

Korpusbasierte Erfassung des Grammatikalisierungsgrades von Indefinitpronomen



Universität Regensburg

Björn Hansen / Roman Fisun
Institut für Slavistik
Fakultät Sprach-, Kultur- und Literaturwissenschaften

DFG Deutsche
Forschungsgemeinschaft

Gliederung

1. DFG-Projekt Indefinitheit
2. Zusammengesetzte Indefinitpronomen
3. Messung der Grammatikalisierung anhand der
Korpusdaten: Methode und konkrete Anwendungen
4. Fazit

1. DFG-Projekt Indefinitheit

Universität Regensburg

DFG-Projekt „Zusammengesetzte Indefinitpronomen in slavischen Sprachen. Ein Beitrag zur semantischen Karte der Idefinitheit der zweiten Generation“ (HA 2659/10-1)

Roman Fisun: „Zusammengesetzte Indefinitpronomen in den slavischen Sprachen“ (RU, UA, PL)

Martina Rybová: Co-Tutelle (Univerzita Karlova) „Die Beziehung zwischen Evaluativität und Indefinitheit am Beispiel der zusammengesetzten Indefinitpronomina im Tschechischen“

Björn Hansen: HR

Irina Maykova: CQL-Abfragen (WHK)

2. Zusammengesetzte Indefinitpronomen

Indefinitpronomen (IPs) in den slavischen Sprachen:

- Nicht-reihenbildende IPs: UA *odyn* 'eins', RU *tot ili inoj* 'irgendwelcher', PL *ten i ów* 'dieser und jener'

- Reihenbildende IPs: kodifiziert=„traditionell“

UA	<i>chto</i> _{wer} [s] _{INDEF}	'jemand'
	<i>ščo</i> _{was} [s] _{INDEF}	'etwas'
	<i>jakyj</i> _{welcher} [s] _{INDEF}	'irgendein'

vs nicht-kodifiziert

UA	<i>chto</i> _{wer} [zavhodno] _{INDEF}	'jeder'
	<i>ščo</i> _{was} [zavhodno] _{INDEF}	'irgendetwas'
	<i>jakyj</i> _{welcher} [zavhodno] _{INDEF}	'jeder' (free-choice)

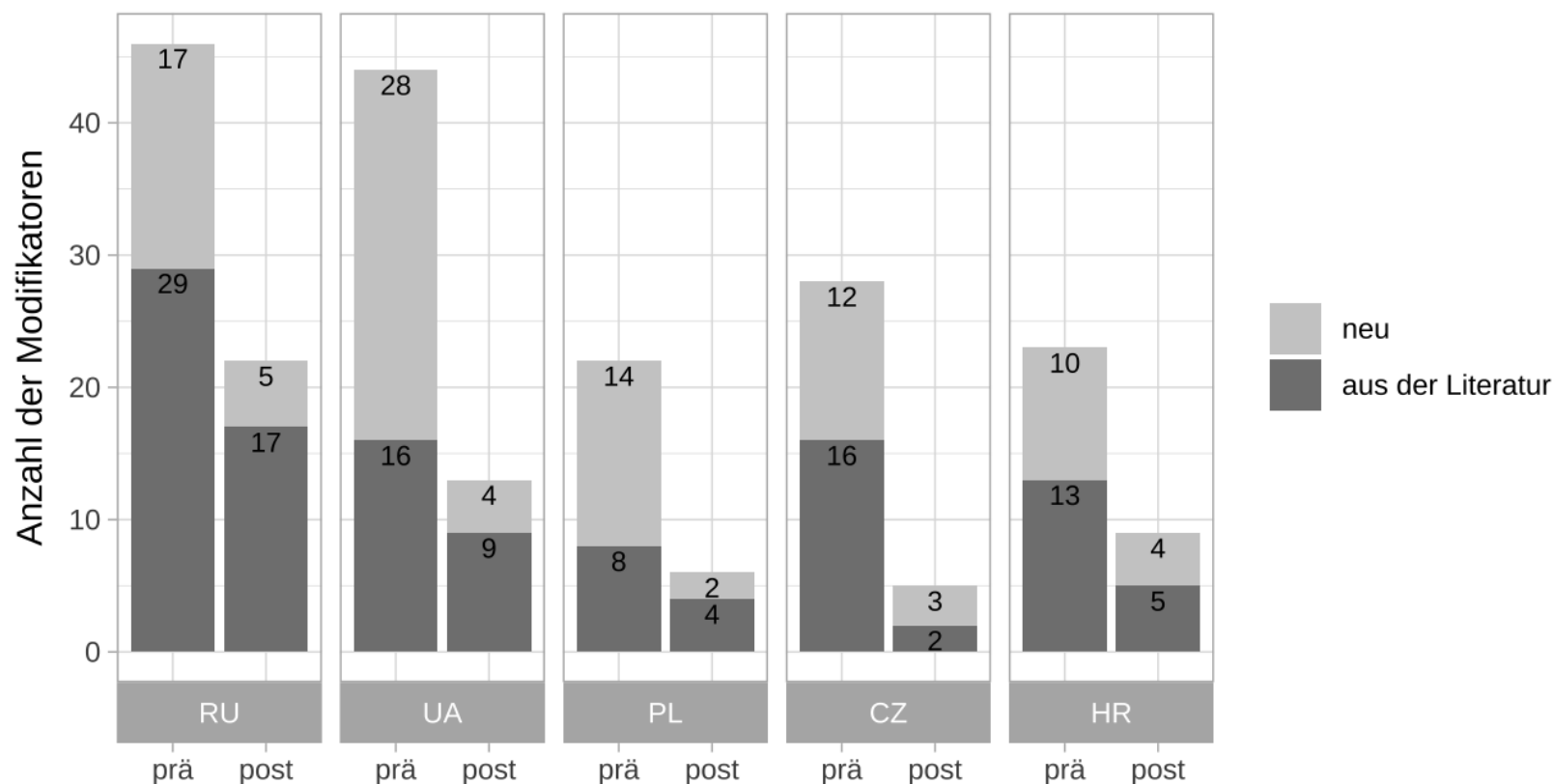
- Pronominale *k-Wurzel + „Modifikator“ (Indefinitmarker)
- Indefinitmarker sind entweder postponiert (*k-Wurzel + INDEF) oder präponiert (INDEF + *k-Wurzel)

