

УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ“ - ШТИП  
ФИЛОЛОШКИ ФАКУЛТЕТ

UDC 81  
UDC 82  
UDC 008



ISSN: 2545-3998  
DOI: 10.46763/palim

# ПАЛИМПСЕСТ

МЕЃУНАРОДНО СПИСАНИЕ ЗА ЛИНГВИСТИЧКИ,  
КНИЖЕВНИ И КУЛТУРОЛОШКИ ИСТРАЖУВАЊА

# PALIMPSEST

INTERNATIONAL JOURNAL FOR LINGUISTIC,  
LITERARY AND CULTURAL RESEARCH

PALMK, VOL X, NO 19, STIP, 2025

ГОД. 10, БР. 19  
ШТИП, 2025

VOL. X, NO 19  
STIP, 2025

# ПАЛИМПСЕСТ

Меѓународно списание за лингвистички, книжевни  
и културолошки истражувања

# PALIMPSEST

International Journal for Linguistic, Literary  
and Cultural Research

Год. 10, Бр. 19  
Штип, 2025

Vol. 10, No 19  
Stip, 2025

PALMK, VOL 10, NO 19, STIP, 2025

DOI: <https://doi.org/10.46763/PALIM251019>

## **ПАЛИМПСЕСТ**

Меѓународно списание за лингвистички, книжевни  
и културолошки истражувања

## **ИЗДАВА**

Универзитет „Гоце Делчев“, Филолошки факултет, Штип

## **ГЛАВЕН И ОДГОВОРЕН УРЕДНИК**

Ранко Младеноски

## **УРЕДУВАЧКИ ОДБОР**

Виктор Фридман, Универзитет во Чикаго, САД  
Толе Белчев, Универзитет „Гоце Делчев“, Македонија  
Нина Даскаловска, Универзитет „Гоце Делчев“, Македонија  
Ала Шешкен, Универзитет Ломоносов, Руска Федерација  
Олга Панкина, НВО Македонски културен центар, Руска Федерација  
Астрид Симоне Хлубик, Универзитет „Крал Михаил I“, Романија  
Алина Андреа Драгоеску Урлика, Универзитет „Крал Михаил I“, Романија  
Сунчана Туксар, Универзитет „Јурај Добрила“ во Пула, Хрватска  
Саша Војковиќ, Универзитет во Загреб, Хрватска  
Шандор Чегледи, Универзитет во Панонија, Унгарија  
Ева Бус, Универзитет во Панонија, Унгарија  
Хусејин Озбај, Универзитет Гази, Република Турција  
Озтурк Емироглу, Универзитет во Варшава, Полска  
Елена Дараданова, Универзитет „Св. Климент Охридски“, Република Бугарија  
Ина Христова, Универзитет „Св. Климент Охридски“, Република Бугарија  
Џозеф Пониах, Национален институт за технологија, Индија  
Сатхарај Венкатесан, Национален институт за технологија, Индија  
Петар Пенда, Универзитет во Бања Лука, Босна и Херцеговина  
Данило Капасо, Универзитет во Бања Лука, Босна и Херцеговина  
Мета Лах, Универзитет во Љубљана, Република Словенија  
Намита Субиото, Универзитет во Љубљана, Република Словенија  
Ана Пеличер-Санчез, Универзитет во Нотингам, Велика Британија  
Мајкл Грини, Универзитет во Нотингам, Велика Британија  
Татјана Гурин, Универзитет во Нови Сад, Република Србија  
Диана Поповиќ, Универзитет во Нови Сад, Република Србија  
Жан Пол Мејер, Универзитет во Стразбур, Република Франција  
Жан Марк Веркруз, Универзитет во Артуа, Република Франција  
Регула Бусин, Швајцарија  
Натале Фиорето, Универзитет во Перуца, Италија  
Оливер Хербст, Универзитет во Вурцбург, Германија  
Шахинда Езат, Универзитет во Каиро, Египет  
Џулијан Чен, Универзитет Куртин, Австралија

## **PALIMPSEST**

International Journal for Linguistic, Literary  
and Cultural Research

## **PUBLISHED BY**

Goce Delchev University, Faculty of Philology, Stip

## **EDITOR-IN-CHIEF**

Ranko Mladenoski

## **EDITORIAL BOARD**

Victor Friedman, University of Chicago, USA  
Tole Belcev, Goce Delchev University, Macedonia  
Nina Daskalovska, Goce Delchev University, Macedonia  
Alla Sheshken, Lomonosov Moscow State University, Russian Federation  
Olga Pankina, NGO Macedonian Cultural Centre, Russian Federation  
Astrid Simone Hlubik, King Michael I University, Romania  
Alina Andreea Dragoescu Urlica, King Michael I University, Romania  
Sunčana Tuksar, Juraj Dobrila University of Pula, Croatia  
Saša Vojković, University of Zagreb, Croatia  
Sándor Czeglédi, University of Pannonia, Hungary  
Éva Bús, University of Pannonia, Hungary  
Husejin Ozbaj, GAZI University, Republic of Turkey  
Öztürk Emiroğlu, University of Warsaw, Poland  
Elena Daradanova, Sofia University “St. Kliment Ohridski”, Republic of Bulgaria  
Ina Hristova, Sofia University “St. Kliment Ohridski”, Republic of Bulgaria  
Joseph Ponniah, National Institute of Technology, India  
Sathyaraj Venkatesan, National Institute of Technology, India  
Petar Penda, University of Banja Luka, Bosnia and Herzegovina  
Danilo Capasso, University of Banja Luka, Bosnia and Herzegovina  
Meta Lah, University of Ljubljana, Republic of Slovenia  
Namita Subiotto, University of Ljubljana, Republic of Slovenia  
Ana Pellicer Sanchez, The University of Nottingham, United Kingdom  
Michael Greaney, Lancaster University, United Kingdom  
Tatjana Durin, University of Novi Sad, Republic of Serbia  
Diana Popovic, University of Novi Sad, Republic of Serbia  
Jean-Paul Meyer, University of Strasbourg, French Republic  
Jean-Marc Vercreysse, Artois University, French Republic  
Regula Busin, Switzerland  
Natale Fioretto, University of Perugia, Italy  
Oliver Herbst, University of Wurzburg, Germany  
Chahinda Ezzat, Cairo University, Egypt  
Julian Chen, Curtin University, Australia

## **РЕДАКЦИСКИ СОВЕТ**

Луси Караниколова-Чочоровска  
Толе Белчев  
Нина Даскаловска  
Билјана Ивановска  
Ева Ѓорѓиевска  
Марија Леонтиќ  
Јована Караникиќ Јосимовска  
Натка Јанкова-Алаѓозовска

