

**Gerd Hentschel, Jan Patrick Zeller, Sviatlana Tesch**

**Das Oldenburger Korpus zur weißrussisch-russischen  
gemischten Rede: OK-WRGR**

Dokumentation

*Stand: 01.04.2014*



BIS-Verlag der Carl von Ossietzky Universität Oldenburg

ISBN: 978-3-8142-2305-6

## Inhaltsverzeichnis

1	Einleitung	1
2	Die Sprecher, ihr sozialer Hintergrund und die Anteile an den Daten der Korpora	5
2.1	Der Anteil einzelner Sprecher an den Daten	5
2.1.1	Im Familienkorpus	5
2.1.2	Im Interviewkorpus	7
2.2	Städte und Sprecher	8
2.2.1	Im Familienkorpus	8
2.2.2	Im Interviewkorpus	12
2.3	Zur Altersstruktur der Respondenten und zum Anteil der Altersgruppen am Korpus	14
2.3.1	Im Familienkorpus	14
2.3.2	Im Interviewkorpus	16
2.4	Zur Beteiligung der Geschlechter	18
2.4.1	Im Familienkorpus	18
2.4.2	Im Interviewkorpus	20
2.5	Zur Bildung	22
2.5.1	Im Familienkorpus	22
2.5.2	Im Interviewkorpus	24
2.6	Zur Land-Stadt-Migration	26
2.6.1	Im Familienkorpus	26
2.6.2	Im Interviewkorpus	28
2.7	Fazit	29
3	Die Struktur der Transkriptionen im System CHAT	30
3.1	Dateinamen	30
3.2	Zur Grobstruktur der Transkriptionen	31
3.3	Übergeordnete Informationen	32
3.3.1	@Languages: Die beteiligten Sprechertypen	32
3.3.2	@Participants: Die Rollen bzw. der Migrationsstatus der beteiligten Sprecher	33
3.3.3	@ID: Die Sprecherinformationen	35
3.3.3.1	Sprechertyp	35
3.3.3.2	Korpus	35
3.3.3.3	Sprecherkürzel	36
3.3.3.4	Beteiligung am Korpus	37
3.3.3.5	Migrationsstatus / Rolle im Interview	37
3.3.3.6	Ausbildung	37
3.3.4	@L1: Die Erstsprache	37
3.3.5	@FL: Die Familiensprache in der Kindheit	38
3.3.6	@MT: Die Muttersprache	38
3.3.7	@Comment: Kommentare zur Transkription	38
3.3.8	@Situation: Die Gesprächssituation	38
3.4	Die Wiedergabe der Gespräche	39
3.4.1	Einteilung in Äußerungen	39
3.4.2	Transkription der Wortformen	40

3.4.2.1	Vokale in betonter Stellung	41
3.4.2.2	Vokale in unmittelbar vorbetonter Stellung	42
3.4.2.3	Vokale in anderen vortonigen und in nachtonigen Stellungen	43
3.4.2.4	Sonderfälle bei Vokalen	44
3.4.2.5	Grundsätzliches zu Konsonanten	45
3.4.2.6	Palatalisiertheit	45
3.4.2.7	Konsonantische Unterschiede zwischen dem Weißrussischen und dem Russischen	46
3.4.2.8	Zur Wiedergabe der Ergebnisse einiger phonologischer Prozesse und weiterer Phänomene	47
3.4.2.9	Besondere Wortformen	48
3.4.3	Die sprachliche Affinität	49
3.4.3.1	Phonische Ebene	50
3.4.3.2	Morphologische Ebene	52
3.4.4	Weitere wortformbezogene Informationen	56
3.4.5	Interpunktion	57
3.4.5.1	Einfache Satzzeichen	57
3.4.5.2	Direkte Rede	57
3.4.5.3	Metasprachliche Referenz	58
3.4.5.4	Wiederholungen, Korrekturen, Umformulierungen, Übersetzungen	59
3.4.5.5	Unvollständige Sätze	60
3.4.5.6	Pausen und Überlappungen im Redefluss	61
3.5	Die Annotationen in der Zeile %mor	61
3.5.1	Die sprachliche Affinität	62
3.5.2	Wortart	63
3.5.3	Lemma	65
3.5.4	Deklination	69
3.5.4.1	Kasus	69
3.5.4.2	Genus	70
3.5.4.3	Numerus	70
3.5.5	Konjugation	71
3.5.5.1	Tempus und Imperativ	71
3.5.5.2	Person	71
3.5.5.3	Genus	72
3.5.5.4	Numerus	72
4	Arbeit mit dem OK-WRGR in CLAN	73
4.1	Einleitung	73
4.2	Vorbereitung des Korpus	73
4.3	Suche nach Wortformen	75
4.4	Suche nach Annotationen in %mor	76
4.5	Suche nach Interpunktionszeichen	77
4.6	Eingrenzung der Suche auf Teilkorpora / Städte / Rollen	78
5	Literatur	79

## 1 Einleitung

Die weißrussisch-russische gemischte Rede, das gemeinsame Auftreten von aus der Außenperspektive „russischen“ und „weißrussischen“ Elementen und Strukturen in engen Skopen der Rede, ist ein Phänomen, das zumindest in bestimmten Gesprächssituationen bei einer großen Anzahl von Weißrussen anzutreffen ist. Das *Oldenburger Korpus zur weißrussisch-russischen gemischten Rede* (OK-WRGR) ist das erste große Korpus zu diesem aktuellen Phänomen des Sprachkontakts zweier eng verwandter Sprachen. Das Korpus entstand im Rahmen des von der Volkswagenstiftung geförderten Projekts *Die Trasjanka in Weißrussland – Eine Mischvarietät als Folge des weißrussisch-russischen Sprachkontakts. Sprachliche Strukturierung, soziologische Identifikationsmechanismen und Sozioökonomie der Sprache*, welches von 2008 bis 2013 von Gerd Hentschel (Institut für Slavistik, Universität Oldenburg) und Bernhard Kittel (Institut für Sozialwissenschaften, Universität Oldenburg / Institut für Wirtschaftssoziologie, Universität Wien) in Kooperation mit Sjarhej Zaprudski (Abteilung für die Geschichte der Weißrussischen Sprache, Weißrussische Staatsuniversität Minsk) und David Rotman (Zentrum für Soziale und Politische Forschungen, Weißrussische Staatsuniversität Minsk) durchgeführt wurde.

Die weißrussisch-russische gemischte Rede ist trotz ihrer weiten Verbreitung in Belarus lange von der Sprachwissenschaft nicht beachtet worden. Erst seit den 1990-er Jahren ist sie Gegenstand sprachwissenschaftlicher Betrachtung, wobei das Verdienst, als erste dieses Phänomen in die wissenschaftliche Diskussion eingeführt zu haben, Henadz' Cychun und Nina Mečkovskaja zukommt (vgl. z.B. CYCHUN 2013 [1998], MEČKOVSKAJA 2006, MJAČKOŪSKAJA 2007). Üblicherweise wird diese Form der Rede in Belarus mit dem Terminus „Trasjanka“ belegt, ursprünglich eine Bezeichnung für mit Stroh gemischtes Heu, also für „schlechtes“, da gestrecktes Viehfutter (vgl. zum Hintergrund dieses Begriffs im Detail CYCHUN 2013 [1998]). Da dieser Begriff – wie die dahinterstehende Metaphorik bereits andeutet – stark negativ konnotiert ist und auch zur Stigmatisierung derer gereicht, die sich „der Trasjanka“ bedienen, verwenden wir im Folgenden durchgehend die Bezeichnung „weißrussisch-russische gemischte Rede“ bzw. abgekürzt „WRGR“. Dies ist auch in folgender Hinsicht angemessen: Von manchen wird die „Trasjanka“ als diejenige Rede definiert, die bei Indi-

viduen zu beobachten ist, die weder das Weißrussische noch das Russische in einer „akzeptablen“ Form beherrschen. Wenn dann umgekehrt die „Trasjanka“ als ein Merkmal wenig gebildeter Sprecher bezeichnet wird, so liegt natürlich ein Zirkelschluss vor (vgl. HENTSCHEL 2013). Unsere Untersuchungen zeigen, dass die Sprecher der WRGR eine sehr viel heterogenere Gruppe bilden und dass die gemischte Rede eben auch von Menschen mit adäquater Beherrschung des Russischen, partiell des Weißrussischen, verwendet wird. Die mangelnde Bildung von Sprechern „der Trasjanka“ ist ein negatives Stereotyp (vgl. KITTEL et al. 2010; HENTSCHEL & KITTEL 2011a/b).

Das OK-WRGR umfasst Datenmaterial aus Aufnahmen weißrussisch-russischer gemischter Rede, die in den Jahren 2007 bis 2008 in insgesamt acht weißrussischen Städten aufgezeichnet wurden (vgl. HENTSCHEL 2008c, 2013). Dabei handelt es sich zum einen um spontane Gespräche im Kreise von weißrussischen Familien und deren Freundeskreisen (das *Familienkorpus*), zum anderen um offene Interviews mit „bekennenden“ Sprechern der gemischten Rede (das *Interviewkorpus*).

Bei dem *Familienkorpus* handelt es sich um nicht-offizielle, spontane Gespräche im Familien- und Freundeskreis innerhalb jeweils einer Familie pro Stadt<sup>1</sup>. Gesprächsteilnehmer sind also Mitglieder der jeweiligen Familie, aber auch weitere Verwandte, Freunde, Nachbarn und evtl. Arbeitskollegen, mit denen die ausgesuchten Familien engen Kontakt haben. Die Aufnahmen wurden bei teilnehmender Beobachtung durchgeführt. Um ein möglichst authentisches Redeverhalten zu gewährleisten, gehörten die Personen, welche die Aufnahmen durchführten, selbst zu den ausgewählten Familien oder zu deren Freundeskreis. Das Aufnahmegerät wurde dabei an einem geeigneten Ort in der Wohnung der Familie platziert. Einige Respondenten wussten vom eingeschalteten Aufnahmegerät, andere wussten während der Aufnahme nicht davon und wurden nachträglich informiert. Da das Aufnahmegerät in der Regel über mehrere Stunden eingeschaltet war, kann auch bei den Sprechern, die prinzipiell von der Aufnahme wussten, von einer weitgehend unbewussten, ungezwungenen Gesprächsführung ausgegangen werden. Ausgewertet wurden Aufnahmen im Umfang von ca. 36 Stunden, und zwar Passagen, in denen die gesprochene Sprache ohne Zweifel ungezwungen war. So wurden zum Beispiel Gesprächsabschnitte unmittelbar nach Einschalten des Aufnahmegeräts nicht ausgewertet. Ebenso nicht berücksichtigt wurden

---

1 In einem Fall, der Stadt Akcjabrski, wurden die Daten in zwei Familien erhoben.

Passagen, die von Störgeräuschen beeinträchtigt sind, in denen viele Personen am Gespräch teilnehmen oder in denen intime oder in anderer Hinsicht heikle Inhalte besprochen werden.<sup>2</sup>

Das *Interviewkorpus* enthält Sprachmaterial aus offenen Interviews, die im Rahmen der sprachsoziologischen Untersuchung des Oldenburger Projekts entstanden sind. Im Zentrum der sprachsoziologischen Untersuchung standen dabei zunächst geschlossene Interviews mit 1.400 zufällig ausgewählten Respondenten (vgl. KITTEL et al. 2010; HENTSCHEL & KITTEL 2011a/b). Von diesen 1.400 Befragten wurden einige Sprecher ausgewählt, die in den geschlossenen Interviews angegeben hatten, die WRGR als primäres Mittel der Kommunikation zu benutzen. Mit diesen Sprechern wurden offene Interviews durchgeführt, die einen weiteren Teil der sprachsoziologischen Komponente des Oldenburger Projekts ausmachen (vgl. KITTEL & LINDNER 2011). Für das Interviewkorpus wurden aus diesen offenen Interviews gezielt Passagen ausgewählt, die als deutliche Instanzen gemischter Rede zu erkennen sind. Die interviewten Respondenten standen in keiner Beziehung zu den Sprechern des Familienkorpus und kannten sich auch untereinander nicht.

Insgesamt enthält das Familienkorpus ca. 212.000 Wortformen bzw. 39.000 Äußerungen in 444 Gesprächen von sehr unterschiedlicher Länge. In das Interviewkorpus sind ca. 170.000 Wortformen in 23.000 Äußerungen aus 54 Gesprächen eingegangen. Eine Übersicht über die Sprecher und ihre soziodemographischen Charakteristika bietet Abschnitt 2.

Das OK-WRGR ist auf der Seite <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> frei verfügbar. Zum einen können einige Audiodateien heruntergeladen werden. Zum anderen liegt das Korpus in drei Versionen in verschriftlichter Form vor:

- Erstens ist eine einfache Version direkt über den Browser lesbar; bei einigen Gesprächen sind hier auch die entsprechenden Audiodateien direkt abspielbar. Diese Version ist (mit einigen Modifikationen, s.u.) in den im Rahmen des CHILDES-Projekts entwickelten Konventionen von CHAT<sup>3</sup> transkribiert. Sie ist auf einfache Lesbarkeit „am Bildschirm“ ausgerichtet und enthält daher keine Annotationen zur sogenannten sprachlichen Affinität (grob: kann die

---

<sup>2</sup> In diversen Auswertungen durch die Projektbeteiligten wurden letztere jedoch nicht ausgeschlossen.

<sup>3</sup> Nach dem Stand der CHAT-Konventionen im Jahr 2013; vgl. MCWHINNEY 2013a.

Wortform dem Weißrussischen oder Russischen zugeordnet werden, ist sie eine „gemeinsame“ oder eine „hybride“; s. Abschnitt 3.4.3) oder zu grammatischen Werten. Sie ist lediglich mit den üblichen Instrumenten eines Internetbrowsers durchsuchbar.

- Zweitens ist der Download des OK-WRGR im CHAT-Format in Form von UTF8-kodierten Textdateien mit den genannten Annotationen zur Grammatik und sprachlichen Affinität möglich. Diese Version des OK-WRGR ist mit den Werkzeugen des Programms CLAN analysierbar (vgl. MCWHINNEY 2013b). Hierauf wird in den Abschnitten 3 und 4 eingegangen.
- Drittens werden zum 15.05.2014 die beiden Teilkorpora des OK-WRGR für lediglich an den Texten interessierte Personen als durchsuchbare PDF-Dateien bereitgestellt werden. Diese Version enthält den reinen Text in der hier beschriebenen Transkription ohne die formalen Kennzeichnungen des CHAT-Formats und ohne Indizierungen von beispielsweise grammatischen Werten. Informationen zu sozialen Charakteristika der jeweils beteiligten Sprecher werden in einer dem Leser zugänglicheren Form als im CHAT-Format gegeben. Anders als in den anderen Versionen erfolgt auch die Kennzeichnung der Wortbetonung.

Das OK-WRGR ist eine eigenständige Publikation. Es ist folgendermaßen zitierbar:

HENTSCHEL, G.; ZELLER, J. P.; TESCH, S. 2014: *Das Oldenburger Korpus zur weißrussisch-russischen gemischten Rede: OK-WRGR*. Oldenburg: BIS [<http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/>] [ISBN: 978-3-8142-2305-6]

Eine Übersicht über die im Rahmen des Oldenburger Projekts auf Basis von Auswertungen des OK-WRGR entstandenen Publikationen (sowie über weitere im Oldenburger Projekt entstandene Publikationen) findet sich auf der Seite <https://www.uni-oldenburg.de/slavistik/forschung/sprachwissenschaft/-schwerpunkt-mischvarietäten/oldenburger-publikationen-zur-weissrussisch-russischen-gemischten-rede/>.

## **2 Die Sprecher, ihr sozialer Hintergrund und die Anteile an den Daten der Korpora**

### **2.1 Der Anteil einzelner Sprecher an den Daten**

#### *2.1.1 Im Familienkorpus*

Insgesamt haben 129 Respondenten Daten zum Korpus von Familiengesprächen beigetragen. Wie es dem Charakter der Kommunikation in Familien entspricht, ist der Beitrag der einzelnen Personen sehr unterschiedlich. In der Regel nehmen von allen Personen, die in der jeweiligen häuslichen Gemeinschaft leben, zwei bis drei extensiver an Gesprächen teil als andere. Bei den üblichen familiären Sozialstrukturen sind das oft die Frauen in der Generation, die den Kern der Familie bildet. Andere Familienmitglieder nehmen weniger umfangreich teil. Darüber hinaus sind Personen involviert, die nicht selbst zur häuslichen Gemeinschaft gehören, sondern sich zu Besuch bei ihr aufgehalten haben. Dazu gehören weitere Familienmitglieder sowie Freunde und Bekannte, oft Nachbarn oder Arbeitskollegen. Daraus resultiert, dass der Umfang der Beteiligung an den Gesprächen stark variiert. Den minimalen Beitrag hat ein Respondent mit nur einer einzigen Wortform geleistet, das Maximum ein anderer mit 12.529 Wortformen.<sup>4</sup> (Da die durchschnittliche Länge einer Äußerung, die meist, aber nicht immer, einem Satz entspricht, ca. fünf bis sechs Wortformen umfasst, liegen von letztgenanntem Respondenten also ca. 2.000 Äußerungen vor.) Folgende Abbildung illustriert die Anzahl der Respondenten, die jeweils einen bestimmten Beitrag zum Korpus geleistet haben:

---

<sup>4</sup> Die quantitativen Angaben in diesem Abschnitt beziehen sich auf die im Oldenburger Projekt verwendete Version des Korpus, in die auch gewisse kurze Passagen intimen oder „heiklen“ Inhalts eingehen, welche in der veröffentlichten Version des Korpus nicht wiedergegeben sind. Für die allgemeinen quantitativen Verhältnisse spielt dieser Unterschied jedoch keine Rolle.