Zusammengesetzte Indefinitpronomen (ZIPs)

- verschiedene Quellkonstruktionen
- unterschiedliche Orthografie
- verschiedene Grammatikalisierungsgrade

RU	<i>neizvestno kto</i> < ‘unbekannt wer’
UA	<i>kazna-cho</i> < ‘der Henker weiß wer’
HR	<i>gdješto</i> < ‘wo was’
PL	<i>mało co</i> < ‘wenig was’
CZ	<i>kdoví-proč</i> < ‘wer weiß warum’
SLO	<i>redko kaj</i> < ‘selten wer’

Im Rahmen des Projekts wurden slavische ZIPs erstmals systematisch und korpusbasiert inventarisiert



Datenbasis: Analyse von Daten aus slavischen Sprachkorpora

- Indefinitpronomen in fünf Sprachen: Russisch, Ukrainisch, Polnisch, Tschechisch und Kroatisch
- Hauptquelle: große Webkorpora, die über Sketch Engine zugänglich sind
 - ruTenTen11 für RU
 - ukTenTen14 und ukTenTen20 für UA
 - plTenTen12 für PL
 - csTenTen12 und csTenTen17 für CZ
 - hrWaC 2.2. für HR

3. Grammatikalisierungsgrad der ZIPs

Problemstellung

- Eine systematische Inventarisierung der ZIPs fehlte bislang
- Viele ZIPs sind nicht kodifiziert und weisen unterschiedliche Grade der Grammatikalisierung auf
- Warum den Grad der Grammatikalisierung messen?
 - um fundierte Vorschläge für die Kodifizierung zu entwickeln,
 - um den tatsächlichen Grad der nichtkodifizierten ZIPs zu messen,
 - um geeignete Elemente für grammatische Beschreibungen auszuwählen und stärker grammatikalisierte ZIPs gezielt analysieren zu können

Grammatikalisierungsparameter von Lehmann (1982) und ihre Operationalisierungen für ZIPs