## **ЈАЗИЧНО УРЕДУВАЊЕ**

Ранко Младеноски (македонски јазик)  
Весна Продановска (англиски јазик)  
Толе Белчев (руски јазик)  
Билјана Ивановска (германски јазик)  
Марија Леонтиќ (турски јазик)  
Ева Ѓорѓиевска (француски јазик)  
Јована Караникиќ Јосимовска (италијански јазик)

## **ТЕХНИЧКИ УРЕДНИК**

Славе Димитров

## **АДРЕСА**

ПАЛИМПСЕСТ  
РЕДАКЦИСКИ СОВЕТ  
Филолошки факултет  
ул. „Крсте Мисирков“ бр. 10-А  
п. фах 201  
МК-2000 Штип

<http://js.ugd.edu.mk/index/PAL>

Меѓународното научно списание „Палимпсест“ излегува двапати годишно во печатена и во електронска форма на посебна веб-страница на веб-порталот на Универзитетот „Гоце Делчев“ во Штип: <http://js.ugd.edu.mk/index.php/PAL>. Трудовите во списанието се објавуваат на следните јазици: македонски јазик, англиски јазик, германски јазик, француски јазик, руски јазик, турски јазик и италијански јазик.

Трудовите се рецензираат.

### **EDITORIAL COUNCIL**

Lusi Karanikolova-Chochorovska  
Tole Belcev  
Nina Daskalovska  
Biljana Ivanovska  
Eva Gjorgjievska  
Marija Leontik  
Jovana Karanikik Josimovska  
Natka Jankova-Alagjozovska

### **LANGUAGE EDITORS**

Ranko Mladenoski (Macedonian language)  
Vesna Prodanovska (English language)  
Tole Belcev (Russian language)  
Biljana Ivanovska (German language)  
Marija Leontik (Turkish language)  
Eva Gjorgjievska (French language)  
Jovana Karanikik Josimovska (Italian language)

### **TECHNICAL EDITOR**

Slave Dimitrov

### **ADDRESS**

PALIMPSEST  
EDITORIAL COUNCIL  
Faculty of Philology  
Krstev Misirkov 10-A  
P.O. Box 201  
MK-2000, Stip

<http://js.ugd.edu.mk/index/PAL>

The International Scientific Journal “Palimpsest” is issued twice a year in printed form and online at the following website of the web portal of Goce Delcev University in Stip: <http://js.ugd.edu.mk/index.php/PAL>  
Papers can be submitted and published in the following languages: Macedonian, English, German, French, Russian, Turkish and Italian language.  
All papers are peer-reviewed.



## СОДРЖИНА / TABLE OF CONTENTS

### 11 ПРЕДГОВОР

Ева Ѓорѓиевска, уредник на „Палимпсест“

### FOREWORD

Eva Gjorgjievska, Editor of “Palimpsest”

### ЈАЗИК / LANGUAGE

### 15 Elena Shalevska

SENTENCE STRUCTURE IN HUMAN AND AI-GENERATED TEXTS: A COMPARATIVE STUDY

### 25 Mehmet Kahraman

SÖZLÜKLERDE ALAN ETİKETLERİNİN İŞLEVLERİ ÜZERİNE BİR İNCELEME: ‘KAZAK TİLİNİN TÜSİNDİRME SÖZDİĞİ’ ÖRNEĞİ

### Mehmet Kahraman

A STUDY ON THE FUNCTIONS OF FIELD LABELS IN DICTIONARIES: THE EXAMPLE OF “KAZAK TILİNİN TÜSİNDİRME SÖZDİĞİ”

### 37 Марија Леонтиќ

СИНТАГМИ СО ГЛАГОЛСКА ПРИДАВКА ВО ТУРСКИОТ ЈАЗИК И НИВНОТО ПРЕДАВАЊЕ ВО МАКЕДОНСКИОТ ЈАЗИК

### Marija Leontik

WORD GROUPS WITH A PARTICIPLE IN TURKISH LANGUAGE AND THEIR EQUIVALENCE IN MACEDONIAN LANGUAGE

### 47 Марија Соколова

УПОТРЕБАТА НА ПАРОНИМСКИТЕ ПАРОВИ ЧИТКО/ЧИТЛИВО И ГЕНЕТСКИ/ГЕНЕТИЧКИ

### Marija Sokolova

THE USE OF PARONYM PAIRS READABLE / LEGIBLE AND GENETIC / GENETICS

### 55 Hana Peloušková

IST DAS MULTILINGUALE KORPUS INTERCORP EINE GEEIGNETE MATERIALQUELLE FÜR KONTRASTIVE UNTERSUCHUNGEN VON PHRASEM-KONSTRUKTIONEN?

### Hana Peloušková

IS THE MULTILINGUAL CORPUS INTERCORP A SUITABLE SOURCE OF MATERIAL FOR CONTRASTIVE ANALYSES OF PHRASEME CONSTRUCTIONS?

- 65 Brikena Kadzadej, Admira Nushi**  
ÜBER GRENZEN HINWEG: EINE EMPIRISCHE ERFORSCHUNG DER  
DEUTSCH-ALBANISCHEN ZWEISPRACHIGKEIT  
**Brikena Kadzadej, Admira Nushi**  
ACROSS BORDERS: AN EMPIRICAL EXPLORATION OF GERMAN  
ALBANIAN BILINGUALISM

