

Tilman Berger

NOCH EINMAL IMPERFEKTIVA IN HANDLUNGSFOLGEN¹

1. Einleitung

In diesem Beitrag möchte ich noch einmal auf ein Thema zurückkommen, zu dem viel Literatur existiert und zu dem ich selbst zweimal beigetragen habe (vgl. Berger 2013a, 2013b), nämlich auf die Verwendung von imperfektiven Verben in Handlungsfolgen im Tschechischen und anderen westslavischen Sprachen. Gegenstand dieses Beitrags soll eine quantitative Untersuchung der Häufigkeit dieser Konstruktionen im Tschechischen, Slovakischen, Polnischen und Russischen sein, wobei ich nicht nur die Verwendung des imperfektiven Aspekts untersuchen will, sondern auch die insbesondere für das Russische charakteristischen „Ersatzkonstruktionen“.

Im Folgenden will ich zunächst einen kurzen Überblick über die bisherige Forschung und ihre Ergebnisse geben, dann einige statistische Angaben aus der Literatur zitieren und davon ausgehend die Fragestellungen formulieren, die anhand von Korpora untersucht werden sollen. Dabei möchte ich neben den Nationalkorpora zu den vier Sprachen auch sog. Web-Korpora heranziehen, die sich für diese Fragestellung als besonders ergiebig erwiesen haben.

2. Die Ergebnisse der bisherigen Forschungen

Als erster hat 1961 Ivančev darauf hingewiesen, dass im Tschechischen (und anderen „westlichen“ slavischen Sprachen) in Handlungsfolgen auch imperfektive Verben verwendet werden können, während dies im Russischen (und anderen „östlichen“ slavischen Sprachen) nur ausnahmsweise möglich ist. Den tschechischen Konstruktionen des Typs *sedl si a četl* entsprechen nach Ivančev (1961, 36f.) komplexere Konstruktionen mit Phasenverben wie *начать/начинать, стать, пойтти* u. a. m., und Ivančev selbst vertrat die These, dass der imperfektive Aspekt im Tschechischen eine ingressive Funktion habe („контекстово обусловлена ингресивна употреба на глаголите на несвършен вид“). Diese

1 Ich bedanke mich bei Stefan Heck, Anja Gattnar, Eugen Kravchenko und Lorenz Geiger für ihre Unterstützung bei der Abfassung dieses Artikels.

Beschreibung deckt aber offenkundig nicht alle Möglichkeiten ab, und in der Literatur wurden deshalb verschiedene weitere Vorschläge gemacht. So spricht Křížková (1963, 287) davon, dass der Sprecher das Ereignis in der Weise stilisiere, dass man es nicht komplex mit Anfang und Ende erfassen könne, statt dessen „seien wir gewissermaßen mitten in ihm“, und Galton (1976, 68ff.) meint, das imperfektive Verb lasse die zweite Handlung unmittelbar und gewissermaßen „nahtlos“ an die ihr vorhergehende anschließen. Stunova (1993, 126f.) hat schließlich die schöne Metapher geprägt, nach der wir uns mit dieser Konstruktion gewissermaßen „in medias res“ begäben. Diesen Terminus verwendet dann auch Dickey (2000, 206), der das Phänomen noch einmal ausführlich diskutiert und es letztlich sehr allgemein als „contextually-conditioned impv past“ (CCIP) bezeichnet, ohne sich also auf eine spezielle Bedeutung wie etwa die ingressive festzulegen. Dickey beschreibt ansonsten im Detail, welche Alternativen im Russischen an die Stelle der tschechischen Konstruktion treten, und konzentriert sich hier auf zwei Möglichkeiten, nämlich erstens die Wiedergabe durch das ingressive Phasenverb *стать* und zweitens auf die Verwendung von ingressiven Verben mit dem Präfix *за-*. Auf beide Fälle werde ich im Folgenden noch genauer eingehen. Ich beziehe auch die ingressiven Verben mit dem Präfix *но-* ein, die Dickey in diesem Kontext nicht behandelt.

In meinen eigenen Artikeln von 2013 ging es mir darum, dass man zwei Typen von CCIP unterscheiden muss. Der häufigere Fall ist derjenige, in dem ein Verlauf auf ein Ereignis folgt, dieser ist in den meisten nordslavischen Sprachen reichlich belegt und nur für das Russische gelten die spezifischen, von Ivančev u. a. beschriebenen Restriktionen. Deutlich seltener ist der zweite Fall, in dem zwei Ereignisse aufeinanderfolgen, von denen aber das zweite mit einem imperfektiven Verb wiedergegeben wird. Diese Konstruktion hat eine spezifische „retardierende“ Funktion und ist nur im Tschechischen, Slovakischen und älteren Obersorbischen belegt. Der zweite Typ von CCIP soll in diesem Beitrag aber nicht behandelt werden, sondern ich konzentriere mich im Weiteren auf den „klassischen Fall“ des CCIP, also die Kombination von Ereignis + Verlauf.

In einem Beitrag von 2014 hat sich schließlich auch Zoroxina-Nilsson mit dieser Konstruktion beschäftigt. Sie beschäftigt sich sehr detailliert mit Fällen, in denen auch im Russischen das CCIP möglich ist, und zwar auch in unmarkierten Kontexten. Dabei geht es im Wesentlichen um zwei Typen von Beispielen, solche, in denen zwischen den beiden Handlungen keine scharfe Grenze existiert,² und solche, in denen mehr als zwei Verben aufeinanderfolgen.³ Auf beide Fälle werde ich im Laufe meiner Untersuchung noch einmal zurückkommen.

2 Vgl. den folgenden Beleg aus einem Text von Ulickaja: „В проходной шестого подъезда его остановили и *просили* оставить портфель.“ (Zoroxina-Nilsson 2014, 181).

3 Vgl. den folgenden Beleg aus einem Text von Gor’kij: „У нас, на Руси, никто не знает, зачем он. Родился, *жил*, помер – как все! Но – зачем?“ (Zoroxina-Nilsson 2014, 184).

3. Übersetzungsvergleich vs. Korpuslinguistik

Fast alle bisherigen Arbeiten zum CCIP bedienen sich des Verfahrens des Übersetzungsvergleichs. Das beginnt schon bei Ivančev und wurde von Stunova und Dickey in ähnlicher Weise fortgesetzt. In allen Fällen wurden in gedruckter Form vorliegende Übersetzungen mit den Originalen verglichen und gewissermaßen manuell ausgewertet. Als Ivančev und Stunova ihre Arbeiten verfassten, gab es noch keine elektronischen Korpora, und auch in der Entstehungszeit der großen Studie von Dickey waren sie erst in den Anfängen.