Lehmans Parameter (2015)	Praktische Operationalisierung
1. Integrität	a. Verlust an semantischer Integrität b. Verlust an morphosyntaktischer Integrität c. Phonologische Erosion
2. Paradigmatizität	a. Indefinitheit b. Präsenz der ontologischen Kategorien
3. Wählbarkeit	a. Unikalität des Modifikators b. Zunahme der Frequenz
4. Skopus	a. Syntaktische Kondensierung b. Koordinative Konstruktionen c. Verletzung der Dependenzbeziehungen zwischen Modifikator und *k-Wurzel
5. Fügungseuge	a. Zusammenschreibung b. Abwesenheit der starken Autonomie c. Abwesenheit der Abtrennbarkeit d. Eingeschränkte Distribution der Modifikatoren
6. Stellungsfreiheit	Fixe Position des Modifikators in der Verbindung

Vorgeschlagene Methode

- Alle zusammengesetzten Reihen wurden nach 14 Grammatikalisierungskriterien annotiert und erhielten einen Grammatikalisierungsscore
- Annotation: Vereinfachung auf eine binäre Skala (+ / – / 0)
- Der Maßstab reicht von 0 (keine Grammatikalisierung) bis 1 (volle Grammatikalisierung) – sowohl für jeden einzelnen Parameter als auch für die Gesamtgrammatikalisierung
- Die Gesamtbewertung erfolgt als einfacher Durchschnitt der Scores nach jedem Lehmann-Parameter
- Mixed-Methods-Ansatz: Kombination aus intuitiver, korpusbasierter, qualitativer und quantitativer Analyse

Grenzen der Methode

- **Unterschiedliche Gewichtung**
Die Grammatikalisierungskriterien (wie auch Lehmanns Parameter selbst) haben unterschiedliches Gewicht
- **Kriterienwahl**
Die Auswahl der einzelnen Grammatikalisierungskriterien ist arbiträr
- **Dennoch:** geeignet für den Vergleich innerhalb der Gruppe

3.1. Einzelne Grammatikalisierungskriterien

I. Integrität: Verlust an semantischer Integrität

- Ausbleichen („Bleaching“) der Bedeutung:
Desemantisierung, Polisemierung

- (1) a. RU *On vzjal jabloko, kotoroe **popalo** emu v ruku.*
‘Er nahm den Apfel, der ihm in die Hand geriet.’
b. RU *Sjuda možeť zajti **kto popalo**.*
‘Jedermann kann hierher kommen.’

Bedeutung des Quellverbs: ‚Zufällig an eine bestimmte Stelle gelangen, sich anschließen, ohne Absicht irgendwo eintreten (Ožegov o.D.)‘

II. Paradigmatizität: Präsenz der ontologischen Kategorien

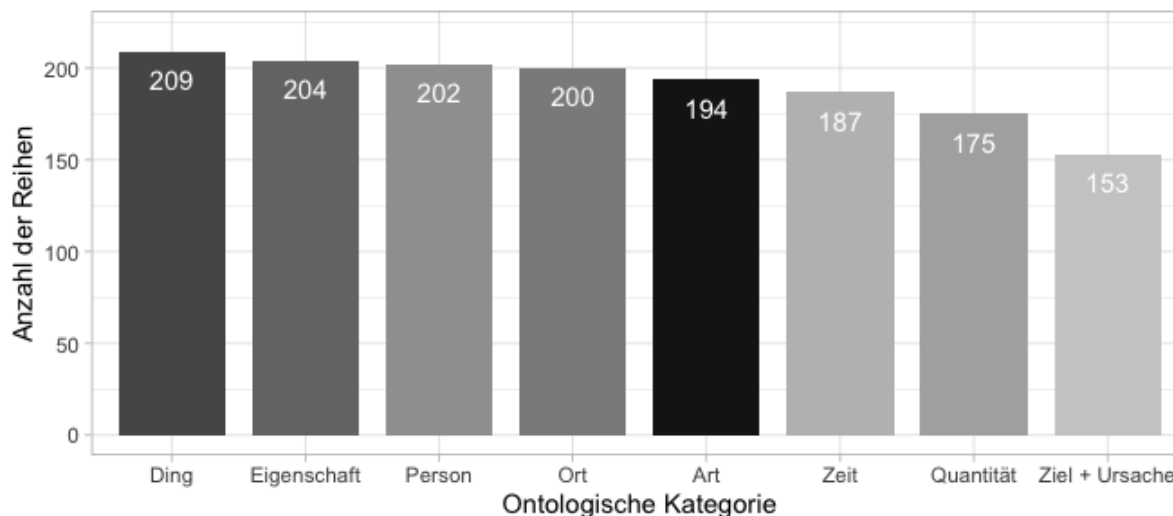
	bądź	Bóg wie	Bóg raczy wiedzieć	bogowie wiedzą	(nie)byle	cholera wie	chuj wie	czort wie	diabeł wie	diabli wiedzą	lada	ledwo	mało	nie wiadomo	niedużo	niewiele	obojętnie	pies wie	rzadko	samwiesz;/samiwiece; /samawiesz;	licho wie	nie:	bądź	:kolwiek	:ś	:kolwiek	popadło	trzeba
PERSON	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+
DING	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EIGENSCHAFT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+
ART	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ORT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
QUANTITÄT	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+		+		+	+	+		+	+
ZIEL + URSACHE	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+		+	+
ZEIT	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Kombinierbarkeit von PL Modifikatoren mit ontologischen Kategorien nach pITenTen12

