodorovich yet.
- d. A beautiful utopian dream of the disappearance of wars, diplomats, banks, and so on.

Notably, following Huddleston’s (1994) distinction between questions and interrogatives, the four sentence types (i.e. exclamatives, imperatives, questions, assertions) are distinguished based on semantic as opposed to syntactic grounds.

⁴² The illustrative examples (83) through (86) are from the corpus parts BK_RO (Russian original ‘Братья Карамазовы’), BK_E1 (English translation ‘The Brothers Karamazov’), BK_E2 (English translation ‘The Karamazov Brothers’).

Moreover, a closer look at the context of the above examples reveals that verbless sentences can be found in both formal and informal situations. For instance, the interruption ‘*Just a moment.*’ in (84b) occurs in the context of a formal meeting, whereas the insult in (83e) ‘*Shameless imposter!*’ is used in a heated argument between father and son. Furthermore, as evident in (86d), length is not a necessary restriction on the form.

3.1.3 Ellipsis and Morphological Case

The productivity and register of **syntactic verbal ellipsis** constitute another profound typological contrast between Russian and English. Russian is known for an extraordinary capacity for verbal ellipsis. Marjorie McShane (2000) compares precise types and licensing conditions of syntactically elliptical structures in English, Russian and other Slavic languages; and particular attention to Russian and English syntactic ellipsis is found in the work of Anna Kirillova (2015).

One of the key reasons that Russian is able to easily ellipt verbs is its highly developed **morphological case system**. By means of case one can effortlessly tell if a noun functions as a subject (*nominative*), an object (*accusative*), an indirect object (*dative*), if it is the possessor of some object (*genitive*), if it is the instrument that was used to carry out some action or by means of which some action was performed (*instrumental*), or if the noun relates to the preposition that came just before (*prepositional*). These case markings are summarized in *Figure 8*.

Russian Morphological Case Marking

subject	NOMINATIVE ИМЕНИТЕЛЬНЫЙ	ребёнок <i>rebënok</i> child.NOM	ручка <i>ručk-a</i> pen.NOM
ownership	GENITIVE РОДИТЕЛЬНЫЙ	ребёнка <i>rebënk-a</i> child.GEN	ручки <i>ručk-i</i> pen.GEN
indirect object	DATIVE ДАТЕЛЬНЫЙ	ребёнку <i>rebenok-u</i> child.DAT	ручке <i>ručk-e</i> pen.DAT
object	ACCUSATIVE ВИНИТЕЛЬНЫЙ	ребёнка <i>rebenok-a</i> child.ACC	ручку <i>ručk-u</i> pen.ACC
‘with’/‘by means of’	INSTRUMENTAL ТВОРИТЕЛЬНЫЙ	ребёнком <i>rebenok-om</i> child.INS	ручкой <i>ručk-oj</i> pen.INS
after preposition	PREPOSITIONAL ПРЕДЛОЖНЫЙ	ребёнке <i>rebenok-e</i> child.PRE	ручке <i>ručk-e</i> pen.PRE

Figure 8. Russian Morphological Case

As a result of the developed morphological case system it is possible to avoid much of the ambiguity that would necessary arise in English which lacks such a developed case system.

The differences in the **productivity** of verbal ellipsis in Russian and English are demonstrated in (87), an example adapted from McShane (2000: 200), and which we also discuss in (Bondarenko, 2019).

(87) *Productivity of Ellipsis*

a. *Russian: Subject case*

Мама попросила Мишу спеть, а отец – сыграть на пианино.
mama poprosila Mišu spet', a otec – sygrat' na pianino
mom asked Misha to_sing, and father.NOM – to_play on piano

b. *Russian: Object case*

Мама попросила Мишу спеть, а отца – сыграть на пианино.
mama poprosila Mišu spet', a otca – sygrat' na pianino
mom asked Misha to_sing, and father.ACC – to_play on piano

c. *English: Object interpretation (default)*

Mom asked Misha to sing and father to play the piano.
Mom asked Misha to sing and [Mom asked] father.OBJ to play the piano.

d. *English: *Subject interpretation*

Mom asked Misha to sing and father to play the piano.
** Mom asked Misha to sing and father.SUBJ [asked Misha] to play the piano.*

The example highlights the importance of morphological case for the verbal ellipsis phenomenon. In instances such as (87), Russian explicitly specifies whether the father is syntactically (a) *the subject* or (b) *the object*, and thus, whether the father is the one asking Misha to play the piano or if it is the father who is being asked by the mother to play the piano. In English, there is no such case marking; only, as noted by McShane (2000), a default preference for such a construction to be interpreted as the latter case, i.e. that father constitutes *the object* being asked to play, as in (c). Based on expectations of a greater cognitive processing required to interpret the structure in English, as compared to Russian which spells out which interpretation is desired via case, it is reasonable to think that there would be a tendency to avoid these types of structures in English.

Russian and English have also been noted to differ in terms of the **register** of verbal ellipsis. In particular, McShane (2000: 198) draws attention to the existence of a register restriction in English on the *gapping* type. This type of ellipsis is illustrated in (88):

(88) *Register of Gapping Ellipsis*

a. *Russian*

Она	идет	в	офис,	а	он	в	парк.
<i>ona</i>	<i>idet</i>	<i>v</i>	<i>ofis,</i>	<i>a</i>	<i>on</i>	<i>v</i>	<i>park</i>
F3SG	PR	PREP	ACC	CONJ	M3SG	PREP	ACC
<i>she</i>	<i>go</i>	<i>to</i>	<i>office</i>	<i>and</i>	<i>he</i>	<i>to</i>	<i>park</i>

b. *English*

She is going to the office, and he to the park.

The above gapping constructions are syntactically possible in both languages. However, the use of the English gapping construction in (88b) is said to be restricted to formal writing and planned speech registers (McShane, 2000: 198). Furthermore, an English structure such as (88b) ‘*he to park*’ or (82) ‘*I to store*’ would normally require an explicit verbal antecedent, such as the one provided by the occurrence of the finite ‘*is going to*’ in the utterance context, except potentially for very marked cases discussed in (82b) above.

In contrast, no such register restriction exists on Russian gapping. The Russian structure ‘он в парк’ (*on v park*; lit. ‘*he to park*’) may be both (i) an elliptical utterance without any obvious register restrictions, as in (88a) where the finite ‘идти’ (*idti*; lit. ‘*go.PR*’) occurs in the preceding utterance, and (ii) it may also occur as a non-elliptical utterance (i.e. without a verbal antecedent) without noticeable restrictions on register.

The typological characteristics of English thus provide a sharp contrast to Russian. Due to the limited English case system, the verb plays an important role in reducing ambiguity, as for instance in (87d). Register restrictions on syntactic gapping, the typical necessity of an overt copula, as well as other important differences such as the existence of articles and strict word order, also make the comparison of English with Russian particularly relevant for the study of verbless sentences.

3.2 ANALYSIS OF TRANSLATIONS

Contrastive analysis of translations is presently undertaken with the aim of uncovering linguistic features of the two languages. Many are optimistic about the unique opportunity provided by parallel corpora for going beyond syntactic description. Studies on parallel corpora and contrastive corpus methods (e.g. Cysouw & Wälchli, 2007; Dahl, 2007; Stolz, 2007a; Wälchli, 2007; Hasselgård, 2010; Zanettin, 2012; Doval & Sánchez Nieto, 2019) have drawn attention to the remedy that parallel texts could provide in terms of contextualized examples and quantitative data, and the consequent benefits for contrastive analysis. A special issue of *STUF 60(2): Language Typology and Universals* (ed. Stolz, 2007b) presents a particularly detailed discussion of the advantages and disadvantages of using parallel corpora for typological investigations.

However, many have also drawn attention to the potential pitfalls of using translations for the purpose of analyzing particularly the typological characteristics of a language. A thorough overview of the most important risks associated with translation analysis and parallel corpora, as well as suggestions for remedies, is provided by Olga Nádvorníková (2017) and Federico Zanettin (2013). These important concerns may be split up into two fundamental challenges outlined below. Ultimately, translations constitute a type of language that deserves to be studied in its own right, alongside with originals.

3.2.1 Source Language Interference

The first issue to keep guard of is the influence of the original source language onto the translated target text. Following Gideon Toury (1995), this concern is commonly referred to as the ‘Law of Interference’:

in translation, phenomena pertaining to the make-up of the source text tend to be transferred to the target text (Toury, 1995: 275, in Nádvorníková, 2017: 12)

The risk of transfer from the source language concerns not only the formal structure, as for instance in a translation that is too literal, but also stylistic issues (e.g. Vinay & Darbelnet, 1977/1995).

The danger *for verbless sentences* is clear: the famous English dependence on the finite verb risks being transferred onto the Russian translation and thereby potentially over specifying the predicative relation in Russian translations as compared with Russian originals. Thus, a contrastive study aiming at linguistic analysis of translations needs to make sure that any identified patterns are not simply an effect of the source language.

3.2.2 Traits of Translation Language

The second concern important to address for any contrastive study that is focused on uncovering typological language-specific constraints is to make sure that the identified patterns are not simply an effect of the act of translation itself. Such interference could occur on the level of:

- (a) *the individual translator*, i.e. a unique personal style particular to the translator, and
- (b) *the historical or social setting*, i.e. unique interpersonal norms particular to a certain time period or domain.

These may be called the ‘**Variant Traits of Translation**’, following Zanettin (2013: 21).

In addition, there exists a heated debate concerning the possibility of ‘**Invariant Traits of Translation**’ (or ‘Translation Universals’). The latter were famously introduced into the discussion by Mona Baker (1993).

The main research strand [of corpus-based translation studies] is perhaps that which investigates the hypothesis of translation universals, i.e. supposedly invariant features which characterize all translated texts independently of source language and translation direction (Baker, 1993). (Zanettin, 2013: 21)

Translation universals is a name for stable features which are hypothesized as being a systematic part of any translated language.

For instance, one potential translation universal that is often discussed is ‘simplification’, i.e. that texts become simplified when they are translated. The linguistic indicators of simplification are typically said to be:

- (a) *lexical variety*, i.e. the overall number of individual word *types* (for example, in ‘*translation is recreation*’ there are three), and
- (b) *lexical density*, i.e. the number of individual word *types* per the total number of word *tokens* (for example, the structure ‘*translation is translation*’, with two types per three tokens, would make a less dense and thus simpler translation than ‘*translation is recreation*’).

If simplification proves to be a universal feature of translation, it would mean that the richness of the vocabulary of translations would be systematically deficient compared to the richness of the vocabulary of an original text in the same language.

The various features that have been hypothesized as being universal with regard to translation language are surveyed by Zanettin (2013), who also draws attention to the controversy surrounding whether or not such invariable features actually exist. In theory, if a systematic difference between source and translation language exists, it should be possible to uncover it and then, just as systematically, control it. The features of translation language make for a rich field of research and it is far from certain that the differences are as systematic as the universal hypotheses make them seem.

3.2.3 Translation Legitimacy

While the precise differences between source and translated language is an enduring controversy, the fundamentally influential insight that was argued by Mona Baker (1993), and has been maintained by many since (e.g. Olohan, 2002; Zanettin, 2013; Krasnopeyeva, 2015), is that translation language must be treated as **a language in its own right**, i.e. on par with the source language as opposed to deficient to it. Crucially, Baker (1993: 235) questions the traditional ‘primacy of the source text’, i.e. the traditional assumption that the language of a translated text is a second-rate inferior and distorted type of language. Such an assumption has previously led many to exclude translations from corpora and from linguistic investigation. Baker convincingly defends translation against derogatory assumptions:

Translated texts record genuine communicative events and as such are neither inferior nor superior to other communicative events in any language. They are however different, and the nature of this difference needs to be explored and recorded. (Baker, 1993: 234)

[Translations] may well influence the recipient culture and language, if only because every translation is initially perceived as a target language utterance. (Toury, 1985: 19, in Baker, 1993: 239)

Translations are argued to constitute genuine instances of language *production* by the translator and, notably, also important instances of language *reception* by the receivers who read or hear the target language of translations. That the language of translations is a priori *different* from that of an original in the same language does not make translated language an illegitimate type of language; on the contrary, it is the key reason that translations must necessarily be included in studies as a type of language to be analyzed on par with non-translated language and other language types. In other words, (i) its potential difference, combined with the fact that (ii) it is produced by and (iii) received by native speakers, drives the legitimacy of studying translation language in monolingual studies of any particular language.

The view of translations as a legitimate source of data is also emphasized in the work of Federico Zanettin, who in addition draws out the implications of this perspective for **contrastive analysis**:

Bilingual and multilingual comparable corpora, together with parallel corpora of translations and source texts, have also substantially contributed to the **revival** of contrastive linguistics. **Comparable** corpora are monolingual subcorpora of texts independently composed in the respective language

communities and selected applying similar sampling techniques, which are used to compare and contrast regularities across languages. **Parallel** corpora are generally not regarded as optimal since the language of translation is seen as a non-standard, deviant variety of language. However, Translation Studies have contributed to an understanding that, while the language of translation may be different from non-translated language, translation is a **common and legitimate** instance of textual production (Zanettin, 2011). Furthermore, since translation establishes a **direct link** between source and target texts, parallel corpora can be used to observe recurrent patterns of correspondence. Thus, a researcher may focus on a specific **source language feature** and look for regularities in the way it is translated (or on a specific **target language feature** and look for regular patterns which give rise to it). The patterns found in translated texts can then be compared to those found in a **reference corpus** of target language texts. With the help of a '**reciprocal**' parallel corpus in the other direction of translation, it is also possible to observe how these patterns, when occurring in source language texts, are translated, and so on in a cyclical fashion. (Zanettin, 2014: 185, emphasis added)

In other words, translations constitute a type of language that deserves to be studied in its own right, alongside with the language of original texts. Providing support for contrastive analysis, the developments in translation studies also reveal ways in which contrastive studies may be strengthened. Of key importance is to multiply the perspectives from which translations patterns are examined.

3.3 MULTIPLE DIRECTIONS

By maintaining the link between the source and the target translation, parallel texts provide a unique opportunity to examine the way that languages encode virtually the same meaning. For verbless sentences, using such an approach is particularly exciting with regard to two languages that are generally known to drastically differ in their use of the phenomenon. The 'direct' link offered by parallel texts may provide unique insight into the potential restrictions on the way that Russian and English use absence.

However, as shown in the previous sections, several issues need to be addressed before it is possible to reap the full benefits of the approach and determine that any apparent patterns are actually indicative of typological linguistic characteristics. For one, it is of utmost importance to make sure that the patterns are not simply an effect of source language interference. Secondly, it is necessary to take all possible measures in order to disentangle

linguistic features from potential translation-related phenomenon. This means making sure that the translation patterns extend beyond a particular translation, genre, or time period. It is also necessary to guard against the possibility that what is being examined may constitute a potential universal feature of translation as opposed to a typological feature of the language. In other words, contrastive analysis that uses translations and aims at uncovering linguistic features through translation patterns faces additional challenges from that of other translation-based studies: it must mitigate the risk of being compromised by revealing features of translation, as opposed to the typological features of the desired languages under study.

In order to guard against these potential pitfalls of using parallel texts for contrastive analysis, the present study has invested particular attention into the corpus design, automatic processing, and the analysis of the translation patterns from multiple directions.

To address the problem of source language interference several particular measures were presently taken.

3.3.1 Directions of Translation

The first is that the verbless sentence translation patterns are presently studied in both directions of translation, i.e. verbless sentences in English originals are compared to their Russian translations, and, *reciprocally*, verbless sentences in Russian originals are compared to their English translations.

(89) *Reciprocal Translations*

- a. English Original Verbless → Russian Translation verbless / verbal
- b. Russian Original Verbless → English Translation verbless / verbal

It is necessary that the patterns in (a) be asymmetrical to the patterns in (b).

3.3.2 Directions of Correspondence

Secondly, the *translations themselves* (being genuine and legitimate language instances following the above discussion) are also presently used as sources of verbless sentence data.

Therefore, verbless sentences in English translations are compared to their Russian source language correspondences, and verbless sentences in Russian translations are compared to English source language correspondences.

(89) *Reciprocal Correspondences*

- c. English Translation Verbless → Russian Original verbless / verbal
- d. Russian Translation Verbless → English Original verbless / verbal

It is necessary that the English-Russian patterns in (a) match with the patterns in (c), and similarly that the Russian-English patterns in (b) match with the patterns in (d).

The involvement of this type of comparison drives the need for the use of the term ‘correspondence’ and ‘correspondence pattern’, as opposed to ‘translation’ and ‘translation pattern’. The fact that a translated text may be the initial source of the verbless phenomenon means that its verbless sentences, just like the verbless sentences from an original text, may have either verbal or verbless correspondences. As a result, from this point on, the term ‘**(translation) correspondence**’ will be used instead of ‘translation’ in order to refer to the link between a verbless sentence (in an original text or a translated text) and its respective counterpart in an original text or a translated text.

Notably, such analysis requires additional work from the point of automatic processing. It is necessary to find not only the verbless sentences in the original text and match them with their translations, but also find the verbless sentences in the translated counterpart and trace them back to their source correspondences.

3.3.3 Third-Language Reciprocal Correspondences

Thirdly, translations from a third language are presently used in order to further mitigate possible Russian or English source language interference. The third language is neither English nor Russian (in the present case, this language is French).

(89) *Third-Language Correspondences*

- e. English Translation-from-3L Verbless → Russian Translation-from-3L verbless / verbal
- f. Russian Translation-from-3L Verbless → English Translation-from-3L verbless / verbal

It is necessary that the patterns in (e) and (f) be asymmetrical with each other. However, a match should exist between the language specific verbless sentence correspondence patterns, i.e. English verbless sentence correspondence patterns should match in (a), (c) and (e), Russian verbless sentence correspondence patterns should match in (b), (d) and (f).

3.3.4 Reference Features

Furthermore, the *features* of translated texts are presently compared to the features of original texts in the same language. While not focused on *translation correspondence patterns* per se, this measure makes it possible to analyze in what respects the language of the translations may be different from source language.

(89) *Reference Translations against Originals within the same Language*

- g. English Translation Verbless → English Original Verbless
- h. Russian Translation Verbless → Russian Original Verbless

Concerning (g), the verbless sentence features of English translations should match with the verbless sentence features of English originals; and the same for the Russian features of verbless sentences in (h).

Furthermore, this measure makes it possible to use translations (and not just originals) in a contrastive analysis that compares Russian and English in terms of (i) verbless sentence use, (ii) the key lexical features of verbless sentences, and (iii) the key segments within verbless sentences.

3.3.5 Multiple Translations, Genre, Time, Number of Originals

In order to control for the interference of translation language itself (as opposed to the source language as above), several measures have also been taken in the present study.

The first is to include multiple translations of the same work whenever possible so as to control the influence of any one particular translator. The following illustrates what this means for one original text with multiple translations.

(90) *Multiple Translations*

One Original and One Translation

1. Russian Original Verbless → English Translation #1 verbless / verbal
2. English Translation #1 Verbless → Russian Original verbless / verbal

One Original and Two Translations

1. Russian Original Verbless → English Translation #1 verbless / verbal
2. Russian Original Verbless → English Translation #2 verbless / verbal
3. English Translation #1 Verbless → Russian Original verbless / verbal
4. English Translation #1 Verbless → English Translation #2 verbless / verbal
5. English Translation #2 Verbless → Russian Original verbless / verbal
6. English Translation #2 Verbless → English Translation #1 verbless / verbal

One Original and Three Translations

1. Russian Original Verbless → English Translation #1 verbless / verbal
2. Russian Original Verbless → English Translation #2 verbless / verbal
3. Russian Original Verbless → English Translation #3 verbless / verbal
4. English Translation #1 Verbless → Russian Original verbless / verbal
5. English Translation #1 Verbless → English Translation #2 verbless / verbal
6. English Translation #1 Verbless → English Translation #3 verbless / verbal
7. English Translation #2 Verbless → Russian Original verbless / verbal
8. English Translation #2 Verbless → English Translation #1 verbless / verbal
9. English Translation #2 Verbless → English Translation #3 verbless / verbal
10. English Translation #3 Verbless → Russian Original verbless / verbal
11. English Translation #3 Verbless → English Translation #1 verbless / verbal
12. English Translation #3 Verbless → English Translation #2 verbless / verbal

One Original and Four Translations

1. Russian Original Verbless → English Translation #1 verbless / verbal
2. Russian Original Verbless → English Translation #2 verbless / verbal
3. Russian Original Verbless → English Translation #3 verbless / verbal
4. Russian Original Verbless → English Translation #4 verbless / verbal

5. English Translation #1 Verbless → Russian Original verbless / verbal
6. English Translation #1 Verbless → English Translation #2 verbless / verbal
7. English Translation #1 Verbless → English Translation #3 verbless / verbal
8. English Translation #1 Verbless → English Translation #4 verbless / verbal

9. English Translation #2 Verbless → Russian Original verbless / verbal
10. English Translation #2 Verbless → English Translation #1 verbless / verbal
11. English Translation #2 Verbless → English Translation #3 verbless / verbal
12. English Translation #2 Verbless → English Translation #4 verbless / verbal

13. English Translation #3 Verbless → Russian Original verbless / verbal
14. English Translation #3 Verbless → English Translation #1 verbless / verbal
15. English Translation #3 Verbless → English Translation #2 verbless / verbal
16. English Translation #3 Verbless → English Translation #4 verbless / verbal

17. English Translation #4 Verbless → Russian Original verbless / verbal
18. English Translation #4 Verbless → English Translation #1 verbless / verbal
19. English Translation #4 Verbless → English Translation #2 verbless / verbal
20. English Translation #4 Verbless → English Translation #3 verbless / verbal

In addition, keeping in mind the directions of comparison (a) – (h) above, the inclusion of multiple translations significantly further multiplies the correspondence patterns to be examined.

Secondly, in order to mitigate potential translation interference, the present study also includes texts from multiple genres. The term ‘genre’ is used in the sense of an externally attributed distinction based on the general function of the text, as defined by Douglas Biber (1989):

Text categories readily distinguished by native speakers of a language [...] defined primarily on the basis of external format [...] However, [...] texts within genres can differ greatly in their linguistic characteristics [...] and different genres can be similar linguistically. (Biber, 1989: 6)

Texts from the following three genres are included in the present study:

(91) *Multiple Genres*

- a. Discourse-based Fiction
- b. Scripted-speech Fiction (i.e. theater plays)
- c. Narrative-based Fiction

The inclusion of several genres means a large additional set of contrastive correspondences since the translation patterns of the texts within each genre must be compared with one another.

Varying the time period is also of key importance with regard to disentangling linguistic traits from traits of translation. Notably, this concerns both the time periods of the original works and those of the translations. The present study includes texts that range from the late 18th to 21st century.

(92) *Multiple Time Periods*

- a. Originals: late 19th to 21st century
- b. Translations: early 20th to 21st century

Finally, the number of works to be examined is obviously another factor that is of utmost importance. The present study includes a total of 32 texts which correspond to 13 separate works.

(93) *Multiple Works*

- | | |
|---------------------------------------|------------------------------|
| a. Russian Originals: | 6 works (6 original texts) |
| b. English Originals: | 6 works (6 original texts) |
| c. Russian Translations from English: | 6 works (8 translated texts) |
| d. English Translations from Russian: | 6 works (8 translated texts) |
| e. Russian Translations from 3L: | 1 work (2 translated texts) |
| f. English Translations from 3L: | 1 work (2 translated texts) |

Each work and each of its translations must be compared with the other works and their translations, and thus integrated into the directions of comparison illustrated in (89) – (93).

From the perspective of corpus methodology, the addition of each supplementary text to a parallel-text study represents much more than what at first may seem to be a single addition. This is because the directions of comparison matter. The directions of comparison are the key way of disentangling linguistic patterns from potential translation induced patterns. The directions of comparison in (89) – (93) are methodological features put in place to guard against translation interference. Consequently, this means that in order to get the benefits from each

additional text, the multiplication of correspondence patterns grows at an exponential rate with each text.

Notably, applying corpus methodology to parallel-texts, for the above reasons, is more demanding than when texts are studied from a monolingual corpus perspective that does not maintain the link between source and target. One must tread carefully so that multiplying the quantity of the works examined does not come at the cost of the other parameters recommended by translation-based studies and, consequently, lead to significant sacrifices in terms of quality.

The above measures strive to mitigate the main risk that has been raised concerning contrastive analysis, i.e. the problem of disentangling linguistic and translation phenomenon. In this way it is hoped that it will be possible to get the fundamental benefits of both contrastive and corpus-based analysis.

3.4 FRAMING TRANSLATION CORRESPONDENCES

As proposed in *Part 1*, verbless sentences are syntactically and semantically indeterminate with respect to a verbal predicate. On the present account, the meaning of the latter may be conversationally implicated but not semantically encoded in any lexical element of the verbless sentence.

This means that semantic properties such as tense and aspect may be conversationally implicated through the pragmatic insertion of a verbless sentence into a specific situation, but, contrary to previous studies (e.g. Lefeuve & Nicolas, 2004; Balvet et al., 2011; Rakhilina, 1999), such meaning cannot be part of the semantically entailed content without the presence of a verb.

The present study proposes to use translations in order to explore the potential meaning that is pragmatically implicated by a verbless sentence. In particular, *verbal* translations of a verbless sentence are presently treated as revealing a potential meaning, which may be activated in a relevant pragmatic context. In other words, the verbal translations *semantically entail* a meaning that is only *conversationally implicated* in a verbless sentence. Therefore, although verbless sentences themselves are semantically incomplete for the types of meaning that require a verb (i.e. tense, aspect, lexical meaning of the predicate), their translations

indicate the *possible pragmatic enrichments* of the semantic content, in the sense of Soames (2008: 456).

Translation provides a unique opportunity to systematically explore the potential pragmatically implicated meaning of a verbless sentence. We discuss alternative approaches to studying the meaning of verbless sentences in (Bondarenko & Celle, 2020: 327–328).

However, it must be stressed that the meanings revealed in translation are *potential*. A verbal translation is an interpretation of the translator regarding the original verbless sentence. In translation, a choice is made regarding the potential semantic predicate. The present study attempts to analyze the choices that were made by several translators, in the systematic way described above, in an attempt to reveal reoccurring patterns.

The ultimate goal of such an approach is to discover the potential predicative meaning of verbless sentences through their verbal translations by comparing what is semantically entailed in a verbal translation and what is implicated in an original verbless sentence.

3.5 CONTEXT OF VERBLESS SENTENCE *TRANSLATIONS*

Another notable issue is yet another third type of context that is required for the contrastive study: the context of the translation correspondence of a verbless occurrence.

Access to the utterance context of the translation correspondence is necessary in order to (a) assess whether or not there exists a verbal antecedent and, consequently, distinguish whether a potentially verbless translation correspondence is an elliptical or a non-elliptical sentence, as well as (b) to compare the semantico-pragmatic features of the verbless sentence translation for contrastive analysis. Without access to the context of the translation, contrastive analysis for semantico-pragmatic features becomes impossible.

Retrieving the *context of a verbless sentence translation correspondence* poses an additional computational challenge. The previously mentioned retrieval of the verbless sentence and its ‘sentence-centered’ context, must be combined with the retrieval of the translation correspondence of the verbless sentence, as well as the ‘sentence-centered’ context of that translation correspondence.

While a lot of progress has been made in the automatic alignment of an original text with its translation (in particular at the paragraph-level, but also at the sentence-level, the latter

however still typically requiring some manual correction), the need to retrieve verbless sentences automatically does not yet integrate into the existing work on alignment. Verbless sentences present a new and unique set of issues from the perspective of alignment. They involve custom segmentation, custom annotation, and, crucially, their alignment must permit the subsequent retrieval of absence, as well as access to the context, in any of the translation correspondences.

This issue once again highlights the need to integrate the computational and the theoretical linguistic domains. What appears obvious from a manual perspective on a small-scale (e.g. access to context) is not always evident in the transfer to automatic processing. Notably, it is qualitative analysis that brings the thornier computational limits to light. Both domains would stand to gain from the integration.

Part 2: Chapter 4 will illustrate the present solution to this third type of context. It involved the development of a particular data structure for sentence level alignment, working with developers on updates to some of the most powerful existing software, as well as the creation of new programs to target precisely contrastive corpus analysis of verbless sentences.

CHAPTER 4

AUTOMATIC PROCESSING

The present chapter proposes a new method of automatic verbless sentence retrieval. The method is replicable on any raw text.

It strives to resolve the problem of the retrieval of absence, while minimizing the biases that may arise from the use of corpus tools for linguistic aims (e.g. Anthony, 2012; *Part 2: Chapter 2* above).

In short, the essence of the present automatic method consists of classifying appropriately delimited and morpho-syntactically tagged sentence units, while maintaining their place in the overall structure of the corpus. All of the steps were recorded in detail and are summarized in the present chapter. Attention is drawn to a typology of specific key issues for automatic-processing which surfaced in the course of the study.

This method was developed with a particular concern for accuracy and involved several pilot studies. The initial proposal underwent several changes brought about progressively as the need for the automatization grew more significant with the expansion of the corpus from a 76,000-word pilot, from which a series of verbless sentences was manually retrieved, to the present 1,4-million-word fully-aligned corpus. The final method, presented here, was carried out on the full corpus from scratch, which means that the same standards apply to the corpus in its entirety.

Before outlining the steps involved, this section starts by presenting how these particular steps were arrived at.

4.1 PILOT STUDIES AND ACCURACY MEASURES

The *initial pilot* aimed at a semantico-pragmatic qualitative study of verbless sentences in Fyodor Dostoyevsky's Russian novel *Brothers Karamazov* and its English translation by Richard Pevear and Larissa Volokhonsky. More precisely, the data involved consisted of the verbless sentences and their translation correspondences in the first fourteen chapters of the novel, which measured 34,425 words of source Russian and 42,050 words of target English.

For this initial study, verbless sentences were manually extracted, and their translations were also manually retrieved. A spreadsheet was used to compile together the data consisting of an exhaustive series of verbless sentences from the extract (as well as their left and right contexts, their translations and the contexts of the translations). This manually-extracted verified series was then used as a gold standard, i.e. truth values, for the verification of each additional automatized step.

Three types of accuracy readings – *Precision*, *Recall*, and *F-Measure* – were used in order to evaluate the results during the automatization process. These accuracy readings were made in accordance with (Bird et al., 2009).

The calculations involve several values. *True Positives* (TP) correspond to the sentences *correctly identified as verbless* in automatic treatment. *False Positives* (FP) refer to sentences that were *incorrectly identified as verbless* in automatic treatment, i.e. in reality they are verbal sentences. *False Negatives* (FN) are *verbless sentences that were missed* in automatic treatment and were not identified as verbless. These values are arrived at by comparing the automatic values, i.e. the verbless sentences found in automatic retrieval, to the manual gold-standard truth values, i.e. the actual verbless sentences.

Regarding the False values, it is notable that False Positive and False Negative types of errors are not of the same gravity when it comes to an entirely *qualitative* study of the phenomenon. The former, FP (also called Type I errors) are verbal sentences that are mislabelled verbless in automatic processing. In qualitative analysis of a series of verbless sentences, some FPs that find themselves among verbless data may be manually eliminated during the analysis.

However, FN (or Type II errors) pose a much more serious problem since these types of errors cannot be so easily spotted in qualitative analysis (i.e. they require a full accuracy evaluation against a gold standard). Errors of this type are extremely dangerous for all types of analysis because they *systematically* exclude *entire types of verbless sentences* based on some

technical specification that remains unknown until the moment that a full accuracy analysis is actually performed.

For instance, as was presently revealed, direct speech quotations systematically pose problems for the automatic retrieval of verbless sentences. The issue is of colossal gravity since, as was also presently observed, verbless sentences are particularly closely correlated with direct speech. This means that without an accuracy analysis against manual values, a dominating proportion of verbless sentences would remain forever hidden from the researcher without one having the slightest suspicion.

Notably, for an analysis that is based on a series of sentences, it is the Type II-FN errors that one must be particularly careful of, but the Type I-FP errors may be corrected during the analysis itself. The importance of this difference in error types becomes especially pronounced when moving from a qualitative analysis of a series to a full corpus and the use of corpus tools. Such a move changes the perspective on accuracy. When corpus tools are used for the analysis of the data (e.g. to determine the key words, n-grams, or correlations) both types of errors become of serious importance. The advantage with respect to False Positives, that one has when analysing a series, disappears when corpus analysis of the data is undertaken. One cannot eliminate the erroneous verbal sentences mixed up in the verbless data, as one could when analyzing a series. However, the silent danger of False Negatives remains for both.

Presently, the Automatic Values (TP, FN, FP) were regularly compared with manual and the accuracy was evaluated according to the following calculations during several stages of retrieval:

(94) *Accuracy Readings*

- a. Precision = $TP / (TP + FP)$
→ *Are the retrieved sentences really verbless?*
- b. Recall = $TP / (TP + FN)$
→ *Are verbless sentences excluded in retrieval?*
- c. F-Measure = $(2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$
→ *All in one score?*

Precision is a measure that indicates how many of the verbless sentences automatically identified were actually verbless. *Recall* indicates how many of the existing verbless sentences were automatically identified. Finally, *F-Measure* combines precision and recall into a single score.

The automatic results values were recorded manually for each sentence, along with the details of the errors. A typology of the errors was compiled from these results. This typology was subsequently used to directly target the sources of error that precisely pose problems for verbless sentence retrieval.

A preliminary accuracy assessment of the very first attempt at automatic retrieval emphasized the inescapable need for further measures. It was made on an extract from the corpus which at that point was segmented into sentences simply by the period and morpho-syntactically tagged. The extract for the preliminary assessment measured 11,146 words and contained 20 true verbless sentences. Although automatic retrieval returned 8 sentences, none of them were actually verbless (i.e. all 8 were False Positives and all of the true verbless sentences were missed). Thus, the initial automatic attempt at extracting verbless sentences showed a *zero score* in terms of all three accuracy readings, i.e. Precision, Recall and F-Score. Such unacceptable results suggested that accuracy should be a serious part of automatic treatment of verbless sentences. They also emphasized that sentence segmentation must necessarily go beyond the period; the latter can only serve as a first step in the delimiting process.

Regular checks against the manual results refined the typology of errors and made it possible to significantly improve the very poor accuracy of automatic verbless sentence retrieval. A full accuracy assessment of the entire 76,000 pilot corpus was carried out three times. We discuss the results of these three assessments in (Bondarenko, 2019).

The most important sources of errors in automatic verbless sentence retrieval, as revealed by the present study, are summarized in *Figure 9*. The figure presents the basic typology.

TYPOLOGY OF ERRORS REVEALED IN AUTOMATIC RETRIEVAL OF VERBLESS SENTENCES

1. Major Punctuation

This corresponds to the period, question mark, exclamation mark, as well as their various combinations (e.g. !?!).

2. Direct Speech

This involves various ways of marking basic direct speech and turn change, including quotations marks, dashes, commas.

3. Embedded Narration

Narration is often embedded within a direct speech sentence. It is necessary to separate the narration from the direct speech. This includes cases when narration is found at the beginning, at the end, or in the middle of direct speech. Particular issues concern the latter situation, i.e. when narration separates a single direct speech sentence into two parts. In such cases, the narration must be separated and the direct speech must be reunited into one sentence. The reuniting of a direct speech sentence that has been divided by narration demands particular attention.

4. Clitics

Verbal clitics and negation must be disambiguated in English. Of particular issue is the need to distinguish the possessive marker from a verbal clitic.

5. Footnotes

This concerns the presence of footnotes and their in-text marking.

6. Abbreviations

The presence of abbreviations is particularly problematic for segmentation.

7. Ellipsis Symbol

The three-dot ellipsis symbol [...] poses particular problems as it may be embedded in the middle of a single sentence. It is responsible for a lot of phantom verbless sentences in automatic results.

8. Capitalization

Capitalized verbs pose problems for morphosyntactic tagging and result in particularly important misclassification of verbless sentences.

9. Homonyms and Other Tagging

Particular word forms, of which homonyms and English verbs in the imperative form are especially important, constitute a consequential source of verb-related tagging errors and thus, for the classification of verbless sentences. Dashes within words are a particular problem with regard to morphosyntactic annotation in Russian.

Figure 9. Typology of Errors for Automatic Processing of Verbless Sentences

The three full accuracy assessments correspond to different stages in the development of the automatic method. The first full evaluation followed basic segmentation by major punctuation marks, i.e. the resolution of the first type of error in the above typology. The second full evaluation followed the resolution of errors two through seven. The third corresponds to the resolution of the final two types of errors. The accuracy results of verbless sentence retrieval at these stages are presented in *Table 1*.

	ENGLISH			RUSSIAN		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
STAGE 1						
Basic segmentation by major punctuation (?!)	.43	.35	.38	.85	.81	.83
STAGE 2						
Additional segmentation and custom tagging to resolve direct speech, embedded narration, clitics, footnotes, abbreviations, ellipsis symbol	.79	.94	.86	.96	.95	.95
STAGE 3						
Capitalization and other tagging issues	.97	.98	.98	.99	.99	.99

Table 1. Automatic Extraction Accuracy

As shown in the table, the resolution of basic sentence segmentation (i.e. period, exclamation mark, question mark) at Stage 1 resulted in Precision score of 0.43 for English. This score indicates that out of the total sentences that were automatically extracted, only 43% were *actually verbless*. The other 57% (more than half) were false positives, that is verbal sentences that were misclassified as verbless.

The Recall score for English at Stage 1 is even worse at 0.35. This indicates that 65% of existing verbless sentences *were not extracted* automatically. In other words, any quantitative or qualitative analysis of verbless sentences at this point could account for only 35% of the existing verbless sentences. What is most dangerous is the fact that the other 65% would remain forever lost without the knowledge of the researcher, unless a manual analysis of the present type was undertaken.

Furthermore, any automatic statistically-based analysis of verbless sentences at this stage (e.g. key words, n-grams) would be very unsound. It would be based on only 35% of the existing verbless sentences, and 57% of this small portion of retrieved sentences, that would be analyzed as being verbless, would actually be *verbal*. What is even worse is that these are systematic errors, which means that the patterns revealed by such data would pertain not to natural language verbless sentences, but would actually be describing an amalgamation between natural language verbless sentences and automatic processing biases.

Clearly, the accuracy results presented in *Table 1* emphasize the necessity of going beyond basic segmentation if automatic retrieval and corpus tools are to be used. Verbless sentence errors of the types 2 through 7 in the typology were resolved (a targeted automatic approach was developed for each type of error) and, as evident from the table, the scores at Stage 2 show significant improvement.

The approach to the automatic resolution of two particularly important types of verbless sentence errors must be mentioned. The first concerns the errors of type 3 in the above typology, i.e. embedded narration. An example of the problem of embedded narration is illustrated in (95a) and the final form arrived at by means of the developed automatic solution in (95b).

(95) *Embedded Narration*

a. Problem

‘Dead on the appointed hour,’ exclaimed Fyodor Pavlovich, ‘but still no sign of my son Dmitry Fyodorovich.’

b. Solution

‘Dead on the appointed hour,’ _ ‘but still no sign of my son Dmitry Fyodorovich.’

_ exclaimed Fyodor Pavlovich _

Sentences such as (95a), in which narration was embedded in the middle of a single direct speech sentence, were automatically separated into two segments. Notably, it was necessary not only to separate the narration, but also to bring together the two parts of the direct speech sentence. This was done automatically in the ‘Segmentation’ step, which will be described below, via a series of rules developed precisely on such data.

The other important type of error with regard to verbless sentences, that concerns particularly English, is the problem of clitics. Clitics include the consistently verbal [*'m*], [*'re*], [*'ll*], [*'ve*], [*'d*] and [*'d've*], the negation [*n't*], as well as the [*'s*]. Clitics are automatically morpho-syntactically annotated, along with their host (e.g. [*You're*], [*don't*], [*when's*], [*wife's*], [*yesterday's*], etc.), as nominal forms. Consequently, this drastically increases the number of false positive errors for verbless sentence retrieval in English.

The present solution involves automatically separating the clitic from its host in the segmentation step, e.g. separating [*You*] from [*'re*]. This thereby creates a new form which is unknown to the tagger, i.e. [*'re*]. The new form then requires special annotation on the grammatical category level (the re-annotation was done with help of the software package Trameur, as will be discussed shortly).

However, the clitic [*'s*] further complicates the problem since in addition to representing a verb form, e.g. ‘*Your wife's a ticklish woman*’, it can also express possession, e.g. ‘*Your wife's shoes*’. To resolve the [*'s*] clitic, a list of host and clitic combinations was put together in which [*'s*] is most likely to only ever constitute a verb form, e.g. [*he's*], [*where's*], [*there's*], etc. This list was then used in automatic segmentation. The cases where [*'s*] corresponds to a condensed verb were transformed into the new form [*'sv*], which meant that they no longer had the same form as the clitic [*'s*] of possession. The new form [*'sv*], unrecognized by the tagger, was then re-annotated on the grammatical category level with the help of Trameur.

Notably, this solution does not resolve all ['s] clitics. Cases such as ‘*Your wife’s a ticklish woman*’, ‘*Everything’s pretty clear*’, ‘*His whole theory’s a sham*’, etc., escape the proposed automatic solution and could only be resolved at the final post-editing stage. Since they constitute false positives, they show up in automatic extraction, were spotted in verification and corrected semi-automatically with the help of Trameur.

Following Stage 2, the English F-Measure score of 86% (i.e. the mean of precision and recall) is a significant improvement from the 38% following basic segmentation of Stage 1. Nevertheless, 21% of English sentences automatically retrieved were still in reality verbal sentences, and 6% of verbless sentences were still systematically lost even at Stage 2. Therefore, further work was undertaken in order to develop a solution and arrive at Stage 3.

The remaining problems following Stage 2 concerned morphosyntactic tagging and were mostly a source of False Positives. These errors corresponded to systematic mis-tagging of sentence initial capitalized verbs, homonyms (notably, more problematic in English than Russian) and also include English imperative verb forms, and other types of errors. Some typical examples in both languages are presented in (96).

(96) *Stage 3 Key Tagging Errors*

- a. Go<NN> back to your husband and care<NN> for him.
- b. Are<NP> you out of your mind, or what?
- c. I love<NN> you. Old lies<VVZ> and conventional bows!
- d. Развеселила<Ncfsnn> ты мое сердце, мать.
- e. О чем плачешь-то<Ncmsnn>?

In (a) both the capitalized-imperative verb [*Go*] and the homonymous-imperative verb [*care*] posed problems for the tagger; as did the capitalization on the verb [*Are*] in (b). Two problematic homonyms are shown in (c). Illustrated in (d) is a problematic capitalized verb in Russian; and in (e), a verb with a particle attached to it by means of a dash.

The tagging errors that caused False Positives were identified, they were recorded, and then corrected using Trameur. Notably, these types of post-editing corrections were carried out on the entire 1,4-million-word corpus as part of the ‘Tagging’ step, as will be discussed below in *Part 2: Section 4.5*. Therefore, the full corpus was resolved for the full typology of errors, including the errors identified at Stage 3.

The final accuracy assessment following Stage 3 resolution showed that the present method, proposed below, makes it possible to automatically recall 98% of English and 99% of

Russian verbless sentences. Furthermore, the precision scores indicate that of 97% of English and 99% of Russian verbless sentences automatically identified are actually verbless.

The present full 1,4-million-word corpus was prepared using a detailed record of steps of which only the essential points are summarized here. All steps were applied in a standardized way to the entire corpus from scratch; minor modifications were sometimes necessary for the Segmentation due to the graphical representation of dialogue and formatting of the different texts (all modifications were recorded). A randomized accuracy sample is beyond the present scope; however, it must be said that the full corpus received even more manual attention than the pilot due to the process of full sentence-alignment. Therefore, although the above accuracy assessments were carried out on the pilot 76,000-word corpus, they are treated as good estimations of the accuracy values for the present retrieval from the full 1,4-million-word corpus.

The present section strove to show that the automatic method of verbless sentence extraction was developed with qualitative considerations in mind at each step. What at first glance may seem to be a minor and purely technical issue, such as for instance the treatment of clitics, translates into significant errors for verbless sentences and therefore seriously compromises the analysis results. The next section will summarize the essential basic steps presently proposed for retrieving verbless sentences automatically from any raw text.

4.2 SENTENCE SEGMENTATION

The first step consists of custom segmenting a raw text into sentences. This means marking the beginning and the end of the sentence using a single ‘delimiter’, i.e. a single graphical symbol (in the present case the [\$]) that represents all of the many ways that the sentence may be marked in natural language.

This was performed automatically using a single sentence delimiting script which was written in regular expressions and Perl.

First, a series of algorithms was developed over the course of the several pilot studies and the regular manual checks. Each algorithm consists of custom combinations of graphical markers (e.g. period, exclamation marks, question marks, quotations, capital letters, apostrophes, commas, dashes, etc.) that each target a particular way of marking the sentence

or a particular problem for verbless sentence retrieval. It is with this *rule-based* segmentation that some of the most important challenges for verbless sentences began to be addressed (including segmenting direct speech, separating embedded narration, clitics, abbreviations, amongst the many other issues).

The writing of the algorithms started from basic Regular Expressions (e.g. Goyvaerts, 2015). In fact, two regular expressions are necessary for each target issue: one to find the right problem cases, and a second to replace the problem cases with a solution.

For a simple example, in order to segment a text into sentences by the period, it is necessary to apply the two regular expressions in (97) to the text:

(97) *Regular Expressions in Period Segmentation*

- a. Find: \. \s
- b. Replace: \. \n\n\\$

In (a), all instances of the period followed by a space would be found. In (b), they are replaced by a period, a space, two new lines and the delimiter [\$] that officially marks the end of the sentence.

This find and replace operation may be done with any text-editor. By substituting, for instance, a question mark for the period in (97), and repeating the find and replace, one further segments questions. The required regular expressions were expanded to target cases when, for instance, the final punctuation is followed by an asterisk [\. *\s], by a numbered footnote [\. \d\s], a quotation mark [\. \?"\s], a quotation followed by footnote [\. \?"\d\s], an ellipsis symbol followed by parenthesis and capital letter [(\.\.\s*)(\([АЯ-Ё])], and so on.

The various ways of marking a sentence which were to be targeted were revealed through the regular accuracy checks with manual results. In this way, solutions were found for instance for various (most consequential in terms of verbless sentence accuracy) cases of embedded narration. Example (98a) shows part of the solution for one of the problem embedded narration cases in English, and (98b) for one of the problem embedded narration cases in Russian.

(98) *Regular Expressions in Embedded Narration Segmentation*

a. *English embedded narration of the type:*

‘Dead on the appointed hour,’ exclaimed Fyodor Pavlovich, ‘but still no sign of my son Dmitry Fyodorovich.’

Find: (')(.*, ')(.*\.)
Replace: \$1 _ ' \$3 \n\n\$ _ \$2 _

b. *Russian embedded narration of the type:*

– Да подожди, подожди, – тревожно перервал Алеша, – из чего ты-то всё это видишь?..

Find: (\\-\\s+[А-ЯЁ\«][А-Яа-яёЁ0-9\\s\\,|~\\(\\)\«\»|;:]*,\\s+\\-\\s+)([А-Яа-яёЁ0-9\\s\\,|~\\(\\)\«\»|;:]*\\-\\s)([А-Яа-яёЁ0-9\\s\\,|~\\(\\)\«\»|;:]*\\.)
Replace: _ \$2 _\n\n\$ \$1 _ - \$3

Using regular expressions, it was possible to develop solutions to target the key problems that were revealed in accuracy checks, without disturbing the rest of the segmentation.

In terms of *Table 1* above, the readings from Stage 1 correspond to a segmentation that targets just the simple period, question and exclamation sentences and consists of 24 regular expressions for English and 30 for Russian. In comparison, the full segmentation consists of 650 regular expressions for English, and 316 regular expressions for Russian. Such a difference is particularly due to the resolution of clitics in English.

All of the regular expressions were then combined into a single Perl script which carried out the operations at once. To do this, the Perl language find-and-replace function in (99a) was applied to each regular expression pair.

(99) *Perl Find-and-Replace*

- a. `$_ =~ s/find_this/replace_with_this/g ;`
- b. `$_ =~ s/\\.\\s/\\. \\n\n$ /g ;`

For instance, the two regular expressions required for the period segmentation in (97) translate to (99b) in Perl.

Two separate scripts were necessary, one for English and one for Russian, with different rules developed for each language. Furthermore, with regard to each text (i.e. novel or

translation), the scripts were run individually since they sometimes needed minor changes. Slight alterations needed to be made depending on the use of graphical markers even within the same language. For example, depending on whether the text used quotations marked as ["], ['], [«], [–], certain lines of the script were activated and others muted.

The sentence-segmented output text was then searched to verify key issues influencing the accuracy of verbless sentences, including underscores (which are part of segmenting embedded narration), different types of quotations (which mark direct speech), the ellipsis symbol (which creates phantom verbless sentences) and the elusive clitic [’s] (to catch the instances uncaught by the rules), and the errors were then manually corrected.

4.3 ALIGNMENT

The next step was to prepare the sentence-segmented texts for their sentence-level alignment. In order to be able to keep the custom sentence delimiting above (including for instance the direct speech and narration divisions), it was necessary to match the sentence segments manually.

Alignment minimally requires an equal number of sentences across the original and its multiple translations. For example, the Russian Original *Brothers Karamazov* extract containing 23,153 sentences, its English Translation 1 containing 24,397 sentences, and its English Translation 2 containing 23,940 sentences, need to be brought to the same number of sentences for alignment. Moreover, the sentences, of course, need to correspond with one another across the original and all translations.

The above *sentence*-segmentation created a good start to the *alignment*-segmentation. However, clearly the number of sentences in an original and its translation do not correspond. Furthermore, in translation, sentences are frequently divided, other sentences are combined together, and, even if the number of sentences happens to be the same, the order of the sentences is very often changed. In order to find the places where there is a problem in correspondence, i.e. an extra sentence, a missing sentence, a change in order, it is necessary to review the *entire* alignment.

The alignment-segmentation was done manually by comparing the source text and all of its translations. In order to align the texts, *without altering* the existing sentences, nor the

original order, the solution undertaken was to insert as many extra ‘empty’ segments as necessary. This solution is illustrated in *Figure 10*.

	ORIGINAL	TRANSLATION 1	TRANSLATION 2	TRANSLATION 3
1	‘Comrades, you have heard already about the strange dream that I had last night.	– Товарищи!	– Товарищи!	– Товарищи, вы все уже слышали, что прошлой ночью мне привиделся странный сон.
2	Z	Вы все уже слышали, что вчера мне приснился удивительный сон.	Как вы слышали, прошлой ночью мне приснился странный сон.	Z
3	But I will come to the dream later.	Но об этом позже.	Но, прежде чем рассказать о нем, позвольте всего несколько слов.	Но к нему я вернусь позже.
4	I have something else to say first.	Сначала я хочу поведать вам вот о чем.	Z	Первым делом я должен вам сказать вот о чем.
5	The life of an animal is misery and slavery: that is the plain truth.	Правда заключается в том, что удел животных – страдание и рабский труд.	Мы живем в рабстве и нищете – вот голая правда.	Жизнь наша характеризуется так – нищета и рабство.
6	Z	Z	Z	Такова истина.
7	But is this simply part of the order of nature?	Но, может быть, таков закон природы?	Но, может быть, это в природе вещей?	Но таков ли должен быть истинный порядок вещей?
8	‘Comrades,’ _ ‘it is half~past six and we have a long day before us.	– Товарищи звери! –	– Товарищи, – _ – сверим, так сказать, наши часы...	– Товарищи, – _ – уже полшестого, и нас ждет долгий день.
9	_ said Snowball _	сказал Снежок.	_ сказал Цицерон, _	_ сказал Сноуболл, _
10	Z	Сейчас половина седьмого, и впереди у нас долгий день.	Сейчас половина седьмого, впереди у нас трудный день – мы начинаем заготовку сена.	Z
11	Today we begin the hay harvest.	Сегодня мы начинаем сенокос.	Z	Сегодня мы начинаем жатву.

Figure 10. Alignment segmentation for verbless sentences

Although extremely costly in terms of time and attention, the solution of inserting ‘empty’ segments in order to fix the alignment is straightforward enough.

However, the fact that the alignment must subsequently be used to search for *verbless* sentences creates another complication. In fact, any empty segments would necessarily be identified as verbless sentences in automatic retrieval. Therefore, the alignment would seriously interfere with the task of verbless sentence retrieval.

As a result, the solution undertaken was to fill the extra segments and subsequently neutralize this content. The ‘empty’ alignment-segments were filled with the character [Z]. This character was then neutralized during the Tagging step by being custom-annotated as a mythical lemma [ZZ] and mythical grammatical category [ZZZ] with the help of the Trameur software package. Thus, the Z-segments, while *permitting alignment*, can be successfully *excluded* from verbless sentences in automatic retrieval.

Furthermore, the [Z] character serves another important purpose in the retrieval of verbless sentences and their translation correspondences. Acting as a *searchable* anchor, this character is one of the keys to the subsequent retrieval and *qualitative* analysis of the verbless sentence *correspondences*.

More specifically, the [Z] is of crucial importance when the verbless sentence has an empty alignment-correspondence. For instance, if the order was changed or a sentence was divided, the correspondence of a verbless sentence may find itself displaced. This problem situation is illustrated in *Figure 10* above, when the correspondences of the verbless sentence in the 6th line ‘Такова истина’ (*takova istina*; lit. ‘*this_way truth*’) are found in the 5th line, as parts of other sentences, e.g. the English correspondence ‘*The life of an animal is misery and slavery: that is the plain truth*’.

In order to subsequently retrieve the displaced translation of a verbless sentence for qualitative analysis, it is necessary to view the context of the ‘empty’ alignment. To do this a series of tests was presently run on a test corpus and a new function was introduced into the Trameur software package. This function is called ‘**Annotation de Sections**’ and is described in a complement to the Trameur manual (Fleury, 2019b).

The new Trameur function targets the specifically retrieved segments within a corpus (in the present case, these are the verbless sentences and their multiple aligned-correspondences). After automatically retrieving the verbless sentences and their correspondences, the function makes it possible to add an extra level of annotation just to these select segments. The first word of each segment is annotated (hence the requirement for the ‘empty’ alignment slot to be filled). Subsequently, concordance lines are found by searching for this new annotation. The

verbless sentence translation correspondences are thus revealed in the left and right context of the concordance lines, despite being displaced in terms of alignment.

4.4 STRUCTURE & TRAMEUR

Once the raw texts were appropriately sentence-segmented for (a) verbless sentence retrieval and (b) alignment, the next step was to unite the individual texts into a single structure which will subsequently be tagged and processed by Trameur.

4.4.1 Trameur

Trameur is a statistical text analysis software package that is particularly specialized for the processing and analysis of multilayer-annotated multilingual parallel and comparable corpora (Fleury, 2013; Fleury & Zimina, 2014; Zimina & Fleury, 2015; Fleury, 2019a).

Attention to the software that one uses is critical for linguistic analysis. As emphasized by Laurence Anthony (2012), the choice and knowledge of the software used is crucial for any corpus investigation – more than that, it *dictates* the potential analysis:

[I]t is important to recognize that corpora are simply linguistic data and that specialized software tools are required to view and analyse them. The **functionality offered by the software tools largely dictates** what corpus linguistic research methods are available to a researcher. (Anthony, 2012: 141, emphasis added)

With this in mind, Trameur is particularly suitable for the natural language processing goal at hand and presents several advantages for the present venture into relatively uncharted territory of absence phenomenon, amongst which are particularly important extended freedoms for the user.

First, the software is extremely flexible in terms of **annotation**. It handles *multiple layers* of annotation. This includes layers for form, lemma, morphosyntactic category, and any additional layers added by the user. The annotation is possible both on the *words* and, using

the new ‘Annotation de Section’ feature discussed above, annotation is now also possible on the delimited *segments*.

Furthermore, both automatic and manual annotation are facilitated. The software is equipped with a built-in TreeTagger which allows for morphosyntactic tagging to be done automatically. A graphical-user-interface permits the automatic annotation to be searched and edited when necessary, and thus, makes it possible for the user to maintain control over the annotation process at all times. In fact, the user is able to perform ‘textometric browsing’, a concept which Maria Zimina (2004; 2005) refers to as:

maintain[ing] control over the entire corpus exploration, from initial segmentation to the extraction and editing of text resources. (Zimina, 2005: 9)

The software greatly simplifies the customizing of annotation which is particularly important from the present perspective of enabling verbless sentence retrieval. It allows us to *search*, *browse (in context)* and *edit* the annotation on any level and adjust both individual items and a series of items. Finally, Trameur makes *statistical* analysis possible on *all* of the annotated layers.

Another major advantage of Trameur is the ‘**Bi-text**’ alignment feature. This feature makes it possible to explore multiple translations in multiple languages from the perspective of any of the annotation layers. Using this feature the verbless sentence alignment correspondences can be found across multiple texts. They can be visualized, annotated, and analyzed within the context of the overall corpus.

Compared to other software that processes annotated corpora, Trameur stands out in terms of its capacities for exploring the overall structure of the corpus:

Several other systems have been already developed for processing annotated corpora, for example: PDT2.0, GATE, ANNIS, Macaon. However, the novelty of Trameur consists in expanding a multi-layered data model to all stages of corpora exploration, including text mapping features and statistical analysis of dependency relations within a single graphical user interface. (Fleury & Zimina, 2014: 61)

Among the wide range of functionalities offered by Trameur, multilayer annotation control, bi-text alignment, projection of morphosyntactic categories onto delimited segments, the annotation of sections, as well as statistical analysis of verbless sentences in terms of characteristic elements and repeated segments, all with full access to the context, are particularly key for the present study.

Appendix 2 provides several illustrations of the key Trameur features from the perspective of the present verbless sentence analysis.

All of these functionalities are possible as a result of a uniquely powerful Thread/Frame data model that is applied by Trameur to the data. It is described as follows:

Multilingual textual space can be explored using a textometric data model (Thread/Frame). A *Thread* is a textual flow represented as a system of items with position identifiers. A *Frame* is used to locate different textual objects (containers and contents) and their contexts. Following these principles all text parts and annotations (including alignments) are stored and exchanged through different computerised procedures. (Zimina & Fleury, 2015: 325)

In essence, the text is treated as a single running thread of data in which every word and inter-word character receives a unique number that identifies its position. The positions are framed in a way that makes them fully explorable. The framing is guided by the user in the structure of the raw text (through the insertion of delimiters, e.g. [\$], and parts, e.g. <volet="BK_R0">). When the structured raw text is processed by Trameur, an XML base is created from the raw text in which all of the positions become fully searchable, modifiable, and open to statistical analysis.

For the present investigation, Trameur presents uniquely powerful features that prove to be extremely valuable for the linguistic study of absence, and, furthermore, its translation correspondences. Thanks to the unique data structure model and the powerful functions offered by Trameur, it was possible to presently develop the customized annotation required for the identification of absence, retrieve the structures and their multiple translation correspondences, and apply the corpus tools to the study of open structures centered on the absence.

4.4.2 Base

It was thus necessary to unite the individual texts (which at this point were appropriately segmented) into a single structure in a particular way that would allow us to (a) keep the texts separated, and (b) align the corresponding texts together. This structure is presented in *Figure 11*.

CORRESPONDENCE 1		CORRESPONDENCE 2		CORRESPONDENCE 3		CORRESPONDENCE 4	
Order	Book	Position	Order	Book	Position	Order	Book
0	<partie="1">		14	<partie="2">		28	<partie="3">
1	<volet="BK_R0"> 3		15	<volet="BK_E1">		29	<volet="BK_E2"> 3037742
2	<volet="TD_R0">		16	<volet="TD_E1">		30	<volet="TD_E2Q">
3	<volet="OD_R0">		17	<volet="OD_E1">		31	<volet="OD_E2">
4	<volet="PO_R0">		18	<volet="PO_E1">		32	<volet="PO_E2Q">
5	<volet="KF_R0">		19	<volet="KF_E1">		33	<volet="KF_E2Q">
6	<volet="OS_R0">		20	<volet="OS_E1">		34	<volet="OS_E2Q">
7	<volet="ST_R1F">		21	<volet="ST_R2F">2381717		35	<volet="ST_E1F">
8	<volet="SR_E0"> 769078		22	<volet="SR_R1">		36	<volet="SR_R2Q"> 3465388
9	<volet="AF_E0">		23	<volet="AF_R1">		37	<volet="AF_R2">
10	<volet="FH_E0">		24	<volet="FH_R1">		38	<volet="FH_R2Q">
11	<volet="JS_E0">		25	<volet="JS_R1">		39	<volet="JS_R2Q">
12	<volet="HP_E0">		26	<volet="HP_R1">		40	<volet="HP_R2Q">
13	<volet="CA_E0">		27	<volet="CA_R1">		41	<volet="CA_R2Q">
						42	<partie="4">
						43	<volet="BK_E3Q"> 3731102
						44	<volet="TD_E3Q">
						45	<volet="OD_E3Q">
						46	<volet="PO_E3Q">
						47	<volet="KF_E3Q">
						48	<volet="OS_E3Q">
						49	<volet="ST_E2F">
						50	<volet="SR_R3Q"> 3849488
						51	<volet="AF_R3">
						52	<volet="FH_R3Q">
						53	<volet="JS_R3Q">
						54	<volet="HP_R3Q">
						55	<volet="CA_R3Q">

Figure 11. Aligned corpus structure

As shown in the figure, the corpus base is structured into 4 major parts with 13 subparts in each part. This is explained by the fact that there are from 2 to 4 translation correspondences for each of the 13 novels. For example, the Russian novel *Brothers Karamazov* (BK_R0) was aligned with 2 English translations (BK_E1 and BK_E2), whereas the English novel *Animal Farm* (AF_E0) was aligned with 3 Russian translations (AF_R1, AF_R2 and AF_R3).

Notably, in order to combine the alignments within a single corpus structure, a number of phantom texts (labeled ‘Q’) needed to be introduced. For example, the phantom text BK_E3Q was added to part 4 in order to complete the alignment structure of *Brothers Karamazov* within the overall corpus.

These phantom Q-texts were filled with the same number of sentence segments as was necessary for the particular alignment. The phantom sentence segments themselves consisted of the character [Q]. This character was then neutralized in the same way as the [Z], discussed above, so as not to affect the automatic retrieval of verbless sentences, nor the statistical analysis, while at the same time permitting full alignment of the corpus.

The figure also indicates the Order of the texts in the overall text structure. The latter complete TXT file is then processed in Trameur and an exportable corpus base is created in XML format.

4.5 TAGGING

Next, automatic morphosyntactic annotation was carried out by means of the TreeTagger (Schmid, 1994). The latter is built into Trameur. *Appendix 3* presents the morphosyntactic specifications for English (PENN tagset) and for Russian (MULTITEXT-East tagset; Sharoff et al., 2008). The automatic tagging was followed by significant post-editing.

4.5.1 A Sensitive Issue for Verbless Sentences

As revealed through the present accuracy assessments, morphosyntactic tagging is one of the key problems for verbless sentence retrieval. Highly accurate morphosyntactic tagging is required particularly with regard to what is and is not tagged as a verb. As mentioned, English clitics (and particularly the clitic [’s]) and other tagging issues in both languages (e.g. capitalized verbs, homonyms, abbreviations) pose significant accuracy problems for the retrieval of verbless sentences.

Although for the study of another phenomenon the accuracy of the tagger may be more than sufficient, i.e. the *inevitable* presence of minor errors is typically offset by simply increasing the size of the corpus, the present results show that the phenomenon of the verbless sentence is particularly affected by tagging errors. Errors that are *minor* in the overall evaluation of the tagger, turn out to constitute a *major* source of errors for verbless sentences.

Without custom tagging, the accuracy of verbless sentences would hardly pass beyond the Stage 1 readings in *Table 1* above (i.e. which showed very poor Precision of 0.43 and Recall of 0.35 for verbless sentence retrieval in English, and better, though still leaving room for improvement, 0.85 Precision and 0.81 Recall scores for Russian).

As an aside, such discrepancies emphasize the benefits of *phenomenon specific* accuracy evaluation of natural language processing tools. A general 95% accuracy for a tagger does not reveal the source of the minor 5% errors, which may turn out to be systematic with regard to a specific phenomenon. As a result, a false sense of security may be present when using a particular tool to study a specific linguistic phenomenon. Furthermore, as shown in the present accuracy assessments, verbless sentences may be useful for improving morphosyntactic

tagging of verbs since they reveal particularly verb related errors. This suggests that the study of absence phenomenon may reveal particular processing issues that are hidden from a ‘presence-only’ perspective.

4.5.2 Tagging of Russian and English

The **automatic morphosyntactic annotation** of the present corpus was performed in two steps. *Figure 11* above includes a column for ‘Position’ which indicates a number corresponding to the particular position in the text when the language of the base file changed from Russian to English, or vice versa. As the base contains two languages, the morphosyntactic annotation had to be performed separately on the Russian and the English texts. Once this was done, the annotated texts were then combined, with the help of the position numbers, into a single morpho-syntactically annotated and fully aligned corpus base.

4.5.3 Post-Editing

The next third step in the tagging process was **semi-automatic post-editing** of the automatic morpho-syntactic annotation. This concerned in particular:

- **a. Customized annotation:** the custom forms that were created in order to resolve the errors discussed above (e.g. special new forms for clitics, etc.) and which were entirely external to the automatic tagger
- **b. Errors:** the specific verbless sentence related errors as revealed through accuracy assessments

After the automatic tagging, it was necessary to resolve (a) customized annotation, i.e. the clitics and the new forms that were created in segmentation to resolve specific errors but were obviously not recognized by the tagger.

Following the resolution of (a), a first retrieval of verbless sentences took place. These verbless sentences were exported and manually reviewed for **false positives** (i.e. the presence of a verb). Errors were recorded in a spreadsheet, and subsequently corrected.

The post-editing corrections were done with the help of Trameur’s user-controlled multi-layer annotation search and edit feature. This feature made it possible to *semi-automate* the post-editing of the tagging. More specifically, using this feature the errors may be searched for, viewed in the context of concordance lines and edited – all of these may be performed on *any of the layers* (including form, lemma, grammatical category and other necessary layers may be added). This opens up wide possibilities to group the errors in various ways and then correct them simultaneously. Semi-automating the post-editing in this way helped to address the manual costs associated with this step.

4.6 CLASSIFICATION

Following custom-segmentation, alignment-structure, and morpho-syntactic annotation, all of the sentences of the corpus were then classified into verbal and verbless.

This task is performed using Trameur’s **‘Projection’** feature. This is a feature that projects the morphosyntactic category of the verb onto the custom sentence segments (which in Trameur constitute a ‘Section Map’). The projected categories include all verb forms in English and in Russian, English modal verbs, as well as the phantom [Z] character. A precise list of the morphosyntactic tags excluded from verbless sentences is found in *Appendix 4*.

Crucially, the context is preserved in the projection process. Due to the Thread/Frame model, the sentences that contain a verb and those that do not were identified without losing their position in the overall corpus structure.

This structure of verbal and verbless segments, in their original order, can then be further explored using a feature called the **‘Section Map’**. This map presents all of the sentences of the corpus graphically in the form of containers in their original sequential order. The sentences that contain a verb are represented graphically as containers with an [x] and the non-[x] containers correspond to the verbless sentences. An illustration is provided in *Figure 12*.

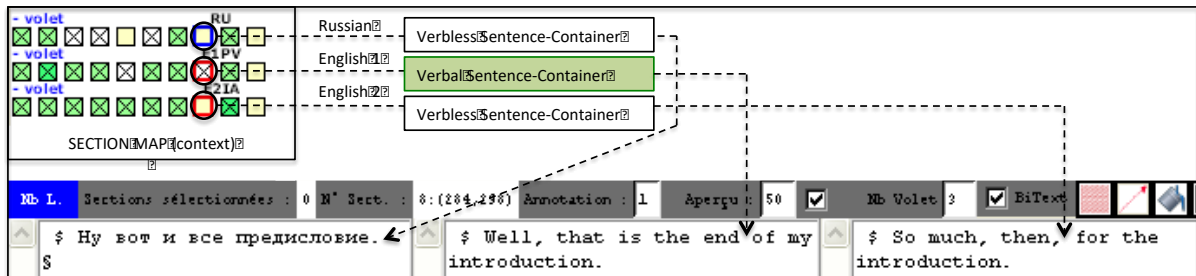


Figure 12. A verbless sentence aligned with multiple translation correspondences in original context in Trameur

The figure shows a mini-simulation of three aligned texts that contain 10 sentences each, as evidenced in the section map in the top-left corner.

The verbless sentences (i.e. the containers without an [x]) are viewable in their original context. This context can be browsed by selecting the desired sentence-container on the section map, and the alignment is preserved and viewable throughout the browsing.

As the verbless sentences are in context, they can be submitted to various statistical analysis against the reference of the rest of the corpus. They can also be exported if necessary.

As also shown in the figure, the verbless sentences are viewed simultaneously with their multiple translations. It is thus not only the verbless sentences, but also their multiple-translation correspondences that are retrieved automatically. The context of the correspondences can also be viewed by browsing through the section map. It is also possible to export the alignments that contain a verbless sentence in any of the translation correspondences.

As a result, verbless sentences, their translations, and the context of both – all of the elements necessary for contrastive analysis – are retrieved, and may be both qualitatively and statistically analyzed.

4.6 ADDITIONAL NEW PROGRAMS

Due to the large number of sentence segments in the present corpus (total 269,456 sentence-segments), it was necessary to develop several scripts in order to carry out the required functions without the graphical user interface provided by the Section Map in

Trameur. These scripts were developed with the help of a team (Borissov et al., 2020). The exported structured-XML-base (i.e. Trameur's unique Thread/Frame model) provided the key foundation to explore the corpus structure.

The Python scripts that were developed include:

- *Characteristic elements computation* for verbless sentences (against the reference corpus) and also for their context
- *Characteristic n-grams computation* for verbless sentences (against the reference corpus)
- *Type-token counter for verbless sentences* and for the corpus (text-specific)
- *Verbal translation correspondence list and counter*

In addition, a *Contrastive analysis graphical-user-interface* was also developed on the basis of the corpus in order to allow the search for the verbless sentence segments and their translation correspondences and to visualize them *one aligned-sentence-segment at a time* for qualitative analysis. Several features were developed that are particularly useful for qualitative contrastive analysis on the sentence level. The new program is illustrated in *Appendix 5*. A contrastive-annotation function and debugging is in progress.

4.7 SUMMARY

The present chapter summarized the results of several pilot studies with regard to automatic retrieval of verbless sentences.

- The **first pilot** started with manual extraction from Dostoyevsky's *Brothers Karamazov* and the Pevear and Volokhonsky translation of the discourse-based novel.
- In the **second pilot**, a competing translation by Avsey was added. This introduced the multiple translation perspective into the methodology.
- The **third pilot** introduced an English original of a different genre – Pinter's play *The Caretaker*, as well as reciprocal parallel-text analysis.

- The **fourth pilot** added two Russian and two English translations of Camus' French original *L'Étranger*, which introduced the genre of narrative-based fiction and the third-language perspective.

The method of verbless sentence retrieval evolved with each pilot. Further issues revealed themselves in the construction of the full 1,4-million-word 32-text parallel and comparable corpus. (As mentioned the latter underwent the latest method from scratch so that all the texts are standardized.)

It is hoped that the methodological results of the present experiences can be valuable in terms of the automatic processing of verbless sentences and lead to further automatization of their retrieval. At present, it was possible to propose a replicable method of automatic verbless sentence retrieval, as well as make the following conclusions concerning their processing.

4.7.1 Three Keys Issues of Verbless Sentence Retrieval

Instead of parsed corpora, which would be typically recommended for grammatical phenomenon, we propose that for the study of verbless sentences it is more appropriate to use a *specialty-segmented and morpho-syntactically-tagged* corpus. This is due to the present results which show that automatically retrieving verbless sentences requires:

1. Customized segmentation

Very particular segmentation is required to delimit sentences and to do so in such a way that direct speech may be disentangled from narration.

2. Highly accurate morphosyntactic tagging

Particularly this concerns what is and what is not tagged as a verb. Clitics and other tagging issues discussed pose specific important problems for verbless sentence retrieval.

3. Classifying custom-annotated and custom-segmented data

Finally, it requires a software that is capable of processing custom-segmented and custom-annotated data and of classifying the sentences into those with the feature of the verb and those without.

At present it was possible to automatize the segmentation process, semi-automatize the post-editing, and automatize the classification. The proposed method makes it possible to avoid the barriers that pose problems for retrieval in most parsed corpora that were discussed above. It is also possible to get a handle on the accuracy, which through the present pilots was revealed to be a serious issue.

The present experience also showed that targeting verbless sentences in automatic processing reveals hidden issues in the existing segmentation, tagging and alignment tools.

Further work remains in automating the process in appropriate ways. It is also of note that the present method will not retrieve *utterances* (i.e. verbless structures that are part of verbal sentences), but focuses on *sentences* whose theoretical applicability to verbless structures was defended in *Part 1*.

CHAPTER 5

QUANTITATIVE AND QUALITATIVE ANALYSIS

Following automatic retrieval, quantitative and qualitative analysis was performed on the verbless sentences, their translation correspondences, as well as their context. Depending on the type of analysis, different parts of the corpus were concerned. The present section summarizes the particular types of analysis and the additional manual qualitative annotation involved.

The analysis may be split up into the following three major perspectives from which the corpus was studied:

1. Monolingual
2. Parallel
3. Third-language translation

5.1 MONOLINGUAL PERSPECTIVE

The monolingual perspective studies the features a single language. This perspective focuses on Russian and English separately from one another. The corpus of Russian texts that were examined from a monolingual perspective involves: Russian originals and Russian translations. Similarly, the corpus of English texts examined from a monolingual perspective

involves texts where English is the source language, as well as texts where English is the target language.

Notably, the present monolingual analysis treats both originals and translations as genuine samples of the examined language. As discussed in *Part 2: Section 3.2* above, concerning the use of translations for monolingual analysis, the present study sides with, for instance, Biber (1993), Baker (1996), Olohan (2002), Zanettin (2014), in arguing that translations constitute an important part of the production and reception of a language. Therefore, the inclusion of translations as part of monolingual analysis does not compromise it, but on the contrary strives to make for a more ‘representative’ sample of the language. Whether one looks to Biber (1993), for whom representativeness in corpus-design consists in targeting linguistic variety, or to Leech (2007), who treats representativeness in terms of matching the proportional usage of a language – on either conception of representativeness, the inclusion of translations in monolingual analysis would seem to be justified. From the former perspective, translations provide another variety of a language, and from the latter, translations, especially those with high exposure, would reflect a certain proportion of language use.

The directions of comparison in monolingual analysis first include comparing (i) Russian originals to Russian translations, (ii) English originals to English translations. Once the language-specific features of Russian and English are thus revealed, the two languages are contrasted with one another in terms of their language-specific features. The analysis is called ‘monolingual’ despite the fact that two languages are contrasted. This terminology emphasizes the fact that the link between the source and the target is not exploited in monolingual analysis; the two languages are considered independently of one another.

From the monolingual perspective, the analysis carried out on the full corpus includes the following.

— *A. Verbless Sentence Frequency*

The frequency of the verbless sentences was normalized and compared across originals, translations and the three genres.

— *B. Key Lexical and Grammatical Elements*

The lexical and grammatical differences between verbless and verbal sentences were explored. The key words (forms, lemmas, morphosyntactic categories) and key n-grams (repeated segments) that characterize verbless sentences were statistically determined. They were calculated against those of the reference corpus (i.e. the entire corpus which includes all of the verbal and verbless sentences). The hypergeometric model used for the calculation is described in (Lebart et al., 1998: 130–136; Lebart & Salem, 1994).

The key words were also calculated for the immediate left and right context of the verbless sentences, against the reference corpus.

— *C. Length and Complexity*

The length of the verbless sentences in terms of types and tokens was determined. A basic assessment of their lexical complexity was calculated.

— *D. Semantico-Pragmatic Lexical Analysis*

A semantico-pragmatic analysis of lexical contents of verbless sentences was carried out by means of a manual classification of their statistically key words and segments. A discussion of the classification scheme is provided in the results.

Due to the fact that the following analysis required a lot of manual annotation, it was performed on much smaller subparts of the corpus.

— *E. Ellipsis*

Classification of verbless sentences into elliptical and non-elliptical was manually performed on a subpart of the corpus.

This involved the manual segmentation of the verbless sentences into utterances (as explained in *Part I*). Subsequently, the utterances were manually annotated in terms of the presence or the absence of an antecedent-based ellipsis

(stripping, gapping, sluicing). The definitions of the ellipsis types were discussed in *Part 1* and correspond to those of McShane (2000).

— *F. Discourse Type*

Verbless sentences were manually classified as belonging to either direct speech or narration in a subpart of the corpus.

— *G. Information Structure and Syntax*

Manual annotation of the Topic, Focus, and Focus-structure Type, as well as the Subject and Predicate, of the verbless sentences was performed on a subpart of the corpus. The definitions used were discussed in detail in *Part 1*. In particular, subject and predicate were defined in accordance with their traditional syntactic definitions (e.g. Touratier, 2009). The information structure features were defined in accordance with Lambrecht (1994), also described in *Part 1*.

— *H. Direct and Indirect Speech Acts in Questions*

Verbless non-elliptical questions were manually annotated in terms of direct and indirect speech act in a subpart of the corpus in accordance with the definitions in (Celle, 2018; Celle et al., 2019; Bondarenko & Celle, 2019). A description of these categories is provided in the results.

Notably, this analysis also required prior manual annotation of the verbless sentences in terms of sentence type (defined in accordance with Huddleston, 1994, and described in *Part 2: Section 3.1.2*), as well as manual annotation of the identified questions in terms of elliptical and non-elliptical (in accordance with the definitions of antecedent-based ellipsis which are described in *Part 1* and correspond to those of McShane, 2000).

5.2 PARALLEL PERSPECTIVE

Secondly, combining contrastive analysis principles with recent developments in parallel corpus methodology described above, the present corpus is studied from a parallel perspective. This perspective differs from the monolingual, in that that it requires a direct link between the original text and the translation correspondence.

In this perspective, patterns between verbless sentences and their translation correspondences are investigated. In particular, the present analysis looks for correspondence patterns that systematically re-occur across multiple translations of the same original, across several works by different authors, across three different genres, and from multiple directions of comparison (as described in *Part 2: Section 3.3* above). Notably, this includes correspondence patterns not only from source to target, but also from target to source. Particular attention is given to verbal correspondences of verbless sentences.

From the parallel perspective, the analysis carried out on the full corpus includes the following.

— *A. Verbal Correspondence Frequency*

The frequency of the verbal translation correspondences of verbless sentences is compared across originals, translations, and the three genres.

— *B. Verbs in Correspondence*

The verbs that are involved in the verbal translation correspondences of verbless sentences are analyzed in terms of their form, lemma and morphosyntactic category. For this macro-analysis each occurrence of the verb is treated independently, e.g. the verb compound ‘*has been going*’ is treated as three separate verb tokens.

The following analysis was performed on a much smaller part of the corpus due to the fact that it required a lot of manual annotation.