Ich bin in meinem ersten Artikel aus dem Jahr 2013, der sich ganz auf das Tschechische konzentrierte, hiervon abgewichen und habe Belege für die Kombination von perfektivem und imperfektivem Verb aus dem Tschechischen Nationalkorpus untersucht, konkret aus dem 100 Millionen Wortformen umfassenden Teilkorpus SYN2010. In meinem zweiten Artikel, der sich mit dem Vergleich mehrerer slavischer Sprachen beschäftigte, habe ich dann zwar mit dem Verfahren des Übersetzungsvergleichs gearbeitet, aber anhand des in elektronischer Form vorliegenden Parallelkorpus ParaSol (vgl. von Waldenfels 2011).

Der Übersetzungsvergleich hat sich als heuristisches Verfahren bewährt, weil man mit seiner Hilfe die Äquivalenz von Konstruktionen in verschiedenen Sprachen feststellen kann. Er ist aber nicht unbedingt geeignet, um die Verhältnisse in einer einzelnen Sprache umfassend zu charakterisieren – in den erwähnten Arbeiten liegt der Schwerpunkt ja auf den Unterschieden zwischen den Sprachen und nicht auf den Konstruktionen, die ihnen gemeinsam sind. Gleichzeitig muss man auch mit der Möglichkeit rechnen, dass Übersetzungen fehlerbehaftet sind, oder aber, dass literarische Übersetzer/innen sich Freiheiten nehmen, die dann zu einer stärkeren Abweichung vom Wortlaut führen. Ein sehr anschauliches Beispiel verdanke ich François Esvan, der dies im Jahr 2012 in einem Vortrag auf der Tagung *Gramatika a korpus 4* an Übersetzungen des CCIP in westeuropäische Sprachen demonstriert hat.

- (1a) „Ten váš příběh není nijak nebezpečný,“ uklidňoval ji vyslanec.
 „Snad ne,“ řekla zaškrceným hlasem a *odcházela* s Kareninem do nočních ulic Prahy.
- (1b) “Don’t worry,” the ambassador comforted her. “Your story doesn’t sound the least bit dangerous.”
 “I suppose it doesn’t,” she said in a tight voice, as she *walked out* into the Prague night with Karenin.
- (1c) L’ambassadeur tentait de la tranquilliser : « Votre mésaventure ne me paraît guère dangereuse ».
 « Ça se peut », dit-elle d’une voix étranglée, et elle *sortit* avec Karénine dans les rues noires de Prague.

- (1d) „Ihre Geschichte ist nicht im geringsten gefährlich“, versuchte sie der Botschafter zu beruhigen.
 „Schon möglich“, antwortete sie mit erstickter Stimme und *ging* mit Karenin *hinaus* auf die nächtlichen Straßen von Prag. (Kundera 1987, 157 f.).

Hier wird der spezielle retardierende Effekt des CCIP in einem Satz aus Kunderas Roman *Nesnesitelná lehkost bytí* (1a) weder bei der Übersetzung ins Englische (1b) noch ins Französische (1c) korrekt wiedergegeben. Dasselbe gilt auch für die deutsche Übersetzung (1d), die ich hier zusätzlich angebe.

Ein weiteres Problem besteht darin, dass es manchmal Asymmetrien zwischen den beiden Richtungen der Übersetzung gibt: So habe ich festgestellt, dass das CCIP zwar in polnischen Originaltexten vorkommt, dass aber bei der Übersetzung von russischen Konstruktionen mit einem Phasenverb ins Polnische meist das Phasenverb übernommen und kein CCIP eingesetzt wird (vgl. Berger 2013b, 63).

Angesichts solcher Probleme meine ich, dass die Ergebnisse des Übersetzungsvergleichs durch Korpusstudien ergänzt werden müssen. Und ich bin überzeugt, dass die großen elektronischen Korpora, die uns heute zur Verfügung stehen, nicht nur zur Auffindung von Beispielen geeignet sind (wie von mir in Berger 2013a praktiziert), sondern dass sie auch die Möglichkeit bieten, die Verhältnisse in unterschiedlichen Sprachen mit Hilfe statistischer Methoden zu vergleichen (worauf ich 2013 noch verzichtet habe).

In einem gewissen Umfang finden sich statistische Angaben auch schon in der bisherigen Literatur. Schon Ivančev äußert sich verschiedentlich dazu, dass bestimmte Konstruktionen häufiger vorkämen als andere,⁴ Stunova führt in dem Kapitel zum CCIP sogar einige Statistiken an (Stunova 1993, 112 und 115). In der einen Tabelle zeigt sie, dass im Tschechischen deutlich mehr imperfektive Verben in Ereignisfolgen vorkommen als im Russischen, in der zweiten Tabelle vergleicht sie das Auftreten bestimmter lexikalischer Gruppen. Die angeführten Zahlen illustrieren Stunovas Thesen gut, freilich werden sie auf der Grundlage sehr kleiner Grundgesamtheiten berechnet: Im ersten Fall legt Stunova 140 Ereignisfolgen aus ihrem aus Übersetzungen gewonnenen Beispielkorpus zugrunde, im zweiten Fall 486. Dickey zitiert schließlich einmal Stunova und berechnet ein weiteres Mal Zahlen auf der Grundlage von Ivančevs und Stunovas Beispie-

4 Vgl. etwa gleich zu Anfang seiner Arbeit: „Наред с това, но, струва ми се, по-рядко в чешкия език последвала начинателност на някакво действие след завършеност на друго действие се изразява и с така наречените фазови глаголи *začítí*, *počítí* и под.“ (Ivančev 1961, 5).

len – hier geht es um das Verhältnis der beiden Phasenverben *стать* und *начать*.⁵

Ich möchte in diesem Beitrag auf der Grundlage von großen Korpora (dazu mehr im folgenden Abschnitt) die folgenden drei Problemfelder statistisch untersuchen:

- 1) die Frequenz des CCIP (Abfolge perfektiv+imperfektiv) in den betrachteten Sprachen;
- 2) die Verwendung von Phasenverben in Konkurrenz zum CCIP;
- 3) die Verwendung ingressiver Verben in Konkurrenz zum CCIP.

Jedem Bereich soll ein eigener Abschnitt gewidmet sein, in dem ich auch genauer auf die Suchverfahren eingehe. Schon hier möchte ich aber darauf aufmerksam machen, dass nur die beiden ersten Fragen in befriedigender Weise behandelt werden können, die dritte dagegen nur mit Einschränkungen (vgl. dazu genauer am Anfang von Abschnitt 7). Zunächst will ich aber noch einige Bemerkungen zu den verwendeten Korpora machen.