Aber: Frequenzinformation zum Ausdruck der ontologischen Kategorien

	PERSON	DING	EIGENSCHAFT	ART	ORT	QUANTITÄT	ZIEL + URSACHE	ZEIT
málo:	415442	31074	7651	130561	2423	29566	26	91
lec:	31076	7455	7567	27811	268	13780	0	0

Absolute Frequenzen der CZ Reihen um *málo*: und *lec*: nach den ontologischen Kategorien, nach csTenTen17. Die Zahlen wurden nicht manuell korrigiert und können teilweise auch Quellkonstruktionen und „data noise“ enthalten.

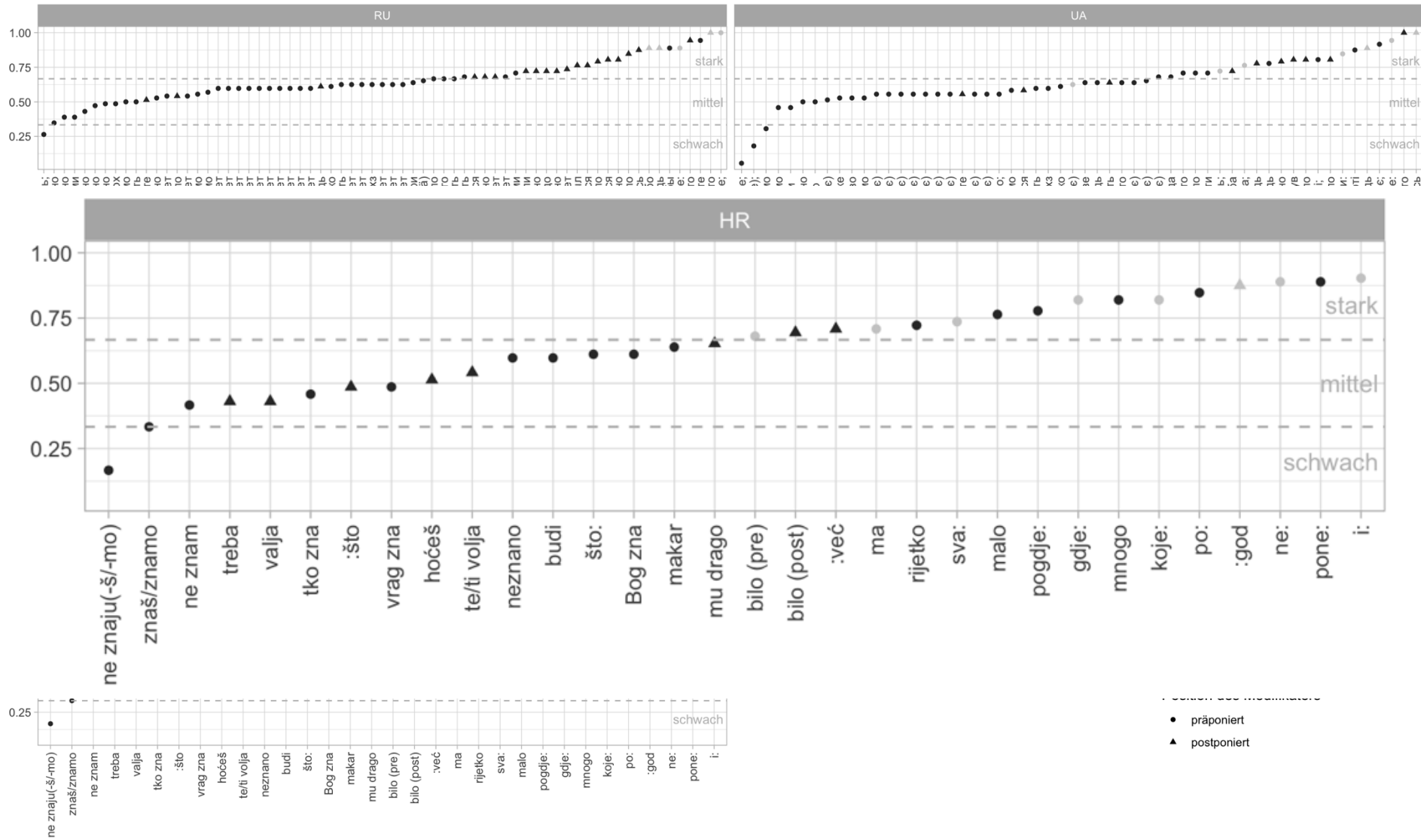


Anzahl der slavischen ZIPs, die jeweilige ontologische Kategorie ausdrücken, nach Webkorpora

III. Skopus: Koordinative Konstruktionen