— *C. Ellipsis in Correspondences*

Translation correspondences of verbless sentences were analyzed for the presence or absence of ellipsis. This analysis was performed on a subpart of the corpus and involved both verbal and verbless correspondences of verbless sentences.

The correspondences were manually segmented into utterances and manually annotated for antecedent-based ellipsis (stripping, gapping, sluicing). The definitions of the utterance segmentation and ellipsis types were discussed in *Part 1* and correspond to McShane (2000).

— *D. Verb Compounds and Verb Tense in Correspondence*

The lexical meaning and the tense of the verbs that are involved in verbal translation correspondences were annotated with regard to a subpart of the corpus. The distinction of compound verbs and the annotation for tense was carried out manually.

This analysis involved prior manual annotation of the correspondences of the subpart in terms of the presence or absence of a verb.

— *E. Aspectual Meaning of Verbal Correspondences*

Verbal translation correspondences of non-elliptical verbless sentences were manually annotated for situation and viewpoint aspect in a subpart of the corpus. The definitions followed were those of Carlota Smith (1997) and are described in the results, as well as in (Bondarenko & Celle, 2020).

This analysis involved prior manual annotation of the verbless sentences of the subpart in terms of elliptical and non-elliptical (in accordance with the definitions in *Part 1* and McShane, 2000).

— *F. Information Structure and Syntax of Correspondences*

Verbal and verbless translation correspondences of verbless sentences were annotated for Topic, Focus, Focus-structure Type, as well as Subject and Predicate, on a subpart of the corpus. Traditional syntactic definitions (e.g. Touratier, 2009) were applied to identify the subject and predicate. Information structure categories were defined in accordance with Lambrecht (1994). The details of the definitions were discussed in *Part 1*. Particular attention was given to the informational structure of the verbal correspondences of verbless sentences, whose verb phrases were manually annotated in terms of whether or not they belong to the informational-structural Focus. Changes between the verbless sentences and their correspondences in terms of Subject, Topic, Focus and Focus-structure Type were also manually annotated.

This analysis involved prior manual annotation of the correspondences of the subpart in terms of the presence or absence of a verb.

— *G. Direct and Indirect Speech Acts of the Verbal Correspondences in Questions*

The verbal correspondences of non-elliptical verbless questions were manually annotated in terms of direct and indirect speech act in a subpart of the corpus. The definitions used correspond to those in (Celle, 2018; Celle et al., 2019; Bondarenko & Celle, 2019), and are described in the results.

This analysis involved prior manual annotation of the correspondences of non-elliptical verbless questions in terms of the presence of the verb. It also required prior manual annotation of the verbless sentences in the subpart in terms of sentence type (Huddleston, 1994; as described in *Part 2: Section 3.1.2*) and the manual classification of the identified verbless questions in terms of elliptical and non-elliptical (McShane, 2000; as described in *Part 1*).

5.3 THIRD-LANGUAGE PERSPECTIVE

The corpus is also examined from the perspective of a third language, that is neither Russian nor English. It involves the study of Russian and English translations from French, and concerns both verbless sentences and the translation correspondences of verbless sentences in these two languages.

The uniqueness of this perspective consists in the fact that (a) it provides another type of data to compare with the originals and the bidirectional translations, and (b) it constitutes a control measure for Russian and English interference on one another.

The particular types of analysis which were performed on third-language translations overlap with the analysis in the monolingual and parallel perspectives described above. More specifically, Russian and English translations from French were subjected to the following types of analysis:

- *Monolingual A: Verbless Sentence Frequency*
- *Monolingual B: Key Lexical and Grammatical Elements*
- *Monolingual C: Length and Complexity*
- *Monolingual D: Semantico-Pragmatic Lexical Analysis*
- *Monolingual H: Direct and Indirect Speech Acts in Questions*
- *Parallel A: Verbal Correspondence Frequency*
- *Parallel B: Verbs in Correspondence*
- *Parallel E: Aspectual Meaning of Verbal Correspondences*
- *Parallel G: Direct and Indirect Speech Acts of Verbal Correspondences in Questions*

Comparison with third-language translations helps to disentangle features and patterns that indicate language specific constraints from those that result from the influence of potential Russian or English source-language influence on the translation.

CHAPTER 6

CONCLUSION: METHODOLOGY

This completes the overview of the interdisciplinary methodology involved in the present study. The key aims of this section were to:

- combine several disciplines in order to push the boundaries of the existing perspectives on verbless sentences
- specify the issues that challenge corpus treatment of verbless sentences
- spotlight the relevant typological differences for the comparison of verbless sentences in Russian and English
- highlight the steps of a contrastive study of verbless sentences from a corpus perspective
- propose a brand-new method of automatic verbless sentence retrieval that overcomes existing processing problems and is reproducible

The next part will present the brand new parallel and comparable corpus created.

PART THREE:

CORPUS

CHAPTER 1

DESIGN AND SELECTION CRITERIA

The notion of representativeness in corpus design is full of intricacies. In a famous paper, Geoffrey Leech (2007) draws attention to the various definitions that the notion can take. The essence of the issue comes down to the following:

There is a crucial difference between claiming that such-and-such is the case in a corpus, and that the same such-and-such is the case in a language. By definition a sample is representative if what we find for the sample also holds for the general population (Manning & Schütze, 1999: 119). Putting this in operational terms ‘representative’ means that the study of a corpus (or combination of corpora) can stand proxy for the study of some entire language or variety of language. (Leech, 2007: 3)

For some, the question of the appropriate design of a language sample, i.e. a corpus, is closed before it is even asked. Noam Chomsky (1986) draws attention to the fact that the key questions considered by linguistics concern language competence, i.e. ‘internal language’, whereas corpus data are instances of language production, i.e. ‘external language’. He therefore insists that studying the latter to get at the former is futile; furthermore, he points out that the textual universe is immense and can never be inventoried.

To defend the legitimacy of the domain against such criticism, corpus linguists convincingly make an analogy to the domain of physics which also studies intangible rules and whose object of study is no less immense in proportions:

It is true that the textual universes associated with a modern language with a large number of native speakers, such as English, can be immense; but no more

bafflingly immense than the universe of the material cosmos, about which physicists construct intelligible theories. (Leech, 2007: 4)

Simultaneously, the amount of work in the corpus linguistic domain, particularly in terms of the integration of statistical methods, is a sign that Chomsky's criticism has been heard, synthesized, and, in a Hegelian dialectical fashion, has resolved in higher standards within the domain. As such, the question concerning the appropriate design of a language sample is a central issue for any corpus study.

Within the design criteria, representativeness is a particularly thorny issue. The basic characterization of a representative language sample (i.e. representative of the textual universe of that particular language) is that the sample should be proportional to the demographics of the population (e.g. in terms of age, profession, etc.) who produce the language (Leech, 2007: 4, 6). However, serious arguments exist against the application of such a standard in corpus linguistics. In a seminal paper on representativeness, Douglas Biber (1993) argues against such a conception of proportional representativeness for the corpus linguistic domain on the basis that it will entirely thwart the variety of language that is studied (i.e. he expects that most of the language that humans produce occurs in basic conversations and these latter are typically not as rich as other types of language). For instance, a proportional demographically oriented corpus would be ill-suited for the study of rare phenomena (Biber, 1993: 245, 256). Biber spotlights the existence of a trade-off between representativeness based on proportionality and that based on variety, and argues for the latter.

For Váradi (2001), such a trade-off is unacceptable; he takes it to signal that corpus compilation, and thus the data, is subjective to the researcher's choice and therefore the results of a corpus study cannot be extended beyond that particular corpus. In response, Leech (2007) proposes to give attention to reception, as opposed to the traditional focus on exclusively production, in corpus design in order to objectivize the process of selection:

[T]he criterion of readership/audience is free of evaluative bias. One of the results, no doubt unpalatable to corpus-builders with a sense of taste, is that tabloid newspapers are more likely to be included in a representative corpus than so-called quality or broadsheet newspapers. But this is something that one has to put up with in the interests of representativeness. (Leech, 2007: 7)

In other words, he proposes to include the number of hearers or readers of a text as measurable selection criteria, and not only the speakers and writers that produce them.

A recent discussion of the developments in the notion of representativeness can be found in a special issue of *Cogni Textes 19: Corpora and Representativeness* (ed. Raineri & Debras,

2019). Ranger (2019) raises the notion of phenomenon-based representativeness and shows it to be particularly relevant for the corpus design of enunciative analysis. The use of parallel texts also carries its own requirements on design, which are given particular attention in a special issue of *STUF 60(2): Language Typology and Universals* (ed. Stolz, 2007b).

Another frequently emphasized aspect is the need to proceed in a cyclical fashion in the building of a corpus, i.e. through pilot studies that motivate further design criteria.

[I]n many aspects of corpus design, work must proceed in a circular fashion, with empirical investigations based on pilot corpora informing the design process. [...] The bottom line in corpus design, however, is that the parameters of a fully representative corpus cannot be determined at the outset. Rather, corpus work proceeds in a cyclical fashion [...] The initial design of a corpus will be more or less advanced depending on the availability of previous research and corpora. (Biber, 1993: 253, 256)

The following section describes the way these concerns have been integrated into the design of the present corpus and will describe its target population. Notably, the objective of the present study being a semantico-pragmatic analysis of verbless sentences from a contrastive perspective sets particular parallel-text and phenomenon-specific requirements.

1.1 THE TARGET POPULATION

The ultimate goal of the present corpus is to represent a wide variety and high frequency of the phenomenon of the verbless sentence. In other words, in a similar vein as Biber (1993), the present target is heterogeneity, as opposed to demographic proportional representativeness. However, the present target is further reduced to a phenomenon-focus. That is to say, the aim is for “an optimally heterogeneous corpus” (Leech, 2007: 9) but specifically for the phenomenon of the verbless sentence.

Therefore, representativeness in terms of general proportional usage was presently traded in favour of the variety of the verbless phenomenon. However, the target is even more specific due to the contrastive aspect of the study. The contrastive translation-analysis goal leads to the more precise design aim of a corpus that represents a wide variety and high frequency of verbless sentences that have been translated.

1.2 THE KINDS OF TEXTS AND THEIR SELECTION

As the available research on verbless sentences in corpora is very limited, it is not initially obvious what kinds of texts and language will best target the phenomenon. Statistics with regard to the sample size that is required in order to observe the phenomenon of the verbless sentence have also not been found to exist.

Consequently, the present design proceeded in stages and was informed by the results of our pilot studies, which motivated the augmentations. This concerned both the issue of further defining the population to be targeted by the texts included in the corpus, as well as determining its required size.

Presented in *Table 2* are the criteria based on which the texts of the present corpus were selected. Potential texts were evaluated as shown. The motivation for the criteria is explained below.

Criteria	Example		
	BK_R0 Original	BK_E1 Translation 1	BK_E2 Translation 2
Basic Requirements			
Realist fiction: realistic prose	+++		
Direct speech dialogue: expected high frequency	+++		
Register: everyday, colloquial	+++	+	+
Time period: late 19 th -21 st century, favour recent	+	+++	+++
Exposure: expected high readership/receivers	+++	++	++
N of existing translations: favour multiple		+++	+++
Identity: translated from original not another translation		+++	+++
Electronic format: txt	+++	+++	+++
Further Reciprocity & Comparability Considerations			
N of Russian & English originals in corpus			
N of Russian & English translations in corpus			
Size relative to corpus			
Genre relative to corpus			
Time period relative to corpus			

Table 2. Present Parallel-Text Selection Criteria

The focus is on **realist fiction** of an **everyday language register** and with anticipated frequent **direct speech dialogue**. In general, such a focus corresponds to the central criteria that have been identified for parallel-text studies (e.g. Stolz, 2007a: 103; the latter provides a particularly thorough discussion of specific parallel-text selection criteria).

The focus on **discourse-based fiction** is further justified by our pilot results which showed that verbless sentences are particularly closely correlated with direct speech. A manual annotation of the verbless utterances in the Russian *Brothers Karamazov* (BK_R0) and its two English translations (BK_E1, BK_E2) showed that 93% of the verbless utterances occurred in direct speech, compared to only 7% in narration (the details of this finding will be discussed in the results section). This reveals that the quantity of speech seems to be an important sampling parameter for the study of verbless sentences. Representing all types of language would thus not be a fruitful corpus-design for a linguistic study of this phenomenon with the present goals: to answer questions about semantico-pragmatic factors, retrieving a high frequency and wide variety is important. A decision was therefore made to target predominately dialogue-based texts so as to examine a wider range of the verbless phenomenon.

A proportional subpart of **narrative-based fiction** was also included: it corresponds to 8.9% of the present corpus of English and 8.8% of Russian (in terms of total word-tokens for each language). This sub-corpus helps to disentangle patterns that may be due to genre-related or other interference (e.g. see Zanettin, 2014, on genre in parallel-texts⁴³). Of note is that the narrative-based fiction genre sample overlaps with the third-language translation sample in the present corpus. Therefore, this sub-part presently serves control purposes.

Two theater plays, and their translations, represent the genre of **scripted-speech fiction** in the present corpus. This sub-corpus helps to further examine the potential impact of genre on the verbless sentence phenomenon. Therefore, it also presently serves a control purpose. Translated theater plays are particularly difficult to find in electronic format. As a result, the size of this sub-corpus is at present limited. One source Russian play and one source English play, and their respective translations, correspond to 4.6% of the Russian corpus and 5.2% of the English corpus (in terms of total word tokens per language).

⁴³ Zanettin draws attention to the particular control benefits for parallel-text studies of including heterogeneous genre, time period, and translation types: "Corpora containing texts translated from different languages may allow for controlling interference from specific source language systems, while subcorpora containing texts belonging to different text types may allow one to distinguish between translation induced and genre-related variation. Subcorpora containing texts collected at different times may instead yield insights into how evolving translation styles and norms relate to evolving language norms." (Zanettin, 2014: 184)

Notably, the **stage directions** have been removed from the theater plays. This is presently justified by the focus on realistic prose and everyday register. Stage directions, such as ‘*Pause*’, ‘*Silence*’ and indications of who is to be speaking, are instances of domain-specific language that is particular to theater-play technical notes. They are not pronounced by the performers, nor are they ever heard by the receivers in the explicit lexical form in which they are indicated (i.e. the spectators observe a physical pause but they do not hear the lexical form ‘pause’). Furthermore, they are so frequent that they would in any case need to be controlled in any linguistic analysis of the verbless sentence results. Therefore, the decision was made to remove all stage directions from the corpus and focus only on the scripted speech for the present study.

In a similar vein, book and chapter **titles** have been removed for all the texts in the corpus. The divisions between all the texts are kept in the corpus and the results (via the Frame structure in Trameur), however the text of the titles was removed. This decision is again justified by the present focus on realistic prose and everyday register. Titles have sometimes been argued to abide by special restrictions; the translation of titles requires special analysis. If included, the presence of titles would need to be controlled in the analysis of the results. Therefore, the decision was made to remove the titles. Neither the titles, nor stage directions make it into the corpus, nor into any of the quantitative descriptions here presented. The present study targets only the contents of the texts.

Notably, the three genres included in the corpus, i.e. dialogue-based fiction, narrative-based fiction and scripted-speech fiction, are distinguished based on what Biber (1989: 5) refers to as an **external format distinction**, i.e. based on the fact that they represent functionally different texts, not on any text-internal linguistic features. Dialogue-based fiction is the main targeted format-based genre and corresponds to 86.6% of the Russian corpus and 85.9% of the English corpus (in terms of the total words per language).

Among the selection criteria is that the texts should be well-known, i.e. of an anticipated high **exposure**. This criterion targets the *works and translations* that are likely to have higher reception rates, the value of which has been emphasized by Leech (2007).

Furthermore, the texts have ideally been **translated multiple times**. This way, even if all of the available translations are not explored in the present study, additional translations can be included in future developments.

Attention was also given to the **identity** of the translations, i.e. the available information concerning the translated texts was verified for the source text from which the translation was made. A preference was given to translations that have the source text as their origin as opposed to having been re-translated from another translation. The importance of this criterion is drawn

attention to by Thomas Stolz (2007a) and consists in identifying and controlling possible source language interference.

The **time period** that is targeted by the present corpus is the last century and a half. The texts range from the late 19th century to the early 21st century. The division between each individual text is maintained and makes it possible for us to examine potential differences that may be associated with the time period of production.

The availability of the text in electronic format is an obvious criterion. A raw text (txt) format was required. Furthermore, the cleanliness of the initial file also had important weight. For instance, serious encoding-related issues with the Russian play *The Love Girl and the Innocent* (OS_R0) demanded a lot of manual attention. These costly issues were resolved in order to balance the corpus in terms of the source language within the scripted speech genre.

Concerning the **size of the text samples**, a decision was made to select texts of various lengths and include them in their entirety. The only exception was for the exclusion of scripted speech and titles as mentioned above, the *Brothers Karamazov* sub-corpus (BK_R0, BK_E1, BK_E2) which was limited to the first 14 chapters of the novel, and the *Quiet Don* sub-corpus (TD_R0, TD_E1) which was limited to part one of the four-part novel. Such a decision was driven by the fact that sample size information concerning the verbless sentence phenomenon does not exist, to the best of our knowledge. It is hoped that such an exhaustive approach to the texts will make it possible to provide some key statistics that, with more work, can be used to eventually determine the required sample size for a verbless sentence study. For the present results, the size of the texts was normalized where necessary so as to allow their comparison.

The results of our pilot studies revealed that a particularly large **corpus size** is required in order to study verbless sentences from a *contrastive* translation-based perspective. Some of the most fascinating questions pertain to the analysis of the verbal correspondences of verbless sentences. However, the differences between the two languages in terms of verbless sentence frequency have the consequence that Russian verbal correspondences of English verbless sentences constitute an important deficit, which we have noted in (Bondarenko & Celle, 2019; 2020). This deficit prevents the necessary analysis of the meanings implicated in Russian verbless sentences. The present augmentations in the size of the corpus attempt to address that deficit.

Additional important criteria in the selection of the texts included in the corpus concern:

- The relative *number of original* texts in Russian and in English
- The relative *number of translations* in Russian and in English

- The relative number of *original and translated texts per genre*, in Russian and in English
- The *time period* of original and translated texts in Russian and in English
- The *size* of the originals, translations, and genre, relative to the size of the corpus, in Russian and in English (in number of words)

These considerations target the requirement of balance in terms of translation reciprocity (i.e. from Russian to English and from English to Russian) and comparability (i.e. in terms of language, source, time period, genre and size).

The next chapters will present the composition of the presently created corpus and analyze its comparability.

CHAPTER 2

COMPOSITION

2.1 GENERAL DOMAIN

The present 1.4-million-word, fully sentence-level aligned and morpho-syntactically tagged parallel and comparable corpus was built specifically for the present study. As described in the previous section, its general domain is realist fiction from the late 19th through early 21st century that has received high exposure and has been translated into the target languages.

2.2 NUMBER OF TEXTS

The corpus consists of a total of 32 texts which correspond to 13 different works of literature. It includes 12 originals (6 in Russian and 6 in English), 16 translations (8 Russian translations from English and 8 English translations from Russian), and 4 third-language translations (2 Russian translations and 2 English translations of the same French original).

Furthermore, three function-based genres are represented within the 32 texts. This includes discourse-based fiction, which corresponds to the main targeted genre and consists of 24 texts (12 in Russian and 12 in English). Scripted-speech fiction and narrative-based fiction

represent control groups that help to disentangle typological patterns from those that may be due to genre interference. The scripted-speech genre contains 4 texts (2 in Russian and 2 in English). The narrative-based fiction genre also contains 4 texts (2 in Russian and 2 in English).

2.3 NUMBER OF WORDS

The exact size of the corpus is 1,388,422 words (i.e. tokens). This corresponds to 581,547 words in Russian and 806,875 words in English.

Original texts correspond to 537,460 words: 262,659 for Russian originals and a similar 274,801 for English originals. Overall, originals represent 38.7% of the total corpus. If considered from a language-specific perspective, original texts correspond to 45.1% of the Russian sub-corpus, and 34.1% of the English sub-corpus. What may appear to be a low proportion of originals is actually due to the inclusion of translations whose size cannot be controlled without losing translation correspondences. Stopping a translation short in order to balance the number of words would be inappropriate, if only because it is precisely the translation correspondences which are necessary for the present contrastive analysis goals.

Translations correspond to 728,403 words: 267,866 for Russian translations from English, and 460,537 for English translations from Russian. In other words, they represent 52% of the total corpus, 46.1% of the Russian sub-corpus, and 57.1% of the English sub-corpus. Notably, the number of words in the Russian translations (i.e. 267,866) is balanced with the number of words in Russian originals (i.e. 262,659). Furthermore, the 267,866 words of the Russian translations corresponds to the translation of 274,801 English source words: this indicates a reduction in the number of words in translation from English to Russian. In contrast, English translations from Russian expanded in terms of the number of words, i.e. the 262,659 words of the Russian originals were translated as 460,537 words in English. This explains the proportional imbalance in terms of words of the English originals (i.e. 274,801) and the English translations (i.e. 460,537).

Third-language translations from French correspond to 122,559 words: 51,022 in Russian and 71,537 in English. They represent 8.8% of the overall corpus, 8.8 % of the Russian sub-corpus, and 8.9% of the English sub-corpus.

In terms of number of words per genre, discourse-based fiction corresponds to 1,196,623 words (503,627 in Russian and 693,005 in English). In other words, it represents 86.2% of the total corpus (86.6% of the Russian sub-corpus and 85.9% of the English sub-corpus). Scripted-speech fiction represents 69,231 words (26,898 in Russian and 42,333 in English) or 5% of the total corpus (4.6% of the Russian sub-corpus and 5.3% of the English sub-corpus). Narrative-based fiction, which as mentioned in *Part 3: Section 1.2* overlaps with the third-language translations, corresponds to 122,559 (51,022 in Russian and 71,537 in English), which represents 8.8% of the overall corpus (8.8% in of the Russian subpart and 8.9% of the English subpart).

A visual illustration of the size of the different parts is provided in *Part 3: Chapter 3* below, which analyzes the comparability of the corpus parts in terms of the various sampling parameters, including size.

2.4 DESCRIPTION OF TEXTS

The present section provides information concerning the particular texts in the corpus. This includes the following metadata and basic summary statistics:

- title, author/translator, and date of publication,
- language (i.e. Russian, English),
- language-type (i.e. original, translation, third-language translation),
- language-genre (i.e. discourse-based, scripted-speech, or narrative-based fiction),
- size of the text in terms of the number of words (i.e. tokens),
- size of the text in terms of the number of unique words (i.e. types),
- size of the text in terms of the number of sentences,
- text code in the present corpus (see *Appendix 1*),
- and other important precisions.

The following visual summary of the created corpus illustrates the texts organized in terms of language, translation correspondence, time period and genre. *Figure 13* provides a basic comparative summary, as well as the legend which corresponds to *Figure 14*. *Table 3* summarizes they key values involved in the present description.

MONOLINGUAL, PARALLEL-RECIPROCAL & LOOSELY-COMPARABLE CORPUS					
		Texts	Time	Words	Genre
English	EO: English Original				
	ET: English Translation (from Russian)	EO: 6	1926-2003, avg 1959	274,801	■ ■
	ETF: English 3 rd Language Translation (from French)	RO: 6	1880-2011, avg 1951	262,659	■ ■
Russian	RO: Russian Original				
	RT: Russian Translation (from English)	ET: 8	1934-2012, avg 1979	460,537	■ ■
	RTF: Russian 3 rd Language Translation (from French)	RT: 8	1956-2006, avg 1987	267,866	■ ■
		ETF: 2	1946-1988, avg 1967	71,537	■
		RTF: 2	1966-1968, avg 1967	51,022	■
■	Discourse-based Fiction (DF)				
■	Scripted Speech (SP)				
■	Narrative-based Fiction (NF)				
		Total texts: 32 (16 Russian; 16 English)			
		Total words: 1,388,422 (581,547 Russian; 806,875 English)			

Figure 13. Summary of Corpus and Legend to Figure 14

ENGLISH ORIGINAL + Russian Translation	RUSSIAN ORIGINAL + English Translation
Scripted Speech Play	
CA-EO <i>The Caretaker</i> , Harold Pinter 1960	OS-RO Олень и Шалашовка, Александр Солженицын 1954 [Olen' i Šalašovka, Aleksandr Solzhenitsyn]
+ CA-R1 Doroševič, Storož 2006	+ OS-E1 Bethel & Burg, <i>The Love-Girl and the Innocent</i> 1969
Discourse-based Fiction	
SR-EO <i>The Sun Also Rises</i> , Ernest Hemingway 1926	BK-RO Братья Карамазовы, Фёдор Достоевский 1880 [Brat'ja Karamazovy, Fyodor Dostoevsky]
+ SR-R1 Toper, Fiesta (I Vosxodit Solnce) 1968	+ BK-E1 Pevear & Volokhonsky, <i>The Brothers Karamazov</i> 1990
	+ BK-E2 Avsey, <i>The Karamazov Brothers</i> 1994
AF-EO <i>Animal Farm</i> , George Orwell 1945	TD-RO Тихий Дон, Михаил Шолохов 1928 [Tixij Don, Mikhail Sholokhov]
+ AF-R3 Polock, Skotnyj Dvor 1980	+ TD-E1 Garry, <i>And Quiet Flows the Don</i> 1934
+ AF-R2 Task, Skotskij Ugolok 1989	
+ AF-R1 Pribylovskij, Zverskaja Ferma: Skazka 2002	
FH-EO <i>Fahrenheit 451</i> , Ray Bradbury 1953	OD-RO Один День Ивана Денисовича, Солженицын 1962 [Oden den' Ivana Denisoviča, Solzhenitsyn]
+ FH-R1 Šinkhar', 451 Gradus po Farengjeju 1956	+ OD-E1 Hingley & Hayward, <i>One Day in the Life of Ivan Denisovich</i> 1963
	+ OD-E2 Willetts, <i>One Day in the Life of Ivan Denisovich</i> 1991
JS-EO <i>Jonathan Livingston Seagull</i> , Richard Bach 1970	PO-RO Пикник на Обочине, Стругацкие [Piknik na Obočine, Strugatsky & Strugatsky] 1972
+ JS-R1 Rodman, Džonatan Livingston Čajka 1989	+ PO-E1 Bouis, <i>Roadside Picnic</i> 1977
HP-EO <i>Harry Potter and the Order of the Phoenix</i> , Joanne Rowling 2003	KF-RO Книга Без Фотографий, Сергей Шаргунов 2011 [Kniga bez Fotografij, Sergej Šargunov]
+ HP-R1 Babkov, Golyšev & Motylev, Garri Potter i Orden Feniksa 2004	+ KF-E1 Patterson, <i>A Book without Photographs</i> 2012
Narrative-based Fiction	
ST-FO L'Étranger, Albert Camus 1942 [FR original not included in corpus]	ST-FO L'Étranger, Albert Camus 1942 [FR original not included in corpus]
+ ST-E1F Gilbert, <i>The Stranger</i> 1946	+ ST-R1F Adamovich, <i>Neznakomec</i> 1966
+ ST-E2F Ward, <i>The Stranger</i> 1988	+ ST-R2F Gal, <i>Postoronnij</i> 1968

Figure 14. Summary of corpus organized by language (red text = Russian; blue text = English), translation correspondence, year of publication and genre

Text	Language	Type	Genre	Year	Word Tokens	Word Types	Sentences
BK_RO	Russian	Original	DF	1880	33984	9381	2367
BK_E1	English	Translation	DF	1990	42853	4982	2459
BK_E2	English	Translation	DF	1994	44225	5580	2401
TD_RO	Russian	Original	DF	1928	93641	29843	11786
TD_E1	English	Translation	DF	1934	130476	11224	11564
OD_RO	Russian	Original	DF	1962	33364	10006	3587
OD_E1	English	Translation	DF	1963	53281	3820	3987
OD_E2	English	Translation	DF	1991	49982	5306	3891
PO_RO	Russian	Original	DF	1972	48143	13402	5768
PO_E1	English	Translation	DF	1977	58858	6298	5840
KF_RO	Russian	Original	DF	2011	37635	14839	5551
KF_E1	English	Translation	DF	2012	54440	6910	5539
OS_RO	Russian	Original	SP	1954	15892	5473	3182
OS_E1	English	Translation	SP	1969	26422	3277	3491
ST_E1F	English	ThirdLngTr	NF	1946	37283	4341	2418
ST_E2F	English	ThirdLngTr	NF	1988	34254	3617	2493
ST_R1F	Russian	ThirdLngTr	NF	1966	25659	6981	2472
ST_R2F	Russian	ThirdLngTr	NF	1968	25363	7110	2518
SR_EO	English	Original	DF	1926	70396	5411	8973
SR_R1	Russian	Translation	DF	1968	55358	12036	8899
AF_EO	English	Original	DF	1945	30486	4223	1706
AF_R1	Russian	Translation	DF	1980	23813	8223	1818
AF_R2	Russian	Translation	DF	1989	21598	8273	1669
AF_R3	Russian	Translation	DF	2002	24069	7622	1747
FH_EO	English	Original	DF	1953	48107	5563	4490
FH_R1	Russian	Translation	DF	1956	42271	11775	5064
JS_EO	English	Original	DF	1970	9410	1694	733
JS_R1	Russian	Translation	DF	1989	9370	3527	927
HP_EO	English	Original	DF	2003	100491	8942	8592
HP_R1	Russian	Translation	DF	2004	80381	18456	9544
CA_EO	English	Original	SP	1960	15911	1722	2021
CA_R1	Russian	Translation	SP	2006	11006	2917	2031

Table 3. Basic Metadata and Summary Statistics

The following provides the key metadata, summary statistics, and additional information concerning the texts one by one.

BK_RO

Title:	Братья Карамазовы <i>Brat'ja Karamazovy / Brothers Karamazov</i>
Author:	Фёдор Михайлович Достоевский <i>Fëdor Mixajlovič Dostoevskij / Fyodor Dostoyevsky</i>
Year of Publication:	1880
Language:	Russian
Language Type:	Original source
Language Genre:	Discourse-based Fiction
Sample:	First 14 chapters
Size:	33,984 word-tokens; 9,381 word-types; 2,367 sentences

Additional notes:

In addition to being a source of innumerable studies in literature and philosophy, Dostoyevsky's writing has been praised as particularly suitable for the study of *spoken* language register. George Thomas (1982: 672) stresses that the important role that Dostoyevsky gives to dialogue makes his novels of particular interest to linguists, especially those interested in the study of speech acts.

The particular novel *Brothers Karamazov* relies, sometimes exclusively, on direct speech dialogue. Furthermore, even the narration is of a conversational oral style. René Wellek (1980) describes the writing as particularly *drama* like:

Dostoyevsky [builds] his novels around scenes in dialogue, on conversations, debates and arguments between three or more persons. Nobody would want to deny the general impression of richness, density and multiplicity of conflicting voices. He represents the trend toward the drama in the novel, toward 'objectivity' and 'impersonality', toward the doctrine of the 'exit author'. (Wellek, 1980: 32)

The register of the Russian novel is not simply that of colloquial everyday speech, it uses spontaneous oral language register even in the voice of the narrator, who is himself an *unidentified character* addressing the reader. This conversational style of writing is called '*skaz*' and is known for being a particularly *oral type of writing*:

Skaz (from the Russian *skazyvat'/skazat'* – to tell, to narrate) refers to “a narrative devised as specifically oral in terms of style” and “fashioned to give the illusion of spontaneous speech” [Prince, 1987]. At the beginning of the twentieth century, formalist critic Boris Eikhenbaum saw the “spoken” word of

skaz as the pattern for the evolution of Russian literature. (Mesropova, 2004: 417)

(For more on *skaz*, and its use in Dostoyevsky's writing, see Mikhail Bakhtin, 1963/1984; 1963/2002).

Furthermore, despite the fact that the novel dates to the late 19th century, it has a very wide audience. It continues being of a conversational register in modern Russia where it is not only widely received, but also part of the basic school curriculum. The novel is therefore presently held to be particularly important also in terms of the parallel-text selection criteria of *reception*, i.e. the influence that a text has through wide exposure.

Furthermore, in terms of the verbless sentence phenomenon, there is evidence to believe that the use of verbless sentences in Russian had stabilized at the time of Dostoyevsky's writing. The extensive work of Mikhail Kopotev (1999; 2007b) into the history and evolution of verbless sentences in Russian suggests that the dropping out of the verb in Russian sentences was a consecutive development that started in the first half of the 11th century and stabilized in the 17th century for informal register, and in the 18th century for formal register (i.e. in terms of both copula and non-copula verbs). Thus, in the late 19th century, at the time of Dostoyevsky's writing, and even in the early 19th century, at the time of Dostoyevsky's birth, the use of verbless sentences in Russian appears to have stabilized in terms of frequency according to available diachronic research.

Finally, the novel has been translated into many languages. Regarding English alone, there exist at minimum 7 published full translations from the original text, i.e. Constance Garnett (1912), David Magarshack (1958), Andrew McAndrew (1970), Julius Katzer (1980), Richard Pevear and Larissa Volokhonsky (1990), David McDuff (1993), Ignat Avsey (1994), as well as at least 6 revisions of the first Constance Garnett translation, i.e. Avrahm Yarmolinsky (1933), William Sharp (1943), Alexandra Kropotkin (1949), Manuel Komroff (1960), Ralph Matlaw (1976), Susan McReynolds Oddo (2011).

For the present study, two *recent* translations have been chosen that have been translated *from the original text*, have *different translation styles*, and are both *popular* and *critically acclaimed* translations. The Pevear and Volokhonsky (1990) translation has a literal style of translation that stays close to the source text. It has been acclaimed as being most true to the original novel. The Ignat Avsey translation is also celebrated, but for the way that it captures stylistic nuances in English (Vasil'chenko, 2007). For instance, in the Avsey translation, the title

of novel is translated as *The Karamazov Brothers* which stylistically reflects the non-marked register of the title in Russian.

BK_E1

Title: The Brothers Karamazov
Translator: Richard Pevear and Larissa Volokhonsky
Year of Publication: 1990
Language: English
Language Type: Translation from Russian
Language Genre: Discourse-based Fiction
Sample: First 14 chapters
Size: 42,853 word-tokens; 4,982 word-types; 2,459 sentences

BK_E2

Title: The Karamazov Brothers
Translator: Ignat Avsey
Year of Publication: 1994
Language: English
Language Type: Translation from Russian
Language Genre: Discourse-based Fiction
Sample: First 14 chapters
Size: 44,225 word-tokens; 5,580 word-types; 2,401 sentences

TD_RO

Title: Тихий Дон
Tixij Don / Quiet Don
Author: Михаил Александрович Шолохов
Mikhail Aleksandrovič Šoloxov / Mikhail Sholokhov
Year of Publication: 1928
Language: Russian
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Book 1 of the four-book novel
Size: 93,641 word-tokens; 29,843 word-types; 11,786 sentences

TD_E1

Title: And Quiet Flows the Don
Translator: Stephen Garry
Year of Publication: 1934
Language: English
Language Type: Translation from Russian
Language Genre: Discourse-based Fiction
Sample: Book 1 of the four-book novel
Size: 130,476 word-tokens; 11,224 word-types; 11,564 sentences

OD_RO

Title: Один День Ивана Денисовича
Odin Den' Ivana Denisoviča / One Day of Ivan Denisovich

Author: Александр Исаевич Солженицын
Aleksandr Isaevič Solženicyn / Aleksandr Solzhenitsyn

Year of Publication: 1962

Language: Russian

Language Type: Original source

Language Genre: Discourse-based Fiction

Sample: Full novel

Size: 33,364 word-tokens; 10,006 word-types; 3,587 sentences

OD_E1

Title: One Day in the Life of Ivan Denisovich

Translator: Ronald Hingley and Max Hayward

Year of Publication: 1963

Language: English

Language Type: Translation from Russian

Language Genre: Discourse-based Fiction

Sample: Full novel

Size: 53,281 word-tokens; 3,820 word-types; 3,987 sentences

OD_E2

Title: One Day in the Life of Ivan Denisovich

Translator: Harry Taylor Willetts

Year of Publication: 1991

Language: English

Language Type: Translation from Russian

Language Genre: Discourse-based Fiction

Sample: Full novel

Size: 49,982 word-tokens; 5,306 word-types; 3,891 sentences

PO_RO

Title: Пикник на Обочине
Piknik na Obočine / Roadside Picnic

Author: Аркадий Н. Стругатский и Борис Н. Стругатский
Arkadij N. & Boris N. Strugatskij / Arkady & Boris Strugatsky

Year of Publication: 1972

Language: Russian

Language Type: Original source

Language Genre: Discourse-based Fiction

Sample: Full novel

Size: 48,143 word-tokens; 13,402 word-types; 5,768 sentences

PO_E1

Title: Roadside Picnic
Translator: Antonina Bouis
Year of Publication: 1977
Language: English
Language Type: Translation from Russian
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 58,858 word-tokens; 6,298 word-types; 5,840 sentences

KF_RO

Title: Книга без Фотографий
Kinga bez Fotografij / Book without Photographs
Author: Сергей Александрович Шаргунов
Sergej Aleksandrovič Šargunov / Sergei Shargunov
Year of Publication: 2011
Language: Russian
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 37,635 word-tokens; 14,839 word-types; 5,551 sentences

KF_E1

Title: A Book without Photographs
Translator: Simon Patterson
Year of Publication: 2012
Language: English
Language Type: Translation from Russian
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 54,440 word-tokens; 6,910 word-types; 5,539 sentences

OS_RO

Title: Олень и Шалашовка
Olen' i Šalašovka / Deer and Tramp
Author: Александр Исаевич Солженицын
Aleksandr Isaevič Solženicyн / Aleksandr Solzhenitsyn
Year of Publication: 1954
Language: Russian
Language Type: Original source
Language Genre: Scripted Speech Fiction
Sample: Full play (stage directions removed)
Size: 15,892 word-tokens; 5,473 word-types; 3,182 sentences

OS_E1

Title: The Love-Girl and the Innocent
Translator: Nicholas Bethel and David Burg
Year of Publication: 1969
Language: English
Language Type: Translation from Russian
Language Genre: Scripted Speech Fiction
Sample: Full play (stage directions removed)
Size: 26,422 word-tokens; 3,277 word-types; 3,491 sentences

ST_E1F

Title: The Stranger
Translator: Stuart Gilbert
Year of Publication: 1946
Language: English
Language Type: Translation from French
Language Genre: Narrative-based Fiction
Sample: Full novel
Size: 37,283 word-tokens; 4,341 word-types; 2,418 sentences

ST_E2F

Title: The Stranger
Translator: Matthew Ward
Year of Publication: 1988
Language: English
Language Type: Translation from French
Language Genre: Narrative-based Fiction
Sample: Full novel
Size: 34,254 word-tokens; 3,617 word-types; 2,493 sentences

ST_R1F

Title: Незнакомец
Neznakomec / Stranger
Translator: Георгий Викторович Адамович
Georgij Viktorovič Adamovič / Georgy Adamovich
Year of Publication: 1966
Language: Russian
Language Type: Translation from French
Language Genre: Narrative-based Fiction
Sample: Full novel
Size: 25,659 word-tokens; 6,981 word-types; 2,472 sentences

ST_R2F

Title: Посторонний
Postoronnij / Stranger
Translator: Нора Галь
Nora Gal' / Nora Gal
Year of Publication: 1968
Language: Russian
Language Type: Translation from French
Language Genre: Narrative-based Fiction
Sample: Full novel
Size: 25,363 word-tokens; 7,110 word-types; 2,518 sentences

SR_EO

Title: The Sun Also Rises
Author: Ernest Hemingway
Year of Publication: 1926
Language: English
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 70,396 word-tokens; 5,411 word-types; 8,973 sentences

SR_R1

Title: Фиеста (И Восходит Солнце)
Fiesta (I Vosxodit Solnce) / Fiesta (And the Sun Rises)
Translator: Вера Максимовна Топер
Vera Maksimovna Toper
Year of Publication: 1968
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 55,358 word-tokens; 12,036 word-types; 8,899 sentences

AF_EO

Title: Animal Farm
Author: George Orwell
Year of Publication: 1945
Language: English
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 30,486 word-tokens; 4,223 word-types; 1,706 sentences

AF_R1

Title: Скотный Двор
Skotnyj Dvor / Animal Farm
Translator: ИланIZEКИИЛОВИЧ ПОЛОЦК
Ilan Izekiilovič Polock / Ilan Polock
Year of Publication: 1980
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 23,813 word-tokens; 8,223 word-types; 1,818 sentences

AF_R2

Title: Скотский Уголок
Skotskij Ugolok / Animal Farm
Translator: Сергей ЭМИЛЬЕВИЧ ТАСК
Sergej Èmil'evič Task / Sergei Task
Year of Publication: 1989
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 21,598 word-tokens; 8,273 word-types; 1,669 sentences

AF_R3

Title: Зверская Ферма: Сказка
Zverskaja Ferma: Skazka / Animal Farm: A Tale
Translator: Владимир Валерианович Прибыловский
Vladimir Valerianovič Pribylovskij / Vladimir Pribylolovsky
Year of Publication: 2002
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 24,069 word-tokens; 7,622 word-types; 1,747 sentences

FH_EO

Title: Fahrenheit 451
Author: Ray Bradbury
Year of Publication: 1953
Language: English
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 48,107 word-tokens; 5,563 word-types; 4,490 sentences

FH_R1

Title: 451 Градусов по Фаренгейту
451 Gradusov po Farengėjtu / 451 Degrees Fahrenheit
Translator: Татьяна Николаевна Шинкаръ
Tat'jana Nikolaevna Šinkar' / Tat'jana Shinkar'
Year of Publication: 1956
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 42,271 word-tokens; 11,775 word-types; 5,064 sentences

JS_EO

Title: Jonathan Livingston Seagull
Author: Richard Bach
Year of Publication: 1970
Language: English
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 9,410 word-tokens; 1,694 word-types; 733 sentences

JS_R1

Title: Джонатан Ливингстон Чайка
Džonatan Livingston Čajka / Jonathan Livingston Seagull
Translator: Юни Самуиловна Родман
Juni Samuilovna Rodman / Uni Rodman
Year of Publication: 1989
Language: Russian
Language Type: Translation from English
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 9,370 word-tokens; 3,527 word-types; 927 sentences

HP_EO

Title: Harry Potter and the Order of the Phoenix
Author: Joanne Rowling
Year of Publication: 2003
Language: English
Language Type: Original source
Language Genre: Discourse-based Fiction
Sample: Full novel
Size: 100,491 word-tokens; 8,942 word-types; 8,592 sentences

HP_R1

Title: Гарри Поттер и Орден Феникса
Garri Potter i Orden Feniksa / Harry Potter and Order of Phoenix

Translator: Владимир Бабков, Виктор Гольшев, Леонид Мотылев
Vladimir Babkov, Victor Golyšev, Leonid Motylev

Year of Publication: 2004

Language: Russian

Language Type: Translation from English

Language Genre: Discourse-based Fiction

Sample: Full novel

Size: 80,381 word-tokens; 18,456 word-types; 9,544 sentences

CA_EO

Title: The Caretaker

Author: Harold Pinter

Year of Publication: 1960

Language: English

Language Type: Original source

Language Genre: Scripted Speech Fiction

Sample: Full play (stage directions removed)

Size: 15,911 word-tokens; 1,722 word-types; 2,021 sentences

CA_R1

Title: Сторож
Storož / Caretaker

Translator: Александр Николаевич Дорошевич
Aleksandr Nikolaevič Doroševič / Aleksandr Doroshevich

Year of Publication: 2006

Language: Russian

Language Type: Translation from English

Language Genre: Scripted Speech Fiction

Sample: Full play (stage directions removed)

Size: 11,006 word-tokens; 2,917 word-types; 2,031 sentences

CHAPTER 3

COMPARABILITY ANALYSIS

The created corpus is presently evaluated as loosely comparable according to several dimensions. The current section analyzes its comparability in terms of the sampling frame, including: language, language type, genre, time period and size.

3.1 RUSSIAN AND ENGLISH IN NUMBER OF TEXTS

It is esteemed that comparability has been fully achieved in terms of the number of texts in *Russian and English overall*: 16 texts in Russian and 16 in English.

Furthermore, the corpus is also comparable in terms of the number of texts in the *sub-categories for each language*. These sub-categories include language type, i.e. originals (O), translations (T), third-language translations (TF) and genre, i.e. discourse-based fiction (DF), scripted speech (SP), narrative-based fiction (NF), where the two languages are fully comparable in the number of texts. Russian and English are also deemed loosely comparable in terms of the number of texts in a 50-year time period (i.e. 1880 to 1929, 1930 to 1979, 1980 to present).

Figure 15 summarizes the number of texts in Russian and English by category and indicates the texts involved in each (see *Appendix 1* and *Part 3: Section 2.4* above for the text labels).

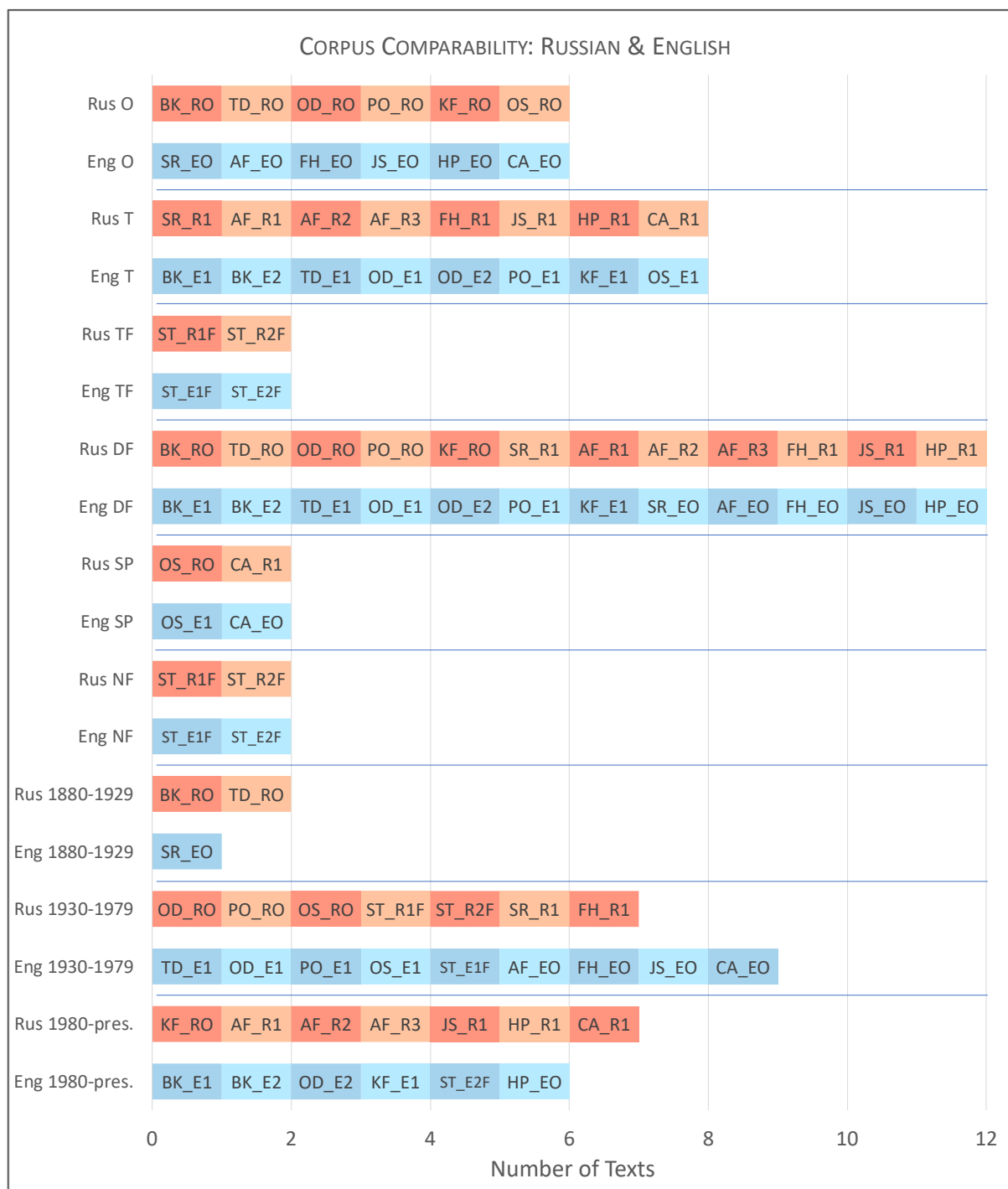


Figure 15. Comparability of Russian and English in Number of Texts

3.2 RUSSIAN AND ENGLISH SOURCE TYPE IN NUMBER OF WORDS

A closer look from the perspective of the number of words (i.e. tokens) shows that the corpus is loosely comparable in terms of Russian and English originals (O), translations (T) and third-language translations from French (TF).

The two languages are balanced in terms of the size of the original subparts ('Rus O' and 'Eng O'). The third-language-translation subparts ('Rus TF' and 'Eng TF') are also balanced. An imbalance exists in the size of the English translation sub-corpus and that of the Russian (i.e. 'Rus T' and 'Eng T'). The reason for this has to do with the contrastive analysis goals at hand and is explained in *Part 3: Section 2.3* above.

Furthermore, the Russian sub-corpus is balanced in terms of the number of words in originals and translations (i.e. 'Rus O' and 'Rus T').

Serving control purposes, the third language subparts are smaller relative to the size of originals and translations.

Figure 16 illustrates the number of words in the Russian and English subparts according to the type of language. It also indicates the relative size, in tokens, of each text involved (see *Appendix 1* and *Part 3: Section 2.4* above for the text labels).

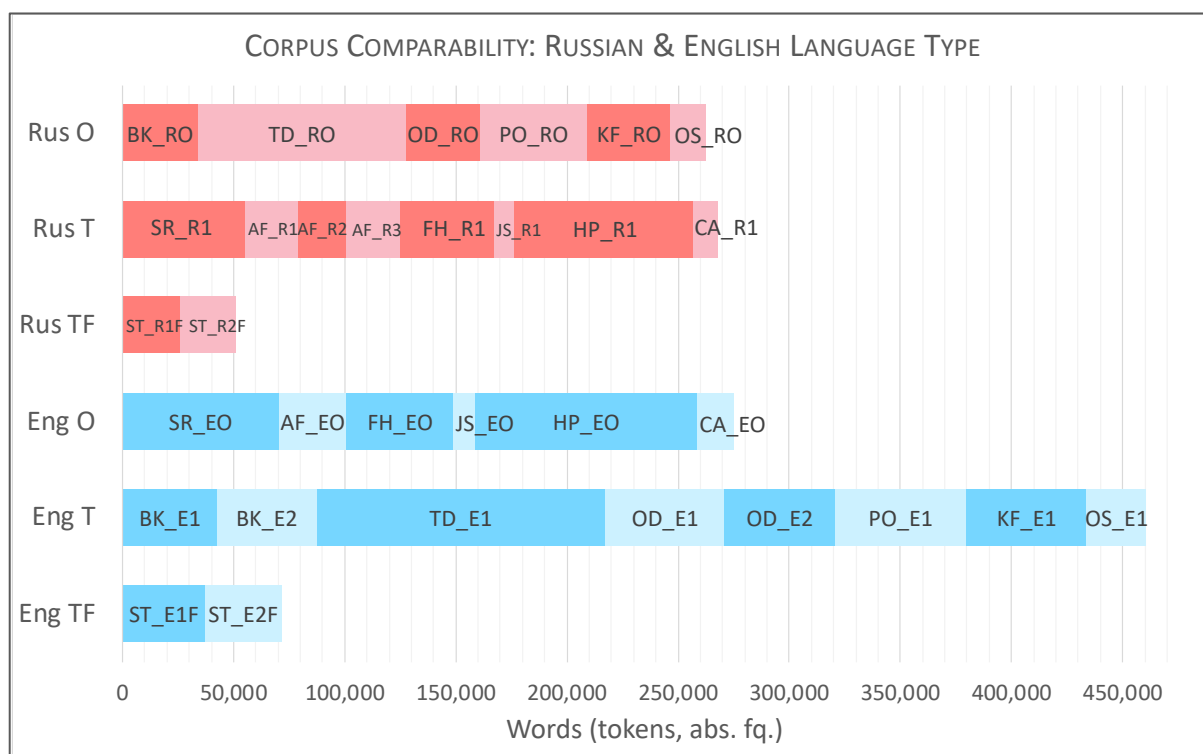


Figure 16. Comparability of Russian and English in Words: Language Source Type

3.3 RUSSIAN AND ENGLISH GENRE IN NUMBER OF WORDS

The Russian and English corpus parts are also loosely comparable in terms of the number of words (i.e. tokens) per function-based genre, i.e. discourse-based fiction (DF), scripted speech (SP) and narrative-based fiction (NP).

Figure 17 illustrates the number of words in Russian and English according to the function-based genre. It also indicates the relative size, in tokens, of each text involved in the genre (see Appendix 1 and Section 2.4 above for the text labels).

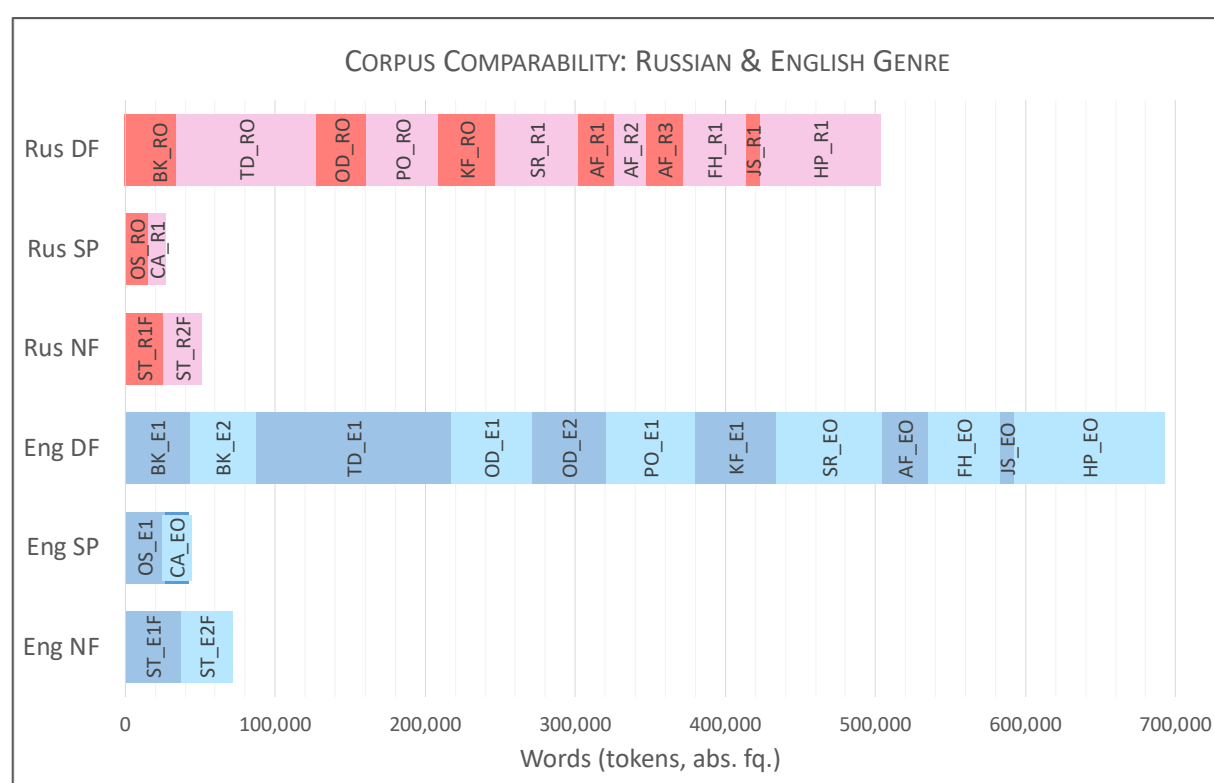


Figure 17. Comparability of Russian and English in Words: Genre

An imbalance exists in the size of English discourse-based fiction ('Eng DF') which exceeds that of the Russian discourse-based fiction ('Rus DF'). This is explained by the English translations. As discussed in *Part 3: Section 2.3*, creating balance with English translations by artificially reducing the size of this corpus part would be inappropriate as it would mean losing the link between the original and the translation and therefore run contrary to the contrastive analysis goals at hand.

Furthermore, although the size of the English discourse-based fiction exceeds that of the Russian discourse-based fiction in terms of the absolute number of words, the proportion that discourse-based fiction has in the Russian and in the English subparts is balanced. Specifically, discourse-based fiction represents 86.6% of the Russian sub-corpus and 85.9% of the English sub-corpus (these values are discussed in *Part 3: Section 2.3* above).

Serving control purposes, the scripted speech fiction and narrative-based fiction subparts are smaller relative to the size of discourse-based fiction. They are both balanced in terms of their size in the two languages.

3.4 RUSSIAN AND ENGLISH TIME PERIOD IN NUMBER OF WORDS

Figure 18 illustrates the number of words in the Russian and English corpus parts according to the time period considered in 50-year spans from 1880 to 1929, 1930 to 1979, 1980 to the present. It also displays the relative size, in tokens, of each text involved in the time period (see *Appendix 1* and *Part 3: Section 2.4* above for the text labels).

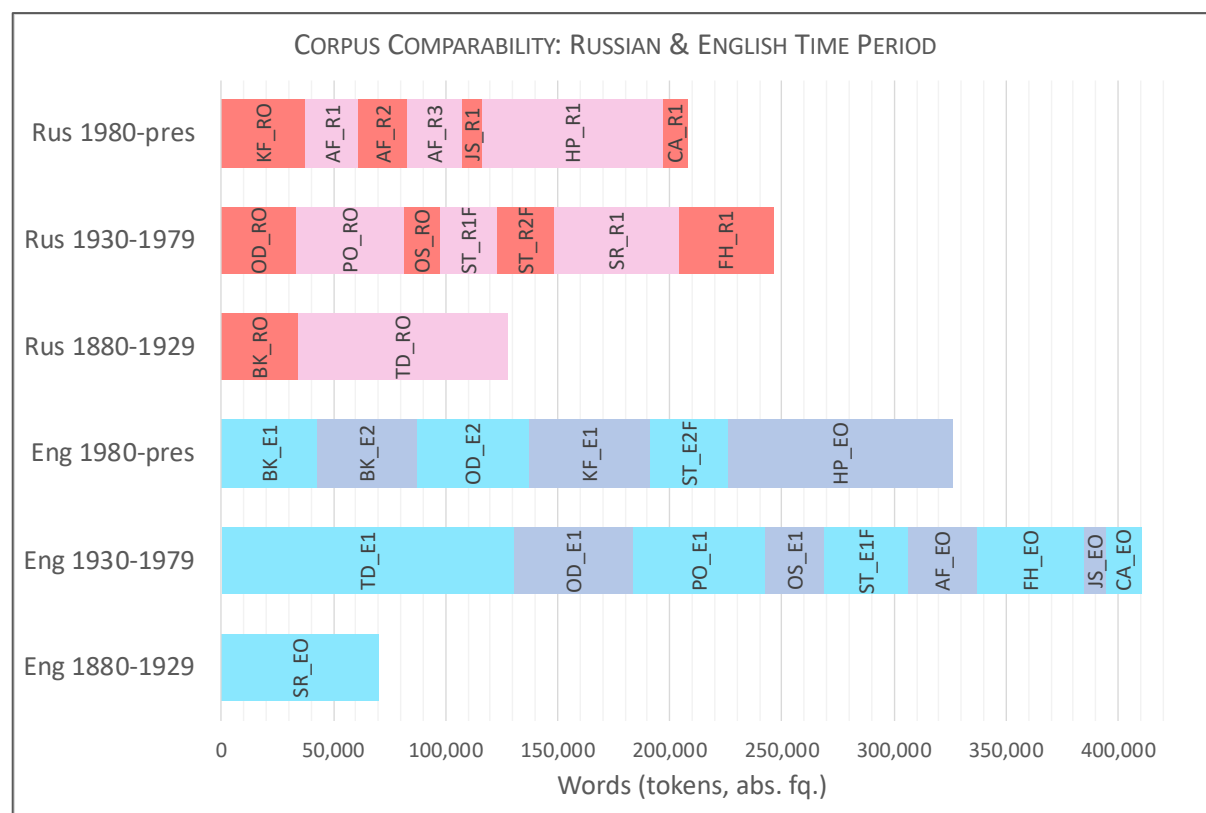


Figure 18. Comparability of Russian and English in Words: Time Period

Although Russian and English differ in terms of the absolute number of word tokens in the three time periods, they are proportionally loosely comparable.

More specifically, the period most represented in both languages is that from the 1930 to 1979. This time period corresponds to 50.8% of the English sub-corpus (i.e. 410,234 words of the 806,875 words of the English sub-corpus) and 42.3% of the Russian sub-corpus (i.e. 246,050 words of the 581,547 words of the Russian sub-corpus).

The next most sizable period in terms of words is the most recent one, between 1980 and the present time. It corresponds to 40.4% of the English sub-corpus (i.e. 326,246 words) and 35.7% of the Russian (i.e. 207,872 words). Furthermore, it is loosely comparable to the first 1930 to 1979 time period in terms of size.

Finally, the period between 1880 and 1929 corresponds to 8.7% of the English sub-corpus (i.e. 70,396 words; 1 text) and 21.9% of the Russian sub-corpus (i.e. 127,625; 2 texts). It is the smallest relative to the other time periods in both languages.

3.5 PARALLEL-TEXT COMPARABILITY MAP

All of the dimension of the sampling frame of the constructed parallel-corpus are summarized in *Figure 19*. Each circle represents a text of a specific size (corresponding to the number of word tokens). The genre is marked by the colour of the circle (green for discourse-based fiction, pink for scripted-speech, yellow for narrative-based fiction). The text-circles are positioned on a timeline corresponding to their year of publication. A red line runs down from the timeline for Russian texts, a blue line for English texts. Finally, the y-axis marks the original, translation, third-language translation distinction by positioning the texts accordingly. In this way, each text has unique coordinates in terms of time, source-type, language, genre and size.

The code of each text (see *Appendix 1*; e.g., BK_E1 corresponds to the *Brothers Karamazov* English translation number 1), the specific year and the size in tokens is further specified under each text on the map. We call it a parallel-text comparability map.

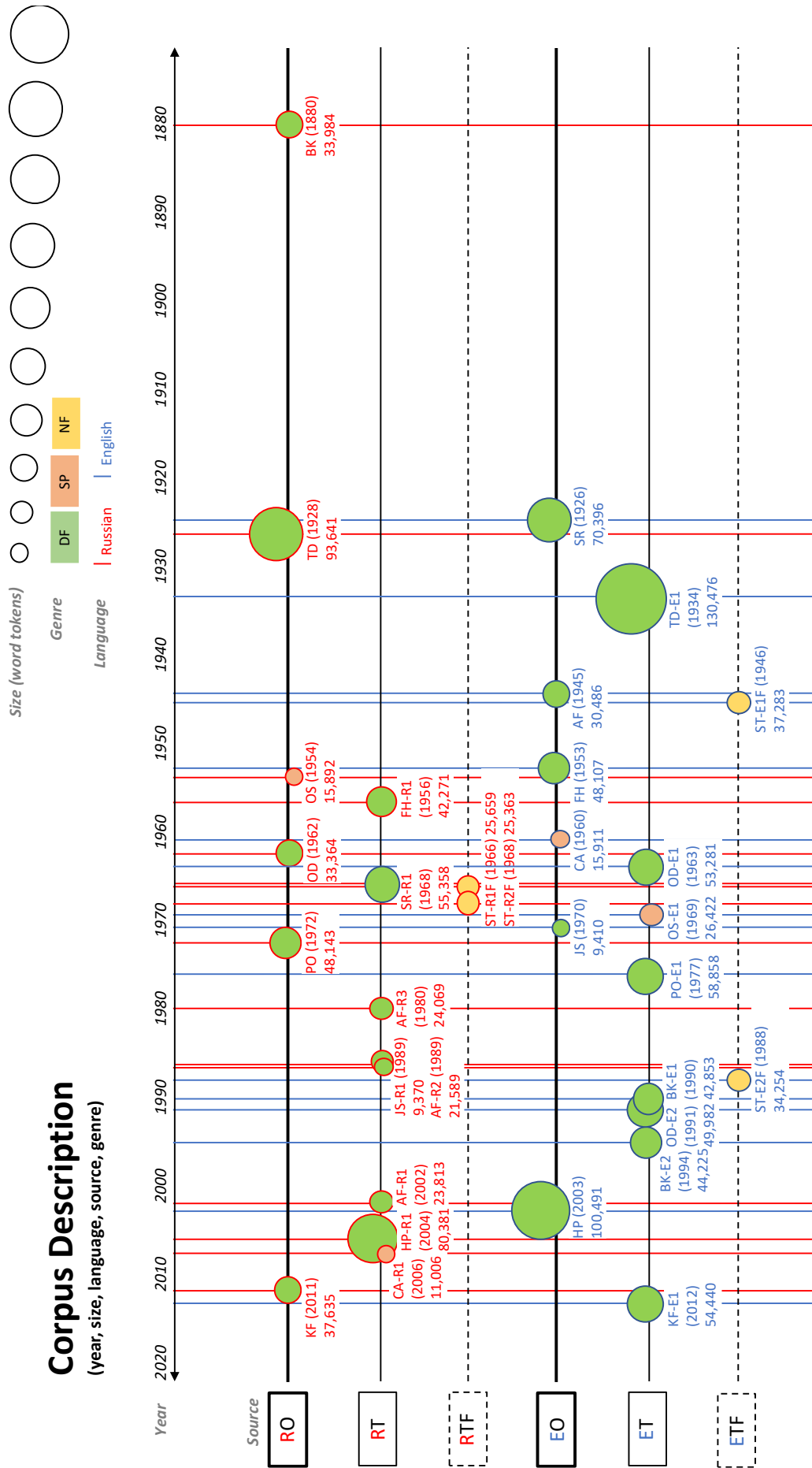


Figure 19. Parallel-Text Comparability Map

PART FOUR:

RESULTS

CHAPTER 1

SYNTACTIC EXPLANATIONS

1.1 FREQUENCY OF VERBLESS SENTENCES

It is not surprising that Russian shows a greater frequency of verbless sentences compared to English, however what is new is that the extent is now quantifiable. According to the present corpus, in general, 17.9% of Russian sentences, and 9.3% of English, are verbless.

The absolute frequency of verbless sentences in the corpus are presented in *Figure 20*. The English sub-corpus measures 70,598 sentences of which 6,581, or 9.3%, are verbless. The Russian sub-corpus contains 68,930 sentences of which 12,354, or 17.9%, are verbless. Thus, overall, Russian shows almost twice (1.92 times) more verbless sentences than English.

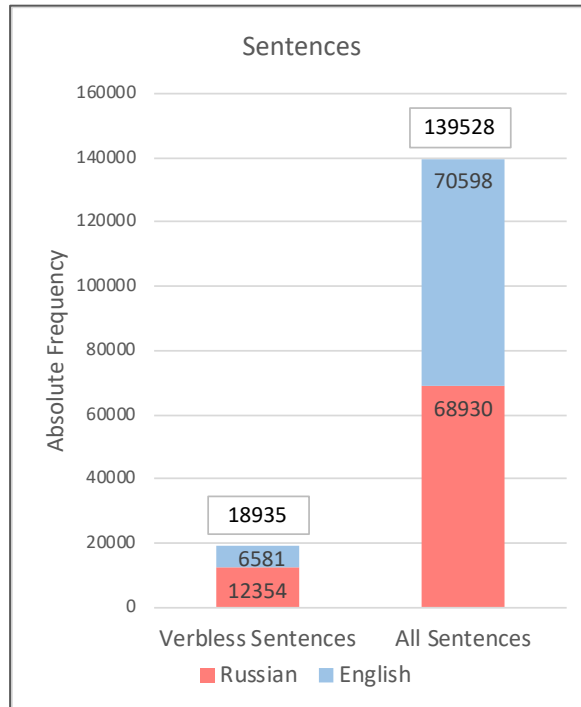


Figure 20. Sentences by language (absolute frequency)

The above results and *Figure 20* concern the entire corpus and hide the differences between the individual texts. Keeping within the absolute frequency perspective, *Figure 21* breaks down the total 18,935 verbless sentences in the corpus according to the text that they come from.

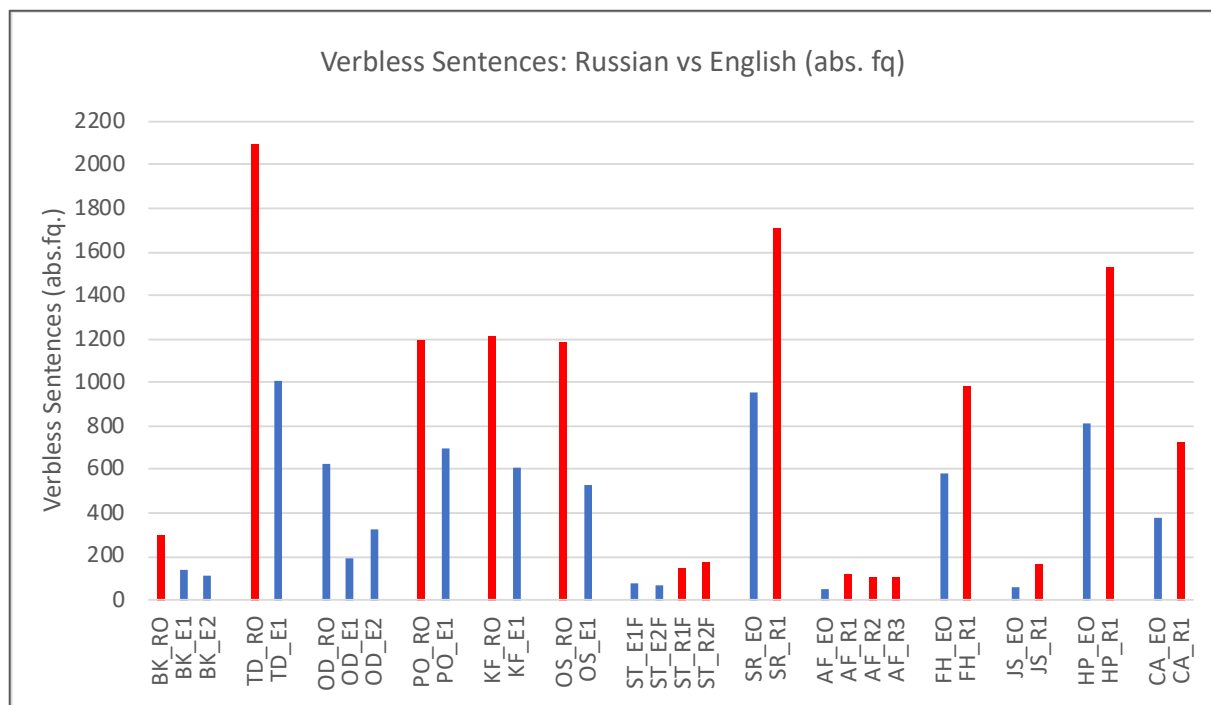


Figure 21. Verbless sentences in Russian and English by text (absolute frequency)

The absolute frequency of verbless sentences per text reveals that *Quiet Flows the Don*, *The Sun Also Rises* and *Harry Potter* provide the largest source of verbless sentences, whereas *The Stranger*, *Animal Farm* and *Jonathan Livingston Seagull* the least, for both languages. This provides key information for the interpretation of any results concerning the Russian and English corpus parts considered as a whole.

However, the absolute numbers of verbless sentences per text cannot be used for the comparison of the texts between themselves. For instance, while *Jonathan Livingston Seagull* provides only 55 verbless sentences, it is also the smallest book in the corpus, with only 9,410 words and 733 sentences total in English, whereas *Harry Potter* provides 815 verbless sentences but also measures 100,491 words and 8,592 sentences in English. In order to compare the proportion of verbless sentences per text, the size of the texts must be controlled. In other words, the absolute frequency must be normalized. *Figure 22* shows the relative frequency of verbless sentences per text, normalized per 100 sentences. More precisely, the verbless sentence absolute values were divided by the total sentences per text and multiplied by 100.

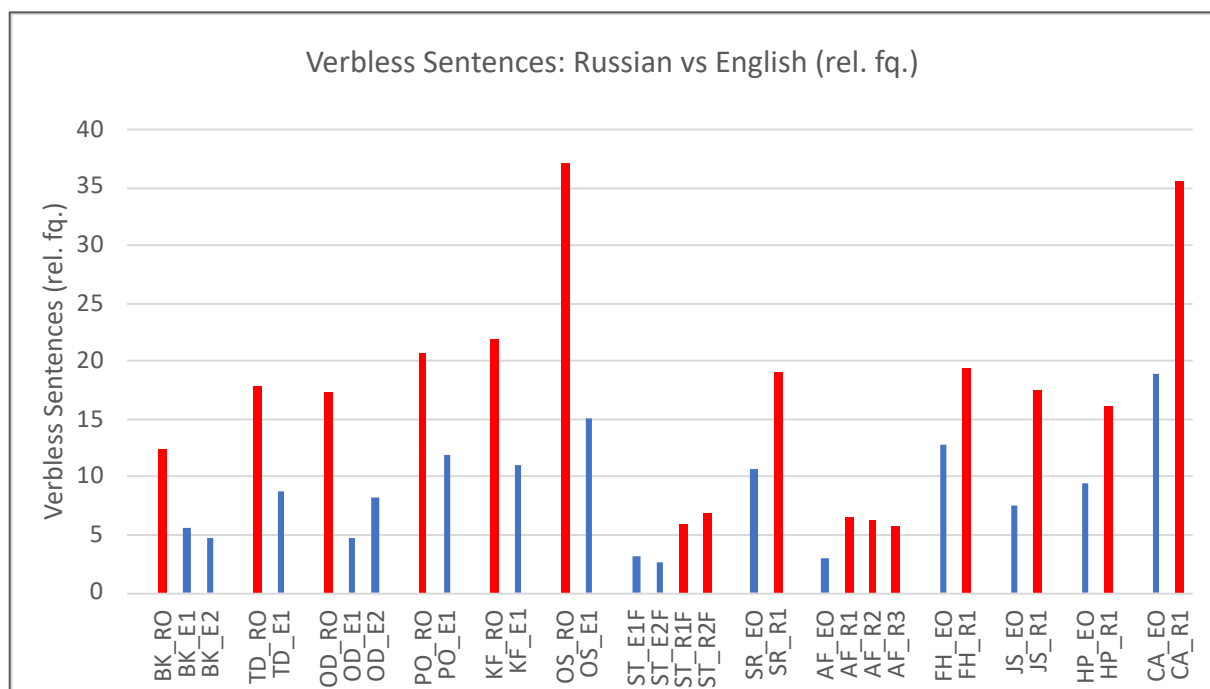


Figure 22. Verbless sentences in Russian and English (normalized per 100 sentences)

The figure reveals that the normalized frequency is not stable across the texts and pinpoints the leading texts in terms of verbless sentence proportions. Furthermore, it highlights that the relative frequency of verbless sentences in the Russian texts (in red) consistently exceeds the relative frequency of the English texts (in blue).

The precise values concerning the number of verbless sentences per text are presented in *Table 4*. The values include the absolute frequency (of sentences, verbless sentences and word tokens), the relative frequency of verbless sentences *normalized per 100 sentences*, as well as the relative frequency of verbless sentences *normalized per 100 words*.

Text	Total Sentences (Abs. Fq.)	Verbless Sentences (Abs. Fq.)	Verbless Sentences (Rel. Fq. per 100 sentences)	Total Words (Tokens, Abs. Fq.)	Verbless Sentences (Rel. Fq. per 100 words)
BK_RO	2367	295	12.46	33984	0.87
BK_E1	2459	139	5.65	42853	0.32
BK_E2	2401	114	4.75	44225	0.26
TD_RO	11786	2099	17.81	93641	2.24
TD_E1	11564	1007	8.71	130476	0.77
OD_RO	3587	624	17.40	33364	1.87
OD_E1	3987	188	4.72	53281	0.35
OD_E2	3891	323	8.30	49982	0.65
PO_RO	5768	1190	20.63	48143	2.47
PO_E1	5840	697	11.93	58858	1.18
KF_RO	5551	1213	21.85	37635	3.22
KF_E1	5539	611	11.03	54440	1.12
OS_RO	3182	1181	37.12	15892	7.43
OS_E1	3491	528	15.12	26422	2.00
ST_E1F	2418	75	3.10	37283	0.20
ST_E2F	2493	67	2.69	34254	0.20
ST_R1F	2472	148	5.99	25659	0.58
ST_R2F	2518	172	6.83	25363	0.68
SR_EO	8973	953	10.62	70396	1.35
SR_R1	8899	1706	19.17	55358	3.08
AF_EO	1706	50	2.93	30486	0.16
AF_R1	1818	119	6.55	23813	0.50
AF_R2	1669	106	6.35	21598	0.49
AF_R3	1747	102	5.84	24069	0.42
FH_EO	4490	578	12.87	48107	1.20
FH_R1	5064	982	19.39	42271	2.32
JS_EO	733	55	7.50	9410	0.58
JS_R1	927	163	17.58	9370	1.74
HP_EO	8592	815	9.49	100491	0.81
HP_R1	9544	1533	16.06	80381	1.91
CA_EO	2021	381	18.85	15911	2.39
CA_R1	2031	721	35.50	11006	6.55
Total RU	68930	12354	17.92	581547	2.12
Total EN	70598	6581	9.32	806875	0.82

Table 4. Verbless sentences per text

Table 4 highlights the *range* in the verbless sentence frequency across the texts. From the perspective of the overall corpus, in Russian 17.9 per 100 sentences are verbless, whereas in English, 9.3 per 100 sentences are verbless. However, a closer look at the texts reveals that the relative frequency in the *Russian texts ranges from 5.84 to 37.12* verbless sentences per 100 sentences, depending on the text. In the *English texts, this range is from 2.69 to 18.85* verbless sentences per 100 sentences.

Normalizing verbless sentences *per number of words*, as opposed to sentences, is important for considerations of required **SAMPLE SIZE** for further studies. In terms of the number of word tokens, according to the present Russian and English corpus parts, one finds *2.12 verbless sentences every 100 Russian words* and *0.82 verbless sentences every 100 English words*. A wide range is revealed from the perspective of individual texts. The relative frequency in the Russian texts ranges from 0.42 to 7.43 verbless sentences per 100 words, and in the English texts from 0.20 to 2.39 verbless sentences per 100 words. To put it in another way, based on the texts of the present corpus, one requires from *238 to 14 Russian words for one verbless sentence*, and from *500 to 42 English words* to encounter one verbless sentence.

The reasons for such a wide range will be explored in *Part 4: Chapter 3* where the texts will be analyzed in terms of genre and source-language type. The focus of the present section however is on Russian and English.