4. Zu den verwendeten Korpora

Wenn man eine statistische Auswertung großer Textmengen plant, liegt nahe, dass man zu jeder Sprache möglichst das größte vorliegende Korpus heranzieht. Im Falle der hier betrachteten slavischen Sprachen bieten sich zunächst die sog. Nationalkorpora an, die schon eine längere Tradition haben (das Tschechische Nationalkorpus wurde 1994 begründet) und die sehr viele Texte umfassen. Am größten ist mit über zwei Milliarden Wortformen in schriftlichen Texten das Tschechische Nationalkorpus⁶ (die mündlichen Texte, das diachrone Korpus und diverse Sonderkorpora nicht eingerechnet), es folgt das Polnische Nationalkorpus mit ca. 1,8 Milliarden Wortformen⁷, das Slowakische Nationalkorpus mit 829 Millionen Wortformen⁸ und das Russische Nationalkorpus mit 364 Millionen Wortformen⁹.

Alle vier Nationalkorpora umfassen eine Vielzahl von Textsorten und sind linguistisch aufbereitet, d. h. zumindest morphologisch annotiert. Es wird auch die Möglichkeit angeboten, sich eigene Subkorpora zu definieren. Die Voraussetzungen für korpusgestützte linguistische Untersuchungen scheinen also her-

5 “Over half of the Cz impv forms in Ivančev’s examples are rendered in Ru with the phase verb *stat’* plus a main verb (to be precise, 63%). Of Stunová’s (1993) examples categorized as ingressive, 75% include *stat’*, whereas only 25% include *načat’*.” (Dickey 2000, 219).

6 Vgl. <http://wiki.korpus.cz/doku.php/cnk:uvod>.

7 Vgl. <http://www.nkjp.uni.lodz.pl/>.

8 Vgl. <http://korpus.juls.savba.sk/res.html>.

9 Vgl. <http://ruscorpورا.ru/corpora-stat.html>.

vorragend. Allerdings unterscheidet sich die Art der linguistischen Bearbeitung je nach nationaler Tradition deutlich, bis hinein in die zur Annotation verwendeten Tagsets, und die Korpora sind sehr unterschiedlich zusammengesetzt. Während im Falle des Tschechischen Nationalkorpus einige Mühen aufgewendet werden, ein repräsentatives oder gar ausgewogenes Korpus anzubieten, verwendet das Polnische Nationalkorpus vor allem Texte, die relativ einfach zu beschaffen sind und für die in der Korpuslinguistik der ironische Ausdruck „opportunistische Korpora“ üblich ist (vgl. Lemnitzer, Zinsmeister 2006, 107). Das Russische Nationalkorpus wiederum holt zeitlich sehr weit aus und enthält auch Texte aus dem frühen 19. Jahrhundert.

Die Unterschiede in der Annotation sind dagegen für die hier untersuchte Fragestellung eher unproblematisch (Verbformen sind – vor allem, wenn man sich auf das Präteritum beschränkt – deutlich einfacher zu annotieren als Substantive oder gar Pronomina). Gewisse Schwierigkeiten entstehen allerdings dadurch, dass die älteste Version des Tschechischen Nationalkorpus (SYN2000) nicht hinsichtlich des Aspekts annotiert war (s. u.).

Die unterschiedliche Zusammensetzung der Korpora wirft dagegen deutlich mehr Probleme auf, vor allem, weil die Annahme naheliegt, dass das CCIP in verschiedenen Textsorten unterschiedlich häufig vorkommen dürfte. Da es um ein Phänomen geht, dass vor allem in narrativen Texten zu erwarten ist, sollte es in belletristischen und möglicherweise auch publizistischen Texten häufiger vorkommen als in Fachtexten oder administrativen Texten. Ob dies tatsächlich so gilt, muss aber überprüft werden – damit wird sich ein Teil von Abschnitt 5 beschäftigen.

Letztlich wäre es das Beste, wenn es einen Typus von Korpus gäbe, der für alle hier behandelten Sprachen existiert und nach einem einheitlichen Prinzip erstellt wurde. In Frage kommen hier momentan vor allem sog. Web-Korpora, eine Art von Korpora, die aus dem Internet extrahiert werden, was u. a. ermöglicht, dass auf relativ einfache Weise sehr große Datenmengen gewonnen werden (vgl. hierzu Schäfer, Bildhauer 2013). Für die hier behandelten slavischen Sprachen hat Vladimír Benko von der Slowakischen Akademie der Wissenschaften die sog. Aranea-Korpora erstellt, diese sind sowohl bei der Akademie¹⁰ als auch über die Seite des Tschechischen Nationalkorpus zugänglich. Für das Tschechische und Slowakische liegen jeweils drei Korpora vor (Araneum Bohemicum Maximum, Maius und Minus sowie Araneum Slovacicum Maximum, Maius und Minus), für das Polnische und Russische je zwei (Maius und Minus). Für das Russische stellt Benko außerdem je zwei Korpora von Texten zur Verfügung, die ausschließlich in Russland (Araneum Russicum Russicum Maius und Minus) und ausschließlich außerhalb von Russland (Araneum Russicum Externum Maius und Minus) entstanden sind.

10 Unter dem Link <http://ucts.uniba.sk/>.

Web-Korpora sind in der Forschung durchaus umstritten,¹¹ und sie sind sicherlich nicht für jede linguistische Fragestellung geeignet, u. a. wegen ihres Oszillierens zwischen Schriftlichkeit und Mündlichkeit. Zur Untersuchung des CCIP erscheinen sie mir jedoch geeignet, insbesondere weil hier kein spezieller Einfluss der mündlichen Sprache zu erwarten ist.

Die Aranea-Korpora sind auch annotiert, allerdings nicht mit einem gemeinsamen Tagset, sondern für das Slovakische, Tschechische und Polnische mit dem Tagset, das im dortigen Nationalkorpus verwendet wird, und für das Russische mit dem sog. MULTEXT-East-Tagset¹². Die Verwendung unterschiedlicher Tagsets ist in diesem Fall unproblematisch, da wir uns für Verbformen interessieren, deren Annotation keine großen Schwierigkeiten mit sich bringt. Sehr problematisch ist jedoch die Tatsache, dass bei der Annotation des tschechischen Web-Korpus nicht der Aspekt berücksichtigt wurde¹³.

Nach reiflicher Überlegung habe ich mich entschlossen, trotzdem für die Untersuchung des Slovakischen, Polnischen und Russischen die jeweiligen Aranea-Korpora zu verwenden. Für das Tschechische verwende ich statt des Aranea-Korpus das Korpus SYN2013PUB – diese Wahl werde ich im Folgenden noch genauer begründen.

5. Die Frequenz des CCIP im Vergleich zu anderen Konstruktionen und ihre Variation

Dieses Kapitel teilt sich in drei Abschnitte. Im ersten werde ich mich mit der Frage beschäftigen, wie in slavischen Korpora nach Belegen für das CCIP und konkurrierenden Konstruktionen gesucht werden kann, im zweiten werde ich anhand der tschechischen Korpora untersuchen, inwieweit die Verwendung der Konstruktion von Textsorten abhängig ist. Im dritten Abschnitt soll schließlich die konkrete Verwendung des CCIP in den vier hier untersuchten Sprachen beschrieben werden.