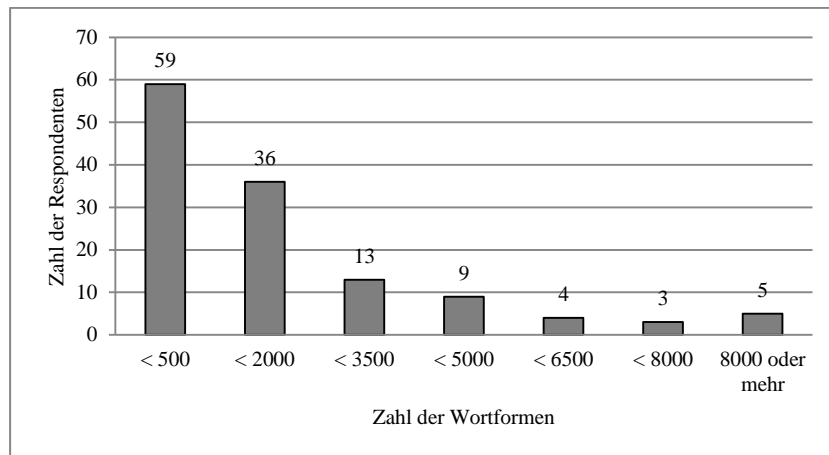


Abb. 1 Anzahl und Umfang der Beteiligung (in Zahl der Wortformen) einzelner Respondenten im Familienkorpus

Angesichts dieser Unterschiede ist es sinnvoll, die 129 beteiligten Respondenten in zwei Gruppen zu unterteilen. Als „periphere Respondenten“ werden die 59 Personen gefasst, die weniger als 500 Wortformen zum Datenbestand beigetragen haben. Alle anderen fallen in die Gruppe der „zentralen Respondenten“, die 70 Personen umfasst. Letztere haben 205.938 der insgesamt vorliegenden 212.279 Wortformen beigetragen (also grob 97 Prozent).

Für die Analyse des sprachlichen Materials spielt der soziale Hintergrund der Respondenten natürlich eine wesentliche Rolle. Um die Bedeutung einzelner sozialer Parameter für die sprachliche Strukturierung der erhobenen Daten prüfen zu können, wurden mit Hilfe eines Fragebogens entsprechende Informationen über die Respondenten erhoben. Dies geschah in der Regel aber nur für die 70 „zentralen“ Respondenten. Die Ausprägungen der entsprechenden Parameter bei den „peripheren“ Sprechern können die Daten aus quantitativen Gründen nicht beeinflussen. Durch widrige Umstände an einem der Erhebungsorte, Akcjabrski, sind für drei der zentralen Respondenten die erhobenen Daten zum sozialen Hintergrund verloren gegangen. Wenn wir diese drei Sprecher vernachlässigen, liefern die verbleibenden 67 zentralen Respondenten 203.110 Wortformen, d.h. knapp 96 Prozent.

### 2.1.2 *Im Interviewkorpus*

Die dem Interviewkorpus zugrunde liegenden Interviews wurden (mit zwei Ausnahmen, s.u.) an denselben Orten wie das Familienkorpus erhoben, und zwar primär zum Zweck qualitativer sozialwissenschaftlicher Analysen (vgl. diejenige in KITTEL & LINDNER 2011). Ausgewählt wurden dafür Respondenten, die sich in den geschlossenen Interviews, die der Untersuchung von HENTSCHEL & KITTEL (2011) zugrunde liegen, dazu bekannten, als primäres Kommunikationsmedium die WRGR zu verwenden. Aus diesen Interviews wurden Fragmente, die Äußerungen in der WRGR in dichter Folge zeigen,<sup>5</sup> genau wie im Familienkorpus transkribiert und indiziert. Wie eingangs erwähnt, standen die interviewten Respondenten in keiner Beziehung zu den Familien des Familienkorpus und waren auch untereinander nicht bekannt.

Im Interviewkorpus ist natürlich zu differenzieren zwischen Äußerungen der Respondenten, und solchen, die von den Interviewern getätigt wurden. Zudem sind in einigen Gesprächen vereinzelt auch Äußerungen anderer Personen aus dem Umfeld der Respondenten vertreten, die während des Interviews anwesend waren. Von den insgesamt 167.523 Wortformen des Interviewkorpus stammen 141.570, also ca. 85 Prozent, von den Respondenten.

Der Anteil einzelner Respondenten an den Sprachdaten variiert im Interviewkorpus natürlich weniger stark als im Familienkorpus. Insgesamt wurden für 54 Respondenten Fragmente aus den halb- bis einstündigen Interviews ausgewählt. Dies sind, wie gesagt, Fragmente, die eine intensive Mischung der Rede zeigen, wohingegen Passagen in russischer Sprache (in der Regel mit mehr oder weniger deutlichen phonischen Interferenzen des Weißrussischen), also der in Weißrussland generell unmarkierten Varietät zur Kommunikation mit Fremden, nicht berücksichtigt wurden. Die folgende Abbildung zeigt die Verhältnisse:

---

<sup>5</sup> Die Bereitschaft der Respondenten, sich der WRGR zu bedienen, war am Anfang der bis zu einstündigen Interviews geringer als im weiteren Verlauf. Ansonsten bediente man sich im Gespräch mit den fremden Interviewern des Russischen, nie des Weißrussischen.

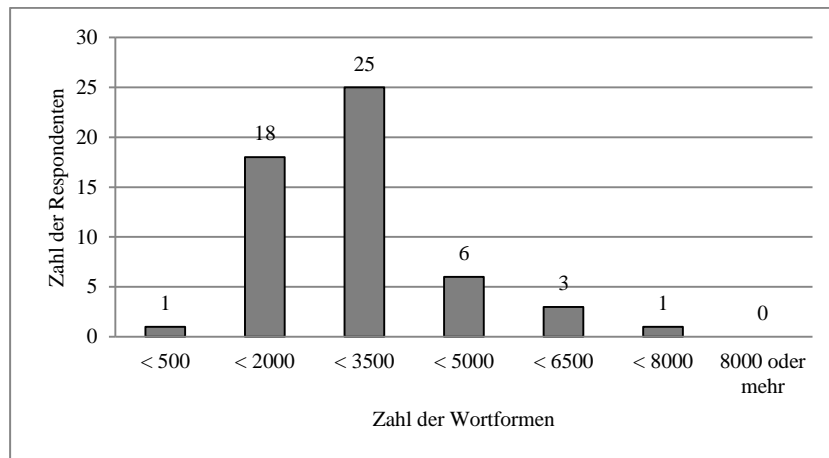


Abb. 2 Anzahl und Umfang der Beteiligung (in Zahl der Wortformen) einzelner Respondenten im Interviewkorpus

Auch für die Respondenten im Interviewkorpus liegen Angaben zum sozialen Hintergrund vor, worauf in den folgenden Abschnitten eingegangen wird.

## 2.2 Städte und Sprecher

### 2.2.1 Im Familienkorpus

Bei den im Familienkorpus untersuchten Städten handelt es sich um sechs kleinere bis mittelgroße Städte, die sich auf die drei traditionell angenommenen weißrussischen Dialektgebiete (das nordöstliche, das südwestliche und das zentrale Übergangsgebiet) verteilen, sowie um die Hauptstadt Minsk (die geographisch im zentralen Dialektgebiet liegt, aber dialektologisch natürlich einen Sonderfall darstellt).<sup>6</sup> In jedem Gebiet wurde eine eher westlich und eine eher östlich gelegene Stadt untersucht:

<sup>6</sup> Viele Respondenten aus Minsk stammen ursprünglich aus dem nordöstlichen Dialektgebiet Weißrusslands.

Tab. 1 Die im Familienkorpus untersuchten Städte und ihre Einwohnerzahl (Stand 2009)

Nordosten		Zentral			Südwesten	
westl.	östl.	westl.	zentral	östl.	westl.	östl.
Šarkoŭ- ščyna (Sh)	Chocimsk (Ch)	Smarhon' (Sm)	Minsk (Mi)	Rahačoŭ (Ra)	Barana- vičy (Ba)	Akcjabr- ski (Ak)
6.900	7.100	36.300	1.836.800	33.700	168.200	7.400
Bezirk Vicebsk	Bezirk Mahilëŭ	Bezirk Hrodna	Bezirk Minsk (Stadt)	Bezirk Homel'	Bezirk Brest	Bezirk Homel'

Auch abgesehen von der Hauptstadt Minsk, die sich als Metropole und Millionenstadt von den anderen Städten extrem abhebt, bestehen einige Unterschiede in der Struktur der untersuchten Städte (vgl. HENTSCHEL & KITTEL 2011). Während Smarhon' und Rahačoŭ mittelgroße Städte sind und Baranavičy eine im weißrussischen Vergleich sogar relativ große Stadt ist, sind die übrigen drei, Šarkoŭščyna, Chocimsk und Akcjabrski, eher klein und offensichtlich ländlich geprägt.

Der ursprüngliche Plan der Datenerhebung hatte das Ziel, für jede der sieben Städte bzw. jeden der sieben Familienkontexte ein Teilkorpus von 25.000 Wortformen zu erheben. Es sollten möglichst Respondenten aus zwei Generationen erfasst werden, dabei Frauen und Männer in Quantitäten, die Vergleiche erlauben. Bei der gewählten Art der Datenaufnahme (die – wie eingangs dargelegt – den Vorteil hat, weitgehend ungezwungene Rede zu erfassen und somit dem Beobachterparadoxon ebenso weitgehend entgegenzuwirken) war dies nicht direkt zu steuern. In vier Städten (Chocimsk, Minsk, Rahačoŭ, Smarhon') musste zusätzliches Material erhoben werden, um deutliche Abweichungen vom genannten Ziel wenigstens abzumildern. Dadurch kamen für diese Städte Sprachaufzeichnungen im Umfang von ca. 8.000 bis 10.000 Wortformen hinzu:

Tab. 2 Wortformen pro Stadt im Familienkorpus

<b>Stadt</b>	<b>Anzahl der Wortformen</b>	<b>in Prozent</b>	<b>Kumulierte Prozente</b>
<b>Ak</b>	24946	11,8	11,8
<b>Ba</b>	25386	12,0	23,7
<b>Ch</b>	34212	16,1	39,8
<b>Mi</b>	33987	16,0	55,8
<b>Ra</b>	33023	15,6	71,4
<b>Sh</b>	25263	11,9	83,3
<b>Sm</b>	35462	16,7	100,0
<b>Gesamt</b>	<b>212279</b>	<b>100,0</b>	

Der Umfang des Datenmaterials, gemessen in Wortformen, variiert zwischen 11,8 und 16,7 Prozent und ist für die Städte somit relativ ausgewogen.

Unterschiedlich ist die Zahl der Sprecher in den einzelnen Erhebungs-orten. Die folgende Tabelle illustriert dies und differenziert gleichzeitig drei Sprechergruppen: die peripheren Respondenten mit jeweils weniger als 500 Wortformen und die zentralen, hier mit zwei Untergruppen: diejenigen mit 500 bis 4.999 Wortformen und diejenigen mit mehr als 5.000.

Tab. 3 Anzahl der Sprecher pro Stadt mit  $n$  Wortformen im Familienkorpus

Stadt	weniger als 500 Wortformen	500 bis 4999	5000 und mehr	
<b>Ak</b>	14	14	0	28
<b>Ba</b>	6	4	2	12
<b>Ch</b>	2	8	2	12
<b>Mi</b>	7	9	2	18
<b>Ra</b>	0	9	1	10
<b>Sh</b>	4	5	2	11
<b>Sm</b>	26	9	3	38
<b>Gesamt</b>	<b>59</b>	<b>58</b>	<b>12</b>	<b>129</b>

Da die Zahl der Wortformen aber auch innerhalb der drei Gruppen stark variieren kann, ist es aufschlussreicher festzustellen, welche Untermengen der Teilkorpora von den drei hier unterschiedenen Sprechergruppen beigetragen wurden. Dies zeigt die folgende Abbildung:

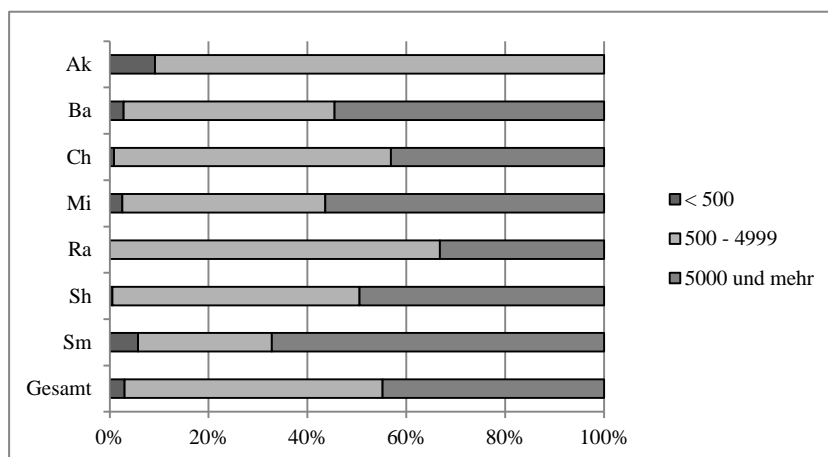


Abb. 3 Anteile von Respondenten mit einer Menge von  $n$  Wortformen am Umfang der Teilkorpora im Familienkorpus

Gemeinsam ist allen sieben Teilkorpora, dass periphere Sprecher, die insgesamt und in den meisten Städten die Mehrheit der Respondenten ausmachen, überall lediglich weniger als zehn Prozent des Datenmaterials beitragen. Deutlich unterschieden sind die Teilkorpora jedoch im Hinblick auf „quantitativ dominante“ Respondenten. Als solche seien Sprecher definiert, die 5.000 Wortformen oder mehr beigetragen haben. Während es in Akcjabrski keinen quantitativ dominanten Respondenten gibt, zeigen die anderen Erhebungsorte einen, zwei oder drei. Die Anteile dieser Sprecher variieren dann zwischen zwei Dritteln in Smarhon', wo drei dominante Sprecher gegeben sind, und ca. einem Drittel in Rahačoŭ, wo nur ein dominanter Sprecher auszumachen ist. In den anderen Städten, in denen jeweils zwei Respondenten dominieren, liegen die Werte entsprechend zwischen einem und zwei Dritteln. Je höher der Anteil der dominanten Sprecher an den Teilkorpora, desto stärker ist das jeweilige Sprachmaterial individuell geprägt.

### 2.2.2 *Im Interviewkorpus*

Das Interviewkorpus umfasst dieselben Städte wie das Familienkorpus, mit zwei Ausnahmen: Zum einen musste Baranavičy durch das benachbarte Slonim ersetzt werden, eine Stadt mit knapp 50.000 Einwohnern. Zum anderen konnte Minsk im Interviewkorpus nicht berücksichtigt werden, da es nicht gelang, von den dortigen Respondenten in nennenswertem Umfang Sprachdaten in der WRGR zu evozieren. Obwohl sich die betreffenden Respondenten dazu bekannt hatten, regelmäßig die WRGR zu praktizieren, sprach man in den Interviews russisch.

Tab. 4 Die im Interviewkorpus untersuchten Städte und deren Einwohnerzahl (Stand 2009)

Nordosten		Zentral		Südwesten	
westl.	östl.	westl.	östl.	westl.	östl.
Šarkoŭščyna (Sh)	Chocimsk (Ch)	Smarhon' (Sm)	Rahačoŭ (Ra)	Slonim (Sl)	Akcjabrski (Ak)
6.900	7.100	36.300	33.700	49.000	7.400
Bezirk Vicebsk	Bezirk Mahilëŭ	Bezirk Hrodna	Bezirk Homel'	Bezirk Hrodna	Bezirk Homel'

In jeder Stadt wurde eine mit dem Familienkorpus vergleichbare Menge an sprachlichem Material erhoben, d.h. zwischen etwa 25.000 und 30.000 Wortformen.

Tab. 5 Wortformen pro Stadt im Interviewkorpus

<b>Stadt</b>	<b>Anzahl der Wortformen</b>	<b>in Prozent</b>	<b>Kumulierte Prozente</b>
<b>Ak</b>	29073	17,4	17,4
<b>Sl</b>	29159	17,4	34,8
<b>Ch</b>	29708	17,7	52,5
<b>Ra</b>	25035	14,9	67,4
<b>Sh</b>	24778	14,8	82,2
<b>Sm</b>	29770	17,8	100,0
<b>Gesamt</b>	167523	100,0	

Der Anteil der Städte am sprachlichen Material des Interviewkorpus schwankt zwischen ungefähr 15 und 18 Prozent, ist also relativ ausgewogen.

Die Zahl der untersuchten Sprecher pro Stadt oszilliert im Interviewkorpus nicht so stark wie im Familienkorpus. Folgende Tabelle zeigt die Zahl der Sprecher aus den sechs Städten:

Tab. 6 Respondenten pro Stadt im Interviewkorpus

<b>Stadt</b>	<b>Anzahl der Respondenten</b>	<b>in Prozent</b>	<b>Kumulierte Prozente</b>
<b>Ak</b>	7	13,0	25,9
<b>Sl</b>	13	24,1	50,0
<b>Ch</b>	11	20,4	70,4
<b>Ra</b>	9	16,7	87,0
<b>Sh</b>	7	13,0	100,0
<b>Sm</b>	7	13,0	100,0
<b>Gesamt</b>	54	100,0	

Natürlich ist es relevant, welchen Anteil die Respondenten und welchen die Interviewer (und sich eventuell sporadisch am Gespräch beteiligende sonstige Sprecher) am Sprachmaterial jeweils aufweisen. Dies zeigt die folgende Tabelle:



Tab. 7 Anteile der Respondenten, Interviewer und sonstigen Sprecher am Sprachmaterial (in Wortformen) im Interviewkorpus

<b>Stadt</b>		<b>Respondenten</b>	<b>Interviewer</b>	<b>Sonstige</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	21776	7297	0	29073
	%	74,9%	25,1%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	26801	2907	0	29708
	%	90,2%	9,8%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	21526	3509	0	25035
	%	86,0%	14,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	20403	3289	1086	24778
	%	82,3%	13,3%	4,4%	100,0%
<b>Sl</b>	<i>N</i>	25122	3739	298	29159
	%	86,2%	12,8%	1,0%	100,0%
<b>Sm</b>	<i>N</i>	25942	3828	0	29770
	%	87,1%	12,9%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	141570	24569	1384	167523
	%	84,5%	14,7%	0,8%	100,0%

Insgesamt stammen 141.570 Wortformen, d.h. ungefähr 85 Prozent des sprachlichen Materials im Interviewkorpus, von den Respondenten. Nur in Akcjabrski sinkt der Anteil unter 80 Prozent. Die Angaben im Folgenden beziehen sich nur auf die von den Respondenten produzierten Sprachdaten.

## 2.3 Zur Altersstruktur der Respondenten und zum Anteil der Altersgruppen am Korpus

### 2.3.1 Im Familienkorpus

Die 70 „zentralen“ von insgesamt 129 Respondenten lassen folgende Altersverteilung in den sieben Städten erkennen (für drei Respondenten aus Akcjabrski fehlen die Altersangaben<sup>7</sup>):

<sup>7</sup> Auch die weiteren Angaben zum sozialen Hintergrund fehlen für diese drei Sprecher.