In addition to quantifying the frequency differences, we are able to show that the famous typological distinction between Russian and English in terms of verbless sentence frequency is statistically significant. This result was shown in (Bondarenko, 2019) for the *Brothers Karamazov* sub-corpus, and it is presently confirmed. A Chi-squared test showed that the distribution of the categorical variable of the verbless sentence is not the same across Russian and English, yielding p-values of less than $2.2e-16$. *Figure 23* presents the contingency tables and results of the test for *Fahrenheit 451* and its Russian translation, and of *Quiet Flows the Don* and its English translation.

<pre>> summary(FH_EO_R1) Sentence Language Length:9554 Length:9554 Class :character Class :character Mode :character Mode :character</pre>	<pre>> table(Sentence,Language) Language Sentence English Russian Verbal 3912 4082 Verbless 578 982</pre>	<pre>> chisq.test(TABLE,correct=T) Pearson's Chi-squared test with Yates' continuity correction data: TABLE X-squared = 73.546, df = 1, p-value < 2.2e-16</pre>

<pre>> summary(TD_RO_E1) Sentence Language Length:23350 Length:23350 Class :character Class :character Mode :character Mode :character</pre>	<pre>> table(Sentence,Language) Language Sentence English Russian Verbal 10557 9687 Verbless 1007 2099</pre>	<pre>> chisq.test(TABLE,correct=T) Pearson's Chi-squared test with Yates' continuity correction data: TABLE X-squared = 418.45, df = 1, p-value < 2.2e-16</pre>

Figure 23. Chi-squared test for the significance of the variation in the frequency of verbless sentences between RU and EN

It is therefore possible to conclude that the difference in the frequency of verbless sentences in Russian and English is no longer a quantitative supposition, but is now confirmed as being statistically significant.

1.2 ELLIPSIS

To explain the statistically significant difference in the frequency of verbless sentences between Russian and English, it is tempting to appeal to the well-known syntactic differences between the two languages concerning the use of ellipsis. As discussed in *Part 3: Section 3.1.3*, Russian has a great capacity for ellipting verbs that is unmatched by any Indo-European language. The capacity is largely due to its very developed morphological case system. In contrast, English has a particularly limited morphological case system which, as illustrated in *Part 3: Section 3.1.3*, results in reduced productivity of ellipsis due to the fact that key syntactic functions, such as subject, are left unmarked and ambiguous without the explicit presence of the verb. Therefore, comparing the two languages, it is entirely reasonable to expect that some portion of the English utterances would be forced into ‘restoring’ a verb that Russian is able to successfully ellipt.

However, the results of the present analysis show that, as tempting as the syntactic explanation may be, it does not explain the difference in frequency of verbless sentences. Furthermore, the verbless sentence phenomenon reveals itself to be largely an issue for

semantico-pragmatic analysis, as opposed to the syntactic domain in which it has typically been explored.

1.2.1 Steps: Utterances and Annotation

The present analysis has for its focus the fundamental elliptical versus non-elliptical distinction which was described in detail in *Part 1*. Notably, the key aspects of the distinction there discussed are: (a) that the distinction is made based on the presence or the absence of a verbal *antecedent* in the (anterior or posterior) linguistic context, and (b) that an analysis of ellipsis requires consideration of smaller *utterance* units, as opposed to *sentences* (i.e. a sentence may contain more than one utterance, and thus more than one ellipsis).

Therefore, the present analysis proceeded as follows. Once the verbless sentences *and* their translation correspondences were identified, they (i.e. both the verbless sentences and the translation correspondences) were (a) segmented into utterance units, and then (b) the utterance units were annotated for the presence or absence of the verb. The *verbless* utterances, thus identified, were then (c) annotated for the presence or absence of ellipsis, and if applicable (d) annotated for the type of ellipsis (i.e. gapping, sluicing, stripping). The definitions in accordance with which the verbless sentences were segmented into utterances and annotated for ellipsis are described in *Part 1: Chapter 4*. In this way, three categories of utterances were identified: *verbal*, *elliptical* (i.e. antecedent-based), and *non-elliptical*.

Both the utterance segmentation and the annotation steps were performed manually. Consequently, the analysis was presently performed only on a small portion of the corpus. More specifically, it was performed on the *Brothers Karamazov* subsection which includes the original (BK_RO) and its two English translations (BK_E1) and (BK_E2). The present section expands on our results in (Bondarenko, 2019).

Table 5 presents the basic description of the sub-corpus that was used in the analysis, including the absolute frequency of verbless sentences and the number of corresponding utterances into which the verbless sentences were segmented for the contrastive analysis of ellipsis.

Text	Verbless Sentences (Abs. Fq.)	Utterances (Abs. Fq.)
BK_RO	315	436
BK_E1	137	436
BK_E2	133	433

Table 5. Summary of sub-corpus used for contrastive ellipsis analysis

Notably, the absolute number of verbless sentences in the table slightly differs from that of the present full-corpus analysis. This is due to the improvements in the automatic method of retrieval (in which sentence segmentation was slightly changed) and its subsequent standardization to the full corpus, including the BK sub-corpus on which the ellipsis analysis was performed. In particular, these changes concern the number of automatically segmented sentences, not the number of utterances into which verbless sentences were delimited.

1.2.2 Ellipsis overrepresented in *English*

Contrary to expectations, ellipses were found to be significantly over-represented not in Russian, the language that stands out among the Indo-European family for permitting the most liberal use of verbless sentences, but rather in English, the language that is typically associated with a dependence on the finite verb phrase.

To start, *Table 6* reveals the absolute frequency of verbless utterances and their correspondences by type. The three types include: verbless antecedent-based elliptical utterance, verbless non-elliptical utterance, and verbal utterance. To be clear, the presence of *verbal* utterances in this table is due to the contrastive aspect of the analysis, i.e. the verbal utterance is a translation correspondence of a verbless utterance in one of the three texts.

Text	Elliptical	Non-Elliptical	Verbal	Total Utterances
BK_RO	27	392	17	436
BK_E1	93	127	216	436
BK_E2	69	138	226	433

Table 6. Absolute frequency of utterances by type

Figure 24 normalizes the results from *Table 6* per 100 utterances. It makes evident the proportion of verbless utterance correspondence types per text.

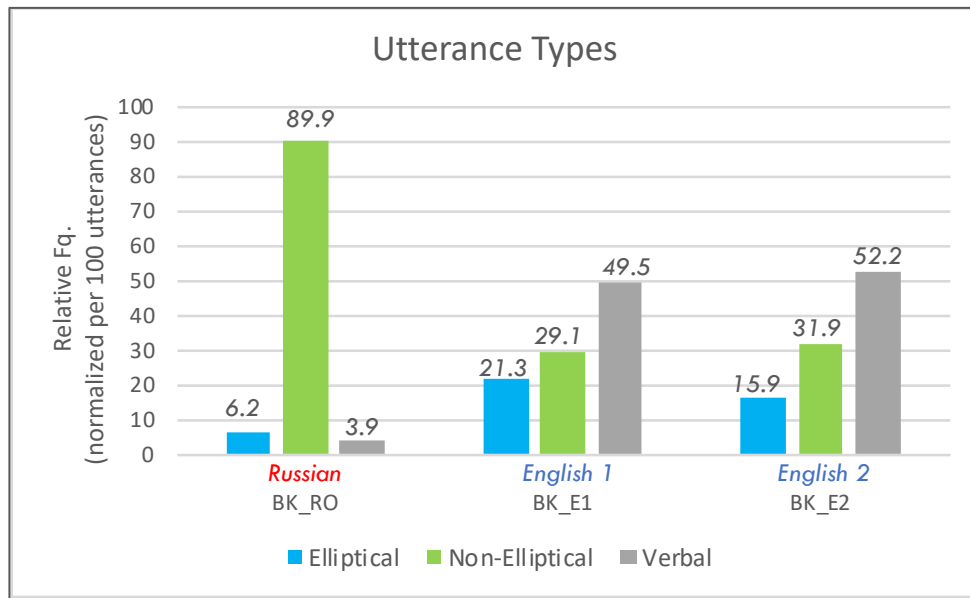


Figure 24. Utterances by type (normalized per 100 utterances)

The figure shows that *verbal* correspondences make up 49.5% and 52.2% of the English utterances (in BK_E1 and BK_E2 respectively) and only 3.9% of Russian utterances. These verbal correspondences are for the moment set aside in order to focus on the verbless utterances and the presence or absence of ellipsis. This perspective is made clearer in *Table 7*, which presents the results in terms of *verbless utterances*. In other words, the table shows only the utterances that are verbless in both languages (i.e. from all directions of comparison).

Text	Elliptical Verbless	Non-Elliptical Verbless	Total Verbless Utterances
BK_RO	27 (6 %)	392 (94 %)	419
BK_E1	93 (42 %)	127 (58 %)	220
BK_E2	69 (33 %)	138 (67 %)	207

Table 7. Absolute frequency of verbless utterances by type

The above results reveal that syntactic ellipsis of the verb explained only 6% of the *Russian verbless* utterances. In contrast, in *English*, on average, 38% of the *verbless* utterances were due to an ellipsis.

The variation in the frequency of ellipsis between Russian and English was tested for statistical significance. The Chi-squared test yielded p-values of less than $2.2e-16$. There is thus evidence to believe that the distribution of the categorical variable of ellipsis is not the same in

Russian and English. In other words, the observed existence of a difference in the frequency of ellipsis in Russian and English is not due to chance, but is statistically significant.

<pre>> summary(ellipsis_E1) Ellipsis Language E:120 E1:220 N:519 RU:419</pre>	<pre>> table(Ellipsis,Language) Language Ellipsis E1 RU E 93 27 N 127 392</pre>	<pre>> chisq.test(TABLE,correct=T) Pearson's Chi-squared test with Yates' continuity correction data: TABLE X-squared = 119.0722, df = 1, p-value <2.20E-16</pre>

<pre>> summary(ellipsis_E2) Ellipsis Language E: 96 E2:207 N:530 RU:419</pre>	<pre>> table(Ellipsis,Language) Language Ellipsis E2 RU E 69 27 N 138 392</pre>	<pre>> chisq.test(TABLE,correct=T) Pearson's Chi-squared test with Yates' continuity correction data: TABLE X-squared = 75.0997, df = 1, p-value <2.20E-16</pre>

Figure 25. Chi-squared test for the significance of the variation in the frequency of ellipsis between RU and EN (BK_RO & BK_E1; BK_RO & BK_E2)

As mentioned, the present results consider all directions of comparison, i.e. not only the verbless sentences from the Russian BK_RO, but the verbless sentences from the English BK_E1 and BK_E2 were also compared with their correspondences. Still, this three-text sub-corpus corresponds to only one main novel whose source language is Russian and which belongs to the discourse-based fiction register. It remains to confirm that these results are not due to the interference of source language, genre, or another factor associated with the presently examined sub-corpus, through further manual annotation of the corpus (i.e. segmentation of verbless sentences and their correspondences into utterances, followed by verbless vs. verbal and elliptical vs. non-elliptical annotation of the utterances).

Nevertheless, the present results are deemed to be striking as they counter the expectation that ellipsis would be more present in a language that is syntactically more adapted for their use. They suggest that the use of ellipsis is over-represented in English, *not* in Russian. Antecedent-based ellipses were found to occur on average three times more frequently in English than in Russian.

An ellipsis of the verb in English, but not in Russian is illustrated in (100).

(100) a. *Russian: non-elliptical utterance*

Дело	в	том,	что	это	пожалуй	и	деятель,
<i>delo</i>	<i>v</i>	<i>tom,</i>	<i>čto</i>	<i>èto</i>	<i>požaluj</i>	<i>i</i>	<i>dejatel'</i>
NN.NOM	PREP	PRO	CONJ	DEM	PART	CONJ	NN.NOM
<i>matter</i>	<i>in</i>	<i>that</i>	<i>what</i>	<i>this</i>	<i>perhaps</i>	<i>and</i>	<i>doer</i>
но	деятель	неопределенный,	не	выяснившийся.			
<i>no</i>	<i>dejatel'</i>	<i>neopredelennyj</i>	<i>ne</i>	<i>vyjasnivšijsja</i>			
CONJ	NN.NOM	ADJ.NOM	NEG	ADJ.NOM			
<i>but</i>	<i>doer</i>	<i>indefinite</i>	<i>not</i>	<i>determinate</i>			

b. *English: elliptical utterance*

The thing is that he does perhaps, make a figure, **but a figure of an indefinite, indeterminate sort.**

The example shows a verbal ellipsis of the stripping type in the English utterance ‘*but a figure of an indefinite, indeterminate sort*’, but not in the Russian ‘но деятель неопределенный, не выяснившийся’ (*no dejatel neopredelennyj, ne vyjasnivšijsja*). The Russian sentence in (a) is entirely verbless, while the English sentence in (b) contains several verbs. The English utterance ‘*he does, perhaps, make a figure*’ contains the verb phrase ‘*does make*’ which serves as the antecedent for the subsequent parallel verbless utterance ‘*but a figure of an indefinite, indeterminate sort*’. In English, the verbal antecedent exists in the linguistic context of the verbless utterance and is then ellipped; in Russian, there is no verbal antecedent that is ellipped in the verbless utterance. This illustrated process results in the significantly higher frequency of ellipsis in English as compared to Russian. Thus, despite Russian syntax allowing for greater productivity of elliptical utterances, it is in English that verbal antecedent-based ellipses occur more frequently.

1.2.3 Comparative Stripping, Sluicing, Gapping

Another striking observation is that despite English having a reduced syntactic capacity for ellipsis compared to Russian, this capacity was sufficient to parallel all of the cases of Russian ellipsis in the present corpus. Virtually all of the occurrences of Russian verbal ellipsis were matched by an ellipsis in the English utterances.

Table 8 breaks down the absolute frequency of ellipsis per text into the stripping, sluicing and gapping types (cf. the definitions in *Part 1: Section 4.5*).

Text	Ellipsis Type			Total
	Stripping	Sluicing	Gapping	
BK_RO	18	9	0	27
BK_E1	73	18	2	93
BK_E2	53	14	2	69

Table 8. Absolute frequency of ellipsis by type per text

A contrastive look at the ellipses from *Table 8* is presented in *Table 9*. This table compares the ellipses from the Russian BK_RO to those of the English translation BK_E1, and a similar comparison is made with the second English translation BK_E2. The table shows the number of ellipsis that match between the two compared texts, as well as the number that do not match and belong either only to the Russian or only to the English text.

Texts Compared	Ellipsis Type			Total
	Stripping	Sluicing	Gapping	
<u>BK_RO vs BK_E1</u>				
Ellipsis only in Russian	0	0	0	0
Ellipsis only in English	55	9	2	66
Ellipsis in Russian and English	18	9	0	27
<u>BK_RO vs BK_E2</u>				
Ellipsis only in Russian	2	1	0	3
Ellipsis only in English	37	6	2	45
Ellipsis in Russian and English	16	8	0	24

Table 9. Absolute frequency of matched ellipsis by type

Table 9 makes evident that all 27 ellipses in Russian BK_RO were matched by an ellipsis in English BK_E1, i.e. none occurred only in Russian. When comparing Russian BK_RO to English BK_E1, only 3 of the 27 ellipses occurred only in Russian and the other 24 were matched by an ellipsis in English.

This suggests that the typological syntactic capacity is not the driving factor behind the use of ellipsis. If it were, one should see a lot more Russian ellipsis that English is unable to match (such as example (87) discussed above). Instead, English not only matches, but *exceeds* Russian in terms of antecedent-based ellipsis.

This result deserves further annotation of the type described above in other types of texts, particularly with English as the source language. Nevertheless, the present contrastive analysis

of the 121,062-word BK sub-corpus reveals that in actual use, English is able to successfully match 94% of Russian ellipses despite typological syntactic differences.

1.2.4 Non-elliptical Dominate

Finally, the results reveal that verbless utterances are predominantly *non-elliptical* in both Russian and English. *Figure 26* illustrates the frequency of elliptical and non-elliptical utterances normalized per 100 verbless utterances (the absolute values are provided in *Table 7* above).

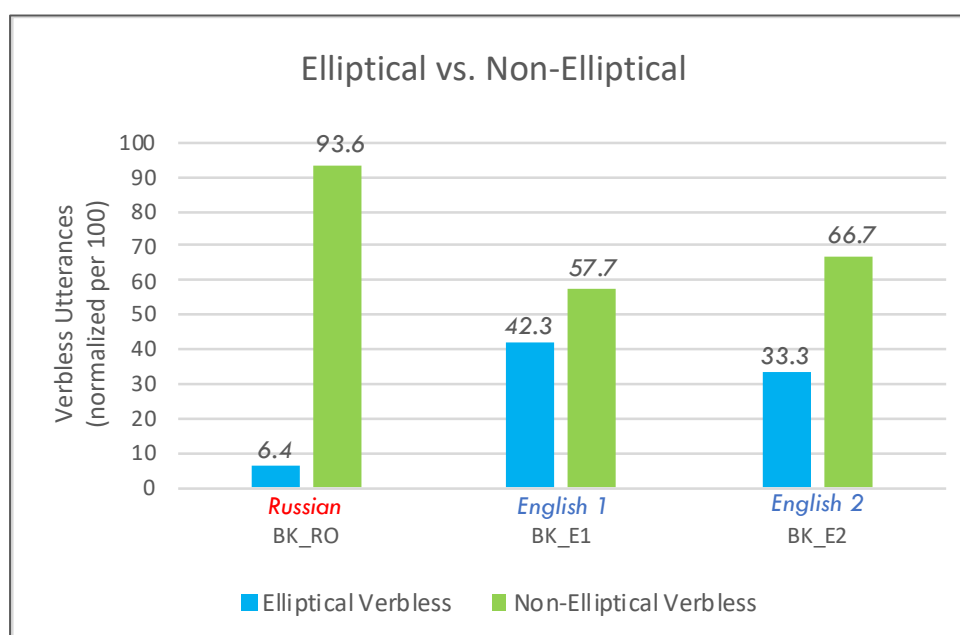


Figure 26. Relative frequency of verbless utterances by type

Non-elliptical utterances represent 94% of the Russian verbless utterances, and an average of 62% of the English verbless utterances. This finding makes clear that it is the non-elliptical type of verbless utterance that dominates in both languages.

Unlike the elliptical type of verbless utterance, the non-elliptical type lacks an explicit antecedent in the linguistic context. As argued in *Part 1*, the semantic properties of the verb, including tense, aspect, as well as a verbal lexical meaning, are also not marked in any way in the non-elliptical type (see *Part 1: Section 3.3 & Section 4.6*, *Part 2: Section 3.4*, and our discussion in Bondarenko & Celle, 2020: 327). In other words, this type of verbless utterance

is entirely semantically undetermined for verbal properties. The finding that it is particularly the non-elliptical type that dominates in both languages indicates that the study of the verbless phenomenon demands particular attention in terms of semantic and pragmatic analysis.

1.2.5 Summary: The Limits of Syntax

These results cast doubt on the extent to which syntactic ellipsis may explain the phenomenon of the verbless sentence. Special attention is required in terms of semantic and pragmatic analysis in order to account for the dominance of the non-elliptical type, and thus the verbless phenomenon in general. The observed frequency differences between the two languages, i.e. the over-representation of ellipsis in English as opposed to Russian, also require an explanation that goes beyond the syntactic productivity of ellipsis.

The findings in this chapter fueled the theoretical argument that runs throughout the present thesis in defense of the independent sentential status of verbless structures and the simultaneous rejection of the elliptical explanations.

The wide variation in the use of verbless sentences and ellipsis between Russian and English, and the unexpected dominance of ellipsis in the latter, suggest that syntax is not the place to look to explain neither the verbless phenomenon, nor sentential status in general. Syntactic explanations lead us into the situation that an antecedent-based elliptical verbless structure would be a sentence in English, but its semantico-pragmatically implicated translation correspondence in Russian would not because it does not involve an antecedent and the costly deletion process. Alternatively, both types of verbless sentences may be called nonsentential, but this does not resolve the issue. One is still left to explain where English verbal sentences disappear to in comparison to Russian verbless correspondences. A definition that yields such overwhelmingly unstable results cross-linguistically does not seem to be appropriate for a general linguistic definition of a sentence.

Furthermore, the idea that a verbless sentence is a reduction of a verbal sentence from which a verbal antecedent has been syntactically deleted, as well as the idea that a verbal antecedent can be semantically restructured back into the verbless sentence, run counter to the present empirical observations that show that most verbless sentences do not have a verbal antecedent at all (i.e. 94% of the Russian and 62% of the English verbless utterances in the present contrastive analysis). At minimum, the hidden structure argument and the semantic

ellipsis argument would be entirely inapplicable to the majority of verbless utterances. Further arguments for rejecting the null element or deletion syntactic explanation, as well as the semantic ellipsis explanation, of the verbless sentence are provided in *Part 1*.

Part 1: Section 4.5.3 also proposed that the way that a zero-predicate could perhaps help to explain the elliptical type of sentence, i.e. the type that does potentially involve an underlying predicate that is indicated by the antecedent. Nevertheless, based on the present results, most verbless sentences appear to be independent, at least syntactically from an existing predicate.

CHAPTER 2

SEMANTIC FEATURES

2.1 STEPS AND CLASSIFICATION SCHEMA

The present chapter analyzes the distinguishing semantic features of verbless sentences. It presents the results of a semantic classification of the 18,935 verbless sentences (6,581 in English and 12,354 in Russian) in the present corpus in terms of the key lexical constituents that statistically distinguish them from verbal sentences.

The statistically key verbless sentence constituents, including words and n-grams, were determined against the reference of 139,528 sentences (70,598 in English and 68,930 in Russian) that make up the full corpus. In terms of words, this corresponds to 55,129 words of the verbless sentences (17,107 words in English and 38,022 words in Russian) compared against the reference of 1,388,422 words (806,875 words in English and 581,547 words in Russian) that make up the full corpus.

To be clear, the key constituents are more than words and n-grams that are frequently found in verbless sentences. These are words and n-grams that are more frequently found in verbless sentences than they are in verbal sentences. This type of analysis would not be possible with an extracted series of verbless sentences, but requires a reference corpus. This means that when, for instance, the indefinite article ‘*a*’ is identified as being key for verbless sentences, it is not just that it is frequently found in verbless sentences and may also just as frequently be found in verbal sentences. Rather, ‘key-ness’ in the present case indicates that it is key for

verbless sentences and is not key for verbal sentences. In other words, it is the statistical notion of key-ness that is presently involved.

The hypergeometric model used for this calculation is described in (Lebart et al., 1998: 130–136; Lebart & Salem, 1994). It is called the specificity calculation and it identifies the statistically key elements that occur more frequently in a selected part of the text as compared to the text in its entirety. This calculation takes into account four variables:

- the size of the corpus,
- the size of the selected text part (i.e. in the present case, all of the verbless sentences),
- the frequency of the element in the corpus,
- the frequency of the element in the selected text part (i.e. in the verbless sentences).

The calculation is performed for each element of the corpus.

An ‘element’ can correspond to the level of the word form (e.g. ‘birds’ or ‘птицы’), of the lemma (e.g. ‘bird’ or ‘птица’), or of the grammatical category (e.g. ‘NNS’ or ‘Ncfnny’). Consequently, the values for the four variables above are determined on the appropriate level depending on the element. As a result of the calculation, each element acquires its own *index of specificity* relative to the other elements of the corpus for the particular selected text part (i.e. in the present case, the verbless sentences).

The specificity calculation is part of the Trameur software package. However, as discussed in *Part 2: Section 4.6*, due to the large number of sentence segments in the present corpus it was not possible to perform it using the function in the software. A special Python script (Borissov et al., 2020) was written and tested to identically match the characteristic elements calculation performed by Trameur.

Therefore, at present, the specificity calculation was performed for the statistically characteristic elements of *verbless sentences* on the level of *forms*, *lemmas* and *grammatical categories*. It was also performed for the *context of the verbless sentences* (i.e. three sentences to the left of the verbless sentence and three sentences to the right).

Not only elements, but also statistically key repeated segments (i.e. *n-grams*) within the verbless sentences were identified. The same hypergeometric model, involving the four variables above, was applied, using a specially developed Python script (Borissov et al., 2020), to analyze multiple-element segments on the level of form (e.g. ‘Good night’ or ‘Спокойной ночи’), lemma (e.g. ‘good night’ or ‘спокойный ночь’), and grammatical category (e.g. ‘JJ NN’ or ‘Afpfsgf Ncfsgn’). As in the case of characteristic elements, a specificity index revealed the n-grams that are statistically key for verbless sentences in comparison to the rest of the corpus.

The statistically key elements were then manually analyzed according to the semantic classification schema in *Table 10*.

SEMANTICO-PRAGMATIC FEATURE	LEXICAL ELEMENTS	
	<i>English</i>	<i>Russian</i>
Immediacy & Deixis	here, this, that, there, those, my, your, yesterday, now, before	тут (<i>tut</i> , 'here'), там (<i>tam</i> , 'there'), это (<i>èto</i> , 'this'), вот (<i>vot</i> , 'this'), здесь (<i>zdes'</i> , 'here'), вчера (<i>včera</i> , 'yesterday'), сейчас (<i>sejčas</i> , 'now'), раньше (<i>ran'she</i> , 'before'), моя (<i>moja</i> , 'my'), её (<i>eë</i> , 'her')
Person Animacy vs. Inanimacy	he, she, you vs. it	он (<i>on</i> , 'he'), она (<i>ona</i> , 'she'), ты (<i>ty</i> , 'you') vs. оно (<i>ono</i> , 'it')
Formality	Mr., Mrs., professional titles	ты (<i>ty</i> , 'you.SG') vs. вы (<i>vy</i> , 'you.PL')
Expressiveness	oh, hey, interjections, swearing	ой (<i>oj</i> , 'oh'), ах (<i>ax</i> , 'ah'), бух (<i>bux</i> , 'bang')
Emphasis Of Intensity vs. Of Contrast	very, quite, even, really, such, indeed, suddenly vs. only, most, more, less, precisely	очень (<i>očen'</i> , 'very'), внезапно (<i>vnezapno</i> , 'suddenly'), уж (<i>už</i> , 'really.PART'), же (<i>že</i> , 'really.PART'), ведь (<i>ved'</i> , 'really.PART'), vs. только (<i>tol'ko</i> , 'only'), больше (<i>bol'she</i> , 'more'), меньше (<i>men'she</i> , 'less')
Referent Specificity: Definite vs. Indefinite	the, Zosima vs. a, an	proper nouns
Quantification	fourteen, all, some, both, no, nothing, not, never, again	пять (<i>pjat'</i> , 'five'), всегда (<i>vsegda</i> , 'always'), оба (<i>oba</i> , 'both'), нет (<i>net</i> , 'no'), не (<i>ne</i> , 'not')
Reasoning & Comparison	well, else, perhaps, if, so, because, than, rather	ну (<i>nu</i> , 'well'), если (<i>esli</i> , 'if'), потому (<i>potomu</i> , 'because'), чем (<i>čem</i> , 'than')
Questioning	what, why, how, when, where, who, whom	почему (<i>počemu</i> , 'why'), где (<i>gde</i> , 'where'), кто (<i>kto</i> , 'who'), когда (<i>kogda</i> , 'when'), ли (<i>li</i> , 'eh.PART')
Characterization	far, kind, old, great, little, good, adjectives	далеко (<i>daleko</i> , 'far'), большой (<i>bol'soj</i> , 'big'), старая (<i>staraja</i> , 'old'), маленькое (<i>malen'koe</i> , 'small')
Content	love, table, birds	любовь (<i>ljubov'</i> , 'love'), стол (<i>stol</i> , 'table'), птицы (<i>pticy</i> , 'birds')

Table 10. Semantic classification schema

The schema was developed for the present analysis. It targets specific semantico-pragmatic features through the lexical elements by way of which they are likely to be expressed. The table lists the semantico-pragmatic features explored and provides examples of the lexical elements which are likely to indicate the presence of the particular feature.

The classification approach to semantic analysis presently applied is inspired by that which we used in (Ferguth et al., 2015), which also starts from the statistical specificities.

The present analysis proceeded as follows. The key elements were first revealed by the specificity calculation. They were then searched within the verbless sentences and examined in context to verify the precise meaning(s) with which the lexical element is used. The source of the particular elements was also verified to make sure that it does not come from a single text. The elements were subsequently manually classified for the features in the table and compared between the languages.

The chapter ends with results concerning verbless sentence length and their lexical complexity.

2.2 A LOOK INSIDE THE VERBLESS SENTENCE

A look at what precisely is happening inside the ‘prototypical’ verbless sentence is found in *Appendix 6*. The appendix reveals the lexical and grammatical elements that characterize verbless sentences in comparison to other sentences. More specifically, it reveals the first 30 lines of the characteristic elements results concerning the forms, lemmas and morphosyntactic categories that are statistically key for verbless structures. It also provides details about the calculation.

The present section discusses the findings of the semantic classification. Starting with similarities, it then identifies what appear to be different uses of verbless sentences in Russian and English.

2.2.1 Agreement and Disagreement

Analysis of the verbless sentences according to their statistically characteristic elements reveals a few features that appear to be equally shared by the verbless sentences of both languages. Of the similarities, the leading feature is the expression of agreement and disagreement. In English, it is signaled by statistically key lemmas ‘yes’, ‘no’, ‘right’, ‘yeah’, ‘okay’, ‘fine’, ‘aha’, ‘sure’ and the n-grams⁴⁴ ‘all right’, ‘of course’. In Russian, by the lemmas ‘нет’ (*net*), ‘да’ (*da*), ‘ладно’ (*ladno*), ‘правда’ (*pravda*), ‘хорошо’ (*xorošo*), ‘так’ (*tak*), ‘конечно’ (*konečno*), ‘верно’ (*verno*), ‘понятно’ (*ponjatno*), ‘ага’ (*aga*), ‘неправда’ (*nepravda*), ‘здорово’ (*zdorovo*), ‘то-то’ (*to-to*), ‘ну-ну’ (*nu-nu*), ‘согласный’ (*soglasnyj*), and the n-grams ‘ещё бы’ (*eščë by*), ‘да нет’ (*da net*). Such elements highlight the strong correlation of verbless sentences to interaction contexts.

The specific values for the mentioned elements, and additional elements, are listed in *Table 11*. The values in brackets indicate the following, in order: index of specificity, number of occurrences in the overall corpus, number of occurrences in the verbless sentence subpart. As discussed in *Appendix 6*, the **Specificity Index (ind.)** indicates the strength with which the element or n-gram is associated with the verbless sentence; a negative value indicates that the element or n-gram is rejected by verbless sentences.

A few examples in (101) illustrate a range of agreement and disagreement uses of verbless sentences. Cases (a), (b) and (c) illustrate basic instances of disagreement with Russian ‘нет’ (*net*; lit. ‘no’) and agreement with English ‘yes’ and Russian ‘так’ (*tak*; lit. ‘sure’).

Example (d) shows an emphatic use of English ‘no’, translated with Russian verbless sentence using the emphatic n-gram ‘да нет’ (*da net*; lit. ‘yes no’).

In (e), English ‘oh no’ is used to reject the content of the previous utterance and express disbelief; it is translated in Russian by the verbless ‘что вы’ (*čto vy*; lit. ‘what you’) which also expresses disbelief but without explicit negation.

⁴⁴ The key n-grams within verbless sentences are also statistically specific to the verbless sentences, i.e. computed against a reference corpus. They will be discussed in *Part 4: Section 2.3* and *Appendix 7*.

English	Russian
'yes' (ind. infinity; 653; 365)	'нет' (<i>net</i> ; lit. 'no'; ind. infinity; 1114; 553)
'no' (ind. 225; 2062; 377)	'да' (<i>da</i> ; lit. 'yes'; ind. 314; 1584; 628)
'right' (ind. 71; 1315; 162)	'ладно' (<i>ladno</i> ; lit. 'all_right'; ind. 54; 147; 79)
'yeah' (ind. 57; 127; 55)	'правда' (<i>pravda</i> ; lit. 'true'; ind. 293; 293; 90)
'okay' (ind. 21; 82; 24)	'хорошо' (<i>xorošo</i> ; lit. 'ok'; ind. 455; 455; 112)
'fine' (ind. 21; 213; 35)	'так' (<i>tak</i> ; lit. 'sure'; ind. 31; 2246; 299)
'aha' (ind. 21; 12; 11)	'конечно' (<i>konečno</i> ; lit. 'of course'; ind. 29; 198; 66)
'amen' (ind. 18; 10; 10)	'верно' (<i>verno</i> ; lit. 'correct'; ind. 26; 74; 38)
'uh-huh' (ind. 14; 8; 8)	'понятно' (<i>ponjatno</i> ; lit. 'understandable'; ind. 24; 65; 34)
'sure' (ind. 11; 310; 28)	'ага' (<i>aga</i> ; lit. 'yeah'; ind. 21; 29; 22)
	'неправда' (<i>nepravda</i> ; lit. 'not true'; ind. 19; 29; 21)
'all right' (ind. 112; 385; 97)	'то-то' (<i>to-to</i> ; lit. 'yeah-yeah'; ind. 14; 28; 17)
'oh yes' (ind. 45; 40; 26)	'ну-ну' (<i>nu-nu</i> ; lit. 'uh-huh'; ind. 13; 12; 11)
'of course' (ind. 37; 303; 41)	'согласный' (<i>soglasnyj</i> ; lit. 'in_agreement'; ind. 10; 32; 15)
'oh no' (25; 28; 15)	
'oh yeah' (27; 17; 13)	'ещё бы' (<i>ešče by</i> ; lit. 'indeed'; ind. 27; 30; 22)
'no no' (25; 36; 16)	'да нет' (<i>da net</i> ; lit. 'nope'; ind. 22; 29; 19)
	'да да' (<i>da da</i> ; lit. 'yes yes'; ind. 17; 24; 15)
	'ну да' (<i>nu da</i> ; lit. 'well yes'; ind. 17; 31; 16)
	'да конечно' (<i>da konečno</i> ; lit. 'yes of course'; ind. 15 ; 17; 12)

Table 11. Agreement and Disagreement

(101) *Agreement and Disagreement Uses*

a. Russian (FH R1: 2719902)

[– Да и кто в наше время будучи в здравом уме, захочет иметь детей? – воскликнула миссис Фелпс, не понимая, почему так раздражает её этот человек.

– Нет, тут я с вами не согласна. – промолвила миссис Бауэлс. – У меня двое. Мне, разумеется, оба раза делали кесарево сечение. Не терпеть же мне родовые муки из-за какого-то там ребенка?]

Нет,	тут	я	с	вами	не	согласна.
<i>net,</i>	<i>tut</i>	<i>ja</i>	<i>s</i>	<i>vami</i>	<i>ne</i>	<i>soglasna</i>
NEG	ADV	PRO	PREP	PRO.INS	NEG	ADJS.F
<i>no</i>	<i>here</i>	<i>I</i>	<i>with</i>	<i>you</i>	<i>not</i>	<i>in_agreement</i>

English (FH EO: 1096708)

["No one in his right mind, the Good Lord knows; would have children!" said Mrs. Phelps, not quite sure why she was angry with this man.

"I wouldn't say that," said Mrs. Bowles. "I've had two children by Caesarian section. No use going through all that for a baby.]

I wouldn't say that.

b. English (TD E1: 1794820)

["So you're off to war, soldiers?"
"Yes, Grandad, off to war."]

Yes, Grandad, off to war.

Russian (TD RO: 253620)

[– На войну, стало быть, служивые?
– На войну, дедушка.]

На	войну,	дедушка.
<i>na</i>	<i>vojnu,</i>	<i>deduška</i>
PREP	NN.ACC	NN
<i>to</i>	<i>war</i>	<i>grandad</i>

c. Russian (BK RO: 18707)

[– Да там нет крючьев, – тихо и серьезно приглядываясь к отцу, выговорил Алёша.
– Так, так, одни только тени крючьев. Знаю, знаю.]

Так,	так,	одни	только	тени	крючьев.
<i>tak,</i>	<i>tak,</i>	<i>odni</i>	<i>tol'ko</i>	<i>teni</i>	<i>krjuč'ev</i>
ADV	ADV	PRO	ADV	NN.NOM	NN.GEN
<i>yes,</i>	<i>yes,</i>	<i>only</i>	<i>exclusively</i>	<i>shadows</i>	<i>of_hooks</i>

English (BK E1: 1494711)

["No, there are no hooks there," Alyosha said quietly and seriously, studying his father.
"Yes, yes. Only shadows of hooks. I know I know.]

Yes, yes.

English (BK E2: 3060207)

['There aren't any hooks there!' said Alyosha softly and seriously, gazing at his father.
'Well, well, only shadows of hooks. I know, I know.]

Well, well, only shadows of hooks.

d. English (HP EO: 1223670)

["I suppose you've been having a real laugh, haven't you, all holed up here together –"
"No, honest –"
"Harry, we're really sorry!"]

No, honest –

Russian (HP R1: 2834592)

[– Я думаю, вы от души надо мной посмеялись в этом уютном гнездышке...
– Да нет же, честно.
– Гарри, нам действительно очень жаль.]

Да	нет	же,	честно.
<i>da</i>	<i>net</i>	<i>že,</i>	<i>čestno</i>
PART	ADV	PART	ADV
<i>yes</i>	<i>no</i>	<i>really</i>	<i>honest</i>

e. English (SR EO: 912665)

["I was hoping he would knock down a waiter," Mike said, "and get arrested. I'd like to see Mr. Robert Cohn in jail."
"No," I said.
"Oh, no," said Edna. "You don't mean that."
"I do, though," Mike said.]

Oh, no,

Russian (SR R1: 2561994)

[– Я все ждал что он ударит официанта, – сказал Майкл, – и его арестуют. Очень был бы рад, если бы мистера Роберта Кона засадили в тюрьму.
– Ну вот ещё, – сказал я.
– Что вы, – сказала Эдна. – Вы шутите?
– Нет, не шучу. – сказал Майкл.]

Что	вы,
<i>čto</i>	<i>vy</i>
PRO	PRO.2PL
<i>what</i>	<i>you</i>

Notably, 'agreement and disagreement' was not the only use of the lemmas in the table. To precise the proportional use of each element, further annotation is required. For the time being, the verbless sentences containing the specific elements were found and qualitatively

examined in context for their use in agreement and disagreement situations. (They were found and examined using the contrastive program discussed in *Appendix 5*.)

The verbless sentences containing the Russian lemma ‘нет’ (*net*; lit. ‘no’) were manually annotated in terms of whether they belong to instances of disagreement or quantification. The latter is illustrated in (102).

- (102) Там миру нет. (BK_RO: 68150)
tam miru net
 ADV NN.GEN NEG
there peace none
- There is no peace there. (BK_E1: 1555849)
 There is no peace between them. (BK_E2: 3122732)

The annotation results in *Table 12*, illustrate that the verbless sentences with this lemma are nearly split between the disagreement use (56%) and the non-disagreement use (43%). (The texts SR_R1 and HP_R1, marked NA, were not annotated for the distinction.)

	Verbless sentences containing lemma 'нет'		Disagreement		Quantification	
	(Abs. Fq.)	(Abs. Fq.)	(Abs. Fq.)	(Abs. Fq.)	(% per text)	(% per text)
BK_RO	16	7	9	44	56	
TD_RO	36	28	8	78	22	
OD_RO	18	4	14	22	78	
PO_RO	48	33	15	69	31	
KF_RO	17	10	7	59	41	
OS_RO	36	20	16	56	44	
ST_R1	24	17	7	71	29	
ST_R2	13	11	2	85	15	
SR_R1	116	NA	NA	NA	NA	
AF_R1	8	5	3	63	38	
AF_R2	5	3	2	60	40	
AF_R3	6	3	3	50	50	
FH_R1	59	34	25	58	42	
JS_R1	7	1	6	14	86	
HP_R1	81	NA	NA	NA	NA	
CA_R1	40	25	14	63	35	
		201	131	56	43	
	530		332			

Table 12. Disagreement annotation: Russian lemma ‘нет’ (*net*) within verbless sentences

Therefore, based on the statistical significance of this element, as well as the other statistically key elements and n-grams from *Table 11*, it seems safe to say that the *agreement and disagreement use* is an important function of verbless sentences in both languages.

2.2.2 Expressiveness

Another feature that is found to be equally important for Russian verbless sentences as it is for English is expressiveness. It is signaled in English by the statistically key presence of interjections, insults and praise. *Table 13*, below, indicates the lemma elements, lemma n-grams and grammatical categories, which were statistically identified as being key for verbless sentences, and verified in context to suggest this semantic function.

Example (103) illustrates a few cases of expressiveness. This includes both *negative* insults, such as (a), and *positive* praise, such as (b) and (c), which the speaker expresses toward an interlocutor. Example (d) illustrates an expressive characterizations of the just mentioned *situation*. Example (e) shows an expressive *sarcastic* characterization of an idea under discussion.

(103) ***Expressiveness***

a. English (OD E1: 1983079)

Bastard, crock, shit-head, no-good sonofabich!

English (OD E2: 3273898)

Filthy swine! Traitor! Rat! Dirty dog!

Russian (OD RO: 394893)

Чу-ма-а!	Шко-одник!	Шушера!	Сука	позорная!
<i>Ču-ma-a!</i>	<i>Ško-odnik!</i>	<i>Šušera!</i>	<i>Suka</i>	<i>pozornaja!</i>
NN.NOM	NN.NOM	NN.NOM	NN.NOM	ADJ.NOM
<i>Pla-agu-ee!</i>	<i>Troo-ublemaker!</i>	<i>Human_trash!</i>	<i>Bitch</i>	<i>shameful!</i>

b. English (TD E1: 1629421)

[The simple, sincere, slightly embarrassed gaze seemed to be saying: “Here am I all, as I am. Judge me as you wish.” “Splendid!” Grigory replied with his eyes and smile.]

Splendid!

Russian (TD RO: 125083)

[Бесхитростный, чуть смущенный, правдивый взгляд словно говорил, «Вот я вся, какая есть. Как хочешь, так и суди меня». – «Славная», – ответил Григорий глазами и улыбкой.]

Славная
slavnaja
ADJ.FSG
glorious

c. English (TD E1: 617676)

“Grisha ... my dearest ... beloved ... let’s go away. My darling!”

Russian (TD RO: 115766)

– Гриша,	дружечка	моя ...	родимый ...	давай	уйдём.
<i>Griša,</i>	<i>družečka</i>	<i>moja ...</i>	<i>rodimyj ...</i>	<i>davaj</i>	<i>ujdëm</i>
NN	NN.NOM	PRO.NOM	ADJ.NOM	V.IMP	V.FT
<i>Grisha,</i>	<i>friend</i>	<i>my ...</i>	<i>darling ...</i>	<i>let’s</i>	<i>leave</i>

d. English (SR EO: 851165)

[Bill was awake and sitting on the edge of the bed. “I saw you out the window,” he said. “Didn’t want to interrupt you. What were you doing? Burying your money?”

“You lazy bum!”

“Been working for the common good? Splendid. I want you to do that every morning.”

“Come on,” I said. “Get up.”]

Splendid.

Russian (SR R1: 2511222)

[– Я тебя видел в окно, – сказал он. – Не хотел мешать тебе. Что ты делал? Зарывал свои деньги?

– Ах ты лентяй!

– Трудился для общего блага? Чудесно! Продолжай в том же духе каждое утро.

– Ну, довольно валяться, – сказал я. – Вставай.]

Чудесно!
čudesno
ADJS.N
splendid

e. English (BK E2: 3106284)

[‘Well, sir, I must admit you have rather gladdened my heart,’ Miusov said with a smile, recrossing his legs. ‘As I understand it, it is a question of realizing some kind of ideal in the infinitely distant future, at the Second Coming perhaps. I suppose you know best. A splendid utopian dream of no more wars, diplomats, banks, and so forth. Sounds remarkably like socialism.]

A splendid utopian dream of no more wars, diplomats, banks, and so forth.

English (BK E1: 1539842)

[“Well, sir, I confess that you have now reassured me somewhat,” Miusov grinned, recrossing his legs again. “So far as I understand it, this, then, would be the realization of some ideal, an infinitely remote one, at the Second Coming. That is as you please. A beautiful utopian dream of the disappearance of wars, diplomats, banks, and so on. Something even resembling socialism.]

A beautiful utopian dream of the disappearance of wars, diplomats banks, and so on.

Russian (BK RO: 54958)

[– Ну-с, признаюсь, вы меня теперь несколько ободрили, – усмехнулся Миусов, переложив ногу на ногу. – Сколько я понимаю, это, стало быть, осуществление какого-то идеала, бесконечно далекого, во втором пришествии. Это как угодно. Прекрасная утопическая мечта об исчезновении вой, дипломатов, банков и проч. Что-то даже похоже на социализм.]

Прекрасная	утопическая	мечта	об	исчезновении
<i>prekrasnaja</i>	<i>utopičeskaja</i>	<i>mečta</i>	<i>ob</i>	<i>isčesznovenii</i>
ADJ.NOM	ADJ.NOM	NN.NOM	PREP	NN.PRE
<i>beautiful</i>	<i>utopian</i>	<i>dream</i>	<i>of</i>	<i>disappearance</i>

войн,	дипломатов,	банков	и	проч.
<i>vojn</i>	<i>diplomatov</i>	<i>bankov</i>	<i>i</i>	<i>proč</i>
NN.GEN	NN.GEN	NN.GEN	CONJ	PRO.GEN
<i>wars</i>	<i>diplomats</i>	<i>banks</i>	<i>and</i>	<i>so_on</i>

English	Russian
oh (ind. 162; 450; 171)	ах (<i>ax</i> ; lit. 'ah'; ind. 35; 74; 45)
hey (ind. 86; 94; 65)	эй (èj; lit. 'hey'; ind. 30; 61; 38)
well (ind. 52; 1351; 139)	черт (<i>čert</i> ; lit. 'damn'; ind. 29; 153; 57)
eh (ind. 50; 80; 43)	Господи (<i>gospodi</i> ; lit. 'God'; ind. 25; 74; 37)
ah (ind. 48; 77; 41)	эх (èx; lit. 'ah'; ind. 22; 47; 28)
hell (ind. 35; 260; 52)	ага (<i>aga</i> ; lit. 'aha'; ind. 21; 29; 22)
God (ind. 32; 369; 58)	милый (<i>milyj</i> ; lit. 'darling'; ind. 21; 95; 38)
huh (ind. 25; 22; 17)	ой (<i>oj</i> ; lit. 'ah'; ind. 17; 34; 21)
darling (ind. 22; 45; 20)	э (è; lit. 'eh'; ind. 15; 49; 23)
bastard (ind. 19; 134; 27)	дорогой (<i>dorogoj</i> ; lit. 'dear'; ind. 13; 91; 29)
aha (ind. 18; 12; 11)	Бог (<i>bog</i> ; lit. 'God'; ind. 13; 207; 45)
amen (ind. 18; 10; 10)	дурак (<i>durak</i> ; lit. 'fool'; ind. 13; 102; 30)
er (ind. 16; 41; 15)	ох (<i>ox</i> ; lit. 'oh'; ind. 12; 54; 21)
bitch (ind. 15; 78; 19)	славный (<i>slavnij</i> ; lit. 'glorious'; ind. 12; 35; 17)
damned (ind. 14; 34; 13)	ужасно (<i>užasno</i> ; lit. 'terrible'; ind. 12; 45; 19)
tut (ind. 11; 13; 8)	молодец (<i>molodec</i> ; lit. 'excellent_person'; ind. 11; 23; 13)
shame (ind. 11; 71; 14)	проклятый (<i>prokljatyj</i> ; lit. 'damned'; ind. 10; 50; 18)
splendid (ind. 10; 25; 9)	хлоп (<i>xlop</i> ; lit. 'bang'; ind. 10; 9; 8)
swine (ind. 10; 26; 9)	цыц (<i>cyc</i> ; lit. 'shh'; ind. 9; 7; 7)
nonsense (ind. 10; 45; 11)	
	І (cat. interjections; ind. infinity; 1354; 804)
UH (cat. interjections; ind. infinity; 1299; 721)	к черт (<i>k čert</i> ; lit. 'to hell'; ind. 22; 34; 20)
	ах ты (<i>ax ty</i> ; lit. 'oh you'; ind. 18; 15; 13)
hell with (ind. 42; 41; 25)	слава Бог (<i>slava Bog</i> ; lit. 'thank God'; ind. 16; 19; 13)
my God (ind. 36; 29; 20)	
to hell with (ind. 30; 26; 17)	
my dear (ind. 29; 86; 23)	
son of (ind. 24; 70; 19)	
to hell (ind. 23; 53; 17)	
good luck (ind. 23; 12; 11)	
son of a (ind. 21; 44; 15)	
the hell (ind. 19; 90; 17)	
son of a bitch (ind. 17; 26; 11)	
of a bitch (ind. 17; 26; 11)	
a bitch (ind. 16; 34; 11)	
the hell with (ind. 14; 15; 8)	

Table 13. Expressiveness

2.2.3 Quantification

Another feature that is revealed to be common to the verbless sentences of both languages is quantification. This is expressed by the characteristic lemma and grammatical category elements, and lemma n-grams, in *Table 14*.

English	Russian
no (ind. 225; 2062; 377)	нет (<i>net</i> ; lit. 'no'; ind. infinity; 1114; 553)
not (ind. 42; 3554; 217)	ничто (<i>ničto</i> ; lit. 'nothing'; ind. 21; 813; 129)
four (ind. 18; 304; 38)	нету (<i>netu</i> ; lit. 'absent'; ind. 21; 55; 29)
one (ind. 18; 2673; 130)	четыре (<i>četyri</i> ; lit. 'four'; ind. 15; 232; 52)
five (ind. 17; 326; 38)	восемь (<i>vosem</i>); lit. 'eight'; ind. 13; 75; 26)
all (ind. 17; 3944; 167)	всё (<i>vsë</i> ; lit. 'all'; ind. 12; 2154; 223)
three (ind. 16; 583; 49)	третья (<i>tret'ja</i> ; lit. 'third'; ind. 10; 17; 11)
two (ind. 15; 1219; 73)	
nothing (ind. 15; 643; 50)	Mc--n (cat. numeral-cardinal-nominative; ind. 85; 2761; 480)
six (ind. 12; 138; 20)	
too (ind. 10; 946; 53)	Mofsn (cat. numeral-ordinal-feminine-singular-nominative; ind. 29; 152; 57)
ten (ind. 10; 242; 24)	
hundred (ind. 9; 240; 22)	Mc---d (cat. numeral-cardinal-digit; ind. 29; 205; 66)
fourteen (ind. 9; 24; 8)	Momsn (cat. numeral-ordinal-masculine-singular- nominative; ind. 13; 426; 71)
CD (cat. cardinal numbers; ind. 147; 6801; 539)	Monsn (cat. numeral-ordinal-neuter-singular-nominative; ind. 11; 49; 19)
one two (ind. 26; 25; 15)	
two three (ind. 22; 17; 12)	
one two three (ind. 22; 17; 12)	
first five (ind. 13; 7; 6)	
how many (ind. 12; 62; 11)	

Table 14. Quantification

The examples in (104), in addition to (102) above, illustrate verbless sentences with a quantificational use. In (a) the presence of trouble is being quantified as to whether or not it exists; in the English example this is done using a zero quantity for trouble, i.e. 'no trouble', while in Russian using a positive quantity for the opposite of trouble, i.e. 'всё спокойно' (*vsë spokojno*; lit. 'all calm').

In (b), two zero quantities are causatively linked through their juxtaposition inside a verbless sentence, i.e. the absence of a password results in the absence of entrance.

Example (c) illustrates a zero quantity is indicated with the typical possessive structure in Russian. In (d), the English 'too' is used to quantify through the repetition of an order.

Finally, (e) and (f) illustrate quantifying questions. In (e), the English verbless elliptical question 'How many copies of Shakespeare and Plato?' (i.e. antecedent 'are left') is translated with a verbal question in Russian. In (f), the Russian verbless question 'Сколько дней в дороге?' (*skol'ko dnej v doroge*) and its English verbless translation 'How many days' journey?' are both non-elliptical quantifying questions.

(104) *Quantificational*

a. English (HP EO: 1315150)

[A porter's cap pulled low over his mismatched eyes, Moody came limping through the archway pushing a trolley loaded with their trunks. "All OK," he muttered to Mrs. Weasley and Tonks, "don't think we were followed..." Seconds later, Mr. Weasley emerged on to the platform with Ron and Hermione. They had almost unloaded Moody's luggage trolley when Fred, George and Ginny turned up with Lupin. "No trouble?" growled Moody. "Nothing," said Lupin. "I'll still be reporting Sturgis to Dumbledore," said Moody, "that's the second time he's not turned up in a week."]

No trouble?

Russian (HP R1: 2911889)

[– Полный порядок, – вполголоса сказал он миссис Уизли и Тонкс. – Никакой слежки ... Они уже почти разгрузили тележку Грюма, когда пришел Люпин с Фредом, Джорджем и Джинни. – Всё спокойно? Прорычал Грюм. – Да, – ответил Люпин.]

Всё	спокойно?
<i>vsë</i>	<i>spokojno</i>
PRO	ADJS.N
<i>everything</i>	<i>calm</i>

b. English (HP EO: 1339735)

["Er ..."] he said glumly, staring up at the Fat Lady, who smoothed the folds of her pink satin dress and looked sternly back at him. "No password, no entrance," she said loftily.]

No password, no entrance.

Russian (HP R1: 2933034)

[– Э ... – с тоской выдавил он из себя, глядя на Полную Даму, которая сурово смотрела на него, разглаживая складки на розовом атласном платье. – Без пароля хода нет, – заявила она надменно.]

Без	пароля	хода	нет.
<i>bez</i>	<i>parolja</i>	<i>xoda</i>	<i>net</i>
PREP	NN.GEN	NN.GEN	NEG
<i>without</i>	<i>password</i>	<i>entrance</i>	<i>none</i>

c. Russian (KF RO: 656467)

[– Ну фотографии-то где?
 – Нет у меня ихних фотографий. У меня и своей нету. Так что помирать нам рановато. Пришли фотографа – пускай отхреначит. Тогда и в землю можно. Правильно говорю? Да и кому они нужны фотографии твои? Цветы растут в поле, иду – и сынков вспоминаю. Трава растет, как будто их волосы. Раз в неделю кошу тут, чтобы не заросли.]

У	меня	и	своей	нету.
<i>u</i>	<i>menja</i>	<i>i</i>	<i>svoej</i>	<i>netu</i>
PREP	PRO.GEN	PART	PRO	NEG
<i>at</i>	<i>me</i>	<i>and</i>	<i>my_own</i>	<i>none</i>

English (KF E1: 2309820)

[“Where are the photographs them?”
 “I don’t have any photographs of them. I don’t even have one of myself. So it’s early for us to die yet. Send a photographer – let him go nuts. Then I can go into the ground. Is that right? Who needs those photos of yours? Flowers grow in the field, I walk and remember my sons. The grass grows, as if it were their hair. Once a week I mow here, so it doesn’t get overgrown.”]

I don’t even have one of myself.

d. English (SR EO: 776777)

[“That’s not good for little girls.”
 “Little girl yourself. Dites garçon, un pernod.”
 “A pernod for me, too.”
 “What’s the matter?” she asked.]

A pernod for me, too.

Russian (SR R1: 2450725)

[– Маленьким девочкам вредно пить перно.
 – Сам маленький. Гарсон, рюмку перно.
 – И мне рюмку перно.
 – Ну как? – спросила она.]

И	мне	рюмку	перно.
<i>i</i>	<i>mne</i>	<i>rjumku</i>	<i>perno</i>
CONJ	PRO.DAT	NN.ACC	NN.GEN
<i>and</i>	<i>for_me</i>	<i>glass_of</i>	<i>perno</i>

e. English (FH EO: 1080475)

[“Professor Faber, I have a rather odd question to ask. How many copies of the Bible are left in this country?”

“I don’t know what you’re talking about!”

“I want to know if there are any copies left at all.”

“This is some sort of trap! I can’t talk to just anyone on the phone!”

“How many copies of Shakespeare and Plato?”

“None! You know as well as I do. None!”]

How many copies of Shakespeare and Plato?

Russian (FH R1: 2704822)

[– Профессор Фабер, у меня к вам не совсем обычный вопрос. Сколько экземпляров Библии осталось в нашей стране?

– Я хочу знать, остался ли у нас хоть один экземпляр Библии?

– Это какая-то ловушка!

– Я не могу со всякими разговаривать по телефону.

– Сколько осталось экземпляров произведений Шекспира, Платона?

– Ни одного! Вы знаете это не хуже меня. Ни одного!

Сколько	осталось	экземпляров	произведений
<i>skol'ko</i>	<i>ostalos'</i>	<i>èkzempljarov</i>	<i>proizvedenij</i>
PRO	V.PS	NN.GEN.PL	NN.GEN.PL
<i>how_many</i>	<i>left</i>	<i>of_copies</i>	<i>of_works</i>

Шекспира,	Платона?
<i>Šekspira,</i>	<i>Platona</i>
NN.GEN	NN.GEN
<i>of_Shakespeare,</i>	<i>of_Plato</i>

f. English (OS E1: 2313969)

[– Just stoke up the hot-house, right?

– Stoke it up hot.

– How about that doctor? It doesn’t matter if there’s no water, does it?

– No-oo.

– How many days’ journey?

– Over a week.]

How many days’ journey?

- [– Прожарку топи, прожарку.
 – Как, доктор, нечего, что без воды?
 – Ничего-о.
 – Сколько дней в дороге?
 – Вторую неделю.]

Сколько	дней	в	дороге?
<i>skol'ko</i>	<i>dnej</i>	<i>v</i>	<i>doroge</i>
PRO	NN.GEN	PREP	NN.NOM
<i>how_many</i>	<i>days</i>	<i>on</i>	<i>road</i>

2.2.4 Formality and Informality

Verbless sentences are found in both formal and informal situations. Although, there is a suggestion in the Russian data that a slight preference for informality exists. This is shown by the more significant index of the second person singular pronoun ‘ТЫ’ (*ty*; lit. ‘*you.2SG*’; ind. 130; 3595; 665) as compared to the second person plural pronoun ‘ВЫ’ (*vy*; lit. ‘*you.2PL*’; ind. 45; 2755; 388). It is also suggested by the slightly higher indices of casual greetings, e.g. English ‘*hi*’ and Russian ‘привет’ (*privet*; lit. ‘*hi*’), forms of address, e.g. English ‘*darling*’ and Russian ‘милый’ (*milyj*; lit. ‘*darling*’), and agreement, e.g. English ‘*yeah*’ and Russian ‘ага’ (*aga*; lit. ‘*yeah*’), as compared to the more formal greetings, forms of address and agreement, e.g. Russian ‘здравствуйте’ (*zdravstvujte*; lit. ‘*hello.2PL*’), English ‘*sir*’, Russian ‘гражданин начальник’ (*graždanin načalnik*; lit. ‘*citizen supervisor*’). Table 15 presents the results showing the key lemma elements and n-grams according to formality and informality uses and their index of specificity to the verbless sentence.

Therefore, it seems important to emphasize that verbless sentences are not entirely of the informal register, as has sometimes been assumed. Simultaneously, based on the present specificity results of their lexical elements, there does appear to be a preference for informality. Further annotation of the situation is required in order to assess the formality proportions and whether the types of verbless sentences differ in formal versus informal contexts.

English	Russian
<i>Informal</i>	<i>Informal</i>
hey (ind. 86; 94; 65)	ты (<i>ty</i> ; lit. 'you.2SG'; ind. 130; 3595; 665)
yeah (ind. 57; 127; 55)	привет (<i>privet</i> ; lit. 'hi'; ind. 32; 42; 33)
hi (ind.31; 23; 20)	эй (<i>èj</i> ; lit. 'hey'; ind. 30; 61; 38)
huh (ind. 25; 22; 17)	твой (<i>tvoj</i> ; lit. 'your.2SG'; ind. 28; 371; 91)
darling (ind. 22; 45; 20)	здравствуй (<i>zdravstvuj</i> ; lit. 'hello.2SG'; ind. 26; 21; 21)
aha (ind. 18; 12; 11)	ага (<i>aga</i> ; lit. 'aha'; ind. 21; 29; 22)
uh-huh (ind. 14; 8; 8)	милый (<i>milyj</i> ; lit. 'darling'; ind. 21; 95; 38)
tut (ind. 11; 13; 8)	здорово (<i>zdorovo</i> ; lit. 'greetings'; ind.15; 58; 25)
hah (ind. 11; 9; 7)	то-то (<i>to-to</i> ; lit. 'yeah-yeah'; ind. 14; 28; 17)
uh (ind. 10; 12; 7)	ребята (<i>rebjata</i> ; lit. 'guys'; ind. 11; 88; 25)
to hell with (ind. 30; 26; 17)	
son of a (ind. 21; 44; 15)	<i>Formal</i>
hello you (ind. 14; 6; 6)	вы (<i>vy</i> ; lit. 'you.2PL'; ind. 45; 2755; 388)
	здравствуйте (<i>zdravstvujte</i> ; lit. 'hello.2PL'; ind. 27; 26; 24)
<i>Formal</i>	ваш (<i>vaš</i> ; lit. 'your.2PL'; ind. 27; 429; 96)
sir (ind. 22 ; 117; 28)	товарищ (<i>tovarišč</i> ; lit. 'comrade'; ind. 22; 302; 72)
comrade (ind. 16; 157; 26)	гражданин (<i>graždanin</i> ; lit. 'citizen'; ind.21; 51; 28)
citizen (ind. 14; 29; 12)	начальник (<i>načalnik</i> ; lit. 'supervisor'; ind. 15; 110; 34)
Mr (ind. 10; 290; 26)	господин (<i>gospodin</i> ; lit. 'sir'; ind. 13; 158; 38)
Honour (ind. 9; 23; 8)	гражданин начальник (<i>graždanin načalnik</i> ; lit. 'citizen supervisor'; ind. 25; 43; 24)
your Honour (ind. 12; 22; 8)	

Table 15. Formality and Informality

2.2.5 Immediacy and Deixis

Another function that characterizes verbless sentences more than it does verbal sentences is the association of the former with marking deixis and immediacy. Although the verbless sentences of both languages show a correlation with deictic and indexical markers, this correlation is more pronounced for Russian verbless sentences than it is for English.

Table 16 presents the specificity results concerning markers of immediacy and deixis in terms of lemma elements and lemma n-grams. Such markers tend to associate verbless sentences with the utterance situation.

English	Russian
your (ind. 16; 1999; 102)	это (<i>èto</i> ; lit. 'this/that'; ind. 222; 3612; 827)
my (ind. 15; 3222; 141)	ты (<i>ty</i> ; lit. 'you.2SG'; ind. 130; 3595; 665)
here (ind. 14; 1983; 98)	вот (<i>vot</i> ; lit. 'here/that'; ind. 122; 1226; 346)
	вы (<i>vy</i> ; lit. 'you.2PL'; ind. 45; 2755; 388)
hey you (ind. 31; 19; 16)	мой (<i>moj</i> ; lit. 'my'; ind. 41; 1204; 216)
my dear (ind. 29; 86; 23)	так (<i>tak</i> ; lit. 'this_way'; ind. 31; 2246; 299)
what about you (ind. 22; 13; 11)	здесь (<i>zdes'</i> ; lit. 'here_place'; ind. 30; 541; 116)
about you (ind. 19; 48; 14)	твой (<i>tvoj</i> ; lit. 'yours.2SG'; ind. 28; 371; 91)
this way (ind. 17; 38; 12)	ваш (<i>vaš</i> ; lit. 'yours.2PL'; ind. 27; 429; 96)
hello you (ind. 14; 6; 6)	вон (<i>von</i> ; lit. 'there/that'; ind. 21; 139; 46)
over here (ind. 13; 38; 10)	там (<i>tam</i> ; lit. 'there_place'; ind. 19; 841; 129)
not me (ind. 13; 12; 7)	наш (<i>naš</i> ; lit. 'ours'; ind. 12; 643; 91)
what about me (ind. 13; 7; 6)	
how about that (ind. 13; 7; 6)	у вы (<i>u vy</i> ; lit. 'at you.2PL'; ind. 47; 120; 53)
you chap (ind. 12; 9; 6)	у ты (<i>u ty</i> ; lit. 'at you.2SG'; ind. 42; 135; 52)
what about my (ind. 11; 5; 5)	ты что (<i>ty čto</i> ; lit. 'you.2SG what'; ind. 38; 116; 46)
over there (ind. 11; 27; 8)	у я (<i>u ja</i> ; lit. 'at me'; ind. 35; 410; 76)
now then (ind. 11; 12; 6)	а ты (<i>a ty</i> ; lit. 'PART you.2SG; ind. 344; 148; 47)
what about your (ind. 10; 7; 5)	это что (<i>èto čto</i> ; lit. 'this what'; ind. 26; 32; 22)
what about this (ind. 10; 7; 5)	у мы (<i>u my</i> ; lit. 'at me; ; ind. 25; 237; 48)
	с ты (<i>s ty</i> ; lit. 'with you.2SG'; ind. 24; 110; 34)
	что это (<i>čto èto</i> ; lit. 'what this'; ind. 24; 291; 52)
	это не (<i>èto ne</i> ; lit. 'this not'; ind. 23; 226; 45)
	ну вот (<i>nu vot</i> ; lit. 'well this'; ind. 22; 33; 20)
	это всё (<i>èto vsë</i> ; lit. 'this all'; ind. 22; 60; 25)
	вот это (<i>vot èto</i> ; lit. 'here this'; ind. 21; 42; 21)
	что с ты (<i>čto s ty</i> ; lit. 'what with you.2SG'; ind. 21; 21; 16)
	это мой (<i>èto moj</i> ; lit. 'this my'; ind. 20; 22; 16)
	ах ты (<i>ax ty</i> ; lit. 'oh you.2SG'; ind. 18; 15; 13)
	это ты (<i>èto ty</i> ; lit. 'this you.2SG'; ind. 17; 52; 20)
	это хороший (<i>èto xorošij</i> ; lit. 'this good'; ind. 17; 14; 12)
	это же (<i>èto že</i> ; lit. 'this PART'; ind. 17; 35; 17)
	же это (<i>že èto</i> ; lit. 'PART this'; ind. 17; 41; 18)
	это такой (<i>èto takoj</i> ; lit. 'this such'; ind. 16; 32; 16)
	кто там (<i>kto tam</i> ; lit. 'who there'; ind. 16; 15; 12)
	ты кто (<i>ty kto</i> ; lit. 'you.2SG who'; ind. 16; 19; 13)
	это очень (<i>èto očen'</i> ; lit. 'this very'; ind. 16; 39; 17)
	нет это (<i>net èto</i> ; lit. 'not this'; ind. 16; 20; 13)
	а вот (<i>a vot</i> ; lit. 'PART this_here'; ind. 14; 72; 20)
	вот так (<i>vot tak</i> ; lit. 'this like_that'; ind. 14; 42; 16)
	это неправда (<i>èto nepravda</i> ; lit. 'this untruth'; ind. 14; 13; 10)
	это ещё (<i>èto eščë</i> ; lit. 'this even_more'; ind. 13; 18; 11)
	это за (<i>èto za</i> ; lit. 'this for'; ind. 13; 23; 12)
	как так (<i>kak tak</i> ; lit. 'how this'; ind. 13; 11; 9)
	вот как (<i>vot kak</i> ; lit. 'this how'; ind. 13; 30; 13)
	что это за (<i>čto èto za</i> ; lit. 'what this PART'; ind. 12; 20; 11)

Table 16. Immediacy and Deixis

The examples in (105) illustrate an instance of this function. The markers ‘*this*’ and ‘*here*’ in English and ‘сюда’ (*sjuda*; lit. ‘*over here*’) in Russian are referring to a place in the ongoing utterance situation. Such markers have a particular dependence on the discourse context for their resolution.

(105) ***Immediacy and Deixis***

English (OD E1: 2003186)

[Shukhov watched to see which bowls he filled before the good part settled back on the bottom of the caldron and which had only the watery stuff off the top. He put ten bowls on the tray and went away. Gopchik was waving at him from a place by the second pair of posts. “This way, Ivan Denisovich, over here!”]

This way, Ivan Denisovich, over here!

English (OD E2: 3292492)

[Shukhov took note which dishes had been filled before the solids had sunk to the bottom of the pail, and which held thin stuff, nothing but water. He lined up ten bowls on his tray and carried it off. Gopchik was waving to him from near the second row of pillars. “Over here, Ivan Denisovich, over here!” Carrying a tray laden with bowls is not as easy as it looks.]

Over here, Ivan Denisovich, over here!

Russian (OD RO: 408514)

[Шухов приметил, какие миски набрать, пока ещё гущина на дно бака не осела, и какие по-холостому – жижа одна. Уставил на своем подносе десять мисок и понес. Гопчик ему машет от вторых столбов: – Сюда, Иван Денисыч, сюда! Миски нести – не рукавом трясти.]

Сюда,	Иван	Денисыч,	сюда!
<i>sjuda</i>	<i>Ivan</i>	<i>Denisyč</i>	<i>sjuda</i>
ADV	NN.NOM	NN.NOM	ADV
<i>over_here</i>	<i>Ivan</i>	<i>Denisych</i>	<i>over_here</i>

2.2.6 Questioning

Verbless sentences appear to be closely correlated with questioning speech acts. This is suggested by their elements and n-grams which statistically set them apart from verbal

sentences. The markers of questioning speech acts that are significant for verbless sentences are shown in *Table 17*, which includes lemma elements and lemma n-grams.

English	Russian
what (ind. 174; 3449; 412)	а (<i>a</i> ; lit. ‘and/but.PART’; ind. 121; 4835; 785)
eh (ind. 50; 80; 43)	какой (<i>kakoj</i> ; lit. ‘which/how/what’; ind. 90; 692; 221)
why (ind. 48; 857; 106)	почему (<i>počemu</i> ; lit. ‘why’; ind. 34; 375; 100)
huh (ind. 25; 22; 17)	кто (<i>kto</i> ; lit. ‘who’; ind. 31; 835; 154)
how (ind. 16; 1268; 76)	где (<i>gde</i> ; lit. ‘where’; ind. 30; 718; 138)
really (ind. 10; 478; 34)	что (<i>čto</i> ; lit. ‘what’; ind. 22; 8254; 764)
	зачем (<i>začem</i> ; lit. ‘why’; ind. 22; 152; 49)
what about (ind. 168; 110; 90)	как (<i>kak</i> ; lit. ‘how’; ind. 21; 4198; 433)
why not (ind. 67; 49; 37)	чего (<i>čego</i> ; lit. ‘of_what’; ind. 17; 128; 40)
how about (ind. 51; 38; 28)	ась (<i>as’</i> ; lit. ‘eh.PART’; ind. 11; 10; 9)
what for (ind. 49; 32; 26)	
what about the (ind. 36; 20; 18)	а что (<i>a čto</i> ; lit. ‘and.PART what’; ind. 46; 136; 55)
but what about (ind. 20; 12; 10)	что такой (<i>čto takoj</i> ; lit. ‘what this’; ind. 40; 84; 42)
but what (ind. 19; 90; 17)	ну что (<i>nu čto</i> ; lit. ‘well.PART what’; ind. 38; 116; 46)
what about it (ind. 16; 7; 7)	что ж (<i>čto ž</i> ; lit. ‘what really.PART’; ind. 33; 153; 47)
what now (ind. 15; 18; 7)	ну как (<i>nu kak</i> ; lit. ‘well.PART what’; ind. 33; 63; 33)
so what (ind. 14; 46; 11)	как же (<i>kak že</i> ; lit. ‘how really.PART’; ind. 32; 82; 36)
and what about (ind. 14; 10; 7)	что же (<i>čto že</i> ; lit. ‘what really.PART’; ind. 27; 126; 38)
what of (ind. 13; 7; 6)	а где (<i>a gde</i> ; lit. ‘and.PART where’; ind. 25; 36; 22)
what about me (ind. 13; 7; 6)	что это (<i>čto èto</i> ; lit. ‘what this’; ind. 24; 291; 52)
how about that (ind. 13; 7; 6)	что за (<i>čto za</i> ; lit. ‘what really.PART’; ind. 24; 66; 27)
how many (ind. 12; 62; 11)	чем дело (<i>čem delo</i> ; lit. ‘what deal’; ind. 23; 27; 19)
what about my (ind. 11; 5; 5)	в чем дело (<i>v čem delo</i> ; lit. ‘in what deal’; ind. 22; 26; 18)
well what about (ind. 11; 5; 5)	что с ты (<i>čto s ty</i> ; lit. ‘what with you.2SG’; ind. 21; 21; 16)
but why (ind. 11; 17; 7)	что с (<i>čto s</i> ; lit. ‘what with’; ind. 19; 104; 28)
	в чем (<i>v čem</i> ; lit. ‘in what’; ind. 18; 75; 24)
when (ind. -11; 1800; 6)	где же (<i>gde že</i> ; lit. ‘where really.PART’; ind. 18; 23; 15)
	кто там (<i>kto tam</i> ; lit. ‘who there’; ind. 16; 15; 12)
	ты кто (<i>ty kto</i> ; lit. ‘you.2SG who’; ind. 16; 19; 13)
	за что (<i>za čto</i> ; lit. ‘for what’; ind. 16; 83; 23)
	что ли (<i>čto li</i> ; lit. ‘what eh.PART’; ind. 15; 63; 20)
	кто такой (<i>kto takoj</i> ; lit. ‘who this_way’; ind. 15; 17; 12)
	а как же (<i>a kak že</i> ; lit. ‘and.PART how really.PART’; ind. 15; 14; 11)
	как так (<i>kak tak</i> ; lit. ‘how this_way’; ind. 13; 11; 9)
	при чем (<i>pri čem</i> ; lit. ‘of what’; ind. 12; 26; 12)
	ну и что (<i>nu i čto</i> ; lit. ‘well.PART and what’; ind. 12; 21; 11)
	когда (<i>kogda</i> ; lit. when; ind. -22; 1466; 20)

Table 17. Questioning

The correlation with questioning is also made transparent by the analysis of the characteristic elements and n-grams within the immediate *right and left context of the verbless sentences*, i.e. the markers that set the context of the verbless sentence apart, statistically, from the rest of the corpus. These include for instance the significant presence in the left context (i.e. three sentences left of the verbless sentence) of the lemma n-grams: ‘спросить он’ (*sprosit’ on*; lit. ‘ask he’; ind. 13; 89; 56) and ‘спросить я’ (*sprosit’ ja*; lit. ‘ask I’; ind. 11; 85; 51) in Russian and ‘do you’ (ind. 36; 890; 282) and ‘he ask’ (ind. 14; 148; 60) in English.

Notably, as evidenced from the comparison of the specificity indices in the above table (and the first lines of the results shown in *Appendix 6*), questioning seems to be particularly important for English. For English, the lemma word ‘what’ is in the top three elements statistically characterizing verbless sentences. While both languages require further disambiguation through manual annotation, the indices of the Russian lemmas in this questioning function, as it stands, though also high, show slightly lower specificity.

It is also notable that this question word ‘what’ is the most general when compared to question words such ‘where’, ‘why’, ‘when’. The latter are much more transparent about the specific information that is desired (i.e. a place, a reason, a time). In contrast, ‘what’ is not necessarily questioning the identity of an object, it may be for instance referring to a situation, e.g. ‘What is going on here?’, or to a more general quality, e.g. ‘A: He’s a genius. B: What genius?’. The question word ‘what’ is therefore much more likely to be a part of indirect speech acts, i.e. in which a specific piece of information is not being asked. Such a high specificity index for the word ‘what’ is in line with the arguments we make in *Chapter 6* below, as well as our study on questions in (Bondarenko & Celle, 2019).

Furthermore, the table also indicates elements with a negative index. The negative index signals elements that are more characteristic of the rest of the corpus than they are of verbless sentences. In particular, the present results show that the temporal question word ‘when’ (ind. -11; 1800; 6) and ‘когда’ (*kogda*; lit. ‘when’; ind. -22; 1466; 20) is *negatively* associated with the verbless sentences (i.e. it is more associated with the rest of the corpus). In other words, ‘when’ is statistically rejected by English verbless sentences and ‘когда’ (*kogda*; lit. ‘when’) is statistically rejected by the Russian verbless sentences. Temporal questions are associated with *verbal* rather than verbless sentences. This finding seems to be very much in line with the theoretical arguments we made in *Part 1* above concerning the ‘a-temporal’ status of verbless sentences, i.e. the result adds support to the argument that verbless sentences are *not* marked for tense or time on any hidden level.

A curious example of the questioning function is illustrated in (106). It illustrates another way that the question function may be performed via a verbless sentence. Though it does not contain any of the question markers in the above table, it is interesting for a different reason.

(106) *Questioning*

English

(OS E1: 2313801)

- [– Citizen commander, bath orderly reporting, sir.
 – Good, now let’s see. I suppose there’s no underwear, is there?
 – Underwear?
 – What about soap?
 – They never brought it.
 – There is water, isn’t there?
 – Water? Who says there’s water? It’s turned off at the main.
 – Hm-mmm.
 – So much the better, citizen sergeant. Won’t take so long to wash them.]

Water?

Russian

(OS RO: 659196)

- [– По вашему вызовы завбаней явился!
 – Ты... вот что... Белья нет?
 – Откуда бельё?
 – А мыло?
 – Не подвезли.
 – Вода-то есть?
 – Где есть! Перекрыли...
 – Гм-м...
 – Ещё лучшее, гражданин начальник, быстрей вымоем.]

Где	есть!
<i>gde</i>	<i>est'</i>
PRO	V.PR
<i>where</i>	<i>is</i>

As we showed in (Bondarenko & Celle, 2019), questions appear to be particularly sensitive to the verb. In *Chapter 6* below, we provide evidence showing that the absence of the verb may be a grammatical marker of indirect speech acts. In both of these studies, we observed that there are very few instances where English verbless questions have a verbal correspondence in Russian, such as this particular instance in (106).

What is notable about this example is that it shows an instance of an English verbless question with an indirect function (i.e. it is not seeking information), which is translated in

Russian with the use of the present tense ‘есть’ (*est*’; lit. ‘*be.V.PR*’). It must be kept in mind that the explicit use of this verb in Russian is marked in the present tense, i.e. ‘быть’ (*byt*’; lit. ‘*be.V.INF*’) is typically not used in the present tense in Russian. In other words, the Russian sentence in this case is using a *marked* verb to indicate the *indirect* function of the sentence. In *Chapter 6*, we suggest that Russian often appears to rely on emphatic particles to mark the direct-indirect distinction in questions. This does not change that observation, though it does suggest that *indirect speech acts* seem to make themselves explicit using particularly *marked structures*, i.e. whether it is done using a marked verb (as in the above example), using the marked absence of the verb (as in the English results discussed in *Chapter 6*), or using marked particles (as in the observation of Russian questions in *Chapter 6*). The question of what is marked being particular to the typological characteristics of the language.

2.2.7 Emphasis

Another semantic feature that characterizes verbless sentences more than it does verbal sentences is emphasis. Emphatic markers in terms of lemma elements and lemma n-grams, and their specificity to the verbless sentence, are presented in *Table 18*. We further distinguish emphasis in terms of emphasis of intensity, e.g. ‘*This soup is very good*’, and emphasis of contrast, e.g. ‘*This soup is the most delicious soup that I’ve ever eaten*’. Verbless sentences appear to be particularly associated with conveying emphasis of intensity. This especially concerns the Russian data, as evidenced from the specificities in the table.