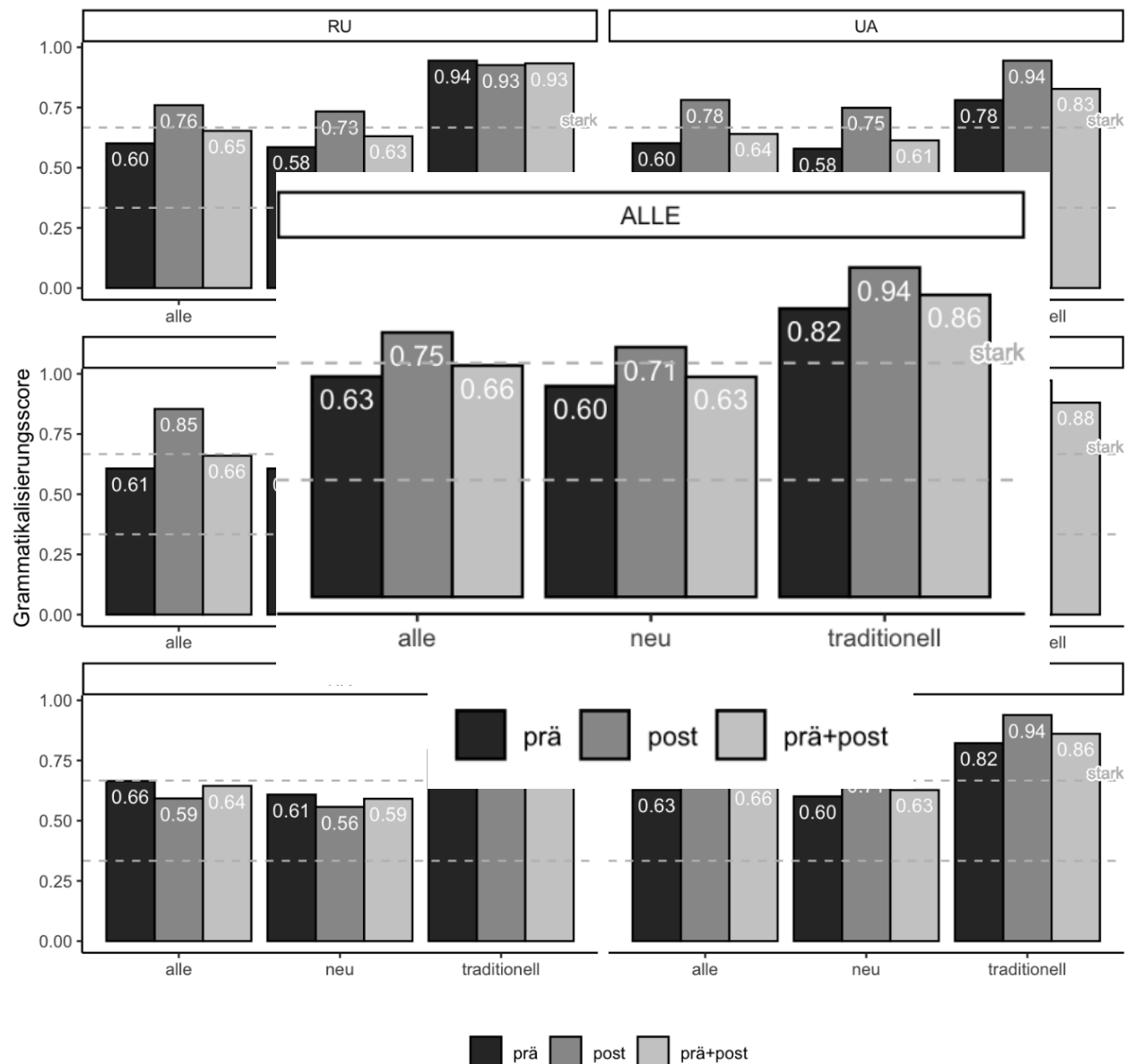
- Koordinative Konstruktionen erlaubt:
- (2) CZ *Zkrátka budou-li nás takovíto tvorové chtít vyhubit, **sotva** [kdo]*_{k-Wurzel} **a**und [co]*_{k-Wurzel} jim v tom zabrání.* (csTenTen17)
 ‘Kurz gesagt, wenn diese Kreaturen uns ausrotten wollen, wird sie kaum jemand oder etwas aufhalten.’
- Verengung des Modifikatorenskopus auf den Stamm:
- (3) HR *Pa cijelo društvo je ovisno o **ne[kome]***_{k-Wurzel} **ili** *oder* o **ne[čemu]***_{k-Wurzel} / ***ne[kome]***_{k-Wurzel} **ili** *oder* **[čemu]***_{k-Wurzel}.* (hrWaC)
 ‘Die gesamte Gesellschaft hängt von jemandem und etwas ab.’