Tab. 8 Die zentralen Respondenten im Familienkorpus nach Geburtsjahr und Stadt

Stadt		vor 1961	1961-1980	nach 1980	k.A.	Gesamt
<b>Ak</b>	<i>N</i>	1	6	4	3	14
	%	7,1%	42,9%	28,6%	21,4%	100,0%
<b>Ba</b>	<i>N</i>	4	2	0	0	6
	%	66,7%	33,3%	0,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	2	6	2	0	10
	%	20,0%	60,0%	20,0%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	5	3	3	0	11
	%	45,5%	27,3%	27,3%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	5	2	3	0	10
	%	50,0%	20,0%	30,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	2	1	4	0	7
	%	28,6%	14,3%	57,1%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	2	7	3	0	12
	%	16,7%	58,3%	25,0%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	21	27	19	3	70
	%	30,0%	38,6%	27,1%	4,3%	100,0%

Deutlich wird, dass die Altersstruktur in den einzelnen Städten bzw. Familienkontexten sehr unterschiedlich ist. Insgesamt sind die Altersgruppen jedoch relativ gleichmäßig besetzt.

Wesentlicher jedoch als die Zahl der Respondenten ist der Anteil der Altersgruppen am sprachlichen Material selbst. Dieser soll im Folgenden an der Zahl der Wortformen verdeutlicht werden, wiederum differenziert nach den sieben Städten. Von den insgesamt vorliegenden 212.297 Wortformen entfallen 205.938 (also 97 Prozent) auf die 70 „zentralen Sprecher“ des Familienkorpus. Letztere verteilen sich, differenziert nach Städten, wie folgt auf die Altersklassen:

Tab. 9 Anteil der Altersgruppen am analysierten Datenmaterial (gemessen in Wortformen), differenziert nach Städten im Familienkorpus (nur zentrale Respondenten)

<b>Stadt</b>		<b>vor 1961</b>	<b>1961-1980</b>	<b>nach 1980</b>	<b>k.A.</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	1432	14162	4245	2828	22667
	%	6,3%	62,5%	18,7%	12,5%	100,0%
<b>Ba</b>	<i>N</i>	17455	7212	0	0	24667
	%	70,8%	29,2%	0,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	4293	24763	4854	0	33910
	%	12,7%	73,0%	14,3%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	18427	10310	4383	0	33120
	%	55,6%	31,1%	13,2%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	22553	2461	8009	0	33023
	%	68,3%	7,5%	24,3%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	6332	5257	13534	0	25123
	%	25,2%	20,9%	53,9%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	1415	29810	2203	0	33428
	%	4,2%	89,2%	6,6%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	71907	93975	37228	2828	205938
	%	34,9%	45,6%	18,1%	1,4%	100,0%

Die Teilkorpora in den einzelnen Erhebungsorten sind also recht unterschiedlich von Sprechern unterschiedlicher Altersgruppen beeinflusst. Während in Baranavičy, Minsk und Rahačoŭ das meiste Material von Sprechern der älteren Altersgruppe kommt, prägt in Akcjabrski, Chocimsk und Smarhon' (im letzten Fall sehr stark) die mittlere Altersgruppe das Bild. In Šarkoŭščyna schließlich dominiert das Material der jüngsten Sprechergruppe.

### 2.3.2 *Im Interviewkorpus*

Die 54 Respondenten des Interviewkorpus lassen folgende Altersverteilung in den sechs Städten erkennen:

Tab. 10 Die Respondenten nach Geburtsjahr und Stadt im Interviewkorpus

<b>Stadt</b>		<b>vor 1961</b>	<b>1961-1980</b>	<b>nach 1980</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	6	1	0	7
	%	85,7%	14,3%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	5	8	0	13
	%	38,5%	61,5%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	7	4	0	11
	%	63,6%	36,4%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	3	5	1	9
	%	33,3%	55,6%	11,1%	100,0%
<b>Sl</b>	<i>N</i>	5	2	0	7
	%	71,4%	28,6%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	6	1	0	7
	%	85,7%	14,3%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	32	21	1	54
	%	59,3%	38,9%	1,9%	100,0%

Die größte Gruppe bilden die älteren Respondenten. Nur ein einziger Respondent ist der jüngsten Altersgruppe zuzuordnen. Zwischen den Städten bestehen allerdings Unterschiede. In Smarhon' und Akcjabrski überwiegen die älteren Respondenten deutlich, in Rahačov und Slonim leicht; in Chocimsk und Šarkouščyna macht die mittlere Altersgruppe den größten Anteil aus.

Was das sprachliche Material (bezogen auf die Anzahl der von den Respondenten geäußerten Wortformen) angeht, so fällt die Verteilung ähnlich aus:

Tab. 11 Anteil der Altersgruppen am analysierten Datenmaterial im Interviewkorpus (gemessen in Wortformen), differenziert nach Städten

<b>Stadt</b>		<b>vor 1961</b>	<b>1961-1980</b>	<b>nach 1980</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	19320	2456	0	21776
	%	88,7%	11,3%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	11701	15100	0	26801
	%	43,7%	56,3%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	12460	9066	0	21526
	%	57,9%	42,1%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	6496	11871	2036	20403
	%	31,8%	58,2%	10,0%	100,0%
<b>Sl</b>	<i>N</i>	15382	9740	0	25122
	%	61,2%	38,8%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	23538	2404	0	25942
	%	90,7%	9,3%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	88897	50637	2036	141570
	%	62,8%	35,8%	1,4%	100,0%

Auch auf die Anzahl der Wortformen bezogen platzieren sich die älteren Sprecher vor der mittleren Altersgruppe. Die jüngste Generation ist nur marginal vertreten.

## 2.4 Zur Beteiligung der Geschlechter

### 2.4.1 Im Familienkorpus

Insgesamt sind unter den 70 „zentralen“ Respondenten des Familienkorpus etwa doppelt so viele Frauen wie Männer:

Tab. 12 Die Respondenten nach Geschlecht und Stadt im Familienkorpus

<b>Stadt</b>		<b>weiblich</b>	<b>männlich</b>	<b>k.A.</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	8	3	3	14
	%	57,1%	21,4%	21,4%	100,0%
<b>Ba</b>	<i>N</i>	4	2	0	6
	%	66,7%	33,3%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	7	3	0	10
	%	70,0%	30,0%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	7	4	0	11
	%	63,6%	36,4%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	7	3	0	10
	%	70,0%	30,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	5	2	0	7
	%	71,4%	28,6%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	7	5	0	12
	%	58,3%	41,7%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	45	22	3	70
	%	64,3%	31,4%	4,3%	100,0%

Noch deutlicher dominiert der Anteil der Frauen bei der Anzahl der geäußerten Wortformen:

Tab. 13 Anteil der Geschlechter am analysierten Datenmaterial im Familienkorpus (gemessen in Wortformen), differenziert nach Städten (nur zentrale Respondenten)

<b>Stadt</b>		<b>weiblich</b>	<b>männlich</b>	<b>k.A.</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	13847	5992	2828	22667
	%	61,1%	26,4%	12,5%	100,0%
<b>Ba</b>	<i>N</i>	18576	6091	0	24667
	%	75,3%	24,7%	0,0%	100,0%

<b>Ch</b>	<i>N</i>	21559	12351	0	33910
	%	63,6%	36,4%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	29451	3669	0	33120
	%	88,9%	11,1%	0,0%	100,0%
<b>Ra</b>	<i>N</i>	26781	6242	0	33023
	%	81,1%	18,9%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	17056	8067	0	25123
	%	67,9%	32,1%	0,0%	100,0%
<b>Sm</b>	<i>N</i>	20366	13062	0	33428
	%	60,9%	39,1%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	147636	55474	2828	205938
	%	71,7%	26,9%	1,4%	100,0%

Die Dominanz gilt für alle Städte, wobei sie mehr oder weniger deutlich ausfallen kann. Am stärksten ist sie in Rahačov.

#### 2.4.2 Im Interviewkorpus

Auch unter den 54 Respondenten des Familienkorpus sind etwa doppelt so viele Frauen wie Männer:

Tab. 14 Die Respondenten im Interviewkorpus nach Geschlecht und Stadt

<b>Stadt</b>		<b>weiblich</b>	<b>männlich</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	7	0	7
	%	100,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	10	3	13
	%	76,9%	23,1%	100,0%
<b>Ra</b>	<i>N</i>	5	6	11
	%	45,5%	54,5%	100,0%
<b>Sh</b>	<i>N</i>	5	4	9
	%	55,6%	44,4%	100,0%

<b>Sl</b>	<i>N</i>	6	1	7
	%	85,7%	14,3%	100,0%
<b>Sm</b>	<i>N</i>	4	3	7
	%	57,1%	42,9%	100,0%
<b>Gesamt</b>	<i>N</i>	37	17	54
	%	68,5%	31,5%	100,0%

Was die Anzahl der von den Respondenten geäußerten Wortformen angeht, so fällt das Überwiegen der Frauen wiederum noch etwas deutlicher aus:

Tab. 15 Anteil der Geschlechter am analysierten Datenmaterial im Interviewkorpus (gemessen in Wortformen), differenziert nach Städten

<b>Stadt</b>		<b>weiblich</b>	<b>männlich</b>	<b>Gesamt</b>
<b>Ak</b>	<i>N</i>	21776	0	21776
	%	100,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	20257	6544	26801
	%	75,6%	24,4%	100,0%
<b>Ra</b>	<i>N</i>	10729	10797	21526
	%	49,8%	50,2%	100,0%
<b>Sh</b>	<i>N</i>	13842	6561	20403
	%	67,8%	32,2%	100,0%
<b>Sl</b>	<i>N</i>	23350	1772	25122
	%	92,9%	7,1%	100,0%
<b>Sm</b>	<i>N</i>	18111	7831	25942
	%	69,8%	30,2%	100,0%
<b>Gesamt</b>	<i>N</i>	108065	33505	141570
	%	76,3%	23,7%	100,0%



## 2.5 Zur Bildung<sup>8</sup>

### 2.5.1 Im Familienkorpus

Die WRGR, die „Trasjanka“, gilt gemeinhin als symptomatisch für eine schlechte Bildung, als Kennzeichen von Menschen, die aufgrund mangelnder Bildung weder „reines Weißrussisch“ noch „reines Russisch“ sprechen können. LISKOVEC (2005) macht letzteres geradezu zum Definitionskriterium für die Träger der Trasjanka. HENTSCHEL (2013) hat dieses bereits zurückgewiesen, und zwar primär mit der Begründung, dass durch die empirische Eingrenzung auf eine solche Sprechergruppe<sup>9</sup> das wahre Ausmaß der WRGR a priori unerkennbar bleibt. Im Familienkorpus des OK-WRGR sind Sprecher mit sehr unterschiedlichem Bildungshintergrund erfasst, wie folgende Tabelle zu erkennen gibt.

Tab. 16 Die Sprecher im Familienkorpus nach Stadt und Bildung

Stadt		noch ohne Schulabschluss	Schule	Berufsschule	Hochschule	kein Schulabschluss	k.A.	Gesamt
<b>Ak</b>	<i>N</i>	0	2	4	4	0	4	14
	%	0,0%	14,3%	28,6%	28,6%	0,0%	28,6%	100,0%
<b>Ba</b>	<i>N</i>	0	2	3	1	0	0	6
	%	0,0%	33,3%	50,0%	16,7%	0,0%	0,0%	100,0%
<b>Ch</b>	<i>N</i>	0	7	3	0	0	0	10
	%	0,0%	70,0%	30,0%	0,0%	0,0%	0,0%	100,0%
<b>Mi</b>	<i>N</i>	0	3	6	1	1	0	11
	%	0,0%	27,3%	54,5%	9,1%	9,1%	0,0%	100,0%

<sup>8</sup> Hier liegen für sechs der „zentralen Respondenten“ keine Angaben vor: zu einem weiteren aus Akcjarbski und zu zwei Respondenten aus Šarkoŭščyna. Die verbleibenden 64 Sprecher lieferten zusammen 193.181 Wortformen. Das sind knapp 91 Prozent der Gesamtmenge.

<sup>9</sup> Ob es tatsächlich Menschen gibt, die nur eine sehr stark gemischte weißrussisch-russische Rede praktizieren können, oder ob nicht doch fast immer außerdem entweder „akzeptable“ Russischkenntnisse (und sei es mit weißrussischen Ausspracheinterferenzen und leichten lexikalischen Interferenzen, letzteres besonders für typisch weißrussische Realien) oder „akzeptable“ Weißrussischkenntnisse (und sei es mit einer gewissen Zahl russischer Entlehnungen besonders in der Lexik) vorhanden sind, wird von Weißrussisten bzw. Sprachwissenschaftlern in Weißrussland uneinheitlich beurteilt.

<b>Ra</b>	<i>N</i>	0	0	2	8	0	0	10
	%	0,0%	0,0%	20,0%	80,0%	0,0%	0,0%	100,0%
<b>Sh</b>	<i>N</i>	3	1	1	0	0	2	7
	%	42,9%	14,3%	14,3%	0,0%	0,0%	28,6%	100,0%
<b>Sm</b>	<i>N</i>	2	2	7	1	0	0	12
	%	16,7%	16,7%	58,3%	8,3%	0,0%	0,0%	100,0%
<b>Gesamt</b>	<i>N</i>	5	17	26	15	1	6	70
	%	7,1%	24,3%	37,1%	21,4%	1,4%	8,6%	100,0%

Das Gros der zentralen Sprecher hat also entweder einen einfachen Schulabschluss oder einen Berufsschulabschluss. Es sind aber auch Sprecher mit einer darüber hinausgehenden Bildung vertreten, geballt in Rahačou. Nur ein Sprecher hat keinen Abschluss, fünf sind noch in der Schulausbildung. Darüber hinaus ist offensichtlich, dass die Familienkontexte in den einzelnen Städten hinsichtlich des Bildungshintergrunds sehr heterogen sind. Dennoch ist die Rede in den erfassten familiären Gesprächen von unserem Gewährsmann S. Zaprudski und den weißrussischen Mitarbeitern in der Projektgruppe intuitiv sehr wohl als Instanz der WRGR bewertet worden. Die bisher vorgelegten Analysen stützen diese Einschätzung klar.<sup>10</sup>

Auch in diesem Zusammenhang stellt sich die Frage nach dem Anteil der einzelnen Gruppen am Material:

Tab. 17 Anteile der Sprecher am Material des Familienkorpus (in Wortformen), differenziert nach Stadt und Bildung

Stadt		noch ohne Schulabschluss	Schule	Berufsschule	Hochschule	kein Schulabschluss	k.A.	Gesamt
<b>Ak</b>	<i>n</i>	0	2491	9591	6645	0	3940	22667
	%	0,0%	11,0%	42,3%	29,3%	0,0%	17,4%	100,0%
<b>Ba</b>	<i>n</i>	0	8382	10863	5422	0	0	24667
	%	0,0%	34,0%	44,0%	22,0%	0,0%	0,0%	100,0%

<sup>10</sup> Vergleiche die Bibliographie unter <https://www.uni-oldenburg.de/slavistik/forschung/-sprachwissenschaft/schwerpunkt-mischvarietäten/oldenburger-publikationen-zur-weissrussisch-russischen-gemischten-rede/>.

<b>Ch</b>	<i>n</i>	0	27727	6183	0	0	0	33910
	%	0,0%	81,8%	18,2%	0,0%	0,0%	0,0%	100,0%
<b>Mi</b>	<i>n</i>	0	16742	12591	2329	1458	0	33120
	%	0,0%	50,5%	38,0%	7,0%	4,4%	0,0%	100,0%
<b>Ra</b>	<i>n</i>	0	0	5184	27839	0	0	33023
	%	0,0%	0,0%	15,7%	84,3%	0,0%	0,0%	100,0%
<b>Sh</b>	<i>n</i>	6295	7239	2169	0	0	9420	25123
	%	25,1%	28,8%	8,6%	0,0%	0,0%	37,5%	100,0%
<b>Sm</b>	<i>n</i>	1480	10477	20612	859	0	0	33428
	%	4,4%	31,3%	61,7%	2,6%	0,0%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	7775	73058	67193	43094	1458	13360	205938
	%	3,8%	35,5%	32,6%	20,9%	,7%	6,5%	100,0%

Auch aus dieser Perspektive bilden allgemein die Sprecher mit einfacher Schulbildung, ggf. ergänzt durch einen Berufsschulabschluss, das Zentrum des Korpus. Nur in Rahačov stammt das meiste Material von Sprechern mit höherer Schulbildung. In Šarkoňšćyna sind auch jüngere Sprecher, die sich noch in der schulischen Ausbildung befinden, stärker beteiligt.

### 2.5.2 *Im Interviewkorpus*

Auch im Interviewkorpus des OK-WRGR sind Sprecher mit sehr unterschiedlichem Bildungshintergrund vertreten, wie folgende Tabelle zeigt.

Tab. 18 Die Respondenten des Interviewkorpus nach Stadt und Bildung

<b>Stadt</b>		<b>kein Schulabschluss</b>	<b>Schule</b>	<b>Berufsschule</b>	<b>Hochschule</b>	<b>Gesamt</b>
<b>Ak</b>	<i>n</i>	0	3	3	1	7
	%	0,0%	42,9%	42,9%	14,3%	100,0%
<b>Ch</b>	<i>n</i>	0	1	10	2	13
	%	0,0%	7,7%	76,9%	15,4%	100,0%
<b>Ra</b>	<i>n</i>	2	2	6	1	11
	%	18,2%	18,2%	54,5%	9,1%	100,0%

<b>Sh</b>	<i>n</i>	0	1	5	3	9
	%	0,0%	11,1%	55,6%	33,3%	100,0%
<b>Sl</b>	<i>n</i>	1	0	5	1	7
	%	14,3%	0,0%	71,4%	14,3%	100,0%
<b>Sm</b>	<i>n</i>	1	4	2	0	7
	%	14,3%	57,1%	28,6%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	4	11	31	8	54
	%	7,4%	20,4%	57,4%	14,8%	100,0%

Vier Sprecher verfügen über keinen Schulabschluss, acht haben Hochschulbildung. Die mit Abstand größte Gruppe machen die Sprecher mit Berufsschulabschluss aus. Wie die folgende Tabelle zeigt, fällt der Anteil der Sprechergruppen am Sprachmaterial sehr ähnlich zur Verteilung der Sprecher aus.

Tab. 19 Anteil der Respondenten am Material des Interviewkorpus (in Wortformen) nach Stadt und Bildung

<b>Stadt</b>		<b>kein Schulabschluss</b>	<b>Schule</b>	<b>Berufsschule</b>	<b>Hochschule</b>	<b>Gesamt</b>
<b>Ak</b>	<i>n</i>	0	7950	12224	1602	21776
	%	0,0%	36,5%	56,1%	7,4%	100,0%
<b>Ch</b>	<i>n</i>	0	1565	20257	4979	26801
	%	0,0%	5,8%	75,6%	18,6%	100,0%
<b>Ra</b>	<i>n</i>	1016	3516	13613	3381	21526
	%	4,7%	16,3%	63,2%	15,7%	100,0%
<b>Sh</b>	<i>n</i>	0	2036	12602	5765	20403
	%	0,0%	10,0%	61,8%	28,3%	100,0%
<b>Sl</b>	<i>n</i>	1527	0	17866	5729	25122
	%	6,1%	0,0%	71,1%	22,8%	100,0%
<b>Sm</b>	<i>n</i>	7939	13395	4608	0	25942
	%	30,6%	51,6%	17,8%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	10482	28462	81170	21456	141570
	%	7,4%	20,1%	57,3%	15,2%	100,0%

Zwischen den Städten variiert die Verteilung allerdings wieder stark.