English	Russian
<u>Intensity</u>	<u>Intensity</u>
what (ind. 174; 3449; 412)	какой (<i>kakoj</i> ; lit. 'which/how/what'; ind. 90; 692; 221)
God (ind. 32; 369; 58)	же (<i>že</i> ; lit. 'really.PART'; ind. 35; 1886; 274)
course (ind. 18; 383; 42)	такой (<i>takoj</i> ; lit. 'how/such/like_this'; ind. 34; 1314; 215)
how (ind. 16; 1268; 76)	очень (<i>očen'</i> ; lit. 'very'; ind. 31; 835; 154)
very (ind. 12; 1072; 61)	вот (<i>vot</i> ; lit. 'here/that'; ind. 122; 1226; 346)
too (ind. 10; 946; 53)	конечно (<i>konečno</i> ; lit. 'of course'; ind. 29; 289; 80)
really (ind. 10; 478; 34)	ж (<i>ž</i> ; lit. 'really.PART'; ind. 23; 309; 74)
just (ind. 9; 1677; 74)	что (<i>čto</i> ; lit. 'what'; ind. 22; 8254; 764)
	эх (<i>èx</i> ; lit. 'ei.PART'; ind. 22; 47; 28)
of course (ind. 37; 303; 41)	как (<i>kak</i> ; lit. 'how'; ind. 21; 4198; 433)
oh God (ind. 11; 11; 6)	настоящий (<i>nastojščij</i> ; lit. 'real'; ind. 17; 105; 55)
	здорово (<i>zdorovo</i> ; lit. 'really'; ind. 15; 58; 25)
<u>Contrast</u>	ведь (<i>ved'</i> ; lit. 'really.PART'; ind. 12; 481; 75)
more (ind. 6; 1240; 52)	ужасно (<i>užasno</i> ; lit. 'terribly'; ind. 12; 45; 19)
	точно (lit. 'precisely'; ind. 9; 176; 35)
	что ж (<i>čto ž</i> ; lit. 'what really.PART'; ind. 33; 153; 47)
	как же (<i>kak že</i> ; lit. 'how really.PART'; ind. 32; 82; 36)
	ещё бы (<i>eščè by</i> ; lit. 'indeed'; ind. 27; 30; 22)
	ну и (<i>nu i</i> ; lit. 'well and'; ind. 26; 87; 32)
	что за (<i>čto za</i> ; lit. 'what really.PART'; ind. 24; 66; 27)
	ну конечно (<i>nu konečno</i> ; lit. 'well of course'; ind. 21; 15; 14)
	ах ты (<i>ax ty</i> ; lit. 'oh you'; ind. 18; 15; 13)
	да конечно (<i>da konečno</i> ; lit. 'yes of course'; ind. 15; 17; 12)
	<u>Contrast</u>
	а (<i>a</i> ; lit. 'and/but/whereas/instead/until.PART'; ind. 121; 4835; 785)

Table 18. Emphasis

A few examples in (107) illustrate the emphatic function of verbless sentences. The Russian verbless sentence in (a) shows contrastive emphasis using the particle 'a' (*a*; lit. 'and/but/whereas/instead/until'). Although this lemma shows a high specificity for Russian verbless sentence, it also has many other uses which are not emphatic. For instance, it is also commonly used in questions, such as (b). Further manual annotation is required in order to disambiguate the different uses of this particle.

(107) *Emphasis*

a. Russian (OD RO: 342746)

[Баланда не менялась ото дня ко дню, зависело – какой овощ на зиму заготовят. В летошнем году заготовили одну соленую морковку – так и прошла баланда на чистой моркошке с сентября до июня. А нонче – капуста черная.]

А	нонче –	капуста	черная.
<i>a</i>	<i>nonče</i> –	<i>kapusta</i>	<i>černaja</i>
PART	ADV	NN.NOM	ADJ.NOM
<i>and</i>	<i>now</i> –	<i>cabbage</i>	<i>black</i>

English (OD E1: 1908935)

[It depended on what vegetables they'd stored for winter. The year before they'd only stocked up with salted carrots, so there was nothing but carrots in the gruel from September to June. And now it was cabbage.]

And now it was cabbage.

English (OD E2: 3202195)

[What was in it depended on which vegetable was stock piled for winter. Last year they'd laid in nothing but carrots in brine – so from September to June it was carrots all the way. This time around it was black cabbage.]

This time around, it was black cabbage.

b. Russian (OD RO: 368339)

[А иной раз подумаешь – дух согрет: срок-то всё ж кончается, катушка-то на размоте ... Господи! Своими ногами – да на волю, а?]

Своими	ногами –	да	на	волю,	а?
<i>svoimi</i>	<i>nogami</i> –	<i>da</i>	<i>na</i>	<i>volju,</i>	<i>a</i>
PRO.INS	NN.INS	PART	PREP	NN.ACC	PART
<i>own</i>	<i>legs</i> –	<i>and</i>	<i>to</i>	<i>freedom,</i>	<i>eh</i>

English (OD E1: 1944952)

[But sometimes you got a kind of funny feeling inside. Maybe your number really would come up one day. God, just to think you might walk out and go home!]

God, just to think you might walk out and go home!

[Sometimes, though, you got thinking and your spirits soared: your sentence was running out, there wasn't much thread left on the spool! Lord! Just think of it! Walking free, on your own two legs!]

Walking free, on your two legs!

The verbless sentences of both (c) and (d) illustrate emphasis of intensity. Example (c) shows ‘*God*’ used in the English sentence to intensify the truth of the speaker’s evaluation. In the Russian sentence in (c), the speaker’s assessment is intensified through the demonstrative ‘э́та’ (*èta*; lit. ‘*this*.DEM.F’) which brings the concept of ‘*majority*’ into the utterance situation in order to criticize it from a closer vantage point.

c. English

(FH EO: 1106880)

[All right, he’s had his say. You must take it in. I’ll say my say, too, in the next few hours. And you’ll take it in. And you’ll try to judge them and make your decision as to which way to jump, or fall. But I want it to be your decision, not mine, and not the Captain’s. But remember that the Captain belongs to the most dangerous enemy of truth and freedom, the solid unmoving cattle of the majority. Oh, God, the terrible tyranny of the majority. We all have our harps to play. And it’s up to you now to know which ear you’ll listen.]

Oh, God, the terrible tyranny of the majority.

Russian

(FH R1: 2729052)

[Хорошо, он сказал всё что хотел. Вы это выслушали. Теперь в ближайший час буду говорить я. Вам придётся выслушать и это. А потом постарайтесь разобраться и решить – с кем вы. Но я хочу, чтобы вы решили это сами, чтобы это решение было вашим собственным, а не моим и не брандмейстера Битти. Одного только не забывайте – брандмейстер принадлежит к числу самых опасных врагов истины и свободы, к тупому и равнодушному стаду нашего большинства. О, эта ужасная тирания большинства! Мы с Битти поем разные песни. От вас самого зависит, кого вы станете слушать.]

О,	эта	ужасная	тирания	большинства!
<i>o,</i>	<i>èta</i>	<i>užaznaja</i>	<i>tiranija</i>	<i>bol’sinstva</i>
INTJ	DEM.F	ADJ.NOM	NN.NOM	NN.GEN
<i>o</i>	<i>this</i>	<i>terrible</i>	<i>tyranny</i>	<i>of_majority</i>

Similarly, in (d), ‘*God*’ and ‘Господи’ (*Gospodi*; lit. ‘*God*’) is used in English and Russian in order to intensify the truth of the statement concerning the resemblance between

Liza and her mother. In addition, the English sentence uses ‘*very*’ to emphasize the degree of the similarity, while in Russian another demonstrative ‘*вот*’ (*vot*; lit. ‘*here/this.PART*’) is used to intensify this extent.

d. English (TD E1: 1661512)

[One evening Sergei Platonovich glanced at his daughter across the tea-table, and was startled. Liza, who had just left high school, had grown into a slender good-looking girl. He looked at her and the saucer filled with ambercoloured tea trembled in his hand. How like her mother she was! God, her very image! “Liza, turn your head sideways!” He had never before noticed how amazingly his daughter resembled her mother.]

God, her very image!

Russian (TD RO: 150683)

[И вот как-то за вечерним чаем несказанно удивился Сергей Платонович, глянув на дочь (Елизавета, к тому времени окончившая гимназию, успела выровняться в видную, недурную девушку), глянул и блюдец с янтарным чаем запрыгало в руках. «На мать-покойницу похожа. Господи, вот сходство!» – Лизка, а ну, повернись! – Проглядел, что дочь с самого детства разительно напоминала мать.]

Господи,	вот	сходство!
<i>Gospodi,</i>	<i>vot</i>	<i>sxodstvo</i>
NN	PART	NN.NOM
<i>God,</i>	<i>this</i>	<i>similarity</i>

Finally, (e) illustrates the emphasis of intensity in the verbless sentences of both languages using ‘*what*’ and ‘*какой*’ (*kakoj*; lit. ‘*how/what*’) to describe the extent of the high pulse.

e. English (FH EO: 1106021)

[He felt beaten unmercifully on brow, eyes, nose, lips, chin, on shoulders on upflailing arms. He wanted to yell, “No! shut up, you’re confusing things, stop it!” Beatty’s graceful fingers thrust out to seize his wrist. “God, what a pulse! I’ve got you going, have I, Montag.]

God, what a pulse!

[Ему казалось, что его нещадно избивают по голове, глазам, лицу, плечам, по беспомощно поднятым рукам. Ему хотелось крикнуть: “Нет! Замолчите! Вы стараетесь всё запутать. Довольно!” Тонкие нервные пальцы Битти схватили Монтэга за руку. – Боже какой пульс! Здорово я вас взвинтил, Монтэг, а?]

Боже	какой	пульс!
<i>Bože</i>	<i>kakoj</i>	<i>pul's</i>
NN	PRO	NN.NOM
<i>God</i>	<i>what</i>	<i>pulse</i>

Recent work on emphasis (Trotzke, 2017; Trotzke & Turco, 2015) has shown that marked word order in combination with discourse particles is typically associated with emphasis and expressiveness. In so far as verbless sentences may be treated as instances of marked word order, their relation to emphasis, which surfaces in our results through the characteristic elements analysis, can be considered as being further highlighted by the noncanonical nature of their syntactic structure.

2.2.8 Reference

Analysis of characteristic elements also suggests the association of English verbless sentences with indefinite reference. *Table 19* reveals the indices of the key lemma and category elements in this regard.

In addition to signaling the existence of an indefinite reference relation through *wh*-pronouns (e.g. Haspelmath, 1997; Kuroda, 1965), the present results also suggest the preference of verbless sentences for indefinite reference through the observed statistical *dispreference for definite* reference. While the indefinite article shows a very moderate specificity index in English relative to the other lemma elements (i.e. lemma ‘*a*’; ind. 16; 18072; 545), it is the negative index of the definite article lemma that is most revealing (i.e. lemma ‘*the*’; ind. -13; 46612; 778). This means that the definite article is shown to be statistically characteristic of *verbal* sentences rather than verbless sentences.

English	Russian
what (ind. 174; 3449; 412)	это (<i>èto</i> ; lit. 'this'; ind. 222; 3612; 827)
why (ind. 48; 857; 106)	ты (<i>ty</i> ; lit. 'you.2SG'; ind. 130; 3595; 665)
darling (ind. 22; 45; 20)	вот (<i>vot</i> ; lit. 'here/that'; ind. 122; 1226; 346)
bastard (ind. 19; 134; 27)	какой (<i>kakoj</i> ; lit. 'which/how/what'; ind. 90; 692; 221)
one (ind. 18; 2673; 130)	вы (<i>vy</i> ; lit. 'you.2PL'; ind. 45; 2755; 388)
Jake (ind. 17; 123; 24)	мой (<i>moj</i> ; lit. 'my'; ind. 41; 1204; 216)
Jenkins (ind. 17; 11; 10)	почему (<i>počemu</i> ; lit. 'why'; ind. 34; 375; 100)
Millie (ind. 17; 29; 14)	такой (<i>takoj</i> ; lit. 'how/such/like_this'; ind. 34; 1314; 215)
all (ind. 17; 3944; 167)	кто (<i>kto</i> ; lit. 'who'; ind. 31; 835; 154)
comrade (ind. 16; 157; 26)	где (<i>gde</i> ; lit. 'where'; ind. 30; 718; 138)
a (ind. 16; 18072; 545)	твой (<i>tvoj</i> ; lit. 'yours.2SG'; ind. 28; 371; 91)
your (ind. 16; 18072; 545)	ваш (<i>vaš</i> ; lit. 'yours.2PL'; ind. 27; 429; 96)
how (ind. 16; 1268; 76)	дело (<i>delo</i> ; lit. 'deal'; ind. 26; 669; 125)
my (ind. 15; 3222; 141)	товарищ (<i>tovarišč</i> ; lit. 'comrade'; ind. 22; 302; 72)
nothing (ind. 15; 643; 50)	что (<i>čto</i> ; lit. 'what'; ind. 22; 8254; 764)
Grisha (ind. 15; 85; 19)	зачем (<i>začem</i> ; lit. 'why'; ind. 22; 152; 49)
citizen (14; 34; 13)	джейк (lit. 'Jake'; ind. 21; 125; 44)
mortar (13; 114; 20)	вон (<i>von</i> ; lit. 'there/that'; ind. 21; 139; 46)
Melekhov (13; 50; 14)	милый (<i>milyj</i> ; lit. 'darling'; 21; 95; 38)
Montag (12; 455; 17)	как (<i>kak</i> ; lit. 'how'; ind. 21; 4198; 433)
Mildred (12; 104; 18)	ничто (<i>ničto</i> ; lit. 'nothing'; ind. 21; 813; 129)
NP (cat. proper nouns; ind. 229; 29892; 1559)	гражданин (<i>graždanin</i> ; lit. 'citizen'; ind. 21; 51; 28)
NN (cat. common nouns; ind. 134; 118187; 3691)	там (<i>tam</i> ; lit. 'there_place'; ind. 19; 841; 129)
WP (cat. wh-pronouns; ind. 124; 5188; 430)	чего (<i>čego</i> ; lit. 'of_what'; ind. 17; 128; 40)
NNS (cat. common plural nouns; ind. 41; 28866; 957)	начальник (<i>načalnik</i> ; lit. 'supervisor'; ind. 15; 110; 34)
WRB (cat. wh-adverbs; ind. 23; 4905; 215)	милли (<i>milli</i> ; lit. 'Millie'; ind. 15; 30; 18)
<u>Negative Specificity Index</u>	дженкинз (<i>dženkinz</i> ; lit. 'Jenkins'; ind. 14; 11; 11)
he (-122; 13569; 3)	дурак (<i>durak</i> ; lit. 'fool'; ind. 13; 102; 30)
I (-104; 14738; 23)	боксер (<i>bokser</i> ; lit. 'Boxer'; ind. 13; 61; 23)
it (-65; 10887; 29)	господин (<i>gospodin</i> ; ind. 13; 158; 38)
his (-43; 9808; 46)	наш (<i>naš</i> ; lit. 'ours'; ind. 12; 643; 91)
they (-43; 5035; 3)	все (<i>vsë</i> ; lit. 'all'; ind. 12; 2154; 223)
she (-29; 2973; 0)	война (<i>vojna</i> ; lit. 'war'; ind. 11; 166; 37)
we (-27; 3346; 3)	молодец (<i>molodec</i> ; lit. 'excellent_person'; ind. 11; 23; 13)
him (-20; 4303; 20)	ребята (<i>rebjata</i> ; lit. 'guys'; ind. 11; 88; 25)
their (-15; 1803; 2)	<u>Negative Specificity Index</u>
the (ind. -13; 46612; 778)	он (<i>on</i> ; lit. 'he'; -53; 12260; 432)
PP (cat. personal pronouns; ind. -318; 77238; 405)	они (<i>oni</i> ; lit. 'they'; -21; 4709; 164)
	её (<i>eë</i> ; lit. 'her'; -20; 1354; 19)
	свой (<i>svoj</i> ; lit. 'one's_own'; -20; 2061; 47)
	себя (<i>sebjä</i> ; lit. 'self'; -18; 1218; 18)
	тот (<i>tot</i> ; lit. 'that_one'; -14; 1748; 46)
	она (<i>ona</i> ; lit. 'she'; -12; 2918; 109)
	я (<i>ja</i> ; lit. 'I'; -4; 12295; 713)

Table 19. Referent Specificity

Furthermore, a very strong *dis*-preference is also observed for personal pronouns, particularly in the third person, in the verbless sentences of both languages. In contrast, second person pronouns have high positive indices, particularly in Russian. Proper nouns are also positively associated with verbless sentences in both languages. A potential reason for this may have to do with the role that proper nouns and second person pronouns have in direct address, which third person pronouns lack. Such a split in terms of pronouns would thus point to the interactive function of verbless sentences. However, direct address is *not* the only purpose with which nouns are used in verbless sentences. Among uses which are non-vocative are also typically descriptions and imperative uses of nouns. However, in order to disambiguate and precise the types and proportions, further annotation is required.

Several uses of indefinite reference within verbless sentences are illustrated in (108). Example (a) shows the wh-pronoun ‘*what*’ used to indefinitely, though emphatically, characterize a particular event that has just occurred. This event, though specific, is attributed to be one amongst the class of ‘*strange meetings on strange nights*’. Example (b) shows a similar use, but in which the speaker is expressing criticism of the event of being woken up by means of attributing it to the indefinite class of ‘*dirty business*’.

In the Russian translation, the indefinite description in (a) is rendered through the use of the adjectival pronoun ‘какой’ (*kakoj*; lit. ‘*which/how/what*’); and the critical description in (b), through the interrogative pronoun ‘что’ (*čto*; lit. ‘*what*’) combined with the particle ‘за’ (*za*; lit. ‘*really.PART*’) which emphasize that the event belongs to some sort of excessive form of ‘безобразия’ (i.e. ‘*nonsense*’).

(108) ***Indefinite Reference***

a. English

(FH EO: 1031712)

[But she was gone – running in the moonlight. Her front door shut gently. “Happy! Of all the nonsense.” He stopped laughing. He put his hand into the glove-hole of his front door and let it know his touch. The front door slid open. Of course I’m happy. What does she think? I’m not? he asked the quiet rooms. He stood looking up at the ventilator grille in the hall and suddenly remembered that something lay hidden behind the grille, something that seemed to peer down at him. He moved his eyes quickly away. What a strange meeting on a strange night. He remembered nothing like it save one afternoon a year ago when he had met an old man in the park and they had talked... Montag shook his head.]

What a strange meeting on a strange night.

[Но девушки перед ним уже не было – она бежала прочь по залитой лунным светом дорожке. В доме тихо затворилась дверь. Счастлив ли я? Что за вздор! Монтэг перестал смеяться. Он сунул руку в специальную скважину во входной двери своего дома. В ответ на прикосновение его пальцев дверь открылась. Конечно, я счастлив. Как же иначе? Она что думает – что я несчастлив? Спрашивал он пустых комнат. В передней взор его пал на вентиляционную решетку. И вдруг вспомнил, что там спрятано. Оно как будто поглядело на него оттуда. Он быстро отвел глаза. Какая странная ночь, какая странная встреча! Такого с ним ещё не случилось. Разве только тогда в парке, год назад, когда он встретился со стариком и они разговорились... Монтэг тряхнул головой.]

Какая	странная	ночь,	и	какая	странная	встреча!
<i>kakaja</i>	<i>strannaja</i>	<i>noč'</i>	<i>i</i>	<i>kakaja</i>	<i>strannaja</i>	<i>vstreča</i>
PRO	ADJ.NOM	NN.NOM	CONJ	PRO	ADJ.NOM	NN.NOM
<i>what</i>	<i>strange</i>	<i>night</i>	<i>and</i>	<i>what</i>	<i>strange</i>	<i>meeting</i>

b. English

(SR EO: 790435)

[There was a row going on outside. I listened and I thought I recognized a voice. I put on a dressing-gown and went to the door. The concierge was talking downstairs. She was very angry. I heard my name and called down the stairs. "Is that you Monsieur Barnes?" the concierge called. "Yes. It's me." "There's a species of woman here who's waked the whole street up. What kind of a dirty business at this time of night! She says she must see you. I've told here you're asleep."]

What kind of a dirty business at this time of night!

[Снаружи доносился шум. Я прислушался, и мне показалось, что я слышу знакомый голос. Я надел халат и подошел к двери. Внизу раздавался голос у консьержки. Она очень сердилась. Услыхав свое имя, я окликнул её.

– Это вы, мосье Барнс? – крикнула консьержка.

– Да я.

– Здесь какая-то женщина, она шумит на всю улицу. Что за безобразие, в такую пору! Говорит, что ей нужно вас видеть.]

Что	за	безобразие,	в	такую	пору!
<i>čto</i>	<i>za</i>	<i>bezobrazie,</i>	<i>v</i>	<i>takuju</i>	<i>poru</i>
PRO	PART	NN.NOM	PREP	DEM	NN.ACC
<i>what</i>	<i>really</i>	<i>nonsense</i>	<i>in</i>	<i>this</i>	<i>time</i>

Example (c) illustrates the description of an indefinite ‘crime’ in BK_E1. This indefinite description constitutes a *prediction* about the future. In Russian, the prediction is carried out with the help of the future tense ‘будет’ (*budet*; lit. ‘be.V.FT’) and the demonstrative ‘эта’

(*èta*; lit. ‘*this*’) which refers to the just mentioned unspecified crime. Thus, the indefinite function in this case appears to be performed by the future tense which predicts an event that is not yet realized; whereas the demonstrative simultaneously brings the unspecified ‘*crime*’ into the utterance situation in order to create proximity and thereby emphasize it. The corresponding translation in BK_E2 uses ‘*be going to*’ for the prediction concerning an indefinite ‘*someone*’.

c. English (BK E1: 1557811)

[Now all the pious frauds in town will start talking and spread it over the whole province, wondering ‘what is the meaning of this dream?’ The old man is really astute, if you ask me: he smelled crime. It stinks in your family.”

“What crime? ”

Rakitin evidently wanted to speak his mind.

“A crime in your nice little family. It will take place between your dear brothers and your nice, rich papa. So Father Zosima bumps his forehead on the ground, for the future, just in case. Afterwards they’ll say, ‘Ah, it’s what the holy elder foretold, prophesied,’ though bumping your forehead on the ground isn’t much of a prophecy.]

A crime in your nice little family.

Russian (BK RO: 69690)

[Вот теперь и заговорят все святоши в городе и по губернии разнесут: «Что дескать сей сон означает?» По моему, старик действительно прозорлив: уголовщину пронюхал. Смердит у вас.

– Какую уголовщину?

Ракитину видимо хотелось что-то высказать.

– В вашей семейке она будет, эта уголовщина. Случится она между твоими братцами и твоим богатеньким батюшкой. Вот отец Зосима и стукнулся лбом на всякий будущий случай. Потом, что случится: «Ах, ведь это старец святой предрек, напороочествовал» , – хотя какое бы в том пророчество, что он лбом стукнулся?]

В	вашей	семейке	она	будет,	эта
<i>v</i>	<i>vašej</i>	<i>semejke</i>	<i>ona</i>	<i>budet,</i>	<i>èta</i>
PREP	PRO.GEN.2PL	NN.PRE	PRO.NOM.3SG	V.FT	DEM
<i>in</i>	<i>your</i>	<i>family</i>	<i>she</i>	<i>be</i>	<i>this</i>

УГОЛОВЩИНА.

ugolovščina

NN.NOM

crime

[It will get all the sanctimonious hypocrites in town talking, and soon the whole province will want to know: “What was all that performance about?” I think the old man's really pretty shrewd: he's smelt a crime in the air. Your family reeks of it.’
‘What crime?’

Rakitin was evidently anxious to have his say.

‘Someone’s going to get hurt in your family. It will be something between your brothers and your wealthy father. So, in anticipation of what might happen, Zosima struck the ground with his forehead. Next thing it’ll be: “After all, the holy man predicted it, his prophecy has come true” – some prophecy, just because he struck the ground with his forehead!]

Someone’s going to get hurt in your family.

Examples (d) and (e) show two imperative uses of verbless sentences with indefinite reference. In (d), a preposition accompanies the invitation to realize an event which would correspond to ‘a duel’. In (e), the indefinite noun phrase alone describes the desires of the speaker which are different from the present state of things. The speaker is inviting the interlocutors to realize a circle, thereby making way for the speaker to dance.

d. English

(BK E1: 1552449)

[Old liars who have been play-acting all their lives have moments when they get so carried away by their posing that they indeed tremble and weep from excitement, even though at that same moment (or just a second later) they might whisper to themselves: “You 're lying; you shameless old man, you 're acting even now, despite all your ' holy ' wrath and ' holy ' moment of wrath.” Dmitri Fyodorovich frowned horribly and looked at his father with inexpressible contempt.

“I thought ... I thought,” he said somehow softly and restrainedly, “that I would come to my birthplace with the angel of my soul, my fiancée, to cherish him in his old age, and all I find is a depraved sensualist and despicable comedian!”

“To a duel!” the old fool screamed again, breathless and spraying saliva with each word.]

To a duel!

English

(BK E2: 3119145)

[There are occasions on which old liars, who have spent their whole lives dissembling, get so carried away that they really do tremble and weep with emotion, despite the fact that at that very moment (or only a split second later) they could well admit to themselves, ‘You’re lying, you old reprobate, you’re play-acting even now, never mind all that “sacred” anger of yours, it’s all a sham.’

Dmitry Fyodorovich scowled viciously and looked at his father with indescribable contempt. ‘I imagined ... I imagined,’ he spoke quietly and with self-control, ‘that I would come home to my father with the angel of my heart, my fiancé, and that I

would cherish him in his old age, but all I see before me is a depraved old roué and the vilest of buffoons!’

‘To a duel!’ the old reprobate shrieked again, panting for breath and spluttering saliva with every word.]

To a duel!

Russian

(BK RO: 65379)

[Есть у старых лгунов, всю жизнь свою проактерствовавших, минуты, когда они до того зарисуются, что уже воистину дрожат и плачут от волнения, несмотря на то, что даже в это самое мгновение (или секунду только спустя) могли бы сами шепнуть себе: «Ведь ты лжешь, старый бесстыдник, ведь ты актер и теперь, несмотря на весь твой “святой” гнев и “святую” минуту гнева». Дмитрий Федорович страшно нахмурился и с невыразимым презрением поглядел на отца:

– Я думал ... я думал, – как-то тихо и сдержанно проговорил он, – что приеду на родину с ангелом души моей, невестою моей, чтобы лелеять его старость, а вижу лишь развратного сладострастника и подлейшего комедианта!

– На дуэль! – завопил опять старикашка, задыхаясь и брызгаясь с каждым словом слюной.]

На дуэль!
na duèl'
PREP NN.NOM
to duel

e. English

(TD E1: 1656150)

[The women were dancing now, to the accompaniment of shouts and whistles. They shook their ample bottoms (there was not a thin one there, for each was wearing five or six skirts), waved lace handkerchiefs, and worked their elbows in the dance. The grating notes of the accordion sounded imperatively. The player began the tune of the Cossack dance. A shout went up: “A circle! Form a circle!” “Squeeze up a bit!” Pyotr begged, pushing the perspiring women aside. Grigory roused himself and winked at Natalya: “Pyotr’s going to dance the ‘Cossack’! You watch him!” “Who with?” “Don’t you see? With your mother.”]

A circle!

Russian

(TD RO: 146229)

[Григорий глянул через головы сидевших за столом в кухню: под уханье и взвизги топтались в круговой бабы. Трясли полными задями (худых не было, на каждой по пять-семь юбок), махали кружевными утирками, сучили в пляске локтями. Требовательно резнула слух трехрядка. Гармонист заиграл казачка с басовыми переливами. – Круг дайте! Круг!

– Потеснитесь, гостечки! – упрашивал Петро, толкая разопревшие от пляса бабьи животы. Григорий, оживившись, мигнул Наталье. – Петро зараз казачка урежет, гляди. – С кем это он? – Не видишь? С матерью твоей.]

Круг	дайте!
<i>krug</i>	<i>dajte</i>
NN.NOM	V.IMP
<i>circle</i>	<i>give</i>

Example (f) illustrates the description of a non-specific referent, i.e. the ‘*terrible*’ quality of the ‘*whisper*’.

f. English (FH E1: 1034455)

[He felt his hand plunge toward the telephone. The jets were gone. He felt his lips move, brushing the mouthpiece of the phone. “Emergency hospital.” A terrible whisper. He felt that the stars had been pulverized by the sound of the black jets and that in the morning the earth would be thought as he stood shivering in the dark, and let his lips go on moving and moving.]

A terrible whisper.

Russian (FH R1: 2663003)

[Рука рванулась к телефону. Бомбардировщики пролетели. Его губы, дрогнув, коснулись телефонной трубки: – Больницу неотложной помощи. Шепот, полный ужаса ... Ему казалось, что от рева черных бомбардировщиков звезды превратились в пыль и завтра утром земля будет вся осыпана этой пылью, словно диковинным снегом. Эта нелепая мысль не покидала его, пока он стоял в темноте возле телефона, дрожа всем телом, беззвучно шевеля губами.]

Шёпот,	полный	ужаса.
<i>šëpot,</i>	<i>polnyj</i>	<i>užasa</i>
NN.NOM	ADJ.NOM	NN.GENT
<i>whisper</i>	<i>full</i>	<i>of_horror</i>

In (g), the ‘*whisper*’ enters the utterance situation as an indefinite referent; as does the ‘*familiarity*’ in (h). In Russian (h), the ‘*знакомое*’ (lit. ‘*familiarity*’) is made less definite with ‘*нечто*’ (lit. ‘*something*’).

[The stars poured over his sight like flaming meteors. He wanted to plunge in the river again and let it idle him safely on down somewhere. This dark land rising was like that day in his childhood, swimming, when from nowhere the largest wave in the history of remembering slammed him down in salt mud and green darkness, water burning mouth and nose, retching his stomach, screaming! Too much water! Too much land! Out of the black wall before him, a whisper. A shape. In the shape, two eyes. The night looking at him. The forest, seeing him. The Hound! After all the running and rushing and sweating it out and half-drowning, to come this far, work this hard, and think yourself safe and sigh with relief and come out on the land at last only to find...]

Out the black wall before him, a whisper.

[Звезды летели ему навстречу, как огненные метеоры. Ему захотелось снова броситься в реку, и пусть волны несут его всё равно куда. Темная громада берега напомнила ему тот случай из его детских лет, когда, купаясь, он был сбит с ног огромной волной (самой большой, какую он когда-либо видел!), она оглушила его и швырнула в зеленую темноту, наполнила рот, нос, желудок солено-жгучей водой. Слишком много воды! А тут было слишком много земли. И внезапно во тьме, стеною вставшей перед ним, – шорох, чья-то тень, два глаза. Словно сама ночь вдруг глянула на него. Словно лес глядел на него. Механический пес! Столько пробежать, так измучиться, чуть не утонуть, забраться так далеко, столько перенести, и, когда уже считаешь себя в безопасности и со вздохом облегчения выходишь наконец на берег, вдруг перед тобой ...]

И	внезапно	во	тьме,	стеною	вставшей	перед
<i>i</i>	<i>vnezapno</i>	<i>vo</i>	<i>t'me</i>	<i>stenoju</i>	<i>vstavšej</i>	<i>pered</i>
PART	ADV	PREP	NN.PRE	NN.INST	V.PART	PREP
<i>and</i>	<i>suddenly</i>	<i>in</i>	<i>darkness,</i>	<i>wall</i>	<i>raised</i>	<i>in_front</i>
ним,	–	шорох,	чья-то	тень,	два	глаза.
<i>nim</i>	–	<i>šorox</i>	<i>č'ja-to</i>	<i>ten'</i>	<i>dva</i>	<i>glaza</i>
PRO.INS.3SG		NN.NOM	PRO.NOM	NN.NOM	NUM	NN.NOM
<i>of_him</i>	–	<i>rustle,</i>	<i>someone's</i>	<i>shadow,</i>	<i>two</i>	<i>eyes</i>

[He stood breathing, and the more he breathed the land in, the more he was filled up with all the details of the land. He was not empty. There was more than enough here to fill him. There would always be more than enough. He walked in the shallow tide of leaves, stumbling. And in the middle of the strangeness, a familiarity. His foot hit something that rang dully. He moved his hand on the ground, a yard this way, a yard that. The railroad track. The track that came out of the city and rusted across the land, through forests and woods, deserted now, by the river. Here was the path to wherever he was going. Here was the single familiar thing, the magic charm he might need a little while, to touch, to feel beneath his feet, as he moved on into the bramble bushes and the lakes of smelling and feeling and touching, among the whispers and the blowing down of leaves.]

And in the middle of the strangeness, a familiarity.

Russian

(FH R1: 2752325)

[Он остановился, глубоко вдыхая запахи земли. И чем глубже он вдыхал их, тем осязаемее становился для него окружающий мир во всем своем разнообразии. У Монтэга уже не было прежнего ощущения пустоты – тут было чем наполнить себя. И отныне так будет всегда. Он брел, спотыкаясь, по сухим листьям. И вдруг в этом новом мире необычного – нечто знакомое. Его нога задела что-то, отозвавшееся глухим звоном. Он пошарил рукой в траве – в одну сторону, в другую. Железнодорожные рельсы. Рельсы, ведущие прочь от города, сквозь рощи и леса, ржавые рельсы заброшенного железнодорожного пути. Путь, по которому ему надо идти. Это было то единственно знакомое среди новизны, тот магический талисман, который ещё понадобится ему на первых порах, которого он сможет коснуться рукой, почувствовать всё время под ногами, пока будет идти через заросли куманики, через море запахов и ощущений, сквозь шорох и шепот леса.]

И	вдруг	в	этом	новом	мире	необычного –
<i>i</i>	<i>vdrug</i>	<i>v</i>	<i>ètom</i>	<i>novom</i>	<i>mire</i>	<i>neobyčnogo –</i>
PART	ADV	PREP	DEM	ADJ.PRE	ADJ.PRE	ADJ.GEN
<i>and</i>	<i>suddenly</i>	<i>in</i>	<i>this</i>	<i>new</i>	<i>world</i>	<i>of_unusual –</i>

нечто	знакомое.
<i>nečto</i>	<i>znakomoe</i>
PRO	ADJ.NOM
<i>something</i>	<i>familiar</i>

Example (i) constitutes a precise character description, with the ‘scarf’ of secondary importance to the description of the eyes. In Russian, the nominative case on ‘смелые серые глаза’ (*smelye serye glaza*; lit. ‘brave grey eyes’) marks their privileged role in the sentence, while the *non*-nominative case on the description of the scarf makes it instrumental to the description.

i. English

(TD E1: 1629123)

[A girl appeared timidly at the door, her dark fingers fidgeting with the frill of her apron. “Come in! Come in! She’s shy,” the mother encouraged her, smiling through her tears. Grigory looked at her. Bold grey eyes under a black lace scarf. A small, rosy dimple in the supple cheek. Grigory turned his eyes to her hands: they were large and marred with hard work.]

Bold grey eyes under a black lace scarf.

Russian

(TD RO: 124815)

[В дверях несмело стала невеста, смуглыми пальцами суетливо перебирая оборку фартука. – Пройди, пройди! Ишь засовестилась, – подбодрила мать и улыбнулась сквозь слезную муть. Григорий, сидевший возле тяжелого – в голубых слинялых цветах – сундука, глянул на неё. Под черной стоячей пылью коклюшкового шарфа смелые серые глаза. На упругой щеке дрожала от смущения и сдержанной улыбки неглубокая розовеющая ямка. Григорий перевел взгляд на руки: большие, раздавленные работой.]

Под	черной	стоячей	пылью	коклюшкового	шарфа
<i>pod</i>	<i>černoj</i>	<i>stojačej</i>	<i>pyl’ju</i>	<i>kokljuškovogo</i>	<i>šarfa</i>
PREP	NN.INS	ADJ.INS	NN.INS	ADJ.GEN	NN.GEN
<i>under</i>	<i>black</i>	<i>stagnant</i>	<i>dust</i>	<i>of_lace</i>	<i>scarf</i>

смелые	серые	глаза.
<i>smelye</i>	<i>serye</i>	<i>glaza</i>
ADJ.NOM.PL	ADJ.NOM.PL	NN.NOM.PL
<i>black</i>	<i>grey</i>	<i>eyes</i>

Finally, the description in (j) is of the behaviour (or rather opinion) which is judged to be an inappropriate sample of the class to which the target of the criticism belongs. The indefinite reference creates distance between the target (i.e. ‘*member of the clergy*’ in BK_E2; ‘*churchman*’ in BK_E1) and the class (i.e. the class of members of the clergy; the class of churchmen).

It appears that in Russian, it is the *non-nominative* case that creates the distance between the target of the criticism and the class, i.e. ‘для духовного лица’ (lit. ‘*for soulful.GEN individual.GEN*’) does not state that the target is a ‘духовное лицо’ (lit. ‘*soulful.NOM individual.NOM*’), rather it states what is inappropriate for members of this class and implicates that the target is a member and in so doing places a distance between the target and the class.

[‘Playing with words,’ Father Païsy again interrupted, unable to contain himself. ‘Most unworthy of a member of the clergy. I have read the book with which you take issue,’ he said, turning to Ivan Fyodorovich, ‘and am astounded that a churchman could claim that “the Church is a kingdom not of this world”.’]

Most unworthy of a member of the clergy.

[“Incidentally, he replies to the following ‘basic and essential’ propositions of his opponent, who, mind you, is a churchman. First, that ‘no social organization can or should arrogate to itself the power to dispose of the civil and political rights of its members.’ Second, that ‘criminal and civil jurisdiction should not belong to the Church and are incompatible with its nature both as divine institution and as an organization of men for religious purposes.’ And finally, third, that ‘the Church is a kingdom not of this world ...’ ”

“A most unworthy play on words for a churchman!” Father Paissy, unable to restrain himself, interrupted again. “I have read this book to which you objected,” he addressed Ivan Fyodorovich, “and was astonished by this churchman saying ‘the Church is a kingdom not of this world.’”]

A most unworthy play on words for a churchman!

[Они отвечают между прочим на следующие «основные и существенные» положения своего противника, духовного лица, заметьте себе. Первое: что «ни один общественный союз не может и не должен присвоивать себе власть – распоряжаться гражданскими и политическими правами своих членов». Второе: что «уголовная и судно-гражданская власть не должна принадлежать церкви и не совместима с природой её и как божественного установления, и как союза людей для религиозных целей » и наконец, в-третьих: что «церковь есть царство не от мира сего» ... – Недостойнейшая игра слов для духовного лица! – не вытерпел и прервал опять отец Паисий. – Я читал эту книгу, на которую вы возражали, – обратился он к Ивану Федоровичу, – и удивлен был словами духовного лица, что «церковь есть царство не от мира сего».]

Недостойнейшая	игра	слов	для	духовного	лица.
<i>nedostojnešaja</i>	<i>igra</i>	<i>slov</i>	<i>dlja</i>	<i>duxovnogo</i>	<i>lica</i>
ADJ.NOM	NN.NOM	NN.GEN	PREP	ADJ.GEN	NN.GEN
<i>unworthy</i>	<i>game</i>	<i>words</i>	<i>for</i>	<i>soulful</i>	<i>individual</i>

The above observations require further annotation in terms of the different types of indefinite reference and the potential ways they are expressed in the two languages; as well as the examination of the existing verbless sentences with definite reference, strongly

uncharacteristic as it may be (i.e. lemma ‘*the*’; ind. -13; 46612; 778), it is nevertheless not entirely excluded by verbless sentences and may perhaps pertain to a specific type and reveal more about reference in verbless sentences. The presented examples of indefinite reference in this section were selected to illustrate the range of functions that are characteristic of the statistical preference of the verbless sentence for indefinite reference.

2.3 CONVENTIONALIZATION AND REPEATED SEGMENTS

An important question that arises with regard to the theoretical interpretation of verbless sentences concerns the extent to which their meaning may be considered conventionalized. Are there verbless sentence types that might be considered constructions and thus be *conventionally*, as opposed to *conversationally*, implicated? In other words, could certain verbless clause types favour conventional implicature?

In this section, we propose that the repeated segments calculation provides insight into this theoretical question. As mentioned, this calculation reveals the n-grams that are statistically key for verbless sentences in comparison to the reference corpus.

Our argument, laid out in more detail in *Part 1*, is that the meaning of the verbless sentences arises in a much similar way as that of a verbal sentence. There is first a *basic assertion* in both verbal and verbless sentences: it consists of the assertion that the lexical elements that constitute the sentence are *relevant* between themselves and to the context. Secondly, the *necessary conversational implicatures* step in for both verbal and verbless sentences. At this point, in the verbless case, conversational implicature also applies to the predicate, whereas in the verbal sentences the predicate is semantically encoded but other elements require conversational implicature (e.g. the resolution of indexicals, particular nuances concerning the meaning of words). Therefore, the verbal sentence does not encode all of its meaning either, but unlike the verbless sentence, it does encode the predicate.

The step from conversational to conventional implicature then occurs in the same way for both verbless and verbal sentences. For both verbless and verbal sentences, if the same structure is used to *conversationally* implicate the same meaning consistently, that meaning becomes *conventionalized*. Thus, in terms of conventional implicature and constructions, the

verbless sentences have the same ‘fate’ as verbal sentences: they need frequency to be conventionalized.

It is the present argument that it does not make sense to treat verbless sentences, as a whole class, as conventionalized constructions simply because they lack the predicate: the meaning of the implicated predicate can change. However, verbless sentences may indeed be conventionalized, just like verbal sentences may be conventionalized. For instance, a certain structure may be conventionalized as ‘motion’ or even ‘present time’ through frequent use in contexts that suggest such an implicature. That said, to refer to the absence of the verb from a sentence as *necessarily* indicating conventionalization, as is sometimes done, is not a good approach because it disregards the fact that the meaning can change. Verbless sentences are productive. Just like verbal sentences they first conversationally implicate and, after frequent use, they may gain conventional implicatures, i.e. a consistent form-function pairing. Certain verbless sentences may constitute conventional implicatures, i.e. constructions, but this does not apply to all verbless sentences, in the same respect as the existence of some constructions among verbal sentences does not make all verbal sentences constructions. (Although, on some accounts, any and every structure is a construction, but even then, it seems there would need to be some distinctions in terms of the level of conventionalization.)

It seems the analysis of verbless sentences as necessarily constituting constructions may to some extent be influenced by two things. First, the type of data that has been previously used, i.e. when one uses fragments from here and there, phrases that stand out, road signs, fridge magnets, etc., it is inevitable that one will get fixed meanings. Secondly, treating all verbless sentences as constructions does not require giving them sentential status and is thus a somewhat convenient position with regard to traditional syntactic conception of the sentence. From the moment that verbless sentences are considered non-conventionalized, an explanation becomes necessary for how it is that they may have a sentential meaning without a syntactic predicate: we are thus led either into hidden structure, or attributing verbal-inflection categories to non-verbal lexical elements, or into admitting that the existence of verbless sentences indicates precisely that syntactic models do not explain the sentence. It is the latter that we are presently arguing for.

The repeated segments calculation, presently carried out, adds interesting evidence both in terms of showing that verbless sentences *overall* are *not* conventional implicatures, and in terms of revealing *which* structures may be conventional implicatures and deserve more attention in this direction. These results are partially presented in *Appendix 7*.

First, the results of this calculation provide proof against treating overall verbless sentences as conventionalized structures. If verbless sentences were to be fixed form-meaning pairings, we should see entire verbless sentences in the list of the repeated segments (i.e. n-grams) that statistically set verbless sentences apart from the rest of the corpus. However, what we see is only *parts* that are repeated. The results, of which the first 30 lines are presented in *Appendix 7*, reveal statistically key lemma n-grams such as in (109a) for English verbless sentences, and, for instance, the lemma n-grams in (109b) for Russian verbless sentences.

(109) *Repeated segments that characterize verbless sentences*

- a. ‘*what about*’ (ind. 168; 110; 90)
‘*all right*’ (ind. 112; 385; 97)
‘*what a*’ (ind. 54; 116; 40)
‘*how about*’ (ind. 51; 38; 28)
‘*just a*’ (29; 113; 25)
- b. ‘у вы’ (*u vy*; lit. ‘*at you.2PL*’; ind. 47; 120; 53)
‘а что’ (*a čto*; lit. ‘*and what*’; ind. 46; 136; 55)
‘у ты’ (*u ty*; lit. ‘*at you.2PL*’; ind. 47; 120; 53)
‘что такой’ (*čto takoj*; lit. ‘*what like_this*’; ind. 40; 84; 42)
‘ты что’ (*ty čto*; lit. ‘*you what*’; ind. 38; 116; 46)

It is important to note that these, as well as most of the segments on the list (see *Appendix 7*), are parts of verbless sentences, not entire sentences. This is similar to the way that in verbal sentences one might find ‘*it be*’, ‘*if you*’, ‘*to the*’, ‘*by the way*’ or other repeated segments, but not entire verbal sentences in the repeated segments list.

Even the forms on the list that may appear at first glance to constitute sentences require a much closer manual analysis before they may be labeled potentially *conventionalized verbless sentence forms*, as opposed to being frequent *parts of* productive verbless sentences. For instance, the search for the English lemma segment ‘*all right*’ (ind. 112; 385; 97) within verbless sentences revealed not only instances where the segment constitutes as sentence, but also a range of occurrences, presented in (110), where the segment is an element within a verbless sentence.

(110) *English lemma n-gram ‘all right’ in verbless sentences*

- a. Oh, all right then. (TD_E1)
- b. All right, so far. (TD_E1)
- c. You all right? (FH_EO)
- d. All right, off with you. (TD_E1)

- e. Warm and tea-coloured, all right, but like dishwater. (OD_E2)
- f. Better than alright, perfect! (FH_EO)
- g. Journey all right, Harry? (HP_EO)
- h. All right – squirt! (HP_EO)

The fact that most of the repeated segments that are specific to verbless sentences *do not* constitute a verbless sentences, but are only parts, is evidence that the verbless sentences, should *not automatically be treated as conventionalized structures any more than verbal* sentences.

Secondly, the repeated segments calculation, which shows the segments (n-grams) that make verbless sentences different from verbal sentences, would reveal if there are certain segments within verbless sentences (i.e. consistent forms) that are fixed. These may then be analyzed as potential conventionalized *forms* of verbless sentences and explored in terms of whether or not they have a fixed *meaning* attached to them, and what that meaning, or set of meanings, may be, in further studies. (Using n-grams as part of the identification of constructions is an approach also taken in Shibuya & Jensen, 2015; Cappelle & Grabar, 2016.)

The identification of the verbless structures that may have conventionalized meaning requires the search of each of the lemma n-grams and examination of the verbless sentences containing them in all of the texts. This analysis for the entire list of n-grams is beyond the present scope.

Nevertheless, for each language, the top six n-grams which are statistically key for verbless sentences were examined (cf. the first six lines of the Russian and English tables in *Appendix 7*). While the top six Russian segments showed a lot of variation (i.e. segments constitute various parts of other verbless sentences), the top six English segments revealed two fixed forms. Therefore, we are presently able to draw attention to the English verbless sentence forms in (111).

- (111) *Potential conventionalized verbless sentence forms*
- a. ‘*why not*’ (ind. 67; 49; 37)
 - b. ‘*what for*’ (ind. 49; 32; 26)

The key n-grams in (111) were examined in the verbless sentences of corpus and found to typically constitute full verbless sentences.

The English segment (111a) ‘*why not*’ occurred 49 times in the corpus and 37 in verbless sentences. The verbless instances all constituted questions. Furthermore, the sentence form showed very little variation: apart from ‘*Why not?*’ only the forms in (112) were observed.

(112)	Well, why not?	(BK_E1)
	And why not?	(OD_E1; AF_EO; FH_EO)
	Yeah, why not?	(OS_E1)
	Why not “The Gypsy Baron” or “The Blue Mazurka”?	(OS_E1)
	Why not me?	(SR_EO)
	Why ... not?	(CA_EO)

Concerning the English key segment (111b), out of the 32 occurrences of ‘*what for*’ in the corpus, 26 were in verbless sentences. All of the verbless sentences also constituted questions and showed very little variation in form from ‘*What for?*’, apart from two occurrences of ‘*But what for?*’ (TD_E1) and ‘*But what for, Comrade Warder?*’ (OD_E1).

These two particular forms may constitute verbless sentence constructions and deserve further exploration and annotation with regard to their possible functional pairings.

Although verbless sentences are possible as construction, it must be emphasized that the meaning of verbless sentences extends beyond the traditional conventionalized role that is often attributed to the structures based on their noncanonical syntax.

Therefore, to summarize an answer to the question this section began with: some types of verbless sentences (such as for instance ‘*Why not?*’ and ‘*What for?*’) may be conventionalized, but on the whole, verbless sentences are not necessarily so. The identification of specific conventionalized constructions, specific types, within the overall category of verbless sentences requires the same approach as verbal sentences: frequency. In the present section, we have proposed, and used, the repeated segments calculation to suggest what some potentially conventionalized verbless sentence forms might be.

2.4 LENGTH AND LEXICAL COMPLEXITY

2.4.1 Quantifying ‘Short’

It is generally assumed that verbless sentences are short or shorter than verbal sentences. At present, it is possible to be a bit more specific about the length of the verbless sentence and

quantify what is meant by ‘short’ and ‘shorter than’. *Table 20* provides the relevant values for the present calculation of sentence length.

TEXT	ALL SENTENCES			VERBLESS SENTENCES		
	Words per Full Text	Sentences per Full Text	Words per Sentence	Words per Verbless Sentence Text Part	Verbless Sentences per Text	Words per Verbless Sentence
	<i>(tokens abs. fq.)</i>	<i>(abs. fq.)</i>	<i>(tokens)</i>	<i>(tokens abs. fq.)</i>	<i>(abs. fq.)</i>	<i>(tokens)</i>
BK_RO	33984	2367	14.4	1377	295	4.7
BK_E1	42853	2459	17.4	491	139	3.5
BK_E2	44225	2401	18.4	425	114	3.7
TD_RO	93641	11786	7.9	6188	2099	2.9
TD_E1	130476	11564	11.3	2782	1007	2.8
OD_RO	33364	3587	9.3	2116	624	3.4
OD_E1	53281	3987	13.4	549	188	2.9
OD_E2	49982	3891	12.8	1025	323	3.2
PO_RO	48143	5768	8.3	3945	1190	3.3
PO_E1	58858	5840	10.1	2025	697	2.9
KF_RO	37635	5551	6.8	3502	1213	2.9
KF_E1	54440	5539	9.8	1604	611	2.6
OS_RO	15892	3182	5.0	3309	1181	2.8
OS_E1	26422	3491	7.6	1389	528	2.6
ST_E1F	37283	2418	15.4	170	75	2.3
ST_E2F	34254	2493	13.7	133	67	2.0
ST_R1F	25659	2472	10.4	526	148	3.6
ST_R2F	25363	2518	10.1	635	172	3.7
SR_EO	70396	8973	7.8	2078	953	2.2
SR_R1	55358	8899	6.2	4721	1706	2.8
AF_EO	30486	1706	17.9	139	50	2.8
AF_R1	23813	1818	13.1	485	119	4.1
AF_R2	21598	1669	12.9	423	106	4.0
AF_R3	24069	1747	13.8	392	102	3.8
FH_EO	48107	4490	10.7	1539	578	2.7
FH_R1	42271	5064	8.3	2976	982	3.0
JS_EO	9410	733	12.8	157	55	2.9
JS_R1	9370	927	10.1	622	163	3.8
HP_EO	100491	8592	11.7	1845	815	2.3
HP_R1	80381	9544	8.4	4877	1533	3.2
CA_EO	15911	2021	7.9	756	381	2.0
CA_R1	11006	2031	5.4	1928	721	2.7
Total RU	581547	68930	8.4	38022	12354	3.1
Total EN	806875	70598	11.4	17107	6581	2.6

Table 20. Sentence Length

The average length of a verbless sentence is **3.1 words in Russian** and slightly shorter in **English at 2.6 words**.

Notably, the length was determined by dividing the number of word tokens in the corpus part which contains only verbless sentences by the number of verbless sentences in that part. (In other words, the length of each sentence was not measured individually, rather the analysis presents the calculated average length based on the number of words and sentences per relevant part.)

Figure 27 compares the average length of a *verbless* sentence in Russian to the length of an average sentence in the full Russian corpus, and the length of an average *verbless* sentence in English to that of an average sentence in the English corpus.

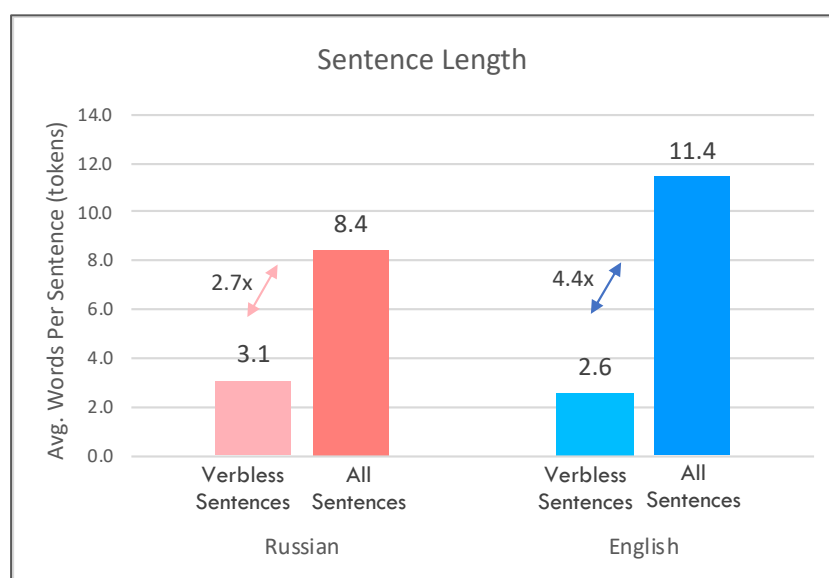


Figure 27. Average sentence length per language

As shown in the figure, in Russian, the verbless sentence is **2.7 times shorter than** the average Russian sentence. In English, the verbless sentence is **4.4 times shorter than** the average English sentence.

Comparing the average verbless sentence lengths across (i) texts, (ii) language-source (i.e. originals translations, third-language translations) and (iii) genre (i.e. discourse-based fiction, scripted speech fiction, narrative-based fiction) does not reveal important variations concerning the length of the verbless sentence.

Concerning (i), Figure 28 illustrates the average *verbless* sentence length *per text*. The average verbless sentence length in the Russian texts is calculated to range from 2.7 to 4.7 average words per verbless sentence (with a mean at 3.1 words), and in the English texts from 2.0 to 3.7 average words per verbless sentence (with a mean at 2.6 words).

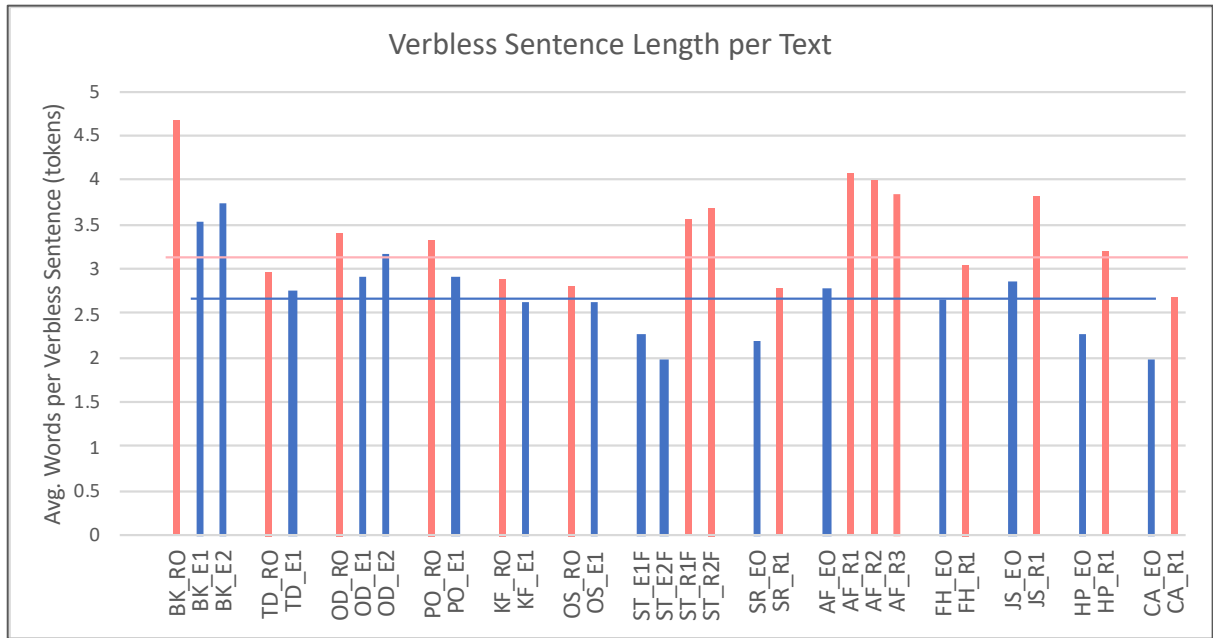


Figure 28. Average verbless sentence length per text

Concerning (ii), *Figure 29* shows the average sentence length *per language-source type*. It illustrates that the average *verbless* sentence length across *Originals*, *Translations*, and *Third-Language Translations* is stable. This is the case despite a slight increase in the length of the average sentence across the language-source type.

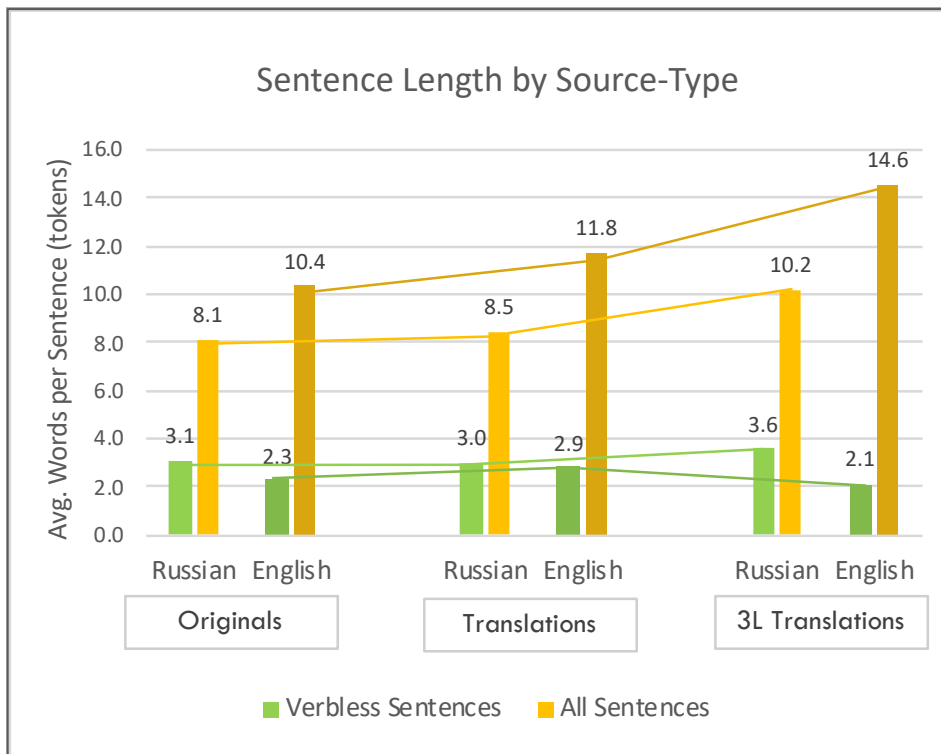


Figure 29. Average sentence length per source-type

The average verbless sentence length *per genre* is revealed in *Figure 30*. It is notable that even for the scripted-speech genre, where the average sentence length is shorter relative to the other genres, the length of the *verbless* sentence remains stable. This is illustrated in *Figure 30* which compares the average length of a *verbless* sentence to the length of an average sentence in the particular *genre* in each language.

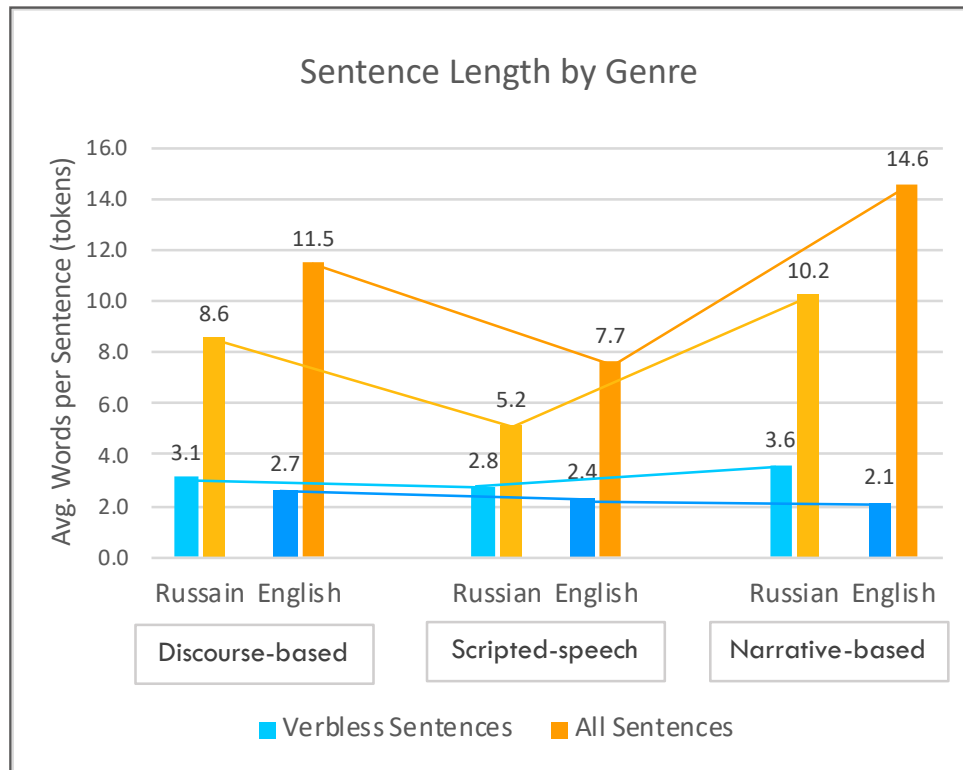


Figure 30. Average sentence length per genre

2.4.2 Lexical Complexity

In spite of their shorter length, verbless sentences reveal surprising results concerning their lexical complexity compared to the average sentence. *Table 21* provides the basic type-token ratio (TTR) for verbless sentences. Notably, the present ratio was determined by dividing the number of types in the verbless sentence corpus subpart by the number of tokens in that same subpart.

Corpus Part	Tokens	Types	Type-Token Ratio
English Verbless Sentences	17107	4091	0.24
Russian Verbless Sentences	38022	10570	0.28

Table 21. Type-token values for verbless sentence subpart

As the overall corpus is much larger than the verbless sentence sub-part, it would be inappropriate to directly compare their type-token values. For larger sized texts, the type-token ratio will necessarily be smaller. A moving-average type-token ratio (MATTR), which adjusts for size, would provide a solution (e.g. Kettunen, 2014), but it is beyond the present scope. Nevertheless, it was possible to determine the type token ratio *per text*. These results are given in *Table 22*.

Text	Tokens	Types	Type-Token Ratio
<i>English</i>			
JS_EO	9410	1694	0.18
CA_EO	15911	1722	0.11
OS_E1	26422	3277	0.12
AF_EO	30486	4223	0.14
ST_E2F	34254	3617	0.11
ST_E1F	37283	4341	0.12
BK_E1	42853	4982	0.12
BK_E2	44225	5580	0.13
FH_EO	48107	5563	0.12
OD_E2	49982	5306	0.11
OD_E1	53281	3820	0.07
KF_E1	54440	6910	0.13
PO_E1	58858	6298	0.11
SR_EO	70396	5411	0.08
HP_EO	100491	8942	0.09
TD_E1	130476	11224	0.09
<i>Russian</i>			
JS_R1	9370	3527	0.38
CA_R1	11006	2917	0.27
OS_RO	15892	5473	0.34
AF_R2	21598	8273	0.38
AF_R1	23813	8223	0.35
AF_R3	24069	7622	0.32
ST_R2F	25363	7110	0.28
ST_R1F	25659	6981	0.27
OD_RO	33364	10006	0.30
BK_RO	33984	9381	0.28
KF_RO	37635	14839	0.39
FH_R1	42271	11775	0.28
PO_RO	48143	13402	0.28
SR_R1	55358	12036	0.22
HP_R1	80381	18456	0.23
TD_RO	93641	29843	0.32

Table 22. Type-token values for full corpus per text

Considering only texts of similar token size to the verbless sub-part (i.e. 38,022 tokens in Russian and 17,107 tokens in English) the following observations are revealed.

Comparing the type-token ratio of English verbless sentences (0.24 TTR, based on 17,107 tokens in the verbless sentence subpart) to that of the average English sentence in texts which have a similar size, it is notable that the verbless sentences have a *greater lexical complexity*. Specifically, the verbless ratio is higher than that of the average sentence in CA_EO (0.11 TTR, based on 15,911 tokens), OS_E1 (0.12 TTR, based on 26,422 tokens), and even the smaller sized JS_EO (0.18 TTR, based on 9,410 tokens).

For Russian, the type-token ratio of verbless sentences (0.28 TTR, based on 38,022 tokens) is similar that of the average sentence in texts of similar size. Specifically, it is the same as the average sentence in BK_RO (0.28 TTR, based on 33,984 tokens), and slightly lower than the average sentence in OD_RO (0.30 TTR, based on 33,364 tokens) and KF_RO (0.39 TTR, based on 37,635 tokens).

Such results are striking because mere fragments and reductions of verbal sentences, as verbless sentences are sometimes treated, are not expected to have higher lexical complexity than the average sentence. A higher type-token ratio indicates that the text part is more dense in terms of the vocabulary that is used, and therefore more informationally rich.

The present comparison suggests that there is interest in performing the more elaborate type-token computation on the other texts. The lexical complexity of verbless sentences provides empirical evidence in support of the argument that verbless sentences are semantically independent structures, not leftovers from verbal sentences.

2.5 STATISTICAL PORTRAIT OF THE VERBLESS SENTENCE

The present portrait of the prototypical verbless sentence is still at this point a sketch in very broad strokes. It requires refinement through an important amount of manual annotation of the specific elements and n-grams in order to semantically disambiguate the uses, in addition to the developments already mentioned in this present section and in the appendices. Nevertheless, it is hoped that the present statistical perspective reveals a few of the lines in which further manual annotation would likely be fruitful.

CHAPTER 3

TRANSLATION LANGUAGE AND GENRE

The present chapter explores the question of translation language and textual genre. What influence does translated language have on verbless sentences? How does textual genre affect verbless sentence use? The verbless sentence is presently shown to be a linguistic phenomenon that is language-specific and *not* significantly affected by translation from both a *monolingual* and a *parallel* perspective. Verbless sentences are found to be more affected by the difference between speech and narration than the source-type, i.e. whether the source of the language is an original or a translation.

3.1 SOURCE-TYPE

3.1.1 Monolingual Approach

To explore the difference between originals and translations with regard to the verbless phenomenon, we take two perspectives. First, from a monolingual perspective one may compare the frequencies in the original texts of a language to the translations in that same language. This perspective is presented in *Figure 31* for Russian.

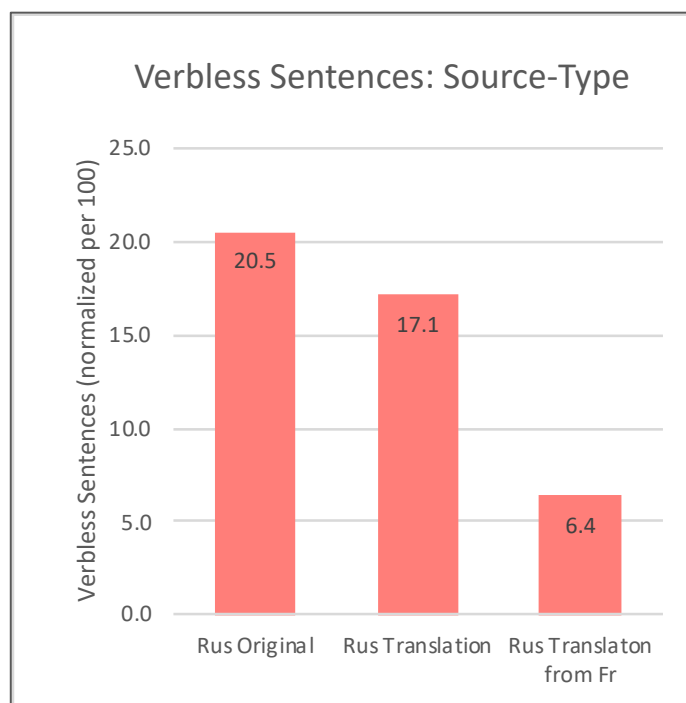


Figure 31. Russian verbless sentences in originals and translations

The figure shows the relative frequency of verbless sentences in Russian texts (normalized per 100 sentences). It shows that in texts where Russian is the source language, the proportion of verbless sentences is 20.5 per 100 sentences. In texts where Russian is the language of translation, 17.1 out of 100 sentences are verbless. For Russian translations from French, 6.4 out of 100 sentences were verbless.

As discussed in *Part 3*, the latter category, i.e. ‘Russian translations from French’, is a small sized corpus-part, and even though it has been normalized in the above figure, this corpus-part of third-language translation also overlaps with the only supply of narrative-based fiction genre in the present corpus. For this reason, it is for the moment set aside. In the next part it will be suggested that it is likely genre, as opposed to the language source-type, that accounts for the low frequency in this sub-part.

As discussed in *Part 3*, the Russian original and Russian translation corpus parts are deemed to be comparable in terms of word-size, number of texts, genre and time period.

Figure 32 shows the relative frequency of verbless sentences in English (normalized per 100 sentences) across originals and translations.

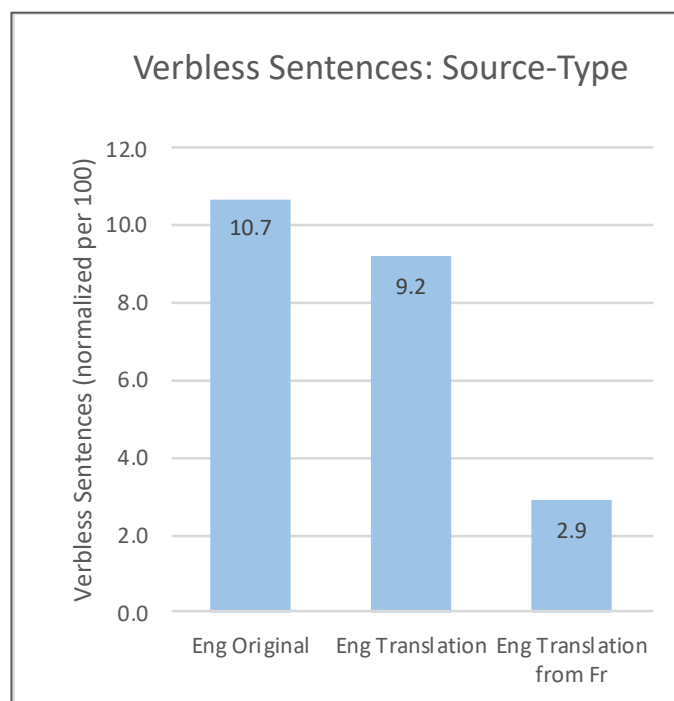


Figure 32. English verbless sentences in originals and translations

In texts where English is the source language, the proportion of verbless sentences is 10.7 per 100 sentences. In texts where English is the language of translation, 9.2 sentences are verbless. For English translations from French, 2.9 out of 100 sentences are verbless, but the same comments made in the above paragraph for Russian also apply to this small sub-section that overlaps with narrative-based genre, and it is therefore set aside for the moment.

It may be noted that when originals and translations are compared within the same language, a slight variation exists. Translations have slightly less verbless sentences than originals both in Russian and in English.

Russian translations from English show 3.4 verbless sentences less than Russian originals. Thus, if we take the 20.5 verbless sentences in original texts as the standard, Russian translations show 16.6% less verbless sentences than Russian originals.

English translations from Russian show 1.5 verbless sentences less than English originals. Thus, if we take the 10.7 verbless sentences in original texts as the standard, English translations show 14% less verbless sentences than English originals.

From the perspective of the overall Russian corpus, 3.4% of sentences that, roughly speaking, would have been verbless in original-Russian are *verbal in translated-Russian*. In English, 1.5% of sentences that would have been verbless in original-English are *verbal in translated-English*.

Thus, from a monolingual perspective the variation in verbless sentence frequency in originals and translations is slight but it does exist.

3.1.2 Parallel Approach

What is striking is that when we approach the question of original-translation differences from a parallel perspective, the variation in verbless sentences between originals and translations virtually disappears. This contrastive perspective is presented in *Figure 33*.

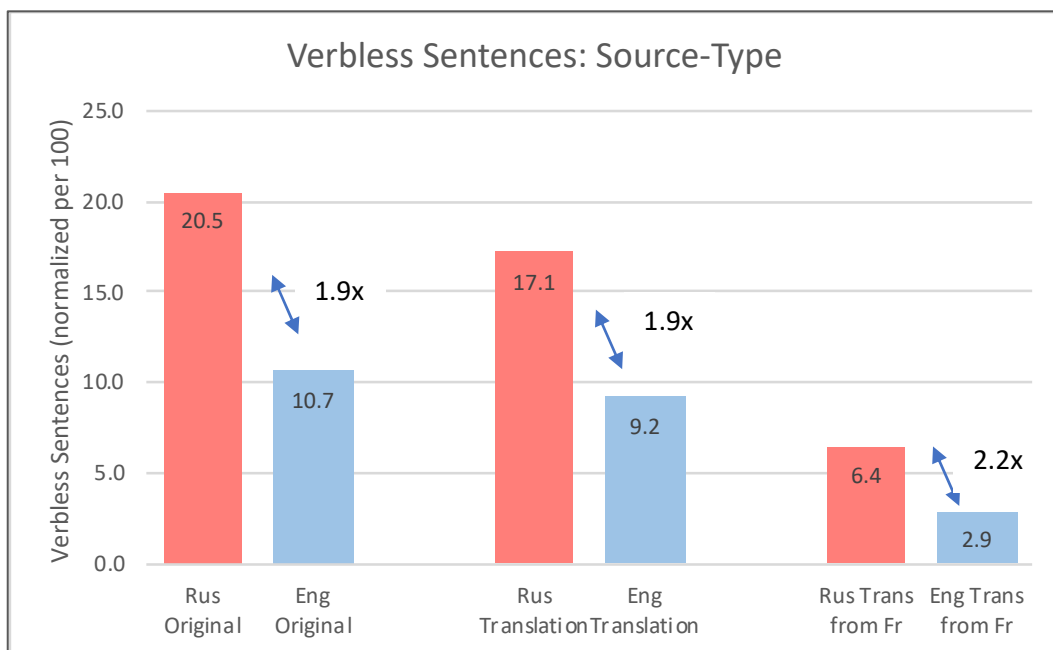


Figure 33. Contrastive perspective on originals and translations

As made evident in the figure, when Russian and English are compared to each other, Russian surpasses English by two regardless whether the source-type is an original or a translation.

More precisely, when the normalized frequency of verbless sentences in Russian originals (i.e. 20.5 verbless per 100 sentences) is compared to that of English originals (i.e. 10.7 verbless per 100 sentences), Russian surpasses English by 9.8 sentences, or *1.92 times more than the number of English verbless sentences*. When the frequency of verbless sentences in Russian translations (i.e. 17.1 verbless per 100 sentences) is compared to that of English

translations (i.e. 9.2 verbless per 100 sentences), Russian again surpasses English by 7.9 sentences, or *1.86 times more than English*.

Even Russian-translations-from-French and English-translations-from-French are revealed to behave similarly to originals and translations when a contrastive perspective is taken. When the frequency of verbless sentences in Russian-translations-from-French (i.e. 6.4 verbless sentences per 100 sentences) is compared to that of English-translations-from-French (i.e. 2.9 verbless sentences per 100 sentences), Russian is revealed to surpasses English by 3.5 sentences, or 2.2 times more than English (i.e. only slightly higher than originals and translations, and not nearly as drastic of a difference as when this this sub-part is considered from a monolingual perspective).

What all of this suggests is that despite the existence of slight differences in the frequency of the phenomenon in monolingual originals and translations, there are stable linguistic rules behind the use of verbless sentences in Russian and English which are not affected by the source language. If for instance, Russian translations of English were to blindly imitate the English source and translate the English verbal sentences as also verbal, one should observe an important difference between the proportions of verbless sentences in Russian and English originals and the proportions of verbless sentences in Russian and English translations (i.e. Russian translations, influenced by the English source, should then show less verbless sentences, while English translations, influenced by the Russian source, would then have a higher proportion of verbless sentences, thereby increasing the gap between the two languages for translations compared to the gap between the two languages for originals). This is clearly not what is observed.

By the present counts, the proportion between originals and translations is stable. Source language interference in terms of verbal-verbless distinction could potentially pertain to only 3.4% of Russian sentences and 1.5% of English sentences.

This constitutes an empirical indication that we can safely reject the argument that translations behave differently from originals with regard to the verbal sentence phenomenon (at least in terms of the verbal-verbless variation). Roughly speaking, Russian originals have twice the verbless sentences of English originals, and Russian translations have twice the verbless sentences of English translations.

Therefore, we may conclude that the phenomenon of the verbless sentences, even when observed in translations, is a *linguistic* one, and *not a translational* one. This is an important factor to establish for a contrastive study, such as the present, since, as discussed in Par 2, contrastive analysis that works with translations is very frequently obliged to defend itself

against the argument that translations constitute a type of language in which linguistic phenomenon may drown under the weight of translational features, i.e. that what is being described is translation related as opposed to linguistically related. When it comes to the verbless sentence phenomenon, the present results suggest that the important concerns with ‘translationese’ have received an adequate response. There certainly exist general differences between originals and translations, perhaps some are systematic, yet, when it comes to the presence or absence of the verb it is the linguistic factors that overpower potential translation universals. Thus, the presence or absence of the verb in a sentence is one for linguistic study whether in originals or translations.

3.2 GENRE

Where there does appear to be a difference with regard to the use of verbless sentences is not source-type related, but rather genre-related. A correlation between verbless sentences and genre becomes apparent both from a monolingual perspective and a contrastive one.

3.2.1 Monolingual Approach

Figure 34 illustrates the monolingual perspective on verbless sentences in Russian across the three types of genre in the present corpus. It summarizes the relative frequency of verbless sentences in Russian texts (normalized per 100 sentences) per genre.