3.2. Ergebnisse



Grammatikalisierungsscores der „traditionellen“ und neuen ZIPs

Universität Regensburg

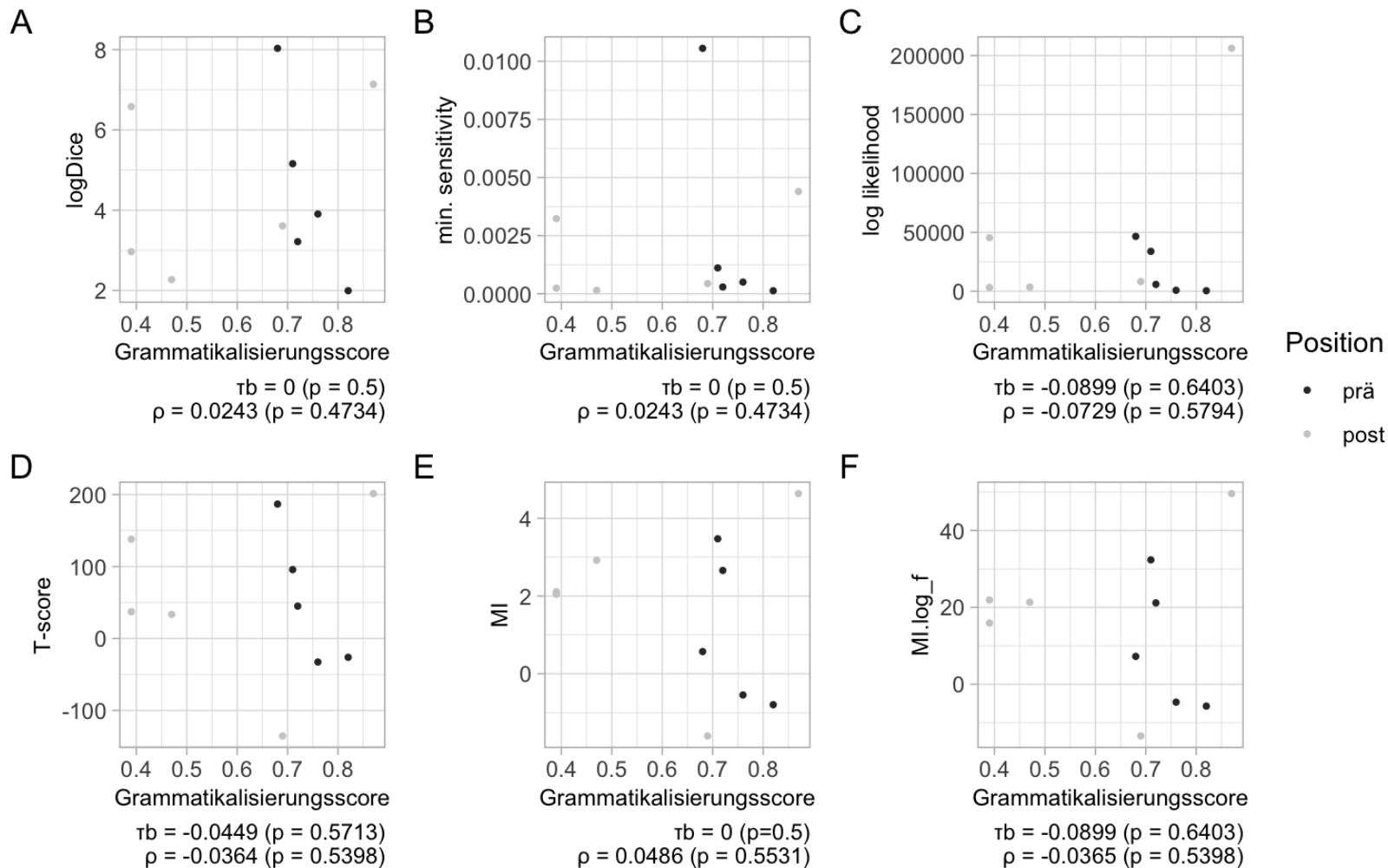


Grammatikalisierungsgrad der Modifikatoren, die auf ähnliche Quellen zurückgehen