#### КНИЖЕВНОСТ / LITERATURE

- 77 Jasmina Mojsieva-Gusheva**  
ПРИСУТСТВОТО НА МАГИЧНИОТ РЕАЛИЗАМ ВО ПРОЗАТА НА ЖИВКО  
ЧИНГО  
**Jasmina Mojsieva-Gusheva**  
THE PRESENCE OF MAGIC REALISM IN THE NOVEL AND SHORT STORY  
OF ŽIVKO ČINGO
- 91 Софија Иванова, Ранко Младеноски**  
СИНОНИМНИТЕ ЖЕНСКИ ЛИКОВИ ВО МАКЕДОНСКАТА ДРАМА ОД  
ПРВАТА ПОЛОВИНА НА 20 ВЕК  
**Sofija Ivanova, Ranko Mladenoski**  
THE SYNONYMOUS CHARACTERS WITH THE ACTING FUNCTION OF  
HELPERS IN THE MACEDONIAN DRAMA FROM THE FIRST HALF OF THE  
20<sup>TH</sup> CENTURY
- 103 Mauro Dujmović, Sunčana Tuksar**  
AI-DRIVEN COMMUNICATION: THE ROLE OF ARTIFICIAL INTELLIGENCE  
IN SHAPING CONSUMERISM AND SOCIAL CONTROL IN *BRAVE NEW  
WORLD* AND *NINETEEN EIGHTY-FOUR*
- 115 Luisa Emanuele**  
LA PRIGIONE DEL LUSSO E DELLA FAMA. L'IDENTITÀ FITTIZIA DA  
FITZGERALD ALLA CONTEMPORANEITÀ DIGITALE  
**Luisa Emanuele**  
THE PRISON OF LUXURY AND FAME. FICTITIOUS IDENTITY FROM  
FITZGERALD TO DIGITAL CONTEMPORANEITY
- 127 Zeki Gürel**  
HALİD ZİYA UŞAKLIGİL'İN NEMİDE ADLI ROMANI ÜZERİNE  
**Zeki Gürel**  
A STUDY ON HALID ZIYA UŞAKLIGİL'S NOVEL "*NEMİDE*"
- 135 Lindita Kazazi, Aterda Lika**  
LATRADUZIONE DELLA POESIA ITALIANA NELL'ALBANIA DITTATORIALE  
– IL CASO DI NIKOLLË DAKAJ TRA LA RESISTENZA E LA PERSECUZIONE  
**Lindita Kazazi, Aterda Lika**  
THE TRANSLATION OF ITALIAN POETRY IN DICTATORIAL ALBANIA - THE  
CASE OF NIKOLLË DAKAJ – BETWEEN RESISTANCE AND PERSECUTION

## КУЛТУРА / CULTURE

- 149 Катица Кулавакова**  
НАРЦИЗМОТ НА МАЛИТЕ РАЗЛИКИ И БАЛКАНИЗАЦИЈАТА  
**Katica Kulavkova**  
THE NARCISSISM OF MINOR DIFFERENCES AND BALKANIZATION
- 159 Туркан Олцај**  
ПИОНЕРИТЕ НА БУГАРСКАТА ПРЕРОДБА НА СТРАНИЦИТЕ НА  
ТУРСКОТО СПИСАНИЕ „ШЕХБАЛ“  
**Türkan Olcaj**  
THE PIONEERS OF THE BULGARIAN REVIVAL ON THE PAGES OF THE  
TURKISH MAGAZINE “SHEHBAL”
- 169 Özge Nur Ünal**  
ŞAMANİZM VE ESKİ TÜRK İNANÇ SİSTEMİNDEN İZLER TAŞIYAN BİR  
DÜĞÜN RİTÜELİ: TAVUK GETİRME OYUNU  
**Özge Nur Ünal**  
A WEDDING RITUAL WITH TRACES OF SHAMANISM AND THE OLD  
TURKISH BELIEF SYSTEM: CHICKEN FETCHING GAME

## МЕТОДИКА НА НАСТАВАТА / TEACHING METHODOLOGY

- 181 Kamran Akhtar Siddiqui**  
ATTITUDES OF BUSINESS UNDERGRADUATES TOWARDS ENGLISH-  
MEDIUM INSTRUCTION FOR THEIR ACADEMIC AND PROFESSIONAL  
CAREER ASPIRATIONS: A CASE STUDY
- 191 Anastasija Anastasova, Nina Daskalovska**  
LANGUAGE PROFICIENCY AND SPEAKING ANXIETY IN MACEDONIAN  
LEARNERS OF ENGLISH
- 203 Martina Mihaljević, Maja Pivčević**  
TRAITEMENT DES ERREURS ORALES EN FLE – ATTITUDES ET  
PRÉFÉRENCES  
**Martina Mihaljević, Maja Pivčević**  
ORAL ERRORS TREATMENT IN FLE CLASSROOM: ATTITUDES AND  
PREFERENCES
- 215 Jonida Bushi, Ema Kristo**  
DIEROLLEDERGESCHICHTEBEIMERLERNENVONFREMDSPRACHEN:  
EIN BESONDERER FOKUS AUF DIE DEUTSCHE SPRACHE  
**Jonida Bushi, Ema Kristo**  
THE ROLE OF HISTORY IN LEARNING FOREIGN LANGUAGES: A  
SPECIAL FOCUS ON THE GERMAN LANGUAGE

- 227 Marisa Janku**  
ARBEIT MIT KÜNSTLICHER INTELLIGENZ IM DEUTSCHUNTERRICHT:  
EINE FALLSTUDIE AM BEISPIEL VON CHATGPT  
**Marisa Janku**  
WORKING WITH ARTIFICIAL INTELLIGENCE IN GERMAN LESSONS: A  
CASE STUDY USING THE EXAMPLE OF CHATGPT

- 235 Kevin Simonov, Nina Daskalovska**  
WHEN ARE STUDENTS AT THEIR PEAK PERFORMANCE?

#### **ПРИКАЗИ / BOOK REVIEWS**

- 247 Марина Димитриева-Ѓорѓиевска**  
ОПОЗИТНОСТА КАКО ОСНОВА НА РОМАНОТ „СВЕТОТ ШТО ГО  
ИЗБРАВ“ ОД КАЛИНА МАЛЕСКА  
**Marina Dimitrieva-Gjorgievska**  
OPPOSITION AS THE FOUNDATION OF THE NOVEL *THE WORLD I  
CHOSE* BY KALINA MALESKA

#### **ДОДАТОК / APPENDIX**

- 259** ПОВИК ЗА ОБЈАВУВАЊЕ ТРУДОВИ  
ВО МЕЃУНАРОДНОТО НАУЧНО СПИСАНИЕ „ПАЛИМПСЕСТ“  
CALL FOR PAPERS  
FOR THE INTERNATIONAL SCIENTIFIC JOURNAL “PALIMPSEST”

## IST DAS MULTILINGUALE KORPUS INTERCORP EINE GEEIGNETE MATERIALQUELLE FÜR KONTRASTIVE UNTERSUCHUNGEN VON PHRASEM-KONSTRUKTIONEN?