In discourse-based fiction, the proportion of Russian verbless sentences is 17.3 per 100 sentences. In the scripted-speech texts, 36.5 out of 100 sentences are verbless. Finally, in the narrative-based fiction genre, 6.4 out of 100 Russian sentences were verbless.

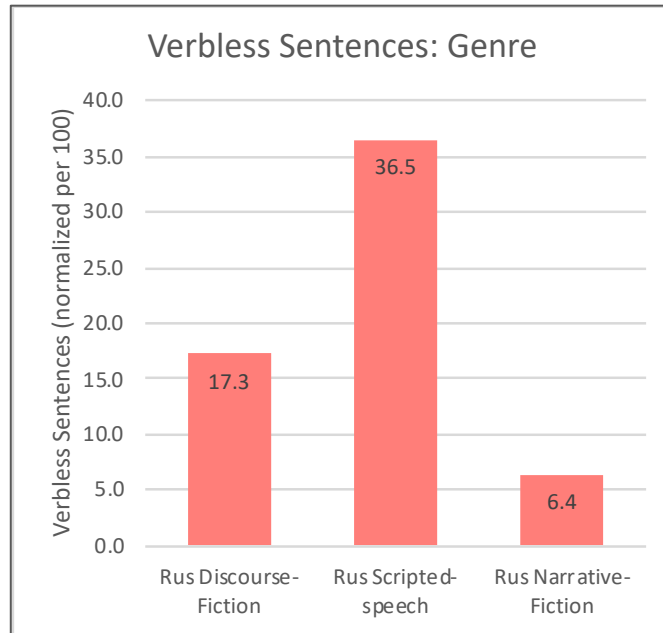


Figure 34. Russian verbless sentences across genre

Figure 35 shows the relative frequency of verbless sentences in the English texts per genre (normalized per 100 sentences).

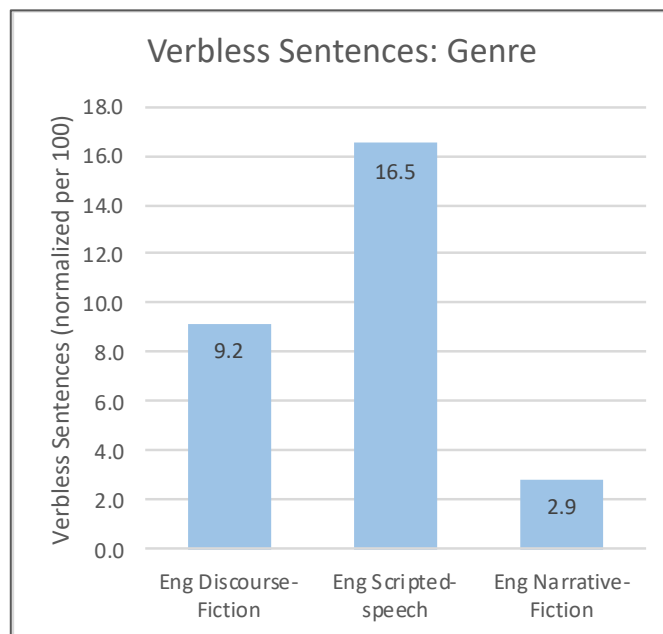


Figure 35. English verbless sentences across genre

The figure makes evident that in discourse-based fiction, the proportion of English verbless sentences is 9.2 per 100 sentences. In the scripted-speech texts, 16.5 out of 100 sentences are verbless in English. Narrative-based fiction shows 2.9 English verbless sentences per 100 sentences.

As discussed in *Part 3*, the texts in the genres are comparable between Russian and English in terms of number of texts, word-size, time period and source-type per genre *in each language*. However, within a single language there are important differences between the genres the details of which were specified in *Part 3*.

At present, it is notable that the frequency of verbless sentences in the scripted speech genre is much higher than for the other genres, both in English and in Russian. Russian scripted speech shows a verbless sentence frequency that is 2.1 times that of Russian discourse-based fiction. In English, the frequency of verbless sentences in scripted speech is 1.8 times that of discourse-based fiction.

The small sup-part of narrative-based fiction shows a particularly low frequency of verbless sentences in both languages relative to the other genres, i.e. in Russian, it shows a verbless sentence frequency that is 2.7 less than that of discourse-based fiction, and in English, 3.1 times less than that of discourse-based fiction.

3.2.2 Parallel Approach

The variation between genres that is made apparent in the monolingual comparison above is also evident from the parallel perspective. *Figure 36* presents the contrastive perspective on genre.

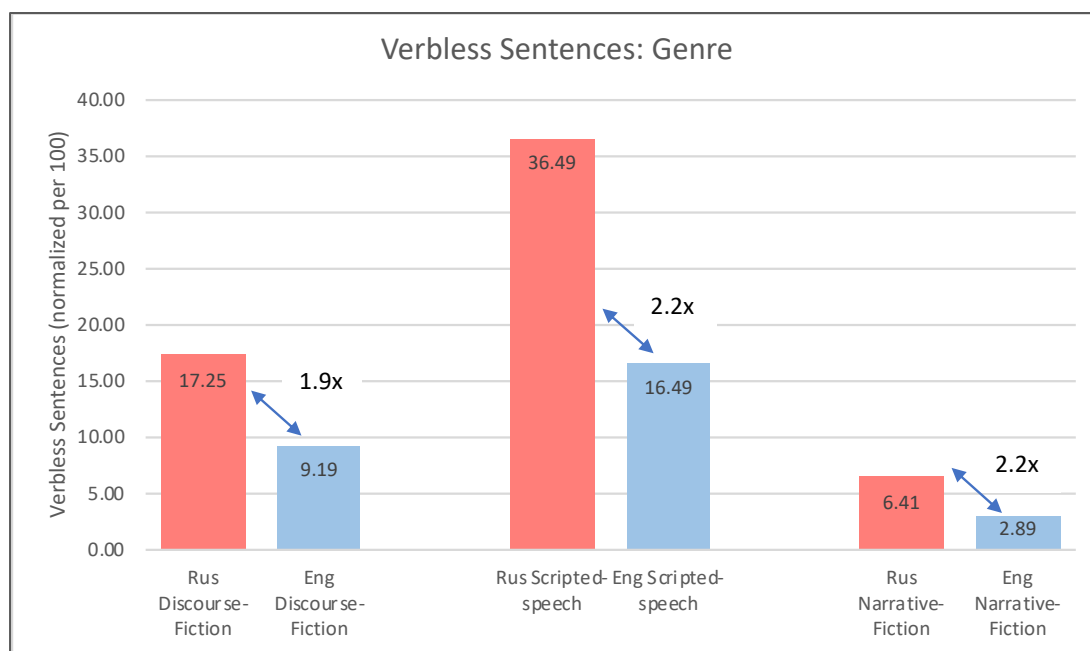


Figure 36. Contrastive perspective on genre

Notably, when the two languages are compared in terms of the number of verbless sentences per genre, it becomes evident that Russian surpasses English in scripted speech to a slightly greater extent than it does in the discourse-based fiction genre. Specifically, in scripted speech Russian shows 2.21 times more verbless sentences than English, while in discourse-based fiction it shows 1.88 times more verbless sentences than English.

Narrative-based fiction, while a much weaker source of verbless sentences in both languages, contrastively shows 2.2 times more verbless sentences in Russian than in English. Thus, in terms of variation between the languages, its virtually identical to that of scripted speech. Though, as mention, the size of this corpus part is much smaller and therefore caution must be taken with regard to its analysis.

What the contrastive perspective on genre suggests is that there is something about scripted-speech that leads Russian to rely on verbless sentences more than English relies on verbless sentences in this genre. If verbless sentences are slightly more important for Russian scripted speech than they are for English scripted speech, it may be that there is a pragmatic difference behind these frequencies that explains why verbless sentences might be more needed in Russian scripted speech than in English scripted speech. To explore this difference further it is necessary to examine more works in the scripted speech sub-part. For now, what we note is that in scripted-speech the difference between Russian and English is slightly more pronounced in terms of verbless sentences.

Taken together, the monolingual and the parallel perspectives on genre indicate that verbless sentences contribute to the linguistic distinction between these three types of genre. The monolingual genre differences in particular, suggest that verbless sentences may be a linguistic feature that helps to establish a functionally-based genre distinction between discourse-based fiction, scripted speech and narrative-based fiction.

3.3 DIRECT SPEECH

Furthermore, manual annotation showed that discourse type is an important factor when it comes to verbless sentence use. This finding was found through the manual annotation of verbless utterances in terms of whether they occur in direct speech or in narration. It was discussed in (Bondarenko, 2019).

The annotation was performed on the verbless utterances of the *Brothers Karamazov* sub-corpus, i.e. BK_RO, BK_E1 and BK_E2, discussed above (*Part 4: Chapter 1: Table 5*). The results showed that, in both languages, 93% of verbless utterances occurred in direct speech, compared to only 7% in narration.

Such a result highlights a pragmatic constraint on the use of verbless sentences in both languages. In direct speech there exists a common ground between the speaker and the addressee, which does not exist in basic narration. It appears that verbless sentences may require a common ground for their use.

This observed correlation with direct speech is in line with the statistical analysis of the constituents of verbless sentences in the previous chapter. As shown in *Part 4: Chapter 2*, as well as in *Appendix 6*, verbless sentences are saturated with elements that require contextual resolution. For instance, the indexical lemma-‘ТЫ’ (*ty*; lit. ‘*you.2SG*’; ind. 13; 35595; 655) and lemma-‘ВЫ’ (*vy*; lit. ‘*you.2PL*’; ind. 45; 2755; 388) are among the top statistically key elements of verbless sentences in Russian. Indexicals also key, but to a lesser extent in English, i.e. the English lemma ‘*your*’ does not make into the top 30 but still has a strong specificity index (‘*your*’; ind. 16; 1999; 102), while the lemma ‘*you*’ even has a slightly negative index in English (‘*you*’; ind. -3; 11178; 203). Furthermore, the high specificity of interjections and question words in both languages also emphasizes interactive contexts of verbless sentence use for both languages.

Taken together, the results point to a strong correlation between verbless sentences and direct speech. This correlation, in turn, highlights the existence of pragmatic requirements on the common ground between interlocutors.

Furthermore, the present results also confirm the choice to focus on spontaneous speech corpora of some of the existing studies of verbless phenomena (e.g. Garcia-Marchena, 2018; Garcia-Marchena, 2016; Landolfi et al., 2010; Fernández et al., 2007; Cresti & Moneglia, 2007; Schlangen, 2005).

CHAPTER 4

INFORMATION STRUCTURE

The present chapter explores the information structure of the verbless sentence. The backbone of the analysis here developed is from our masters studies (Bondarenko, 2014; 2015). Verbless sentences and their translations were annotated for topic, focus and information structure type in accordance with the definitions of Lambrecht (1994), discussed above in *Part 1*. The annotation was manually carried out on the verbless sentences in the Russian *The Brothers Karamazov* (BK_RO) and the Pevear and Volokhonsky English translation (BK_E1). The details of this corpus are presented above in *Table 5*. The present chapter summarizes the key qualitative findings and situates them within the presently proposed model.

4.1 INFORMATION STRUCTURE TYPE

We start with the question of whether verbless sentences may be associated with one of the information structure sentence types discussed by Lambrecht (1994), i.e. the Topic-Comment, Identificational or the Thetic information structure (or ‘focus structure’) type. The findings do not reveal an association between information-structure type and a specific verbless structure. Verbless sentences were found across all three focus structure types.

The Russian and English verbless sentences in (113) illustrate instances of the Topic-Comment type.

(113) *Topic-Comment Verbless Sentence in Russian and English*

a. Russian (BK RO)

[– Молчать! – закричал Дмитрий Федорович, – подождите, пока я выйду, а при мне не смейте марать благороднейшую девицу... Уж одно то, что вы о ней осмеливаетесь заикнуться, позор для нее... Не позволю! Он задыхался. – Митя! Митя! – слабонервно и выдавливая из себя слезы вскричал Федор Павлович, – а родительское-то благословение на что? А ну прокляну, что тогда будет?
– Бесстыдник и притворщик! – неистово рявкнул Дмитрий Федорович.]

Бесстыдник	и	притворщик!
<i>besstydnik</i>	<i>i</i>	<i>prityvorščik</i>
NN.NOM	CONJ	NN.NOM
<i>shameless_man</i>	<i>and</i>	<i>pretender</i>

b. English (BK E1)

["Silence!" Dmitri Fyodorovich shouted. "Wait until I'm gone. Do not dare in my presence to sully the noblest of girls... that you are even so bold as to mention her is shameful enough... I will not allow it!" He was gasping for breath.
"Mitya! Mitya!" Fyodor Pavlovich cried tremulously, trying to squeeze out a tear. "Don't you care about a father's blessing? And what if I should curse you?"
"Shameless impostor!" Dmitri Fyodorovich roared in a fury.]

Shameless impostor!

The topic of the verbless sentence insults in (a) and (b) is the addressee, i.e. Fyodor Pavlovich. Although he is not mentioned in either the Russian nor the English verbless sentence, the addressee is active and presupposed in the minds of both the speaker and the hearer. The verbless sentences can be reasonably interpreted as being about no one other than the addressee. A new piece of information is created from the combination of this presupposed topic and the words '*shameless imposter*' or 'бесстыдник и притворщик' (*besstydnik i prityvorščik*) which constitute the focus of the sentence (i.e. the unpredictable elements). The example therefore constitutes the **Topic-Comment** type of focus structure, even though the topic is not explicitly mentioned.

However, were the context to be different the same sentence could be Identificational, Thetic or ambiguous. For example, if '*Shameless imposter!*' or 'Бесстыдник и притворщик!' (*Besstydnik i prityvorščik!*) was uttered in response to the question '*What is the best thing to call*

him?’, the utterance would be of the **Identificational** type (i.e. providing the missing element in the open proposition ‘*the best thing to call him is X*’).

The information structure of the sentence would also change if Dmitri had screamed ‘*Shameless imposter!*’ or ‘Бесстыдник и притворщик!’ (*Besstydnik i pritvorščik!*) outside of a context in which the addressee can be presupposed as a Topic. For instance, if in the middle of a conversation, say one in which the discourse topic is elephants in Africa, Fyodor Pavlovich suddenly walked into the room and Dmitri shouted ‘*Shameless imposter!*’ or ‘Бесстыдник и притворщик!’ (*Besstydnik i pritvorščik!*). Here, the verbless sentences would be announcing the event of the arrival of someone who is a shameless impostor. The predication would be supported by the event of the arrival. The sentence would not have a Topic because the person’s presence is entirely unexpected. The sentence would only have a Focus. Therefore, the information structure of the verbless sentences in such a context would be **Thetic**.

The annotation of verbless sentences in this corpus part did not reveal any association between syntactically similar verbless sentences and information structure type. Furthermore, we emphasize that as information structure necessarily requires the consideration of context, verbless sentences are capable of carrying out any of the information structure types.

4.2 TOPIC

The information structural Topic was analyzed from a contrastive perspective, i.e. by comparing verbless sentences with their translation correspondences. The annotation results suggest a trend for Topics that are contextually (linguistically or extra-linguistically) *implied* in Russian verbless sentences to be lexically *evoked* in the English sentences. Furthermore, the instantiation of the Topic is observed to correlate with the instantiation of the verb. The observations taken together suggest that some of the frequency differences between Russian and English verbless sentences may in part be due to typological differences in Topic activation.

One of the ways in which an already active Topic was evoked in English is illustrated in example (61), reproduced with expanded context below as (114). This is the case of the subject representing the Topic added in translation.

(114) *Subject-as-Topic added*a. Russian (BK RO)

[– Вдовею я, третий год, – начала она полупшепотом, сама как бы вздрагивая. – Тяжело было замужем-то, старый был он, больно избил меня. Лежал он больной; думаю я, гляжу на него: а коль выздоровеет, опять встанет, что тогда? И вошла ко мне тогда эта самая мысль... – Постой, – сказал старец и приблизил ухо свое прямо к ее губам. Женщина стала продолжать тихим шепотом, так что ничего почти нельзя было уловить. Она кончила скоро. – Третий год? – спросил старец. – Третий год. Сперва не думала, а теперь хворать начала, тоска пристала. – Издалека? – За пятьсот верст отселева. – На исповеди говорила? – Говорила, по два раза говорила. – Допустили к причастию-то? – Допустили. Боюсь; помирать боюсь. – Ничего не бойся, и никогда не бойся, и не тоскуй.]

Издалека?	За	пятьсот	верст	отселева.
<i>Izdaleka</i>	<i>za</i>	<i>pjat'sot</i>	<i>verst</i>	<i>otseleva</i>
ADV	PREP	NUM	NN.PL	ADV
<i>from_far_away</i>	<i>over</i>	<i>five_hundred</i>	<i>miles</i>	<i>from_here</i>

b. English (BK E1)

["I'm three years a widow," she began in a half-whisper, with a sort of shudder. "My married life was hard, he was old, he beat me badly. Once he was sick in bed; I was looking at him and I thought: what if he recovers, gets up on his feet again, what then? And then the thought came to me ..." "Wait," said the elder, and he put his ear right to her lips. The woman continued in a soft whisper, almost inaudibly. She soon finished. "It's the third year?" the elder asked. "The third year. At first I didn't think about it, and now I've begun to be ill, grief has caught hold of me." "Have you come from far away?" "Over three hundred miles from here." "Did you tell it at confession?" "I did. Twice I confessed it." "Were you allowed to receive communion?" "I was. I'm afraid, afraid to die." "Do not be afraid of anything, never be afraid, and do not grieve.]

Have you come from far away? Over three hundred miles from here.

The question in (114) is about the active Topic referent: the woman. In the Russian sentence, the Topic referent is not lexically evoked, whereas in the English sentence, the Topic is made explicit by means of the addition of the subject 'you'.

It is notable that the addition of the subject in the English question is not a structural necessity: one may simply ask '*From far away?*'. The reason for the reactivation of the Topic may be motivated by the fact that the subsequent utterance '*Over three hundred miles from here*' is also both subjectless and verbless. The verb '*come*' is then added into the English sentence as a result of the addition of the subject. This example illustrates the way that

differences in Topic activation, through the addition of the subject, may influence the realization of the verb.

There were also two other ways in which an active Topic was evoked in English. Illustrated in (115) is a Topic that is linguistically hinted at in the Russian sentence by means of morphological case, but is made fully explicit in the English sentence as a subject.

(115) *Topic made more explicit*

a. Russian (BK RO)

[– Преназойливый старичишка, – заметил вслух Миусов, когда помещик Максимов побежал обратно к монастырю.
– На фон-Зона похож, – проговорил вдруг Фёдор Павлович.]

Ha	фон-Зона	похож
<i>na</i>	<i>fon-Zona</i>	<i>poxož</i>
PREP	NN.MSG.ACC	ADJS.MSG
<i>on</i>	<i>von Sohn</i>	<i>similar</i>

b. English (BK E1)

[“A most obnoxious old fellow,” Miusov remarked aloud, as the landowner Maximov ran back to the monastery. “He looks like von Sohn,” Fyodor Pavlovich declared suddenly.]

He looks like von Sohn

The masculine singular form of the short-adjective in the Russian sentence agrees with the active Topic referent (i.e. the male landowner Maximov). In the English sentence, this Topic referent is linguistically evoked through the subject pronoun ‘*he*’. Again, the evocation of the Topic, and subsequent subject, brings a verb into the sentence.

The third way in which English was observed to evoke an already active topic referent that was not evoked in Russian is through the addition of a *Topical expression* (as opposed to a Topic as in the previous two types). This case is illustrated in (116).

(116) *Topical expression added*a. Russian (BK RO)

[С первого мгновения старец ему не понравился. В самом деле было что-то в лице старца, что многим бы и кроме Миусова не понравилось. Это был невысокий сгорбленный человек с очень слабыми ногами, всего только шестидесяти пяти лет, но казавшийся от болезни гораздо старше, по крайней мере лет на десять. Все лицо его, впрочем, очень сухенькое, было усеяно мелкими морщинками, особенно было много их около глаз. Глаза же были небольшие, из светлых, быстрые и блестящие, в роде как бы две блестящие точки. Седенькие волосики сохранились лишь на висках, борода была крошечная и реденькая, клином, а губы, часто усмехавшиеся - тоненькие, как две бечевочки. Нос не то чтобы длинный, а востренький, точно у птички.]

Нос	не	то	чтобы	длинный,
<i>nos</i>	<i>ne</i>	<i>to</i>	<i>čtoby</i>	<i>dlinnyj</i>
NN.MSG.NOM	NEG	DEM	CONJ	ADJ.MSG.NOM
<i>nose</i>	<i>not</i>	<i>that</i>	<i>so_as</i>	<i>long</i>

а	востренький,	точно	у	птички.
<i>a</i>	<i>vostren'kij</i>	<i>točno</i>	<i>u</i>	<i>ptički</i>
CONJ	ADJ.MSG.NOM	CONJ	PREP	F.S.GEN.DIM
<i>but</i>	<i>sharp</i>	<i>as_though</i>	<i>at</i>	<i>bird</i>

b. English (BK E1)

[He disliked the elder from the first moment. Indeed, there was something in the elder's face that many other people besides Miusov might have disliked. He was a short, bent little man, with very weak legs, who was just sixty-five, but, owing to his illness, appeared much older, by at least ten years. His whole face, which, by the way, was quite withered, was strewn with little wrinkles, especially numerous around his eyes. His eyes themselves were small, pale, quick and bright like two bright points. A few white hairs remained only on his temples, his pointed beard was tiny and sparse, and his often smiling lips were as thin as two threads.]

His nose was not so much long as sharp, like a little bird's beak.

Though not a Topic, the possessive determiner 'his' in the English sentence is *topical* in the sense that it refers to the discourse topic (i.e. the male character being described). It is what Lambrecht (1994: 86) calls an 'anchor'. It anchors the subject 'nose' to the discourse topic.

These three ways of activating a contextually implied subject representing the Topic were found to occur 39 times in the English verbless sentence translation correspondences in BK_RO and BK_E1: 20 for the addition of the subject-as-topic, 8 for a more explicit topic, and 11 for the topic expression. It seems that in English, the cognitive "task of assessing the topic

referent by retrieving it from long-term memory or by drawing inferences leading to its assessment” (Lambrecht, 1994: 165) is simplified. In each case, the Topic becomes more accessible in English.

The finding demands further annotation. At present it provides a hypothesis for a further quantitative study. The English Topic activation, and its link to the subject and hence the verb, suggests that some of the frequency differences between Russian and English concerning verbless sentences may in part be due to typological differences in Topic activation. This constitutes a pragmatic explanation that may account for some portion of the variation in the verbless sentences between the languages.

4.3 FOCUS

Although all sentences do not require a Topic as defined by Lambrecht (1994), all sentences must have a Focus. The Focus is a relation that is non-predictable and non-recoverable (Lambrecht, 1994: 273). The *verb* can be either a part of the Focus or not included in it.

Example (117) illustrates an English sentence where the verb is *not* in Focus.

(117) *Implicated Verb not in Focus*

a. Russian (BK RO)

[– Я был, был, я уже был... Un chevalier parfait! – и помещик пустил на воздух щелчок пальцем. – Кто это chevalier? – спросил Миусов.]

Кто это chevalier?
kto èto chevalier
 PRO DEM NN
who this chevalier

b. English (BK E1)

[“I’ve been, I’ve been already... Un chevalier parfait!” And the landowner loosed a snap of his fingers into the air. “Who is a chevalier?” asked Miusov.]

Who is a chevalier?

In (117b), the English verb is presupposed as a part of the open proposition ‘*X is a chevalier*’ and therefore is *not* in Focus.

More specifically, ‘*Who is a chevalier?*’ constitutes a question by means of which one is inquiring about the identity of ‘*X*’ in the open proposition ‘*X is a chevalier*’. The open proposition that someone is a chevalier is active in the linguistic context (i.e. someone has just been referred to as being a chevalier) and thus presupposed in the question ‘*Who is a chevalier?*’. In other words, the question constitutes an Identificational focus structure type. Notably, the only element that is in the Focus domain of the question ‘*Who is a chevalier?*’ is the word ‘*who*’, i.e. it is the only element that has an unpredictable, non-recoverable relation within the sentence. Since the verb ‘*is*’ is part of the presupposed open proposition, it is *not* in the Focus domain.

When one takes a contrastive analysis perspective and considers the Russian correspondence in (117a), it becomes evident that the English verb, which is not in Focus, is in this instance pragmatically implicated by the Russian *verbless* sentence ‘*Кто это chevalier?*’ (*Kto èto chevalier?*). The example therefore shows a case in which a verbless sentence is found to pragmatically implicate a predictable and recoverable predicate.

What is striking is that the contrastive analysis annotation results show that the above, very much anticipated, case in which the implicated predicate is information structurally predictable and recoverable corresponds to the *minority* of cases. In fact, in 62% of the utterances of the BK_RO and BK_E1 corpus the pragmatically implicated verb was found *to be a part of the Focus* domain, in other words *not predictable* and *not recoverable*.

To illustrate, examples (118) through (120) show cases of an English verb that *is* in Focus pragmatically implicated by Russian *verbless* sentences.

(118) *Implicated Verb in Focus*

- a. Russian (BK RO)

[Ну, а здесь ничего, здесь нет монастырских жен, а монахов штук двести. Честно. Постники. Сознаюсь... Гм. Так ты к монахам хочешь?]

Постники.
postniki
NN.MPL.NOM
people_who_fast

- b. English (BK E1)

[Well, there's nothing like that here, no monastery wives, and about two hundred monks. It's honest. They fast. I admit it...H'm. So you want to go to the monks?]

They fast.

In (118), the Topic is the referent of 'monks' or 'монахи' (*monaxi*; lit. 'monks'). In the English sentence the Topic referent is lexically evoked with the pronoun 'they', while in Russian it remains in the linguistic context. The lexical element 'fast' in English and 'постники' (*postniki*; lit. 'people_who_fast.NN') in Russian has an unpredictable relation to the sentence which means that they constitute the Focus domain. Thus, the information structure of both sentences is of the Topic-Comment type. While the Russian sentence (118a) is verbless, in the English sentence (118b), the verb 'fast' is in the Focus domain. Therefore, taking into account the discussion of pragmatic implicature and translation correspondences in *Part 1* and *Part 2*, the example illustrates an instance of a *verbless sentence* pragmatically implicating a predicate that is in the Focus domain, in other words *not predictable and not recoverable*.

Example (119) illustrates a similar case.

(119) *Implicated Verb in Focus*a. Russian (BK RO)

[Но действительно Федор Павлович всю жизнь свою любил представляться, вдруг проиграть пред вами какую-нибудь неожиданную роль, и, главное, безо всякой иногда надобности, даже в прямой ущерб себе, как в настоящем например случае. Черта эта впрочем свойственна чрезвычайно многим людям и даже весьма умным, не то что Федору Павловичу.]

Черта	эта	впрочем	свойственна	чрезвычайно	
<i>čerta</i>	<i>èta</i>	<i>vpročem</i>	<i>svojstvenna</i>	<i>črezvyčajno</i>	
NN.FSG.NOM	DEM	CONJ	ADJS.FSG	ADV	
<i>trait</i>	<i>this</i>	<i>however</i>	<i>peculiar</i>	<i>extremely</i>	
многим	людям	и	даже	весьма	
<i>mnogim</i>	<i>ljudjam</i>	<i>i</i>	<i>daže</i>	<i>ves'ma</i>	
PRO.PL.DAT	NN.PL.DAT	CONJ	PART	ADV	
<i>many</i>	<i>people</i>	<i>and</i>	<i>even</i>	<i>very</i>	
умным,	не	то	что	Федору	Павловичу.
<i>umnym</i>	<i>ne</i>	<i>to</i>	<i>čto</i>	<i>Fedoru</i>	<i>Pavloviču</i>
ADJ.PL.DAT	NEG	DEM	CONJ	NN.DAT	NN.DAT
<i>intelligent</i>	<i>not</i>	<i>this</i>	<i>what</i>	<i>Fyodor</i>	<i>Pavlovich</i>

b. English (BK E1)

[But all his life, as a matter of fact, Fyodor Pavlovich was fond of play-acting, of suddenly taking up some unexpected role right in front of you, often when there was no need for it, and even to his own real disadvantage, as for instance, in the present case. This trait, however, is characteristic of a great many people, even rather intelligent ones, and not only of Fyodor Pavlovich.]

This trait, however, is characteristic of a great many people, even rather intelligent ones, and not only of Fyodor Pavlovich.

In (119), the Topic is the referent of ‘*trait*’ or ‘*черта*’ (*čerta*; lit. ‘*trait*’). The sentence is a new piece of information that is created from the combination of this predictable element, that is active in the linguistic context, with the non-predictable characterization that is given to it, i.e. ‘*is characteristic of a great many people, even rather intelligent ones, and not only of Fyodor Pavlovich*’ in English, and ‘*впрочем свойственна чрезвычайно многим людям и даже весьма умным не то что Федору Павловичу*’ (*vpročem svojstvenna čresvyčajno mnogim ljudjam i daže ves'ma umnym ne to čto Fedoru Pavloviču*) in Russian. In the English sentence (119b), the verb is thus included in the Focus domain ‘*is characteristic of a great many people,*

even rather intelligent ones, and not only of Fyodor Pavlovich'. The Russian sentence (119a) therefore pragmatically implicates a predicate that is *non-predictable* and *non-recoverable*.

Example (120) illustrates the case of an English verb in Focus in an information structurally Thetic sentence.

(120) *Implicated Verb in Focus*

a. Russian (BK RO)

[– Да, вот вы тогда обедали, а я вот веру-то и потерял! – поддразнивал Федор Павлович. – Какое мне дело до вашей веры!]

Какое	мне	дело	до	вашей	веры!
<i>kakoe</i>	<i>mne</i>	<i>delo</i>	<i>do</i>	<i>vašej</i>	<i>very</i>
PRO	PRO.DAT	NN.NOM	PREP	PRO.GEN	NN.GEN
<i>what</i>	<i>to_me</i>	<i>business</i>	<i>for</i>	<i>your</i>	<i>faith</i>

b. English (BK E1)

[“So you were having dinner then, and I just lost my faith!” Fyodor Pavlovich went on teasing him. “What do I care about your faith!”]

What do I care about your faith!

In (120) the Focus domain extends over the entire sentence. The lexical element ‘*your faith*’ in English and ‘*вашей веры*’ (*vašej very*) in Russian is a topical element that is included in the Focus domain. Its referent is not considered to be the Topic of the sentence since the purpose of the sentence is not to make a comment about faith. Rather, the purpose is to express a comment about the speaker, who is the subject but not the Topic (i.e. the comment that the speaker does not care). Notably, the predication in both (120a) and (120b) does not have a Topic; it is supported by the prior linguistic and extra-linguistic context (i.e. Fyodor Pavlovich’s utterance that the speaker deems irrelevant). The example thus illustrates a Russian verbless sentence that pragmatically implicates a non-predictable and non-recoverable predicate in contrastive analysis.

The present analysis revealed that 62% of the verbless sentence correspondences were of the type illustrated in (118) to (120), i.e. with verbs belonging to the Focus domain.

These results counter the intuitive conception of the verbless sentence as constituting a structure with a pragmatically predictable predicate. The verbs that are pragmatically implicated by verbless sentences were found through contrastive analysis to be a part of Focus. This finding challenges the extent to which the verbless sentence can be accounted for in terms

of the omission of a predictable and reconstructable element. It provides further evidence to support the argument that is presently defended, that is that the constituents of a verbless sentence can be sufficient to express a complete thought and satisfy the requirements for constituting full instances of predication.

CHAPTER 5

TEMPORAL AND ASPECTUAL MEANING

In what sense can verbless sentences be associated with meaning that is temporal, aspectual or otherwise prototypically verbal and what could this meaning look like? In (Bondarenko & Celle, 2020) we explore this question focusing particularly on aspectual meaning (or rather potential pragmatic implicature). In *Section 5.1* of the present chapter, we summarize the arguments and present the key results concerning our small-scale study on aspect in the *Stranger* sub-corpus. Then, in *Section 5.2*, a macro full-corpus perspective is taken on the verbal correspondences of verbless sentences with regard to tense and lexical meaning.

5.1 ASPECT

The essence of the aspectual issue for verbless sentences is that they constitute structures in which the typical marker of temporal and aspectual meaning is absent. This marker is what Smith (1997: 10) calls the ‘*verb constellation*’, in the sense that it involves more than the verb. Most current analyses agree that to determine aspectual meaning, the entire sentence must be considered. For instance, despite the shared verb in (121), (a) and (b) have different aspectual meanings (Corre, 2009: 57).

- (121) a. She is painting. [Activity]
 b. She is painting a picture. [Accomplishment]

Example (a) emphasizes a temporary ongoing process and corresponds to an *activity* situation-type, while (b) indicates that there is a change of state that occurs as a part of this process and indicates an *accomplishment* situation type. The distinction concerns the semantic property of telicity, however, in order to perceive it, it is clearly necessary to look beyond the verb. The complements and the entire sentence must be considered, otherwise the semantic properties distinguishing aspectual meaning appear to be irrelevant. The necessity of the sentence-level analysis, and the insufficiency of the verb phrase alone, for determining aspectual meaning is an argument that was first made by Henk Verkuyl (1972).

Yet, some take a more radical step and question the very necessity of the verb for aspectual meaning. For instance, some studies of the verbless sentence phenomenon approach the question of its aspectual properties by assign aspectual value to non-verbal elements. On such accounts, the verbless sentence in (122a) would represent an *achievement* situation type in which the semantic properties of *instantaneity* and *telicity* are assigned to the noun ‘*vengeance*’.

- (122) a. Vengeance!
 b. Vengeance in an hour [Adverbial of completion]
 c. ? Vengeance for an hour [Adverbial of duration]
 d. ? Vengeance from 2 to 3 in the afternoon [Adverbial of duration]

The first premise for justifying such a perspective is the existence of a verbal counterpart for the noun. For example, the noun ‘*vengeance*’ is associated with the verb ‘*to avenge*’, the noun ‘*victory*’ with ‘*to vanquish*’, the noun ‘*loss*’ with ‘*to lose*’, etc. The hypothesis that some nouns inherent the aspectual properties of the verbs from which they originate is explored in research on *deverbal nouns* (e.g. Balvet et al., 2011; Rakhilina, 1999).

The second essential premise of justifying the approach is that that the typical tests for aspectual properties may be applied not only to verb constellations, but also to nominal sentences. This perspective is developed by Lefeuvre and Nicolas (2004). The typical tests for aspectual meaning are discussed by David Dowty (1979). They include, for instance, the *competition test*: the possibility of rephrasing the sentence using the adverbial ‘*in an hour*’ indicates that the sentence has the property of *completion*. The property of *duration* is indicated by the possibility of using the sentence with the adverbials ‘*for an hour*’ or ‘*from 2 to 3 in the afternoon*’. Lefeuvre and Nicolas (2004) show that the tests may be successful applied not only

in verbal sentences, but also in verbless sentences. This is illustrated with the verbless sentence in (122) which is felicitous with (b) ‘*in an hour*’, but is minimally very marked with (c) ‘*for an hour*’ and (d) ‘*from 2 to 3 in the afternoon*’. Thus, on such an account, the verbless sentence ‘*Vengeance!*’ would be analyzed as having the property of *completion*, but not *duration*.

The driving goal behind such an approach appears to be a key perspective that is also fully shared in the present thesis: the defence of the existence of a coherent and complete semantics for sentences without verbs. However, despite it being a refreshing approach, several reasons lead us to take a step away from it. The present reasons for the distancing from this approach concern the nature of the semantic transfer from the verb to the noun. First, it has been shown by Huyghe (2011) that the grammatical constraints on the use of verbs and nouns prevent a straight forward inheritance of (*a*)*telic* properties from verbs to count and mass nouns. The second issue is the direction of transfer, i.e. whether it is the verb or the noun that inherits the properties of the other. The third concern is that by attributing aspectual meaning to deverbal nominalizations, the approach seems to assume that it is the *verb*, rather than the *verb constellation*, that is sufficient to be the origin of aspectual meaning. This runs counter to for instance example (121) which showed that aspectual meaning is a feature of the sentence as opposed to any lexical element. Finally, the ascribing of aspectual features to nouns would seem to create competition for aspectual meaning in sentences which also contain a verb.

Therefore, though we share the argument that verbless sentences have a complete semantics, we approach its semantic properties from a different perspective, i.e. the perspective of pragmatic implicature rather than semantic entailment. When it comes to the presence of the verb, we take the traditional view that it is a necessary, though not sufficient, condition to determine the aspectual meaning of the sentence. As discussed in *Part 1*, the aspectual and temporal value of a sentence without a verb, on our perspective, is semantically undetermined. Therefore, in (122a), the noun ‘*vengeance*’ does not semantically entail neither the idea of completion nor aspectual or temporal meaning. However, its pragmatic insertion into a specific situation is capable of *conversationally implicating* aspectual and temporal properties when ‘*Vengeance!*’ is uttered in context.

In (Bondarenko & Celle, 2020), we applied this conception to analyze the implicated aspectual meaning of verbless sentences in a corpus consisting of Albert Camus’ French novel *L’Étranger*, two English and two Russian translations. The Russian and English translations are included in the present corpus as ST_E1F, ST_E2F, ST_R1F and ST_R2F (see *Part 3*).

While the full details of the method and annotation definitions are found in the article, a few key points must be mentioned. The analysis focused on *verbal correspondences* of *non-*

elliptical verbless sentences. Therefore, to start, all verbless sentences were manually annotated as elliptical and non-elliptical. Notably, the ellipsis annotation occurred on the level of the sentence, not utterance, with the sole goal of setting aside the sentences in which the antecedent may be used to help assess tense and aspect.

The verbal correspondences of non-elliptical verbless sentences were then manually annotated for *situation type* and *viewpoint type* in accordance with the *two-component theory of aspect* proposed by Smith (1997). The verbal correspondences were annotated for the properties of *stativity/dynamism*, *durativity/instantaneity*, *telicity/atelicity*. The presence of the properties was determined according to the Russian-specific and English-specific tests developed by Smith (1997). The latter develops an intricate series of language-specific tests for the properties and builds a framework that extends Vendler's (1957) classification of semantic types.⁴⁵ The *situation type* of each sentences (i.e. either *state*, *activity*, *accomplishment*, *achievement*, or *semelfactive*) was then assigned based on the annotated properties.

The *viewpoint type* was also analyzed according to Smith's (1997) language-specific definitions. The *perfective viewpoint* is generally defined as one that focuses on the situation in its entirety, while the *imperfective viewpoint* focuses on a part of the situation and presents no information concerning end points, i.e. it is informationally open (Smith, 1997: 73). In line with Smith (1997), the distinction was made according to the following criteria:

— *Russian:*

The *perfective* or *imperfective viewpoint* is marked on each verb either by means of an affix or by alteration of the verb stem (Smith, 1997: 228; Maslov, 1984). In Russian, the *perfective viewpoint* presents only closed situations and does not apply to *statives* (Smith, 1997: 69–71).

— *English:*

The *imperfective viewpoint* corresponds to the progressive tense signaled by the progressive aspect 'be + ing', while the *perfective viewpoint* typically corresponds to

⁴⁵ For the general conceptual differences between the properties see Smith (1997: 17-35); for the language-specific realization of the properties in English – Smith (1997: 176-184) and for Russian – Smith (1997: 241-250).

its absence (Smith, 1997: 170).⁴⁶ In English, the *perfective viewpoint* is applicable to all situation types; however, it does not include endpoints of *states* and therefore may present stative situations as open or closed (Smith, 1997: 69–71).

Figure 37 illustrates Smith’s (1997) two-component system of aspectual meaning and the annotation carried out (with the present exclusion of French).

ASPECTUAL MEANING OF A SENTENCE							
SITUATION TYPE				VIEWPOINT TYPE			
	State	Activity	Accomplishment	Achievement	Semelfactive		
						Russian	
						Perfective	
						Imperfective	
						English	
						Perfective	
						Imperfective	
Situation	Static	Durative	Telic		Viewpoint	Russian	English
ST	+	+	-		Perfective	prefix, verb stem	all other tenses
ACT	-	+	-		Imperfective	absence of prefix, verb stem	progressive tense
ACC	-	+	+				
ACH	-	-	+				
SEM	-	-	-				

Figure 37. Situation type and viewpoint type annotation

While the full results are provided in the article, the following summarizes only a few key observations concerning the aspectual meaning implicated by Russian and English verbless sentences. The reference to *verbless sentences* in the sections below refer specifically to the non-elliptical type.

⁴⁶ It is notable that the past simple, i.e. without ‘*be + ing*’, is not always *perfective*. For instance, in ‘*He stood in the middle of the room*’ the act of standing does not include a final endpoint. Nevertheless, for the present annotation we take only the progressive grammatical aspect as the marker of English *imperfective viewpoint*, and all other tenses are taken to mark the *perfective viewpoint*. In doing so, we follow Smith’s analysis that:

[English viewpoint] is indicated by the presence or absence of the verbal auxiliary. The perfective viewpoint is phonetically zero, contrasting with the auxiliary morpheme which conveys the imperfective, progressive viewpoint. [...] The analysis of English is based on a distribution: there is a consistent, obligatory contrast between the auxiliary ‘*be + ing*’ and the simple verb form, with a zero morpheme. Since all verbs have one of these forms, all verbs in English have a viewpoint morpheme. (Smith, 1997: 170)

Table 23 presents a summary of the contrastive results. More specifically, it shows the results for the verbal correspondences (VC) of verbless non-elliptical sentences (VlessNE) in the 8 different directions of comparison of the 2 Russian and 2 English texts. Column (a) ‘Correspondences’ shows the number of correspondences of verbless sentences (separated into ‘verbal’, which were further annotated, and ‘other’, i.e. the verbless or elliptical correspondences set aside). Column (b) ‘Situation Type’ breaks down the verbal correspondences by situation type (*state, activity, accomplishment, achievement, semelfactive*). Column (c) ‘Viewpoint’ reveals the verbal correspondences that were of the *perfective* or the *imperfective* viewpoint. The frequencies are shown in terms of absolute values in (i), and normalized per 100 verbless non-elliptical sentences in (ii).

Direction	(i) Absolute Frequency									(ii) Normalized per 100 Verbless NE								
	a.		b.					c.		a.		b.			c.			
	Verbal	Other	ST	ACT	ACC	ACH	SEM	P	I	Verbal	Other	ST	ACT	ACC	ACH	P	I	
Russian Verbless																		
I	E1VC of R1VlessNE	78	39	70	5	1	2	0	78	0	67	33	60	4	1	2	67	0
II	E2VC of R1VlessNE	73	44	61	4	1	7	0	72	1	62	38	52	3	1	6	62	1
III	E1VC of R2VlessNE	103	43	95	4	2	2	0	101	2	71	29	65	3	1	1	69	1
IV	E2VC of R2VlessNE	94	52	85	5	0	4	0	92	2	64	36	58	3	0	3	63	1
English Verbless																		
V	R1VC of E1VlessNE	5	27	1	1	1	2	0	2	3	16	84	3	3	3	6	6	9
VI	R2VC of E1VlessNE	4	28	0	2	0	2	0	2	2	13	88	0	6	0	6	6	6
VII	R1VC of E2VlessNE	3	28	1	0	1	1	0	1	2	10	90	3	0	3	3	3	6
VIII	R2VC of E2VlessNE	1	30	0	0	0	1	0	0	1	3	97	0	0	0	3	0	3

Table 23. Situation type and viewpoint aspect implicated by verbless sentences

It is notable that the number of Russian verbal correspondences of non-elliptical English verbless sentences (i.e. directions V to VIII in the table) include only a total of 13 occurrences. Consequently, the analysis of the aspectual meaning implicated by English verbless sentences is at present extremely limited.

5.1.1 States and Events

The first observation is that, based on the present corpus, verbless sentences are frequently used in contexts which call for a stative situation. On average, Russian verbless sentences implicate a stative situation type in 89% of verbal correspondences. An instance is illustrated in (123).

(123) *Stative Implicature*

[Court was adjourned and my lawyer sat back down. He looked exhausted. But his colleagues came over to shake his hand. I heard:]

- a. Превосходно, дорогой мой! [Russian verbless]
prevosходno, dorogoj moj
excellent, dear my
- b. That was brilliant! [English state]

‘*That was brilliant!*’ semantically entails a stative situation and, thus, exposes the possible conversationally implicated stative aspectual meaning of the verbless Russian sentence.

Dynamic events are also within the range of implicature of verbless sentences. Russian verbless sentences were observed to implicate activities, achievements and accomplishments. An instance is illustrated in (124).

(124) *Dynamic Implicature*

[Then he wanted to know if I had hired an attorney. I admitted I hadn’t and inquired whether it was really necessary to have one. “Why do you ask?” he said.]

- a. В каком смысле? [Russian verbless]
v kakom smysle
in what sense
- b. Why do you ask? [English activity]

Although the annotation must be extended to a much larger corpus, at present we are able to note that a correlation of verbless sentences with dynamic situation types is possible. The only type not observed in any of the results is the semelfactive (i.e. instantaneous telic events such as ‘*He sneezed*’ appear to be excluded from the aspectual potential of verbless sentences).

5.1.2 Viewpoint

Russian verbless sentences appear to be restricted to implicating situations presented in their entirety in English. This is suggested by a strong and high frequency correlation for the perfective viewpoint, such as illustrated in (123) above. In terms of absolute frequency, all 78 correspondences of verbless R1 in E1, 72 of the 73 correspondences of verbless R1 in E2, 101 of 103 correspondences of verbless R2 in E1, and 92 of 94 correspondences of verbless R2 in E1, correlated with the perfective viewpoint in English.

At this point, such a result is taken to indicate that verbless sentences are *not* free in terms of the viewpoint that they may implicate. In other words, they appear to have a particular aspectual implicature. However, further work is required to uncover the nature of these aspectual limits through manual annotation.

5.2 TENSE

Working with the full 1.4-million-word corpus, it was possible to automatically identify and summarize the verbal correspondences of verbless sentences. Although the automatic results present the important limit that a compound verb such as ‘*has been going*’ is treated as three separate verbs, they nevertheless provide an interesting macro perspective to be refined in further studies.

When all of the originals and translations are considered together, a total of 53 translation correspondences exist in the corpus (e.g. verbless in BK_RO and verbal in BK_E1 constitutes one set of correspondences).

Figure 38 shows the results of the verbal English correspondences of Russian verbless sentences.

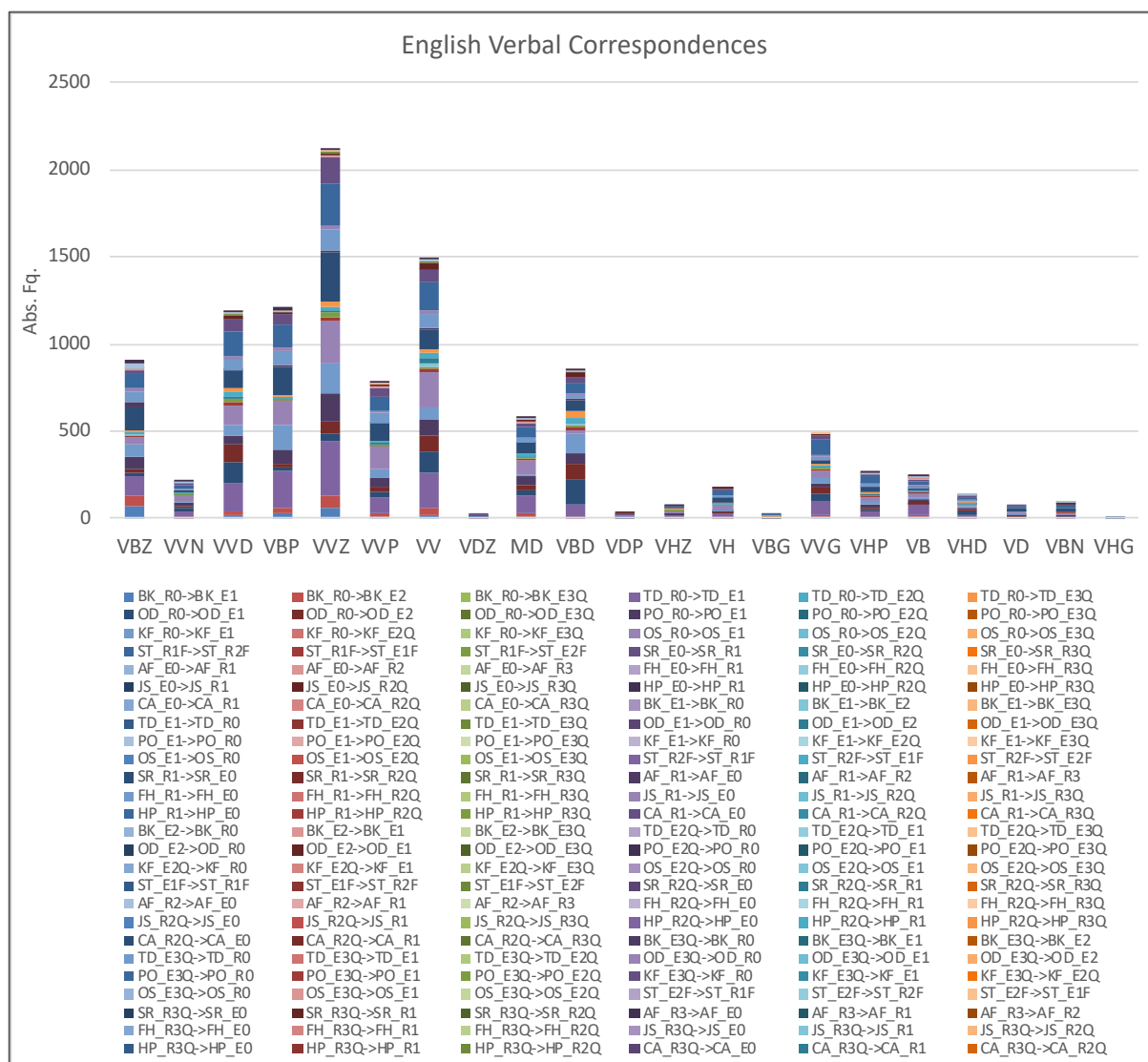


Figure 38. English verbal correspondences by category (absolute frequency)

Verbal English correspondences (of Russian verbless sentences) pertain to all verbal morphosyntactic categories. Even though the *present tense* is the most frequent correspondence, i.e. particularly VVZ (third-person-singular present form of verbs other than ‘be’, ‘do’ and ‘have’), VBP (‘am’, ‘are’), VBZ (‘is’), VVP (non-third-person-singular present forms of verbs other than ‘be’, ‘do’ and ‘have’), the *past tense* is also strongly represented, i.e. VVD (past forms of verbs other than ‘be’, ‘do’ and ‘have’), VBD (‘was’, ‘were’).

The Russian verbal correspondences of English verbless sentences are presented in Figure 39.

Concerning Russian verbs, it is evident that the *present tense*, the *imperative form* and the *past tense* dominate the correspondences.

As discussed in *Appendix 3* and *Appendix 6*, Russian has a lot of morphosyntactic specifications. The next step is to create a customized grouping of some of the morphosyntactic specifications in Trameur in order to get a better idea concerning the relevant categories. The present results may be used to help make the choice concerning what may be grouped and what should remain distinct for the verbless sentence phenomenon. For instance, it may be relevant to custom group the gender and some of the person categories.

Furthermore, this question also requires manual annotation in terms of determining the tense of the compound verbs, as opposed to the present calculations which consider only the form.

Nevertheless, at this point, the results present clear evidence that verbless sentences, in both languages, implicate *a range of tenses*, and are far from only *present* implicature.

5.3 VERBAL LEMMAS

On the lemma level, the verbs ‘*be*’, ‘*have*’, ‘*do*’, ‘*get*’ and ‘*go*’ are strongly implicated by the Russian verbless sentences. As shown in *Figure 40*, they represent more than half of the verbal correspondences. Furthermore, verbs of *possession*, *motion*, *communication* and *perception* are among the top implicatures of Russian verbless sentences.

Note that ‘*sv*’ in the figure corresponds to the verbal clitic [’*s*].

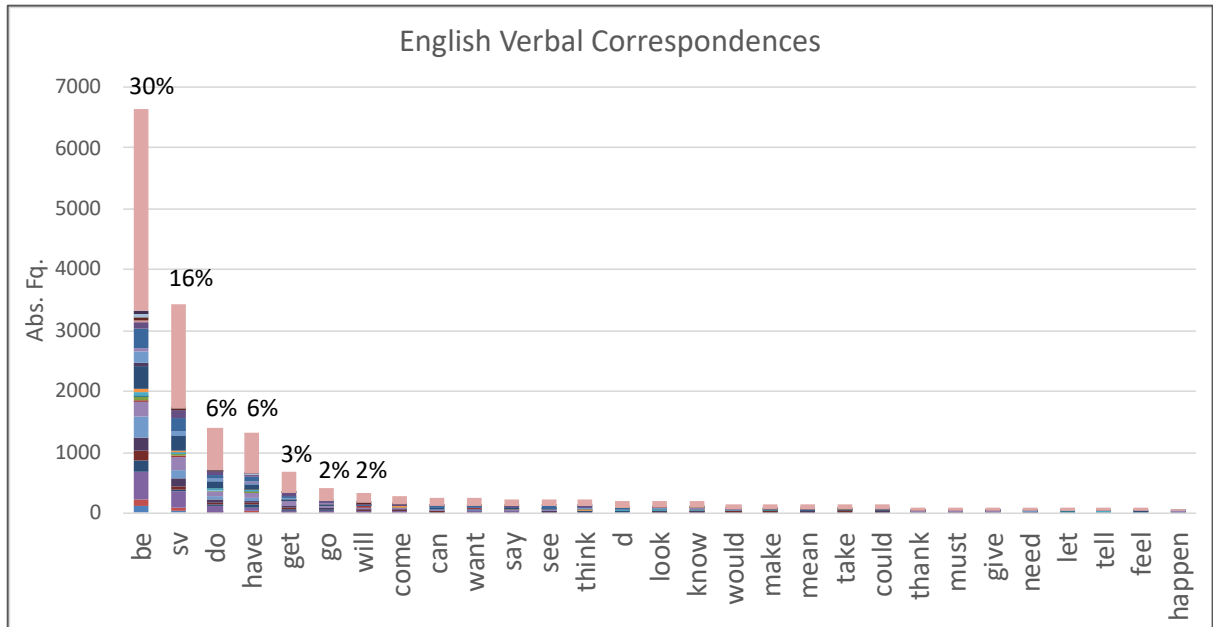


Figure 40. English verbal correspondences by lemma (absolute frequency)

In Russian, as in English, the verb ‘быть’ (*byt’*; lit. ‘be’) is also the most frequent verbal correspondence. As evident in *Figure 41*, the English verbless sentence also tends to pragmatically implicate verbs of *communication, possession, motion, and perception*.

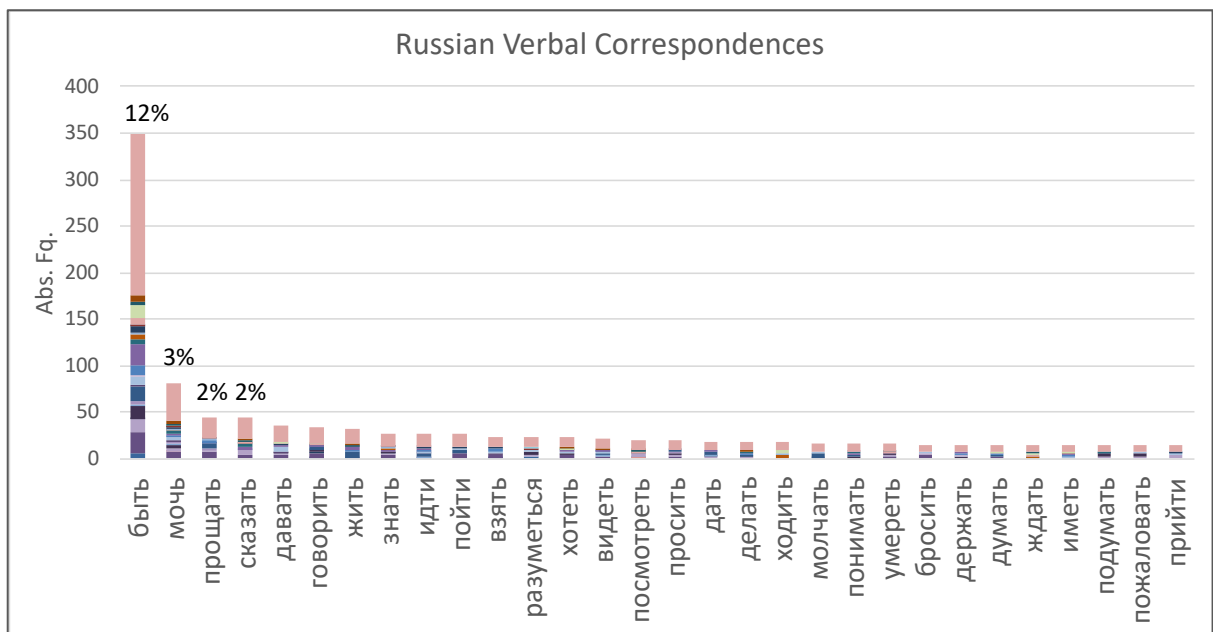


Figure 41. Russian verbal correspondences by lemma (absolute frequency)⁴⁷

⁴⁷ Translation of Russian verbal correspondences (in order of frequency from highest to lowest as in the graph):

The same warnings apply as in the previous section, i.e. a compound verb such as ‘*has been going*’ is treated as three separate lemmas. Further manual annotation in terms of determining main lexical verb in the compound is the next step.

Nevertheless, the present results provide evidence that, in both languages, the semantic categories of verbs of *communication*, *possession*, *motion* and *perception* make it possible for the speaker to pragmatically implicate them without overt semantic encoding.

byt' (lit. 'be'),
moč' (lit. 'can'),
proščat' (lit. 'forgive'),
skazat' (lit. 'say.PFV'),
davat' (lit. 'give.IPFV'),
govorit' (lit. 'speak.IPFV'),
žit' (lit. 'live'),
znat' (lit. 'know'),
idti (lit. 'walk.IPFV'),
pojti (lit. 'walk.PFV'),
vzjat' (lit. 'take'),
razumet'sja (lit. 'be_meant'),
xotet' (lit. 'want'),
videt' (lit. 'see.IPFV'),
posmotret' (lit. 'look_at.PFV'),
prosit' (lit. 'ask'),
dat' (lit. 'give.PFV'),
delat' (lit. 'do'),
xodit' (lit. 'walk.IPFV'),
molčat' (lit. 'keep_silent'),
ponimat' (lit. 'understand'),
umeret' (lit. 'die'),
brosit' (lit. 'throw'),
deržat' (lit. 'hold'),
dumat' (lit. 'think.IPFV'),
ždat' (lit. 'wait'),
imet' (lit. 'have'),
podumat' (lit. 'think.PFV'),
požalovat' (lit. 'come'),
prijti (lit. 'arrive')

CHAPTER 6

INDIRECT SPEECH ACTS IN QUESTIONS

Another question about which it was possible to gain insight is: *What type of sentences are most concerned by verbal correspondences?* There are some sentences that remain verbless in both languages. However, there are others that show variation concerning the verb. The present results show that *questions*, as opposed to other types of sentences, are particularly sensitive to this variation.

The issue at hand was explored in the *Brothers Karamazov*, *Caretaker* and *Stranger* sub-corpora:

- Russian original *dialogue-based fiction* novel BK_RO and its two English translations BK_E1 and BK_E2,
- English original *scripted-speech fiction* novel CA_EO and its Russian translation CA_R1;
- two Russian and two English translations from French, the *narrative-based fiction* novel ST_R1, ST_R2, ST_E1, ST_E2.

The details of these corpus parts are found in *Part 3*.

The verbless sentences and their translation correspondences in this sub-corpus were first manually annotated for the presence of the verb and the sentence type. This annotation identified 345 examples that have a question in at least one of the parallel-texts:

- 95 in BK (across 3 parallel-texts)
- 213 in CA (across 2 parallel-texts)
- 37 in ST (across 4 parallel-texts)

Further annotation was then performed on these questions and their translation correspondences (i.e. a total of 859 sentences).

This further manual annotation included: the presence of antecedent-based ellipsis, direct or indirect speech act, and question type. The present results focus particularly on the 284 non-elliptical verbless questions and their translation correspondences. Precisions concerning method and definitions are provided in *Part 2*.

6.1 SENSITIVITY TO THE VERB

The results reveal that verbal variation is a particularly salient issue for questions, as compared to other types of sentences.

Figure 42 summarizes the verbal correspondences of Russian verbless utterances. More specifically, the figure shows that in the corpus part where English is the original-source language (i.e. CA_EO), 78% of the *question* correspondences are *verbal* (and 22% are verbless), whereas only 51% of *all-sentence-types* correspondences are verbal (and 49% of all-sentence-types remain verbless). In the corpus part where English is the language of translation (i.e. BK_E1 and BK_E2), questions show 60% verbal correspondence rate (40% of questions remain consistently verbless), whereas all-sentence-types show a 49% verbal correspondence rate. When the English is of the third-language translation type (i.e. ST_E1 and ST_E2), 65% of the questions are verbalized (35% remain consistently verbless) compared to a verbalization of 56% of all-sentence-types combined (46% of all-sentence-types remain consistently verbal).

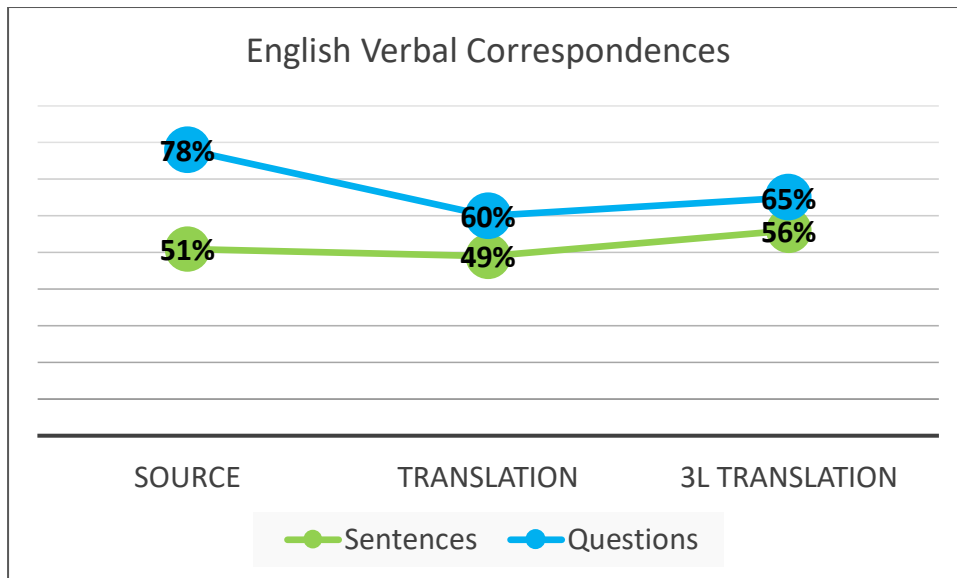


Figure 42. Russian verbless sentences and their verbal correspondences

Clearly, Russian verbless questions with verbal correspondences (represented by the blue line) consistently exceed the other types of Russian verbless utterances with verbal correspondences (represented by the green line). In other words, concerning verbless sentences in Russian, questions have more verbal correspondences than the other types of utterances put together.

This pattern proved to be reciprocal. As shown in *Figure 43*, English verbless questions consistently have fewer verbal correspondences than the other types of utterances combined.

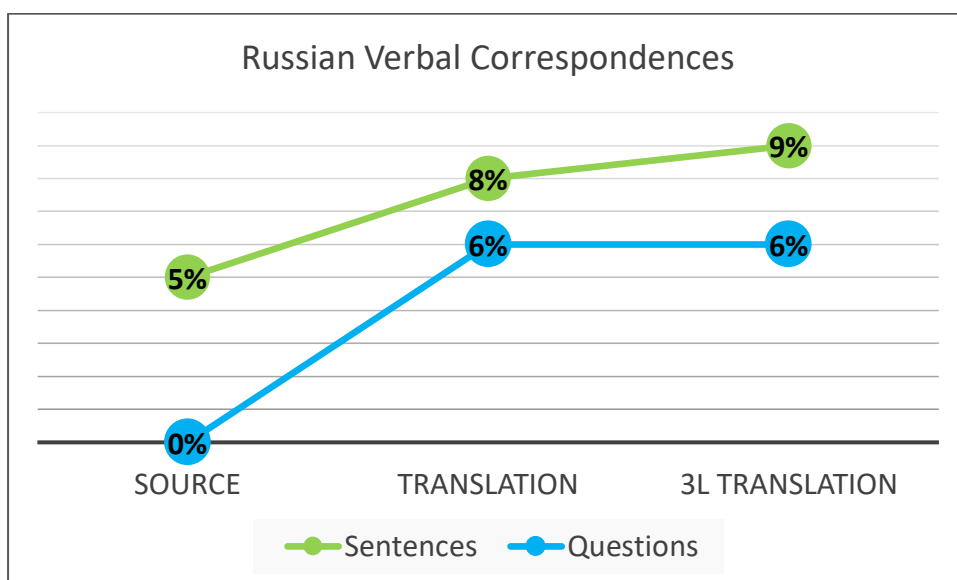


Figure 43. English verbless sentences and their verbal correspondences

It is also notable that the pattern holds true regardless of whether one is looking at the verbal correspondences of the source language, the translation language, or the third-language translation. Therefore, it is safe to say that *questions*, compared to other sentence types, are more sensitive to the verb.

The full details of the calculation are provided in *Table 24*. Notably, it makes explicit the absolute values involved in the calculation of the verbalization rate of *verbless utterances* (i.e. the utterances of all of the verbless sentences in the particular text, including questions) compared to the verbalization rate of the *verbless questions* (i.e. the non-elliptical questions within the verbless utterances).

	Verbless Utterances	Verbless Utterances with Verbal Correspondences				Verbalization of Verbless Utterances
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> BK_E1 BK_E2 BK_RO </div>				
BK_RO	419	207	220			0.510
BK_E1	220			8		0.047
BK_E2	207			12		
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> CA_R1 CA_EO </div>				
CA_EO	385	30				0.078
CA_R1	710		351			0.494
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> ST_R1 ST_R2 ST_E1 ST_E2 </div>				
ST_E1	77	8	8			0.090
ST_E2	79	6	6			
ST_R1	168			93	89	0.564
ST_R2	195			118	111	
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> BK_E1 BK_E2 BK_RO </div>				
BK_RO	58	44	46			0.776
BK_E1	10			0		0.000
BK_E2	7			0		
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> CA_R1 CA_EO </div>				
CA_EO	33	2				0.061
CA_R1	131		79			0.603
		<div style="display: flex; justify-content: space-around; font-weight: normal;"> ST_R1 ST_R2 ST_E1 ST_E2 </div>				
ST_E1	4	0	1			0.063
ST_E2	3	0	0			
ST_R1	17			10	9	0.648
ST_R2	21			16	15	

Table 24. Verbalization of utterances and questions (absolute frequency)

6.2 SPEECH ACT OF VERBLESS QUESTIONS

Having observed that verbalization is particularly variable with regard to verbless questions, the next issue to explore is then why could this be? Having annotated the questions in terms of speech act and question type, we found a pragmatic association of the verb in English. The association was uncovered in two steps.

First, a monolingual perspective revealed that verbless (non-elliptical) questions, in both languages, tend to correlate with indirect speech acts. This means that the verb tends to be omitted mostly in situations where the speaker is *not* requesting an informational answer from the addressee.

Figure 44 shows that indirect speech acts represent 62% of the Russian verbless questions (i.e. an absolute frequency of 118 out of the 189 verbless non-elliptical Russian questions in originals and translations) and 64% of the English verbless questions (i.e. an absolute frequency of 32 out of the 50 verbless non-elliptical Russian questions in originals and translations).

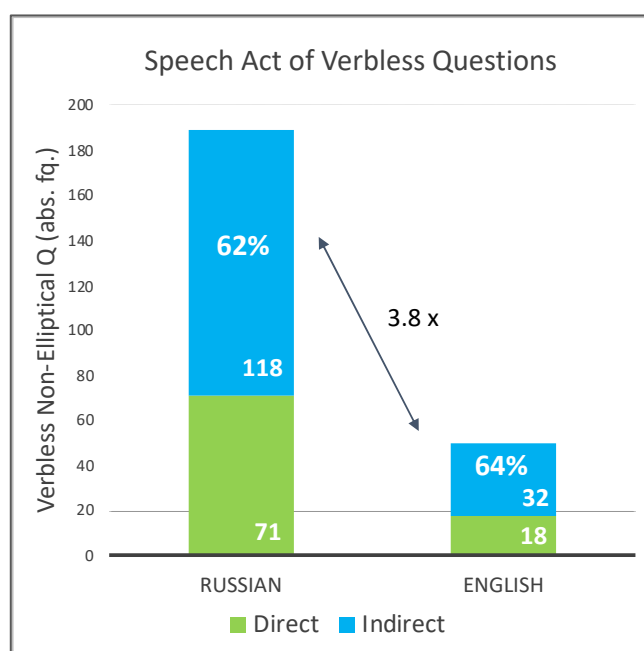


Figure 44. Speech act of verbless questions

This finding suggests that not seeking an answer seems to be a pragmatic characteristic of verbless (non-elliptical) questions, that goes beyond cross-linguistics boundaries.

To illustrate, two particular types of questions used as indirect speech acts were observed in the present corpus: *surprise questions* and *indirect rhetorical questions*.

Surprise questions such as (125), express the speaker's response to some unexpected situation. In this particular case, the question constitutes a response to the repetitive taunting by the speaker's interlocutor.

(125) *Indirect Surprise Question*

{Fyodor Pavlovich starts taunting Miusov again. Miusov says:}

Russian (BK RO)

Вы опять?
vy opjat'
you again

English (BK E1)

What, again?

By asking '*What, again?*' and 'Вы опять?' (vy opjat') the speaker shows that he disapproves of this situation and refuses to cognitively assimilate it. The addressee is being challenged. By using this question the speaker implicates '*I can't believe it!*'.

Also notable is the use of '*what*' in (125). It is not a referential pronoun here, i.e. the question is not about the identity of a referential entity. The interrogative pronoun used here is the most indeterminate one, and this allows for the surprise reading. In terms of *wh*-words, '*what*' is one of the most statistically characteristic elements of verbless sentences in general.

The second type of verbless question used as an indirect speech act which was observed in the present data is the *indirect rhetorical question*. In (126) the absence of the verb emphasizes that the question is associated with *bias*.

(126) *Indirect Rhetorical Question*

Russian (BK RO)

[– Стыдно! – вырвалось вдруг у отца Иосифа.
– Стыдно и позорно! – своим отроческим голосом, дрожащим от волнения, и весь покраснев, крикнул вдруг Калганов, все время молчавший.
– Зачем живет такой человек? – глухо прорычал Дмитрий Федорович, почти уже в иступлении от гнева, как-то чрезвычайно приподняв плечи и почти от того сгорбившись, – нет, скажите мне, можно ли еще позволить

ему бесчестить собою землю, – оглядел он всех, указывая на старика рукой. Он говорил медленно и мерно.

– Слышите ли, слышите ли вы, монахи, отцеубийцу, – набросился Федор Павлович на отца Иосифа. – Вот ответ на ваше “стыдно”! Что стыдно? Эта “тварь”, эта “скверного поведения женщина” может быть святее вас самих, господа спасающиеся иеромонахи!]

Что	стыдно?
<i>čto</i>	<i>stydno</i>
<i>what</i>	<i>shame</i>

English

(BK E2)

[‘For shame!’ Father Yosif could not contain himself.

‘Shame and a disgrace indeed,’ Kalganov, who had been silent until now, shouted suddenly, blushing to the roots of his hair, his adolescent voice quivering with emotion.

‘A man like him doesn’t deserve to live!’ Dmitry Fyodorovich growled, beside himself with rage, his awkwardly raised shoulders lending him a hunched appearance. ‘Tell me, is he to be allowed to continue to contaminate the earth with his existence?’ He looked at everyone in turn, pointing at the old man with his finger. He spoke slowly and deliberately.

‘Did you hear him, holy fathers, did you hear the murderer?’ Fyodor Pavlovich turned on Father Yosif. ‘Here’s the answer to your “for shame”! What shame? This “creature”, this “disreputable woman”, is holier, perhaps, than any of you monks seeking salvation here!]

What shame?

The questions ‘*What shame?*’ and ‘Что стыдно?’ (*čto stydno*) indicate bias in the sense that an assertion is conveyed. The question conveys the belief that the answer to it is that ‘*there is no shame*’. In this way the disagreement with the interlocutor is being highlighted.

The example illustrates the sense of ‘bias’ discussed by Brian Jon Reese (2007), who argues that bias is an assertion, not a conversational implicature, and as a result, biased questions can be thought of as a type of indirect speech act. It is precisely this element of bias that makes (126), the *rhetorical question*, different from the *surprise question* in (125). Although this distinction is not traditionally made, we argue that these do belong to two different question types. Both of these questions serve a challenging function, but the function of (125) is different. Here we follow Obenauer (1994) and Celle (2018). In (125), ‘*What again?*’ and ‘Вы опять’ (*vy opjat*), it is the fact that the whole situation repeats itself that is found to be surprising. The question does not carry a bias or an assertion, but rather implicates ‘*I can’t believe it!*’.

6.3 SPEECH ACT OF VERBAL CORRESPONDENCES

Second, from a contrastive perspective, i.e. when the parallel correspondences of verbless questions are examined, a particular use of the verb surfaces in English. The English verb was found in the question correspondences particularly when the speech act is *direct* (i.e. when the speaker is seeking an informational answer).

Figure 45 summarizes the English verbal correspondences of Russian verbless questions in terms of whether they belong to direct or indirect speech acts. The numbers shown are the absolute frequencies (averaged in the case of multiple translations).

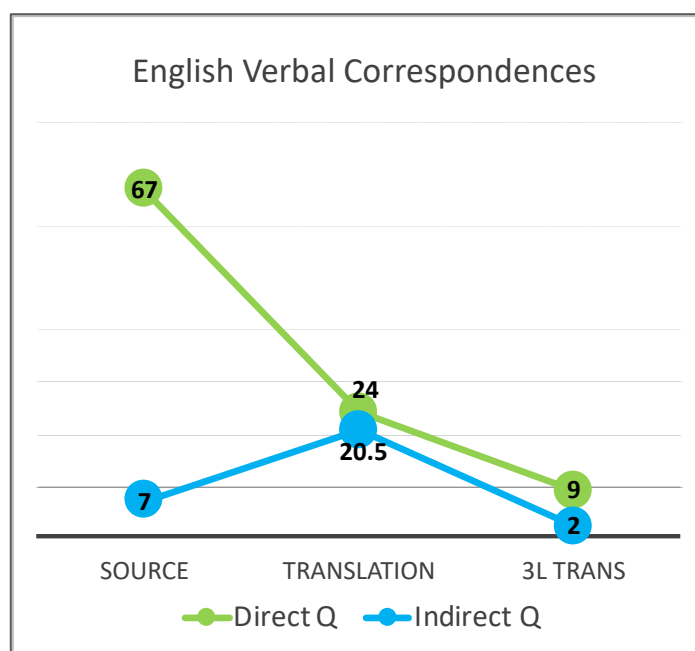


Figure 45. Speech act of English verbal correspondences (avg.)

As observed in the figure, the English verbal correspondences (of verbless non-elliptical Russian questions) correlate with direct speech acts more than with indirect speech acts across all three types of language considered (i.e. originals, translations, and third-language translations).

The pattern is particularly pronounced when English is the original-source language, i.e. 67 cases verbal correspondences in direct speech acts and only 7 cases of verbal correspondences when the speech act is indirect. In other words, when English is the source-language, 90% of the English verbal correspondences occur with direct speech acts.

When English is the language of translation, the pattern persists but appears to even out. However, it must be emphasized that in the present results, the difference between the source-type and genre overlap, i.e. the English source language is from the scripted speech of CA_EO whereas the translation language belongs to discourse-based fiction (BK_E1 and BK_E2) and the third-language translation to narrative-based fiction (ST_E1, ST_E2). Therefore, the disambiguation between source-language and genre is at present not possible. Annotation on a wider variety of texts is needed to investigate whether the reason that the distinction is not as pronounced in the BK and ST subpart has to do with translation language or their genre, or perhaps another factor. In any case, what is key is that the pattern persists across all the examined directions. Thus, the conclusions that are presently drawn cannot concern the distinction between the types of languages, but only the fact that across all the types, English verbal correspondences of questions consistently associate with direct speech acts.

This analysis of verbal correspondences in terms of direct and indirect speech act reveals that in English the verb allows to distinguish the two types of speech acts in the way that is illustrated in example (127).

(127) *English verbalization of questions in direct speech acts*

{The context is a debate and ‘excommunication’ is an argument in the debate. The speaker interrupts his interlocutor and utters two questions, (a) and (b), that immediately follow one another.}

a. Russian (BK RO)

Как	это	отлучение,
<i>kak</i>	<i>èto</i>	<i>otlučenie</i>
<i>how</i>	<i>this</i>	<i>excommunication</i>

English (BK E1)

What do you mean excommunication?

b. Russian (BK RO)

что	за	отлучение?
<i>čto</i>	<i>za</i>	<i>otlučenie</i>
<i>what</i>	<i>PART</i>	<i>excommunication</i>

English (BK E1)

What excommunication?

The context indicates that both of the questions are indirect, i.e. the speaker knows what ‘*excommunication*’ means and is not looking for an answer either in question (a) or in question (b). However, it is notable that in English, the first verbal question, i.e. (a) ‘*What do you mean excommunication?*’, is immediately reworded with a verbless one, i.e. (b) ‘*What excommunication?*’. The present results indicate that the reason for this rewording, i.e. the omission of the verb, is likely to emphasize the nature of the speech act. This example illustrates the association of the English verbal correspondences with indirect speech acts which was quantitatively discovered through the annotation above.

In other words, the absence of the verb is observed to be used in English in order to insist on the fact that the speech act is *indirect*, that it does not seek a response, but rather seeks to express disagreement with the addressee. The correspondence results show that the absence of the verb tends to be used as a grammatical marker in English in order to signal that a questioning speech act is indirect.

Note that in the Russian example in (127), both the question in (a) and the question in (b) are verbless. Unlike in English, where the absence of the verb is used to mark the difference, in Russian it is the emphatic particle ‘*za*’ (*za*) that is observed to insist on the rhetorical intent of the question.

For the contrastive analysis of Russian in terms of its correspondences, it is necessary to annotate a much larger corpus in the same way as has been done in this chapter. In the texts considered, there were only *three* verbal correspondences of English verbless non-elliptical questions. The present suspicion is that the verb will not play this central role in marking the contrast between direct and indirect speech acts in Russian. What we see in Russian is the frequent presence of emphatic particles, and we therefore make the hypothesis that it is through particles that Russian may mark this contrast. If we suppose that in Russian this distinction is not marked by the verb, while in English, as we have seen it is, this would help to explain why Russian has almost four times more verbless questions than English. Annotation in order to verify this hypothesis is the next step in a future study.