- Viele Modifikatoren zeigen vergleichbaren Grammatikalisierungsgrad in RU, UA, PL, CZ
- Quantitative Modifikatoren ('viel' und 'wenig') variieren stärker

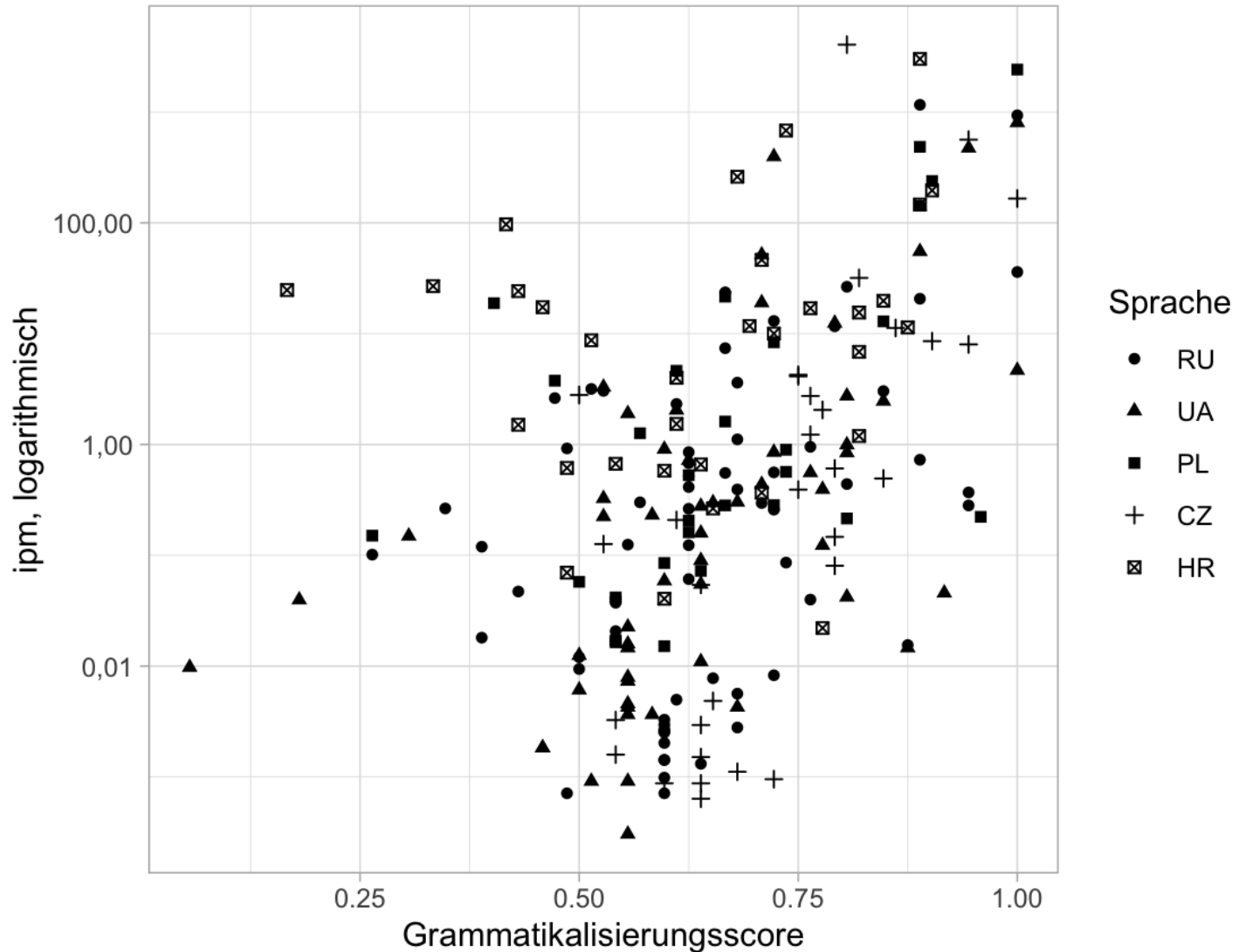
Korrelation zwischen dem Grammatikalisierungsgrad und anderen Parametern