11 Die Vorbehalte gegen die Verwendung von Google, wie sie Kilgariff (2007) formuliert hat, sind durch neuere Methoden wohl ausgeräumt, vgl. aber die Kritik von Aston (2011).

12 Dieses Tagset ist schon etwas älter und wurde von 1995–1997 entwickelt, eigentlich für eine größere Menge von Sprachen, darunter auch nichtslavische wie das Englische und Persische (vgl. <http://nl.ijs.si/ME/>). In den Aranea-Korpora wird es deshalb für das Russische verwendet, weil der entsprechende Tagger „open-source“ ist, anders als der Tagger des Russischen Nationalkorpus.

13 Dies liegt daran, dass nur der ältere Tagger MorphoDiTa, der den Aspekt noch nicht berücksichtigt, „open-source“ ist, nicht hingegen der ansonsten im Tschechischen Nationalkorpus verwendete neuere Tagger (persönliche Mitteilung von Vladimír Benko).

5.1. Die Suche nach dem CCIP und konkurrierenden Konstruktionen

Bei einer statistischen Untersuchung des CCIP stellt sich als erstes die Frage, mit welcher anderen Konstruktion es verglichen werden soll. Beim CCIP selbst liegt, wie oben dargestellt, stets die Verbindung eines perfektiven und eines imperfektiven Verbs vor, und zum Vergleich eignen sich dann die Verbindungen zweier perfektiver Verben. Dies erscheint vor allem deshalb sinnvoll, da ja nach Ivančev und anderen dem westslavischen CCIP im Russischen zumeist Konstruktionen mit perfektiven Phasenverben des Beginnens (also *начать* oder *стать*) entsprechen, d.h. wir erwarten im Prinzip unterschiedliche Aufteilungen der Kombination zweier Verben, von denen das erste perfektiv ist. Sehr schematisch könnte man sich die Lage folgendermaßen vorstellen:

Tschechisch:	pf + pf	pf + impf
Russisch:	pf + pf	pf + impf

Tab. 1: Schematische Aufteilung von Perfektiva und Imperfektiva nach vorherigem Perfektivum

Zu erwarten ist also für das Tschechische ein höherer Anteil der Verbindung „pf+impf“ gegenüber „pf+pf“.

Was die konkrete Suche angeht, so erscheint es mir zunächst sinnvoll, sich ganz auf das Präteritum zu beschränken. Weiterhin möchte ich nur nach Beispielen suchen, in denen die beiden Verben durch die Konjunktion *a* (bzw. im Polnischen und Russischen *i*) verbunden sind. In ihrer einfachsten Form lauten die beiden Suchformeln dann für das Tschechische folgendermaßen:

- (2a) [tag="Vp.....P"] "a" [tag="Vp.....I"]
 [tag="Vp.....P"] "a" [tag="Vp.....P"]

Der Tag V steht hier für die Wortart Verbum, der Tag p für das Präteritum¹⁴ und die beiden Tags P bzw. I bezeichnen den Aspekt.

Bei dieser einfachen Form der Suche fallen allerdings viele Beispiele durch das Raster, insbesondere diejenigen, bei denen zwischen der ersten Verbform und der Konjunktion ein Enklitikum steht, wie etwa in Ivančevs Beispiel „sedl si a četl“. Ähnliches ist auch in den anderen slavischen Sprachen zu erwarten, wobei sich die Verhältnisse jeweils etwas unterscheiden. So dürfte es im Russischen weniger Fälle geben, in denen zwischen Verb und Konjunktion etwas eingeschoben wird, weil dort das Präteritum ohne Auxiliar gebildet wird. Im Polni-

¹⁴ Statt dieses Tags kann auch an der neunten Position der Tag *R* verwendet werden.

schen erwarten wir dagegen mehr solche Fälle, weil heute das Flexionssuffix gewordene Auxiliar fast ausschließlich nach dem Verb steht.¹⁵

Sicherlich wäre es möglich, Suchformeln zu schreiben, die auch Fälle berücksichtigen, in denen zwischen dem ersten Verb und der Konjunktion eine Wortform eingeschoben ist. Damit würde sich die Suche aber deutlich komplizieren (weil genau festgelegt werden müsste, welche Elemente dort stehen können und welche nicht), ohne dass garantiert wäre, dass alle Fälle von CCIP gefunden werden können. Mir erscheint es sinnvoller, sich ganz bewusst auf die Teilmengen aller CCIP-Belege und der konkurrierenden Verbindungen zweier perfektiver Verben zu beschränken, in der das erste Verb, die Konjunktion und das zweite Verb aufeinanderfolgen, denn die Proportionen zwischen den beiden Klassen ändern sich nicht prinzipiell. Der einzige Schönheitsfehler bei dieser Vorgehensweise ist, dass keine Aussage über die relative Häufigkeit des CCIP in der Grundgesamtheit getroffen werden kann, sondern nur eine Aussage über das Verhältnis des CCIP zur konkurrierenden Konstruktion. Dies ist aber für den hier beabsichtigten Sprachvergleich nicht entscheidend.

Im Tschechischen und Slowakischen wird die Negationspartikel *ne-* immer mit dem Verb zusammengeschrieben, während ihr polnisches und russisches Äquivalent (*nie* bzw. *не*) als eigene Wortform gilt. Um die Vergleichbarkeit zu gewährleisten, wurde die Suchformel so modifiziert, dass in der Zweitposition nur affirmative Verbformen zugelassen werden. Die Suchformeln erhalten dadurch folgende Gestalt:

(2b) [tag="Vp.....P"] "a" [tag="Vp.....A....I"]
 [tag="Vp.....P"] "a" [tag="Vp.....A....P"]

Wenn man mit Hilfe dieser Formeln beispielsweise das Teilkorpus SYN2010 (das letzte „ausgewogene“ Teilkorpus) durchsucht, so erhält man die folgenden Ergebnisse:

(3) perfektiv + imperfektiv: 9725 Belege
 perfektiv + perfektiv: 39545 Belege

Wie wir sehen, sind die Verbindungen zweier perfektiver Verben ungefähr viermal so häufig wie das CCIP. Zur Charakterisierung verwende ich im Weiteren jeweils den Quotienten von pf+impf zu pf+pf – im vorliegenden Fall lautet er 0,245.

15 Orthografisch wird das Auxiliar mit dem Verb zusammengeschrieben, im Polnischen Nationalkorpus werden die entsprechenden Elemente (*em*, *m*, *eś*, *ś* usw.) aber als eigene Wortformen behandelt.