**Hana Peloušková**

Pedagogická fakulta Masarykovy univerzity, Brno, Česká republika  
[pelouskova@ped.muni.cz](mailto:pelouskova@ped.muni.cz)

**Abstract:** Das Projekt PhraConRep (COST) erstellt ein Repository deutscher Phrasem-Konstruktionen (satzwertige oder nichtsatzwertige Strukturen mit freien Slots) und ihrer Entsprechungen in 14 Projektsprachen (Albanisch, Bosnisch, Bulgarisch, Kroatisch, Mazedonisch, Polnisch, Russisch, Serbisch, Slowakisch, Slowenisch, Tschechisch, Ukrainisch, Ungarisch und Weißrussisch), in dem authentische Belege gesammelt und detailliert beschrieben werden. Dieser Beitrag setzt sich zum Ziel festzustellen, ob das vielsprachige Parallelkorpus InterCorp (mit Tschechisch als „Pivotsprache“) eine geeignete und ausreichende Quelle für alle oder zumindest einige der Projektsprachen sein kann. Untersucht werden die Größe der Subkorpora jeder Sprache mit Deutsch als Ausgangssprache, das Vorhandensein der notwendigen Metainformationen über die Texte und das Vorkommen von drei Phrasem-Konstruktionen in diesen Subkorpora. Zufriedenstellende Ergebnisse werden, wie zu erwarten war, für das deutsch-tschechische Korpus erzielt, aber auch für Polnisch, Ukrainisch und Kroatisch kann InterCorp sehr nützlich sein. Für andere Sprachen ist es möglich, linguistisches Material aus den sog. „Sammlungen“ zu extrahieren, d. h. aus mehrsprachigen, automatisch bearbeiteten Korpora, in denen jedoch die Metainformationen (unbekannte Originalsprache) unvollständig sind.

**Schlüsselwörter:** *Phrasem-Konstruktionen, Phraseoschablonen, Anker, Slots, Parallelkorpora, InterCorp, kontrastive Untersuchungen.*

### 1. Zielsetzung

Dieser Beitrag<sup>1</sup> hat einen technisch-informativen Charakter. Er setzt sich zum Ziel zu überprüfen, ob die Parallelkorpora im Rahmen des Projekts InterCorp eine geeignete bzw. sogar ausreichende Materialquelle für das Projekt „A multilingual repository of Phraseme Constructions in Central and East European languages“(COST) darstellen.

Auf einer allgemeineren Ebene stellt er sich die Frage, ob diese Korpora eine geeignete Materialquelle für die kontrastive Forschung über Phrasem-Konstruktionen und folglich über Phraseologismen oder andere sprachliche Phänomene sein können.

<sup>1</sup> Dieser Artikel entstand im Rahmen der COST Action CA 22115 „A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages (PhraConRep)“, s. Action CA22115 - COST.

## 2. Phrasemkonstruktionen (PhK) – Begriffsbestimmung

Unter dem Begriff Phrasem-Konstruktionen (PhK) verstehen wir idiomatische Strukturen (feste Anker - Sätze oder Wortgruppen) mit offenen Stellen/Slots. Diese Strukturen tragen schon an sich eine Bedeutung (Ablehnung, Betonung, Warnung, negative oder positive Beurteilung u. a.), die beim Füllen der Slots durch konkrete Lexeme (Filler) präzisiert/konkretisiert wird (vgl. z. B. Pavlova 2022, Dobrovol'skij/Mellado Blanco 2021 u. a.).

z. B.

*Die Wirklichkeit/Bescheidenheit/die Realität/die Wahrheit/... sieht anders aus.* (Enttäuschung, Kritik)

*Fragen über Fragen, Gründe über Gründe, Torheit über Torheit ...* (große Anzahl)

Diesen Strukturen wurde bis zum Anfang des 21. Jahrhunderts wenig Aufmerksamkeit gewidmet.

Seit den 60er Jahren des 20. Jahrhunderts wurden sie erwähnt, klassifiziert, jedoch nicht besonders tiefgehend behandelt und mit verschiedenen Begriffen bezeichnet, z. B. „modellierte Bildungen“ (Černyševa 1975), „Phraseoschablonen“ (Fleischer 1997), „Modellbildungen“ (Burger 1998).

Eingehender werden die Problematik der PhK und die entsprechenden Strukturen monolingual und kontrastiv aus der Sicht der Konstruktionsgrammatik untersucht. Mit zunehmender Verbreitung dieses neuen Ansatzes sind auch neue Termini in Umlauf gekommen, wie z. B. „lexically open idioms“, „formal idioms“ (Fillmore/Kay/O'Connor 1988), „constructional idioms“ (Langacker 1987; Taylor 2016), „schematic idioms“ (Croft & Cruse 2004), „idiomatischen Konstruktionsmuster“ (Finkbeiner 2008), „Phrasem-Konstruktionen“ (Dobrovol'skij 2011) u. a. (mehr dazu Mellado Blanco/Mollica/Schafroth 2022).

## 3. Das Projekt „A multilingual repository of Phraseme Constructions in Central and East European languages“(PhraConRep)

**PhraConRep** ist ein internationales kooperatives Projekt, das darauf abzielt, eine mehrsprachige Datenbank (Repository) von Phrasem-Konstruktionen in mittel- und osteuropäischen Sprachen zu schaffen, um Forschung, Bildung und Übersetzung zu verbessern und das sprachliche Erbe Europas zu bewahren.

Aktuell sind 130 Teilnehmer aus mehr als 30 Ländern an diesem Projekt beteiligt, darunter Linguisten, Übersetzer, Datenwissenschaftler und andere Experten, die in verschiedenen Arbeitsgruppen kooperieren.

Ausgehend von der „Pivotsprache“ Deutsch werden in die Datenbank Äquivalente von PhK in Albanisch, Bosnisch, Bulgarisch, Kroatisch, Mazedonisch, Polnisch, Russisch, Serbisch, Slowakisch, Slowenisch, Tschechisch, Ukrainisch, Ungarisch und Weißrussisch eingetragen und beschrieben. Bei jeder PhK werden ihre Funktion, Bedeutung, semantische Einschränkungen, morphologische und syntaktische Beschreibung, Gebrauch, Stil, Prosodie, Äquivalente in Projektsprachen, Kommentare zu Äquivalenten, Varianten, Synonyme, Homonyme und evtl. Paronyme angeführt.

Das Projekt wird seit 2023 von der Europäischen Kooperation für Wissenschaft und Technologie (COST) finanziert (mehr Informationen über das Projekt auf <https://www.phraconrep.com/>).

Die geplante Datenbank soll lediglich authentische Belege enthalten, die aus verschiedenen Quellen (Belletristik, Publizistik, einsprachige Korpora u. a.) stammen können. Die optimalen Quellen für das Sprachmaterial scheinen Parallelkorpora zu sein, die jedoch in einigen Ländern nicht vorhanden sind.

#### **4. InterCorp (IC)**

##### **4.1. Das Projekt IC**

**InterCorp** ist ein Teil des tschechischen Nationalkorpus, ein Projekt, das vom tschechischen Ministerium für Bildung, Jugend und Sport im Rahmen verschiedener Programme 2005-2022 unterstützt wurde. An seiner Erstellung waren und sind Lehrende und Studierende der Karlsuniversität und anderer akademischer Einrichtungen beteiligt. Seit 2022 wird InterCorp ohne finanzielle Unterstützung erweitert und weiterentwickelt.