## 2.6 Zur Land-Stadt-Migration

### 2.6.1 Im Familienkorpus

Traditionell wird die WRGR bzw. die „Trasjanka“ als Charakteristikum des Dorfbewohners gesehen, den es auf Dauer in die Stadt verschlagen hat. Von weiterem, besonderen Interesse sind natürlich die Kinder dieser Land-Stadt-Migranten. Folgende Tabelle zeigt die Verteilung der Respondenten nach dem Kriterium der Binnenmigration.

Tab. 20 Die Respondenten des Familienkorpus nach Stadt und Binnenmigration

		Spätmigranten und Dorfbewohner	typische Land- Stadt-Migranten	geb. Städter	k.A.	Gesamt
<b>Ak</b>	<i>n</i>	1	6	3	4	14
	%	7,1%	42,9%	21,4%	28,6%	100,0%
<b>Ba</b>	<i>n</i>	0	4	2	0	6
	%	0,0%	66,7%	33,3%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	1	4	5	0	10
	%	10,0%	40,0%	50,0%	0,0%	100,0%
<b>Mi</b>	<i>n</i>	1	7	3	0	11
	%	9,1%	63,6%	27,3%	0,0%	100,0%
<b>Ra</b>	<i>n</i>	2	4	4	0	10
	%	20,0%	40,0%	40,0%	0,0%	100,0%
<b>Sh</b>	<i>n</i>	0	0	0	7	7
	%	0,0%	0,0%	0,0%	100,0%	100,0%
<b>Sm</b>	<i>n</i>	0	8	4	0	12
	%	0,0%	66,7%	33,3%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	5	33	21	11	70
	%	7,1%	47,1%	30,0%	15,7%	100,0%

Den größten Teil der Sprecher (fast die Hälfte) machen „typische Land-Stadt-Migranten“ aus, die ihre Jugend weitgehend auf dem Land verbrachten und

frühestens für die Berufsschule in die Stadt kamen, dort aber auf jeden Fall mehr als zehn, meistens viel mehr Jahre blieben. Die zweitgrößte Gruppe sind Sprecher, die schon in der jeweiligen Stadt geboren bzw. aufgewachsen sind; meist sind dies Kinder von Vertretern der ersten Gruppe. Die dritte Gruppe bilden Dorfbewohner, die während der Aufnahme zu Besuch sind, oder Respondenten, die im hohen Lebensalter vom Dorf in die Stadt gekommen sind. In der Regel handelt es sich um die Großeltern in der jeweiligen Familie. Folgende Tabelle zeigt den Anteil dieser drei Gruppen am Sprachmaterial:

Tab. 21 Anteil der Respondenten am Material des Familienkorpus (in Wortformen), differenziert nach Stadt und Binnenmigration

		Spätmigranten und Dorfbewohner	typische Land- Stadt-Migranten	geb. Städter	k.A.	Gesamt
<b>Ak</b>	<i>n</i>	1432	13421	3186	4628	22667
	%	6,3%	59,2%	14,1%	20,4%	100,0%
<b>Ba</b>	<i>n</i>	0	17455	7212	0	24667
	%	0,0%	70,8%	29,2%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	1543	18366	14001	0	33910
	%	4,6%	54,2%	41,3%	0,0%	100,0%
<b>Mi</b>	<i>n</i>	1458	27279	4383	0	33120
	%	4,4%	82,4%	13,2%	0,0%	100,0%
<b>Ra</b>	<i>n</i>	3880	20075	9068	0	33023
	%	11,7%	60,8%	27,5%	0,0%	100,0%
<b>Sh</b>	<i>n</i>	0	0	0	25123	25123
	%	0,0%	0,0%	0,0%	100,0%	100,0%
<b>Sm</b>	<i>n</i>	0	21471	11957	0	33428
	%	0,0%	64,2%	35,8%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	8313	118067	49807	29751	205938
	%	4,0%	57,3%	24,2%	14,4%	100,0%

Gemessen an den Anteilen der genannten drei Gruppen am Material (in Wortformen) vergrößert sich die Bedeutung der typischen Land-Stadt-Migranten noch um ca. 10 Prozentpunkte, wohingegen die Anteile der beiden übrigen Gruppen jeweils etwas geringer werden.

## 2.6.2 Im Interviewkorpus

Folgende Tabelle zeigt die Verteilung der Respondenten im Interviewkorpus nach dem Kriterium der Binnenmigration.

Tab. 22 Die Respondenten im Interviewkorpus nach Stadt und Binnenmigration

		Spätmigranten und Dorfbewohner	typische Land- Stadt-Migranten	geb. Städter	k.A.	Gesamt
<b>Ak</b>	<i>n</i>	0	5	2	0	7
	%	0,0%	71,4%	28,6%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	0	8	4	1	13
	%	0,0%	61,5%	30,8%	7,7%	100,0%
<b>Ra</b>	<i>n</i>	0	6	1	4	11
	%	0,0%	54,5%	9,1%	36,4%	100,0%
<b>Sh</b>	<i>n</i>	0	8	1	0	9
	%	0,0%	88,9%	11,1%	0,0%	100,0%
<b>Sl</b>	<i>n</i>	0	3	3	1	7
	%	0,0%	42,9%	42,9%	14,3%	100,0%
<b>Sm</b>	<i>n</i>	1	4	2	0	7
	%	14,3%	57,1%	28,6%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	1	34	13	6	54
	%	1,9%	63,0%	24,1%	11,1%	100,0%

Etwa zwei Drittel der Respondenten sind Land-Stadt-Migranten. Nur ein Sprecher ist der Gruppe der Spätmigranten und Dorfbewohner zuzuordnen. Jeder vierte Respondent ist in der Stadt geboren und aufgewachsen.

Folgende Tabelle zeigt die Anteile dieser drei Gruppen am sprachlichen Material:

Tab. 23 Anteile der Respondenten am Interviewkorpus (in Wortformen), differenziert nach Stadt und Binnenmigration

		Spätmigranten und Dorfbewohner	typische Land- Stadt-Migranten	geb. Städter	k.A.	Gesamt
<b>Ak</b>	<i>n</i>	0	14923	6853	0	21776
	%	0,0%	68,5%	31,5%	0,0%	100,0%
<b>Ch</b>	<i>n</i>	0	17278	7958	1565	26801
	%	0,0%	64,5%	29,7%	5,8%	100,0%
<b>Ra</b>	<i>n</i>	0	12113	858	8555	21526
	%	0,0%	56,3%	4,0%	39,7%	100,0%
<b>Sh</b>		0	18367	2036	0	20403
	%	0,0%	90,0%	10,0%	0,0%	100,0%
<b>Sl</b>		0	10709	11267	3146	25122
	%	0,0%	42,6%	44,8%	12,5%	100,0%
<b>Sm</b>	<i>n</i>	7939	13395	4608	0	25942
	%	30,6%	51,6%	17,8%	0,0%	100,0%
<b>Gesamt</b>	<i>n</i>	7939	86785	33580	13266	141570
	%	5,6%	61,3%	23,7%	9,4%	100,0%

Die Anteile der drei Gruppen am Sprachmaterial fallen wieder sehr ähnlich aus wie die Verteilung der Sprecher auf die drei Gruppen. Lediglich der Vertreter der Gruppe der Spätmigranten ist überdurchschnittlich am gesammelten Material beteiligt.

## 2.7 Fazit

Unterschiede wie die hier dargelegten sind bei der Nutzung von Mitschnitten ungewollter Gespräche bei den üblicherweise begrenzten praktischen, nicht zuletzt finanziellen Möglichkeiten der Datenerhebung und dem Bedarf an einer breiten Materialbasis nicht zu vermeiden, schränken die „direkte“ Vergleichbarkeit der Teilkorpora mit Methoden der einfachen deskriptiven Statistik jedoch ein. Durch Nutzung entsprechender statistischer Verfahren kann man diesen „Asymmetrien“ jedoch begegnen.



### 3 Die Struktur der Transkriptionen im System CHAT

Das OK-WRGR ist in den im Rahmen des CHILDES-Projekts entwickelten Konventionen von CHAT transkribiert und annotiert (Stand 2013; vgl. MCWHINNEY 2013a). Das Korpus ist daher mit den Werkzeugen des Programms CLAN analysierbar (vgl. MCWHINNEY 2013b, siehe Abschnitt 4). Es handelt sich um UTF8-kodierte Textdateien, in denen die Informationen in verschiedenen Zeilen (sogenannten Tiers) organisiert sind.

Im Folgenden wird vor allem auf einige Abweichungen vom üblichen CHAT-System eingegangen, ferner auf die Prinzipien bei der Transkription, bei der Bestimmung der „sprachlichen Affinität“ der Wortformen (grob: kann die Wortform dem Weißrussischen oder Russischen zugeordnet werden; ist sie eine „gemeinsame“ oder eine „hybride“ Wortform) und bei der grammatischen Annotation. Für die sonstigen Konventionen des CHAT-Systems sei der Nutzer auf das CHAT-Manual (MCWHINNEY 2013a) verwiesen.<sup>11</sup>

#### 3.1 Dateinamen

Die Dateinamen der Transkripte sind nach folgendem Schema aufgebaut: Ein dreistelliges Kürzel des Subkorpus (*fam* für *Familienkorpus*, *int* für *Interviewkorpus*), ein zweistelliges Kürzel für die Stadt, aus der die Aufnahme stammt, und eine Nummer innerhalb der Stadt, jeweils getrennt durch einen Unterstrich. Die Kürzel der Städte sind folgende:

<i>Kürzel</i>	<i>Stadt</i>
ak	Akcjabrski
ba	Baranavičy (nur im Familienkorpus)
ch	Chocimsk
sl	Slonim (nur im Interviewkorpus)

---

<sup>11</sup> Die folgenden Darstellungen zur prinzipiellen Vorgehensweise im OK-WRGR schließen nicht aus, dass im OK-WRGR punktuelle Abweichungen von den Prinzipien auftreten. Angesichts der großen Datenmenge, die teilweise von studentischen und wissenschaftlichen Hilfskräften zu bewältigen war, sind vereinzelte Widersprüche oder Fehler nicht völlig vermeidbar. Eine Überprüfung der Transkription der über 400.000 Wortformen in den beiden Teilkorpora anhand akustischer Analysen ist praktisch natürlich völlig ausgeschlossen.

mi	Minsk (nur im Familienkorpus)
ra	Rahačou
sm	Smarhon'
sh	Šarkoŭščyna

### 3.2 Zur Grobstruktur der Transkriptionen

Folgendes Beispiel verdeutlicht die Struktur der Transkriptionen:

*Beispiel:*

```
@UTF8
@Begin
@Languages:   hybe
@Participants: RFB Village
@ID:   hybe|fam|RFB|72;|female|central||Village|higher||
@L1 of RFB:   be
@MT of RFB:   be
@FL of RFB:   brms
@Situation:   Conversation at home
*RFB:   pryvet@r@pb , dzetkâ@c@pb .
%mor:   r:pb#n:inanim|_0_privet_&NOM-M:SG
        c:pb#n:anim|_dzetka_detka_&NOM-GC:SG .
*RFB:   nu@c@pc , bjažy@b@pb , a@c@pc to@c@pc (.) ja@c@pc s@c@pc
        xolâdu@b@pc .
%mor:   c:pc#ptl|_nu_nu_ b:pb#v:ipf|_behčy_0_&IMPER-2:SG
        c:pc#conj|_a_a_ c:pc#ptl|_to_to_
        c:pc#pro:pers:dei|_ja_ja_&NOM-1:SG c:pc#prep|_z_s_
        b:pb#n:inanim|_cholad_cholod_&GEN-M:SG .
*RFB:   xalodnâ@c@pc .
%mor:   c:pc#adv|_choladna_cholodno_ .
*RFB:   zâstužu@r@pc cjabe@b@pb .
%mor:   r:pc#v:pf|_0_zastudit'_&FUT-1:SG
        b:pb#pro:pers:dei|_ty_ty_&ACC-2:SG .
@End
```

Jedes Transkript beginnt mit der Angabe der Kodierung **@UTF8**, gefolgt von einer Zeile **@Begin**, die den Beginn des Transkripts markiert. Hiernach folgen in einem Header einige übergeordnete Informationen zum Gespräch und

zu den beteiligten Respondenten – von der Zeile **@Languages:** bis zur Zeile **@Situation:**. Es folgt die eigentliche Transkription des Gesprächs, inklusive einiger Annotationen. Jedes Transkript endet mit der Zeile **@End**.

### 3.3 Übergeordnete Informationen

Zunächst werden in einem Header einige übergeordnete Informationen zum Gespräch und zu den beteiligten Respondenten gegeben. Diese Informationen werden im Folgenden erläutert.

#### 3.3.1 @Languages: Die beteiligten Sprechertypen

In der Zeile **@Languages:** erfolgt in CHAT üblicherweise die Angabe der in dem Gespräch benutzten Sprachen. Im OK-WRGR werden hier die beteiligten *Sprechertypen* aufgelistet. Hinter dem Sprechertyp verbirgt sich Folgendes: HENTSCHEL & ZELLER (2013) zeigen mithilfe einer Clusteranalyse, dass sich die zentralen Sprecher des Familienkorpus – solche, die mit einer ausreichend großen Anzahl an Äußerungen im Korpus vertreten sind – anhand ihrer Anteile an (bei Abstraktion von phonischen Merkmalen) ‚weißrussischen‘, ‚russischen‘ und ‚hybriden‘ Äußerungen in vier Sprechertypen einteilen lassen.<sup>12</sup> Der Sprechertyp ist also für einen Sprecher in allen Gesprächen konstant und sagt damit nichts darüber aus, wie sich der Sprecher gerade in dem gegebenen Gespräch sprachlich verhält.

Die Sprecher, die im Interviewkorpus vertreten sind, sind „bekenkende“ Sprecher der gemischten Rede, d.h. sie hatten in einer Umfrage mit 1.400 Respondenten in Weißrussland zu erkennen gegeben, dass die WRGR ihr regelmäßiges, meist primäres Kommunikationsmittel ist. Zudem wurden für das Interviewkorpus – anders als im Familienkorpus – gezielt nur solche Passagen aus den Interviews ausgewählt, die als Instanzen der WRGR zu werten sind. Für diese Sprecher verbietet sich daher eine ähnliche, d.h. statistisch-basierte deduktive Klassifizierung wie für die Sprecher im Familienkorpus. Alle Respondenten im Interviewkorpus erhalten daher das Kürzel **brms**.

---

<sup>12</sup> Wie HENTSCHEL & ZELLER (2014) zeigen, unterscheiden sich diese Sprechertypen auch in ihren phonetisch-phonologischen Charakteristika.

<i>Korpus</i>	<i>Kürzel</i>	<i>Erläuterung</i>	<i>Anzahl der Sprecher im OK-WRGR</i>
Familienkorpus	hybe	Hohe Frequenz sowohl ‚hybrider‘ als auch ‚weißrussischer‘ Äußerungen, niedrige Frequenz ‚russischer‘ Äußerungen (Sprechertyp HW bei HENTSCHEL & ZELLER (2013))	10
Familienkorpus	hyb	Klares Übergewicht ‚hybrider‘ Äußerungen über sowohl ‚russische‘ als auch ‚weißrussische‘ Äußerungen (Sprechertyp H bei HENTSCHEL & ZELLER (2013))	23
Familienkorpus	hyru	Hohe Frequenz vor allem ‚hybrider‘, aber auch ‚russischer‘ Äußerungen, niedrigere Frequenz ‚weißrussischer‘ Äußerungen (Sprechertyp HR bei HENTSCHEL & ZELLER (2013))	25
Familienkorpus	rus	Klares Übergewicht ‚russischer‘ Äußerungen sowohl über ‚hybride‘ als auch – noch deutlicher – über ‚weißrussische‘ Äußerungen (Sprechertyp R bei HENTSCHEL & ZELLER (2013))	12
Familienkorpus	nan	„Keine Angabe“: Sprecher mit insgesamt <b>wenigen Äußerungen</b>	59
Interviewkorpus	brms	„Belarusian-Russian mixed speech“: Angabe für alle Respondenten des Interviewkorpus	54
Interviewkorpus	nan	„Keine Angabe“ (Interviewer und in kurzen Passagen vertretene Nicht-Zielpersonen)	nicht differenziert

*Beispiel:*

@Languages:    hyb, hyru

An diesem Gespräch nehmen also Sprecher teil, die im Allgemeinen ein Übergewicht ‚hybrider‘ Äußerungen zeigen, und solche, die sowohl viele ‚hybride‘ als auch viele ‚russische‘ Äußerungen aufweisen.

*3.3.2 @Participants: Die Rollen bzw. der Migrationsstatus der beteiligten Sprecher*

In der Zeile **@Participants**: erfolgt in CHAT üblicherweise die Angabe der „Rollen“ der am Gespräch beteiligten Respondenten. Im OK-WRGR erfolgt hier für die Sprecher, für die dies relevant ist, die Angabe des Migrationsstatus (s.u. 3.3.3.5).

Prototypisch sind drei Gruppen von Sprechern der gemischten Rede zu unterscheiden, die sich nach dem Kriterium der Migration unterscheiden:

- Die erste Gruppe bilden *Land-Stadt-Migranten*, d.h. Personen, die in den 1960er und 1970er, eventuell 1980er Jahren im frühen Erwachsenenalter als Arbeitsmigranten aus weißrussisch-dialektal geprägten ländlichen Gebieten in die sich rasch entwickelnden Städte gezogen sind. Ihre schulische Ausbildung war noch wesentlich weißrussisch geprägt.
- Die zweite Gruppe sind *geborene Städter*, d.h. in der Regel die Kinder dieser Migranten. Sie sind bereits in den Städten zur Welt gekommen oder haben dort zumindest den größten Teil ihrer Kindheit verbracht. Ihre erste sprachliche Sozialisierung erfuhren sie durch die gemischte Rede ihrer Eltern, d.h. gemischt-monolektal, ohne eine Differenzierung des sprachlichen Inputs. Anders als ihre Eltern wurden Angehörige dieser Generation allerdings in der schulischen Ausbildung, teilweise bereits im vorschulischen Bereich, viel früher und intensiver im Russischen sozialisiert. Sie sind daher in aller Regel in der Lage, Russisch in seiner Standardform, ohne tiefere Interferenzen des Weißrussischen, zu sprechen. Der Kontakt zum dialektalen Weißrussischen ist für diese Generation schwächer.
- Vor allem im Familienkorpus vertreten ist außerdem eine weitere Gruppe, und zwar die Gruppe der *Spätmigranten und Dorfbewohner*, d.h. in der Regel die Generation der Eltern von Land-Stadt-Migranten. Vertreter dieser Generation haben ihr gesamtes Leben auf dem Land verbracht oder sind erst im höheren Alter zu ihren Kindern in die Städte gezogen.