Wenn man einige der so gefundenen Belege durchsieht, stellt sich allerdings schnell heraus, dass hierunter auch Fälle sind, die man nicht als CCIP ansehen kann. Vgl. etwa die folgenden Belege aus SYN2010:

- (4a) Policisté místo zajistili a *byl* přivolán pyrotechnik, který si municí převzal.
 (4b) Cizinec vstal a *měl* se k odchodu.
 (4c) Pak se na policistu vrhl a *chtěl* ho strhnout k zemi [...].

In diesen Beispielen steht in der Zweitposition ein Verb, das zwar traditionell als imperfektiv eingeordnet wird, zu dem aber kein perfektives Pendant existiert. Teilweise geht es auch wie in (4a) um periphrastische Konstruktionen mit *být* oder (seltener) *mít*, oder das Verb ist Bestandteil eines Funktionsverbgefüges wie in (4b).

Die erste Suchformel wurde daher in der Form modifiziert, dass in der Zweitposition die Verben *být* und *mít* sowie die Modalverben *muset*, *mocet* und *chtít* ausgeschlossen wurden. Die Suchformel erhielt dadurch die etwas umfangreichere Form:

- (2a) [tag="Vp.....P"] "a" [tag="Vp.....A....I"&lemma!="být" & lemma!="mít" & lemma!="muset" & lemma!="mocet" & lemma!="chtít"]

Wendet man diese Formel an, so sinkt die Anzahl der Belege mit imperfektivem Zweitglied auf 8018 und der Quotient sinkt auf 0,202.

5.2. Verwendung der CCIP in Abhängigkeit von Textsorten

Mit Hilfe der beiden Formeln wurden nun sechs synchrone schriftliche Teilkorpora des Tschechischen Nationalkorpus untersucht, konkret die drei „ausgewogenen“ Korpora SYN2000, SYN2005¹⁶ und SYN2010 sowie die publizistischen Korpora SYN2006PUB, SYN2009PUB und SYN2013PUB. Die Ergebnisse der Suche stelle ich in einer Tabelle zusammen, die folgende Spalten enthält: Zunächst die absolute und relative Anzahl von Verbformen im Präteritum, die mit *a* verbunden sind (die relative Häufigkeit bezieht sich jeweils auf eine

¹⁶ In den beiden ersten „ausgewogenen“ Korpora war der Aspekt zwar nicht annotiert, das Nationalkorpus erlaubt aber auch die Durchsuchung einer neueren Version der Korpora, die Informationen über den Aspekt enthält.

Million Wortformen¹⁷), dann die absolute und relative Anzahl der beiden Fälle „pf+impf“ zu „pf+pf“¹⁸ und schließlich deren Quotient.

	V+V		pf+impf		pf+pf		Quotient
	absolut	i.p.m.	absolut	i.p.m.	absolut	i.p.m.	
SYN2000	44885	378,11	4492	37,84	20893	176,00	0,2150
SYN2005	77737	640,73	8003	65,96	36814	303,43	0,2173
SYN2010	78424	644,58	8018	65,90	39545	325,03	0,2027
SYN2006PUB	93232	258,10	8636	23,91	45093	124,83	0,1915
SYN2009PUB	239413	283,37	21356	25,28	115550	136,76	0,1848
SYN2013PUB	338849	302,54	29461	26,30	162794	145,35	0,1809
SYN	872629	324,99	77958	29,03	421420	156,95	0,1849

Tab. 2: Verhältnis „pf+impf“ : „pf+pf“ in sechs Teilkorpora des Tschechischen Nationalkorpus

Wenn wir versuchen, diese Zahlen zu interpretieren, so können wir zunächst feststellen, dass der Anteil von zwei Verbformen des Präteritums, die mit der Konjunktion *a* verbunden sind, zwischen den Korpora stark schwankt. Zwar kann festgehalten werden, dass solche Verbindungen in den publizistischen Korpora seltener sind als in den „ausgewogenen“ Korpora, aber innerhalb beider Gruppen von Korpora schwanken die Zahlen stark. Offenkundig sind die Verhältnisse in verschiedenen Textsorten sehr unterschiedlich, und die Kombination verschiedener Textsorten in Korpora erlaubt keine Verallgemeinerung.

Anders sieht die Sachlage aus, wenn wir nur die Beispiele der beiden Klassen und den Quotient ihrer Frequenzen betrachten. Hier sind die Zahlen sowohl innerhalb der ersten wie auch der zweiten Klasse von Korpora relativ ähnlich und man kann davon ausgehen, dass die Proportion der beiden Klassen innerhalb der „ausgewogenen“ Korpora wie auch publizistischen Korpora ungefähr gleich ist. Auf eine exakte statistische Analyse will ich hier aber verzichten, da ich letztlich nur mit einem der Korpora arbeiten werde, und beschränke mich auf den Hinweis, dass der Unterschied zwischen SYN2000 und SYN2005 nicht statistisch relevant ist, und ebenso auch der Unterschied zwischen SYN2009PUB und SYN2013PUB.

Welches der tschechischen Korpora kommt nun in seiner Verteilung den Netzkorpora am nächsten? Es erscheint mir sinnvoll, hier das letzte „ausgewo-

17 Ich verwende hier die im Tschechischen Nationalkorpus verwendete Abkürzung *i.p.m.*, die für *instances per million* steht.

18 Ich weise ausdrücklich darauf hin, dass der Wert in der zweiten Spalte nicht mit der Summe der vierten und der sechsten Spalte übereinstimmt, weil in ihm auch die durch die komplexere Suchformel ausgeschlossenen Verbindungen mit *být*, *mít* und Modalverben und negierte Verben in Zweitposition enthalten sind.

gene“ Korpus, also SYN2010 zu wählen. Es enthält eine Vielfalt von Textsorten und nicht nur publizistische Texte und es ist mit 120 Millionen Wortformen ungefähr genauso groß wie die kleinsten Webkorpora der Aranea-Gruppe.