Das InterCorp-Korpus ist das Hauptergebnis des gleichnamigen Projekts, das darauf abzielte, ein großes paralleles, synchrones Korpus zu erstellen, das so viele Sprachen wie möglich abdecken sollte. Die „Pivotsprache“ ist Tschechisch, d. h. jeder Text im Korpus hat eine tschechische Parallele. Das Korpus dient als Datenquelle für theoretische Studien, die Lexikographie, studententische Qualifikationsarbeiten, den Unterricht, insbesondere den Fremdsprachunterricht sowie die Computerverarbeitung natürlicher Sprache und soll z. B. Übersetzern und der breiteren Öffentlichkeit als Informationsquelle dienen. InterCorp kann nach kostenloser Registrierung und Anmeldung vollständig genutzt werden (vgl. z. B. Káňa, 2014; Peloušková, 2013).

##### **4.2. Zusammensetzung von IC**

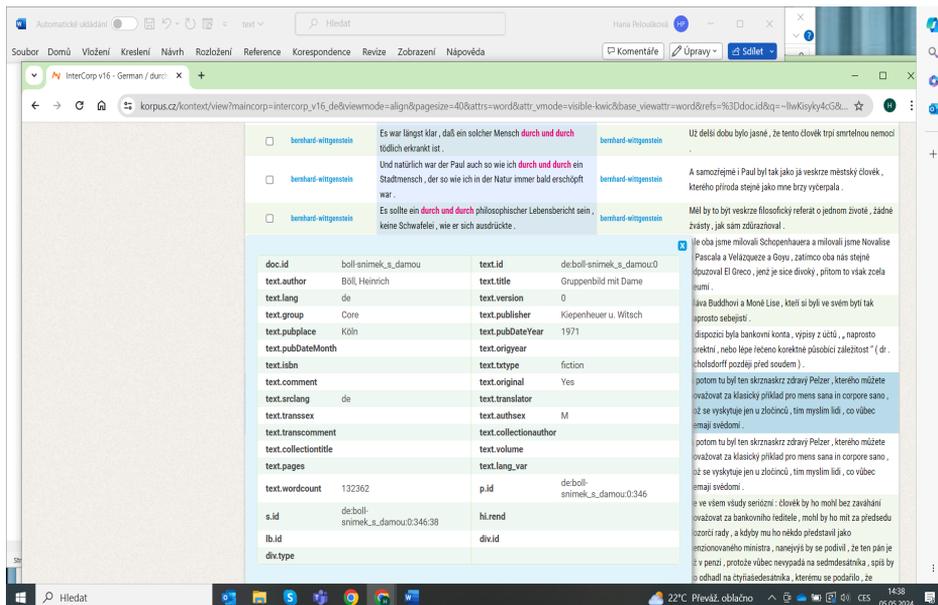
Das InterCorp-Korpus besteht aus dem sog. Kern (Core) und aus den sog. Sammlungen/Kollektionen.

Der Kern des InterCorp-Korpus (Core) besteht aus handkorrigierten, meist belletristischen Texten. Zu jedem Text können detaillierte Metainformationen angezeigt werden, z. B.:

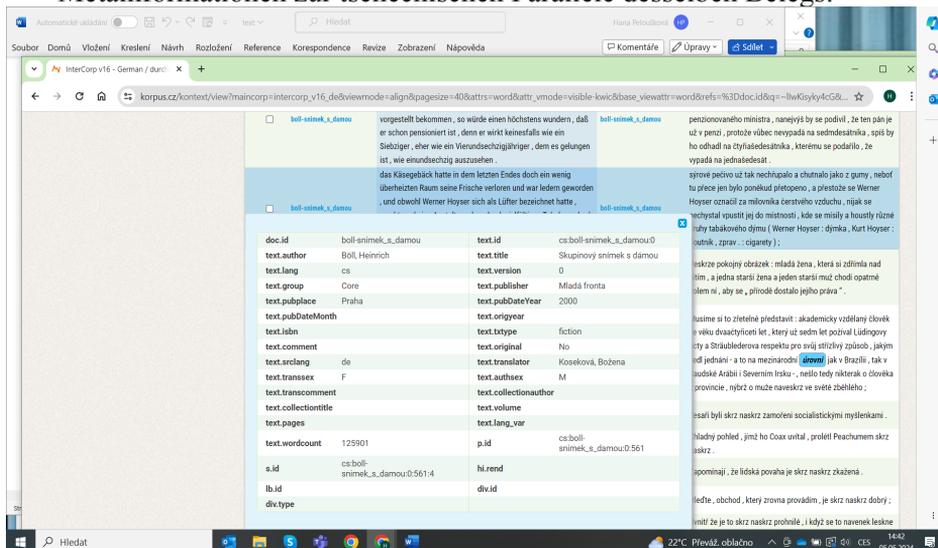
Metainformationen zur deutschen Parallele eines Belegs:<sup>2</sup>

---

<sup>2</sup> Die Einträge ins PhraConRep-Repositoryum (im GitHub) benötigen die folgenden Angaben (Metadaten): Name des Autors, Titel der Quelle, Name des Übersetzers, Jahr der Publikation. Die Originalsprache einer Reihe von Parallelstellen (z. B. Deutsch) wird in den Metainformationen angegeben und kann auch bei der Erstellung eines Subkorpus (ein Subkorpus, das nur deutsche Originaltexte und deren Übersetzungen in eine bestimmte Sprache enthält) als Auswahlkriterium festgelegt werden.



### Metainformationen zur tschechischen Parallele desselben Belegs:



Darüber hinaus enthält das Korpus auch ein Angebot an vielsprachigen automatisch bearbeiteten Texten, die sog. Sammlungen/Kollektionen. In der Version 16 sind die folgenden Kollektionen verfügbar:

- journalistische Artikel und Nachrichten von den Websites Project **Syndicate** und VoxEurop (ehemals PressEurop)
- Rechtstexte der Europäischen Union aus dem Korpus des **Acquis Communautaire**
- Protokolle des Europäischen Parlaments von 2007-2011 aus dem **Europarl**-Korpus
- Filmuntertitel aus der **OpenSubtitles**-Datenbank
- **Bibel**übersetzungen

Diese Texte werden nur automatisch aligniert (manchmal fehlerhaft). Daher können sich bei den Konkordanzen Treffer ergeben, die keine tatsächlichen Parallelstellen darstellen. Bei manchen Texten fehlen auch Parallelen in einer oder in mehreren Sprachen. Die Metainformationen sind nicht vollständig. Die Originalsprache wird hier nicht angegeben. Es lässt sich zwar der Name des Autors, nicht aber der Name des Übersetzers ins Deutsche/Tschechische/andere Sprachen feststellen. Oft werden Texte in weniger verbreiteten Sprachen zuerst ins Englische und dann vom Englischen in andere Sprachen übersetzt.