Im Projekt standen die ersten beiden Gruppen im Mittelpunkt des Interesses. Die dritte Gruppe bilden zufällig an den Familiengesprächen Beteiligte; sie sind zahlenmäßig nur schwach vertreten (vgl. 2.6).

Im OK-WRGR erfolgt in der Zeile **@Participants**: für die Sprecher im Familienkorpus und die Respondenten im Interviewkorpus soweit bekannt die Angabe, zu welcher dieser Gruppen sie gehören. Die Interviewer im Interviewkorpus erhalten die Rolle **Interviewer**, andere Sprecher im Interviewkorpus die Angabe **Other**.

<i>Kürzel</i>	<i>Bedeutung</i>
Village	Spätmigrant oder Dorfbewohner
Migrant	Land-Stadt-Migrant
Town	Geborener Städter
Na	keine Angabe
Interviewer	Interviewer im Interviewkorpus
Other	Zufällig während der Interviews anwesende weitere Personen

**Beispiel:**

@Participants: BFA Migrant, BFB Town

Bei dem Sprecher BFA handelt es sich um einen Vertreter der Land-Stadt-Migranten, BFB ist ein geborener Städter.

**3.3.3 @ID: Die Sprecherinformationen**

Zu den einzelnen Sprechern finden sich in den Zeilen @ID: folgende Informationen:

@ID: Sprechertyp|Korpus|Sprecherkürzel|Alter|Geschlecht|Beteiligung||Rolle/Migrationsstatus|Ausbildung||

Die Angaben zu *Alter* (in Jahren; Stand von 2008) und *Geschlecht* (male/female) sind selbsterklärend, die übrigen Informationen seien kurz erläutert.

**3.3.3.1 Sprechertyp**

An der Stelle, an der in CHAT üblicherweise die Sprache angegeben wird, die von dem jeweiligen Sprecher im Gespräch verwendet wird, erfolgt im OK-WRGR die Angabe der sprachlichen Orientierung des Sprechers, d.h. des *Sprechertyps* (s.o. 3.3.1).

**3.3.3.2 Korpus**

Das OK-WRGR besteht, wie eingangs erwähnt, aus zwei Teilkorpora: dem Familienkorpus und dem Interviewkorpus. Zu welchem der beiden Korpora das Gespräch gehört, wird an dieser Stelle festgehalten.

<i>Kürzel</i>	<i>Bedeutung</i>
fam	Familienkorpus
int	Interviewkorpus

### 3.3.3.3 Sprecherkürzel

Die Sprecher im Familienkorpus und die Respondenten im Interviewkorpus sind eineindeutig einem über die einzelnen Gespräche hinweg konstanten drei- oder vierstelligen Sprecherkürzel zugeordnet. Der erste Buchstabe dieses Kürzels gibt die Stadt an.

<i>Kürzel</i>	<i>Stadt</i>
A	Akcjabrski
B	Baranavičy (nur im Familienkorpus)
C	Chocimsk
L	Slonim (nur im Interviewkorpus)
M	Minsk (nur im Familienkorpus)
R	Rahačoŭ
S	Smarhon'
H	Šarkoŭščyna

Im Familienkorpus folgt an zweiter Stelle ein **F**, im Interviewkorpus ein **I**. An dritter Stelle folgt ein in der Stadt eindeutiges Kürzel des Sprechers, in der Regel einstellig (im Familienkorpus Buchstabe oder Zahl, im Interviewkorpus Zahl), in einigen Fällen (wenn mehr Sprecher in einer Stadt gegeben sind, als Ziffern bzw. Buchstaben und Ziffern zur Verfügung stehen) auch zweistellig.

#### *Beispiele:*

CI12: Sprecher 12 aus Chocimsk im Interviewkorpus

HFN: Sprecher N aus Šarkoŭščyna im Familienkorpus

Im Interviewkorpus erhalten alle Interviewer das Kürzel **INT**. Sprecher, die zufällig während des Interviews anwesend waren und sich nur mit einigen wenigen Äußerungen am Gespräch beteiligten, erhalten das Kürzel **OTH** („Other“), bei mehreren solcher Sprecher in einem Gespräch durchnummeriert (**OT1**, **OT2** usw.).

### 3.3.3.4 Beteiligung am Korpus

Für jeden Sprecher im Familienkorpus wird angegeben, ob der Sprecher zu den „zentralen“ Respondenten gehört. „Zentral“ sind, wie eingangs dargelegt, die 70 Sprecher des Familienkorpus, die mit mehr als 500 Wortformen im Korpus vertreten sind (was ungefähr 100 Äußerungen entspricht); diese erhalten den Eintrag **central**. Den übrigen Sprechern wird die Kennzeichnung **peripher** zugewiesen.

Im Interviewkorpus erhalten die Interviewten die Kennzeichnung **Informant**. Interviewer und sonstige Personen im Interviewkorpus haben keinen Eintrag; die Angabe, dass es sich um den Interviewer handelt, erfolgt im nächsten Schritt.

### 3.3.3.5 Migrationsstatus / Rolle im Interview

An dieser Stelle erfolgt in CHAT üblicherweise die Angabe der „Rollen“ der am Gespräch beteiligten Sprecher. Im OK-WRGR erfolgt hier für die Sprecher, für die dies relevant ist (die Sprecher des Familienkorpus, die Respondenten des Interviewkorpus), die Angabe des Migrationsstatus (**Village**, **Migrant** oder **Town**), die Interviewer erhalten die Angabe **Interviewer**, andere Sprecher im Interviewkorpus die Angabe **Other** (siehe oben, Abschnitt 3.3.2).

### 3.3.3.6 Ausbildung

Hinsichtlich der Ausbildung werden folgende Gruppen unterschieden:

<i>Kürzel</i>	<i>Erläuterung</i>
<i>in education</i>	Schüler
<i>incomplete/primary</i>	8-9 Schuljahre und weniger
<i>secondary</i>	10-11 Schuljahre
<i>professional</i>	Berufsschule, Technikum, College
<i>higher</i>	Hochschulausbildung
Na	keine Angabe

### 3.3.4 @L1: Die Erstsprache

Sofern Informationen vorliegen, sind in der Zeile **@L1 of #**: die Sprachen angegeben, die der Sprecher als seine Erstsprachen angegeben hat (als Ant-



wort auf die Frage: „In welcher Sprache haben Sie begonnen zu sprechen?“).  
Folgende Werte sind möglich:

<i>Kürzel</i>	<i>Erläuterung</i>
be	Weißrussisch
ru	Russisch
brms	Weißrussisch-russische gemischte Rede

Vereinzelte sind auch andere Angaben wie **p1** (*Polnisch*) möglich. Da Mehrfachantworten zugelassen wurden, können auch Kombinationen dieser Werte auftreten.

### 3.3.5 @FL: Die Familiensprache in der Kindheit

Sofern Informationen vorliegen, sind in der Zeile **@FL of #**: die Sprachen angegeben, für die der Sprecher angibt, dass sie in seiner Kindheit in seiner Familie gesprochen wurden.

### 3.3.6 @MT: Die Muttersprache

Sofern Informationen vorliegen, sind in der Zeile **@MT of #**: die Sprachen angegeben, die der Sprecher als seine Muttersprache („rodnoj jazyk“) angibt.

### 3.3.7 @Comment: Kommentare zur Transkription

In der Zeile **@Comment**: erfolgen allgemeine Kommentare zur Transkription. So sind z.B. in einigen wenigen Gesprächen Gesprächspausen oder Überlappungen nicht vermerkt, was hier angegeben wird.

### 3.3.8 @Situation: Die Gesprächssituation

Die Zeile **@Situation**: enthält Angaben zur Gesprächssituation. Typische Angaben sind **Conversation at home** oder **X talks on the phone** o.ä.

### 3.4 Die Wiedergabe der Gespräche

Die mit einem Sternchen ( \* ) beginnenden Zeilen geben die Äußerungen der Sprecher in chronologischer Reihenfolge wieder. Nach dem Sternchen steht das Kürzel des Sprechers (vgl. 3.3.3.3). Darunter erfolgt in der Zeile %mor: die Annotation der Wortformen dieser Äußerung (s.u. 3.5).

#### Beispiel:

```
*BFA: jana@b@pb ne@c@pc máladajê@c@pc užé@r@pc takajê@c@pc
      pârikmaxêršý@r@pr .
%mor: b:pb#pro:pers:ana|_jana_0_&NOM-F:SG c:pc#ptl|_ne_ne_
      c:pc#adj|_malady_molodoj_&NOM-F:SG r:pc#adv|_0_uže_
      c:pc#pro:dem|_taki_takoj_&NOM-F:SG
      r:pr#n:anim|_0_parikmacherša_&NOM-F:SG .
*BFA: ne@c@pc dzevâčkâ@r@pb .
%mor: c:pc#ptl|_ne_ne_ r:pb#n:anim|_0_devočka_&NOM-F:SG .
```

Pausen und Unterbrechungen in der Aufnahme und (aus Datenschutzgründen o.ä.) nicht-transkribierte Passagen sind durch eine spezielle Zeile @New Episode: gekennzeichnet. Die ausgelassenen Passagen können auch sehr kurz sein; mitunter handelt es sich nur um eine einzige Äußerung.

#### Beispiel:

```
*AFD: on@r@pc takovâ@r@pc slovâ@c@pc ni@c@pr znajêt@r@pc .
%mor: r:pc#pro:pers:ana|_0_on_&NOM-M:SG
      r:pc#pro:dem|_taki_takoj_&GEN-N:SG
      c:pc#n:inanim|_slova_slovo_&GEN-N:SG c:pr#ptl|_ne_ne_
      r:pc#v:ipf|_znac'_znat'_&PRAES-3:SG .
@New Episode
*AFJ: nu@c@pc tak@c@pc nâbiri@c@pr jemu@c@pc .
%mor: c:pc#ptl|_nu_nu_ c:pc#ptl|_tak_tak_
      c:pr#v:pf|_nabrac'_nabrat'_&IMPER-2:SG
      c:pc#pro:pers:ana|_ên_on_&DAT-M:SG .
```

#### 3.4.1 Einteilung in Äußerungen

Die Einteilung des Redeflusses in Äußerungen entspricht in der Regel der syntaktischen Einteilung.

- Im Zweifelsfall wird bei der Entscheidung, ob eine Äußerung vorliegt oder zwei Äußerungen gegeben sind, die syntaktische Gliederung vorgezogen: Eine Äußerung entspricht in der Regel einem einfachen Satz (wobei elliptische Strukturen natürlich häufig sind). Äußerungen aus nur einem „Teilsatz“ i.w.S. (mit oder ohne Prädikat, elliptisch oder nicht) machen knapp zwei Drittel der Äußerungen im Familienkorpus aus; ein gutes Viertel besteht aus zwei Teilsätzen (mit Ko- oder Subordination). Nicht selten liegen aber auch prädikatslose Äußerungen oder Teiläußerungen mit einer einzigen Wortform oder nur sehr wenigen Wortformen vor, wie z.B. Affirmations- oder Negationspartikel und / oder Diskursmarker (wie dt. *Ja!*, *Nein!*, *Natürlich!*, *Aha!*) etc. Solche Konstruktionen machen z.B. 8% des Familienkorpus aus.
- Nur wenn intonatorisch starke Evidenz besteht, dass ein syntaktischer Nebensatz als eine eigenständige Äußerung aufzufassen ist (lange Pause und / oder klar auf Abgeschlossenheit hindeutende Intonation im übergeordneten Teilsatz), wird dieser als eigene Äußerung behandelt.
- Nur wenn intonatorisch starke Evidenz besteht (keine Pause und klar auf Nichtabgeschlossenheit hindeutende Intonation im ersten Teilsatz), werden zwei syntaktische Hauptsätze zu einer einzigen Äußerung zusammengefasst.

### 3.4.2 *Transkription der Wortformen*

Die im OK-WRGR vorgenommene Transkription ist eine grobe: Die Notation stellt einen Kompromiss zwischen der herkömmlichen wissenschaftlichen Transliteration (vom weißrussischen bzw. russischen Kyrillisch ins lateinische Alphabet) und einer (sehr breiten) Transkription (Umsetzung von Lautung in Schrift) dar. Die einschlägigsten, prominentesten weißrussisch-russischen Unterschiede in Phonetik und Phonologie werden dabei erfasst.<sup>13</sup>

---

<sup>13</sup> Es handelt sich dabei um ohrenphonetische Klassifizierungen, die zu einem großen Teil von weißrussischen studentischen Hilfskräften durchgeführt wurden. Angesichts der phonetischen Gradualität einiger hier relevanter Erscheinungen, etwa beim unbetonten Vokalismus, besteht natürlich hinsichtlich einer gegebenen einzelnen Transkription stets eine gewisse Unsicherheit. Wie jedoch ZELLER (i.Vorb.) mithilfe akustischer Analysen des Materials des OK-WRGR zeigt, sind die Transkriptionen insgesamt sehr belastbar.

### 3.4.2.1 Vokale in betonter Stellung

In betonten Silben können die Vokalzeichen ( **a** ), ( **e** ), ( **i** ), ( **o** ), ( **ë** ), ( **u** ), ( **y** ) und ( **é** ) auftreten. Zu beachten ist Folgendes:

- /i/ ([i]) und /e/ ([ɛ]) nach nicht-palatalisierten Konsonanten werden mit ( **y** ) und ( **é** ) wiedergegeben.

<i>Transkription:</i>	
mašyna	wr. <i>mašyna</i> ; ru. <i>mašina</i>
šélest	ru. <i>šelest</i>

- ( **i** ) und ( **e** ) sowie ( **ë** ) zeigen die Palatalisiertheit des vorherigen Konsonanten an, ( **y** ) und ( **é** ) die Nicht-Palatalisiertheit. Bei den übrigen Vokalen steht ein ( **j** ) oder – im Falle eines vorangehenden ( **č** ) oder ( **šč** ) – ein ( **'** ), wenn der vorherige Konsonant palatalisiert ist. Dies gilt auch bei Konsonanten, die im Russischen palatalisiert, im Weißrussischen nicht-palatalisiert sind (s.u.).

<i>Transkription bei nicht-palatalisiertem Konsonant</i>	<i>Transkription bei palatalisiertem Konsonant</i>	
Radâm	rjadâm	ru. <i>rjadom</i>
Trox	trëx	wr. <i>troch</i> ; ru. <i>trëch</i>
učorâ	učërâ	wr. <i>učora</i>
častâ	č'astâ	wr. <i>časta</i> ; ru. <i>často</i>
Ńaču	xač'u	wr. <i>chaču</i> ; ru. <i>choču</i>

- Betontes /o/ nach palatalisierten Konsonanten wird mit ( **ë** ) transkribiert (ASCII 235). /jo/ wird mit ( **jö** ) transkribiert:

Transkription	
vsë	ru. <i>vse</i>
trëx	wr. <i>troch</i> ; ru. <i>trëch</i>
jön	wr. <i>jon</i>
vtrajom	ru. <i>vtroem</i>
pap'jöm	wr. <i>pap'ëm</i> ; ru. <i>pop'em</i>

### 3.4.2.2 Vokale in unmittelbar vorbetonter Stellung

Hier stehen die Zeichen ( **a** ), ( **e** ), ( **i** ), ( **u** ), ( **y** ) und ( **é** ). ( **i** ) und ( **e** ) zeigen die Palatalisiertheit des vorherigen Konsonanten an, ( **y** ) und ( **é** ) die Nicht-Palatalisiertheit. Bei den übrigen Vokalen steht ein ( **j** ) oder – im Falle eines vorangehenden ( **č** ) oder ( **šč** ) – ein ( **´** ), wenn der vorherige Konsonant palatalisiert ist.

- nach palatalisierten Konsonanten können die Zeichen ( **i** ), ( **e** ) sowie – nach ( **j** ) – ( **a** ) und ( **u** ) auftreten. Es wird also zwischen [i], [e], [a] und [u] unterschieden. Bei unmittelbar vortonigem /a/, /e/, /o/ wird dementsprechend differenziert nach mit ‚russischem‘ Ikanje übereinstimmenden (in bestimmten Fällen auch mit ‚weißrussisch-dialektalem‘ dissimilativem Jakanje übereinstimmenden) Realisierungen im Bereich [i], mit ‚weißrussischem‘ Jakanje übereinstimmenden Realisierungen im Bereich [a] und intermediären Realisierungen im Bereich [e].<sup>14</sup>

<i>Transkription (je nach phonetischer Realisierung)</i>	
ciper / ceper / cjaper	wr. <i>cjaper</i> ; ru. <i>teper'</i>
nima / nema / njama	wr. <i>nima/njama</i>

- nach nicht-palatalisierten (nicht verhärteten) Konsonanten können in der Regel die Zeichen ( **a** ), ( **u** ) und ( **y** ) auftreten (zu Ausnahmefällen s.u.). Akanje, d.h. die Neutralisierung der Opposition von /a/ und /o/ durch ihren Zusammenfall in einem [a]-artigen Laut, wird in der Transkription wiedergegeben. Allerdings wird nicht differenziert zwischen stärker zentrierten Realisierungen als [ɐ], [ʌ] oder auch [ə], wie sie für das Russische charakteristisch sind, und offeneren Realisierungen im Bereich [a], wie sie für das Weißrussische kennzeichnend sind.

<i>Transkription</i>	
vada	wr. v[a]da; ru. v[ɐ]da

<sup>14</sup> Für eine Analyse der phonischen Variativität auf Basis der auf ohrenphonetischen Klassifizierungen beruhenden Transkriptionen des OK-WRGR vergleiche HENTSCHEL & ZELLER (2014). Eine akustische Analyse des Phänomens findet sich in HENTSCHEL & ZELLER (2011) sowie in ZELLER (i.Vorb.).

- Nach verhärteten Konsonanten wird analog zur Position nach palatalisierten Konsonanten bei /a/, /e/, /o/ mithilfe von ( **ѣ** ), ( **ѐ** ) und ( **а** ) unterschieden zwischen [i, ε, a]. Dies gilt auch für Vokale nach nicht-palatalisiertem [ʃ] sowie nach nicht-palatalisiertem [r], wenn letzterem im Russischen ein palatalisiertes /r'/ entspricht.