5.3. Die Verwendung des CCIP im Tschechischen, Slovakischen, Polnischen und Russischen

Im Folgenden sollen nun die Ergebnisse für das Tschechische mit denen der drei anderen Sprachen verglichen werden. Bei diesen Sprachen gebe ich jeweils das kleinste Webkorpus (in der Terminologie von Benko „Araneum Minus“) und das mittlere Webkorpus („Araneum Maius“) an. Auf das größte Webkorpus („Araneum Maximum“) verzichte ich, weil es nicht für alle Sprachen vorliegt. Es wurde wie beim Tschechischen zunächst nach Verbindungen zweier Verbformen im Präteritum gesucht¹⁹, danach nach der Verbindung eines perfektiven Präteritums mit einem imperfektiven (unter Ausschluss der Verben ‘sein’ und ‘haben’ und der Modalverben)²⁰ und schließlich nach der Verbindung zweier perfektiver Präterita²¹. Die Ergebnisse sind in der folgenden Tabelle zusammengestellt:

	V+V		pf+impf		pf+pf		Quotient
	absolut	i.p.m.	absolut	i.p.m.	absolut	i.p.m.	
SYN2010	78424	644,58	8018	65,90	39545	325,03	0,2027
Slov. Minus	38105	315,31	3530	29,21	16755	138,64	0,2106
Slov. Maius	357897	298,21	33313	27,76	157237	131,01	0,2118
Poln. Minus	9496	79,90	485	4,08	5227	43,98	0,0927
Poln. Maius	3906	69,92	3919	3,27	46161	38,47	0,0848
Russ. Minus	25428	211,65	918	7,64	13698	114,02	0,0607
Russ. Maius	244968	204,14	8737	7,28	131971	109,98	0,0662

Tab. 3: Verhältnis „pf+impf“ : „pf+pf“ in slovakischen, polnischen und russischen Webkorpora

19 Hier lautet die Suchformel für das Slovakische [tag="VL..*"] [lemma="a"] [tag="VL....+"], für das Polnische [tag="praet.*"] [lemma="i"] [tag="praet.*"] und für das Russische [tag="Vmis.*"] [lemma="и"] [tag="Vmis.*"].

20 Hier lautet die Suchformel für das Slovakische [tag="VLd.*"] [lemma="a"] [tag="VLe...+\&lemma!="byt"&lemma!="mat"&lemma!="musiet"&lemma!="môct"&lemma!="chciet"], für das Polnische [tag="praet.*:perf.*"] [lemma="i"] [tag="praet.*:imperf.*" & lemma!="być"&lemma!="mieć"&lemma!="musieć"&lemma!="móc"&lemma!="chcieć"] und für das Russische [tag="Vmis....p"] [lemma="и"] [tag="Vmis....e"&lemma!="быть"&lemma!="мочь"&lemma!="хотеть"].

21 Hier lautet die Suchformel für das Slovakische [tag="VLd.*"] [lemma="a"] [tag="VLd...+\&lemma!="byť"&lemma!="mieć"&lemma!="musieć"&lemma!="móc"&lemma!="chcieć"], für das Polnische [tag="praet.*:perf.*"] [lemma="i"] [tag="praet.*:perf.*"] und für das Russische [tag="Vmis....p"] [lemma="и"] [tag="Vmis....p"].

Wenn wir die Ergebnisse genauer betrachten, so stellt sich als erstes die Frage, ob die Verhältnisse in den für alle Sprachen (außer dem Tschechischen) angeführten zwei Korpora homogen sind. Diese Frage kann positiv beantwortet werden, wie eine Anwendung des Chi-Quadrat-Tests zeigt.²² Ich werde deshalb im Weiteren auch nur noch die jeweils größten Korpora (der Maius-Reihe) berücksichtigen und auf die kleineren Korpora verzichten.

Weiterhin stellen wir fest, dass der Anteil der Verwendung des imperfektiven Aspekts erkennbar von Westen nach Osten hin abnimmt. Dabei liegen Tschechisch und Slowakisch sehr nahe beieinander, Polnisch nimmt eine Zwischenstellung ein und Russisch liegt am anderen Ende der Skala. Interessant ist, dass es den Anschein hat, als sei das CCIP im Slowakischen noch etwas ausgeprägter als im Tschechischen. Angesichts der Probleme mit dem tschechischen Korpus würde ich das zunächst nicht überwerten (die beiden Sprachen lassen sich derzeit nicht wirklich objektiv vergleichen), allerdings weisen weitere Daten, die im Verlauf der Untersuchung erhoben wurden, in eine ähnliche Richtung.

Weiterhin können wir feststellen, dass das Polnische dem Russischen näher steht als dem Tschechischen und Slowakischen, auch dies passt gut zu den Ergebnissen meines Artikels aus dem Jahr 2013 (Berger 2013b). Etwas überraschend ist aber auf der anderen Seite der geringe Anteil von Verbindungen zweier Verben im Polnischen. Hier müsste im Einzelnen noch geklärt werden, ob dies tatsächlich nur an der Position der Enklitika liegt (s. o.) oder ob noch weitere Faktoren hierfür verantwortlich sind.

6. Die Frequenz von Phasenverben in Konkurrenz zum CCIP

Als nächstes möchte ich nun die Frequenz der Phasenverben in ihrem Verhältnis zum CCIP untersuchen. Zu diesem Zweck habe ich in den Korpora ermittelt, wie häufig jeweils ein perfektives Verb mit einem Verb des Anfangens + Infinitiv kombiniert wird. Ich untersuche auch hier wieder nur den einfachsten Fall, also etwa für das Tschechische die Kombination eines perfektiven Verb im Präteritum mit der Konjunktion *a* und dem Präteritum des Verbs *začít* ‘anfangen’ + Infinitiv²³. Es ist klar, dass auch hier wieder eine Reihe von relevanten Beispielen herausfallen, etwa weil zwischen *začít* und dem Infinitiv weitere Wortformen stehen, aber ich gehe davon aus, dass dieser Fall in allen behandelten Sprachen gleichermaßen auftritt und die Ergebnisse deshalb nicht verfälscht

22 Im Falle des Slowakischen beträgt die Wahrscheinlichkeit, dass beide „Stichproben“ aus derselben Grundgesamtheit stammen, 78,51%, im Falle des Russischen 74,70%. Im Falle des Polnischen ist der Unterschied zwischen den beiden Korpora am größten, die Wahrscheinlichkeit, dass beide Stichproben aus derselben Grundgesamtheit stammen, liegt mit 8,165% aber immer noch über der üblichen Grenze von 5%.