Mit einem befriedigenden Alignment ist das aus den Sammlungen extrahierte Material eine brauchbare Ressource für Untersuchungen von konkreten PhK, aber für den Aufbau einer Datenbank von Phrasem-Konstruktionen sind die Metainformationen nicht ausreichend.<sup>3</sup>

### 4.3. Umfang von IC

Jeder fremdsprachige Text hat im Korpus seine tschechische Version (Tschechisch ist der sog. Pivot). Die tschechische Version des Textes (Original oder Übersetzung) wird mit einer oder mehreren fremdsprachigen Versionen aligniert. Der Gesamtumfang des zugänglichen Teils des Korpus in der Version 16 von 2023 beträgt 387 Millionen Wörter im Kern (Core) und 4.506 Millionen Wörter in den Kollektionen der alignierten fremdsprachigen Texte. Bei den tschechischen Texten sind es 125 Millionen Wörter im Kern (Core) und 273 Millionen Wörter in den Kollektionen.

Eine umfangreiche Beschreibung der aktuellen Version (16), wie z. B. die Verteilung der Texte innerhalb des Kerns/Core und in den einzelnen Sammlungen/Kollektionen genau wie die Vertretung der einzelnen Sprachen im Korpus in Millionen von Wörtern ist auf <https://wiki.korpus.cz/doku.php/cnk:intercorp:verze16> zu finden.

### 4.4. IC und PhraConRep

#### 4.4.1. IC und mehrsprachige Parallelkorpora

Eine kleine Stichprobe aus der Textdatenbank von IC zeigt, dass einige deutsche Originaltexte (aus dem Core) zahlreiche fremdsprachige Parallelen aufweisen, was theoretisch die gleichzeitige Suche in mehreren Sprachen ermöglicht.

Als Beispiel geben wir das Auftreten dreier literarischer Texte in den Projektsprachen:

<b>Sprache</b>	<b>Kafka: Der Prozess</b>	<b>Hesse: Der Steppenwolf</b>	<b>Grass: Die Blechtrommel</b>
Albanisch			
Bulgarisch	ja		
Bosnisch			
Kroatisch	ja	ja	
Mazedonisch	ja	ja	ja
Polnisch	ja	ja	ja

<sup>3</sup> Das gleiche Problem lässt sich bei der Extraktion von Material aus SketchEngine beobachten.

Russisch	ja		
Serbisch	ja	ja	ja
Slowakisch			
Slowenisch	ja		
Tschechisch	ja	ja	ja
Ukrainisch	ja	ja	ja
Ungarisch	ja		
Weißrussisch	ja	ja	

Bei der gleichzeitigen Suche nach PhK in mehreren Sprachen wird das Ergebnis auf Ressourcen eingegrenzt, die allen Sprachen gemeinsam sind. So ergab beispielsweise die Suche nach den deutschen Parallelen **N´ für N´** auf Weißrussisch, Kroatisch und Tschechisch nur 17 Belege aus drei Quellen (*Tag für Tag, Mann für Mann, Monat für Monat, Woche für Woche, Nacht für Nacht, Schritt für Schritt, Zug für Zug*). Eine parallele Suche in mehreren Sprachen erwies sich als für das Projekt weniger geeignet. Zweisprachige Recherchen sind eindeutig für das Projekt effektiver.

#### 4.4.2. Parallele Subkorpora des Deutschen mit den Projektsprachen

In Übereinstimmung mit den Regeln für die Erstellung der Datenbank der PhK erscheint es sinnvoll, zweisprachige Subkorpora mit Deutsch als Ausgangssprache nach den folgenden Parametern zu erstellen:

- Beschränkung der Auswahl der Texte auf Core
- Beschränkung der Texte in Core auf Texte mit Deutsch als Ausgangssprache/Originalsprache

Untersuchen wir nun die Subkorpusgrößen der Projektsprachen:

Sprache	Anzahl der Tokens	Anmerkungen
Albanisch	0	kein Core, nur Kollektionen
Bulgarisch	88 128	
Bosnisch	0	kein Core, nur Kollektionen
Kroatisch	2 363 059	
Mazedonisch	552 783	
Polnisch	2 506 957	
Russisch	548 376	
Serbisch	1 371 586	
Slowakisch	4 239	
Slowenisch	88 128	
Tschechisch	17 315 727	
Ukrainisch	2 126 096	
Ungarisch	345 485	
Weißrussisch	876 676	

Diese Übersicht zeigt, dass das tschechisch-deutsche Subkorpus am umfangreichsten und deshalb besonders brauchbar ist (Tschechisch ist im IC immer präsent). Das IC hat bisher für alle für die Datenbank bearbeiteten PhK eine ausreichende Anzahl an Belegen geliefert.

Die Subkorpora mit Polnisch, Kroatisch, Ukrainisch und Serbisch sind brauchbar.

Wesentlich kleinere Teilkorpora mit anderen Projektsprachen (außer Albanisch und Bosnisch)<sup>4</sup> sind nur sehr eingeschränkt nutzbar.

#### 4.4.3. Umfang der Subkorpora und Vertretung von PhK

In diesem kleinen Exkurs wird gezeigt, inwieweit der Umfang der parallelen Subkorpora mit der quantitativen Repräsentation/Vertretung bestimmter PhK korrespondiert.

Zu den häufigeren PhK gehören Strukturen mit rekkurrenten Substantiven, die durch Präpositionen verbunden sind, z. B.: **N' für N', N' um N' und N' über N'**.

##### 4.4.3.1. N' für N'

Diese PhK (z. B. *Jahr für Jahr*) dient zum Ausdruck einer kontinuierlichen Aufeinanderfolge (siehe die Beschreibung der Konstruktion im Repositoryum)

Von den drei oben genannten PhK kommt sie am häufigsten vor. Aus den oben angeführten Parallelkorpora der einzelnen Projektsprachen mit Deutsch als Ausgangssprache wurde diese PhK durch folgende Abfrage extrahiert:

**1:[tag="N.\*"] [word="für"] 2:[tag="N.\*"] & 1.lemma = 2.lemma**

Ergebnisse (Anzahl der Vorkommen dieser PhK):

Sprache	Anzahl der Vorkommen	Anmerkungen
Albanisch	0	kein Core, nur Kollektionen
Bulgarisch	2	
Bosnisch	0	kein Core, nur Kollektionen
Kroatisch	97	
Mazedonisch	25	
Polnisch	54	
Russisch	32	
Serbisch	71	
Slowakisch	0	
Slowenisch	2	
Tschechisch	<b>536</b>	
Ukrainisch	<b>76</b>	
Ungarisch	14	
Weißrussisch	45	

Für manche Sprachen (Kroatisch, Polnisch, Serbisch, Ukrainisch und natürlich für Tschechisch) könnte die Anzahl der Belege eine ausreichende Basis für einen Eintrag in die Datenbank und für die allseitige Beschreibung sein.