<i>Transkription (je nach phonetischer Realisierung)</i>	
žylanêj / žélanêj / žalanêj	ru. <i>želanij</i>
ryka / réka / raka	wr. <i>raka</i> ; ru. <i>reka</i>
učyra / učéra / učara	ru. <i>včera</i>

### 3.4.2.3 Vokale in anderen vortonigen und in nachtonigen Stellungen

Im absoluten Anlaut und bei Hiatt können folgende Zeichen stehen: ( **â** ), ( **і** ), ( **u** ), ( **ѣ** ), ( **ѐ** ).

<i>Transkription</i>	
inâhda	ru. <i>inogda</i>
nêinteresnâ	ru. <i>neinteresno</i>
ykânamist / ékanamist	wr. <i>ékanamist</i> ; ru. <i>ekonomist</i>
ânanas	wr. <i>ananas</i> ; ru. <i>ananas</i>

Abgesehen von diesen beiden Positionen (Anlaut und Hiatt) können folgende Zeichen auftreten: ( **ѐ** ), ( **u** ), ( **ѣ** ), ( **ѣ̂** ), ( **â** ).

- Nach palatalisierten Konsonanten stehen ( **ѐ** ) und – nach ( **ј** ) – ( **u** ). Stärker [i]-, [e]- oder [a]-artige Realisierungen anstelle von /a/, /e/, /o/ werden in anderen als unmittelbar vorbetonten Silben also nicht differenziert. Dementsprechend wird auch zwischen /i/ einerseits und /a/, /e/, /o/ andererseits nicht unterschieden. Es steht stets ( **ѐ** ). Die Realisierung von Vokalen an diesen Stellen ist auch abgesehen von Phänomenen der Mischung sowohl im Weißrussischen als auch im Russischen variativ und ohrenphonetisch nur begrenzt fassbar.

<i>Transkription</i>	
étê / écê	ru. <i>eti</i>
mamê	wr. <i>mame</i> ; ru. <i>mame</i>
vamê	wr. <i>vami</i> ; ru. <i>vami</i>

cjanê	wr. <i>cjahne</i> ; ru. <i>tjanet</i>
xodêt	wr. <i>chodzic'</i> ; ru. <i>chodit</i>
sinêm	wr. <i>sinim</i> ; ru. <i>sinem</i>

- Nach nicht-palatalisierten (nicht verhärteten) Konsonanten können in der Regel ( **â** ), ( **ɣ** ) und ( **u** ) auftreten. ( **ɣ** ) steht für /i/; für /a/ und /o/ steht ( **â** ).

<i>Transkription</i>	
mâlako	wr. <i>malako</i> ; ru. <i>moloko</i>
horâdâ	wr. <i>horada</i> ; ru. <i>goroda</i>
krasnyjê	ru. <i>krasnye</i>

- Nach verhärteten Konsonanten wird analog zur Position nach palatalisierten Konsonanten bei /a/, /e/, /o/ nicht unterschieden zwischen phonetisch unterschiedlichen Realisierungen. Es steht stets ( **ɣ** ). Dies gilt auch für nicht-palatalisiertes [ʃ] und nicht-palatalisiertes [r], wenn letzterem im Russischen ein palatalisiertes /r'/ entspricht. Nach nicht-palatalisiertem [ʃ] sowie nach nicht-palatalisiertem [r] wird auch /i/ mit ( **ɣ** ) wiedergegeben.

<i>Transkription:</i>	
bol'šý	wr. <i>bol'sy</i>
čatyrý	wr. <i>čatry</i> ; ru. <i>četyre</i>
čýlavek	wr. <i>čalavek</i> ; ru. <i>čelovek</i>
xočý	wr. <i>choča</i>
vyhrýlâ	wr. <i>vyjhrala</i> ; ru. <i>vyigrala</i>
zakonêýlâ	wr. <i>zakončyla</i> ; ru. <i>zakončila</i>

#### 3.4.2.4 Sonderfälle bei Vokalen

- Fälle des sogenannten „nepolnoe akanje“ werden durch ( **o** ) bzw., nach palatalisiertem Konsonanten, als ( **jo** ) wiedergegeben (*Beispiele*: maslô, pâčakajcjô).
- In Lehnwörtern kann ( **e** ) in unbetonten Silben auch nach nicht-palatalisierten Konsonanten auftreten (*Beispiele*: dékanat, kamp'jutér).

- Bei Nebenbetonungen werden die stärker differenzierenden Vokalzeichen ( **ı** ), ( **ʏ** ), ( **ɛ** ), ( **é** ), ( **ɑ** ), ( **o** ) und ( **ë** ) benutzt (*Beispiele*: kilahram, diagânal´, fotââparat, šestičýsâvoj).

#### 3.4.2.5 Grundsätzliches zu Konsonanten

Die Transkription der Konsonanten erfolgt in Anlehnung an die Konventionen der wissenschaftlichen Transliteration, mit folgenden Ausnahmen:

- [x] wird durch ( **ɣ** ) wiedergegeben.
- /j/ im (Silben-)Anlaut wird in der Transkription wiedergegeben. Dies gilt auch für prothetisches [j] vor /i/ (*Beispiele*: máladajê, jêdinêcâ, list´jêv, jix).
- Unsilbisches [ɥ] wird durch ( **ú** ) wiedergegeben.

#### 3.4.2.6 Palatalisiertheit

Die Palatalisiertheit eines Konsonanten wird auf unterschiedliche Arten signalisiert:

- Folgende Vokalzeichen signalisieren die Palatalisiertheit des vorhergehenden Konsonanten: ( **ı** ), ( **ɛ** ), ( **ë** ), ( **é** ).
- Vor den übrigen Vokalen drückt ein auf das Konsonantzeichen folgendes ( **j** ) die Palatalisiertheit des Konsonanten aus. Ausnahmen hiervon sind ( **šč** ) und ( **č** ). Hier drückt ein folgendes Apostroph ( **´** ) die Palatalisiertheit aus (*Beispiele*: ljudzê, šč´aslivâjê).
- Vor Konsonanten oder im Auslaut drückt ein auf das Konsonantzeichen folgendes Apostroph ( **´** ) die Palatalisiertheit des Konsonanten aus (*Beispiele*: tol´kê, byvajuc´).



### 3.4.2.7 Konsonantische Unterschiede zwischen dem Weißrussischen und dem Russischen

Folgende (im weiteren Sinne) phonische Unterschiede zwischen dem Weißrussischen und dem Russischen werden in den Transkriptionen festgehalten.<sup>15</sup>

- ru. [dʲ], [tʲ] vs. wr. [ɗʲ], [tʲ] (das sogenannte Dzekanje/Tsekanje des Weißrussischen):

Transkription bei ‚weißrussischer‘	Transkription bei ‚russischem‘	
<i>Affrikate:</i>	<i>Plosiv:</i>	
Dzeczê	detê	wr. <i>dzeci</i> ; ru. <i>deti</i>

- ru. plosives [g] vs. wr. frikatives [ɣ] oder [ɦ] (wobei zwischen den frikativen Realisierungen des Weißrussischen nicht differenziert wird. Auch Auslautverhärtung und Stimmtonassimilation werden – wie generell – nicht wiedergegeben):

Transkription bei ‚weißrussischem‘	Transkription bei ‚russischem‘	
<i>Frikativ</i>	<i>Plosiv</i>	
Глаз	glaz	ru. <i>glaz</i>
Мог	mog	wr. <i>moh</i> ; ru. <i>mog</i>

- ru. [l] vs. wr. [ɫ] im Präteritum Maskulinum und bei einigen Lexemen wie ru. *volk* – wr. *voŭk*:

Transkription bei ‚weißrussischem‘	Transkription bei ‚russischem‘	
<i>[ɫ]</i>	<i>[l]</i>	
Даў	dal	wr. <i>daŭ</i> ; ru. <i>dal</i>
Воўк	voŭk	wr. <i>voŭk</i> ; ru. <i>volk</i>

- ru. /v/ vs. wr. [u] oder [ɫ] wie in ru. *lavka* vs. wr. *lauka*; ru. *vsë* vs. wr. *ŭsë/usë*. Nicht unterschieden wird dagegen zwischen der Realisierung als [v] oder als [f]:

<sup>15</sup> Für eine Analyse der phonischen Variativität auf Basis der auf ohrenphonetischen Klassifizierungen beruhenden Transkriptionen des OK-WRGR vergleiche HENTSCHEL & ZELLER (2014). Eine akustische Analyse der phonischen Variation in der WRGR liefert ZELLER (i.Vorb.).

Transkription bei ‚weißrussischem‘ Vokal oder Halbvokal	Transkription bei ‚russischem‘ Frikativ	
uremê, úremê	vremê	ru. <i>vremja</i>
Laúkâ	lavka	wr. <i>laŭka</i> ; ru. <i>lavka</i>
usê, úsê	vsê	wr. <i>ŭsê/usê</i> ; ru. <i>vsê</i>

- ru. [u] vs. wr. [ŭ] für anlautendes /u/ nach Vokal:

Transkription bei ‚weißrussischem‘ unsilbischem [u]	Transkription bei ‚russischem‘ silbi- schem [u]	
úžé	užé	ru. <i>uže</i>

- Ru. palatalisiertes [ʃʲ] und [ʃʲ:] vs. wr. [ʃ] und [ʃʲ] sowie ru. palatalisiertes /rʲ/ vs. wr. nicht-palatalisiertes /r/. Die Palatalisiertheit wird im Auslaut und vor Konsonanten durch ein Apostroph ( ´ ) (ASCII 146) ausgedrückt, das Fehlen der Palatalisiertheit durch das Fehlen dieses Zeichens. Vor Vokal wird Palatalisiertheit durch die nachstehenden Vokalzeichen ( *i* ), ( *e* ), ( *ë* ), ( *ê* ) oder – im Falle von /rʲ/ – durch ein ( *j* ) bzw. – im Falle von [ʃʲ] und [ʃʲ:] – ein Apostroph ( ´ ) ausgedrückt. Fehlen der Palatalisiertheit wird durch folgendes ( *a* ), ( *é* ), ( *y* ), ( *o* ), ( *u* ), ( *ŷ* ) signalisiert:

Transkription bei ‚weißrussischen‘ nicht-palatalisierten Konsonanten	Transkription bei ‚russischen‘ palatalisierten Konsonanten	
Učorâ	učërâ	wr. <i>učora</i>
častâ	č´astâ	ru. <i>často</i> ; wr. <i>časta</i>
pamoč	pamoč´	ru. <i>pomoč´</i>
Radâm	rjadâm	ru. <i>rjadom</i>
Trox	trëx	wr. <i>troch</i> ; ru. <i>trëch</i>
ciper	ciper´	wr. <i>cjaper</i> ; ru. <i>teper´</i>

#### 3.4.2.8 Zur Wiedergabe der Ergebnisse einiger phonologischer Prozesse und weiterer Phänomene

- Es wird zwischen dem Suffix *-čsk-* und dem Suffix *-c-* unterschieden (*Beispiele*: *baranâvêčskêj* vs. *baranâvêckê*).
- Assimilationen von Sibilanten werden oft, aber von den Transkriptoren nicht durchgehend wiedergegeben. Häufig ist ohrenphonetisch,

d.h. für Transkriptoren, die Feststellung, ob Assimilation vorliegt oder nicht, problematisch. Die Transkription ist also vielfach „nicht belastbar“, etwa in der zweiten Person Sg. bei Reflexivverben (*Beispiele*: naješšê vs. najessê für ru. *naješ'sja*) oder im Lok. bzw. Dat. bei femininen Substantiven mit velarem Stammauslaut und Diminutivsuffix *-č-* (*Beispiele*: réčcŷ vs. réccŷ für wr. *rěčcy*).

- Palatalitätsassimilationen wie in [s']*liva* oder [z']*beh* werden in der Transkription nicht festgehalten (*Beispiele*: slivâ, zbeh).
- Stimmtonassimilationen und Auslautverhärtung wie in *no*[ʃ]*ka* oder *koro*[f] werden in der Transkription nicht wiedergegeben (*Beispiele*: nožkâ, karov).
- Die Folge aus der Infinitivendung und dem Reflexivpostfix (ru. graphemisch *-t'sja*, wr. *-cca*) wird als ( **ccý** ) wiedergegeben (*Beispiel*: dâgâvaricčý).
- Bei dem Laut [ts] wird nicht differenziert, ob er phonologisch (oder sogar morphologisch) weiter zerlegbar ist; er wird stets durch ( **c** ) wiedergegeben (*Beispiele*: saldackê für ru. *soldatskij* und wr. *sal-dacki*; dvaccýc´ für ru. *dvadcat'* und wr. *dvaccac'*).

#### 3.4.2.9 Besondere Wortformen

- Abkürzungen werden in der Regel transkribiert, aber nicht von allen Transkriptoren durchgehend. Neben Beispielen wie *ésémés*, *béémvé* finden sich auch Beispiele wie *bgu*.
- Klitika, die im Russischen/Weißrussischen mit Bindestrich geschrieben werden, werden im OK-WRGR ebenfalls mit Bindestrich wiedergegeben (*Beispiel*: dze-tâ).
- Auch andere Wortformen mit Bindestrich werden im OK-WRGR mit Bindestrich wiedergegeben (*Beispiele*: sinê-sinê, kojê-xto).
- Interjektionen, die aus mehreren Teilen bestehen, werden ebenfalls mit Bindestrich notiert (*Beispiele*: aj-aj-aj).
- Komposita, die im Russischen/Weißrussischen getrennt geschrieben werden, werden in der Regel durch Unterstrich verbunden (*Beispiele*: sjuda\_tuda).

- Unverständliches Material wird durch ( **xx** ) gekennzeichnet.

*Beispiel:*

\*RFK: za@c@pc xx@u jim@b@pc zâpissalê@c@pc , vylêčýúsê@b@pb  
toj@b@pc čêlavêk@c@pr .

- Eigennamen nicht-öffentlicher Personen sind aus Datenschutzgründen nicht transkribiert; stattdessen steht ( **persname** ) als Platzhalter.

*Beispiel:*

\*S3B: +" vyxadzi@r@pb za@c@pc persname persname !

- Vulgarismen sind nicht transkribiert, stattdessen steht ( **swearword** ) als Platzhalter.

*Beispiel:*

\*RFL: to@c@pc sâviščanêjê@r@ph , to@c@pc pêdsavet@c@pc ,  
to@c@pc iščê@b@ph jakajê@b@pb swearword .

### 3.4.3 Die sprachliche Affinität

Für jede Wortform im Korpus wird mithilfe von Tags ihre sprachliche Affinität zum Weißrussischen und / oder Russischen angegeben. Es sei deutlich hervorgehoben, dass die allgemeine Frage, was in gemischter Rede aus zwei nah verwandten und strukturell sehr ähnlichen Sprachen der einen oder der anderen Sprache zugeordnet oder als gemeinsam bzw. hybrid klassifiziert wird, in keiner Weise eine rein empirische Frage ist, sondern einer ganzen Reihe von theoretischen Festlegungen bedarf. Mit anderen Worten, hier sind prinzipiell mehrere Möglichkeiten der Beschreibung denkbar. Die Angaben im Korpus basiert auf einem strengen Klassifikationsalgorithmus, der im Detail in HENTSCHEL (2008) vorgestellt wurde. Die Bestimmung der Affinität und deren Kennzeichnung im Korpus verläuft getrennt auf zwei Ebenen: auf der phonetisch-phonologischen (im Folgenden *phonische Ebene*) und einer strukturell tieferen Ebene, die Phänomene der Lexik, Morphologie, Morphologie und Morphosyntax berücksichtigt (im Folgenden verkürzend *morphologische Ebene*).

Der Terminus „Affinität“ hebt primär auf die Ähnlichkeit / Identität einer Wortform zur bzw. mit der weißrussischen oder russischen Standardsprache ab bzw., wenn keine Zuordnung zu einer der beiden Standardsprachen getroffen werden kann, auf eine Klassifikation der Wortform als ‚gemeinsam‘ oder

‚hybrid‘ im Sinne des angesprochenen Algorithmus. Es ist sicherlich so, dass weißrussische Dialekte in manchen Fällen z.B. lexikalische Einheiten zeigen, die eher mit der funktional oder semantisch äquivalenten Einheit der russischen Standardsprache zusammenfallen und nicht mit der jeweiligen Entsprechung des weißrussischen Standards (vgl. RAMZA 2008). Von derartigen Verhältnissen wird abstrahiert. Dies hat eine Reihe von methodischen Gründen: Erstens verzeichnen einschlägige dialektale Nachschlagewerke nur einen Teil der im Korpus erfassten Lexik. Ein das ganze weißrussische Sprachgebiet umfassendes Dialektwörterbuch fehlt übrigens. Zweitens wären ohnehin nur die Mundarten einschlägig, die für die Respondenten im Korpus einschlägig sind. Drittens – und das ist der entscheidende Grund – ist zu beachten: Wenn ein Sprecher der WRGR in einer konkreten Äußerung ein Lexem verwendet, das in seiner Mundart und in der russischen Standardsprache das einschlägige ist, nicht aber in der weißrussischen, so ergibt sich in solchen konkreten Fällen in einer Äußerung in der Regel keinerlei Möglichkeit der Entscheidung, ob der betreffende Respondent das Lexem seiner Mundart oder der russischen Standardsprache „entnommen“ hat. Nur wenn ein sprachliches Zeichen standardsprachlich nicht vorliegt, werden die Verhältnisse in weißrussischen Dialekten und dem russischen Prostorečie berücksichtigt. Die Klassifizierung in die Kategorien ‚weißrussisch‘ und ‚russisch‘, die sich primär an der Affinität zu den standardsprachlichen Verhältnissen orientiert und nur bei „Fehlanzeigen“ in den Standardsprachen auf die Affinität zu Subvarietäten zugreift, stellt einen wissenschaftlich prüfbar (falsifizierbaren) methodischen Weg dar. Vielfach kann natürlich keine Entscheidung getroffen werden: Dann liegt entweder ein ‚gemeinsames‘ oder ein ‚hybrides‘ sprachliches Zeichen vor. Dazu an anderer Stelle mehr.

#### 3.4.3.1 Phonische Ebene

Auf der phonischen Ebene werden folgende sechs rein phonetisch-phonologisch beschreibbare Unterschiede zwischen dem Weißrussischen und dem Russischen berücksichtigt. Diese gehören sicherlich zu den prominentesten phonischen Unterschieden zwischen beiden Sprachen.