PART FIVE:

CONCLUSION

VERBLESS SENTENCES

One of the key goals of the present work has been to defend the phenomenon of the verbless sentence as a legitimate one, both on theoretical and empirical grounds. We also hope to have made, broad, but nevertheless key, strokes in the semantico-pragmatic portrait of the prototypical verbless sentence and its use in Russian and English. If we were successful in these aims, it means that a lot of work lies ahead, including expanded targeted annotation.

More specifically, the following contributions are esteemed to have been made with the present dissertation:

1. We have provided a *sentential account* that justifies verbless sentences as constituting a legitimate linguistic phenomenon. As shown in *Part I*, the status of verbless structures has fueled a rich debate throughout history. By taking stock of the existing arguments on the issue, and considering them from both a theoretical perspective and the perspective of the current data, we have attempted to move the ‘*Krylovian cart*’ toward the recognition of the *sentential* status of *verbless* structures, *without*, however, attributing to them hidden verbal properties. Verbless sentences exist, and they are entirely verbless – both in terms of their syntax and their semantics. The fact of recognizing their existence, and of admitting their lack of a syntactic predicate, reveals that the frontiers of syntactic models of language are drawn at idealized canonical structures. Attempts to introduce hidden structure or attribute verbal inflectional categories to non-verbal elements, only delay the inevitable conclusion that is already powerfully coming to the surface in arguments such as made by Ellen Barton, Ljiljana Progovac, Robert Stainton and Reinaldo Elugardo. Only, in their justified arguments, these latter seem to have chosen the flight option by relegating the structures to ‘*nonsentential*’, whereas we, while

agreeing with their critiques of existing attempts, think it is nevertheless necessary, and possible, to defend the status of verbless structures, whatever the cost, in the interest that our constructed models reflect language use.

2. Having defended their existence, we also esteem to have made progress in the *semantic and pragmatic account* of verbless sentences in Russian and English. Particularly, we have combined the strengths of several disciplines into one method and investigated the semantico-pragmatic features of the phenomenon using a *contrastive corpus approach*. In *Part 2*, we outlined the uniqueness of our methodological framework. It concerns not only the particular choice of languages, Russian and English, for their typological differences. It also involves the design and construction of the corpus in a way that will address the concerns with using translations for the study of linguistic phenomenon. Furthermore, the concerns that arise with using corpora for more qualitative-based semantico-pragmatic analysis were addressed through careful development of the corpus in terms of accuracy, searchability, context retrieval and the use of calculations. In this way, we hope to have begun to lift some of the data limits on previous research and quenched some of the thirst for the use of corpora in the study of the verbless sentence as a linguistic phenomenon.
3. The task of *retrieving verbless sentences in corpora* is a major roadblock that has been presently dispersed. We have developed a method for the automatic retrieval of verbless sentences. In addition, this method not only finds verbless sentences but also finds them in such a way that allows their further analysis – semantically, contrastively and from a textometric perspective. It retrieves the verbless sentence, preserves the context, aligns the verbless sentence and the context with multiple translations, finds the translations of the verbless sentence automatically, permits the access to data that is necessary for semantico-pragmatic analysis and annotation of originals and translations, and makes possible textometric computations concerning the verbless sentence, its translations and its context against a reference corpus.
4. Finally, a *new parallel and comparable corpus* of thirteen works of 19th – 21th century Russian and English realist fiction has been born. Measuring 1,4-million-words, it is of a size that is particularly large both for parallel-corpus studies and for

verbless phenomenon studies. Multiple translations per original work comprise the 32 texts. The entire corpus is morpho-syntactically annotated (with particular attention to the accuracy concerning the verbless and verbal phenomenon), it is sentence segmented (including the segmentation of direct speech from narration), and it is entirely sentence-level aligned across multiple translations.

In our analysis, we have striven to combine qualitative and quantitative perspectives in order to address several questions that pertain to the typological similarities and differences between Russian and English. The implications of the present analysis concern in particular the following questions.

- A. To what extent can syntactic ellipsis explain the phenomenon of the verbless sentence?

We started by establishing that the difference between Russian and English in terms of verbless sentences is statistically significant, as was expected. Russian was found to show almost twice as many verbless sentences as English across the present 1.4-million-word corpus. We then investigated the well-known typological difference between Russian and English which pertains to a particularly heightened productivity of syntactic ellipsis in the former language. Having annotated verbless sentences (and their translation correspondences) in terms of syntactically ‘elliptical’ and ‘non-elliptical’ utterances, we observed that syntactic productivity of ellipsis is entirely unsuitable for explaining the variation in verbless sentence frequency between the two languages. First, we found that ellipses were significantly overrepresented in English, and not in the syntactically more elliptically-productive Russian. Secondly, verbless sentences were found to be predominantly *non-elliptical* in both languages. Additional results are found in *Part 4.1*. Such results suggest that syntactic ellipsis cannot explain the phenomenon of the verbless sentence, and call for its semantic and pragmatic analysis.

Furthermore, these results also support the treatment of verbless sentences as largely syntactically independent structures. If verbless sentences were mere syntactic reductions of verbal sentences, the Russian productivity in this domain should have explained the frequency differences between the languages. However, it does not. Consequently, the present results also imply that verbless sentences are *not* syntactic reductions of verbal sentences.

- B. What semantic and pragmatic features distinguish verbless sentences from verbal sentences in the two languages?

One of the ways in which we provided insight into this question is by using a combination of textometric corpus tools and manual annotation. The discussion in *Part 4.2* takes a look inside the prototypical verbless sentence. Characteristic elements and n-grams that set verbless sentences statistically apart from the rest of the corpus were calculated, subsequently observed in context and in parallel with their translations, and then classified according to a semantico-pragmatic schema. The results explore the various statistically key functions of verbless sentences in the two languages, including the marking of agreement and disagreement, expressiveness, quantification, formality and informality, immediacy and deixis, questioning, and emphasis. Comparing the functions in the two languages, indefinite reference and *wh*-words are found to be key in English verbless sentences, whereas the emphasis of intensity is revealed to be particularly important for Russian verbless sentences. Results in this section also include an analysis of length and lexical complexity. Finally, we assess the potential level of semantic conventionalization of verbless sentences based on an analysis of their statistically characteristic repeated segments (the n-grams inside that set them apart from other sentences). Based on these findings, we argue that verbless sentences are not nearly as conventionalized as they are sometimes assumed to be.

- C. How does translation language and genre affect the verbless sentence phenomenon?

We show that from a contrastive perspective verbless sentences behave similarly in originals, translations and third-language translations in terms of verbal translation correspondences. From a monolingual perspective, only a slight difference in frequency is observed. Verbless sentences are therefore shown to be a language-specific linguistic phenomenon that is not significantly affected by translation. Rather, our results show that considerations of genre, and the difference between direct speech and narration, is of greater significance for the verbless sentence. A correlation is exposed between verbless sentences and direct speech (through several means including manual annotation for ‘direct speech’ and ‘narration’, frequency variation across genre, and statistically characteristic lexical elements). This correlation highlights the centrality of pragmatic function, common ground and interactive settings for verbless sentence use. We further propose that verbless sentences may be a linguistic feature of genre differences.

D. Do verbless sentences omit predictable and recoverable information?

We explored this question from a contrastive perspective by analyzing the information structure of the verbless sentences and their translations. Through annotation of verbless sentences and their translation correspondences, we show that the verbs that are conversationally implicated by verbless sentences are largely a part of the information structural *Focus*. This finding challenges the extent to which the verbless sentence can be accounted for in terms of a predictable and reconstructable predicate, and provides further evidence in support of the argument that runs throughout the present thesis, that verbless sentences can be sufficient to express a complete thought and satisfy the requirements for constituting full instances of predication.

With regard to information structure, we also show that verbless sentences are not restricted to a particular information structure type and are observed as topic-comment, identificational andthetic sentences in both languages.

Furthermore, annotation reveals a contrastive difference between the instantiation of the informational topic in the Russian verbless sentences and their English translations. The instantiation of the topic in English is observed to correlate with the instantiation of the verb. We propose that some of the frequency differences between the languages concerning verbless sentences may be in part be due to typological pragmatic differences in topic activation.

E. How do semantic categories associated with the verb, including temporal and aspectual meaning, relate to verbless sentences?

The semantic properties typically associated with the verb are presently argued to be a matter of pragmatic implicature as opposed to semantic entailment. Exploring the question from a contrastive approach, we show that verbless sentences in both languages have the potential to conversationally implicate a wide range of temporal, aspectual and verbal lexical meaning. Manual annotation was used to examine situation and viewpoint aspect of verbal correspondences. In addition, the verbal correspondences of the full corpus provide insight into the potential lexical and temporal meaning of the conversationally implicated verbs of the verbless sentences in both languages.

F. How are verbless sentences used in speech acts?

We show that in the verbless non-elliptical questions of both languages, the absence of the verb tends to be associated with indirect speech acts, i.e. situations in which the speaker is not requesting information from the addressee.

Furthermore, we present evidence that the absence of the verb may be used as a grammatical marker in English to signal that a question is an indirect speech act. Contrastive analysis shows that questions, compared to other types of utterances, are particularly sensitive to the presence and absence of the verb. In English, the verb allows to distinguish between direct and indirect speech acts, whereas in Russian the direct-indirect distinction is not observed to be related to the verb. Such a result suggests that contributing to the explanation of verbless sentence frequency differences between the languages is a typological difference in the pragmatic use of the verb in questions.

To conclude, in addition to language specific insights concerning the semantico-pragmatic features of verbless sentences, the wider implications of this dissertation pertain to the theoretical account of the verbless sentence. The present analysis has contributed arguments toward more general questions, including: *Are verbless sentences reductions of verbal sentences? To what extent is their meaning a matter of assertion versus implicature? Where do verbless sentences stand in the debate between compositional versus conventional meaning?*

We have proposed an account of the sentential status of verbless structures that does not rely on a hidden syntactic structure, but rather insists on the primacy of information structural focus as a key sentential criteria. Our argument is that the semantic content of a verbless sentence should be treated in terms of linguistically explicit asserted elements and contextually-revealed conversation implicatures.

The final contribution of the present thesis concerns *corpus treatment of the absence of the verb*. The present method of verbless sentence retrieval and corpus design was developed through a multi-step process, consisting of several smaller pilot studies that had qualitative aims at every step. This has enabled us to draw attention to the specific challenges facing the automatic processing of the open class of grammatical structures centered on the absence of the verb, and also to overcome a lot of them in the present study. In doing so, we hope to have provided perspectives for further contrastive treatment of verbless sentences in corpora.

APPENDICES

APPENDIX 1

Corpus Labels

Each text of the corpus has a unique label of 5 to 6 characters, e.g., BK_R0 or ST_E1F. These labels are used to structure the corpus parts and also to specify the source of the examples used in-text. The label includes the following categories.

Characters 1-2: Title of the Aligned Novel

BK	Братья Карамазовы (Brothers Karamazov)
TD	Тихий Дон (Quiet Flows the Don)
OD	Один День Ивана Денисовича (One Day in the Life of Ivan Denisovich)
PO	Пикник на Обочине (Roadside Picnic)
KF	Книга без Фотографий (A Book without Photographs)
OS	Олень и Шалашовка (The Love-girl and the Innocent)
ST	The Stranger
SR	The Sun Also Rises
AF	Animal Farm
FH	Fahrenheit 451
JS	Jonathan Livingston Seagull
HP	Harry Potter and the Order of the Phoenix
CA	The Caretaker

Character 3: _

Character 4: Language

R Russian

E English

Character 5: Original or Translation Number

0 Original source language

1 Translation #1

2 Translation #2

3 Translation #3

Character 6: Additional Information about the Translation

F The translation is from the third language (i.e. French)

Q The translation is a phantom one required for alignment

In addition, the examples used in the text specify a ‘position number’ which corresponds to the number of the first word of the particular sentence in the corpus (see *Part 2: Section 4.4.2*).

APPENDIX 2

Trameur Illustration

This appendix illustrates some of the key Trameur features from the perspective of the present analysis, as discussed in *Part 2*. The illustrations include:

- Frame and Thread data model (*Figure 46*)
- User-controlled multilayered annotation: search and edit (*Figure 47*)
- User-controlled multilayered annotation: concordance line editing (*Figure 48*)
- Projection of annotations onto Section Map (*Figure 49*)
- Multiple Correspondence Bi-Text Alignment (*Figure 50*)
- The new Section Annotation feature (*Figure 51*)
- Concordance on the sentence level (*Figure 52*)
- Summary parameters of the present corpus (*Figure 53*)

Frame & Thread

The screenshot shows the 'Le Trameur' software interface. The main window displays a list of frames under the heading 'Le Cadre Lexicométrique'. Each frame is represented by a checkbox and a code string, such as '<partie(1|pos=1)>>' and '<volet(BK_R0|pos=21)>>'. A blue arrow points from the text 'Frame (the structure)' to the first frame. Below the list, a detailed view of a thread is shown, with a blue arrow pointing from the text 'Thread (the text)' to it. The thread details include: 'Position: <24>', 'Forme: <недоуменная> | Freq: 4', 'Lemme: <недоумение> | Freq: 21', and 'Cat.: <Ncnsln> | Freq: 1945'. A blue arrow points from the text 'The form 'недоумению' has position 24 in the string of text-&-delimiters that constitute the corpus. This string runs from 1 to 4,015,790 positions. This form occurs 3 times in the corpus; its lemma 'недоумение' 21 times; its morphosyntactic category 'Ncnsln' 1945 times.' to the thread details. The interface also includes a menu bar with options like 'Cadre', 'Ventilation', 'Section', 'Forme-Lemme', 'Catégorie-Tag', 'Segment', 'Cgoc', 'Stat', 'Concordance', 'Patron', 'Graphe', 'Relation', 'Sélection', 'Rapport', and 'Param'. A toolbar at the bottom contains various icons for file operations and search. The status bar at the bottom indicates 'NB.L. Partie : volet=BK_R0 (1,44813) Annotation : 1 Aperçu : 50' and 'Навигация жизнеописания героя моего, Алексея Федоровича Карамазова, находится в некотором недоумении. \$'.

Figure 46. Trameur: Frame & Thread

User-Controlled Multilayered Annotation

Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V. 12.170

Cadre Ventilation Section **Forme-Lemme** Catégorie-Tag Segment Cgoc Stat Concordance Patron Graphe Relation Sélection Rapport Param

Recherche Forme/Lemme

Recherche Forme(s) (Dico Forme):

(Ecrire motif supra puis Entrée)

Recherche Forme(s) (Dico Lemme):

(Ecrire motif supra puis Entrée)

Aides

TagCloud sur Annotation sélectionnée

Génération TagCloud

Annotation sélectionnée:

Délimiteur de sections:

Stop-liste (édition)

Seuil Min | Seuil Max

Formes à filtrer (regeexp)

Motif:

Pour générer un usage:

- Sélectionner une annotation
- Charger une stop-liste
- Définir un motif de filtrage
- Modifier les seuils
- Lancer la génération

Forme		Lemme	
Fq	Forme	Fq	Lemme
1208	more	2694511	RETURN
1206	yme	269456	\$
1198	some	126280	'
1191	man	119906	'
1165	only	118029	QQ
1154	us	46612	the
1130	He	34311	"
1120	konna	29595	-
1110	two	27947	be
1089	cane	24471	'
1077	m	23678	and
1066	Hero	20492	x
1062	here	18436	to
1061	en	18073	a
1052	no	15410	of
1039	away	14738	I

List of forms and their frequency in the corpus

List of lemmas and their frequency in the corpus

Modification des annotations sur toutes les occurrences (m)

Forme.....

Lemme.....

Catégorie.....

Pour les champs Lemme, Catégorie et annotations importées, si la zone de saisie reste vide, la valeur initiale sera conservée.

Here the form 'm', i.e. the verbal clitic, is edited as 'be' on the lemma layer and 'VBP' on the morphosyntactic category layer

Figure 47. Trameur: User-Controlled Multilayered Annotation – search and edit

Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V.12.170

Cadre Ventilation Section Forme-Lemme Catégorie-Tag Segment Cgoc Stat Concordance Patron Graphe Relation Sélection Rapport Param

contexte-partie | Tri-Concordance | Shift-Clic(pôle) : sélection | Clic-droite(pôle) : édition des annotations | Shift-Clic-droite(pôle) : relation | Ctrl-Clic-droite(pôle) : recherche relation

Concordance de < sélection Positions >

partie=1
 never read a book about it. \$ Go on and read a book all full
 sure. \$ Possible. " \$ Go on. \$ Tell me about it.
 anything I could remember. " \$ Go on. \$ Take that drink and
 Remember the nigger perfectly. " \$ Go on. " \$ " Wonderful nigger.
 some more travel stories. " \$ Go on. " \$ " Let 'us
 I asked \$ Go up to the café and see Brett
 It ' sv a lovely nose. \$ Go on, point it at me. \$
 are a lovely piece. " \$ Go on to the fight, " \$
 " Comment? " \$ Go on to hell! " \$ said Bill.
 I laughed. " Yes. \$ Go on and laugh, " \$ said
 mouth full of pity. " \$ Go on, " \$ I said. \$
 reflect discredit on me. " \$ Go on. \$ Tell it, Brett.
 Can you blame them? " \$ Go on to hell, Mike. " \$ "
 I can ' t. " \$ Go on. \$ We ' re all
 " \$ Bill said. " \$ Go on, tell him. \$ Tell him
 Brett and listening to her. " \$ Go on. \$ Tell him! " \$
 ' re not wanted, Cohn? \$ Go away. \$ Go away, for God
 wanted, Cohn? \$ Go away. \$ Go away, for God's sake. \$
 \$ He turned to Cohn: " \$ Go away! \$ Go away now! "
 away! \$ Go away now! " \$ " But
 to bed. \$ Go on to bed. " \$ "
 Go on in there and dance with Mike
 me. " \$ Go on to hell! " \$ " She
 is? " \$ Go on to hell! " \$ " I
 Y. \$ Go on. " \$ The waiters moved
 m. " \$ Go on. \$ Go on up and
 n. \$ Go on up and see him. "
 t you. \$ Go on up and see Cohn. \$
 . \$ " Go sit at it, " \$ Bill
 reached. \$ Go on, anyway, shove the bore down
 in? " \$ Go ahead. " \$ " How did
 y. \$ " Go on, " \$ said the woman.
 chest. " \$ Go on, " \$ said the woman,
 m. \$ 2 \$ " Go let him in, will you? \$
 things. " \$ Go take the beetle. " \$ 2
 hit dogs. \$ Go take the beetle. " \$ "
 home. \$ Go to bed. \$ Why waste your
 Montay's head was clear. " \$ Go to the firehouse when it ' sv

Multilayer Annotation Control

Here the form 'Go' is edited for a specific concordance line

Correction Item
(position : 902119)

Forme..... Go
 Lemme..... Go
 Catégorie..... NF

Enregistrer
 Annuler

Annotations : 1 2 3

Concordance d'item :
 (sélection po) RegExp
 (Ecrire motif supra puis Entrée)
 Sélection Annotation :
 Forme Lemme Catégorie
 n'Annotation
 Annotation sélectionnée :
 1 Forme
 Fenêtre concordance : 15
 Coloration des annotations

Parties

partie
volet

Export Concordance :
 Ajouter au rapport

Figure 48. Trameur: User-Controlled Multilayered Annotation – concordance line editing

Projection of Annotations onto the Section Map

Le Trameur - Le Métier: Lexicométrique @CLA2T-P3 v. 12.170

Ajout Nouvelle Couverte Annotation
 Importation d'une annotation :
 Visualisation d'annotation (>3)
 Toutes les annotations
 N° 4

Cadre Ventilation Section Forme-Lemme Catégorie-Tag Segment Cooc Stat Concordance Patropl Graphe Relation Sélection Rapport Param

Dictionnaire de catégories (via tree-tagger) Dictionnaire de annotations importés: co-présence

Catégorie (tree-tagger) Projection sur la carte des sections des catégories sélectionnées in importée Nom Nom

iq	Forme-Lemme	Catégorie-Tag	Segment	Cooc	Stat	Concordance	Patropl	Graphe	Relation	Sélection	Rapport	Param
4905	WRB		3									
82	Wp\$		3									
5188	WP		3									
1135	WDT		3									
8843	WVZ		3									
8590	WVP		3									
5126	WVN		3									
16403	WVG		3									
56400	WVD		3									
22681	WV		3									
485	Waps-snp		3									
2	Waps-snpse		3									
4	Waps-snp-pn		3									
6	Waps-snp-pl		3									
8	Waps-snp-pl		3									
12	Waps-snp-pg		3									
15	Waps-snp-pa		3									
37	Waps-snp-fpn		3									
7	Waps-snp-fpl		3									
12	Waps-snp-fpi		3									
13	Waps-snp-fpg		3									
4	Waps-snp-fpd		3									
42	Waps-snp-fpa		3									
2	Waps-snp-feg		3									
1	Waps-snm-pn		3									
6	Waps-snm-fpn		3									
1	Waps-snm-fpg		3									
7	Waps-snm-fpa		3									
1	Waps-snm-en		3									
2	Waps-sma-pl		3									
1	Waps-sma-pi		3									
3	Waps-sma-pa		3									
9	Waps-sma-fpn		3									
1	Waps-sma-fpl		3									
2	Waps-sma-fpi		3									
2	Waps-sma-fpg		3									

[Projection of Annotations onto Section Map](#)
 Here the verbal morphosyntactic categories are selected to be projected onto the sentence-containers on the Section Map. This separates all of the sentences in the corpus into those that contain and those that do not contain the selected categories. In other words, into verbal and verbless sentences.

Figure 49. Trameur: Projection of annotations onto the Section Map

Multiple Correspondence Bi-Text Alignment

The screenshot displays the Le Métrier Lexicométrique software interface, which is used for multiple correspondence bi-text alignment. The interface is divided into several panels:

- Top Panel (Section Map):** Titled "Section Map (4 texts containing 16 sentences each)", it shows a grid of checkboxes for mapping between "English Original" and three "Russian Translation" sections. A "context" label points to a specific area in the grid. The text "Verbless sentences and their multiple translation correspondences are retrieved" is displayed above the grid.
- Left Panel (Configuration):** Contains settings for "Changement de la Carte des sections", "Définition de sections", "Recherche Forme sur la carte", and "Spécificités sur Sections".
- Main Panel (Bi-Text Alignment):** Shows the alignment of the selected sections. It displays the text "But alas!" in English and "Но, увы, сими уже оставили еро." in Russian. A tooltip provides morphological data for the word "alas":


```
Position: <617>
Forme: <alas>|Freq:1
Lemme: <alas>|Freq:1
Cat: <OH>|Freq:1
a-00004: <unknown>|Freq:913
a-00005: <-1>|Freq:855
```
- Right Panel (Annotations):** Shows the alignment of the three Russian translations: "Russian 1: Verbal", "Russian 2: Displaced", and "Russian 3: Verbless".

Figure 50. Trameur: Multiple Correspondence Bi-Text Alignment

The New Section-Annotation Feature

Changement de la Carte des sections :

Délimiteur de sections : § Partie

La carte des sections peut être construite soit en choisissant un délimiteur soit en choisissant une partie du cache (nom de partie)

Parties

wo.let

Recherche Forme sur la carte :

(*WZZZ\$W*) RegExp

Spécificités sur Sections

BI-TEXT

V1 1 V2 ?

TMX

Sélection Annotation :

Forme Lemme Catégorie

Annotation sélectionnée :

Forme 1

(New) Section Annotation

This adds another annotation layer to all of the sentences after projection. The first word of a verbal sentence is marked 1, the first word of verbless sentence is marked 0. The 0 annotation is then selected to view concordance lines of verbless sentences.

Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)

Control-clic sur marqueur de page : sélection 5 sections | Shift-control-clic sur marqueur de page : sélection 25 sections (1 ligne)

Ho, ysm, cymu yme ocraasuru ero.

Ho ysm! - §

Annotations : 1 2 3 4 5

Figure 51. Trameur: The new Section Annotation feature

Concordance on the Sentence Level & Displaced Alignment

Le Trameur - Le Métier Lexicométrique @CLA2I-P3 V. 12.170

Shift-Click(pôle) : sélection | Clic-droit(pôle) : édition des annotations | Shift-Clic-droit(pôle) : relation | Ctrl-Clic-droit(pôle) : recherche relation

contexte-partie Tri-Concordance

Concordance de < sélection Positions >

volet=EN

no animal had killed another animal. \$ Not even a rat had been killed.
 : that is the plain truth. \$ 7 \$ But is this simply part
 have smashed the van to matchwood. \$ But alas \$ His strength had left
 volet=RU1

животные не убивали друг друга. \$ Даже крыс никто не трогал. \$ Они
 животных - страдания и рабский труд. \$ 7 \$ Но, жемет-бьют, такое закон
 , чтобы разнести фурун в щепки. \$ Но, уям, силн уже оставили его
 volet=RU2

чтобы одно животное убило другое. \$ Даже крысу. \$ И вот все поднялись
 ищите -- вот голая правда. \$ 7 \$ Но, может быть, это в
 но сейчас он был слишком слаб. \$ 2 \$ И очень быстро затих. \$
 volet=RU3

не покупалось на жизнь своего соплеменника. \$ Перестали убивать даже крыс. \$ Все побрели
 характеризуется так - нищета и рабство. \$ Такова жизнь. \$ Но такое ли должен
 командами Екхсера фурун разлетится в щепки. \$ Но уям! - \$ Силн уже покинули

English: Verbless

Russian 1: Verbal

Russian 2: Displaced

Russian 3: Verbless

Position: <2153>
 Forme: <но> | Freq: 8
 Lemme: <но> | Freq: 11
 Cat: <C> | Freq: 66
 a-000004: <unknown> | Freq: 913
 a-000005: <0> | Freq: 12

Concordance Lines for Verbless Sentences
 & their Alignment Correspondences

Viewing concordance lines of the verbless sentences
 and their correspondences makes it possible to find
 translations that have been displaced

Concordance d'item :
 sélection po FreqExp
 (Ecrire motif supra puis Entrée)
 Sélection Annotation :
 Forme Lemme Catégorie
 n'Annotation
 Annotation sélectionnée :
 5 a-000005
 Fenêtre concordance :
 15
 Coloration des annotations
 Parties
 Export Concordance :
 Ajouter au rapport

Annotations : 1 2 3 4 5

Figure 52. Trameur: Concordance on the Sentence Level and Displaced Alignment

Summary Parameters of the Present Corpus

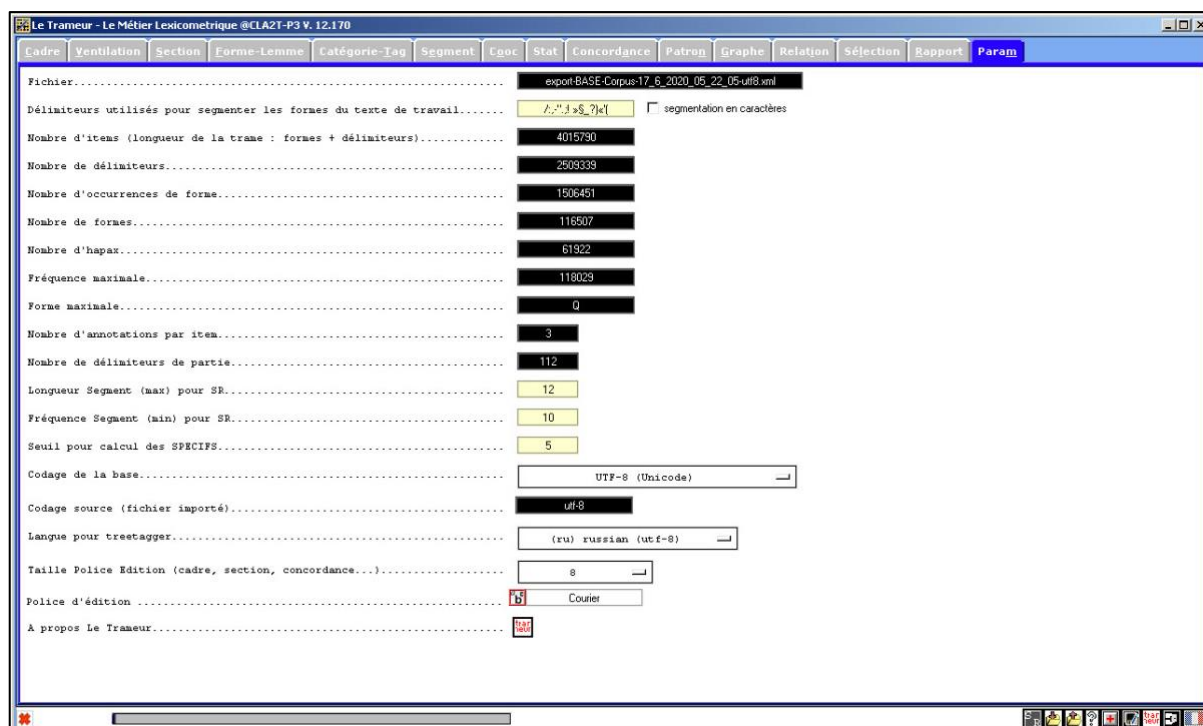


Figure 53. Trameur: Summary parameters of the present corpus

Number of items (word-forms + delimiters)	4,015,790
Delimiters	2,509,339
Number of form occurrences (i.e. tokens)	1,506,451
Number of forms (i.e. types)	116,507
Number of hapax	61,922
Max frequency	118,029
Max form	Q
Number of annotations per item	3
Number of part delimiters	112
Encoding	UTF-8

APPENDIX 3

Morphosyntactic Specifications

English

The English Part-Of-Speech specifications used in automatic tagging correspond to the TreeTagger Penn Treebank Tagset which includes 58 tags and can be found on:

<https://courses.washington.edu/hypertext/csar-v02/penntable.html>

Russian

The Russian morphosyntactic specifications used in automatic tagging correspond to Version 4 (2010) of the MULTITEXT-East framework (documented in Sharoff et al., 2008). It includes approximately 600 tags and can be found on: *<http://nl.ijs.si/ME/Vault/V4/msd/html/>*

MORPHOSYNTACTIC SPECIFICATIONS: ENGLISH
(Tree Tagger Tag Set)

POS Tag	Description	Example
CC	coordinating conjunction	<i>and, but, or, &</i>
CD	cardinal number	<i>1, three</i>
DT	determiner	<i>the</i>
EX	existential there	<i>there is</i>
FW	foreign word	<i>d'œuvre</i>
IN	preposition/subord. conj.	<i>in, of, like, after, whether</i>
IN/that	complementizer	<i>that</i>
JJ	adjective	<i>green</i>
JJR	adjective, comparative	<i>greener</i>
JJS	adjective, superlative	<i>greenest</i>
LS	list marker	<i>(1),</i>
MD	modal	<i>could, will</i>
NN	noun, singular or mass	<i>table</i>
NNS	noun plural	<i>tables</i>
NP	proper noun, singular	<i>John</i>
NPS	proper noun, plural	<i>Vikings</i>
PDT	predeterminer	<i>both the boys</i>
POS	possessive ending	<i>friend's</i>
PP	personal pronoun	<i>I, he, it</i>
PP\$	possessive pronoun	<i>my, his</i>
RB	adverb	<i>however, usually, here, not</i>
RBR	adverb, comparative	<i>better</i>
RBS	adverb, superlative	<i>best</i>
RP	particle	<i>give up</i>
SENT	end punctuation	<i>?, !, .</i>
SYM	symbol	<i>@, +, *, ^, , =</i>
TO	<i>to</i>	<i>to go, to him</i>
UH	interjection	<i>uhhuhhuhh</i>
VB	verb <i>be</i> , base form	<i>be</i>
VBD	verb <i>be</i> , past	<i>was were</i>
VBG	verb <i>be</i> , gerund/participle	<i>being</i>
VBN	verb <i>be</i> , past participle	<i>been</i>
VBZ	verb <i>be</i> , pres, 3rd p. sing	<i>is</i>
VBP	verb <i>be</i> , pres non-3rd p.	<i>am are</i>
VD	verb <i>do</i> , base form	<i>do</i>
VDD	verb <i>do</i> , past	<i>did</i>
VDG	verb <i>do</i> gerund/participle	<i>doing</i>
VDN	verb <i>do</i> , past participle	<i>done</i>

VDZ	verb <i>do</i> , pres, 3rd per.sing	<i>does</i>
VDP	verb <i>do</i> , pres, non-3rd per.	<i>do</i>
VH	verb <i>have</i> , base form	<i>have</i>
VHD	verb <i>have</i> , past	<i>had</i>
VHG	verb <i>have</i> , gerund/participle	<i>having</i>
VHN	verb <i>have</i> , past participle	<i>had</i>
VHZ	verb <i>have</i> , pres 3rd per.sing	<i>has</i>
VHP	verb <i>have</i> , pres non-3rd per.	<i>have</i>
VV	verb, base form	<i>take</i>
VVD	verb, past tense	<i>took</i>
VVG	verb, gerund/participle	<i>taking</i>
VVN	verb, past participle	<i>taken</i>
VVP	verb, present, non-3rd p.	<i>take</i>
VVZ	verb, present 3d p. sing.	<i>takes</i>
WDT	wh-determiner	<i>which</i>
WP	wh-pronoun	<i>who, what</i>
WP\$	possessive wh-pronoun	<i>whose</i>
WRB	wh-abverb	<i>where, when</i>
:	general joiner	<i>;; , --</i>
\$	currency symbol	<i>\$, £</i>

Figure 54. English morphosyntactic specifications

MORPHOSYNTACTIC SPECIFICATIONS: RUSSIAN
(Multitext-East Version 4)

Table of Categories			1. Specification for Noun (N)			2. Specification for Verb (V)		
Name	Code	Attributes	P Attribute	Value	Code	P Attribute	Value	Code
Noun	N	6	0 CATEGORY	Noun	N	0 CATEGORY	Verb	V
Verb	V	10	1 Type	common	c	1 Type	main	m
Adjective	A	6		proper	p		auxiliary	a
Pronoun	P	7	2 Gender	masculine	m	2 VForm	indicative	i
Adverb	R	1		feminine	f		imperative	m
Adposition	S	3		neuter	n		conditional	c
Conjunction	C	4		common	c		infinitive	n
Numeral	M	6	3 Number	singular	s		participle	p
Particle	Q	1		plural	p		gerund	g
Interjection	I	1	4 Case	nominative	n	3 Tense	present	p
Abbreviation	Y	4		genitive	g		future	f
Residual	X	0		dative	d		past	s
				accusative	a	4 Person	first	1
				vocative	v		second	2
				locative	l		third	3
				instrumental	i	5 Number	singular	s
			5 Animate	no	n		plural	p
				yes	y	6 Gender	masculine	m
			6 Case2	partitive	p		feminine	f
				locative	l		neuter	n
						7 Voice	active	a
							passive	p
							medial	m
						8 Definiteness	short-art	s
							full-art	f
						9 Aspect	progressive	p
							perfective	e
							biaspectual	b
			10 Case	nominative	n			
				genitive	g			
				dative	d			
				accusative	a			
				locative	l			
				instrumental	i			

3. Specification for Adjective (A)

P Attribute	Value	Code
0 CATEGORY	Adjective	A
1 Type	qualificative	f
	possessive	s
2 Degree	positive	p
	comparative	c
	superlative	s
3 Gender	masculine	m
	feminine	f
	neuter	n
4 Number	singular	s
	plural	p
5 Case	nominative	n
	genitive	g
	dative	d
	accusative	a
	locative	l
	instrumental	i
6 Definiteness	short-art	s
	full-art	f

4. Specification for Pronoun (P)

P Attribute	Value	Code
0 CATEGORY	Pronoun	P
1 Type	personal	p
	demonstrative	d
	indefinite	i
	possessive	s
	interrogative	q
	relative	r
	reflexive	x
	negative	z
	nonspecific	n
	2 Person	first
second		2
third		3
3 Gender	masculine	m
	feminine	f
	neuter	n
4 Number	singular	s
	plural	p
5 Case	nominative	n
	genitive	g
	dative	d
	accusative	a
	vocative	v
	locative	l
6 Syntactic_Type	nominal	n
	adjectival	a
	adverbial	r
7 Animate	no	n
	yes	y

5. Specification for Adverb (R)

P Attribute	Value	Code
0 CATEGORY	Adverb	R
1 Degree	positive	p
	comparative	c
	superlative	s

6. Specification for Adposition (S)

P Attribute	Value	Code
0 CATEGORY	Adposition	S
1 Type	preposition	p
2 Formation	simple	s
	compound	c
3 Case	genitive	g
	dative	d
	accusative	a
	locative	l
	instrumental	i

7. Specification for Conjunction (C)

P Attribute	Value	Code
0 CATEGORY	Conjunction	C
1 Type	coordinating	c
	subordinating	s
2 Formation	simple	s
	compound	c
3 Coord_Type	sentence	p
	words	w
4 Sub_Type	negative	z
	positive	p

8. Specification for Numeral (M)

P Attribute	Value	Code
0 CATEGORY	Numeral	M
1 Type	cardinal	c
	ordinal	o
	multiple	m
	collect	l
2 Gender	masculine	m
	feminine	f
	neuter	n
3 Number	singular	s
	plural	p
4 Case	nominative	n
	genitive	g
	dative	d
	accusative	a
	locative	l
	instrumental	i
5 Form	digit	d
	roman	r
	letter	l
9 Animate	no	n
	yes	y

9. Specification for Particle (Q)

P Attribute	Value	Code
0 CATEGORY	Particle	Q
1 Formation	simple	s
	compound	c

10. Specification for Interjection (I)

P Attribute	Value	Code
0 CATEGORY	Interjection	I
2 Formation	simple	s
	compound	c

11. Specification for Abbreviation (Y)

P Attribute	Value	Code
0 CATEGORY	Abbreviation	Y
1 Syntactic_Type	nominal	n
	adverbial	r
2 Gender	masculine	m
	feminine	f
	neuter	n
3 Number	singular	s
	plural	p
	paucal	c
4 Case	nominative	n
	genitive	g
	dative	d
	accusative	a
	locative	l
	instrumental	i

12. Specification for Residual (X)

P Attribute	Value	Code
0 CATEGORY	Residual	X

Figure 55. Russian morphosyntactic specifications

APPENDIX 4

Verbal Tags

The morphosyntactic tags excluded from verbless sentences include the following.

Russian

V	all verbal categories whose tags start with a V (e.g. Vmps-snp, etc.)
ZZZ	the phantom alignment segments for existing texts
QQQ	the phantom alignment segments for non-existing texts
FW	foreign language

English

V	all verbal categories whose tags start with a V (e.g. VVZ, VBP, etc.)
MD	modal verbs
ZZZ	the phantom alignment segments for existing texts
QQQ	the phantom alignment segments for non-existing texts
FW	foreign language

APPENDIX 5

Contrastive Search & Interface

The new program was developed by Borissov et al. (2020) in Java. The program works together with Trameur's Thread/Frame data structure. Trameur's Section Map requires all sentence-segments to be displayed before verbless sentences can be projected and retrieved. The 269,456 sentence-segments of the present corpus proved to be too large.

The graphical interface of the new program neutralizes the need to display a very large number of interconnected sentence segments simultaneously before retrieval. In addition, two key features could, with further development, make it an effective complement to Trameur particularly for qualitative contrastive analysis on the sentence level.

The Search

The search feature finds verbless sentences containing specific elements (e.g. forms, lemmas, or morphosyntactic categories annotated in Trameur) and verbal correspondences.

Alignment-in-Context Display

Another valuable feature for qualitative analysis is that the context and sentence-level alignment may be displayed simultaneously.

The Search

Dashboard

Annotate

[Search](#)

- * Verbless sentences containing specific elements
- * Verbless sentences with verbal correspondences

← [Verbless Sentence](#)

s.find((w) => w.c.startsWith('V') || w.c == 'MD' || w.c == 'QQQ' || w.c == 'ZZZ' || w.c == 'FW') == null

Total

52 Titles, 4 translation variants	548
The Brothers Karamazov (Fyodor Dostoevsky)	3106
And Quiet Flows the Don (Mikhail Sholokhov)	1135
One Day in the Life of Ivan Denisovich (Solzhenitsyn)	1887
Roadside Picnic (Arkady Strugatsky)	1824
Book Without Pictures (Sergey Shargunov)	1709
Deer and Shaloshovka (Solzhenitsyn)	462
Stranger (Albert Camus)	2659
The Sun Also Rises (Ernest Hemingway)	

BK_RO

Matched: 295 | Pct: 11% | Size: 2529

Frequency in Russian Brothers Karamazov

11 | Violet Contribution: 53%

BK_E1

Matched: 139 | Pct: 5% | Size: 2529

Frequency in English BK Translation 1

5 | Violet Contribution: 25%

BK_E2

Matched: 114 | Pct: 4% | Size: 2529

Frequency in English BK Translation 2

Figure 56. Contrastive search

Alignment In-Context Display

The screenshot displays an alignment tool with two main panels. The top panel, titled "Alignment and Context Display", shows two segments of text in English, labeled BK_E1 and BK_E2. The bottom panel shows the corresponding Russian text. Annotations include "Eng 1: Verbal" and "Eng 2: Verbal" pointing to specific English segments, and "Russian: Verbless" pointing to a Russian segment. Blue arrows indicate the alignment between the Russian and English segments.

Panel 1: BK_E1

- 1472953 I would not, in fact, venture into these rather vague and uninteresting explanations but would simply begin without any introduction - if they like it, they'll read it as it is - but the trouble is that while I have just one biography, I have two novels.
- 1473058 The main novel is the second one - about the activities of my hero in our time, that is, in our present, current moment.
- 1473112 As for the first novel, it already took place thirteen years ago and is even almost not a novel at all but just one moment from my hero's

Panel 2: BK_E2

- 3038590 Now, I would not have indulged in these tedious and obscure explanations, I would simply have got on with my story, without any preamble - if they like it, they'll read it - but the trouble is, I have one life story and two novels.
- 3038692 The second novel is the main one; this concerns my hero's actions right up to the present time.
- 3038734 But the action of the first novel takes place as long as thirteen years ago, and is not so much a novel as a single episode in my hero's early

Panel 3: Russian Text

- 562 Ибо не только чудак не всегда чуждость и обособление, а напротив бывает так, что он-то пожалуй и носит в себе иной раз сердцевину целого, а остальные люди его эпохи - все, какими-нибудь наплывным ветром, на время почему-то от него оторвались ...
- 657 Я бы впрочем не пускаясь в эти весьма мелочливые и смутные объяснения и начал бы просто-зв-просто без предисловия: поправится, так и так прочтут; но беда в том, что жизнеописание-то у меня одно, а романов два.
- 735 Главный роман второй, - это деятельность моего героя уже в наше время, именно в наш теперешний текущий момент.
- 776 Первый же роман произошел ещё тринадцать лет назад, и есть почти даже и не роман, а лишь один момент из первой юности моего героя.
- 829 Обойтись мне без этого первого романа невозможно, потому что многое во втором романе стало бы непонятным.
- 865 Но таким образом ещё усложняется первоначальное мое затруднение: если уж я, то ~ есть сам биограф, нахожу, что и одного-то романа может быть было бы для такого скромного и неопределенного героя излишне, то каково же являться с двумя и чем объяснить такую с моей стороны звносчивость?
- 965 Теряясь в разрешении сих вопросов, решаюсь их обойти безо всякого разрешения.

Figure 57. Contrastive alignment in-context display

APPENDIX 6

Characteristic Elements

The present appendix shows what is happening inside the verbless sentence. More specifically, it reveals the lexical and grammatical elements that characterize verbless sentences in comparison to other sentences, as described in *Part 4: Section 2.1* and *Part 2: Chapter 5*. It shows the first 30 lines of the characteristic elements results which concern the forms, lemmas and morphosyntactic categories that are statistically key for verbless structures.

The *Specificity Index* indicates the strength with which the element is associated with the verbless sentence. The index also includes negative values, which reveal that the elements negatively associated with the verbless sentence (i.e. rejected by the structures).

In addition to the index, each element also shows a value for the total number of times that it occurred in the corpus (*Frequency Total*) and the number of times that it occurred in verbless sentences (*Frequency Part*). These values also merit attention in the analysis, i.e. a high specificity with low overall frequency is not as revealing as that with a high frequency.

As discussed in *Part 4: Chapter 2*, two other variables enter in the calculation of the index, including the *overall size of the corpus* and the *size of the corpus part*. The precise calculation is described in (Lebart et al., 1998: 130–136; Lebart & Salem, 1994). These two key numbers are also indicated in the tables below. The label ‘*totals*’ refers to the number of tokens in the full corpus (i.e. ‘nombre d’occurrences du texte global’). The label ‘*totalS*’ reveals the number of tokens in the verbless sentence subpart (i.e. ‘nombre d’occurrences dans la partie visée’).

The *verbless sentence corpus part* (i.e. ‘partie visée’) more specifically corresponds to all of the sections that do not contain: ‘V’, ‘MD’, ‘QQQ’, ‘ZZZ’, ‘FW’, as described in *Appendix 4*.

The calculation of the Russian characteristic elements, involved *only the Russian texts*, which corresponds to the following book numbers:

booksList = [1,2,3,4,5,6,7,21,22,23,24,25,26,27,37,51]

The calculation of the English characteristic elements, involved *only the English texts*, which corresponds to the following book numbers:

booksList = [8,9,10,11,12,13,15,16,17,18,19,20,29,31,35,49]

These book numbers correspond to the order of the texts in the structure of the overall corpus, which is given in *Figure 11 (Part 2: Section 4.4.2)*.

The characteristic elements of the *verbless sentences* are shown in the middle column. The left column shows the characteristic elements of the *left context* of the verbless sentences, i.e. 3 sentences to the left. The right column shows the elements that statistically characterize the *right context*, i.e. 3 sentences to the right of the verbless sentence.

The results shown include the following levels:

- a) ENGLISH *FORMS*
- b) RUSSIAN *FORMS*
- c) ENGLISH *LEMMAS*
- d) RUSSIAN *LEMMAS*
- e) ENGLISH *MORPHOSYNTACTIC CATEGORIES*
- f) RUSSIAN *MORPHOSYNTACTIC CATEGORIES*

The specifications for the morphosyntactic category level are provided in *Appendix 3*. Additional discussion of the method of analysis is found in *Part 2*.

Notably, the present results show the values for all 16 Russian texts and all 16 English texts. Calculation and comparison with the results of the texts considered individually, as well as with only translations, originals, and the specific genres, is the next step.

Further analysis would also benefit from the re-grouping of specific morphosyntactic categories, keeping in mind the present results. For instance, grouping the masculine and feminine nouns in Russian would give a more precise value for the specificity of nouns. A customized regrouping of some of the distinctions made in the 600 Russian morphosyntactic tags and the 28 English tags, using Trameur's editing feature, would give a better idea of the specificity of morphosyntactic categories; provided of course that a grouped distinction does not reveal itself as being significant to the verbless sentence phenomenon. At present the morphosyntactic categories were *not* grouped in order to view the full results and get a better idea of the groupings that would be relevant.

ENGLISH FORMS																		
Left Context				Verbless Sentences				Right Context										
<i>(sentlist = [-3,-2,-1]; totals = 806875; totalS = 132548)</i>																		
Form	Exact Index	Rounded Index	Frequency	Part	Form	Exact Index	Rounded Index	Frequency	Total	Frequency	Part	Form	Exact Index	Rounded Index	Frequency	Total	Frequency	Part
you	52.31368	53	8804	1999	Yes			541	318	said	92.43536	93	4513	1280				
said	45.41672	46	4513	1115	No	319.0212	320	702	312	you	48.74744	50	8804	1961				
What	43.50845	45	1484	461	What	218.6329	220	1484	325	What	42.5521	44	1484	455				
asked	35.86061	37	880	298	A	185.0496	186	1006	252	I	40.1773	41	14738	3006				
sv	35.34706	36	4525	1070	Good	165.2831	166	137	114	sv	39.08467	40	4525	1079				
You	34.62935	36	2344	622	Oh	155.4365	156	411	162	You	31.23176	32	2344	604				
How	24.04281	25	493	175	Not	145.8353	147	216	124	Yes	22.47156	23	541	182				
I	22.50616	24	14738	2871	Hey	80.87783	82	90	62	Il	16.96713	18	1695	412				
Yes	20.31503	21	541	178	Hello	78.94991	80	51	49	do	15.74344	17	2804	622				
Who	15.33327	16	237	91	All	74.10024	75	473	107	Well	15.23816	16	602	177				
Have	14.27005	15	149	65	Why	55.47077	56	462	89	Harry	15.21462	16	1525	370				
Where	13.561	15	291	101	Very	54.81731	56	50	39	your	12.66637	14	1812	414				
do	12.75333	14	2804	609	Right	52.96734	54	85	46	m	12.32292	13	1077	267				
is	11.32316	12	207	75	Nothing	51.46811	52	79	44	Oh	11.78358	13	411	124				
ll	10.17226	11	1695	381	Well	48.30408	49	602	92	re	11.03587	12	1414	328				
are	9.303412	10	1452	329	yes	47.53881	49	111	47	How	11.03448	12	493	140				
Jake	9.031546	10	121	48	Just	45.66071	47	232	60	Where	10.53221	12	291	93				
here	9.011901	10	1062	251	Thanks	44.49943	45	54	35	No	10.32499	11	702	182				
Want	8.727134	10	28	19	My	42.83473	44	338	67	Let	9.546833	11	311	95				
A	8.631088	10	1006	238	Fine	41.54058	43	33	28	voice	9.419982	10	532	143				
name	8.328229	9	236	75	Your	40.15022	41	178	50	Not	9.233622	10	216	72				
Do	8.16209	9	728	180	right	39.80101	41	1208	116	Do	9.226845	10	728	183				
Brett	7.695429	9	429	116	Ah	38.85786	40	70	35	A	9.05177	10	1006	238				
Article	7.498145	8	26	17	Me	35.75417	37	35	26	are	8.933807	10	1452	324				
She	7.238956	8	1012	232	Eh	34.7772	36	33	25	t	8.716439	10	6199	1182				
Well	7.163099	8	602	150	Four	33.82678	35	48	28	Brett	8.247779	9	429	117				
Jenkins	6.8733	8	11	10	To	33.0602	34	208	47	Montag	7.655226	9	453	120				
Montag	6.603611	8	453	117	Yeah	32.8657	34	96	35	Noonan	7.654389	9	161	55				
have	6.436684	7	2806	561	Sure	31.18355	32	36	24	Have	7.628173	9	149	52				
That	6.277447	7	890	203	Really	31.13204	32	32	23	That	7.384635	8	890	207				

Figure 58. ENGLISH FORMS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN FORMS												
Left Context				Verbless Sentences				Right Context				
<i>(sentList = [-3,-2,-1]; totalS = 581547; totalI = 173428)</i>				<i>(sentList = [0]; totalS = 581547; totalI = 38022)</i>				<i>(sentList = [1,2,3]; totalS = 581547; totalI = 174276)</i>				
Form	Exact Index	Rounded Index	Frequency	Form	Exact Index	Rounded Index	Frequency	Form	Exact Index	Rounded Index	Frequency	Part
А	41.581	43	2081	Да	321.35	322	955	Ты	52.453	53	1518	740
ты	38.648	40	1518	А	240.26	241	2081	сказал	49.228	50	2337	1040
сказал	36.838	38	2337	Нет	223.49	224	462	А	39.379	40	2081	909
Да	35.529	37	955	Ну	205.33	206	867	Да	35.326	36	955	471
спросил	34.268	35	685	Это	197.08	198	833	Ну	32.673	34	867	429
Что	31.188	32	679	Что	133.48	134	679	Я	31.015	32	3149	1251
Ты	28.582	30	711	нет	128.25	129	650	Ты	26.33	27	711	350
Ну	28.278	29	867	Вот	107.6	109	543	вы	24.555	26	1061	478
Я	25.05	26	3149	это	81.157	82	2479	сказала	24.35	25	573	290
Вы	23.022	24	479	У	65.554	67	587	Вы	19.601	21	479	240
вы	18.861	20	1061	Как	64.48	65	583	Что	19.4	20	679	317
Не	14.991	16	1130	Спасибо	54.234	55	70	говору	17.439	18	264	147
Как	14.598	16	583	Ты	52.431	53	1518	Это	15.382	16	833	360
сказала	14.226	15	573	Где	52.381	53	144	тебе	14.185	15	523	241
тебя	14.22	15	627	Очень	47.963	49	115	спросил	13.58	15	685	299
Где	13.535	15	144	Ничего	45.304	46	148	я	13.306	14	4993	1740
тебе	12.093	13	523	Так	42.637	44	451	Гарри	13.305	14	1494	583
кто	10.977	12	184	Ты	42.469	43	711	здесь	12.895	14	468	216
Нет	10.557	12	462	И	40.957	42	1740	ж	12.444	13	308	153
Брет	10.508	12	430	Привет	35.541	37	30	вас	11.225	12	532	234
Рон	10.442	11	385	Какая	35.49	36	48	Нет	10.909	12	462	207
У	10.428	11	587	Ладно	34	35	88	Вот	10.434	11	543	235
Гарри	10.194	11	1494	же	33.943	35	1886	тебя	10.343	11	627	265
вас	10.141	11	532	О	33.436	34	143	Как	10.307	11	583	249
говору	9.1277	10	264	кто	30.665	32	184	не	10.299	11	1130	441
Это	8.9821	10	833	Ах	30.514	32	68	Мне	9.8036	11	328	153
Вот	8.9257	10	543	такое	29.816	31	255	Билл	9.7255	11	292	139
спросила	8.6557	10	129	да	29.482	30	615	Брет	9.2055	10	430	189
Ах	8.5261	10	68	Эй	28.91	30	58	Гермиона	9.1659	10	307	143
здесь	8.1682	9	468	мой	27.938	29	277	Рон	9.0951	10	385	172

Figure 59. RUSSIAN FORMS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

ENGLISH LEMMAS												
Left Context				Verbless Sentences				Right Context				
<i>(sentList = [-3,-2,-1]; totals = 806875; totalS = 132548)</i>				<i>(sentList = [0]; totals = 806875; totalS = 17107)</i>				<i>(sentList = [1,2,3]; totals = 806875; totalS = 131130)</i>				
Lemma	Exact Index	Rounded Index	Frequency	Exact Index	Rounded Exponent	Frequency	Lemma	Exact index	Rounded Index	Frequency	Part	
you	82.9727	84	11178	inf	inf	653	say	85.4582	86	5694	1509	
say	42.1691	43	5694	223.729	225	2062	you	76.3021	77	11178	2570	
sv	35.6295	37	4533	172.925	174	3449	I	40.1773	41	14738	3006	
what	33.5828	35	3449	160.742	162	450	sv	38.9615	40	4533	1080	
ask	30.8205	32	1254	90.3075	91	1201	what	26.0814	27	3449	803	
yes	22.7951	24	653	84.9291	86	94	yes	22.6001	24	653	209	
I	22.5062	24	14738	81.753	83	67	Harry	15.1704	16	1526	370	
do	15.779	17	6900	70.3386	71	1315	will	14.064	15	2615	576	
how	9.85498	11	1268	59.383	60	82	nt	14	15	11591	2192	
will	9.23634	10	2615	56.1542	57	127	your	13.3699	14	1999	454	
Jake	8.75061	10	123	50.9982	52	1351	do	12.3258	13	6900	1345	
name	8.44206	9	313	49.3894	50	80	well	11.3055	12	1351	317	
nt	8.24286	9	11591	46.8985	48	77	oh	10.8593	12	450	130	
count	7.68609	9	188	46.5292	48	857	Brett	8.64865	10	432	119	
Brett	7.51622	9	432	41.077	42	3554	voice	8.42328	9	617	157	
well	7.50012	9	1351	36.1005	37	2047	Montag	8.0954	9	455	122	
paragraph	6.8733	8	11	34.848	36	181	Noonan	7.65439	9	161	55	
Jenkins	6.8733	8	11	33.6594	35	260	let	7.39772	8	905	210	
she	6.72732	8	2973	31.4836	32	369	get	7.19195	8	3566	699	
your	6.65131	8	1999	30.2557	31	23	paragraph	6.9192	8	11	10	
Montag	6.49779	7	455	24.0785	25	22	Fred	6.8063	8	161	53	
get	6.23007	7	3566	21.1974	22	45	five	6.78857	8	326	90	
call	6.16804	7	629	20.8223	22	370	shout	6.77905	8	464	119	
uncle	6.14847	7	201	20.7714	22	117	us	6.64666	8	1155	254	
Harry	6.11184	7	1526	20.1777	21	82	Mildred	6.3429	7	104	38	
look	5.66495	7	2421	20.002	21	213	nigger	6.02998	7	17	12	
sir	5.65887	7	117	18.6874	20	36	Bill	5.97886	7	363	95	
zone	5.63433	7	155	17.98	19	134	fascist	5.64274	7	13	10	
here	5.53219	7	1983	17.4348	18	304	call	5.60192	7	629	147	
oh	5.29594	6	450	17.3407	18	12	yesterday	5.4241	6	61	25	
aha												

Figure 60. ENGLISH LEMMAS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN LEMMAS												
Left Context				Verbless Sentences				Right Context				
<i>(sentList = [-3,-2,-1]; totals = 581547; totalS = 173428)</i>				<i>(sentList = [0]; totals = 581547; totalS = 38022)</i>				<i>(sentList = [1,2,3]; totals = 581547; totalS = 174276)</i>				
Lemma	Exact Index	Rounded Index	Frequency	Exact Index	Rounded Index	Frequency	Exact Index	Rounded Index	Frequency	Total	Frequency	Partie
ты	99.427	100	3595	inf	inf	1114	inf	inf	553	106.21	107	1703
вы	55.245	56	2755	313.32	314	1584	313.32	314	628	71.635	73	1583
сказать	51.394	52	3585	221.28	222	3612	221.28	222	827	59.509	61	1232
спросить	42.249	43	923	205.54	207	1004	205.54	207	405	44.214	45	4401
ну	30.403	31	1004	128.87	130	3595	128.87	130	665	34.556	36	488
я	29.583	31	12295	120.82	122	1226	120.82	122	346	25.98	27	675
да	26.351	27	1584	120.38	121	4835	120.38	121	785	16.524	18	466
а	14.919	16	4835	88.973	90	692	88.973	90	221	16.469	17	1127
нет	13.227	14	1114	74.632	76	111	74.632	76	84	15.475	16	252
какой	12.95	14	692	52.688	54	147	52.688	54	79	спросить	16	923
вот	11.904	13	1226	46.077	47	251	46.077	47	95	говорить	15	1404
здесь	10.846	12	541	45.132	46	3012	45.132	46	415	а	14	4835
брет	10.402	11	431	44.458	45	2755	44.458	45	388	гарри	14	1495
у	10.391	11	3012	40.336	41	1204	40.336	41	216	ж	13	309
это	10.358	11	3612	35.336	36	293	35.336	36	90	ответить	12	451
гарри	10.297	11	1495	33.943	35	1886	33.943	35	274	рон	12	547
майкл	9.2607	10	276	33.765	35	455	33.765	35	112	твой	11	371
ах	8.6643	10	74	33.651	35	74	33.651	35	45	это	11	3612
джейк	8.505	10	125	33.36	34	375	33.36	34	100	вот	11	1226
ваш	8.177	9	429	32.798	34	1341	32.798	34	215	брет	10	431
рон	8.1376	9	547	30.703	32	42	30.703	32	33	воскликнут	10	95
кто	7.329	8	951	30.412	31	835	30.412	31	154	знать	9	1452
ж	7.0331	8	309	30.295	31	951	30.295	31	167	билл	9	366
там	6.703	8	841	30.196	31	2246	30.196	31	299	ваш	9	429
мы	6.6011	8	3043	29.218	30	718	29.218	30	138	какой	9	692
гермиона	6.5246	8	409	29.102	30	61	29.102	30	38	нунан	9	194
воскликнут	6.4332	7	95	29.063	30	541	29.063	30	116	отозваться	8	46
дженниз	5.7802	7	11	28.497	29	198	28.497	29	66	фред	8	163
крикнуть	5.6956	7	216	28.077	29	289	28.077	29	80	гермиона	8	409
граф	5.6602	7	68	27.6	29	153	27.6	29	57	повторить	8	79

Figure 61. RUSSIAN LEMMAS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

ENGLISH CATEGORIES													
Left Context				Verbless Sentences				Right Context					
<i>(sentList = [-3,-2,-1]; totals = 806875; totalS 132523)</i>													
Category	Exact Index	Rounded Index	Frequency Total	Frequency Partie	Exact Index	Rounded Index	Frequency Total	Frequency Partie	Category	Exact Index	Rounded Index	Frequency Total	Frequency Partie
PP	35.32319	36	77238	13922	inf	inf	1299	721	UH	51.0165	52	29892	5825
NP	34.89998	36	29892	5705	inf	inf	38466	2277	VBP	34.0997	35	4375	1025
WP	25.80936	27	5188	1146	227.655	229	29892	1559	VVZ	30.1892	31	8843	1848
UH	24.84174	26	1299	363	146.097	147	6801	539	PP	29.956	31	77238	13680
VBP	21.85293	23	4375	966	132.541	134	118187	3691	UH	29.7664	31	1299	376
VVZ	21.16019	22	8843	1793	123.055	124	5188	430	VVP	22.4111	23	8590	1741
VVP	10.41808	11	8590	1638	84.1871	85	47206	1643	WP	18.7607	20	5188	1089
VBZ	9.984547	11	2720	574	40.0008	41	28866	957	VV	14.1117	15	22681	4113
VH	9.377476	10	1373	314	21.6241	23	4905	215	VH	11.2459	12	1373	321
WRB	8.042757	9	4905	955	3.96412	5	1543	56	RP	11.0639	12	15762	2875
RP	8.00113	9	15762	2851	3.64614	5	1409	51	VD	8.2024	9	700	173
VD	7.41895	8	700	171	2.05738	3	80568	1801	RB	7.74052	9	38466	6643
VV	6.942256	8	22681	4013	1.313	2	222	9	VBZ	7.21494	8	2720	547
VHP	6.665403	8	2828	567	0.9118	0	749	21	VHP	6.70331	8	2828	562
RB	5.633103	7	38466	6645	0.87018	0	200	7	CD	6.20892	7	6801	1255
FW	4.885507	6	108	36	0.08227	0	82	1	WRB	2.37236	3	4905	866
ZZZ	2.921926	4	6697	1193	5.8E-10	0	1	0	FW	2.16212	3	108	28
VVD	2.545269	4	56400	9499	5.8E-10	0	1	0	VHZ	2.03208	3	535	108
NPS	2.146136	3	222	51	5.8E-10	0	1	0	VDZ	1.83137	3	47	14
VHZ	2.029915	3	535	109	3.6E-10	0	4	0	VDP	1.79377	3	95	24
EX	1.808582	3	4	3	-2E-10	0	13	0	VVD	1.26375	0	56400	9301
VDZ	1.792843	3	47	14	-0.43733	0	47	0	MD	1.09843	0	10308	1728
MD	1.737738	3	10308	1772	-0.62678	0	1135	20	VVG	0.90623	0	16403	2720
VDP	1.486517	2	95	23	-0.88398	0	95	0	VHN	0.78915	0	1	1
VB	0.695046	0	3167	538	-1.00496	0	108	0	WP\$	0.33195	0	82	14
RBS	0.620726	0	200	37	-2.06589	-3	222	0	EX	0.29411	0	4	1
CD	0.424422	0	6801	1127	-2.14035	-3	230	0	RBR	0.29239	0	1409	229
VVG	0.30504	0	16403	2695	-4.97959	-6	535	0	PDT	0.18794	0	13	2
RBR	0.281388	0	1409	231	-6.12663	-7	20367	337	VDN	7.1E-10	0	1	0
VHN	-7.2E-11	0	1	0	-6.51602	-8	700	0	VDD	7.1E-10	0	1	0
VD													

Figure 62. ENGLISH MORPHOSYNTACTIC CATEGORIES that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN CATEGORIES													
Left Context				Verbless Sentences				Right Context					
<i>(sentList = [-3,-2,-1]; totals = 581547; totalS = 173262)</i>				<i>(sentList = [0]; totals = 581547; totalS = 37972)</i>				<i>(sentList = [1,2,3]; totals = 581547; totalS = 174124)</i>					
Category	Exact Index	Rounded Index	Frequency Total	Frequency Total	Frequency Total	Frequency Total	Frequency Total	Category	Exact Index	Rounded Index	Frequency Total		
P-2-snn	65.6167	67	2240	1054	inf	inf	29514	4200	P-2-snn	77.4992	78	2240	1093
Q	60.1682	61	29514	10066	inf	inf	5424	1312	Q	70.8749	72	29514	10224
Npmsny	49.2164	50	13827	4920	inf	inf	1354	804	Npmsny	53.483	54	13827	4977
I	41.6766	43	1354	642	214	215	7096	1230	Vmip1s-a-e	48.4617	49	2894	1240
P-2-pnn	39.0548	40	1545	706	211.19	212	8291	1354	I	43.5928	45	1354	650
Nemsny	36.3256	37	7096	2609	170.54	172	11889	1617	P-2-pnn	42.5566	44	1545	720
Vmip2s-a-e	36.1622	37	1265	591	117.98	119	5920	892	Vmip2s-a-e	40.37	41	1265	606
P-1-snn	28.1284	29	8167	2896	95.451	96	2240	442	P-1-snn	39.8266	41	8167	3002
Vmip1s-a-e	25.5776	27	2894	1128	86.265	87	3225	536	Npfsny	29.4668	30	2418	987
Vmis-sma-p	25.5167	27	18638	6207	84.903	86	31863	2967	Nemsny	26.5011	28	7096	2544
Vmip2p-a-e	23.6245	25	816	381	83.921	85	2761	480	P-----r	26.4853	27	11889	4098
Vmis-sfa-p	19.6951	21	4427	1603	76.965	78	1585	329	Vmm-2s-a-p	26.1774	27	1111	502
Vmm-2s-a-p	19.1822	20	1111	474	56.93	58	2791	423	Vmip2p-a-e	23.199	24	816	381
Npfsny	18.8432	20	2418	927	54.096	55	13827	1381	Vmis-sma-p	22.1441	23	18638	6188
P-2-san	18.3818	19	570	271	47.548	49	1308	242	Vmip3s-a-e	21.6457	23	5840	2091
P-----r	18.1159	19	11889	3981	44.807	46	11962	1185	Vmif1s-a-p	21.2343	22	986	438
P--nsnn	17.8101	19	5424	1913	44.806	46	3463	457	Vmm-2s-a-e	19.9855	21	1479	611
Vmip3s-a-e	15.9548	17	5840	2030	39.396	40	767	162	P--nsnn	19.1159	20	5424	1933
P-2-sdn	15.17	16	581	265	37.23	38	1041	191	Vmm-2p-a-p	17.4829	18	731	330
Vmif1s-a-p	11.7945	13	986	397	31.217	32	5593	599	P-2-sdn	16.0942	17	581	269
P-2-pan	11.6574	13	555	243	29.148	30	441	103	P-2-san	13.4656	14	570	256
Vmm-2s-a-e	10.747	12	1479	560	29.073	30	1131	184	Vmis-sfa-p	12.8604	14	4427	1551
Ncfsny	10.5114	12	2044	747	27.803	29	152	57	P-2-pan	12.2246	13	555	246
Vmm-2p-a-p	9.96439	11	731	299	27.528	29	205	66	Vmip3p-a-e	12.1002	13	1965	735
Vmm-2p-a-e	9.35591	10	494	212	26.32	27	613	120	Vmip1p-a-e	11.6116	13	488	219
Vmif2s-a-p	8.16594	9	533	221	25.242	26	2044	265	Vmif2s-a-p	9.87333	11	533	229
Vmip	7.95171	9	170	86	24.846	26	1545	216	P-1-pan	9.00723	10	582	243
P-2-pdn	7.13181	8	535	217	22.99	24	581	110	Vmm-2p-a-e	8.90538	10	494	211
P-2-sin	6.83164	8	142	72	19.905	21	1726	219	V	7.6339	9	577	235
Vmip1s-m-e	5.53127	7	320	134	16.12	17	919	131	P-1-pnn	7.24247	8	1910	680

Figure 63. RUSSIAN MORPHOSYNTACTIC CATEGORIES that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

APPENDIX 7

Characteristic N-Grams

The present appendix provides the first 30 lines of the ‘characteristic repeated segments’ (or ‘characteristic n-grams’) calculation results. This calculation is similar to the characteristic elements in that it is also based on the hypergeometric model and includes all four variables discussed above (*Appendix 6*) but for n-grams, i.e. (1) the frequency of the n-gram in the entire corpus, (2) the frequency of the n-gram in the verbless sentence corpus part, (3) the total number of n-grams of that length in the corpus and (4) the total number of n-grams of that length in the verbless sentence corpus part. It determines the statistically key n-grams that characterize verbless sentences against a reference corpus (i.e. the entire corpus consisting of verbal and verbless sentences).

N-grams (or repeated segments) are adjacent words of various lengths. The results presented in this appendix account for all of the possible lengths of n-grams in the corpus. That is to say, the calculation was run for 2-grams, 3-grams, 4-grams, 5-grams, 6-grams and 7-grams, at which point no other n-grams resulted. These individual results were then combined into one list which organizes the n-grams of various lengths according to their specificity index.

The columns of the results are to be read in a similar way as those of the characteristic elements above. The ‘*Specificity Index*’ indicates the statistical specificity that the n-gram has for the verbless sentence as compared to the rest of the corpus. The ‘*Fq Total*’ indicates the frequency of the n-gram in the entire corpus. The ‘*Fq Part*’ indicates the frequency of the n-gram in the verbless sentence corpus part. In addition, the right most column indicates the length of the n-gram.

The results shown include the following levels:

- a) ENGLISH *FORM N-GRAMS*
- b) ENGLISH *LEMMA N-GRAMS*
- c) ENGLISH *MORPHOSYNTACTIC CATEGORY N-GRAMS*
- d) RUSSIAN *FORM N-GRAMS*
- e) RUSSIAN *LEMMA N-GRAMS*
- f) RUSSIAN *MORPHOSYNTACTIC CATEGORY N-GRAMS*

The morphosyntactic category specifications are found in *Appendix 3*.