Für die anderen Sprachen können diese Belege eher als eine Bereicherung des Materials aus anderen Quellen dienen.

Mit niedrigerer Häufigkeit der PhK sinkt natürlich die Anzahl der vorgekommenen Belege.

<sup>4</sup> Bei diesen Sprachen ist kein Core vorhanden.

#### 4.4.3.2. N' um N'

Auch diese PhK (z. B. *Jahr um Jahr*) dient zum Ausdruck einer kontinuierlichen Aufeinanderfolge (siehe die Beschreibung der Konstruktion im Repositorium). Die Abfrage ist:

**1:[tag="N.\*"] [word="um"] 2:[tag="N.\*"] & 1.lemma = 2.lemma**

Ergebnisse (nur bei umfangreicheren Subkorpora):

Sprache	Anzahl der Vorkommen
Kroatisch	51
Polnisch	24
Serbisch	44
Tschechisch	415
Ukrainisch	64

#### 4.4.3.3. N' über N'

Noch seltener kommt diese PhK vor, die eine große Anzahl bezeichnet (z. B. *Gründe über Gründe*, d. h. sehr viele Gründe).

Die Abfrage ist:

**1:[tag="N.\*"] [word="über"] 2:[tag="N.\*"] & 1.lemma = 2.lemma**

Ergebnisse (nur bei umfangreicheren Subkorpora)<sup>5</sup>

Sprache	Anzahl der Vorkommen
Kroatisch	12
Polnisch	13
Serbisch	5
Tschechisch	63
Ukrainisch	6

Diese Anzahl der Belege könnte für einen Eintrag in die Datenbank reichen, für eine detaillierte Beschreibung könnte ein Subkorpus aus dem ganzen IC (Core + Kollektionen) als akzeptable Quelle dienen.

#### 4.4.4. Das ganze IC und Vertretung von PhK

Vergleichen wir die Anzahl der extrahierten Belege der PhK N' um N' aus dem Core und aus dem ganzen IC<sup>6</sup>. Die Abfrage ist gleich wie bei Core:

**1:[tag="N.\*"] [word="um"] 2:[tag="N.\*"] & 1.lemma = 2.lemma**

Sprache	Anzahl der Belege - Core	Anzahl der Belege – das ganze IC
Kroatisch	97	691
Polnisch	54	657
Serbisch	71	650
Tschechisch	536	1508
Ukrainisch	76	355

<sup>5</sup> Die Größe der Subkorpora ist nicht direkt proportional zur Anzahl der gefundenen PhK, entscheidend ist, aus welchen konkreten Texten (Werken) das jeweilige Subkorpus zusammengesetzt ist.

<sup>6</sup> Zu den Nachteilen, siehe oben im Text.

Meines Erachtens ist die die Anzahl der aus dem gesamten IC gewonnenen Belege für eine gründlichere Untersuchung dieser relativ häufigen PhK bereits ausreichend. Das auf diese Weise gewonnene Material könnte auch für andere Projektsprachen interessant sein (aber nicht für die Zusammenstellung des Repositoriums, s. weiter oben).

#### **4.4.5. Weitere Möglichkeiten der Verwendung von IC im Projekt**

Neben der eingehenden Untersuchung einzelner PhK werden die umfangreicheren Korpusdaten es ermöglichen, wertvolle statistische Daten zu erhalten, z. B.

- Häufigkeit der einzelnen PhK
- die Häufigkeit der eigenen Strukturen der PhK
- die Häufigkeit bestimmter Slotbesetzungen in PhK

#### **Fazit:**

Diese einfache Analyse hat gezeigt:

- IC enthält alle Projektsprachen.
- Für die Aufnahme der identifizierten PhK in das Repositorium ist es ratsam, ein deutsches Subkorpus zu erstellen, das Texte aus dem IC-Kern (CORE) und nur Texte enthält, deren Originalsprache Deutsch ist. Die Kerntexte sind mit vollständigen Metainformationen versehen, was für die Kollektionen nicht der Fall ist (z. B. kann die Originalsprache nicht ermittelt werden).
- Der Umfang der Parallelkorpora mit Deutsch als Ausgangssprache variiert von Projektsprache zu Projektsprache. Die Korpora mit Tschechisch, Polnisch, Kroatisch und Ukrainisch sind am größten.
- Für eine detaillierte Beschreibung der PhK ist es möglich, das Korpus auch auf die Kollektionen zu erweitern.
- Umfangreichere Korpusdaten erlauben es, die Häufigkeit einzelner PhK, ihre Strukturen und andere sprachliche Entsprechungen statistisch auszuwerten.

#### **Literatur**

- Burger, H. (2010): *Phraseologie. Eine Einführung anhand des Deutschen*. 4. Aufl.. Berlin, Erich Schmidt.
- Dobrovol'skij, D./ Mellado Blanco, C (2021): Von Jahr zu Jahr. Das Pattern [von X1sg zu X1sg] und seine Entsprechungen im Russische und Spanischen: eine Korpusstudie. *Aussiger Beiträge 15/2021*, Sonderheft *Phraseologie im digitalen Zeitalter – Neue Fragestellungen, Methoden und Analysen*, 113-138.
- Fleischer, W. (1997): *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Káňa, T. (2014): *Sprachkorpora in Unterricht und Forschung DaF/DaZ*. Brno, Masarykova univerzita
- Pavlova, A. (2022): Mehrsprachige Datenbank der Phrasem-Konstruktionen. In Klosa-Kückelhaus, A./Engelberg, S./ Möhrs, Ch./Storjohann, P. (eds.): *Dictionaries and Society*. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag, 594-604.

- Peloušková, H. (2013): Das Projekt InterCorp und seine Rolle in der Deutschlehrerausbildung und Forschung. In *Slowakische Zeitschrift für Germanistik*. Banská Bystrica: Verband der Deutschlehrer und Germanisten der Slowakei, 5(2)/2013, 55-64.
- Mellado Blanco, C./Mollica, F./Schaffroth, E. (2022): Phrasem-Konstruktionen in der heutigen Forschung: ein Überblick. In: Mellado Blanco, C./Mollica, F./Schaffroth, E. (2022): *Konstruktionen zwischen Lexikon und Grammatik. Phrasem-Konstruktionen monolingual, bilingual und multilingual*. Berlin, De Gruyter, 1-18.
- Mellado Blanco, C./Mollica, F./Schaffroth, E. (2022): *Konstruktionen zwischen Lexikon und Grammatik. Phrasem-Konstruktionen monolingual, bilingual und multilingual*. Berlin, De Gruyter.