keine Korrelation mit der Kollokationsstärke der Verbindung ‚Modifikator+*k-Wurzeln‘



Universität Regensburg

schwache, statistisch signifikante positive Korrelation mit der Frequenz



5. Fazit und Beitrag zur allgemeinen Linguistik

Beitrag zur Grammatikalisierungstheorie von Indefinitpronomen:

1. Entwicklung eines Verfahrens zur Analyse von Grammatikalisierungsprozessen bei Indefinitpronomen, das auch auf weitere Sprachen übertragbar ist.
2. Es konnte keine Korrelation zwischen dem Grammatikalisierungsgrad von Modifikatoren und der Kollokationsstärke ihrer Verbindung mit Pronominalstämmen festgestellt werden.
3. Zwischen dem Grammatikalisierungsgrad von Modifikatoren und der Frequenz konnte dagegen eine schwache, statistisch signifikante positive Korrelation gefunden werden.
4. Es wurde gezeigt, dass Indefinitmarker, die in verwandten Sprachen auf parallele Quellen zurückgehen, auch ähnliche Grammatikalisierungsgrade entwickeln.

Ergebnisse mit Bedeutung für die Slavistik:

1. Grammatikalisierungsgrade der ZIPs wurden bestimmt.
2. Die Hypothese, dass die kodifizierten Indefinitpronomen auch stärker grammatikalisiert sind, wurde bestätigt;
3. andererseits wurden auch Ausnahmen gefunden, die gute Kandidaten für eine zukünftige Kodifizierung darstellen.

Literatur

- Fisun, R. S. (2024) Zusammengesetzte Indefinitpronomen in den slavischen Sprachen (Doktorarbeit, Universität Regensburg)
- Fisun, R. & Hansen, B. (2025) A corpus based study on the inventory and the grammaticalisation of compound indefinite pronouns in Croatian. *Die Welt der Slaven* 69(2), 293-330. URL: https://www.researchgate.net/publication/390201462_A_corpus-based_study_on_the_inventory_and_grammaticalisation_of_compound_indefinite_pronouns_in_Croatian [25.05.2025]
- Lehmann, Ch. (2015[1982]). *Thoughts on Grammaticalization* (3rd edition). (Classics in Linguistics 1). Berlin
- Maykova, Irina (2023) Zur Funktionalstilistik ausgewählter Indefinitpronomen im Russischen. Publikationssever UR. URL: <https://epub.uni-regensburg.de/54410/2/Masterarbeit%20Maykova%20digitale%20Version.pdf> [25.05.2025]
- Ožegov, S. I. (o. D.). Tolkovyj slovar' Ožegova onlajn. URL: <https://slovarozhegova.ru/> [12.03.2023]
- Rybová, M. (2023) Die Beziehung zwischen Evaluativität und Indefinitheit am Beispiel der zusammengesetzten Indefinitpronomina im Tschechischen (Doktorarbeit, co-tutelle UK Prag – Universität Regensburg)
- Sketch Engine (o. D.). URL: <https://www.sketchengine.eu> [12.03.2023]