ENGLISH FORM N-GRAMS																	
Left Context (sentList = [-3,-2,-1])				Verbless Sentences (sentList = [0])				Right Context (sentList = [1,2,3])									
Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth
('I', 'said')	17.37	18	563	165	2	('All', 'right')	117.9	119	162	79	2	('said', 'Harry')	28.56	30	265	115	2
('he', 'asked')	15.02	16	120	55	2	('What', 'about')	114.4	115	66	59	2	('I', 'said')	27.06	28	563	186	2
('What', 'sv')	14.57	16	259	90	2	('Why', 'not')	60.07	61	35	31	2	('I', 'I')	18.99	20	724	201	2
('Do', 'you')	12.94	14	194	71	2	('what', 'about')	54.1	55	44	31	2	('I', 'm')	17.2	18	1068	264	2
('do', 'n', 't')	12.63	14	1188	273	3	('What', 'a')	52.1	53	60	33	2	('n', 't')	15.54	17	5377	1008	2
('do', 'n')	12.63	14	1188	273	2	('What', 'for')	50.15	51	24	24	2	('It', 'sv')	14.86	16	681	180	2
('n', 't', 'you')	12.35	13	356	107	3	('Oh', 'yes')	44.93	46	38	26	2	('That', 'sv')	14.58	16	476	137	2
('did', 'you')	12.25	13	167	63	2	('hell', 'with')	41.28	42	41	25	2	('t', 'you')	11.09	12	407	113	2
('Have', 'you')	12.2	13	69	36	2	('How', 'about')	38.69	40	31	22	2	('do', 'n', 't')	11.06	12	1188	262	3
('t', 'you')	11.26	12	407	115	2	('Just', 'a')	37.83	39	28	21	2	('do', 'n')	11.06	12	1188	262	2
('n', 't')	11.15	12	5377	988	2	('Good', 'night')	30.14	31	16	15	2	('Let', 'us')	10.31	11	190	64	2
('That', 'sv')	10.17	11	476	126	2	('Hey', 'you')	28.44	29	18	15	2	('Do', 'n', 't')	10.21	11	464	122	3
('do', 'you')	10.12	11	423	115	2	('One', 'two')	28.08	29	15	14	2	('Do', 'n')	10.21	11	464	122	2
('your', 'name')	10.11	11	24	18	2	('To', 'the')	26.21	27	34	17	2	('What', 'do')	10.07	11	175	60	2
('you', 'know')	9.964	11	274	83	2	('No', 'not')	25.07	26	12	12	2	('sv', 'a')	9.767	11	436	115	2
('are', 'you')	9.604	11	306	89	2	('My', 'God')	24.21	25	21	14	2	('What', 'sv')	9.683	11	259	78	2
('It', 'sv')	9.48	10	681	164	2	('Of', 'course')	23.87	25	121	23	2	('are', 'you')	9.664	11	306	88	2
('I', 'm')	9.358	10	1068	236	2	('Oh', 'no')	23.81	25	28	15	2	('do', 'you')	9.641	11	423	112	2
('sv', 'the')	9.344	10	314	90	2	('To', 'hell', 'with')	23.8	25	17	13	3	('you', 'going')	9.368	10	59	29	2
('I', 'asked')	9.039	10	135	49	2	('To', 'hell')	23.8	25	17	13	2	('What', 'do', 'you')	9.266	10	148	52	3
('Is', 'that')	8.565	10	55	27	2	('Not', 'a')	23.24	24	18	13	2	('n', 't', 'you')	9.168	10	356	97	3
('He', 'sv')	8.536	10	349	95	2	('One', 'two', 'three')	21.91	23	12	11	3	('have', 'n', 't')	9.156	10	179	59	3
('I', 'I')	8.435	9	724	168	2	('Oh', 'yeah')	21.82	23	16	12	2	('have', 'n')	9.156	10	179	59	2
('Want', 'to')	8.316	9	18	14	2	('two', 'three')	21.3	22	17	12	2	('I', 'have', 'n', 't')	8.934	10	68	31	4
('he', 'said')	8.058	9	569	137	2	('No', 'no')	21.14	22	30	14	2	('I', 'have', 'n')	8.934	10	68	31	3
('sv', 'your', 'name')	8.052	9	12	11	3	('What', 'about', 'the')	20.89	22	10	10	3	('I', 'have')	8.714	10	304	85	2
('you', 'go')	7.77	9	77	32	2	('Not', 'bad')	20.89	22	10	10	2	('All', 'right')	8.644	10	162	54	2
('have', 'to')	7.443	8	341	90	2	('Good', 'luck')	20.89	22	10	10	2	('said', 'Brett')	8.641	10	25	17	2
('You', 'know')	7.404	8	158	51	2	('Not', 'so')	20.43	21	14	11	2	('What', 'about')	8.631	10	66	30	2
('What', 'sv', 'your')	7.303	8	13	11	3	('son', 'of')	19.94	21	56	16	2	('he', 'said')	8.59	10	569	137	2

Figure 64. ENGLISH FORM N-GRAMS

that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

ENGLISH LEMMA NGRAMS													
Left Context (sentList = [-3,-2,-1])				Verbless Sentences (sentList = [0])				Right Context (sentList = [1,2,3])					
Lemma	Exact Index	Rounded Index	Fq Total	Exact Index	Rounded Index	Fq Total	Lemma	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	
('do', 'you')	35.3	36	890 282 2		166.5	168	110 90 2	('I', 'say')	29.08	30	684 218 2		
('I', 'say')	21.1	22	684 201 2	('what', 'about')	110.7	112	385 97 2	('say', 'Harry')	28.56	30	265 115 2		
('what', 'sv')	19.02	20	380 128 2	('all', 'right')	66.39	67	49 37 2	('be', 'you')	18.2	19	521 156 2		
('be', 'you')	19	20	521 160 2	('why', 'not')	52.5	54	116 40 2	('that', 'sv')	17.9	19	868 227 2		
('have', 'you')	17.59	19	189 78 2	('what', 'a')	49.86	51	38 28 2	('I', 'will')	17.53	19	809 214 2		
('you', 'be')	16.91	18	1301 312 2	('how', 'about')	48.39	49	32 26 2	('do', 'you')	17.16	18	890 229 2		
('you', 'know')	16.28	17	443 136 2	('what', 'for')	44.01	45	40 26 2	('you', 'be')	16.54	18	1301 306 2		
('he', 'ask')	13.32	14	148 60 2	('oh', 'yes')	41.28	42	41 25 2	('it', 'sv')	16.44	17	1266 299 2		
('nt', 'nt', 'you')	12.57	14	358 108 3	('hell', 'with')	35.6	37	303 41 2	('I', 'be')	16.09	17	2295 485 2		
('you', 'go')	12.12	13	160 61 2	('of', 'course')	35.34	36	20 18 3	('nt', 'nt')	15.56	17	5382 1009 2		
('nt', 'you')	11.46	12	409 116 2	('what', 'about', 'the')	34.82	36	29 20 2	('what', 'do')	15.37	16	343 110 2		
('nt', 'nt')	11.17	12	5382 989 2	('my', 'God')	30.52	32	28 18 2	('what', 'do', 'you')	14.34	15	211 77 3		
('all', 'right')	11.13	12	385 110 2	('good', 'night')	30.45	31	19 16 2	('let', 'us')	12.13	13	296 92 2		
('he', 'say')	10.2	11	698 170 2	('hey', 'you')	29.05	30	26 17 3	('do', 'nt', 'nt')	11.46	12	2985 577 3		
('your', 'name')	9.62	11	25 18 2	('to', 'hell', 'with')	27.65	29	113 25 2	('do', 'nt')	11.46	12	2985 577 2		
('what', 'sv', 'your')	9.592	11	16 14 3	('just', 'a')	27.6	29	86 23 2	('nt', 'you')	10.94	12	409 113 2		
('sv', 'the')	9.344	10	314 90 2	('my', 'dear')	25.78	27	18 14 2	('all', 'right')	10.93	12	385 108 2		
('what', 'do')	9.303	10	343 96 2	('no', 'not')	24.86	26	25 15 2	('sv', 'a')	9.767	11	436 115 2		
('what', 'do', 'you')	9.14	10	211 67 3	('one', 'two')	23.81	25	28 15 2	('you', 'go')	9.348	10	160 55 2		
('article', '@card@')	8.937	10	24 17 2	('oh', 'no')	23.8	25	17 13 2	('nt', 'nt', 'you')	9.028	10	358 97 3		
('do', 'nt', 'nt')	8.917	10	2985 568 3	('oh', 'yeah')	23.63	25	36 16 2	('what', 'sv')	8.875	10	380 101 2		
('do', 'nt')	8.917	10	2985 568 2	('no', 'no')	23.07	24	70 19 2	('have', 'you')	8.588	10	189 60 2		
('what', 'be')	8.812	10	379 102 2	('son', 'of')	22.13	23	53 17 2	('you', 'get')	8.49	9	190 60 2		
('what', 'sv', 'the')	8.803	10	82 35 3	('to', 'hell')	21.91	23	12 11 2	('you', 'have')	8.374	9	685 158 2		
('that', 'sv')	8.685	10	868 196 2	('good', 'luck')	21.3	22	17 12 2	('you', 'Ron')	8.362	9	143 49 2		
('where', 'be')	8.619	10	122 45 2	('two', 'three')	21.3	22	17 12 3	('say', 'Fred')	8.219	9	61 28 2		
('you', 'know', 'what')	8.423	9	49 25 3	('one', 'two', 'three')	21.3	22	13 11 3	('say', 'Brett')	8.165	9	29 18 2		
('I', 'will')	8.227	9	809 183 2	('what', 'about', 'you')	21.1	22	14 11 3	('he', 'say')	8.031	9	698 159 2		
('do', 'you', 'have')	8.113	9	26 17 3	('just', 'a', 'minute')	20.43	21	44 15 3	('you', 'know')	7.811	9	443 110 2		
('sv', 'your', 'name')	8.052	9	12 11 3	('son', 'of', 'a')	20.08	21	44 15 3	('what', 'about')	7.783	9	110 40 2		
				('about', 'the')	19.97	21	276 27 2						

Figure 65. ENGLISH LEMMA N-GRAMS

that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

ENGLISH CATEGORY NGRAMS																	
Left Context (sentList = [-3,-2,-1])				Verbless Sentences (sentList = [0])				Right Context (sentList = [1,2,3])									
Category	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Category	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Category	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth
('PP', 'VVP')	33.54	35	2894	683	2	('JJ', 'NN')	231.5	233	18510	693	2	('VVD', 'NP')	61.14	62	2519	693	2
('PP', 'VV')	29.54	31	5672	1170	2	('WP', 'IN')	179.4	180	273	124	2	('PP', 'VVP')	46.93	48	2894	723	2
('PP', 'VVD')	22.06	23	17402	3077	2	('DT', 'NN')	174.2	175	44702	1009	2	('PP', 'VBP')	36.81	38	3075	720	2
('VVD', 'NP')	21.03	22	2519	558	2	('RB', 'RB')	162.6	164	2015	217	2	('NP', 'VVD')	27.72	29	6729	1322	2
('VV', 'PP')	20.75	22	5841	1143	2	('NP', 'NP')	151.8	153	3100	245	2	('PP', 'VWZ')	25.8	27	2973	654	2
('PP', 'VBP')	18.93	20	3075	648	2	('UH', 'NP')	133	134	125	80	2	('PP', 'VV')	25.68	27	5672	1128	2
('NP', 'VVD')	18.44	19	6729	1277	2	('UH', 'RB')	131.2	132	147	83	2	('VV', 'PP')	22.91	24	5841	1139	2
('WP', 'VWZ')	18.22	19	693	195	2	('CD', 'CD')	101.2	102	331	87	2	('VWZ', 'DT')	19.83	21	1514	359	2
('VBP', 'PP')	17.67	19	564	166	2	('NN', 'NN')	93.25	94	6147	253	2	('NP', 'RB')	17.76	19	1203	292	2
('PP', 'VWZ')	16.99	18	2973	619	2	('WP', 'NN')	91.46	92	233	73	2	('VVD', 'NP', 'RB')	15.75	17	411	126	3
('PP', 'VV', 'PP')	16.83	18	1230	298	3	('NN', 'NP')	80.94	82	3271	174	2	('RP', 'RP')	15.56	17	5382	1009	2
('VWZ', 'DT')	16.54	18	1514	351	2	('UH', 'NN')	79.17	80	83	49	2	('VBP', 'PP')	15.51	17	564	158	2
('WP', 'VWZ', 'DT')	15.19	16	174	70	3	('IN', 'NN')	75.89	77	9260	284	2	('PP', 'VHP')	15.3	16	840	213	2
('RP', 'RP', 'PP')	14.72	16	632	172	3	('DT', 'JJ')	74.18	75	15978	386	2	('RP', 'PP')	14.34	15	780	198	2
('RP', 'PP')	14.7	16	780	202	2	('CD', 'NNS')	67.98	69	2115	130	2	('RP', 'RP', 'PP')	14.17	15	632	168	3
('VVP', 'RP', 'RP')	13.36	14	1214	281	3	('UH', 'UH')	67.06	68	57	39	2	('VVD', 'NN', 'NP')	13.04	14	336	103	3
('VVP', 'RP')	13.36	14	1214	281	2	('RB', 'DT')	64.88	66	1620	113	2	('VWZ', 'DT', 'NN')	12.74	14	845	205	3
('VD', 'PP')	12.94	14	198	72	2	('RB', 'NN')	64.81	66	2021	124	2	('VVP', 'RP', 'RP')	12.01	13	1214	271	3
('MD', 'PP')	12.21	13	717	181	2	('RB', 'JJ')	59.63	61	1974	117	2	('VVP', 'RP')	12.01	13	1214	271	2
('NN', 'NP')	12.2	13	3271	641	2	('WP', 'DT')	57.16	58	275	55	2	('WP', 'VWZ')	11.99	13	693	173	2
('WP', 'VWZ', 'DT', 'NN')	11.82	13	116	49	4	('CD', 'CD', 'CD')	55.09	56	64	35	3	('PP', 'VVD')	11.09	12	17402	2877	2
('VH', 'PP')	11.69	13	88	41	2	('WRB', 'RB')	53.09	54	169	45	2	('WP', 'NP', 'PP')	10.91	12	189	65	3
('VWZ', 'DT', 'NN')	11.42	12	845	203	3	('WRB', 'IN')	49.33	50	94	36	2	('VVD', 'NP', 'VWG')	10.77	12	225	73	3
('RP', 'RP')	11.17	12	5382	989	2	('NN', 'IN')	46.94	48	27005	464	2	('PP', 'MD')	10.75	12	6410	1135	2
('NP', 'NN')	10.16	11	1537	325	2	('WP', 'IN', 'PP')	46.61	48	80	33	3	('CD', 'NN', 'CD')	10.69	12	28	20	3
('PP', 'MD')	9.796	11	6410	1145	2	('NP', 'NN')	46.57	48	1537	91	2	('PP', 'VH')	10.65	12	1182	259	2
('VHP', 'PP')	9.694	11	223	71	2	('WP', 'DT', 'NN')	46.14	47	173	41	3	('VWZ', 'RB')	10.3	11	890	204	2
('WRB', 'VBP')	9.196	10	206	66	2	('DT', 'JJ', 'NN')	46.11	47	10463	247	3	('VBP', 'RB')	9.88	11	712	169	2
('PP', 'VVP', 'RP', 'RP')	8.887	10	831	190	4	('RB', 'IN')	43.19	44	4132	141	2	('PP', 'VVP', 'RP', 'RP')	9.804	11	831	191	4
('PP', 'VVP', 'RP')	8.887	10	831	190	3	('IN', 'DT')	42.26	43	32952	519	2	('PP', 'VVP', 'RP')	9.804	11	831	191	3

Figure 66. ENGLISH MORPHOSYNTACTIC CATEGORY N-GRAMS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN FORM NGRAMS																	
Left Context (sentList = [-3,-2,-1])					Verbless Sentences (sentList = [0])					Right Context (sentList = [1,2,3])							
Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth	Form	Exact Index	Rounded Index	Fq Total	Fq Part	Len gth
('сказал', 'я')	17.25	18	274	140	2	('А', 'что')	43.93	45	104	49	2	('сказал', 'я')	39.87	41	274	179	2
('ну', 'как')	12.34	13	52	39	2	('что', 'такое')	37.81	39	29	27	2	('сказала', 'Брет')	15.58	17	134	81	2
('спросил', 'он')	11.53	13	73	48	2	('ну', 'что')	34.7	36	79	38	2	('сказал', 'Билл')	15.37	16	149	87	2
('спросил', 'я')	11.34	12	67	45	2	('ну', 'как')	34.04	35	52	32	2	('сказал', 'Гарри')	14.02	15	85	57	2
('А', 'что')	10.65	12	104	60	2	('у', 'меня')	33.56	35	121	44	2	('что', 'такое')	10.47	11	29	25	2
('что', 'такое')	10.53	12	29	25	2	('что', 'это')	32.91	34	55	32	2	('сказал', 'Майкл')	10.05	11	94	55	2
('сказал', 'он')	9.52	11	309	133	2	('Спокойной', 'ночи')	29.92	31	20	20	2	('сказал', 'он')	9.039	10	309	132	2
('что', 'ли')	9.043	10	63	40	2	('А', 'ты')	28.19	29	83	34	2	('А', 'что')	8.885	10	104	57	2
('сказал', 'Майкл')	8.993	10	94	53	2	('ну', 'и')	26.81	28	71	31	2	('не', 'могу')	8.295	9	112	59	2
('сказала', 'Брет')	8.194	9	134	67	2	('у', 'вас')	25.19	26	85	32	2	('А', 'ты')	8.059	9	83	47	2
('А', 'вы')	7.92	9	67	40	2	('у', 'тебя')	24.42	25	38	23	2	('сказал', 'Рон')	8.049	9	57	36	2
('у', 'тебя')	7.874	9	38	27	2	('как', 'же')	24.27	25	53	26	2	('у', 'нас')	7.513	9	176	81	2
('с', 'тобой')	7.728	9	100	53	2	('А', 'это')	23.52	25	36	22	2	('сказала', 'она')	7.3	8	107	55	2
('А', 'ты')	7.052	8	83	45	2	('что', 'с')	23.11	24	22	18	2	('с', 'тобой')	7.155	8	100	52	2
('мистер', 'Уизли')	6.736	8	67	38	2	('у', 'вас')	22.23	23	35	21	2	('говору', 'я')	6.893	8	32	23	2
('я', 'думал')	6.511	8	21	17	2	('чем', 'дело')	22.18	23	27	19	2	('ответил', 'Гарри')	6.861	8	28	21	2
('у', 'тебя')	6.361	7	97	49	2	('Вот', 'это')	21.88	23	24	18	2	('сказал', 'Рэдрик')	6.704	8	62	36	2
('вы', 'не')	6.125	7	32	22	2	('ну', 'ладно')	21.72	23	32	20	2	('я', 'и')	6.271	7	38	25	2
('да', 'нет')	6.002	7	26	19	2	('А', 'где')	21.72	23	32	20	2	('вы', 'не')	6.234	7	74	40	2
('что', 'с')	5.992	7	22	17	2	('с', 'тобой')	21.49	22	100	31	2	('А', 'это')	6.202	7	36	24	2
('сказал', 'Билл')	5.984	7	149	67	2	('ну', 'вот')	21.33	22	33	20	2	('что', 'вы')	6.091	7	117	56	2
('произнес', 'он')	5.757	7	10	10	2	('но', 'это')	21.18	22	38	21	2	('гражданин', 'начальник')	6.082	7	32	22	2
('сказал', 'Рэдрик')	5.641	7	62	34	2	('в', 'чем')	20.99	22	19	16	2	('да', 'ты')	6.082	7	32	22	2
('что', 'ты')	5.313	6	56	31	2	('до', 'свидания')	20.94	22	14	14	2	('я', 'не')	6.002	7	197	84	2
('что', 'это')	4.989	6	55	30	2	('у', 'нас')	20.94	22	61	25	2	('что', 'ли')	5.908	7	63	35	2
('сказала', 'она')	4.825	6	107	49	2	('да', 'нет')	20.84	22	26	18	2	('же', 'мне')	5.89	7	13	12	2
('я', 'тебе')	4.673	6	32	20	2	('ну', 'конечно')	19.78	21	15	14	2	('у', 'тебя')	5.839	7	97	48	2
('в', 'Сидкап')	4.606	6	8	8	2	('у', 'тебя')	19.66	21	97	29	2	('Мне', 'нужно')	5.734	7	10	10	2
('на', 'Гарри')	4.511	6	60	31	2	('в', 'чем', 'дело')	19.44	20	13	13	3	('Гарри', 'и')	5.622	7	85	43	2
('сказал', 'Фред')	4.416	5	26	17	2	('Ещё', 'бы')	18.64	20	23	16	2	('что', 'ты')	5.284	6	134	60	2

Figure 67. RUSSIAN FORM N-GRAMS

that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN LEMMA NGRAMS												
Left Context (sentList = [-3,-2,-1])				Verbless Sentences (sentList = [0])				Right Context (sentList = [1,2,3])				
Lemma	Exact Index	Rounded Index	Fq Total	Lemma	Exact Index	Rounded Index	Fq Total	Lemma	Exact Index	Rounded Index	Fq Total	
	Part	Len	gth		Part	Len	gth		Part	Len	gth	
('сказать', 'я')	15.05	16	339	('у', 'вы')	45.58	47	120	('сказать', 'я')	33.25	34	339	
('у', 'ты')	12.52	14	135	('а', 'что')	44.7	46	136	('сказать', 'брат')	15.58	17	134	
('спросить', 'он')	12.14	13	89	('у', 'ты')	41.01	42	135	('сказать', 'билл')	15.37	16	149	
('ну', 'как')	11.23	12	63	('что', 'такой')	39.18	40	84	('сказать', 'гарри')	13.68	15	86	
('а', 'ты')	10.36	11	148	('ты', 'что')	37.06	38	116	('а', 'ты')	12.25	13	148	
('сказать', 'он')	10.27	11	387	('ну', 'что')	36.8	38	93	('а', 'что')	9.596	11	136	
('спросить', 'я')	10.02	11	85	('у', 'я')	34.26	35	410	('сказать', 'майкл')	9.57	11	96	
('сказать', 'майкл')	9.663	11	96	('а', 'ты')	32.68	34	148	('у', 'мы')	9.439	10	237	
('а', 'что')	9.227	10	136	('что', 'ж')	31.94	33	153	('что', 'ты')	9.389	10	243	
('что', 'ли')	9.043	10	63	('ну', 'как')	31.84	33	63	('что', 'вы')	9.323	10	199	
('с', 'ты')	8.278	9	110	('как', 'же')	31.11	32	82	('ты', 'не')	9.044	10	226	
('ты', 'быть')	8.199	9	59	('спокойный', 'ночь')	27.36	28	25	('сказать', 'он')	8.912	10	387	
('сказать', 'брат')	8.194	9	134	('ещё', 'бы')	26.25	27	30	('я', 'ты')	8.867	10	127	
('что', 'ты')	8.131	9	243	('не', 'надо')	26.2	27	100	('что', 'такой')	8.393	9	84	
('а', 'вы')	7.912	9	94	('вот', 'и')	25.85	27	95	('вы', 'не')	8.285	9	176	
('это', 'что')	7.816	9	32	('что', 'же')	25.66	27	126	('у', 'ты')	7.928	9	135	
('ты', 'не')	7.759	9	226	('это', 'что')	25.23	26	32	('сказать', 'рон')	7.845	9	60	
('вы', 'не')	7.622	9	176	('ну', 'и')	24.82	26	87	('как', 'же')	7.73	9	82	
('да', 'нет')	7.362	8	29	('гражданин', 'начальник')	24.25	25	43	('с', 'ты')	7.702	9	110	
('у', 'вы')	6.975	8	120	('до', 'свидание')	24.19	25	18	('что', 'ж')	7.243	8	153	
('что', 'с', 'ты')	6.511	8	21	('у', 'мы')	23.73	25	237	('это', 'мой')	6.963	8	22	
('на', 'война')	6.446	7	25	('а', 'где')	23.52	25	36	('ответить', 'гарри')	6.861	8	28	
('это', 'не')	6.152	7	226	('с', 'ты')	23.44	24	110	('сказать', 'она')	6.729	8	121	
('как', 'вы')	6.124	7	80	('что', 'это')	22.97	24	291	('сказать', 'рэдрик')	6.704	8	62	
('ты', 'хотеть')	6	7	39	('что', 'за')	22.53	24	66	('да', 'ты')	6.609	8	37	
('сказать', 'она')	5.985	7	121	('чем', 'дело')	22.18	23	27	('гражданин', 'начальник')	6.119	7	43	
('сказать', 'билл')	5.984	7	149	('это', 'не')	21.99	23	226	('ты', 'же')	5.954	7	39	
('нет', 'не')	5.883	7	44	('ну', 'ладно')	21.72	23	32	('что', 'ли')	5.908	7	63	
('что', 'такой')	5.865	7	84	('ну', 'вот')	21.33	22	33	('мистер', 'уизли')	5.898	7	102	
('что', 'ж')	5.839	7	153	('да', 'нет')	21.25	22	29	('не', 'знать')	5.841	7	450	
			68				19				166	

Figure 68. RUSSIAN LEMMA N-GRAMS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RUSSIAN CATEGORY NGRAMS															
Left Context (sentList = [-3,-2,-1])				Verbless Sentences (sentList = [0])				Right Context (sentList = [1,2,3])							
Category	Exact Index	Rounded Index	Fq Total	Fq Part	Fq Len	Exact Index	Rounded Index	Fq Total	Fq Part	Fq Len	Exact Index	Rounded Index	Fq Total	Fq Part	Fq Len
('Vmis-sma-p', 'Npmsny')	74.01	75	2097	947	2	210.9	212	1730	398	2	122.3	123	2097	1069	2
('Vmis-sma-p', 'P-1-snn')	27.86	29	510	253	2	157.6	159	1159	283	2	43.65	45	510	286	2
('Vmis-sma-p', 'P-3msnn')	22.36	23	757	327	2	154.1	155	1677	331	2	31.91	33	454	241	2
('Vmis-sfa-p', 'Npfsny')	17.58	19	454	207	2	113.7	115	1878	299	2	23.6	25	757	332	2
('Q', 'P-2-snn')	16.71	18	317	155	2	107.7	109	1394	251	2	21.47	22	317	166	2
('Q', 'Q')	15.21	16	1677	595	2	98.19	99	326	126	2	18.46	19	313	158	3
('Q', 'C')	14.72	16	1878	654	2	97.27	98	741	177	2	18.18	19	1878	676	2
('R', 'Vmis-sma-p', 'Npmsny')	14.44	15	313	148	3	91.27	92	1142	209	2	16.9	18	780	318	2
('Sp-g', 'P-2-san')	12.96	14	204	104	2	89.7	91	2389	303	2	16.72	18	1142	436	2
('P-2-snn', 'Vmip2s-a-e')	12.67	14	190	98	2	88.45	89	230	103	2	15.76	17	219	116	2
('Q', 'Vmip1s-a-e')	12.54	14	780	300	2	87.12	88	534	144	2	15.49	16	1159	436	2
('Vmis-sfa-p', 'Ncfsny')	12.42	13	194	99	2	72.89	74	985	173	2	15.06	16	1937	677	2
('P-2-snn', 'P-----r')	12.39	13	109	65	2	69.67	71	170	79	2	14.76	16	370	170	2
('Vmis-sma-p', 'Ncmsny')	11.87	13	410	175	2	69.3	70	807	153	2	14.55	16	281	137	2
('P--nsnn', 'P--nsna')	11.6	13	75	49	2	67.78	69	375	107	2	14.45	15	943	362	2
('Q', 'P-2-pnn')	11.17	12	223	107	2	60.16	61	101	59	2	14.38	15	1677	593	2
('Vmis-sfa-p', 'P-3fsnn')	10.6	12	249	115	2	58.41	59	171	71	2	13.81	15	327	152	2
('Q', 'Ncmsny')	10.22	11	326	141	2	55	56	1654	198	2	13.53	15	177	95	2
('P-2-pnn', 'R')	9.852	11	125	67	2	53.8	55	60	45	2	13.47	14	190	100	2
('R', 'Vmis-sma-p')	9.798	11	1937	640	2	53.65	55	253	79	2	13.33	14	624	253	2
('Q', 'P-----r')	9.69	11	1159	405	2	53.52	55	1859	209	2	13.3	14	356	161	2
('C', 'P-2-snn')	9.29	10	370	153	2	52.91	54	130	60	2	13.29	14	410	180	2
('P-1-snn', 'Q')	9.233	10	943	336	2	48.91	50	305	81	2	13.24	14	298	140	2
('Q', 'P-1-snn')	9.219	10	578	221	2	47.27	48	58	41	2	12.83	14	249	121	2
('Npmsny', 'Npmsny')	9.203	10	1142	397	2	46.83	48	75	45	2	12.28	13	1730	597	2
('Sp-i', 'P-2-sin')	8.564	10	124	64	2	46.49	47	317	80	2	11.64	13	227	110	2
('Vmis-sma-p', 'Npmsny', 'Vmip2p-a-e')	8.473	9	109	58	3	45.56	47	74	44	2	10.59	12	136	73	2
('P-1-snn', 'Vmip1s-a-e')	8.378	9	624	232	2	45.43	46	58	40	2	10.38	11	326	142	2
('Q', 'R')	8.359	9	1730	568	2	44.48	45	64	41	2	9.703	11	151	77	2
('Mc-n', 'Mc-n')	8.328	9	171	81	2	42.45	43	927	128	2	9.583	11	307	133	2

Figure 69. RUSSIAN MORPHOSYNTACTIC CATEGORY N-GRAMS that statistically characterize the verbless sentence (middle column) and its left and right context (left and right columns)

RÉSUMÉ LONG EN FRANÇAIS

MOTIVATION DE L'ÉTUDE

Cette thèse porte sur les structures dans lesquelles le marqueur typique de la phrase – le prédicat verbal – est absent. Les caractéristiques sémantiques et pragmatiques des phrases averbales sont étudiées en anglais et en russe en utilisant une approche interdisciplinaire qui intègre la linguistique contrastive, les méthodes de corpus et l'analyse énonciative.

A quel moment les mots deviennent-ils des phrases ? La question de savoir ce qui constitue précisément une phrase en langue naturelle est au cœur de toute théorie linguistique. Une phrase exige quelque chose de plus qu'une succession aléatoire de mots. Cela a été noté pour la première fois par Platon il y a plus que deux mille ans dans le *Sophiste*. La déclaration que « *les verbes sont mêlés avec les noms leur première combinaison est l'énoncé* » (Platon, 360 B.C./1921 : *Sophiste* 262b–d) est généralement identifiée comme le catalyseur de la discussion sur l'activité fondamentale qui a lieu dans une phrase. Depuis Platon, cette activité centrale, appelée la prédication, est associée à la présence du verbe. L'exigence traditionnelle du verbe rend contradictoire la notion de la phrase averbale et, par conséquent, relègue souvent les structures sans verbe en marge de la discussion linguistique, les considérant comme incomplètes, atypiques et non pertinentes pour l'analyse linguistique.

Pourtant, le phénomène des phrases averbales a été constaté dans de nombreuses langues naturelles, tout particulièrement par Émile Benveniste dans *Problèmes de Linguistique Générale* (1966). Parmi la famille des langues indo-européennes, la langue russe est généralement connue pour permettre l'utilisation la plus libérale de phrases averbales (McShane, 2000 ; Kopotev, 2007b). Elle utilise couramment des phrases telles que (i) ci-dessous, littéralement « *I Tonya* », où le verbe copule « *be* » serait typiquement attendu en anglais, ainsi que des phrases sans verbe lexical complet, telles que (ii), littéralement « *I to store* ». Les caractéristiques typologiques du russe comprennent un système de cas

morphologique très développé qui contribue à l'extraordinaire capacité de la langue pour l'ellipse du sujet et l'ellipse verbale ; la flexibilité de l'ordre des mots et de l'intonation ; l'inexistence d'articles ; la suppression typique du verbe copule « БЫТЬ » (*byt'* ; lit. « être ») au présent, comme l'illustre la construction « zéro copule » (Stassen, 2013) en (i), et l'autorisation de structures averbales telles que (ii) sans exiger un verbe antécédent dans le contexte.

- (i) Я Тоня.
ja *Tonja*
 PRO NN.NOM
I *Tonya*
je *Tonya*

'*I am Tonya.*'
 « *Je suis Tonya.* »

- (ii) Я В магазин.
ja *v* *magazin*
 PRO PREP NN.ACC
I *to* *store*
je *au* *magasin*

'*I am going to the store.*'
 « *Je vais au magasin.* »

Bien que la langue anglaise soit connue pour sa dépendance à l'égard du groupe verbal fini, on trouve également des phrases averbales dans divers actes de langage, comme l'illustrent la déclaration, la directive, l'exclamation et la question dans (iii).

- (iii) a. No sign of Dmitry Fyodorovich yet.
 NEG NN PREP NN NN ADV
pas signe de Dmitry Fyodorovich encore
 « *Aucun signe de Dmitry Fyodorovich pour le moment.* »

- b. Just a moment.
 ADV DET NN
seulement un instant
 « *Un instant.* »

- c. How terrible!
 ADV ADJ
quelle horreur
 « *Quelle horreur !* »

- d. What about my parental blessing?
 PRO PREP PRO ADJ NN
quoi concernant ma bénédiction parentale
 « *Et ma bénédiction parentale ?* »

Les caractéristiques de l'anglais contrastent fortement avec celles du russe et comprennent un système de cas limité qui est aidé par le verbe pour réduire l'ambiguïté ; des restrictions de registre sur l'ellipse verbale syntaxique, en particulier le gapping (McShane, 2000) ; la nécessité typique d'une copule verbale explicite, la présence d'articles et un ordre strict des mots. Ces profondes différences typologiques entre les deux langues rendent leur contraste particulièrement pertinent pour l'étude des phrases averbales.

Bien que les structures averbales aient fasciné de nombreux linguistes, la difficulté de les retrouver automatiquement a fait que la plupart des analyses se sont appuyées sur des données fragmentaires et des exemples inventés. Le défi pour toute recherche automatisée du phénomène est que l'élément fondateur de la requête est une structure grammaticale indéterminée et centrée sur l'absence. Dans le cadre du traitement automatique des langues, les structures averbales constituent des chaînes ouvertes composées d'éléments grammaticaux qui ne peuvent être spécifiés a priori. Autrement dit, ils représentent une recherche quasi absurde de quelque chose d'indéterminé qui contient quelque chose qui n'existe pas. La présence d'un verbe par le biais d'un élément nul, et la position éventuelle d'un tel élément dans une structure linguistique, est un sujet de controverse théorique. L'annotation du corpus ne comprend pas de tels marqueurs zéro (Loock, 2016 : 33). Les corpus arborés, qui sont typiquement recommandés pour le traitement des structures grammaticales, sont pour la plupart construits autour de modèles syntaxiques centrés sur le verbe et ne permettent pas la recherche de clauses sans verbe (Landolfi et al., 2010). En outre, peu d'attention a été accordée à la fiabilité de l'extraction automatique des structures dans les corpus existants. Par conséquent, la plupart des travaux de corpus sur le phénomène se sont concentrés sur des structures averbales particulières, qui sont prédéterminées et comprennent un certain marqueur formel interrogeable.

La phrase averbale a principalement été étudiée dans le domaine de la syntaxe. Le fait de s'écarter de la structure canonique des clauses (c'est-à-dire une structure en deux parties composée d'un sujet et d'un prédicat verbal lié au sujet par le biais de l'inflexion de l'accord et du temps) lui a apporté diverses appellations, telles que « fragment », « phrase elliptique », « énoncé non-phrastique », « énoncé nominal », entre autres, mais pas une phrase à part entière. La plupart des discussions syntaxiques sur les structures averbales portent sur la récupération

d'un verbe. Afin de donner aux structures averbales un statut phrastique, bien qu'elliptique, des modèles ont été proposés pour une structure cachée sous-jacente qui est verbale et qui est soit (a) transformée en une structure averbale par une suppression à la surface, soit (b) inclut des éléments spéciaux nuls qui remplacent le verbe (Elugardo & Stainton, 2005). Alternativement, les structures averbales sont proposées comme étant générées sans aucun élément caché dans leur structure, mais leur statut phrastique est retiré en raison du manque d'inflexion pour l'accord et le temps (Barton & Progovac, 2005). Cette dernière analyse semble préférable car elle n'attribue pas de structure supplémentaire à la construction averbale au-delà des mots qui s'y trouvent. Cependant, dans les deux cas, les mots d'une structure averbale sont traités comme s'ils n'étaient pas suffisants pour constituer une phrase complète : soit ils sont les restes incomplets d'une phrase complète et reçoivent le statut de « phrase elliptique », soit ils sont des « énoncés non-phrastiques » autosuffisants.

D'un point de vue sémantique, la phrase averbale pose davantage d'énigmes. À première vue, il est raisonnable de supposer que les éléments qui peuvent être omis dans une phrase sont les éléments qui sont prévisibles et dont le sens peut donc être facilement récupéré afin de compléter le sens de la phrase. Après tout, le sens d'une phrase averbale est généralement clair pour les interlocuteurs. Cependant, récupérer le sens du verbe s'avère être tout aussi problématique que de lui trouver une place dans la structure syntaxique.

Tout d'abord, il est loin d'être clair quel prédicat doit être rétabli ; la même phrase averbale peut être réussie avec un éventail de prédicats qui peuvent être réduits mais pas complètement résolus par le contexte. Par exemple, la phrase en (iv), littéralement « *He to hospital* », peut être utilisée dans le contexte (a) du mouvement de l'agent vers une destination (volontaire, par ex. marcher, conduire, et involontaire, par ex. arriver), (b) de la violence sur une cible militaire illégale (réelle, par ex. tirer, bombarder, et prospective, par ex. viser), (c) des fleurs pour les patients à l'intérieur (acquisition, par ex. acheter, commander, et distribution, par ex. donner, livrer), ainsi que dans d'autres contextes qui réduisent l'éventail des possibilités pour le sens lexical d'un prédicat.

(iv)	Он	в	больницу.
	<i>on</i>	<i>v</i>	<i>bol'nicu</i>
	PRO	PREP	NN.ACC
	<i>he</i>	<i>to</i>	<i>hospital</i>
	<i>il</i>	<i>à</i>	<i>l'hôpital</i>

De même, le contexte peut restreindre l'éventail des sens temporel et aspectuel possibles qu'une phrase peut prendre, par ex., observer un agent qui (a) vient de partir, par opposition à (b) qui se prépare à partir, changerait l'éventail des possibilités temporelles et aspectuelles pour le prédicat. Néanmoins, la structure de la phrase est à elle seule compatible avec de multiples possibilités, et le contexte, au-delà de la délimitation d'une fourchette, ne précise pas quel prédicat particulier de cette fourchette doit être adopté pour être réintroduit.

De surcroît, même si un prédicat particulier pouvait être récupéré, le rétablir semble changer le sens de la phrase. Les analyses sémantiques existantes mettent en évidence les changements associés à l'explicitation d'un prédicat saillant pour certaines structures russes. Des différences concernant la valeur référentielle et les liens situationnels ont été relevées par Selivërstova (1973) et Paillard (1984). Ainsi, un exemple célèbre de Selivërstova (1973) illustre différentes lectures pour les phrases averbales en (v) « У него седые волосы » (*u nego sedye volosy*) et sa contrepartie verbale en utilisant le présent « БЫТЬ » (*byt'* ; lit. « être »).

(v) a.	У	него	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>sedye</i>	<i>volosy</i>
	PREP	PRO.GEN	ADJ.NOM	NN.NOM
	<i>at</i>	<i>him</i>	<i>gray</i>	<i>hair</i>
	<i>à</i>	<i>lui</i>	<i>blancs</i>	<i>cheveux</i>

'He has gray hair.'
« Il a les cheveux blancs. »

b.	У	него	есть	седые	волосы.
	<i>u</i>	<i>nego</i>	<i>est'</i>	<i>sedye</i>	<i>volosy</i>
	PREP	PRO.GEN	V.PR	ADJ.NOM	NN.NOM
	<i>at</i>	<i>him</i>	<i>is</i>	<i>gray</i>	<i>hair</i>
	<i>à</i>	<i>lui</i>	<i>est</i>	<i>blancs</i>	<i>cheveux</i>

'He has some gray hair.'
« Il a des cheveux blancs. »

Tandis que la phrase averbale en (a) est typiquement utilisée pour attribuer la qualité « gris » à l'ensemble de la chevelure, l'alternative verbale en (b) est une structure marquée qui bloque une telle lecture. Cette dernière ne peut être utilisée que pour affirmer l'existence d'une quantité limitée de cheveux gris. Récupérer un prédicat verbal pour une structure averbale, même dans des situations où le verbe semble totalement prévisible, n'est pas simple syntaxiquement et encore moins sémantiquement.

La présente thèse vise à franchir une nouvelle frontière dans la description sémantico-pragmatique des phrases averbales en adoptant une approche contrastive de corpus sur le phénomène en russe et en anglais. Les caractéristiques des phrases averbales et les restrictions possibles de leur utilisation et de leur signification sont analysées à travers un corpus parallèle comparable de 1,4 million de mots, spécialement créé pour permettre le développement et l'utilisation d'une nouvelle méthode précise de récupération des phrases averbales, l'analyse quantitative des phrases averbales par rapport à un corpus de référence, et leur analyse sémantique et pragmatique en contexte.

FEUILLE DE ROUTE

PARTIE 1 : LE CADRE THÉORIQUE

La **première partie** expose le contexte théorique de l'étude. On y commence par définir le phénomène cible dans une perspective qui fait un minimum de présomptions. Les phrases averbales sont formalisées comme des chaînes de texte qui (a) se trouvent en contexte, (b) ne comprennent pas de verbe fini, ni aucune autre forme verbale, et (c) sont délimitées par un marquage de ponctuation initial et final ou par un changement de tours de parole. Les avantages et les limites de la définition actuelle sont explorés. La section examine la terminologie essentielle relative aux structures averbales, y compris les diverses utilisations du terme « ellipse » et la notion de prédicat zéro.

PARTIE 2 : LA MÉTHODOLOGIE

La **deuxième partie** décrit le cadre méthodologique. Trois méthodes séparées pour l'étude du langage – l'approche contrastive, l'approche de corpus et l'approche énonciative – sont présentées en fonction de leurs forces et faiblesses individuelles, ainsi que le bénéfice potentiel de leur union pour l'étude sémantique du phénomène averbale.

Cette étude est guidée par la méthode d'analyse contrastive développée par Jacqueline Guillemin-Flescher (2003) et le principe selon lequel l'analyse de modèles de traduction systématiquement récurrents peut mettre en évidence des contraintes linguistiques qui restent cachées du point de vue d'une seule langue. Afin d'atténuer l'influence de l'interférence de la langue source et de démêler les modèles linguistiques des universaux potentiels de traduction (Nádvořníková, 2017 ; Zanettin, 2013 ; Olohan, 2002 ; Baker, 1993), des mesures sont prises pour inclure des traductions multiples, réciproques et d'une troisième langue dans la composition du corpus. Le corpus est entièrement aligné avec les traductions au niveau de la phrase, ce qui permet de retrouver systématiquement non seulement les phrases averbales, mais aussi leurs correspondances de traduction à travers de multiples textes.

Les aspects sémantiques et pragmatiques peu étudiés des phrases averbales sont actuellement ciblés par l'analyse des facteurs contextuels de la situation énonciative dans laquelle se trouvent les phrases averbales. Les études existantes dans ces domaines, en raison de difficultés technologiques, se sont pour la plupart éloignées des corpus. Par conséquent, ils sont vulnérables aux critiques auxquelles sont confrontées les données qui ont été sélectionnées de manière éclectique à partir de diverses sources par le chercheur pour illustrer un modèle préconçu (Kohnen, 2015 ; McEnery et al., 2006 ; McEnery & Wilson, 2001 ; Garside et al., 1997). Cependant, afin de combler la lacune dans l'analyse sémantique et pragmatique du point de vue du corpus, il est nécessaire de rendre le contexte central à toutes les étapes de la conception du corpus et de l'extraction des données. Cela ajoute une autre complication à l'extraction des phrases averbales dans des corpus. Maintenir le lien vital avec le contexte afin de permettre l'analyse sémantique n'est pas encore une tâche facile car elle ne concerne pas les lignes de concordance typiques centrées sur un mot ou une expression, mais des phrases averbales et simultanément leurs correspondances de traduction. Ce défi est actuellement surmonté grâce à une conception de corpus et une méthode d'extraction qui résout tous les trois : phrases averbales, correspondances de traduction et contexte.

La nouvelle méthode actuelle d'extraction automatique a été développée avec un souci particulier de précision afin de minimiser certains des biais liés à l'utilisation des outils de corpus (Anthony, 2012). La méthode vise à être reproductible sur n'importe quel texte brut et, dans sa substance, consiste à classer des unités de phrase correctement délimitées et étiquetées POS tout en maintenant leur place dans la structure globale. Tout d'abord, les textes sont segmentés en phrases à l'aide d'un script qui réunit une série particulière d'algorithmes qui ciblent les marqueurs graphiques afin de distinguer le discours direct, les clitics et d'autres enjeux qui posent des problèmes pour le rappel et la précision de l'extraction de phrases

averbales. Les textes sont ensuite alignés manuellement sur plusieurs traductions au niveau des phrases et réunis dans une structure de textes parallèles. Tous les mots sont ensuite automatiquement séquencés pour former un seul fil de données, étiquetés morpho-syntaxiquement, les segments de phrase sont classés en verbal et non-verbal, et l'alignement est visualisé et exploré à l'aide du logiciel de traitement de texte et d'analyse statistique multi-niveaux Trameur (Fleury, 2019a; Fleury & Zimina, 2014). Ce dernier traite des données segmentées sur mesure, permet la correction automatique d'une grande partie des erreurs morphosyntaxiques de l'étiquetage, permet la visualisation de phrases averbales alignées avec des traductions multiples dans leur contexte original, ainsi que l'analyse statistique par rapport à un corpus de référence de phrases verbales.

Après l'extraction, les phrases averbales et leurs correspondances de traduction dans les sous-parties du corpus sont segmentées manuellement en unités propositionnelles plus petites et analysées manuellement pour les caractéristiques syntaxiques, sémantiques et pragmatiques. Parmi les catégories annotées manuellement figurent l'ellipse verbale basée sur l'antécédent, le sujet et le prédicat, la correspondance verbale de traduction et le temps et l'aspect de celle-ci, le type de discours, la structure de l'information, l'acte de langage direct et indirect, ainsi que d'autres catégories dont les définitions sont présentées dans cette partie. La partie définit également l'analyse statistique qui a été effectuée sur les phrases averbales et leur contexte dans une perspective monolingue, y compris les éléments caractéristiques et les n-grammes, et fournit également un guide pour la classification sémantique des résultats identifiés statistiquement.

Le corpus est examiné sous trois angles : monolingue, parallèle et de traduction de langue tierce. Premièrement, dans une perspective monolingue, ce ne sont pas seulement les textes originaux, mais aussi les traductions, qui sont traités à part entière comme de véritables échantillons de production langagière. Concernant l'utilisation des traductions dans l'analyse monolingue, la présente étude se situe du côté de Zanettin (2014), Olohan (2002), Baker (1996), Biber (1993), entre autres, en soutenant que les traductions constituent une partie importante de la production et de la réception d'une langue, et qu'en conséquence, elles doivent être incluses dans toute conception de corpus qui se veut représentative d'une langue, que ce soit en termes de variété linguistique (Biber, 1993) ou d'usage proportionnel (Leech, 2007). Ainsi, les originaux russes sont comparés non seulement aux originaux anglais, mais aussi aux traductions russes, et de même pour l'anglais. Les quatre types de langues – originaux russes, originaux anglais, traductions russes, traductions anglaises – sont comparés en termes de

fréquences normalisées, de classification sémantique des éléments caractéristiques et des n-grammes, et des catégories annotées manuellement.

Deuxièmement, en combinant les principes de l'analyse contrastive de Guillemín-Flescher (2003) avec les critères des corpus parallèles fiables (Nádvořníková, 2017 ; Stolz, 2007a ; McEnery & Xiao, 2008 ; Malmkjær, 1998), le corpus est étudié dans une perspective parallèle. La présente analyse recherche des modèles de correspondance qui se répètent dans plusieurs traductions du même original, dans plusieurs œuvres d'auteurs différents, dans des genres différents et dans des sens de traduction réciproques. En plus des deux directions de traduction typiques, (a) des originaux russes vers les traductions anglaises, et (b) des originaux anglais vers les traductions russes, la volonté de traiter les traductions comme un type de langue à part entière et la nature du phénomène conduisent la présente étude à inclure également l'analyse des modèles de correspondance (c) des traductions anglaises vers les originaux russes et (d) des traductions russes vers les originaux anglais. Pour la présente analyse parallèle, une attention particulière est accordée aux correspondances verbales des phrases averbales sans antécédent et à leur corrélation avec les catégories annotées manuellement.

Enfin, le corpus est examiné du point de vue de la traduction d'une langue tierce. Afin d'identifier les modèles qui indiquent des contraintes spécifiques à la langue, il est nécessaire de les distinguer des modèles qui résultent de l'influence de la langue source sur la traduction (par exemple, la traduction littérale d'une phrase verbale anglaise peut aboutir à une phrase verbale en russe, même si dans le contexte de la traduction d'une autre langue, ou dans un contexte non traductionnel, une phrase averbale serait typiquement utilisée), ainsi que ceux qui peuvent être dus à l'acte même de la traduction (par exemple, la tendance à simplifier dans le passage d'une langue à l'autre et d'autres caractéristiques universelles potentielles de la traduction). Pour tenter de contrôler l'interférence de la langue source et les universaux potentiels de la traduction, un sous-corpus composé de traductions du français en russe et en anglais est ajouté. Ce sous-corpus fait l'objet des mêmes analyses monolingues et parallèles. La perspective de la troisième langue permet de contrôler les interférences entre le russe et l'anglais et fournit un autre type de données traduites à comparer avec les originaux.

PARTIE 3 : LE CORPUS

La **troisième partie** présente le corpus parallèle et comparable de 1,4 million de mots, entièrement aligné au niveau de la phrase et étiqueté morpho-syntaxiquement, qui a été construit pour la présente étude. Son domaine général est la fiction réaliste de la fin du 19^{ème} siècle au début du 21^{ème} siècle qui a reçu une forte exposition et a été traduite dans les langues cibles. Le total de 32 textes se compose de 13 œuvres différentes et comprend 12 originaux, 16 traductions et 4 traductions de troisième langue, qui se répartissent équitablement entre le russe et l'anglais. Trois genres fonctionnels sont représentés, y compris la fiction discursive, qui représente le principal genre ciblé, ainsi que la pièce de théâtre et la fiction narrative, ces deux dernières représentant des groupes de contrôle afin de distinguer les modèles dus à l'interférence des genres.

Cette partie décrit la composition du corpus, précise les critères de sélection et évalue les textes en fonction de ces critères. La comparabilité du corpus est analysée et évaluée comme étant librement comparable, selon le cadre d'échantillonnage, y compris la taille, la période, le genre, la langue (i.e. russe et anglais) et le type de langue (i.e. source, traduction, traduction de troisième langue).

Le corpus vise à être représentatif d'une grande variété de phrases averbales et à en assurer leur fréquence élevée. La représentativité en termes de proportion de locuteurs ou de destinataires des langues ciblées (par exemple, la proportion d'anglophones représentée par la composition du présent corpus) est échangée en faveur de la variété linguistique du phénomène ciblé.

PARTIE 4 : LES RÉSULTATS

La **quatrième partie** regroupe la description des résultats et l'analyse quantitative et qualitative en plusieurs chapitres qui ciblent chacun une question différente.

Chapitre 1 : Les Explications Syntaxiques

Le **premier chapitre** de cette partie aborde les explications syntaxiques des différences typologiques entre les langues. Il analyse les résultats concernant la fréquence des phrases averbales dans les deux langues, ainsi que la distinction fondamentale entre les phrases averbales elliptiques et non elliptiques. Sur la base des données présentes, ce chapitre révèle d'abord la fréquence d'utilisation des phrases averbales dans les deux langues et établit que la surreprésentation attendue des phrases averbales en russe par rapport à l'anglais est statistiquement significative. Il montre ensuite que l'utilisation plus fréquente de phrases averbales en russe par rapport à l'anglais ne peut pas être expliquée par le recours à des différences typologiques dans la productivité des ellipses verbales syntaxiques. Contrairement aux attentes, les ellipses à l'antécédent sont significativement surreprésentées non pas en russe, la langue qui se distingue dans la famille des langues indo-européennes par le fait de permettre l'utilisation la plus libérale de phrases averbales, mais plutôt en anglais, la langue qui est typiquement associée à une dépendance à la phrase verbale finie. De plus, les résultats révèlent que les énoncés averbaux sont principalement non-elliptiques dans les deux langues. Afin de rendre compte du phénomène averbal en général (c'est-à-dire la forte proportion du type non-elliptique) et des différences de fréquence interlinguistique observées, une attention particulière est requise en termes d'analyse sémantique et pragmatique. Cette attention est portée dans les résultats des chapitres suivants qui se concentrent sur la détermination du sens sémantique et pragmatique et des possibles restrictions des phrases averbales.

Le premier chapitre pose également les fondations de l'argument qui traverse toute la thèse, à savoir que les structures averbales constituent un type de phrase à part entière et ne doivent pas être traitées comme des réductions syntaxiques ou sémantiques des phrases verbales. Ce chapitre fournit des preuves empiriques montrant que le phénomène des phrases averbales ne s'explique pas par un élément nul, ni par la suppression d'un prédicat, qui existe dans la phrase averbale et qui est reconstructible à partir du contexte linguistique. Il est proposé que le prédicat zéro puisse aider à expliquer le type de phrase elliptique, c'est-à-dire le type qui implique potentiellement la suppression d'un prédicat sous-jacent. Toutefois, d'après les résultats actuels, la plupart des phrases averbales semblent être indépendantes, au moins syntaxiquement, d'un prédicat existant.

Chapitre 2 : Les Caractéristiques Sémantiques

Le **chapitre deux** fournit des preuves empiriques à l'appui de l'argument selon lequel elles sont également des phrases indépendantes sur le plan sémantique. Malgré leur longueur réduite, les phrases averbales présentent une plus grande complexité lexicale que la phrase moyenne en anglais. Un tel résultat n'est pas attendu pour de simples réductions et fragments de phrases verbales.

Ce chapitre analyse également les caractéristiques sémantiques distinctives des phrases averbales. Il présente les résultats d'une classification sémantique de près de 20 000 phrases averbales en fonction des constituants lexicaux qui les distinguent statistiquement des phrases verbales. Les constituants qui distinguent les phrases averbales des phrases verbales dans les deux langues se révèlent statistiquement inclure des éléments et segments lexicaux indexicaux, de quantification négative, expressifs et informels. Alors que la référence indéfinie et les mots « *wh* » sont clés dans les phrases averbales anglaises, les éléments marquant la deixis et l'emphase de l'intensité sont clés pour le russe.

Chapitre 3 : La Langue de Traduction et le Genre

Le **chapitre trois** explore la question de la langue de traduction et du genre textuel. Quelle est l'influence de la langue traduite sur les phrases averbales ? Quelle est l'influence du genre textuel sur l'utilisation des phrases averbales ? La phrase averbale s'avère être un phénomène linguistique qui est spécifique à la langue et qui n'est pas sensiblement affecté par la traduction, tant d'un point de vue monolingue que parallèle. On constate que les phrases averbales sont plus affectées par la différence entre le discours et la narration plutôt que par le fait que la source de la langue soit un original ou une traduction.

Ces résultats sont favorables à une analyse contrastive du phénomène. Une analyse linguistique basée sur les traductions est typiquement confrontée à la nécessité de se défendre contre l'argument selon lequel les traductions représentent un type de langage dans lequel un certain phénomène particulier peut disparaître, submergé par les caractéristiques universelles potentielles du langage traduit. Les résultats actuels suggèrent que cet argument peut être rejeté avec confiance pour l'étude des phrases averbales. D'un point de vue contrastif, les phrases sans verbe se comportent de manière similaire dans les originaux et les traductions en termes

de correspondances de traduction verbale ; d'un point de vue monolingue, seule une légère différence est observée dans la fréquence des phrases sans verbe.

Ce chapitre analyse également la corrélation entre les phrases averbales et le discours direct. Cette corrélation est exposée à travers (a) l'annotation manuelle des phrases averbales pour le discours direct versus la narration, (b) l'analyse statistique des constituants des phrases averbales qui révèlent la saturation des éléments qui nécessitent une résolution contextuelle (par exemple les indexicaux) et de l'interaction (par exemple les interjections, les mots d'interrogation), et (c) la variation de la fréquence des phrases averbales à travers les genres examinés. Cette corrélation met en évidence les exigences pragmatiques sur les connaissances communes entre les interlocuteurs. En outre, il est proposé que la phrase averbale puisse être une caractéristique linguistique permettant d'établir une distinction de genre fonctionnelle entre la fiction discursive, le discours scénarisé et la fiction narrative.

Chapitre 4 : La Structure Informationnelle

Le **chapitre quatre** analyse la structure informationnelle (Lambrecht, 1994) des phrases averbales d'un point de vue monolingue et contrastif. Considérées de manière monolingue, les phrases averbales se retrouvent dans tous les types de structures informationnelles, y compris les phrases « topique-commentaire », « identificateur » et « thétiqes ».

Les résultats de ce chapitre montrent également que les verbes qui sont pragmatiquement impliqués par les phrases averbales s'avèrent, au moyen de l'analyse contrastive, faire partie du focus. Cette constatation remet en question la mesure dans laquelle la phrase averbale peut être expliquée par l'omission d'un élément prévisible et reconstituable. Elle fournit des preuves supplémentaires à l'appui de l'argument selon lequel les constituants d'une phrase averbale peuvent être suffisants pour exprimer une pensée complète et satisfaire aux exigences pour constituer des instances complètes de prédication.

Le chapitre révèle également une différence dans la manifestation du topique informationnel dans les phrases averbales des deux langues. Il est observé que la manifestation du topique est corrélée avec la manifestation du verbe. Ce résultat suggère que certaines des différences de fréquence entre les deux langues concernant les phrases averbales peuvent en partie être dues à des différences typologiques dans l'activation du topique. Ce chapitre apporte

ainsi une explication pragmatique qui peut expliquer une partie de la variation des phrases averbales entre les langues.

Chapitre 5 : Le Sens Temporel et Aspectuel

Le **chapitre cinq** aborde la question des potentielles restrictions sémantiques, temporelles et aspectuelles des phrases averbales. Que sont les sens temporel et aspectuel sans leur marqueur typique – le verbe « constellation » (Smith, 1997 : 10) ? Ce chapitre explore la relation entre les phrases averbales et les propriétés sémantiques qui sont typiquement associées aux clauses verbales canoniques – le temps et l’aspect (lexical « situation » et grammatical « viewpoint »), ainsi que leur relation avec le sens lexico-verbal. Les présentes données parallèles et la méthodologie contrastive permettent d’aborder cette question sous un angle différent de celui qui est généralement adopté dans les analyses existantes des phrases averbales. Au lieu de tester les propriétés sémantiques sur des éléments non verbaux, dans la présente analyse les traductions verbales sont utilisées pour étudier le potentiel temporel, aspectuel et lexical qui est impliqué conversationnellement dans une phrase averbale.

Les résultats montrent que les phrases averbales ont le potentiel d’impliquer un large éventail de sens temporel, aspectuel et lexico-verbal. Les corrélations spécifiques à chaque langue sont utilisées pour révéler les différences typologiques. En défendant une sémantique complète pour les phrases averbales, les propriétés sémantiques typiquement associées au verbe sont actuellement considérées comme une question d’implicature pragmatique par opposition à l’implication sémantique.

Chapitre 6 : L’Acte de Langage

Le **chapitre six** révèle une autre explication pragmatique de la variation de la phrase averbale non elliptique entre les langues. Le chapitre montre que l’absence de verbe peut être utilisée comme un marqueur grammatical en anglais pour signaler qu’une question est un acte de langage indirect.

Il montre tout d’abord que l’analyse des correspondances de traduction distingue les questions, par rapport aux autres types de phrases, comme étant particulièrement sensibles à la

présence et à l'absence du verbe : quel que soit le type de langue (source, traduction ou traduction d'une troisième langue), les verbes varient plus fréquemment dans les correspondances de questions averbales que dans tous les autres types d'énoncés réunis.

Deuxièmement, d'un point de vue monolingue, dans les deux langues, les questions averbales sont corrélées à des actes de langage indirect. Autrement dit, l'absence du verbe dans les questions tend à être associée à des situations dans lesquelles le locuteur ne demande pas d'informations à son destinataire (par exemple, les questions de type rhétorique, les questions de type surprise) dans les deux langues.

Cependant, une analyse contrastive met en évidence des différences typologiques dans l'utilisation du verbe dans les questions. Les correspondances verbales anglaises des questions averbales russes sont corrélées aux actes de langage directs dans tous les types de langue. En d'autres termes, on observe que l'anglais utilise l'absence du verbe comme un marqueur grammatical des actes de langage indirect dans le questionnement. Ces résultats suggèrent qu'en anglais, le verbe permet de distinguer les questions en tant qu'actes de langage directs ou indirects, alors que dans les questions russes, la distinction directe-indirecte n'est pas liée au verbe. Ainsi, plutôt que la syntaxe, ces résultats suggèrent qu'une différence typologique dans l'utilisation pragmatique du verbe dans les questions contribue à expliquer les différences de fréquence des phrases averbales entre les langues.

PARTIE 5 : CONCLUSION

En **conclusion**, les principales contributions de la présente thèse sont les suivantes :

- une explication phrastique qui justifie les phrases averbales comme constituant un phénomène linguistique légitime;
- le développement de l'analyse sémantique et pragmatique des phrases averbales en russe et en anglais par une approche contrastive de corpus ;
- une méthode semi-automatique d'extraction de phrases averbales pour l'analyse sémantique ;

- un nouveau corpus parallèle et comparable de treize œuvres de fiction réaliste russe et anglaise des 19^e – 21^e siècles.

Plusieurs questions portant sur les différences typologiques entre le russe et l'anglais sont explorées, y compris les preuves empiriques concernant :

- Dans quelle mesure les différences syntaxiques dans la productivité de l'ellipse parviennent-elles à expliquer la variation entre les langues ?
- Quelles caractéristiques sémantiques et pragmatiques distinguent les phrases averbales par rapport aux phrases verbales dans les langues ? Quels facteurs contextuels influencent leur utilisation ?
- Comment est-ce que le genre textuel et la langue de traduction affectent les phrases averbales ?
- Que se passe-t-il avec les phrases averbales dans la traduction du russe vers l'anglais et vice versa ? Comment est-ce que les catégories sémantiques associées avec le verbe se rapportent aux phrases averbales ? Les phrases averbales sont-elles restreintes sur le plan sémantique, pragmatique ou structurel ?
- Comment les phrases averbales sont-elles utilisées dans les actes de langage ?

Les implications plus larges de la thèse ont trait à l'explication théorique de la phrase averbale. Les implications plus larges de la thèse ont trait à l'explication théorique de la phrase averbale. En plus des résultats spécifiques à chaque langue, l'analyse apporte des arguments concernant des questions plus générales, notamment :

- Les phrases averbales sont-elles des réductions de phrases verbales ?
- Dans quelle mesure leur sens relève-t-il de l'assertion versus de l'implicature ?
- Où se situent les phrases averbales dans le débat entre le sens compositionnel et le sens conventionnel ?

Nous proposons une explication du statut phrastique des structures averbales qui ne repose pas sur une structure syntaxique cachée, mais qui insiste plutôt sur la primauté du focus de la structure informationnelle comme le critère phrastique nécessaire. On propose que le contenu

sémantique d'une phrase averbale soit traité en termes d'éléments assertifs linguistiquement explicites et d'implicatures conversationnelles révélées par le contexte.

Une dernière contribution de la présente thèse porte sur le traitement du corpus concernant l'absence du verbe. Le développement de la conception du corpus nécessaire est généralement un processus en plusieurs étapes, comme le souligne Biber (1993), entre autres. La conception du corpus actuel et la méthode d'extraction sont élaborées à partir de plusieurs études pilotes de moindre envergure. L'attention est attirée sur les défis spécifiques auxquels est confrontée la récupération automatique de la classe ouverte des structures grammaticales centrées sur l'absence du verbe. Des suggestions méthodologiques pour la poursuite du traitement des phrases averbales dans les corpus sont proposées, y compris les développements nécessaires et les prochaines étapes dans l'étude contrastive du phénomène.

LES IMPLICATIONS

L'un des principaux objectifs de ce travail a été de défendre le phénomène de la phrase averbale en tant que phénomène légitime, à la fois sur des fondements théoriques et empiriques. Nous espérons également avoir fait des traits larges, mais néanmoins clés, dans le portrait sémantico-pragmatique de la phrase averbale prototypique et de son utilisation en russe et en anglais. Si nous avons réussi à atteindre ces objectifs, cela signifie qu'il y aura beaucoup de travail à venir, dont un élargissement des annotations ciblées.

Plus précisément, les contributions suivantes sont estimées avoir été apportées par la présente thèse.

1 UNE EXPLICATION PHRASTIQUE

Nous avons fourni une *explication phrastique* qui justifie les phrases averbales en tant que phénomène linguistique légitime. Comme l'a montré la première partie, le statut des structures averbales a alimenté un riche débat au cours de l'histoire. En faisant le point sur les arguments existants sur la question, et en les considérant à la fois d'un point de vue théorique et du point de vue des données actuelles, nous avons tenté de faire avancer le « *chariot de Krylov* » vers la reconnaissance du statut phrastique des structures averbales, sans toutefois leur attribuer des propriétés verbales cachées. Les phrases averbales existent, et elles sont entièrement averbales – tant sur le plan de leur syntaxe que de leur sémantique. Le fait de reconnaître leur existence, et d'admettre leur absence de prédicat syntaxique, nous révèle que

les frontières des modèles syntaxiques du langage sont tracées aux structures canoniques idéalisées. Les tentatives d'introduire une structure cachée ou d'attribuer des catégories flexionnelles verbales à des éléments non verbaux ne font que repousser la conclusion inévitable qui se dessine déjà avec force dans des arguments comme ceux d'Ellen Barton, Ljiljana Progovac, Robert Stainton et Reinaldo Elugardo. Seulement, dans leur argumentation justifiée, ces derniers semblent avoir choisi l'option de la fuite en reléguant les structures au statut de « non phrastique », alors que nous, tout en étant d'accord avec leurs critiques des tentatives existantes, pensons qu'il est néanmoins nécessaire, et possible, de défendre le statut des structures averbales, coûte que coûte, dans l'intérêt que nos modèles construits reflètent l'usage de la langue.

2 UNE APPROCHE INTERDISCIPLINAIRE À LA DESCRIPTION SÉMANTICO-PRAGMATIQUE

Ayant défendu leur existence, nous estimons également avoir fait des progrès dans la description *sémantique et pragmatique* des phrases averbales en russe et en anglais. En particulier, nous avons combiné les forces de plusieurs disciplines en une seule méthode et étudié les caractéristiques sémantico-pragmatiques du phénomène en utilisant une *approche contrastive de corpus*. Dans la deuxième partie, nous avons souligné le caractère unique de notre cadre méthodologique. Cela ne concerne pas seulement le choix particulier des langues, le russe et l'anglais, pour leurs différences typologiques. Elle implique également la conception et la construction du corpus de manière à répondre aux problématiques liées à l'utilisation des traductions pour l'étude des phénomènes linguistiques. En outre, les inquiétudes soulevées par l'utilisation de corpus pour une analyse sémantico-pragmatique plus qualitative ont été adressées par le développement du corpus de manière attentive en termes de l'exactitude, de la pertinence de la requête, de la récupération du contexte et de la mise en œuvre des calculs. De cette manière, nous espérons avoir commencé à lever certaines des limites liées aux données imposées aux travaux de recherche précédents et avoir éteint une partie de la soif pour l'utilisation de corpus dans l'étude de la phrase averbale en tant que phénomène linguistique.

3 LA RECHERCHE D'ABSENCE

La récupération des phrases averbales dans les corpus est un obstacle majeur qui a été actuellement dispersé. Nous avons développé une méthode pour l'extraction automatique de phrases averbales. En outre, cette méthode ne se contente pas de repérer les phrases averbales, elle les repère également de manière à permettre la poursuite de leur analyse – sur le plan sémantique, contrastif et d'un point de vue textométrique. Elle récupère la phrase averbale, préserve le contexte, aligne la phrase averbale et le contexte avec des traductions multiples, trouve automatiquement les traductions de la phrase averbale, permet l'accès aux données nécessaires pour l'analyse sémantico-pragmatique et l'annotation des originaux et des traductions, et rend possible les calculs textométriques concernant la phrase averbale, ses traductions et son contexte par rapport à un corpus de référence.

4 UN NOUVEAU CORPUS PARALLÈLE ET COMPARABLE

Enfin, un *nouveau corpus parallèle et comparable* de treize œuvres de fiction réaliste russe et anglaise des 19^e – 21^e siècles a été créé. Comptant 1,4 million de mots, il est d'une taille particulièrement importante tant pour les études de corpus parallèles que pour les études de phénomènes averbales. Les 32 textes comprennent plusieurs traductions par œuvre originale. L'ensemble du corpus est annoté morpho-syntaxiquement (avec une attention particulière à l'exactitude concernant le phénomène de l'averbale et la verbalité), il est segmenté en phrases (y compris la segmentation du discours direct par rapport à la narration), et il est entièrement aligné au niveau de la phrase sur plusieurs traductions. L'ensemble du corpus est annoté morpho-syntaxiquement (avec une attention particulière à l'exactitude concernant le phénomène averbal et le verbe), il est segmenté en phrases (y compris la segmentation du discours direct par rapport à la narration), et il est entièrement aligné au niveau de la phrase sur plusieurs traductions.

Dans notre analyse, nous nous sommes efforcés de combiner les perspectives qualitatives et quantitatives afin d'aborder plusieurs questions relatives aux similitudes et aux différences

typologiques entre le russe et l'anglais. Les implications de la présente analyse concernent en particulier les questions suivantes.

5 DANS QUELLE MESURE L'ELLIPSE SYNTAXIQUE PEUT-ELLE EXPLIQUER LE PHÉNOMÈNE DE LA PHRASE AVERBALE ?

Nous avons commencé par établir que la différence entre le russe et l'anglais en termes de phrases averbales est statistiquement significative, comme on s'y attendait. On a constaté que le russe présentait presque deux fois plus de phrases averbales que l'anglais dans le corpus actuel de 1,4 million de mots. Nous avons ensuite étudié la différence typologique bien connue entre le russe et l'anglais qui concerne une productivité particulièrement élevée de l'ellipse syntaxique dans la première langue. Ayant annoté les phrases averbales (et leurs correspondances de traduction) en termes d'énoncés syntaxiquement « elliptiques » et « non-elliptiques », nous avons observé que la productivité syntaxique de l'ellipse est totalement inadéquate pour expliquer la variation de la fréquence des phrases averbales entre les deux langues. Premièrement, nous avons constaté que les ellipses étaient significativement surreprésentées en anglais, et non dans le russe, une langue syntaxiquement plus propice aux ellipses. Deuxièmement, les phrases averbales se sont avérées être principalement non elliptiques dans les deux langues. Des résultats supplémentaires se trouvent dans la partie 4.1. De tels résultats suggèrent que l'ellipse syntaxique ne peut pas expliquer le phénomène de la phrase averbale, et appellent à son analyse sémantique et pragmatique.

Par ailleurs, ces résultats soutiennent également le traitement des phrases averbales comme des structures largement indépendantes sur le plan syntaxique. Si les phrases averbales étaient de simples réductions syntaxiques des phrases verbales, la productivité russe dans ce domaine aurait dû expliquer les différences de fréquence entre les langues. Or, ce n'est pas le cas. Par conséquent, les présents résultats impliquent également que les phrases averbales ne sont pas des réductions syntaxiques des phrases verbales.

6 QUELLES CARACTÉRISTIQUES SÉMANTIQUES ET PRAGMATIQUES DISTINGUENT LES PHRASES AVERBALES DES PHRASES VERBALES DANS LES DEUX LANGUES ?

L'une des façons dont nous avons abordé cette question est en utilisant une combinaison d'outils textométriques de corpus et d'annotation manuelle. La discussion dans la partie 4.2 se penche sur ce qui se passe à l'intérieur de la phrase averbale prototypique. Les éléments et les n-grammes caractéristiques qui distinguent statistiquement les phrases averbales du reste du corpus ont été calculés, puis observés en contexte et en parallèle avec leurs traductions, et enfin classés selon un schéma sémantico-pragmatique. Les résultats explorent les diverses fonctions statistiquement clés des phrases averbales dans les deux langues, y compris :

- le marquage de l'accord et du désaccord,
- l'expressivité,
- la quantification,
- la formalité et l'informalité,
- l'immédiateté et la deixis,
- l'interrogation
- et l'emphase.

En comparant les fonctions dans les deux langues, on constate que la référence indéfinie et les mots « *wh* » sont essentiels dans les phrases averbales anglaises, tandis que l'accentuation de l'intensité se révèle particulièrement importante pour les phrases averbales russes.

Les résultats de cette section comprennent également une analyse de la longueur et de la complexité lexicale.

Enfin, nous évaluons le niveau potentiel de conventionnalisation sémantique des phrases averbales sur la base d'une analyse de leurs segments répétés statistiquement caractéristiques (les n-grammes internes qui les distinguent des autres phrases). Sur la base de ces résultats, nous soutenons que les phrases averbales ne sont pas aussi conventionnelles qu'on le pense parfois.

7 COMMENT EST-CE QUE LA LANGUE DE LA TRADUCTION ET LE GENRE INFLUENCENT-ILS LE PHÉNOMÈNE DE LA PHRASE AVERBALE ?

Nous montrons que, du point de vue contrastif, les phrases averbales se comportent de manière similaire dans les originaux, les traductions et les traductions d'une langue tierce en termes de correspondances verbales de traduction. D'un point de vue monolingue, seule une légère différence de fréquence est observée. Les phrases averbales s'avèrent donc être un phénomène linguistique spécifique à une langue, qui n'est pas affecté de manière significative par la traduction. Nos résultats montrent plutôt que les considérations de genre, et la différence entre le discours direct et la narration, sont plus importantes pour la phrase averbale.

Une corrélation est mise en évidence entre les phrases averbales et le discours direct (par le biais de plusieurs moyens, notamment par l'annotation manuelle de « discours direct » et de « narration », par la variation de fréquence entre les genres, et par les éléments lexicaux statistiquement caractéristiques). Cette corrélation met en évidence la centralité de la fonction pragmatique, des connaissances partagées et des paramètres interactifs pour l'utilisation de phrases averbales. Nous proposons en outre que les phrases averbales puissent être une caractéristique linguistique des différences de genre.

8 LES PHRASES AVERBALES OMETTENT-ELLES DES INFORMATIONS PRÉVISIBLES ET RÉCUPÉRABLES ?

Nous avons étudié cette question dans une perspective contrastive en analysant la structure informationnelle des phrases averbales et de leurs traductions. Au moyen de l'annotation de phrases averbales et de leurs correspondances de traduction, nous montrons que les verbes qui sont impliqués conversationnellement par les phrases averbales font largement partie du *Focus* de la structure informationnelle. Cette découverte remet en question la mesure dans laquelle la phrase averbale peut être expliquée en termes de prédicat prévisible et restructurable, et fournit des preuves supplémentaires à l'appui de l'argument qui sous-tend la présente thèse, à savoir que les phrases averbales peuvent être suffisantes pour exprimer une

pensée complète et satisfaisante aux exigences pour constituer des instances complètes de prédication.

En ce qui concerne la structure informationnelle, nous montrons également que les phrases averbales ne sont pas restreintes à un type particulier de structure informationnelle et qu'elles sont observées comme des phrases de type « topique-commentaire », « identificateur » et « thétiques » dans les deux langues.

De plus, l'annotation révèle une différence contrastive entre la réalisation du topique informationnel dans les phrases averbales russes et leurs traductions anglaises. On observe que la manifestation du topique en anglais est en corrélation avec la manifestation du verbe. Nous proposons que certaines des différences de fréquence entre les langues concernant les phrases averbales puissent être en partie dues à des différences pragmatiques typologiques dans l'activation du topique.

9 COMMENT EST-CE QUE LES CATÉGORIES SÉMANTIQUES ASSOCIÉES AU VERBE, Y COMPRIS LA SIGNIFICATION TEMPORELLE ET ASPECTUELLE, SE RAPPORTENT AUX PHRASES AVERBALES ?

Les propriétés sémantiques typiquement associées au verbe sont actuellement considérées comme relevant de l'implicature pragmatique et non de l'implication sémantique. En étudiant la question par une approche contrastive, nous montrons que les phrases averbales dans les deux langues ont le potentiel d'impliquer conversationnellement un large éventail de sens temporel, aspectuel et lexico-verbal. L'annotation manuelle a été effectuée afin d'examiner l'aspect de « situation », ainsi que celui de « viewpoint », des correspondances verbales. De plus, les correspondances verbales du corpus entier donnent un aperçu du sens lexical et temporel potentiel des verbes impliqués conversationnellement dans les phrases averbales dans les deux langues.

10 COMMENT EST-CE QUE LES PHRASES AVERBALES SONT UTILISÉES DANS LES ACTES DE LANGAGE ?

Nous montrons que dans les questions averbales non-elliptiques des deux langues, l'absence du verbe tend à être associée à des actes de langage indirect, c'est-à-dire des situations dans lesquelles le locuteur ne demande pas d'information à l'allocutaire.

En outre, nous présentons des résultats indiquant que l'absence du verbe peut être utilisée comme un marqueur grammatical en anglais pour signaler qu'une question est un acte de langage indirect. L'analyse contrastive montre que les questions, par rapport aux autres types d'énoncés, sont particulièrement sensibles à la présence et à l'absence du verbe. En anglais, le verbe permet de distinguer entre les actes de langage directs et indirects, alors qu'en russe, la distinction direct-indirect n'est pas observée comme étant liée au verbe. Ce résultat suggère que les différences de fréquence des phrases averbales entre les langues s'expliquent en partie par une différence typologique dans l'utilisation pragmatique du verbe dans les questions.

11 LES IMPLICATIONS PLUS LARGES

Pour conclure, en plus des éclaircissements propres à chaque langue concernant les caractéristiques sémantico-pragmatiques des phrases averbales, les implications plus larges de cette thèse concernent le fondement théorique de la phrase averbale. La présente analyse a apporté des arguments concernant des problématiques plus générales, notamment : *Les phrases averbales sont-elles des réductions de phrases verbales ? Dans quelle mesure est-ce que leur sens est une question d'assertion versus d'implicature ? Où se situent les phrases averbales dans le débat entre le sens compositionnel et le sens conventionnel ?*

Nous avons proposé une explication du statut phrastique des structures averbales qui ne fait pas appel à une structure syntaxique cachée, mais qui insiste plutôt sur la primauté du focus de la structure informationnelle comme étant le critère phrastique clé. Notre argument est que le contenu sémantique d'une phrase averbale doit être traité en termes d'éléments assertifs linguistiquement explicites et d'implicatures conversationnelles révélées par le contexte.

La contribution finale de la présente thèse concerne le *traitement de l'absence du verbe dans les corpus*. La présente méthode d'extraction de phrases averbales et de conception de corpus a été développée par un processus à plusieurs étapes, consistant en des études pilotes plus petites qui avaient des objectifs qualitatifs à chaque étape. Cela nous a permis d'attirer l'attention sur les défis spécifiques auxquels est confronté le traitement automatique de la classe ouverte des structures grammaticales centrées sur l'absence du verbe, et aussi de surmonter un grand nombre d'entre eux dans la présente étude. Ce faisant, nous espérons avoir ouvert des perspectives la poursuite du traitement contrastif des phrases averbales dans les corpus.

LIST OF FIGURES

Figure 1.	Šaxmatov’s Typology of One-Part and Two-Part Sentences.....	79
Figure 2.	The syntactic structure of a canonical sentence	84
Figure 3.	Target scope of syntactic sentence models	113
Figure 4.	Multiple potential propositional meanings in single context	121
Figure 5.	Multiple propositional meanings in different contexts	126
Figure 6.	Reintroduction of the predicate changes meaning	129
Figure 7.	The five necessary criteria of a Topic	154
Figure 8.	Russian Morphological Case	239
Figure 9.	Typology of Errors for Automatic Processing of Verbless Sentences.....	260
Figure 10.	Alignment segmentation for verbless sentences.....	268
Figure 11.	Aligned corpus structure	273
Figure 12.	Verbless sentence aligned with multiple translations in context	277
Figure 13.	Summary of Corpus and Legend to Figure 14.....	302
Figure 14.	Summary of Corpus by language, translation, year of publication, genre	303
Figure 15.	Comparability of Russian and English in Number of Texts	316
Figure 16.	Comparability of Russian and English in Words: Language Source Type	317
Figure 17.	Comparability of Russian and English in Words: Genre	318
Figure 18.	Comparability of Russian and English in Words: Time Period	319
Figure 19.	Parallel-Text Comparability Map	321
Figure 20.	Sentences by language (abs. fq.).....	324
Figure 21.	Verbless sentences in Russian and English by text (abs. fq.).....	324
Figure 22.	Verbless sentences in Russian and English (normalized per 100 sentences)	325
Figure 23.	Chi-squared test for significance of variation in freq of verbless sentences.....	328
Figure 24.	Utterances by type (normalized per 100 utterances)	331
Figure 25.	Chi-squared test for significance of variation in frequency of ellipsis	332
Figure 26.	Relative frequency of verbless utterances by type.....	335
Figure 27.	Average sentence length per language	385
Figure 28.	Average verbless sentence length per text	386
Figure 29.	Average sentence length per source-type.....	386
Figure 30.	Average sentence length per genre	387

Figure 31.	Russian verbless sentences in originals and translations.....	391
Figure 32.	English verbless sentences in originals and translations.....	392
Figure 33.	Contrastive perspective on originals and translations.....	393
Figure 34.	Russian verbless sentences across genre.....	396
Figure 35.	English verbless sentences across genre.....	396
Figure 36.	Contrastive perspective on genre.....	397
Figure 37.	Situation type and viewpoint type annotation.....	416
Figure 38.	English verbal correspondences by category (abs. fq.).....	420
Figure 39.	Russian verbal correspondences by category (abs. fq.).....	420
Figure 39.	Russian verbal correspondences by category (abs. fq.).....	421
Figure 40.	English verbal correspondences by lemma (abs. fq.).....	423
Figure 41.	Russian verbal correspondences by lemma (abs. fq.).....	423
Figure 42.	Russian verbless sentences and their verbal correspondences.....	427
Figure 43.	English verbless sentences and their verbal correspondences.....	427
Figure 44.	Speech act of verbless questions.....	429
Figure 45.	Speech act of English verbal correspondences (avg.).....	432
Figure 46.	Trameur: Frame & Thread.....	446
Figure 47.	Trameur: User-Controlled Multilayered Annotation – search and edit.....	447
Figure 48.	Trameur: User-Controlled Multilayered Annotation – concordance line.....	448
Figure 49.	Trameur: Projection of annotations onto the Section Map.....	449
Figure 50.	Trameur: Multiple Correspondence Bi-Text Alignment.....	450
Figure 51.	Trameur: The new Section Annotation feature.....	451
Figure 52.	Trameur: Concordance on the Sentence Level and Displaced Alignment.....	452
Figure 53.	Trameur: Summary parameters of the present corpus.....	453
Figure 54.	English morphosyntactic specifications.....	456
Figure 55.	Russian morphosyntactic specifications.....	459
Figure 56.	Contrastive search.....	462
Figure 57.	Contrastive alignment in-context display.....	463
Figure 58.	English forms.....	467
Figure 59.	Russian Forms.....	468
Figure 60.	English Lemmas.....	469
Figure 61.	Russian Lemmas.....	470
Figure 62.	English Morphosyntactic Categories.....	471
Figure 63.	Russian Morphosyntactic Categories.....	472
Figure 64.	English form N-Grams.....	475
Figure 65.	English Lemma N-Grams.....	476
Figure 66.	English Morphosyntactic Category N-Grams.....	477
Figure 67.	Russian Form N-Grams.....	478
Figure 68.	Russian Lemma N-Grams.....	479
Figure 69.	Russian Morphosyntactic Category N-Grams.....	480

LIST OF TABLES

Table 1.	Automatic Extraction Accuracy	260
Table 2.	Present Parallel-Text Selection Criteria	294
Table 3.	Basic Metadata and Summary Statistics	304
Table 4.	Verbless sentences per text	326
Table 5.	Summary of sub-corpus used for contrastive ellipsis analysis	330
Table 6.	Absolute frequency of utterances by type	330
Table 7.	Absolute frequency of verbless utterances by type	331
Table 8.	Absolute frequency of ellipsis by type per text	334
Table 9.	Absolute frequency of matched ellipsis by type	334
Table 10.	Semantic classification schema	340
Table 11.	Agreement and Disagreement	343
Table 12.	Disagreement annotation: Russian lemma 'нет' (<i>net</i>)	346
Table 13.	Expressiveness	350
Table 14.	Quantification	351
Table 15.	Formality and Informality	356
Table 16.	Immediacy and Deixis	357
Table 17.	Questioning	359
Table 18.	Emphasis	363
Table 19.	Referent Specificity	368
Table 20.	Sentence Length	384
Table 21.	Type-token values for verbless sentence subpart	387
Table 22.	Type-token values for full corpus per text	388
Table 23.	Situation type and viewpoint aspect implicated by verbless sentences	417
Table 24.	Verbalization of utterances and questions (abs. fq.)	428

REFERENCES

- Aijmer, Karin, & Rühlemann, Christoph (Eds.). (2015). *Corpus Pragmatics: A Handbook*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139057493>
- Aksakov [Аксаков], Konstantin Sergeevič [Константин Сергеевич]. (1855). *О Русских Глаголах [On Russian Verbs]*. Типография Л. Степановой.
- Alston, William. (2000). *Illocutionary Acts and Sentence Meaning*. Cornell University Press. <https://doi.org/10.7591/9781501700422>
- Anthony, Laurence. (2012). A critical look at software tools in corpus linguistics. *Linguistic Research*, 30(2), 141–161. <https://doi.org/10.17250/khisli.30.2.201308.001>
- Antonova [Антонова], Irina Anatol'evna [Ирина Анатольевна]. (2004). О терминах “категория состояния” и “предикативы”: Предикативы как предикат для выражения эмоционального состояния/отношения [About the terminology “Category of State” and “Predicatives”: Predicatives as predicates for the expression of emotional states/relationships]. *Язык, Сознание, Коммуникация*, 26, 128–133. https://www.philol.msu.ru/~slavphil/books/jsk_26_10antonova.pdf

- Apresjan [Апресян], Jurij Derenikovič [Юрий Дереникович]. (1985). Синтаксические признаки лексем [Syntactic features of lexemes]. *Russian Linguistics*, 9, 289–317.
<https://doi.org/10.1007/BF03380065>
- Apresjan [Апресян], Jurij Derenikovič [Юрий Дереникович]. (1986). Типы синтаксического отсутствия в русском языке [Types of syntactic absence in Russian]. In *Тезисы Краевой Научной Конференции “Проблемы Семантики Предложения: Выраженный и Невыраженный Смысл”* (pp. 111–114). Красноярский Государственный Педагогический Институт.
- Apresjan [Апресян], Valentina Jur’evna [Валентина Юрьевна]. (2017). Русские посессивные конструкции с нулевым и выраженным глаголом: правила и ошибки [Russian Possessive Constructions with overt “be” and Zero Copula: Rules and Errors]. *Русский Язык в Научном Освещении*, 1(33), 86–115.
<https://rjano.ruslang.ru/ru/archive/2017-1/86-115>
- Ariel, Mira. (2001). Accessibility theory: An overview. In Ted Sanders, Joost Schilperoord, & Wilbert Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (pp. 29–88). John Benjamins Publishing Company.
<https://doi.org/10.1075/hcp.8.04ari>
- Aristotle. (1989). *Prior Analytics* (Robin Smith, Trans.). Hackett Publishing Company, Inc. (Original work published ca. 350 B.C.)
- Arnauld, Antoine, & Lancelot, Claude. (1660). *Grammaire Générale et Raisonnée*. Republications Paulet.
- Arnauld, Antoine, & Nicole, Pierre. (1662). *La Logique ou L’Art de Penser*. Flammarion.
- Babajceva [Бабайцева], Vera Vasil’evna [Вера Васильевна], & Maksimov [Максимов], Leonard Ur’evič [Леонард Юрьевич]. (1981). *Современный Русский Язык: Часть*

III - Синтаксис, Пунктуация [Modern Russian Language: Part III - Syntax, Punctuation]. Просвещение.