### **Internetquellen**

- Dovalil, V./ Káňa, T./Peloušková, H./Zbytovský, Š./Vavřín, M./Zasina, A. J.: Korpus InterCorp – Deutsch, Version 16 z 12. 10. 2023. Ústav Českého národního korpusu FF UK, Praha 2023. Abgerufen am 2. 2. 2024. <http://www.korpus.cz>
- Webseite des Tschechischen Nationalkorpus. Abgerufen am 1. 3. 2024. [www.korpus.cz](http://www.korpus.cz)
- Webseite von InterCorp. Abgerufen am 15. 1. 2024. <https://intercorp.korpus.cz/>
- Beschreibung der 16. Version von InterCorp. Abgerufen am 20. 1. 2024. <https://wiki.korpus.cz/doku.php/cnk:intercorp:verze16>

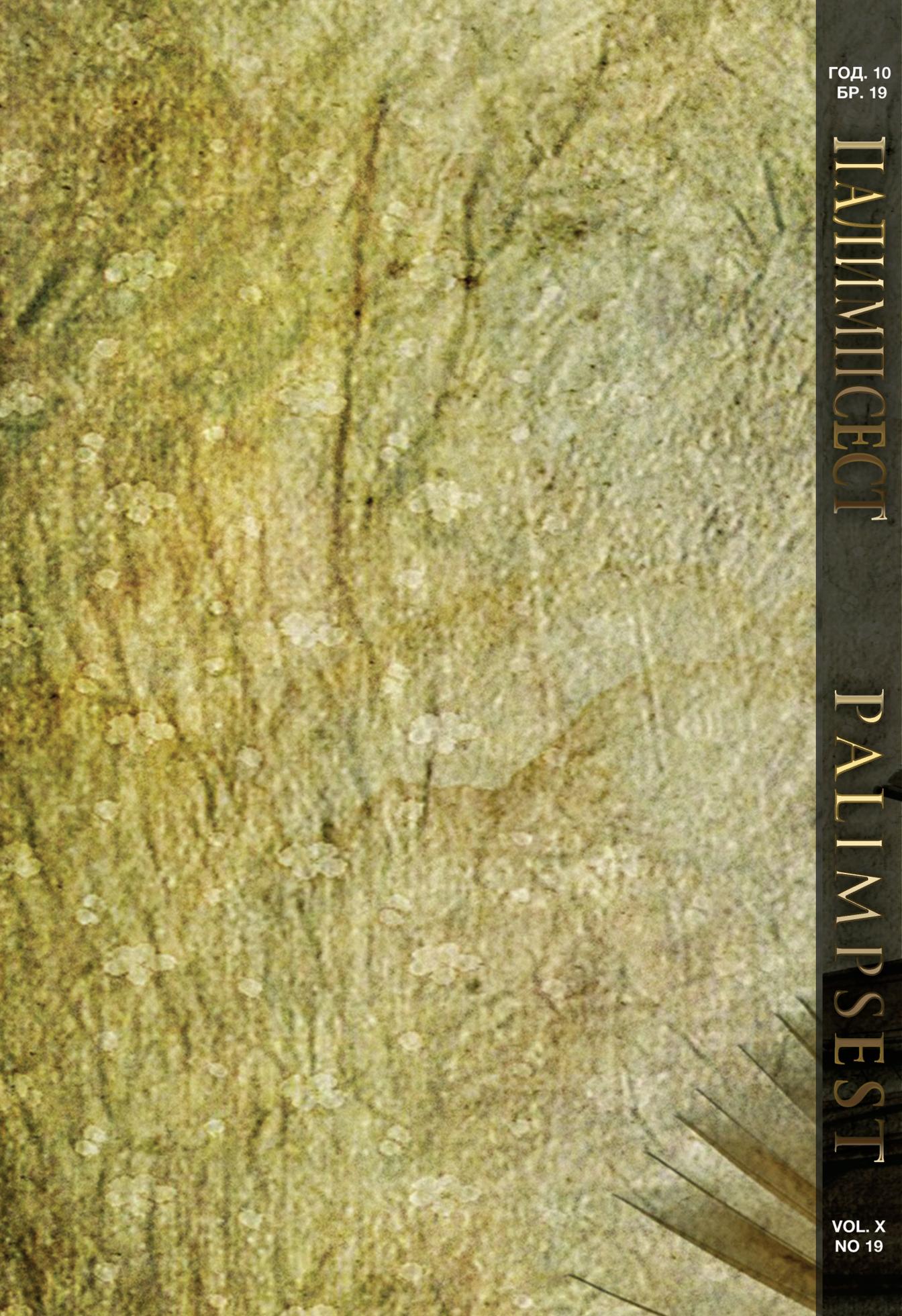
### **Hana Peloušková**

Faculty of Education, Masaryk University, Brno, Czech Republic

### **Is the Multilingual Corpus Intercorp a Suitable Source of Material for Contrastive Analyses of Phraseme Constructions?**

**Abstract:** The PhraConRep (COST) project is creating a repository of German phraseme constructions (sentential or non-sentential structures with free slots) and their equivalents in 14 project languages (Albanian, Bosnian, Bulgarian, Croatian, Macedonian, Polish, Russian, Serbian, Slovak, Slovenian, Czech, Ukrainian, Hungarian and Belarusian), in which authentic evidence of these structures will be collected and described in detail. This article aims to determine whether the multilingual parallel corpus InterCorp (with Czech as „pivot“ language) can serve as a suitable and sufficient source for all, or at least some, of the project languages. The analysis focuses on the size of the subcorpora for each language with German as the source language, the availability of necessary meta-information about the texts, and the occurrence of three phrasal constructions within these subcorpora. As expected, satisfactory results are obtained for the German-Czech corpus, however, InterCorp can also be very useful for Polish, Ukrainian and Croatian. For the remaining languages, the material can be extracted from so-called collections, i.e. from multilingual, automatically processed corpora which, however, lack complete meta-information (unknown source language).

**Keywords:** *Phraseme constructions; anchors; slots; parallel corpora; InterCorp; contrastive analyses.*



ГОД. 10  
БР. 19

ПАЛИМПСЕСТ

PALIMPSEST

VOL. X  
NO 19