23 Die Suchformel lautet: [tag="Vp.....P"] "a" [tag="Vp.....A....P" & lemma="začít"] [tag="Vf.*"].

werden. In analoger Weise habe ich auch im slovakischen und im polnischen Korpus gesucht, jeweils für die Verben *začat*²⁴ und *zacząć*²⁵. Dabei ist zuzugestehen, dass es auch weitere Verben des Anfangens gibt, etwa im Tschechischen das veraltete Verb *počít* und im Polnischen *począć* und *rozpocząć*, aber diese verbinden sich vernachlässigbar selten mit dem Infinitiv. Anders sieht es im Russischen aus, wo *начать* und *стать* miteinander konkurrieren²⁶. Die folgende Tabelle enthält daher für das Russische drei Zeilen, je eine für *начать* und *стать* und eine für die Summe der beiden Verwendungen. In der letzten Spalte steht schließlich der Quotient, der das Verhältnis der Konstruktion mit dem Verb 'anfangen' zum CCIP angibt:

	pf+'anfangen'+Inf			pf+impf		Quotient
	Verb	absolut	i.p.m.	absolut	i.p.m.	
SYN2010	začít	643	5,28	8018	65,90	0,0801
Slov. Korpus	začať	2376	1,98	33313	27,76	0,0713
Poln. Korpus	zacząć	937	0,78	3919	3,27	0,2390
Russ. Korpus	начать	1821	1,52	8737	7,28	0,2084
	стать	2523	2,10	"	"	0,2887
	Summe	4344	3,62	"	"	0,4971

Tab. 4: Frequenz der Phasenverben in ihrem Verhältnis zum CCIP

Diese Ergebnisse zeigen wieder anschaulich die Unterschiede zwischen den vier betrachteten Sprachen, wobei auch hier das Slovakische die Extremposition mit dem stärksten Übergewicht des CCIP einnimmt, wenn auch in geringem Abstand vom Tschechischen. Das Polnische steht hier deutlich zwischen Tschechisch/Slovakisch und Russisch, im Russischen wird die große Bedeutung der Verben des Anfangens überdeutlich. Das Verhältnis von *начать* und *стать* ist hingegen etwas ausgeglichener, als die Angaben von Stunova und Dickey erwarten lassen. Wie oben erwähnt, kommt Dickey in seiner Berechnung auf einen Anteil von 63% für *стать*, und Stunova sogar auf 75%, in dem von mir herangezogenen Korpus beträgt die Proportion von *стать* zu *начать* 58:42.

In der obigen Tabelle wurden die Verben des Anfangens nur mit dem CCIP verglichen. Man kann sich aber auch die Frage stellen, wie hoch ihr Anteil in-

24 Hier lautet die Suchformel: [tag="VLd.*"] [lemma="a"] [tag="VLd...\|+" & lemma="začať"] [tag="VI.*"].

25 Und hier lautet die Formel: [tag="praet.*:perf.*"] [lemma="i"] [tag="praet.*:perf.*" & lemma="zacząć"] [tag="inf.*"].

26 Die Suchformeln lauten:

[tag="Vmis....p"] [lemma="и"] [tag="Vmis....p" & lemma="начать"] [tag="Vmn.*"] bzw. [tag="Vmis....p"] [lemma="и"] [tag="Vmis....p" & lemma="стать"] [tag="Vmn.*"]

nerhalb der Kombination von zwei perfektiven Verben im Präteritum aussieht. Diese Zahlen finden sich in der nächsten Tabelle:

	pf+pf	'anfangen ₁ '		'anfangen ₂ '	
		absolut	prozentual	absolut	prozentual
SYN2010	39545	643	0,0162	–	–
Slovak. Korpus	157237	2376	0,0151	–	–
Poln. Korpus	46161	937	0,0202	–	–
Russ. Korpus	131971	1821	0,0137	2523	0,0191

Tab. 5: Anteil der Anfangsverben in der Kombination zweier perfektiver Verben im Präteritum

Wir sehen hier, dass das „normale“ Verbum des Anfangens (also *začít, začat'* und *начать*) im Tschechischen, Slovakischen und Russischen relativ ähnliche Frequenzen aufweist. Das passt sehr gut zu Dickey's Aussage (2000, 220f.), dass gerade *стать* mit einer speziellen, von *начать* abweichenden Bedeutung (vgl. hierzu auch Flank 1987 und Paillard, Fici Giusti 1998) besonders prädestiniert ist, das tschechische und slowakische CCIP wiederzugeben. Im Polnischen, das mit *zacząć* nur über ein Verbum des Anfangens verfügt, muss dieses – obwohl im Polnischen ja auch das CCIP belegt ist (allerdings seltener als im Tschechischen und Slovakischen) – eine höhere Frequenz aufweisen, weil es sowohl die Funktion von russ. *начать* wie auch bis einem gewissen Grade die von russ. *стать* erfüllt.

7. Die Frequenz ingressiver Verben in Konkurrenz zum CCIP

Die Untersuchung der mit dem Präfix *za-* gebildeten ingressiven Verben in den vier Korpora stößt leider auf gewisse Schwierigkeiten. Es ist zwar sicher richtig, dass man im Russischen faktisch jedes intransitive Verb mit *za-* präfigieren kann (vgl. Dickey 2000, 223, mit Verweis auf Avilova 1976, 278), während entsprechende Bildungen im Tschechischen und Slovakischen fehlen und im Polnischen sehr selten sind (vgl. Dickey 2000, 225–227). Aber da das Präfix *za-* noch eine Reihe weiterer Funktionen hat, kann man nicht einfach im Korpus nach Verben suchen, die mit *za-* beginnen.

Um zumindest eine gewisse Abschätzung zu ermöglichen, welche Rolle die mit *za-* gebildeten Ingressiva im Russischen spielen, habe ich im Korpus „Araneum Russicum Maius“ zunächst ermittelt, welche mit *za-* gebildeten perfektiven Verben im Korpus überhaupt vorkommen.²⁷ Hier ergab sich eine Liste von über 2500 präfigierten und als perfektiv annotierten „Lexemen“, die freilich von Hand nachbearbeitet werden musste. Und unter den hundert ersten Verben der

27 Die Suchformel hierfür lautet: [tag="V.....p" & lemma="за.*"].

Liste waren dann überhaupt nur sieben eindeutige Ingressiva.²⁸ Die folgende Tabelle fasst zusammen, wie häufig sie als Zweitglied auftreten:²⁹

Lexem	Frequenz im Korpus		in der Verbindung pf+pf	
	absolut	i.p.m.	absolut	i.p.m.
заболеть	19593	16,33	48	0,04
заговорить	15880	13,23	118	0,10
загореться	5337	4,45	53	0,04
закричать	7992	6,66	156	0,13
замолчать	5074	2,23	104	0,09
заплакать	4616	3,85	121	0,10
засмеяться	4348	3,62	46	0,04

Tab. 6: Häufigkeit der mit *за-* gebildeten russischen Ingressiva als Zweitglied

Aus der Tabelle wird deutlich, dass die Verwendung ingressiver Verben mit dem Präfix *за-*, deren prinzipielle Bedeutung natürlich nicht bestritten werden soll, zahlenmäßig nicht ins Gewicht fällt.