- ru. [dʲ], [tʲ] vs. wr. [ɕʲ], [ʂʲ] (das sogenannte Dzekanje/Tsekanje des Weißrussischen); vgl. ru. *deti* vs. wr. *dzeci*
- ru. plosives [g] vs. wr. frikatives [ɣ] oder [h]; vgl. ru. *noga* vs. wr. *naha*

- ru. palatalisiertes [rʲ] vs. wr. nicht-palatalisiertes [r], vgl. ru. *kričat* vs. wr. *kryčac*, ru. *rezat* vs. wr. *řezac*
- ru. [v] im Auslaut und vor Konsonant vs. wr. [ũ] oder positionsbedingt [u]; vgl. ru. *vrač* vs. wr. *urač* oder *ůrač*. Der Laut [ũ] wird auch dann als ‚weißrussisch‘ eingestuft, wenn er ru. [l] oder [u] entspricht, wie z.B. in *daũ* (wr. *daũ*, ru. *dal*) oder *ũžé* (ru. *uže*). Umgekehrt wird [l] oder [u], das wr. [ũ] entspricht, nicht als phonisch ‚russisch‘ bewertet, da [l] und [u] in vergleichbaren phonologischen Kontexten prinzipiell auch im Weißrussischen möglich sind.
- ru. Ikanje vs. wr. Jakanje, d.h. die Realisierung von unmittelbar vorbetontem /a/, /e/, /o/ nach palatalisierten Konsonanten als [a]-ähnlicher Laut im Weißrussischen, als [i]-ähnlicher Laut im Russischen; vgl. ru. *le[ɪ]žat* vs. wr. *ljažac*. Unterschiede zwischen beiden Sprachen bei Vokalreduktionen nach nicht-palatalisierten Konsonanten und nach palatalisierten Konsonanten in anderen unbetonten Silben werden nicht berücksichtigt.
- ru. palatalisiertes [ʃʲ] und [ʃʲ:] vs. wr. nicht-palatalisiertes [ʃ] und [ʃʲ]; vgl. ru. *č[ʃʲ]ert* vs. wr. *č[ʃ]ort*.

Enthält eine Wortform eine oder mehrere ‚weißrussische‘ Merkmale, so wird sie als phonisch ‚weißrussisch‘ klassifiziert. Enthält sie eines oder mehrere der ‚russischen‘ Merkmale, so wird sie als phonisch ‚russisch‘ klassifiziert. Enthält sie sowohl mindestens ein ‚russisches‘ als auch ein ‚weißrussisches‘ Merkmal, so wird sie als phonisch ‚hybrid‘ klassifiziert. Enthält sie keines der Merkmale, so wird sie als ‚gemeinsam‘ klassifiziert.

Kürzel	Erläuterung	Beispiele
@pb	phonisch ‚weißrussisch‘	<i>jamu@c@pr; try@c@pb</i>
@pr	phonisch ‚russisch‘	<i>jimuc@c@pr; tri@c@pr</i>
@pc	phonisch ‚gemeinsam‘	<i>jemu@c@pc; da@r@pc</i>
@ph	phonisch ‚hybrid‘	<i>jiščé@b@ph; havorêc´@b@ph</i>

### 3.4.3.2 Morphologische Ebene

Auf der sogenannten morphologischen Ebene, die – wie gesagt – Phänomene der Lexik, Morphologie, Morphologie und Morphosyntax umfasst, wird von Unterschieden auf der phonischen Ebene abstrahiert. So wird z.B. die Wortform *dzelâ* (gegenüber ru. *delo*, wr. *sprava*) trotz ‚weißrussischem‘ Dzekanje morphologisch nicht als ‚weißrussisch‘ oder ‚hybrid‘ qualifiziert, sondern als ‚russisch‘. Genauso wird *sibry* trotz (vermeintlich – vgl. HENTSCHEL & ZELLER 2014) ‚russischem‘ Ikanje anstelle von ‚weißrussischem‘ Jakanje (ru. *druz’ja*, wr. *sjabry*) als morphologisch ‚weißrussisch‘ und nicht als ‚russisch‘ oder ‚hybrid‘ bestimmt.

Die elementare Einheit für die Bestimmung der Affinität auf der morphologischen Ebene ist das Morph, d.h. die Wurzel, und – sofern gegeben – die Präfixe, Suffixe, Endung(en), Postfixe einer Wortform. Für jedes Morph wird gefragt, ob es in dieser Kombination aus Ausdruck und Inhalt aktueller Bestandteil des Weißrussischen oder des Russischen oder beider Sprachen ist. D.h., es gibt folgende Möglichkeiten, die Affinität der Morphe zu bewerten. Für diese Analyse wurden bestimmte Wörterbücher<sup>16</sup> herangezogen, und zwar das TSBM 1977, das BRS 2003 und das RBS 2005.

- spezifisch ‚russisch‘, wie z.B. die Wurzel des Substantivs *jazyk* (ru. *jazyk*; wr. *mova*) oder die Konjunktion *no* (ru. *no*; wr. *ale*).
- spezifisch ‚weißrussisch‘, wie z.B. die Wurzel des Substantivs *bac’kâ* (ru. *otec*; wr. *bac’ka*) oder die Präposition *dzelê* (ru. *radi*; wr. *dzelja*).

---

<sup>16</sup> Weißrussische Wörterbücher können im Einzelnen mehr oder weniger puristisch hinsichtlich der Anerkennung von Russismen als Lexeme des weißrussischen Standards sein. Es gibt also unterschiedliche Auffassungen in der Kodifizierung, die im Detail stark über den Gegensatz zwischen der „Taraškevica“ und der „Narkomaŭka“ hinausgehen. Da die Zahl der „aktiven“ Träger des Standardweißrussischen sehr begrenzt ist, gibt es (wie die Arbeit im hier besprochenen Forschungsprojekt zeigte) selbst unter weißrussischen Weißrussisten in Einzelfällen erhebliche Meinungsverschiedenheiten, ob ein gegebenes Lexem ein zu akzeptierendes Lehnwort aus dem Russischen oder ein russischer „Barbarismus“ ist. Methodisch bleibt hier nur die Anlehnung der Klassifizierung an bestimmte Wörterbücher. Es soll an dieser Stelle darüber hinaus nicht unerwähnt bleiben, dass in vielen Fällen auch unsere Kollegen an der Weißrussistik der Weißrussischen Staatlichen Universität unterschiedliche Einschätzungen zur Frage aufweisen, ob ein bestimmtes Wort ein Russismus ist, der nicht zur hochsprachlichen Norm gehört, oder ob es als akzeptierte Entlehnung aus dem Russischen zu gelten hat. Diese Normenunsicherheit ist angesichts der sprachlichen Gegebenheiten im Land nicht verwunderlich.

- ‚gemeinsam‘ (in beiden Sprachen vorhanden, also nicht zu spezifizieren), z.B. die Partikel *tak* 'so' (ru. *tak*; wr. *tak*; *tak* 'ja' hingegen wäre ‚weißrussisch‘ und *da* in derselben Bedeutung ‚russisch‘).
- ‚hybrid‘, wie z.B. die Wurzel bei dem Demonstrativpronomen *étyx* (ru. *étich*; wr. *hétych*)

Zum letztgenannten Beispiel: Es gibt auf der Ebene der Morphe einige ‚Hybride‘, deren Hybridität darin besteht, dass ihre morphonemische Repräsentation teilweise russische, teilweise weißrussische Züge trägt. So ist die Wurzel der Wortform *étyx* hybrid, da ihre russische Entsprechung *étich* einen palatalisierten Stammauslaut hat und die weißrussische Entsprechung *hétych* einen prothetischen Laut aufweist.

Wenn die vorliegenden Wortformen in ihren Morphen mit keiner der beiden Ausgangssprachen im Sinne der Standardsprache zusammenfallen, wird geprüft, ob sie in den weißrussischen Dialekten vorkommen. Ist das der Fall, so erhalten die Wortformen bzw. das Morph zusätzlich zu dem Wert ‚weißrussisch‘ den Wert ‚dialektal‘. Liegt auch in den weißrussischen Dialekten keine Entsprechung vor, so wird in einschlägigen Beschreibungen des russischen *Prostorečie* geprüft. Wenn die Wortform oder das Morph in der *Prostorečie* vorhanden ist, bekommt es zusätzlich zu dem Wert ‚russisch‘ den Wert ‚*prostorečie*‘. Der Wert ‚*prostorečie*‘ wird also nur dann vergeben, wenn eine Form weder im Standardrussischen noch im Standardweißrussischen noch in den weißrussischen Dialekten vorhanden ist. Wenn auch dann keine Einstufung möglich ist, wird das Morph oder die Wortform als ‚unklar‘ klassifiziert.

Die Affinität der einzelnen Wortformen wird anschließend (in der Regel; zu Ausnahmen s.u.) auf der Basis der Affinität ihrer morphologischen Bestandteile ‚errechnet‘. Folgende Kombinationen der Affinitäten von Morphen ergeben sich für Wortformen: Eine Wortform wird als ‚russisch‘ eingestuft, wenn sie ausschließlich aus ‚russischen‘ Morphen oder aus ‚russischen‘ und ‚gemeinsamen‘ Morphen besteht. Zu den ‚weißrussischen‘ Wortformen zählen entsprechend solche, die entweder nur ‚weißrussische‘ oder ‚weißrussische‘ und ‚gemeinsame‘ Morphe enthalten. ‚Gemeinsame‘ Wortformen bestehen entsprechend nur aus ‚gemeinsamen‘ Morphen. Außer ‚weißrussischen‘, ‚russischen‘ und ‚gemeinsamen‘ Wortformen gibt es schließlich auch ‚hybride‘ Wortformen, die sowohl ‚russische‘ als auch ‚weißrussische‘ (und ggf. ‚gemeinsame‘) Morphe bzw. ein ‚hybrides‘ Morph enthalten. D.h., dass



‚gemeinsame‘ Morphe bei der Bestimmung der Affinität keine Rolle spielen, wenn außerdem ein spezifisch ‚weißrussisches‘ oder ein spezifisch ‚russisches‘ Morph (oder ein ‚hybrides‘) in der Wortform enthalten ist.

Kürzel	Bedeutung	mögliche Morphkombinationen
@b	‚weißrussisch‘	‚weißrussisch‘ (+ ‚gemeinsam‘)
@r	‚russisch‘	‚russisch‘ (+ ‚gemeinsam‘)
@c	‚gemeinsam‘	‚gemeinsam‘
@h	‚hybrid‘	‚weißrussisch‘ + ‚russisch‘ (+ ‚gemeinsam‘); ‚hybrid‘ (+ ‚russisch‘ / ‚weißrussisch‘ / ‚gemeinsam‘)
@u	unklar	‚unklar‘ (+ ‚gemeinsam‘ / ‚russisch‘ / ‚weißrussisch‘ / ‚hybrid‘)

Zusätzlich erhalten die Wortformen gegebenenfalls Kürzel für ‚dialektal‘ oder ‚prostorečie‘.

Kürzel	Bedeutung
@pros	‚(russisch-)prostorečie‘
@dial	‚(weißrussisch-)dialektal‘

Es kann allerdings der Fall auftreten, dass alle Morphe einer Wortform ‚gemeinsamer‘ Affinität sind, die Wortform selbst jedoch in dieser Bedeutung nur in einer der beteiligten Sprachen auftritt, wie z.B. die Wortform *žonkâ*. Die Wurzel *žon/žen-*, das Suffix *-k-* und die Endung *-a* bei *žonkâ* (ru. *žena*, Pl. *žěny*, wr. *žonka*) gibt es sowohl im Russischen als auch im Weißrussischen, die Kombination aus diesen Morphen kommt aber als unmarkierte Bezeichnung für die Ehefrau nur im Weißrussischen vor. In diesem Falle wird die Wortform *žonkâ* „holistisch“ als ‚weißrussisch‘ qualifiziert.

Ähnliches kann sich durch die Berücksichtigung der Betonungsverhältnisse ergeben. Die Betonung hat sowohl im Russischen als auch im Weißrussischen bedeutungsunterscheidende Funktion. Sie ist in beiden Sprachen morphologisiert, d.h., sie muss mit den Lexemen erlernt werden, und zwar sowohl lexemorientiert wie bei wr.=ru. *muka* 'Qual' vs. *muka* 'Mehl' als auch formenorientiert wie bei ru. *goroda* 'Stadt; Gen.Sg.' vs. *goroda* 'Stadt; Nom.Pl.'. Die Betonung im Weißrussischen und Russischen kann nicht als ausschließlich phonisches Phänomen modelliert werden, da sie bekannterweise in beiden Sprachen lexikalisch bzw. morphologisch motiviert ist. So gilt eine Wortform *vesci* mit Betonung auf der ersten Silbe (vgl. ru. *vesti*

mit der Betonung auf der letzten Silbe, wr. *vesci* dagegen mit Betonung auf der ersten Silbe) als ‚weißrussisch‘, obwohl ihre morphologischen Bestandteile, also die Wurzel *ves-* und die Endung *-ti / -ci* ‚gemeinsam‘ sind.<sup>17</sup>

Bei freien Elementen wird paradigmatisch klassifiziert, wie z.B. bei Wurzeln wie in ru. *bumag-a* vs. wr. *paper-a*, das heißt, dass die Wurzeln ohne syntagmatischen Kontext als ‚weißrussisch‘ oder ‚russisch‘ bestimmt werden. Die gebundenen Elemente dagegen – Endungen, Präfixe oder auch die meisten Präpositionen – sind oft nur nach syntagmatischen Kriterien einer der beiden Sprachen zuzuordnen, wenn sie in beiden Sprachen homonym im Inventar vorliegen. Aus diesem Grund wird in solchen Fällen die syntagmatische Einbettung des Morphs berücksichtigt: So ist z.B. die Präposition *pa/po* ‚als solche‘, d.h. paradigmatisch, in beiden Sprachen vorhanden und stimmt in den meisten Verwendungen überein. In einer Konstruktion wie in *pajsci pa xleb* ‚Brot holen gehen‘ liegt jedoch eine spezifisch weißrussische Verwendung der Präposition *pa* vor, so dass sie in solchen Belegen als ‚weißrussisch‘ bestimmt würde. Im Russischen wird der gleiche Inhalt mit Hilfe der Präposition *za* ausgedrückt: *pojti za xlebom*. In einer solchen Äußerung würde die Präposition *za* entsprechend als ‚russisch‘ bestimmt. Ähnlich wird bei Flexionsendungen verfahren: Bei der Bestimmung einer Endung, die generell in beiden Sprachen vorhanden ist, als ‚weißrussisch‘, ‚russisch‘ oder ‚gemeinsam‘ wird der syntagmatische Kontext berücksichtigt. Entscheidend ist, ob eine Verbindung aus einem gegebenen Stamm und einer gegebenen Endung entweder nur im Weißrussischen, nur im Russischen oder in beiden Sprachen möglich ist: So ist bei Substantiven im Nominativ Plural die Endung *-y* (bzw. *-i*) prinzipiell in beiden Sprachen möglich: ru. *student-y* und wr. *studěnt-y*. Bei der Wortform *profesârŷ* wird die Endung *-y* dagegen als ‚weißrussisch‘ bestimmt, die ‚russische‘ Entsprechung wäre mit der Endung *-a*; vgl. ru. *professor-a*; wr. *profesar-y*.

Grundsätzlich verschiedene Elemente, wie z.B. wr. *kalja* und ru. *okolo*, sind den beiden Sprachen natürlich ohne Berücksichtigung des Kontexts zuzuordnen.

Da die Angaben zur sprachlichen Affinität auf der phonischen und der „tieferen“ morphologisch-morphosyntaktischen Ebene zentral sind für jede Analyse der weißrussisch-russischen gemischten Rede, werden sie – um eine

---

<sup>17</sup> In der veröffentlichten Version des OK-WRGR sind die Betonungsverhältnisse nicht abgebildet.

möglichst flexible Suche im Korpus zu gewährleisten – im OK-WRGR an zwei Stellen angegeben. Zum einen werden sie in der Zeile %mor (der Zeile der Annotationen) angegeben, worauf in Abschnitt 3.5 eingegangen wird. Zum anderen werden sie in der Zeile der transkribierten Äußerungen an die Wortformen angehängt. Zunächst erfolgt dabei die Angabe der morphologischen Affinität, dann die der phonischen Affinität, dann gegebenenfalls die Angabe, ob die Wortform als weißrussisch-, ‚dialektal‘ oder als ‚prostorečie‘ einzustufen ist.

Beispiel	Erläuterung
ja@c@pc	‚gemeinsam‘; phonisch ‚gemeinsam‘
dumâjêcê@c@pb	‚gemeinsam‘; phonisch ‚weißrussisch‘
kudy@b@pc	‚weißrussisch‘; phonisch ‚gemeinsam‘
vot@r@pc	‚russisch‘; phonisch ‚gemeinsam‘
pakudâvâ@r@pc@pros	‚(russisch-)prostorečie‘; phonisch ‚gemeinsam‘
brynčéc´@b@pb@dia1	‚(weißrussisch-)dialektal‘; phonisch ‚weißrussisch‘

#### 3.4.4 Weitere wortformbezogene Informationen

- *Weakforms* wie *tada* (ru. *togda*) oder *tokâ* (ru. *tol'ko*) erhalten zusätzlich zu dem Sprachkürzel die Kennzeichnung ( @w ) (Beispiel: *tada@r@pc@w*<sup>18</sup>).
- *Metasprachliche Nennungen* von Lauten und Buchstaben erhalten zusätzlich zu dem Sprachkürzel die Kennzeichnung ( @1s ). (Beispiel: *o@c@pc@1s*; siehe auch unten 3.4.5.3)
- Abgebrochene oder nur teilweise verständliche Wortformen erhalten das Kürzel ( @k ) (Beispiel: *dastu@k@pc*).

<sup>18</sup> Es ist bekannt, dass für eine Form wie *tada* in weißrussischen oder russischen Subvarietäten eine gewisse Unsicherheit besteht, ob „noch“ von einer phonetisch schwachen Form oder „bereits“ von einer verfestigten (morphologisierten) phonologischen Form ausgegangen werden kann. Da aus den oben dargelegten methodischen Gründen ein Abgleich mit den standardsprachlichen Normen im Vordergrund steht, wurde die erste Lösung gewählt.

### 3.4.5 Interpunktion

#### 3.4.5.1 Einfache Satzzeichen

- Nach den üblichen Zeichensetzungsregeln stehen die Interpunktionszeichen ( , ), ( . ), ( ! ) und ( ? ).
- Der Doppelpunkt ( : ) steht vor Aufzählungen (nicht bei wörtlicher Rede, s.u.):

*Beispiel:*

```
*SFA: my@c@pc što@c@pc kupilê@c@pc sa@b@pc svadzêbnyx@r@pb  
dzenêh@r@pb : škaf@r@pc , tryl'jaž@c@pb i@c@pc  
kravac'@r@pb .
```

- In Kopulasätzen / in elliptischen Sätzen, in denen die weißrussische und russische Orthographie einen Gedankenstrich fordern, steht zusätzlich zu diesen in Chat gebräuchlichen Interpunktionszeichen ( [- ] ).