Baker, Mona. (1993). Corpus linguistics and translation studies: Implications and applications. In Mona Baker, Gill Francis, & Elena Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233–250). John Benjamins Publishing Company. <https://doi.org/10.1075/z.64.15bak>

Baker, Mona. (1996). Corpus-based translation studies: The challenges that lie ahead. In Harold Somers (Ed.), *Terminology, LSP & Translation: Studies in Language Engineering in Honour of Juan C. Sager*. John Benjamins Publishing Company. <https://doi.org/10.1075/btl.18.17bak>

Bakhtin [Бахтин], Mikhail [Михаил Михайлович]. (2002). Проблемы Поэтики Достоевского [Problems of Dostoevsky's Poetics]. In Sergej Georgievič [Сергей Георгиевич] Воџаров [Бочаров] & Ljudmila Arčilovna [Людмила Арчиловна] Gogotišvili [Гоготишвили] (Eds.), *М. М. Бахтин - Собрание Сочинений в Семи Томах* (Vol. 6). Русские Словари: Языки Славянской Культуры. (Original work published in 1963). https://www.fedordostoevsky.ru/pdf/bakhtin_ppd.pdf

Bakhtin, Mikhail. (1984). *Problems of Dostoevsky's Poetics* (Caryl Emerson, Trans.). University of Minnesota Press. (Original work "Проблемы Поэтики Достоевского" published in 1963). https://monoskop.org/images/1/1d/Bakhtin_Mikhail_Problems_of_Dostoevskys_Poetics_1984.pdf

Balvet, Antonio, Barque, Lucie, Condette, Marie-Hélène, Haas, Pauline, Huyghe, Richard, Marín, Rafael, & Merlo, Aurélie. (2011). Nomage: An electronic lexicon of French deverbal nouns based on a semantically annotated corpus. In Benoît Sagot (Ed.),

Proceedings of WoLeR 2011 at ESSLLI, International Workshop on LEXical Resources, August 1-5, 2011, Ljubljana, Slovenia (pp. 8–15). <https://hal.archives-ouvertes.fr/halshs-01078047>

Barton, Ellen, & Progovac, Ljiljana. (2005). Nonsententials in Minimalism. In Reinaldo Elugardo & Robert J Stainton (Eds.), *Ellipsis and Non-Sentential Speech* (pp. 71–94). Springer. https://doi.org/10.1007/1-4020-2301-4_4

Barxudarov [Бархударов], Leonid Stepanovič [Леонид Степанович]. (1966). *Структура Простого Предложения Современного Английского Языка [The Structure of a Simple Sentence of Modern English]*. Высшая Школа.

Behr, Irmtraud, François, Jacques, Lacheret-Dujour, Anne, & Lefevre, Florence. (2005a). Introduction. *Syntaxe & Sémantique - Aux Marges de La Prédication*, 6, 7–12. <https://doi.org/10.3917/ss.006.0007>

Behr, Irmtraud, François, Jacques, Lacheret-Dujour, Anne, & Lefevre, Florence (Eds.). (2005b). *Syntaxe & Sémantique: Aux Marges de la Prédication* (Vol. 6). Presses Universitaires de Caen. <https://www.cairn.info/revue-syntaxe-et-semantique-2005-1.htm>

Belošarkova [Белошаркова], Vera Arsen'evna [Вера Арсеньевна], Bryzgunova [Брызгунова], Elena Andreevna [Елена Андреевна], Zemskaja [Земская], Elena Andreevna [Елена Андреевна], Miloslavskij [Милославский], Igor' Grigor'evič [Игорь Григорьевич], Novikov [Новиков], Lev Alekseevič [Лев Алексеевич], & Panov [Панов], Mixail Victorovič [Михаил Викторович]. (1989). *Современный Русский Язык [The Modern Russian Language]* (Vera Arsen'evna [Вера Арсеньевна] Belošarkova [Белошаркова], Ed.; 2nd ed.). Высшая Школа.

- Bencini, Giulia, & Goldberg, Adele. (2000). The contribution of argument structure constructions to sentence meaning. *Journal of Memory and Language*, 43, 640–651. <https://doi.org/10.1006/jmla.2000.2757>
- Benson, Morton. (1954). The problem of predicatives in Russian. *Slavic and East European Journal*, 1(4), 284–289. <https://doi.org/10.2307/304306>
- Benveniste, Émile. (1966). La phrase nominale. In Émile Benveniste (Ed.), *Problèmes de Linguistique Générale* (pp. 151–167). Gallimard.
- Benveniste, Émile. (1971). *Problems in General Linguistics* (Mary Elizabeth Meek, Trans.). University of Miami Press. (Original work 'Problèmes de Linguistique Générale' published 1966)
- Biber, Douglas. (1989). A typology of English texts. *Linguistics*, 27, 3–43. <https://doi.org/10.1515/ling.1989.27.1.3>
- Biber, Douglas. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257. <https://doi.org/10.1093/lc/8.4.243>
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey N, Conrad, Susan, & Finegan, Edward. (1999). *The Longman Grammar of Spoken and Written English*. Pearson Education Limited.
- Bird, Steven, Klein, Ewan, & Loper, Edward. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly. <https://www.nltk.org/book/>
- Böhmová, Alena, Hajič, Jan, Hajičová, Eva, & Hladká, Barbora. (2003). The Prague Dependency Treebank. In *Treebanks: Building and Using Parsed Corpora* (Vol. 20, pp. 103–127). Springer. https://doi.org/10.1007/978-94-010-0201-1_7

- Bondarenko, Antonina. (2014). *Verbless Sentences in “The Brothers Karamazov”: A Russian-English Contrastive Study* [Masters Thesis: Master 1 Recherche – Études Anglophones Linguistique Anglaise, Université Paris Diderot - Paris 7]. Thesis Director: Agnès Celle.
- Bondarenko, Antonina. (2015). *Verbless Sentences: A Russian-English Parallel Corpus Study* [Masters Thesis: Master 2 Recherche – Études Anglophones Linguistique Anglaise, Université Paris Diderot - Paris 7]. Thesis Directors: Agnès Celle and Maria Zimina.
- Bondarenko, Antonina. (2018). Predication transformation: A parallel corpus-based study of Russian verbless sentences and their English translations. *Annales Universitatis Mariae Curie-Sklodowska Philologiae: La Linguistique Contrastive*, 36(1), 43–54.
<https://doi.org/10.17951/ff.2018.36.1.43-54>
- Bondarenko, Antonina. (2019). A corpus-based contrastive study of verbless sentences: Quantitative and qualitative perspectives. *Studia Neophilologica*, 91(2), 175–198.
<https://doi.org/10.1080/00393274.2019.1616221>
- Bondarenko, Antonina, & Celle, Agnès. (2019). Traduire l’absence: Les questions à prédicat zéro dans un corpus parallèle russe et anglais. *Des Mots Aux Actes: Sémantique, Sémiotique et Traductologie*, 7, 341–364. <https://doi.org/10.15122/isbn.978-2-406-08745-8.p.0341>
- Bondarenko, Antonina, & Celle, Agnès. (2020). Verbless sentences in L’Étranger: A French-Russian-English contrastive study. In Eric Corre, Danh-Thành Do-Hurinville, & Huy-Linh Dao (Eds.), *The Expression of Tense, Aspect, Modality and Evidentiality in Albert Camus’s “L’Étranger” and Its Translations* (Vol. 35, pp. 325–352). John Benjamins.
<https://doi.org/10.1075/lis.35.17bon>

- Bonnot, Christine. (1999). Pour une définition formelle et fonctionnelle de la notion de thème: Sur l'exemple du russe modern. In Claude Guimier (Ed.), *La Thématization dans les Langues* (pp. 15–31).
- Borissov, Anton, Borissov, Oleg, Borissova, Elena, & Bondarenko, Antonina. (2020). *Pickle Scripts and Contrastive Search & Annotate Interface* [Python, Java]. McMaster University: Department of Physics and Astronomy; Moscow Institute of Physics and Technology; Université de Paris.
- Burnard, Lou. (2007). *Reference Guide for the British National Corpus (XML Edition)*. Oxford University Computing Services. Available online at <http://www.natcorp.ox.ac.uk/docs/URG/>
- Buslaev [Буслаев], Fëdor Ivanovič [Фёдор Иванович]. (1858). *Опыт Исторической Грамматики Русского Языка [Historical Grammar of the Russian Language]*. Москва: Университетская Типография.
- Cappelle, Bert, & Grabar, Natalia. (2016). Towards an n-grammar of English. In Sabine De Knop & Gaëtanelle Gilquin (Eds.), *Applied Construction Grammar* (pp. 271–302). De Gruyter Mouton. <https://doi.org/10.1515/9783110458268-011>
- Castagnoli, Luca, & Di Lascio, Ermelinda Valentina. (2012). Ancient philosophy of language. In Gillian Russell & Dalia Graff Fara (Eds.), *The Routledge Companion to Philosophy of Language* (pp. 811–826). Routledge. 10.4324/9780203206966.ch7_1
- Celle, Agnès. (2018). Questions as indirect speech acts in surprise contexts in English. In Dalila Ayoun, Agnès Celle, & Laure Lansari (Eds.), *Tense, Aspect, Modality, Evidentiality: Crosslinguistic Perspectives* (pp. 213–238). John Benjamins. <https://doi.org/10.1075/slcs.197.10cel>

- Celle, Agnès, Jugnet, Anne, Lansari, Laure, & Peterson, Tyler. (2019). Interrogatives in surprise contexts in English. In Natalie Depraz & Agnès Celle (Eds.), *Surprise at the Intersection of Phenomenology and Linguistics*. John Benjamins.
<https://doi.org/10.1075/ceb.11.07cel>
- Chafe, Wallace. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N Li (Ed.), *Subject and Topic* (pp. 25–56). Academic Press.
- Chauvin, Catherine. (2009). Enoncés sans sujet et / ou sans verbe en anglais: prédication et processus de récupération contextuelle. *Faits de Langues: La Prédication*, 31–32, 99–108. <https://doi.org/10.1163/19589514-31-32-01-900000011>
- Chomsky, Noam. (1986). *Knowledge of Language: Its Nature, Origin and Use*. Praeger.
- Chomsky, Noam. (1989). Some notes on economy of derivation and representation. *MIT Working Papers in Linguistics*, 10, 43–74.
- Chomsky, Noam. (1995). *The Minimalist Program*. MIT Press.
- Chomsky, Noam. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511811937>
- Chvany, Chatherine. (1975). *On the Syntax of BE-Sentences in Russian*. Slavica Publishers.
- Cohen, S Marc. (2008). Grice: “Logic and Conversation.” *University of Washington Philosophy of Language Lectures*, 1–13.
- Comrie, Bernard, Haspelmath, Martin, & Balthasar, Bickel (Eds.). (2015). *The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses*. Department of Linguistics of the Max Plank Institute for Evolutionary Anthropology & Department of Linguistics of the University of Leipzig.
<http://eva.mpg.de/lingua/resources/glossing-rules.php>

- Cornish, Francis. (2009). L'Absence de prédication, le topique et le focus: Le cas des phrases "thétiques." *Faits de Langues: La Prédication*, 31–32, 121–131.
<https://doi.org/10.1163/19589514-31-32-01-900000013>
- Corre, Eric. (2009). *De l'Aspect Sémantique à la Structure de l'Événement: Les Verbes Anglais et Russes*. Presses Sorbonne Nouvelle. <http://books.openedition.org/psn/3331>
- Creissels, Denis. (1993). Les notions primitives de la syntaxe: Réflexions sur l'articulation logique entre les notions sur lesquelles se fondent les descriptions grammaticales. *Lidil*, 8, 51–83.
- Creissels, Denis. (1995). *Éléments de Syntaxe Générale*. Presses Universitaires de France-PUF.
- Creissels, Denis. (2004). Les notions de base de l'analyse syntaxique. In *Cours de Syntaxe Générale* (pp. 1–12).
https://is.muni.cz/el/1421/podzim2012/FJIIA103/um/35710539/Chapitre_1_-_Les_notions_de_base_de_l_analyse_syntaxique__D._Creissels_.pdf
- Cresti, Emanuela, & Moneglia, Massimo (Eds.). (2005). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. John Benjamins Publishing Company.
<https://doi.org/10.1075/scl.15>
- Cresti, Emanuela, & Moneglia, Massimo. (2007). C-ORAL-ROM: Comparing Romance languages in spontaneous speech corpora. In Thaïs Cristófaró Silva & Heliana Mello (Eds.), *Conferências do V Congresso Internacional da Associação Brasileira de Lingüística* (pp. 211–228). Universidade Federal de Minas Gerais.
- Cysouw, Michael, & Wälchli, Bernhard. (2007). Parallel texts: Using translation equivalents in linguistic typology. *STUF-Sprachtypologie Und Universalienforschung*, 60(2), 95–99. <https://doi.org/10.1524/stuf.2007.60.2.95>

- Dahl, Östen. (2007). From questionnaires to parallel corpora in typology. *STUF-Sprachtypologie Und Universalienforschung*, 60(2), 172–181.
<https://doi.org/10.1524/stuf.2007.60.2.172>
- Dalrymple, Mary. (2005). Against reconstruction In ellipsis. In Reinaldo Elugardo & Robert J Stainton (Eds.), *Ellipsis and Non-Sentential Speech* (pp. 31–56). Springer.
https://doi.org/10.1007/1-4020-2301-4_2
- Dalrymple, Mary, Shieber, Stuart M, & Pereira, Fernando C N. (1991). Ellipsis and Higher-order Unification. *Linguistics & Philosophy*, 14(4), 399–452.
<https://doi.org/10.1007/BF00630923>
- Danon-Boileau, Laurent, & Morgenstern, Aliyah. (2009). Peut-on parler de prédication dans les premiers énoncés de l'enfant? *Faits de Langues: La Prédication*, 31–32, 57–65.
<https://doi.org/10.1163/19589514-31-32-01-900000007>
- Davies, Mark. (2008). *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>
- Declerck, Renaat, Reed, Susan, & Cappelle, Bert. (2006). *The Grammar of the English Verb Phrase*. Mouton de Gruyter. <https://doi.org/10.1515/9783110199888>
- Divjak, Dagmar, Sharoff, Serge, & Erjavec, Tomaž. (2017). Slavic corpus and computational linguistics. *Journal of Slavic Linguistics*, 25(2), 1–26.
<https://doi.org/10.1353/jsl.2017.0008>
- Dmitrevskij [Дмитревский], Aleksandr Alekseevič [Александр Алексеевич]. (1877a). Практические Заметки О Русском Синтаксисе, 1: Определение предложения [Practical Notes On Russian Syntax: Defining the sentence]. *Филологические Записки*, 3, 1–15.

Dmitrevskij [Дмитревский], Aleksandr Alekseevič [Александр Алексеевич]. (1877b).

Практические Заметки О Русском Синтаксисе, 2: Два ли главных члена в предложении? [Practical Notes On Russian Syntax, 2: Are there two principal members of the sentence?]. *Филологические Записки*, 4, 15–37.

Dmitrevskij [Дмитревский], Aleksandr Alekseevič [Александр Алексеевич]. (1878a). Еще

несколько слов о второстепенности подлежащего [A few more words about the secondary status of the subject]. *Филологические Записки*, 6, 15–27.

Dmitrevskij [Дмитревский], Aleksandr Alekseevič [Александр Алексеевич]. (1878b).

Практические Заметки О Русском Синтаксисе, 3: Сказуемое и его определение [Practical Notes On Russian Syntax, 3: The predicate and its identification].

Филологические Записки, 1, 37–53.

Dolin [Долин], Jurij Trofimovič [Юрий Трофимович]. (2001). К вопросу о

грамматическом статусе односоставных предложений в русском языке [On the grammatical status of one-part sentences in Russian]. *Вестник ОГУ*, 2001(1), 54–

56. <https://cyberleninka.ru/article/n/k-voprosu-o-grammaticheskom-statuse-odnosostavnyh-predlozheniy-v-russkom-yazyke>

Dolin [Долин], Jurij Trofimovič [Юрий Трофимович]. (2008). *Вопросы Теории*

Односоставного Предложения [Theoretical Questions Concerning the One-part Sentence] (2nd ed.). ИПК ГОУ ОГУ.

http://www.orenport.ru/images/doc/1156/Dolin_russ.pdf

Doval, Irene, & Sánchez Nieto, Teresa M. (2019). Parallel Corpora in Focus: An Account of

Current Achievements and Challenges. In Irene Doval & Teresa M. Sánchez Nieto

(Eds.), *Parallel Corpora for Contrastive and Translation Studies: New Resources and*

Applications. John Benjamins Publishing Company.

<https://doi.org/10.1075/scl.90.01dov>

Dowty, David Roach. (1979). *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and In Montague's PTQ*. Reidel.

<https://doi.org/10.1007/978-94-009-9473-7>

Dymarskij [Дымарский], Mikhail Yakovlevič [Михаил Яковлевич]. (2018). А у меня в кармане гвоздь: Нулевая связка или эллипсис сказуемого? [And I have a nail in my pocket: A zero ligament or an ellipsis of the predicate?]. *Мир Русского Слова*, 2018(3), 5–12. <https://doi.org/10.24411/1811-1629-2018-13005>

Elugardo, Reinaldo, & Stainton, Robert J. (2003). Grasping objects and contents. In Alex Barber (Ed.), *Epistemology of Language* (pp. 257–302). Oxford University Press.

Elugardo, Reinaldo, & Stainton, Robert J. (2005). Introduction. In Reinaldo Elugardo & Robert J Stainton (Eds.), *Ellipsis and Non-Sentential Speech* (pp. 1–26). Springer. https://doi.org/10.1007/1-4020-2301-4_1

Ferguth, Johan, Jouannet, Aurélie, Zamiti, Asma, Wu, Yunhe, Li, Jia, Bondarenko, Antonina, Nouvel, Damien, & Valette, Mathieu. (2015). Proposition méthodologique pour la détection automatique de Community Manager: Étude multilingue sur un corpus relatif à la Junk Food [Methodological proposal for automatic detection of Community Manager: Multilingual study based on a junk food corpus]. *Actes de La 22ème Conférence Sur Le Traitement Automatique de Langues Naturelles, Caen, France, 22-25 Juin 2015*, 1–7. <https://aclanthology.org/2015.jeptalnrecital-court.46>

Fernández, Raquel. (2006). *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use* [Doctoral Dissertation, Department of Computer Science, King's College

London, University of London].

https://staff.fnwi.uva.nl/r.fernandezrovira/papers/RaquelFernandez_PhD.pdf

Fernández, Raquel, & Ginzburg, Jonathan. (2002). Non-sentential utterances: A corpus study. *Traitement Automatique de Langues*, 43(2), 13–42.

[https://kclpure.kcl.ac.uk/portal/en/publications/nonsentential-utterances\(11214e78-4f44-4544-b6ad-b17efbfc8ab5\).html](https://kclpure.kcl.ac.uk/portal/en/publications/nonsentential-utterances(11214e78-4f44-4544-b6ad-b17efbfc8ab5).html)

Fernández, Raquel, Ginzburg, Jonathan, & Lappin, Shalom. (2007). Classifying non-sentential utterances in dialogue: A machine learning approach. *Association for Computational Linguistics*, 33(3), 397–427. <https://doi.org/10.1162/coli.2007.33.3.397>

Feuillet, Jack. (2009). La notion de prédicat est-elle nécessaire dans certains types d'énoncés? *Faits de Langues: La Prédication*, 31–32, 133–144.

<https://doi.org/10.1163/19589514-31-32-01-900000014>

Fleury, Serge. (2013). Le Trameur: Propositions de description et d'implémentation des objets textométriques. *Université Paris 3 Online Publication*, 1–53.

<http://www.tal.univ-paris3.fr/trameur/trameur-propositions-definITIONS-objets-textometriques.pdf>

Fleury, Serge. (2019a). *Le Trameur (Version 12.174, Full Version: Windows XP 32bits) [Software]*. Université Sorbonne Nouvelle-Paris 3: SYLED-CLA2T. Available from <http://www.tal.univ-paris3.fr/trameur>

Fleury, Serge. (2019b). Trameur 12.170: Mise à Jour - Annotation des Sections. *Le Trameur: Documentation Complémentaire*, 2019(June), 1–3. <http://www.tal.univ-paris3.fr/trameur/trameur-annotation-section.pdf>

Fleury, Serge, & Zimina, Maria. (2014). Trameur: A framework for annotated text corpora exploration. In Lamia Tounsi & Rafal Rak (Eds.), *Proceedings of COLING 2014, the*

25th International Conference on Computational Linguistics: System Demonstrations

(pp. 57–61). Dublin City University and Association for Computational Linguistics.

<http://www.aclweb.org/anthology/C14-2013.pdf>

Fougeron, Irina. (1986). Analyse prosodique de la phrase russe: L'organisation du message dans les phrases assertives à un, deux ou trois composants. *Revue Des Études Slaves*, LVIII(4), 661–665. https://www.persee.fr/doc/slave_0080-2557_1986_num_58_4_5593

Fougeron, Irina. (1988). Accent “de phrase”, accent “logique”, et verbe noyau du rhème. *Bulletin de La Société Linguistique de Paris*, LXXXIII(1), 163–188. <https://doi.org/10.2143/BSL.83.1.2013660>

Fougeron, Irina. (1984). La paradigme des phrases à deux composants. *IVème Colloque de Linguistique Russe, Institut d'Études Slaves*, 139–151.

Frege, Gottlob. (1967). Begriffsschrift: A Formula Language, Modeled Upon that of Arithmetic, for Pure Thought. In Jean Van Heijenoort (Ed.), & Stefan Bauer-Mengelberg (Trans.), *From Frege to Gödel: A Source Book in Mathematical Logic*. Harvard University Press. (Original work “Begriffsschrift, eine der Arithmetischen nachgebildete Formelsprache des reinen Denkens” published in 1879). <https://doi.org/10.4159/harvard.9780674864603.c2>

Frolova [Фролова], Tat'jana [Татьяна]. (2012). Описание синтаксического отсутствия в корпусах СинТагРус и ХАНКО [Description of syntactic absence in the SinTagRus and Xanko corpora]. *Институт Проблем Передачи Информации им. А. А. Харкевича РАН*, 437–442. <http://itas2012.iitp.ru/pdf/1569601687.pdf>

- Gak [Гак], Vladimir Grigor'evič [Владимир Григорьевич]. (1975). *Русский Язык в Сопоставлении с Французским [Contrastive Study of Russian and French]*. Москва: Русский Язык.
- Gak, Vladimir Grigor'evič. (1994). Traduire: Art ou science? *Équivalences*, 24(2), 5–18.
<https://doi.org/10.3406/equiv.1994.1184>
- Galkina-Fedoruk [Галкина-Федорук], Evdokija Miĥajlovna [Евдокия Михайловна]. (1958). *Безличные Предложения в Современном Русском Языке [Impersonal sentences in modern Russian language]*. Издательство Московского Университета.
- Garcia-Marchena, Oscar. (2016). Spanish verbless clauses and fragments: A corpus analysis. In Antonio Moreno Ortiz & Chantal Pérez-Hernández (Eds.), *EPiC Series in Language and Linguistics: CILC 2016 8th International Conference on Corpus Linguistics* (Vol. 1, pp. 130–143). Université de Malaga. <https://easychair.org/publications/open/G9MF>
- Garcia-Marchena, Oscar. (2018). Polar verbless clauses and gapping subordination in Spanish. In Eric Fuß, Marek Konopka, Beata Trawiński, & Ulrich H. Waßner (Eds.), *Grammar and Corpora 2016* (pp. 169–182). Heidelberg University Publishing.
<https://library.oapen.org/bitstream/handle/20.500.12657/29526/361-68-81164-3-10-20180607.pdf?sequence=1#page=171>
- Gardiner, Alan Henderson. (1989). *Langage & Acte de Langage: Aux Sources de la Pragmatique* (C Douay, Trans.). Presses Universitaires de Lille. (Original work published in 1932)
- Garside, Roger G, Leech, Geoffrey N, & McEnery, Anthony (Eds.). (1997). *Corpus Annotation: Linguistic Information from Computer Text Corpora* (Longman).
- Gasparov, Boris. (1995). La linguistique slavophile. *Histoire Epistémologie Langage*, 17(2), 125–145. <https://doi.org/10.3406/hel.1995.2415>

- Gécseg, Zsuzsanna. (2011). Être ou ne pas être: Prédication averbale et localisation spacio-temporelle en hongrois. In Florence Lefeuve & Irmtraud Behr (Eds.), *Les Énoncés Averbaux Autonomes Entre Grammaire et Discours* (pp. 31–46). Ophrys.
- Gellerstam, Martin. (1986). Translationese in Swedish novels translated from English. In Lars Wollin & Hans Lindquist (Eds.), *Translation Studies in Scandinavia* (pp. 88–95). CWK Gleerup.
- Gellerstam, Martin. (2005). Fingerprints in translation. In Gunilla Anderman & Margaret Rogers (Eds.), *In and Out of English: For Better, For Worse?* (pp. 201–213). Multilingual Matters. <https://doi.org/10.21832/9781853597893-016>
- Gilquin, Gaëtanelle. (2002). Automatic retrieval of syntactic structures: The quest for the Holy Grail. *International Journal of Corpus Linguistics*, 7(2), 183–214. <https://doi.org/10.1075/ijcl.7.2.03gil>
- Goldberg, Adele. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago University Press.
- Goldberg, Adele. (1997). The relationships between verbs and constructions. In Marjolijū Verspoor, Kee Dong Lee, & Eve Sweetser (Eds.), *Lexical and Syntactical Constructions and the Construction of Meaning: Proceedings of the Bi-Annual ICLA Meeting in Albuquerque, July 1995* (Vol. 150, pp. 383–398). John Benjamins. <https://doi.org/10.1075/cilt.150.27gol>
- Goldberg, Adele. (2003). Constructions: A new theoretical approach to language. *TRENDS in Cognitive Sciences*, 7(5), 219–224. [https://doi.org/10.1016/s1364-6613\(03\)00080-9](https://doi.org/10.1016/s1364-6613(03)00080-9)
- Goldberg, Adele, & Casenhiser, Devin. (2006). English constructions. In Bas Aarts & April McMahon (Eds.), *The Handbook of English Linguistics* (pp. 343–355). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470753002.ch15>

- Goldberg, Adele, & Perek, Florent. (2019). Ellipsis in construction grammar. In Jeroen Van Craenenbroeck & Tanja Temmerman (Eds.), *The Oxford Handbook of Ellipsis* (pp. 1–21). Oxford Handbooks Online.
<https://doi.org/10.1093/oxfordhb/9780198712398.013.8>
- Gournay, Lucie, & Merle, Jean-Marie (Eds.). (2004). *Contrastes: Mélanges Offerts à Jacqueline Guillemin-Flescher*. Ophrys.
- Goyvaerts, Jan. (2015). *Regular Expressions Dot Info: The premier website about regular expressions*. Retrieved from <http://www.regular-expressions.info>
- Grandy, Richard E, & Warner, Richard. (2020). Paul Grice. In Edward N Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2020).
<https://plato.stanford.edu/archives/sum2020/entries/grice/>
- Green, Mitchell S. (2017). Assertion. In *Oxford Handbooks Online* (pp. 1–23). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199935314.013.8>
- Grice, Herbert Paul. (1975). Logic and conversation. In Peter Cole & Jerry Morgan (Eds.), *Syntax and Semantics 3: Speech Acts* (pp. 41–58). Academic Press. (Original work published 1967)
- Grice, Herbert Paul. (1989). *Studies in the Way of Words*. Harvard University Press.
- Guillemin-Flescher, Jacqueline. (1981). *Syntaxe Comparée du Français et de l'Anglais: Problèmes de Traduction*. Ophrys.
- Guillemin-Flescher, Jacqueline. (1999). Qualification and point of view. In Akio Kamio & Ken-ichi Takami (Eds.), *Function and Structure: In Honour of Susumu Kuno* (pp. 159–190). John Benjamins. <https://doi.org/10.1075/pbns.59.09gui>
- Guillemin-Flescher, Jacqueline. (2003). Théoriser la traduction. *Revue Française de Linguistique Appliquée*, 8(2), 7–18. <https://doi.org/10.3917/rfla.082.07>

- Guillemin-Flescher, Jacqueline. (2004). Les énoncés averbaux: De l'identification à l'évaluation. *Syntaxe & Sémantique - Aux Marges de La Prédication*, 6, 139–162.
<https://doi.org/10.3917/ss.006.0139>
- Guillemin-Flescher, Jacqueline. (2011). L'énoncé averbal: Répérage et subjectivité. In Florence Lefevre & Irmtraud Behr (Eds.), *Les Énoncés Averbaux Autonomes Entre Grammaire et Discours* (pp. 11–30). Ophrys.
- Guiraud-Weber, Margarite. (2009). La prédication sans sujet: Le cas du russe. *Faits de Langues: La Prédication*, 31–32, 77–86. <https://doi.org/10.1163/19589514-31-32-01-900000009>
- Gvozdev [Гвоздев], Aleksandr Nikolaevič [Александр Николаевич]. (1968). *Современный Русский Литературный Язык: Часть 2 - Синтаксис [Modern Russian Literary Language: Part 2 – Syntax]*. Просвещение.
- Hankamer, Jorge, & Sag, Ivan A. (1976). Deep and surface anaphora. *Linguistic Inquiry*, 7(3), 391–428. <http://www.jstor.org/stable/4177933> .
- Haspelmath, Martin. (1997). *Indefinite Pronouns*. Oxford University Press.
<https://doi.org/10.1093/oso/9780198235606.001.0001>
- Hasselgård, Hilde. (2010). Parallel corpora and contrastive studies. In Richard Xiao (Ed.), *Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS2010)* (pp. 1–30). Lancaster University.
<https://www.lancaster.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/papers/Hasselgard.pdf>
- Hengeveld, Kees. (1992). *Non-Verbal Predication: Theory, Typology, Diachrony*. Mouton de Gruyter. <https://doi.org/10.1515/9783110883282>

- Henkel, Daniel. (2018). De quoi sont représentatifs les corpus de textes traduits au juste? *AFLiCo JET 2018: Corpora and Representativeness, 3-4 May 2018, Nanterre, France.*
- Huddleston, Rodney. (1994). The contrast between interrogatives and questions. *Journal of Linguistics*, 30(2), 411–439. <https://doi.org/10.1017/S0022226700016704>
- Huddleston, Rodney, & Pullum, Geoffrey K. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press. <https://doi.org/10.1017/9781316423530>
- Huyghe, R. (2011). (A)Telicity and the mass-count distinction: The case of French activity nominalizations. *Recherches Linguistiques de Vincennes*, 40, 101–126. <https://doi.org/10.4000/rlv.1952>
- Ildefonse, Frédérique. (1994). Sujet et prédicat chez Platon, Aristote et les Stoïciens. *Histoire Epistémologie Langage*, 3–34. <https://doi.org/10.3406/hel.1994.3391>
- Iomdin [Иомдин], Leonid Lejbovič [Леонид Лейбович]. (2003). Большие Проблемы Малого Синтаксиса [The big problems of small syntax]. Доклады Международной Конференции “Диалог 2003,” 1–6. <https://www.dialog-21.ru/media/2635/iomdinl.pdf>
- Isačenko [Исаченко], Aleksandr Vasil’evič [Александр Васильевич]. (1955). О категории состояния в славянских языках [About the category of state in Slavic languages]. *Вопросы Языкознания*, 6, 48–65.
- Janda, Laura, Koro, Mikhail, & Nessel, Tore. (2020). Constructions, their families and their neighborhoods: the case of durak durakom “a fool times two” [Конструкции, их семьи и соседи: ‘Дурак дураком’]. *Russ Linguist*, 44, 109–127. <https://doi.org/10.1007/s11185-020-09225-y>
- Kahane, Sylvain, & Osborne, Timothy. (2015). Translator’s introduction. In Lucien Tesnière, *Elements of Structural Syntax*. John Benjamins. <https://doi.org/10.1075/z.185.002int>

- Kettunen, Kimmo. (2014). Can Type-Token Ratio be used to show morphological complexity of languages? *Journal of Quantitative Linguistics*, 21(3), 223–245.
<https://doi.org/10.1080/09296174.2014.911506>
- Kim, Jong-Bok, & Abeillé, Anne. (2019). Why-Stripping in English: A corpus-based perspective. *Linguistic Research*, 36(3), 365–387.
<https://doi.org/10.17250/khisli.36.3.201912.002>
- Kim, Jong-Bok, Hee, Kyung, Kim, Jungsook, & Abeillé, Anne. (forthcoming). Aggressively non-D-linked wh-phrases and ellipsis: A direct interpretation Approach. *Journal of Linguistics*, 1–6.
- Kirillova [Кириллова], Anna Vladimirovna [Анна Владимировна]. (2015). Эллиптические Предложения в Русском и Английском Языках [Elliptical Sentences in Russian and English]. *Вестник Ленинградского Государственного Университета Им. А. С. Пушкина*, 163–168. <https://cyberleninka.ru/article/n/ellipticheskie-predlozheniya-v-russkom-i-angliyskom-yazykah>
- Klassovskij [Классовский], Vladimir Ignat'evič [Владимир Игнатьевич]. (1870). *Нерешенные Вопросы в Грамматике [Unresolved Questions in Grammar]*. Типография М. О. Вольфа.
- Kohnen, Thomas. (2015). Speech acts: A diachronic perspective. In Karin Aijmer & Christoph Rühlemann (Eds.), *Corpus Pragmatics: A Handbook* (pp. 52–83). Cambridge University Press. <https://doi.org/10.1017/CBO9781139057493.004>
- Korotev [Копотев], Mikhail Vjačeslavovič [Михаил Вячеславович], & Gurin [Гурин], Grigorij Borisovič [Григорий Борисович]. (2007). Разметка синтаксической неполноты в корпусе [The marking of syntactic incompleteness in a corpus].

Труды международной конференции “Диалог 2007,” 307–310.

<https://www.dialog-21.ru/digests/dialog2007/materials/pdf/46.pdf>

Kopotev, Mikhail. (1999). *Безглагольные Предложения В Истории Русского Языка* [Verbless Sentences in the History of the Russian Language]. [Doctoral Dissertation, Петрозаводский Государственный Университет].

<https://cheloveknauka.com/bezglagolnye-predlozheniya-v-istorii-russkogo-yazyka>

Kopotev, Mikhail. (2007a). What a difference a verb makes! Russian and Finnish verbless sentences. In Marja Nenonen & Sinikka Niemi (Eds.), *Collocations and Idioms 1: Papers from the First Nordic Conference on Syntactic Freezes, May 19-20, 2006* (pp. 177–192). University of Joensuu.

https://www.mv.helsinki.fi/home/kopotev/what_a_difference.pdf

Kopotev, Mikhail. (2007b). Where Russian syntactic zeros start: Approaching Finnish? *Slavica Helsingiensia*, 32, 116–137. <https://blogs.helsinki.fi/slavica-helsingiensia/files/2019/11/sh32-kopotev.pdf>

Kopotev, Mikhail. (2015). Reconstruction and idiomaticity: The origin of Russian verbless clauses reconsidered. *Folia Linguistica Historica*, 36(1), 219–243.

<https://doi.org/10.1515/flih-2015-0007>

Kopotev, Mikhail, Mustajoki, Arto, & Bonch-Osmolovskaya, Anastasia. (2021). Corpora in text-based Russian studies. In Daria Gritsenko, Mariëlle Wijermars, & Mikhail Kopotev (Eds.), *The Palgrave Handbook of Digital Russia Studies* (pp. 299–317).

https://doi.org/10.1007/978-3-030-42855-6_17

Kornilov [Корнилов], Nikolaj Viktorovič [Николай Викторович]. (2015). Алексей Афанасьевич Дмитриевский или Александр Алексеевич Дмитриевский? [Aleksej Afanas’evič Dmitrievskij or Aleksandr Alekseevič Dmitrevskij?]. *Владимирский*

Государственный Университет Имени Александра Григорьевича и Николая Григорьевича Столетовых, 1–6. <https://cyberleninka.ru/article/n/aleksey-afanasievich-dmitrievskiy-ili-aleksandr-alekseevich-dmitrevskiy>

Kornilov [Корнилов], Nikolaj Viktorovič [Николай Викторович]. (2017). Из лингвистического наследия А. А. Дмитриевского: Доминанта предложения [From Linguistic Heritage of A. A. Dmitrevsky: The Dominant of the Sentence]. *Известия ВГПУ*, 4(277), 182–188.

[http://izvestia.vspu.ac.ru/content/izvestia_2017_v277_N4/Izv%20VGPU%202017%20Issue%204%20\(277\)_из%20лингвистич%20182-188.pdf](http://izvestia.vspu.ac.ru/content/izvestia_2017_v277_N4/Izv%20VGPU%202017%20Issue%204%20(277)_из%20лингвистич%20182-188.pdf)

Kozlova [Козлова], Ljubov' Aleksandrovna [Любовь Александровна]. (2019).

Сравнительная Типология Английского и Русского Языков [Contrastive Typology of English and Russian]. АлтГПУ.

Krasnoperueva [Краснопереева], Ekaterina Sergeevna [Екатерина Сергеевна]. (2015). О месте понятия “переводческие универсалии” в исследованиях переводного дискурса [On the place of “translation universals” in translation discourse studies]. *Вестник Челябинского Государственного Университета: Филологические Науки*, 27, 108–113. <https://cyberleninka.ru/article/n/o-meste-ponyatiya-perevodcheskie-universalii-v-issledovaniyah-perevodnogo-diskursa>

Krylov [Крылов], Ivan Andreevič [Иван Андреевич]. (1816). Лебедь, Щука и Рак [The Swan, The Pike and The Crab]. In *Новые Басни [New Fables]* (Vol. 4, p. 10). Въ Типографии Императорского Театра. <https://www.prlib.ru/item/848788>

Kuroda, Shige-Yuki. (1965). *Generative Grammatical Studies in the Japanese Language* [Doctoral Dissertation, Massachusetts Institute of Technology]. <https://dspace.mit.edu/handle/1721.1/13006>

- Lambrecht, Knud. (1994). *Information Structure and Sentence Form: Topic, Focus and the Mental Representation of Discourse Referents*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511620607>
- Landolfi, Annamaria, Sammarco, Carmela, & Voghera, Miriam. (2010). Verbless clauses in Italian, Spanish and English: A Treebank annotation. In Sergio Bolasco, Isabella Chiari, & Luca Guiliano (Eds.), *JADT 2010: Proceedings of the 10th International Conference on Statistical Analysis of Textual Data* (pp. 1187–1194). LED Edizioni Universitarie.
http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-1187-1194_066-Landolfi.pdf
- Lebart, Ludovic, & Salem, André. (1994). *Statistique Textuelle*. Dunod.
<http://lexicometrica.univ-paris3.fr/livre/st94/st94-tdm.html>
- Lebart, Ludovic, Salem, André, & Berry, Lisette. (1998). *Exploring Textual Data*. Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-017-1525-6>
- Leech, Geoffrey N. (2004). *Meaning and the English Verb* (3rd ed.). Longman.
<https://doi.org/10.4324/9781315835464>
- Leech, Geoffrey N. (2007). New resources, or just better old ones? The Holy Grail of representativeness. In Marianne Hundt, Nadja Nesselhauf, & Carolin Biewer (Eds.), *Corpus Linguistics and the Web* (pp. 133–149). Rodopi.
https://doi.org/10.1163/9789401203791_009
- Lefevre, Florence. (1999). *La Phrase Averbale en Français*. L'Harmattan.
- Lefevre, Florence. (2001). La phrase averbale en français. *L'Information Grammaticale*, 88, 47–48. <https://doi.org/10.3406/igram.2001.2732>
- Lefevre, Florence. (2007). Le segment averbal comme unité syntaxique textuelle. In Michel Charolles, Nathalie Fournier, Catherine Fuchs, & Florence Lefevre (Eds.), *Parcours*

- de la Phrase: Mélanges en l'Honneur de Pierre Le Goffic* (pp. 143–158). Ophrys.
<https://halshs.archives-ouvertes.fr/halshs-00138297>
- Lefeuve, Florence, & Behr, Irmtraud. (2011). Préface. In *Les Énoncés Averbaux Autonomes Entre Grammaire et Discours* (pp. 1–10). Ophrys.
- Lefeuve, Florence, & Nicolas, David. (2004). La phrase nominale existentielle et la distinction aspectuelle télique/atélique. *Revue de Sémantique et Pragmatique*, 14, 101–119. <https://halshs.archives-ouvertes.fr/halshs-00138090/>
- Leinonen, Marja. (1985). Finnish and Russian as they are spoken. *Scando-Slavica*, 31, 117–144. <https://doi.org/10.1080/00806768508600880>
- Lekant [Лекант], Pavel Aleksandrovič [Павел Александрович]. (1986). *Синтаксис Простого Предложения В Современном Русском Языке [Syntax Of The Simple Sentence In Modern Russian Language]*. Высшая Школа.
- Letuchiy [Летучий], Aleksander [Александр Борисович]. (2013). Свойства нулевой связки в русском языке в сопоставлении со свойствами выраженного глагола [Properties of zero copula in Russian in comparison with properties of non-zero verbs]. *Компьютерная Лингвистика и Интеллектуальные Технологии*, 12, 420–234. <https://publications.hse.ru/en/articles/85645528>
- Letuchiy [Летучий], Aleksander [Александр Борисович]. (2018). Нулевая Связка [The Zero Copula]. In Vladimir Aleksandrovič [Владимир Александрович] Plungjan [Плу́нган], Ekaterina Rolandovna [Екатерина Роландовна] Dobrušina [Добрушина], Natal'ja Markovna [Наталья Марковна] Stojnova [Стойнова], & Ekaterina Vladimirovna [Екатерина Владимировна] Rahilina [Рахилина] (Eds.), *Русская Корпусная Грамматика [Russian Corpus Grammar]* (pp. 1–37). Материалы Для

- Проекта Корпусного Описания Русской Грамматики (<http://rusgram.ru>). На правах рукописи 2018. http://rusgram.ru/Нулевая_связка
- Letuchiy, Alexander. (2015). Russian zero copulas and lexical verbs: Similar or different? *Lingue e Linguaggio*, 2015(2), 233–250. <https://doi.org/10.1418/81750>
- Li, Charles N, & Thompson, Sandra A. (1976). Subject and topic: A new typology of language. In Charles N Li (Ed.), *Subject and Topic* (pp. 457–489). Academic Press.
- Loock, Rudy. (2016). *La Traductologie de Corpus*. Presses Universitaires du Septentrion. <https://doi.org/10.4000/books.septentrion.31714>
- Lyons, John. (1980). *Sémantique Linguistique* (Jacques Durand & Dominique Boulonnais, Trans.). Larousse. (Original work “Semantics I” published 1978). <https://doi.org/10.1017/CBO9781139165693>
- Maillard, Michel. (2009). Y a-t-il prédication sans sujet ni verbe? Approche interlinguistique. *Faits de Langues: La Prédication*, 31–32, 23–31. <https://doi.org/10.1163/19589514-31-32-01-900000004>
- Malmkjaer, Kirsten. (1998). Love thy neighbour: Will parallel corpora endear linguists to translators? *Meta*, 43(4), 534–541. <https://doi.org/10.7202/003545ar>
- Manning, Christopher D, & Schütze, Hinrich. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press. <https://dl.acm.org/doi/10.5555/311445>
- Maslov [Маслов], Jurij Sergeevič [Юрий Сергеевич]. (1984). *Очерки по Аспектологии [Essays on Aspect]*. Издательство Ленинградского Университета.
- Mažara, Jekaterina. (2010). Swiss cheese for lazy speakers: verb omissions in Russian and Czech. In Atle Gronn & Irena Marijanovic (Eds.), *Oslo Studies in Language: Russian in Contrast 2(1)* (pp. 231–242). <https://doi.org/10.5617/osla.116>

- Mažara, Jekaterina. (2011). Missing verbs in colloquial Russian: ambiguity or vagueness? In K B Karl, G Krumbholz, & M Lazar (Eds.), *Beiträge der europäischen Slavistischen Linguistik (POLYSLAV)* (pp. 137–145). Verlag Otto Sagner.
<https://doi.org/10.5167/uzh-81439>
- McEnery, Anthony, & Wilson, Andrew. (2001). *Corpus Linguistics: An Introduction: Vol. Edinburg Textbooks in Empirical Linguistics* (2nd ed.). Edinburgh University Press.
- McEnery, Anthony, & Xiao, Richard. (2008). Parallel and comparable corpora: What is happening? In Gunilla Anderman & Margaret Rogers (Eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 18–31). Multilingual Matters.
<https://doi.org/10.21832/9781853599873-005>
- McEnery, Anthony, Xiao, Richard, & Tono, Yukio. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. Routledge.
- McShane, Marjorie. (2000). Verbal ellipsis in Russian, Polish and Czech. *The Slavic and East European Journal*, 44(2), 195–233. <https://doi.org/10.2307/309950>
- Mel'čuk [Мельчук], Igor' Alkesandrovič [Игорь Александрович]. (1974). О синтаксическом нуле [About the syntactic zero]. In Александр Алексеевич Холодович (Ed.), *Типология Пассивных Конструкций: Диатезы и Залог* (pp. 343–361). Наука.
- Mel'čuk, Igor. (1979). Syntactic, or lexical, zero in natural language. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society*, 224–260.
<https://doi.org/10.3765/bls.v5i0.2167>
- Mel'čuk, Igor. (2019). Russian U Y-A EST'– X constructions. *Русский Язык в Научном Освещении [Russian Language and Linguistic Theory]*, 1, 6–45.
<https://doi.org/10.31912/rjano-2019.1.1>

- Merle, Jean-Marie (Ed.). (2009a). *Faits de Langues: La Prédication* (Vols. 31–32). Ophrys.
<https://brill.com/view/journals/fdl/31-32/1/fdl.31-32.issue-1.xml>
- Merle, Jean-Marie. (2009b). Présentation générale. *Faits de Langues: La Prédication*, 31–32, 5–12. <https://doi.org/10.1163/19589514-31-32-01-900000002>
- Mesropova, Olga. (2004). Between literary and subliterate paradigms: “Skaz” and contemporary Russian “Estrada” comedy. *Canadian Slavonic Papers / Revue Canadienne Des Slavistes*, 46(3–4), 417–433.
<https://doi.org/10.1080/00085006.2004.11092367>
- Milovidov [Миловичев], Grigorij Alekseevič [Григорий Алексеевич]. (1878). Второстепенный ли член подлежащее? Заметка на Заметки г. Дмитревского [Is the subject a secondary element?]. *Филологические Записки*, 5, 14–18.
- Mitkov, Ruslan (Ed.). (2003). *The Oxford Handbook of Computational Linguistics*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199573691.001.0001>
- Montemagni, Simonetta, Barsotti, Francesco, Battista, Marco, Calzolari, Nicoletta, Corazzari, Ornella, Lenci, Alessandro, Zampolli, Antonio, Fanciulli, Francesca, Massetani, Maria, Raffaelli, Remo, Basili, Roberto, Pazienza, Maria Teresa, Saracino, Dario, Zanzotto, Fabio, Mana, Nadia, Pianesi, Fabio, & Delmonte, Rodolfo. (2003). Building the Italian Syntactic-Semantic Treebank. In Anne Abeillé (Ed.), *Treebanks: Building and Using Parsed Corpora* (Vol. 20, pp. 189–210). Springer. https://doi.org/10.1007/978-94-010-0201-1_11
- Moro, Andrea. (1997). Appendix: A brief history of the copula. In *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure* (pp. 248–302). Cambridge University Press. <https://doi.org/10.1017/CBO9780511519956.007>

- Moro, Andrea. (2006). Chapter 18: Copular sentences. In Martin Everaert & Henk Van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 1–23). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470996591.ch18>
- Muller, Claude. (2013). Le prédicat, entre (méta)catégorie et fonction. *Cahiers de Lexicologie*, 1(102), 51–65. <https://doi.org/10.15122/isbn.978-2-8124-1259-2.p.0051>
- Nádvorníková, Olga. (2017). Pièges méthodologiques des corpus parallèles et comment les éviter. *Corela [Online] HS*, 21, 1–28. <https://doi.org/10.4000/corela.4810>
- Novakova, Iva, & Guentchéva, Zlatka. (2008a). Présentation. *Revue de Linguistique et de Didactique Des Langues: Syntaxe et Sémantique Des Prédicats*, 37(June), 5–22. <https://doi.org/10.4000/lidil.2681>
- Novakova, Iva, & Guentchéva, Zlatka (Eds.). (2008b). *Revue de Linguistique et de Didactique des Langues: Syntaxe et Sémantique des Prédicats* (Vol. 37). Open Edition Journals. <https://doi.org/10.4000/lidil.2678>
- Obenauer, Hans-Georg. (1994). *Aspects de la Syntaxe A-barre: Effets d'Intervention et Mouvements des Quantifieurs* [Thèse de doctorat d'état].
- Olohan, Maeve. (2002). Corpus linguistics and translation studies: Interaction and reaction. *Linguistica Antverpiensia*, 1, 419–429. <https://doi.org/10.52034/lanstts.v1i.29>
- Paillard, Denis. (1984). *Énonciation et Détermination en Russe Contemporain*. Institut d'Études Slaves.
- Palmer, Frank Robert. (1974). *The English Verb*. Longman.
- Perry, John, & Blackburn, Simon. (1986). Thought without representation. *Proceedings of the Aristotelian Society: Supplemental Volume, LX*, 263–283. <https://doi.org/10.1093/aristoteliansupp/60.1.137>

- Peřkovskij [Пешковский], Aleksandr Matveevič [Александр Матвеевич]. (1928). *Русский Синтаксис В Научном Освещении [Russin Syntax from an Academic Perspective]*. Государственное Издательство. <https://www.prlib.ru/item/363106>
- Plato. (1921). *Sophist* (Harold N. Fowler, Trans.). Harvard University Press. (Original work published ca. 360 B.C.).
<http://data.perseus.org/citations/urn:cts:greekLit:tlg0059.tlg007.perseus-eng1:216>
- Poreau, Bastien. (2020). *Modalité, Énonciation et Aspect: Analyse de Prédicats et Structures Exprimant l'Obligation et Régissant un Datif en Russe Contemporain* [Doctoral Dissertation, Institut National des Langues et Civilisations Orientales].
<https://halshs.archives-ouvertes.fr/tel-03153684v1>
- Pospelov [Поспелов], Nikolaj Semenovič [Николай Семенович]. (1955). В защиту категории состояния [In defence of the category of state]. *Вопросы Языкознания*, 2, 55–65.
- Ротарова [Потапова], S X [С X]. (2017). *Изучение Односоставных Предложений на Уроках Русского Языка в Школе: На Материале Произведения А. С. Пушкина [The Study of One-part Sentences in Russian Language Classes: On the Basis of A. S. Puřkin's "The Belkin Tales"]*. Вятский Государственный Университет.
- Potebnja [Потеня], Aleksandr Afanas'evič [Александр Афанасьевич]. (1888). *Из Записок По Русской Грамматике [From Notes on Russian Grammar]*. Издание Книжного Магазина Д. Н. Полуехтова.
- Prince, Gerald. (1987). *A Dictionary of Narratology*. University of Nebraska Press.
- Raineri, Sophie, & Debras, Camille (Eds.). (2019). *Cogni Textes: Corpora and Representativeness* (Vol. 19). *Revue de l'Association Française de Linguistique Cognitive*. <https://doi.org/10.4000/cognitextes.1301>

- Rakhilina, Ekaterina. (1999). Aspectual classification of nouns: A case study of Russian. In Abraham Werner & Leonid Kulikov (Eds.), *Tense-aspect, Transitivity and Causativity: Essays in Honour of Vladimir Nedjakov* (pp. 341–350). John Benjamins.
<https://doi.org/10.1075/slcs.50.23rak>
- Ranger, Graham. (2019). Énonciation, corpus et représentativité : Le cas de « along ». *Cogni Textes: Corpora and Representativeness*, 19, 1–15.
<https://doi.org/10.4000/cognitextes.1517>
- Recanati, François. (1987). *Meaning and Force: The Pragmatics of Performative Utterances*. Cambridge University Press.
- Reese, Brian Jon. (2007). *Bias in Questions* [Doctoral Dissertation, University of Texas at Austin]. <http://hdl.handle.net/2152/3280>
- Rudnev [Руднев], Aleksej Georgievič [Алексей Георгиевич]. (1968). *Синтаксис Современного Русского Литературного Языка [The Syntax of Modern Literary Russian]*. Высшая Школа.
- Sag, Ivan Andrew. (1976). *Deletion and Logical Form* [Doctoral Dissertation, Massachusetts Institute of Technology]. <http://hdl.handle.net/1721.1/16401>
- Šapiro [Шапиро], Abram Borisovič [Абрам Борисович]. (1955). Есть ли в русском языке категория состояния как часть речи? [Is there in the Russian language a part of speech called the category of state?]. *Вопросы Языкознания*, 2, 42–54.
- Šaxmatov [Шахматов], Aleksej Aleksandrovič [Алексей Александрович]. (1925). *Синтаксис Русского Языка [Syntax of the Russian Language]*. Российская Государственная Академическая Типография. <https://www.prlib.ru/item/365824>

- Šaxmatov [Шахматов], Aleksej Aleksandrovič [Алексей Александрович]. (2015).
Синтаксис Русского Языка [Syntax of the Russian Language] (2nd ed.).
 Издательство “Флинта.” (Original work published in 1941)
- Ščerba [Щерба], Lev Vladimirovič [Лев Владимирович]. (1928). О частях речи в русском языке [About parts of speech in the Russian language]. *Русская Речь: Новая Серия*, 2. <http://www.ruthenia.ru/apr/textes/sherba/sherba1.htm>
- Schlangen, David. (2005). Towards finding and fixing fragments: Using ML to identify non-sentential utterances and their antecedents in multi-party dialogue. In Kevin Knight (Ed.), *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics* (pp. 247–254). Association for Computational Linguistics.
<https://doi.org/10.3115/1219840.1219871>
- Schmid, Helmut. (1994). Probabilistic part-of-speech tagging using decision trees. *Proceedings of International Conference on New Methods in Language Processing*, 1–9. <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>
- Sedel'nikov [Седельников], Е А [Е А]. (1961). Структура простого предложения с точки зрения синтагматических и парадигматических отношений [The structure of the simple sentence from the perspective of syntagmatic and paradigmatic relations]. *Филологические Науки*, 3, 66–77.
- Selivërstova [Селивёрстова], Ol'ga [Ольга]. (1973). Семантический анализ предикативных притяжательных конструкций с глаголом быть [Semantic analysis of predicative possessive constructions with the verb byt']. *Вопросы Языкознания*, 5, 95–105.

- Sériot, Patrick. (2000). Le combat des termes et des relations (à propos des discussions sur les constructions impersonnelles dans la linguistique en Russie). In Patrick Sériot & Alain Berrendonner (Eds.), *Le Paradoxe du Sujet: Les Propositions Impersonnelles Dans Les Langues Slaves et Romanes, Cahiers de l'ILSL* (Vol. 12, pp. 235–251).
<https://crecleco.seriot.ch/recherche/biblio/00IMPERS/00Impers.html>
- Sériot, Patrick. (2004). L'affaire du petit drame: filiation franco-russe ou communauté de pensée? (Tesnière et Dmitrievskij). *Slavic Occitania*, 17, 93–118.
<https://crecleco.seriot.ch/recherche/biblio/04TESNIERE/txt.html>
- Sharoff, Serge. (2006). Methods and tools for development of the Russian Reference Corpus. In Dawn Archer, Andrew Wilson, & Paul Rayson (Eds.), *Corpus Linguistics Around the World* (pp. 167–180). Brill. https://doi.org/10.1163/9789401202213_014
- Sharoff, Serge, Kopotev, Mikhail, Erjavec, Tomaz, Feldman, Anna, & Divjak, Dagmar. (2008). Designing and evaluating a Russian tagset. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, & Daniel Tapias (Eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)* (pp. 279–285). http://www.lrec-conf.org/proceedings/lrec2008/pdf/78_paper.pdf
- Shibuya, Yoshikata, & Jensen, Kim Ebensgaard. (2015). Mining for constructions in texts using N-gram and network analysis. *Globe: A Journal of Language, Culture and Communication*, 2, 23–54. <https://doi.org/10.5278/ojs.globe.v2i0.1113>
- Sičina [Сичинава], Dmitri Vladimirovič [Дмитрий Владимирович]. (2005). Национальный корпус русского языка: Очерк предыстории [RNC: A historical sketch]. In *Национальный Корпус Русского Языка: 2003—2005* (pp. 21–30). Индрик. <https://ruscorpora.ru/old/sbornik2005/03sitch.pdf>

- Siemund, Peter. (2018). *Speech Acts and Clause Types: English in a Corpus-Linguistic Context*. Oxford University Press.
- Skoblikova [Скобликова], Elena Sergeevna [Елена Сергеевна]. (2006). *Современный Русский Язык: Синтаксис Простого Предложения [Modern Russian Language: Syntax of the Simple Sentence]* (3rd ed.). Флинта: Наука.
- Smith, Carlota. (1997). *The Parameter of Aspect* (2nd ed.). Kluwer.
<https://doi.org/10.1007/978-94-011-5606-6>
- Soames, Scott. (2008). Drawing the line between meaning and implicature - and relating both to assertion. *NOÛS*, 42(3), 440–465. <https://doi.org/10.1111/j.1468-0068.2008.00691.x>
- Soames, Scott. (2012). Propositions. In Gillian Russell & Dalia Graff Fara (Eds.), *The Routledge Companion to Philosophy of Language* (pp. 209–220). Routledge.
https://doi.org/10.4324/9780203206966.ch2_3
- Stainton, Robert J. (1993). *Non-Sentential Assertions* [Doctoral Dissertation, Massachusetts Institute of Technology]. <http://hdl.handle.net/1721.1/12501>
- Stainton, Robert J. (1995). Non-sentential assertions and semantic ellipsis. *Linguistics and Philosophy*, 18, 281–296. <https://doi.org/10.1007/BF00985446>
- Stainton, Robert J. (1997). What assertion is not. *Philosophical Studies*, 85, 57–73.
<https://doi.org/10.1023/A:1017922124403>
- Stainton, Robert J. (2000). The meaning of “sentences.” *NOÛS*, 34(4), 441–454.
<https://doi.org/10.1111/0029-4624.00219>
- Stainton, Robert J. (2004). The pragmatics of non-sentences. In Ward Horn (Ed.), *The Handbook of Pragmatics* (pp. 266–287).
<https://doi.org/10.1002/9780470756959.ch12>

- Stassen, Leon. (2013). Zero copula for predicate nominals. In Matthew S. Dryer & Martin Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/120>
- Stolz, Thomas. (2007a). Harry Potter meets Le Petit Prince: On the usefulness of parallel corpora in crosslinguistic investigations. *STUF - Sprachtypologie Und Universalienforschung: Language Typology and Universals*, 60(2), 100–117. <https://doi.org/10.1524/stuf.2007.60.2.100>
- Stolz, Thomas (Ed.). (2007b). *STUF - Sprachtypologie Und Universalienforschung: Language Typology and Universals: Vol. 60 (2)*. De Gruyter. <https://www.degruyter.com/journal/key/stuf/60/2/html>
- Švedova [Шведова], Natal'ja Ul'evna [Наталья Юльевна], Bryzgunova [Брызгунова], Elena Andreevna [Елена Андреевна], Gabučan [Габучан], K V [К В], Ickovič [Ицкович], V A [В А], Kovtunova [Ковтунова], [И И], Kručinina [Кручинина], I N [И Н], Ljaron [Ляпон], M V [М В], Prijatkina [Прияткина], Alla Fëdorovna [Алла Фёдоровна], & Svjatogor [Святогор], I P [И П]. (1980). *Русская Грамматика: Том II - Синтаксис [Russian Grammar: Volume 2 - Syntax]* (Natal'ja Ul'evna [Наталья Юльевна] Švedova [Шведова], Nina Davidovna [Нина Давидовна] Arutjunova [Арутюнова], Aleksandr Vladimirovič [Александр Владимирович] Bondarko [Бондарко], Valerij Vasil'evič [Валерий Васильевич] Ivanov [Иванов], Vladimir Vladimirovič [Владимир Владимирович] Lopatin [Лопатин], Igor' Stepanovič [Игорь Степанович] Uluxanov [Улуханов], & Fedot Petrovič [Федот Петрович] Filin [Филин], Eds.). Наука. <http://rkiff.philol.msu.ru/wp-content/uploads/2020/05/Русская-грамматика.-Том-II.pdf>

- Taylor, Ann, Marcus, Marcus, & Santorini, Beatrice. (2003). The Penn Treebank: An overview. In Anne Abeillé (Ed.), *Treebanks: Building and Using Parsed Corpora* (Vol. 20, pp. 5–22). Springer. https://doi.org/10.1007/978-94-010-0201-1_1
- Tesnière, Lucien. (1959). *Éléments de Syntaxe Structurale*. Klincksieck.
- Testelec [Тестелец], Jakov Georgievič [Яков Георгиевич]. (2008). Структура предложений с невыраженной связкой в русском языке [The structure of sentences with an unexpressed copula in Russian]. In Aleksandr Vladimirovič [Александр Владимирович] Bondarko [Бондарко], Galina Ivanovna [Галина Ивановна] Kustova [Кустова], & Raisa Iosifovna [Раиса Иосифовна] Rozina [Розина] (Eds.), *Динамические Модели: Слово, Предложение, Текст. Сборник Статей в Честь Е.В. Падучевой [Dynamic Patterns: Word, Sentence, Text. A Collection of Articles in Honour of Elena Padučeva]* (pp. 773–789). Языки Славянских Культур. <https://www.dialog-21.ru/digests/dialog2007/materials/html/Testelelets.htm>
- Thomas, George. (1982). Aspects of the study of Dostoyevsky's vocabulary. *The Modern Language Review*, 77(3), 670–678. <https://doi.org/10.2307/3728074>
- Timberlake, Alan. (2004). *A Reference Grammar of Russian*. Cambridge University Press.
- Тиханова [Тиханова], О V [О В]. (2019). О нулевом сказуемом в номинативных предложениях [About the zero predicate in nominative sentences]. 1–2. <https://doi.org/10.18411/spc-26-03-2019-05>
- Тихонов [Тихонов], Aleksandr Nikolaevič [Александр Николаевич]. (1960). *Категория Состояния в Современном Русском Языке [The Category of State in Modern Russian Language]*. Издательство Самаркандского Государственного Университета.

- Touratier, Christian. (2009). Que faut-il entendre par prédication et prédication seconde? *Faits de Langues: La Prédication*, 31–32, 13–22. <https://doi.org/10.1163/19589514-31-32-01-900000003>
- Toury, Gideon. (1985). A rationale for descriptive translation studies. In Theo Hermans (Ed.), *The Manipulation of Literature: Studies in Literary Translation* (pp. 16–41). Croom Helm.
- Toury, Gideon. (1995). *Descriptive Translation Studies - and Beyond*. John Benjamins Publishing Company. <https://doi.org/10.1075/btl.4>
- Traum, David, & Heeman, Peter. (1997). Utterance units in spoken dialogue. In Elisabeth Maier, Marion Mast, & Susann LuperFoy (Eds.), *Dialogue Processing in Spoken Language Systems* (pp. 125–140). Springer-Verlag. https://doi.org/10.1007/3-540-63175-5_42
- Travniček [Травничек], František [Франтишек]. (1956). Заметки о категории состояния [Notes about the category of state]. *Вопросы Языкознания*, 3, 46–53.
- Trotzke, Andreas. (2017). *The Grammar of Emphasis: From Information Structure to the Expressive Dimension*. De Gruyter Mouton. <https://doi.org/10.1515/9781501505881>
- Trotzke, Andreas, & Turco, Giuseppina. (2015). The grammatical reflexes of emphasis: evidence from German wh-questions. *Lingua*, 168, 37–56.
- Valgina [Валгина], Nina Sergeevna [Нина Сергеевна]. (2000). *Синтаксис Современного Русского Языка [Syntax Of The Modern Russian Language]*. Издательство Агар.
- Van Heijenoort, Jean. (1967). Editor's introduction to Begriffsschrift. In Jean Van Heijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic*. Harvard University Press.

Váradi, Tamás. (2001). The linguistic relevance of corpus linguistics. In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, & Shereen Khoja (Eds.), *Proceedings of the Corpus Linguistics 2001 Conference: Lancaster University UCREL Technical Papers 13* (pp. 587–593).

Vasil'čenko [Васильченко], Tat'jana Valerievna [Татьяна Валериевна]. (2007).

Современные англоязычные переводы романа Ф.М. Достоевского «Братья Карамазовы»: К вопросу о переводческих стратегиях в англоязычной культуре конца XX в [Modern English translations of F.M. Dostoyevsky's "Bratya Karamazovy": regarding the question of translation strategies in English culture at the end of the XX century]. *Вестник Томского Государственного Университета*, 305, 7–10. <https://cyberleninka.ru/article/n/sovremennye-angloyazychnye-perevody-romana-f-m-dostoevskogo-bratya-karamazovy-k-voprosu-o-perevodcheskih-strategiyah-v-angloyazychnoy>

Vendler, Zeno. (1957). Verbs and times. *The Philosophical Review*, 66(2), 143–160.

<https://doi.org/10.2307/2182371>

Verkuyl, Henk J. (1972). *On the Compositional Nature of the Aspects*. Kluwer.

<https://doi.org/10.1007/978-94-017-2478-4>

Vilinbakhova [Вилинбахова], Elena Leonidovna [Елена Леонидовна]. (2016).

Сопоставительные тавтологии в Русском языке [Coordinated tautologies in Russian]. *Вопросы Языкознания*, 2, 61–74.

<https://doi.org/10.31857/S0373658X0000981-8>

Vilinbakhova [Вилинбахова], Elena Leonidovna [Елена Леонидовна], & Koptev [

Коптев], Mikhail Vjačeslavovič [Михаил Вячеславович]. (2017). "X есть X"

значит "X это X"? Ищем ответ в синхронии и диахронии [Does "X est' X" mean 'X

èto X"? Looking for an answer in synchrony and diachrony]. *Вопросы Языкознания*, 2017(3), 110–124. <https://doi.org/10.31857/s0373658x0001003-2>

Vinay, Jean-Paul, & Darbelnet, Jean. (1995). *Comparative Stylistics of French and English: A Methodology for Translation [Stylistique Comparée du Français et de l'Anglais]* (Juan C Sager & M-J Hamel, Trans.). John Benjamins. (Original work published 1977). <https://doi.org/10.1075/btl.11>

Vinogradov [Виноградов], Viktor Vladimirovič [Виктор Владимирович]. (1958).

Проблема сказуемости как основы предложения и углубление противоречий в понимании других членов предложения [The problem of predication as a sentential basis and the expansion of contradictions in the understanding of other elements of the sentence]. In *Из Истории Изучения Русского Синтаксиса: от Ломоносова до Потебни и Фортунатова [Within the History of the Study of Russian Syntax: From Lomonosov to Potebnja and Fortunatov]* (pp. 284–295).

Издательство Московского Университета.

http://crecleco.seriot.ch/textes/Vinogradov58.html#_ftn8

Vinogradov [Виноградов], Viktor Vladimirovič [Виктор Владимирович]. (1986). *Русский Язык: Грамматическое Учение о Слове [The Russian Language: Grammatical Study of the Word]* (3rd ed.). Высшая Школа. (Original work published 1947)

Vinogradov [Виноградов], Viktor Vladimirovič [Виктор Владимирович], & Istrina [Истрина], Evgenija Samsonovna [Евгения Самсоновна]. (1954). *Грамматика Русского Языка: Том 2 - Синтаксис [Grammar of the Russian Language: Volume 2 - Syntax]*. Издательство Академии Наук СССР.

Vishenkova, Anna, & Zevakhina, Natalia. (2019). Wh-Exclamatives with and without predicates in Russian [Бесказуемые и сказуемые восклицательные

конструкции в русском языке]. *Russ Linguist*, 43, 107–125.

<https://doi.org/10.1007/s11185-019-09213-x>

Voghera, Miriam. (2010). The constants of spoken discourse: What evidence for grammar models? In Massimo Pettorino, Antonella Gianninin, Isabella Chiari, & Francesca M. Dovetto (Eds.), *Spoken Communication* (pp. 77–109). Cambridge Scholars Publishing.

Wälchli, Bernhard. (2007). Advantages and disadvantages of using parallel texts in typological investigations. *STUF-Sprachtypologie Und Universalienforschung*, 60(2), 118–134. <https://doi.org/10.1524/stuf.2007.60.2.118>

Weiss, Daniel. (2011a). Безглагольные конструкции русской разговорной речи: Их типология и статус в лингвистическом описании [Verbless constructions in colloquial Russian: Their typology and status in linguistic descriptions]. In Igor' Mihajlovič [Игорь Михайлович] Boguslavskij [Богуславский], Leonid Lejbovič [Леонид Лейбович] Iomdin [Иомдин], & Leonid Petrovič [Леонид Петрович] Krysin [Кысин] (Eds.), *Слово и Язык: Сборник Статей к Восмидесятилетию Академика Ю. Д. Апресяна [Word and Language: A Collection of Papers for the 80th Birthday of Academic Ju. D. Apresjan]* (pp. 138–155). Языки Славянских Культур. <https://doi.org/10.5167/uzh-46347>

Weiss, Daniel. (2013). The lazy speaker and the fascination of emptiness: Colloquial Russian from a typological perspective. In Irina Kor Chahine (Ed.), *Current Studies in Slavic Linguistics*. John Benjamins Publishing Company.

<https://doi.org/10.1075/slcs.146.06wei>

Weiss, Daniel. (2011b). MTT meets construction grammar: The treatment of argument structure. In Igor Boguslavsky & Leo Wanner (Eds.), *Proceedings of the 5th*

- International Conference on Meaning-Text Theory* (pp. 295–305). Recurso Electrónico. <https://doi.org/10.5167/uzh-57786>
- Wellek, René. (1980). Bakhtin's view of Dostoevsky: Polyphony and carnivalesque. *Dostoevsky Studies*, 1, 31–39. <http://sites.utoronto.ca/tsq/DS/01/031.shtml>
- Wertz, Christopher A. (1994). The pedagogical case for predicatives in Russian. *Slavic and East European Journal*, 38(2), 303–318. <https://doi.org/10.2307/308808>
- Williams, Edwin S. (1977). Discourse and logical form. *Linguistic Inquiry*, 8(1), 101–139. <https://www.jstor.org/stable/4177974>
- Xrakovskij [Храковский], Victor Samuilovič [Виктор Самуилович]. (1983). Истоки вербоцентрической концепции предложения в русском языкознании [The origins of the verbo-centric conception of the sentence in Russian linguistics]. *Вопросы Языкознания*, 3, 110–117.
- Zakharov [Захаров], Victor [Виктор], & Bogdanova [Богданова], Svetlana [Светлана]. (2020). *Корпусная Лингвистика [Corpus Linguistics]* (3rd ed.). Издательство Санкт-Петербургского Университета.
- Zakharov, Victor. (2013). Corpora of the Russian language. In Ivan Habernal & Václav Matoušek (Eds.), *Text, Speech, and Dialogue: TSD 2013* (pp. 1–13). Springer. https://doi.org/10.1007/978-3-642-40585-3_1
- Zanettin, Federico. (2011). Translation and corpus design. *Synaps*, 26, 14–23. https://openaccess.nhh.no/nhh-xmlui/bitstream/handle/11250/2394011/Zanettin_26.pdf?sequence=1
- Zanettin, Federico. (2012). *Translation-driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Routledge. <https://doi.org/10.4324/9781315759661>

- Zanettin, Federico. (2013). Corpus methods for descriptive translation studies. *Procedia - Social and Behavioral Sciences*, 95, 20–32.
<https://doi.org/10.1016/j.sbspro.2013.10.618>
- Zanettin, Federico. (2014). Corpora in translation. In Juliane House (Ed.), *Translation: A Multidisciplinary Approach* (pp. 178–199). Palgrave Macmillan.
https://doi.org/10.1057/9781137025487_10
- Zdorenko, Tatiana. (2010). Subject omission in Russian: A study of the Russian National Corpus. In Stefan Th Gries, Stefanie Wulff, & Mark Davies (Eds.), *Corpus-linguistic Applications: Current Studies, New Directions* (Vol. 71, pp. 119–133). Rodopi.
https://doi.org/10.1163/9789042028012_009
- Zimina, Maria. (2004). *Approches Quantitatives de l'Extraction de Ressources Traductionnelles à Partir de Corpus Parallèles* [Doctoral Dissertation, Université Sorbonne Nouvelle-Paris 3]. <https://tel.archives-ouvertes.fr/tel-00008311/document>
- Zimina, Maria. (2005). Bi-text topography and quantitative approaches of parallel text processing. *Corpus Linguistics Conference Series*, 1, 1–15. <https://hal.archives-ouvertes.fr/hal-01224587/>
- Zimina, Maria, & Fleury, Serge. (2015). Perspectives de l'architecture Trame/Cadre pour les alignements multilingues. *Nouvelles Perspectives En Sciences Sociales: Sur Le Thème de l'analyse de Données Textuelles Informatisée*, 11(1), 325–353.
<https://doi.org/10.7202/1035940ar>
- Zimmerling [Циммерлинг], Anton [Антон]. (2018). Предикативы и предикаты состояния в русском языке [Predicatives and predicates of state in the Russian]. *Slavistična Revija*, 66(1), 45–64. https://srl.si/ojs/srl/article/view/COBISS_ID-66614114

CONDENSED TABLE OF CONTENTS

ACKNOWLEDGEMENTS – i

CONTENTS – 1

ABSTRACTS – 9

PRELIMINARIES – 20

PART ZERO: INTRODUCTION

MOTIVATION AND ROAD MAP – 26

PART ONE: THEORETICAL FRAMEWORK

CHAPTER 1: THE PHENOMENON – 40

CHAPTER 2: PREDICATION AND SENTENTIAL STATUS – 42

CHAPTER 3: ELLIPTICAL SENTENCES AND NONSENTENTIALS – 82

CHAPTER 4: VERBLESS SENTENCES – 142

PART TWO: METHODOLOGY

CHAPTER 1: AN INTERDISCIPLINARY APPROACH – 216

CHAPTER 2: THIRST FOR CORPORA – 217

CHAPTER 3: CONTRASTIVE LINGUISTICS – 231

CHAPTER 4: AUTOMATIC PROCESSING – 255

CHAPTER 5: QUANTITATIVE AND QUALITATIVE ANALYSIS – 281

CHAPTER 6: CONCLUSION: METHODOLOGY – 289

PART 3: CORPUS

CHAPTER 1: DESIGN AND SELECTION CRITERIA – 291

CHAPTER 2: COMPOSITION – 299

CHAPTER 3: COMPARABILITY ANALYSIS – 315

PART 4: RESULTS

CHAPTER 1: SYNTACTIC EXPLANATIONS – 323

CHAPTER 2: SEMANTIC FEATURES – 338

CHAPTER 3: TRANSLATION LANGUAGE AND GENRE – 390

CHAPTER 4: INFORMATION STRUCTURE – 400

CHAPTER 5: TEMPORAL AND ASPECTUAL MEANING – 412

CHAPTER 6: INDIRECT SPEECH ACTS IN QUESTIONS – 425

PART 5: CONCLUSION

VERBLESS SENTENCES – 436

APPENDICES – 442

RÉSUMÉ LONG EN FRANÇAIS – 481

LIST OF FIGURES – 509

LIST OF TABLES – 511

REFERENCES – 512

LES PHRASES AVERBALES ET LES PHRASES À PRÉDICAT ZÉRO : ÉTUDE CONTRASTIVE ANGLAIS-RUSSE BASÉE SUR CORPUS

Cette thèse porte sur les structures dans lesquelles le marqueur typique de la phrase – le prédicat verbal – est absent, et convoque des méthodes contrastives, énonciatives ainsi que la linguistique de corpus pour explorer les traits sémantico-pragmatiques des phrases averbales en anglais et en russe. Les résultats montrent que la phrase averbale ne s'explique pas par la productivité syntaxique et militent pour une explication sémantico-pragmatique. Des implications plus larges portent sur la justification phrastique des structures averbales, ainsi que sur le traitement automatique de l'absence.

VERBLESS AND ZERO-PREDICATE SENTENCES: AN ENGLISH AND RUSSIAN CONTRASTIVE CORPUS STUDY

Focusing on structures in which the typical marker of sentential status – the verbal predicate – is absent, the present dissertation takes an interdisciplinary approach that combines contrastive linguistics, corpus methods and enunciative analysis in order to explore the semantic and pragmatic characteristics of verbless sentences in English and Russian. The results show that syntactic ellipsis does not explain the phenomenon of the verbless sentence and reveal a semantico-pragmatic explanation. Implications pertain to the theoretical account of the sentence and the sentential status of verbless structures, as well as the automatic processing of the absence of the verb.

БЕЗГЛАГОЛЬНЫЕ ПРЕДЛОЖЕНИЯ И ПРЕДЛОЖЕНИЯ С НУЛЕВЫМ ГЛАГОЛЬНЫМ СКАЗУЕМОМ: КОНТРАСТИВНОЕ КОРПУСНОЕ ИССЛЕДОВАНИЕ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ

Диссертация посвящена структурам, в которых канонический критерий предложения – глагольное сказуемое отсутствует. Сочетая методы контрастивного, корпусного и дискурсивного анализа, исследуются семантико-прагматические характеристики безглагольных предложений в английском и русском языках. Результаты показывают, что безглагольность предложения не объясняется типологической продуктивностью синтаксического эллипсиса и свидетельствуют в пользу семантико-прагматического объяснения. Последствия данного исследования относятся к теории предложения и обоснования безглагольных структур в качестве предложений, а также к автоматической обработке отсутствия глаголов.