Anders sieht es bei den Ingressiva mit dem Präfix *но-* aus, die Dickey in seiner Studie von 2000 nicht erwähnt und die auch in einem späteren Beitrag zu den *но-*Präfigierungen nur kurz vorkommen (vgl. Dickey 2011, 194). Der Autor sieht diese insbesondere von imperfektiven determinierten Verben der Bewegung gebildeten Ableitungen synchron offenbar als deren Aspektpartner an und erwähnt deshalb nicht speziell ihre ingressiv Bedeutung. Genau diese *но-*Ingressiva haben aber in der Verwendung als Ersatzkonstruktionen für das CCIP eine relativ hohe Frequenz:

28 Das Verb *захотеть* und sein reflexives Gegenstück *захотеться* wurde nicht berücksichtigt, weil ich oben die Modalverben ganz ausgeschlossen habe. Ferner wurden die Verben *забить* und *заработать* ausgeklammert, weil sie nur selten mit ingressiver Bedeutung auftreten – diese fehlt im Übrigen auch bei beiden Verben im Verblexikon von Apresjan und Páll (1982).

29 Für die Suche wurde hier die Formel [tag="Vmis.....p"] [lemma="и"] [tag="Vmis.....p" & lemma="заболеть"] verwendet. An die Stelle von *заболеть* traten dann jeweils auch die anderen Verben.

Lexem	Frequenz im Korpus		in der Verbindung pf+pf	
	absolut	i.p.m.	absolut	i.p.m.
пойти	235385	196,15	2510	2,09
поехать	55830	46,52	685	0,57
повезти	30757	25,63	60	0,05
повести	15827	13,19	143	0,12
побежать	12501	10,42	514	0,43

Tab. 7: Häufigkeit der mit *no-* gebildeten russischen Ingressiva als Zweitglied

Die Bedeutung der mit *no-* gebildeten Ingressiva als Ersatzkonstruktion ist also wesentlich größer als diejenige der mit *za* gebildeten Ingressiva (allein *noÿmu* kommt in dieser Funktion ungefähr viermal so häufig wie alle oben angeführten *za*-Verben zusammen!). Ein detaillierter Vergleich, wie ich ihn für die Verben des Anfangens geliefert habe, kann aber leider angesichts der fehlenden Suchmöglichkeiten nicht durchgeführt werden.

8. Zusammenfassung

Die quantitativen Untersuchungen anhand von slovakischen, polnischen und russischen Webkorpora sowie des tschechischen Korpus SYN2010 haben im Wesentlichen die Aussagen der bisherigen Arbeiten bestätigt. Dies betrifft zunächst das Verhältnis der vier behandelten slavischen Sprachen: Tschechisch und Slowakisch verhalten sich im Hinblick auf die Verwendung von Imperfektiva in Handlungsfolgen nahezu identisch, Russisch verhält sich dezidiert anders und das Polnische nimmt eine Zwischenposition ein.

Weiterhin ist deutlich geworden, dass als Ersatzkonstruktion für das tschechische und slowakische CCIP im Russischen vor allem Verbindungen mit dem Verb *стать* verwendet werden, während das andere Verb des Anfangens, *начать*, in ähnlichem Umfang verwendet wird wie tschechisch *začít* und slowakisch *začať*. Im Polnischen steht hingegen das CCIP (das aber deutlich seltener vorkommt als im Tschechischen und Slowakischen) neben einer leicht erhöhten Frequenz des Verbs *zacząć*.

Eine detaillierte Untersuchung der Verwendung von Ingressiva (mit den Präfixen *za-* und *no-*) im Russischen als Ersatzkonstruktion für das CCIP war hingegen nicht möglich, da die Korpora keine entsprechenden Suchmöglichkeiten zulassen. Es ist aber deutlich geworden, dass die *no*-Ingressiva eine wesentlich wichtigere Rolle als Ersatzkonstruktion für das CCIP spielen als die in der Literatur häufiger behandelten *za*-Ingressiva.

Literatur

- Apresjan, Ju. D., Páll, E. 1982. *Russkij glagol, vengerskij glagol*. Bd. 1, 2. Budapest.
- Aston, G. 2011. Applied Corpus Linguistics and the learning experience. In: Viana V., Zyngier S., Barnbrook G. (eds.) *Perspectives on Corpus Linguistics*. Amsterdam, 1–16.
- Avilova, N. S. 1976. *Vid glagola i semantika glagol'nogo vida*. Moskva.
- Benko, V. 2014. Aranea: Yet Another Family of (Comparable) Web Corpora. In: Sojka P. et al. (eds.) *Text, Speech and Dialogue. 17th International Conference*. Brno, 257–264.
- Berger, T. 2013a. Ungewöhnliche Verwendungen des Aspekts im Tschechischen – der imperfektive Aspekt in Handlungssequenzen. *Zeitschrift für Slawistik* 58, 31–42.
- Berger, T. 2013b. Imperfektive Verben in Handlungsfolgen im Westslavischen. In: Kempgen S. et al. (eds.) *Deutsche Beiträge zum 15. Internationalen Slavistenkongress in Minsk*. München, 57–66.
- Dickey, S. M. 2000. *Parameters of Slavic Aspect. A Cognitive Approach*. Stanford.
- Dickey, S. M. 2011. The Varying Role of *po-* in the Grammaticalization of Slavic Aspectual Systems: Sequences of Events, Delimitatives, and German Language Contact. *Journal of Slavic Linguistics* 19, 175–230.
- Esvan, F. 2012. Ke střídání vidu při vyprávění v češtině. Handout zum Vortrag auf der Tagung *Gramatika a korpus 4*.
- Flink, S. 1987. Phase Subdivisions and Russian Inceptives. *Die Welt der Slaven* 32, 310–316.
- Galton, H. 1976. *The main functions of the Slavic verbal aspect*. Skopje.
- Ivančev, S. 1961. *Kontekstovo obuslovljena ingresivna upotreba na glagolite ot nesvāršen vid v českija ezik*. Sofija. (Godišnik na Sofijskija universitet, Filologičeski fakultet 65/3).
- Kilgariff, A. 2007. Googleology is bad science. *Computational Linguistics* 33, 147–151.
(<http://www.mitpressjournals.org/doi/pdf/10.1162/coli.2007.33.1.147>)
- Lemnitzer, L., Zinsmeister, H. 2006. *Korpuslinguistik. Eine Einführung*. Tübingen.
- Paillard, D., Fici Giusti, F. 1998. L'inchoation en russe: entre préverbes et auxiliaires. *Le Langage et l'Homme* 33, 79–94.
- Schäfer, R., Bildhauer, F. 2013. *Web Corpus Construction*. San Francisco.
- Stunova, A. 1993. *A contrastive study of Russian and Czech aspect: Invariance vs. discourse*. Amsterdam.
- von Waldenfels, R. 2011. Recent developments in ParaSol: Breadth for depth and XSLT based web concordancing with CWB. *Natural Language Processing, Multilinguality. Proceedings of Slovko 2011*. Bratislava, 156–162.
- Zoroxina-Nilsson, N. 2014. Nesoveršennyj vid i sukcesivnost'. K voprosu o netrivial'nych kontekstach v russkom jazyke v plane prošlogo. *Scando-Slavica* 60, 172–188.