*Beispiel:*

```
*AFP: tam@c@pc balotâ@c@pc , paétâmu@r@pc , v@r@pr  
pryncýpê@c@pb , samâ@b@pc âpcimal'nâ@r@pb [-]  
izmel'čenêjê@r@pr i@c@pc razbrasyvânêjê@r@pc .
```

#### 3.4.5.2 Direkte Rede

Direkte, wörtliche Rede wird doppelt gekennzeichnet, am übergeordneten Satz (sofern vorhanden) und an der wörtlichen Rede selbst:

- Wenn dem übergeordneten Satz wörtliche Rede folgt: ( +"/. )
- Wenn dem übergeordneten Satz wörtliche Rede vorausgeht: ( +". )

- Am Anfang der wörtlichen Rede und am Anfang jeder weiteren Äußerung in wörtlicher Rede: ( "+" )

*Beispiele:*

\*S3B: i@c@pc vsë@c@pr ja@c@pc (.) na@c@pc bac'ku@b@pb  
hâvaru@c@pb +"/.

\*S3B: +" bac'kâ@b@pb , vaz'mi@c@pc ty@c@pc âtruhaj@r@pb .

\*S3B: +" i@c@pc kak-tâ@r@pc vot@r@pc vyžýlê@c@pc .

\*S3B: hâvaryt@r@pb +".

- Anders als in CHAT üblich: Wenn in die wörtliche Rede ein kurzes übergeordnetes Element (wie z.B. dumâju, kažý, havoryc' o.ä.) eingeschoben ist, wird dieses übergeordnete Element nicht als eine eigene Äußerung gewertet. Stattdessen wird der eingeschobene Teil innerhalb der wörtlichen Rede folgendermaßen gekennzeichnet: ( ["-"] )

*Beispiel:*

\*SAF: +" ja@c@pc haryt@r@pbêw ["-"] iščo@r@ph sutkê@c@pc  
pabudu@c@pc .

### 3.4.5.3 Metasprachliche Referenz

Metasprachliche Elemente werden durch das Zeichen ( [""] ) gekennzeichnet. Dieses Zeichen zeigt sowohl metasprachliche Elemente im eigentlichen Sinne als auch appositive Namen, Titel usw an. Bei Bezug auf mehrere Wortformen wird der Skopus durch Spitzklammern angegeben.

*Beispiele:*

\*S3B: nu@c@pc , slovâ@c@pc bëlâruskâjê@c@pc , zarâz@b@pb [""] .

\*L2B: u@c@pc nix@r@pc bolêjê@r@pc tako@c@pc o@c@pc@ls [""]  
vydzêlênâ@c@pb , v@r@pr râsian@c@pc .

\*SI2: nu@c@pc my@c@pc ix@r@pc zavém@c@pc ix@r@pc <kléckê@c@pc  
z@b@pc mjasâm@c@pc>["] .

\*INT: a@c@pc vot@r@pc znajêcê@c@pb , možý@b@pc , takuju@c@pc  
hrupu@c@pb krâmbambulê@c@pc ["] ?

#### 3.4.5.4 Wiederholungen, Korrekturen, Umformulierungen, Übersetzungen

Das CHAT-System verwendet als Markierungen für Wiederholungen ( [/\_] ), für Korrekturen ( [/\_/] ), für Umformulierungen ( [/\_//] ). CHAT sieht vor, dass das sprachliche Material, das im Skopus dieser Markierungen steht, nicht (grammatisch usw.) annotiert wird. Daher werden diese Zeichen im OK-WRGR nicht verwendet. Dies verhindert, dass Wiederholungen usw. bei einigen CLAN-Analysewerkzeugen ausgeschlossen werden, ermöglicht aber andererseits die grammatische Annotation dieses Materials. Folgende Markierungen werden verwendet:

- Wiederholungen: ( [/\_] )
- Korrekturen: ( [/\_/] )
- Umformulierungen: ( [/\_//] )
- Übersetzungen: ( [/\_///] )

„Übersetzung“ ist in einem weiten Sinne zu verstehen: Als Übersetzungen werden Stellen markiert, in denen eine Wortform oder mehrere Wortformen „gleichen“ Inhalts nacheinander stehen, die unterschiedlicher sprachlicher Affinität sind. Dabei wird nicht differenziert nach wörtlichen und freien Übersetzungen.

Diese vier Zeichen stehen nach dem Element, das wiederholt, umformuliert, korrigiert, übersetzt wird, also vor der Wiederholung, Korrektur, Umformulierung, Übersetzung. Ist der Skopus dieser Zeichen weiter als eine Wortform, so wird der Skopus durch Spitzklammern deutlich gemacht.

#### *Beispiele:*

\*CFK: jany@b@pb ščé@b@pb@w tut@c@pc máladéjê@b@pc@dialog [/\_/]  
máladéjê@b@pc@dialog saúsem@r@pb .

- \*CI4: synâ@c@pc vydava@k@pc zam@k@pc [\_//] žynilâ@c@pc ,  
nu@c@pc .
- \*AI2: da@r@pc , étym@h@pc ja@c@pc [\_///] bez@c@pc étâvâ@r@pc  
ne@c@pc magu@c@pcr .
- \*AFA: nu@c@pc jano@b@pb , znajěš@c@pc , jak@b@pc što@c@pc  
<pad@c@pc škurâju@b@pc>[\_///] pad@c@pc kožýju@h@pc  
jak@b@pc što@c@pc tam@c@pc , o@c@pc tak@c@pc  
pêrýcirajê@b@pb .

### 3.4.5.5 Unvollständige Sätze

Hier gibt es keine Unterschiede zu den CHAT-Konventionen. Es wird unterschieden nach Sätzen, die vom Sprecher selbst abgebrochen werden, und solchen, die von einem zweiten Sprecher unterbrochen werden.

- Nach abgebrochenen, nicht vom Gesprächspartner unterbrochenen Äußerungen: ( + . . . )
- Nach abgebrochenen, nicht vom Gesprächspartner unterbrochenen Fragen: ( + . . ? )
- Nach vom Gesprächspartner unterbrochenen Äußerungen: ( + / . )
- Nach unterbrochenen Fragen: ( + / ? )
- Wird eine unterbrochene Äußerung fortgesetzt, so steht am Anfang der Fortsetzung: ( + , )

#### *Beispiele:*

- \*S7B: budut@r@pc učyccý@r@pb , možý@b@pc , kto-tâ@r@pc ú@b@pb  
minsku@b@pc astaněccý@r@pc , možý@b@pc +...
- \*S5B: pajedzêm@c@pb pa@c@pc rasijê@c@pc , za@c@pc hranicâj@c@pb  
+ / .
- \*INT: nu@c@pc , vot@r@pc râskažycê@c@pb .
- \*S5B: + , pajez´dzêú@b@pb .

### 3.4.5.6 Pausen und Überlappungen im Redefluss

Hier gibt es keine Unterschiede zu CHAT. Zu vermerken ist lediglich, dass eine eher laxe Notierung von Pausen und Überlappungen erfolgt.

- Pausen werden nach der Länge unterschieden. Kurze Pausen werden durch ( . ) gekennzeichnet, lange durch ( . . ). Nur extrem lange Pausen werden durch ( . . . ) gekennzeichnet.
- Bei Überlappungen, d.h., wenn ein zweiter Sprecher seine Äußerung beginnt, während der vorherige Sprecher noch spricht, steht am Anfang der zweiten Äußerung ( +< ).

*Beispiele:*

\*RFN: ja@c@pc pašla@r@pc (..) v@r@pc adzinâccýty@h@pb klas@c@pc .

\*CFA: nja@c@pb budzêc'@b@pb@dial žý@c@pc jna@b@pc úsem@r@pb  
xadzic'@c@pb raskazyvâc'@r@pb .

\*CFC: +< a@c@pc mamkâ@b@pc nikomu@b@pc ne@c@pc havorêc'@b@ph .

## 3.5 Die Annotationen in der Zeile %mor

Die Zeile %mor enthält das weißrussische und / oder das russische Lemma, Angaben zur morphologischen und phonischen Affinität sowie zu Wortart und grammatischen Werten.

CHAT sieht für die Angabe von „Suffixen“ einen Bindestrich ( - ) vor, für „fusional or infixed morphology“ ein ( & ), für Präfixe ein ( # ) (vgl. MCWHINNEY 2013a, 109). Diese Zeichen werden im OK-WRGR für die Angabe der Affinität und der grammatischen Kategorien verwendet. Anders als im CHAT-System angedacht, geht es im OK-WRGR nicht um eine morphologische Analyse, sondern um die Angabe der Werte der grammatischen Kategorien. Es wird daher nicht unterschieden zwischen „fusional suffixes“ und „suffixes“, also nicht differenziert, wie die grammatische Information kodiert ist.

Die im OK-WRGR verwendeten Kürzel für Wortarten und grammatische Werte orientieren sich an den CHAT-Konventionen, weichen jedoch in einigen Aspekten von ihnen ab.



Das Schema der Annotationen sieht folgendermaßen aus:

```
MorphologischeAffinität:PhonischeAffinität#Wortart|_Weißrussi-
schesLemma_RussischesLemma_&(Kasus/Tempus)-(Genus/Per-
son):Numerus
```

Es erfolgt also

- zunächst die Angabe der morphologischen Affinität;
- dann nach einem Doppelpunkt ( : ) die Angabe der phonischen Affinität;
- dann nach einem ( # ) die Angabe der Wortart;
- dann nach einem ( | ) die Angabe des weißrussischen und/oder des russischen Lemmas;
- dann (bei flektierten Wortarten) nach einem ( & ) die Angabe von Kasus oder Tempus;
- dann nach einem ( - ) die Angabe von Genus oder Person;
- dann nach einem ( : ) die Angabe des Numerus.

Bei nicht-flektierten Wortarten erfolgt nur die Angabe der sprachlichen Affinität, der Wortart und des Lemmas.

### 3.5.1 Die sprachliche Affinität

Die Angabe der Affinität erfolgt mithilfe des Markers ( # ), der in CHAT eigentlich zur Angabe von Präfixen vorgesehen ist. Es erfolgt zunächst die Angabe der morphologischen Affinität, dann die der phonischen Affinität, getrennt durch einen Doppelpunkt.

```
MorphologischeAffinität:PhonischeAffinität#Wortart|_Weißrussi-
schesLemma_RussischesLemma_&(Kasus/Tempus)-(Genus/Per-
son):Numerus
```

Die verwendeten Kürzel sind identisch mit den Kürzeln auf den Sprecherzeilen (vgl. 3.4.3.).

*Beispiel:* **r:pg#pro:pers:ana|\_0\_oni\_&NOM-0:PL**

### 3.5.2 Wortart

Die Angaben zur Wortart stehen nach der Angabe zur sprachlichen Affinität und vor der Angabe des Lemmas, getrennt durch den Marker ( | ).<sup>19</sup>

Affinität#**Wortart**|\_WeißrussischesLemma\_RussischesLemma\_&(Kasus/Tempus)-(Genus/Person):Numerus

An dieser Stelle erfolgt bei Adjektiven und Adverbien auch die Angabe, ob es sich um eine Komparativform handelt.

<i>Kürzel</i>	<i>Bedeutung</i>
adj:abrev	Abbraviatur als Adjektiv (z.B. ru. <i>žédé</i> als Kurzform für ru. <i>železnodorožnyj</i> )
adj:brev	Adjektiv in Kurzform
adj:cp	Adjektiv im Komparativ
adj:indecl	nicht-deklinierbares Adjektiv
adj:loc	Von einem Ortsnamen abgeleitetes Adjektiv
adj:poss	Possessivadjektiv
adj:prop	Adjektivischer Eigenname
adv	Adverb
adv:conj	Adverb/Konjunktion
adv:cp	Adverb im Komparativ
adv:pro	Adverb/Pronomen
adv:ptl	Adverb/Partikel
co	Interjektion
conj	Konjunktion
conj:ptl	Konjunktion/Partikel
conj:pro	Konjunktion/Pronomen
disc	Diskursmarker
n:abrev	Abbraviatur
n:adj	Substantiviertes Adjektiv

<sup>19</sup> Ein notorisches Problem bei der Wortartbestimmung ist die Abgrenzung nicht-flektierter Wortarten untereinander. Da angesichts der großen Datenmenge ein Großteil der Wortartbestimmung von studentischen Hilfskräften durchgeführt wurde, wurde im OK-WRGR ein einfaches, nicht ganz widerspruchsfreies Modell gewählt. Daher kann es in Einzelfällen zu Variation bei der Einordnung kommen.

n:adj:anim	Substantiviertes Adjektiv, belebt
n:adj:inanim	Substantiviertes Adjektiv, unbelebt
n:adj:prop	Substantiviertes, von einem Eigennamen abgeleitetes Adjektiv
n:anim	Substantiv belebt
n:inanim	Substantiv unbelebt
n:loc	Ortsbezeichnung
n:prod	Produktbezeichnung
n:prop	Eigennamen
na	keine Angabe
num:kard	Kardinalzahl
num:koll	Kollektivzahl
num:ord	Ordinalzahl
num:pro	Zahlwort/Pronomen
part:akt	Partizip Aktiv
part:pass	Partizip Passiv
part:pass:brev	Partizip Passiv Kurzform
pred	Prädikativum
prep	Präposition
pro:dem	Demonstrativpronomen
pro:dem:kop	<i>éto / hěta</i> in Kopulasätzen
pro:det	Determinativpronomen (z.B. ru. <i>vsjakij, každyj</i> )
pro:ind	Indefinitpronomen
pro:int	Interrogativpronomen
pro:neg	Negationspronomen
pro:pers:ana	Anaphorisches Personalpronomen
pro:pers:dei	Deiktisches Personalpronomen
pro:poss:ana	Anaphorisches Possessivpronomen
pro:poss:dei	Deiktisches Possessivpronomen
pro:poss:refl	Reflexives Possessivpronomen
pro:ptl	Pronomen/Partikel
pro:refl	Reflexivpronomen
pro:rez	Reziprokes Pronomen
pro:sonst	Sonstiges Pronomen
ptl	Partikel
v:comasp	biaspektuelles Verb
v:ger	Adverbialpartizip

v:ipf	Verb imperfektiv
v:pf	Verb perfektiv

### 3.5.3 Lemma

Dort, wo CHAT den „Stamm“ vorsieht, steht im OK-WRGR die Nennform (also der Infinitiv bei Verben, der Nominativ Singular bei Substantiven; Genaueres siehe unten).

Affinität#Wortart|\_WeißrussischesLemma\_RussischesLemma\_&(Kasus/Tempus)-(Genus/Person):Numerus

Einige Wortformen werden sowohl einem weißrussischen als auch einem russischen Lemma zugeordnet, andere nur einem weißrussischen oder nur einem russischen. Entscheidend bei dieser Zuordnung ist die Affinität des Stamms; die Klassifikation von Endungen spielt keine Rolle. Es wird nur die morphologische Ebene berücksichtigt.

- Bei einem als ‚russisch‘ bestimmten Stamm erfolgt die Angabe eines russischen Lemmas, keines weißrussischen. Z.B. wird die Wortform *jizyk*, deren Stamm als ‚russisch‘ bestimmt wird, dem russischen Lemma *jazyk* zugeordnet, jedoch keinem weißrussischen Lemma.
- Bei einem ‚weißrussischen‘ Stamm wird die Wortform einem weißrussischen Lemma zugeordnet. So wird der Stamm in *mová* als ‚weißrussisch‘ bestimmt und die Wortform dem weißrussischen Lemma *mová* zugeordnet.
- Bei einem ‚gemeinsamen‘ Stamm wird die Wortform sowohl einem weißrussischen als auch einem russischen Lemma zugeordnet. So wird z.B. der Stamm in *syn* als ‚gemeinsam‘ bestimmt und die Wortform sowohl einem russischen als auch einem weißrussischen Lemma *syn* zugeordnet. Wenn der Stamm als ‚gemeinsam‘ bestimmt wird, aber die Wortform holistisch als ‚russisch‘ klassifiziert wird, wird die Wortform nur einem russischen Lemma zugeordnet. Bei der Wortform *áb’jasnilá* sind alle Morphe und entsprechend der Stamm als ‚gemeinsam‘ bestimmt. Holistisch wurde die Wortform jedoch als ‚russisch‘ klassifiziert, d.h., dass die Wortform nur dem russischen Lemma *ob’jasnjat’* zugeordnet wird. Wenn umgekehrt der Stamm als ‚gemeinsam‘ bestimmt wird, holistisch die Wortform

aber als ‚weißrussisch‘ klassifiziert wird, wird die Wortform nur einem weißrussischen Lemma zugeordnet. Z.B. wird die Wortform *žonkâ* nur dem weißrussischen Lemma *žonka* zugeordnet.

- Bei einem ‚hybriden‘ Stamm gilt das Ähnlichkeitsprinzip: Je nachdem, ob die Wortform dem Äquivalent in der weißrussischen oder in der russischen Sprache „ähnlicher“ ist, wird sie einem weißrussischen oder einem russischen Lemma zugeordnet. Z.B. wird die Wortform *âbižaúsê* einem russischen Lemma zugeordnet, und zwar *obižat'sja*, die Wortform *sënnëšněj* dagegen einem weißrussischen Lemma, und zwar *sënnjašni*. Wo es nicht zu entscheiden ist, wird die Wortform einem weißrussischen und einem russischen Lemma zugeordnet. So wird z.B. *hétix* sowohl unter das russische Lemma *étot* als auch unter das weißrussische Lemma *hëty* subsumiert.
- Weder einem russischen noch einem weißrussischen Lemma werden Wortformen zugeordnet, die einen Stamm mit ‚unklarer‘ Affinität aufweisen.

Die Darstellung der Angaben zu den Lemmata im OK-WRGR in der Zeile **mor%** sieht wie folgt aus: Ist die Wortform sowohl einem weißrussischen als auch einem russischen Lemma zuzuordnen, so wird zunächst das weißrussische, dann das russische Lemma angegeben, getrennt durch einen Unterstrich (    ). Vor dem weißrussischen und nach dem russischen Lemma steht ebenfalls ein Unterstrich. Ist die Wortform nur einem weißrussischen oder nur einem russischen Lemma zuzuordnen, so wird die Position des anderen Lemmas durch ( 0 ) gefüllt. (Ist sie keinem Lemma zuzuordnen, so steht an beiden Positionen ( 0 ).)

<i>Wortform</i>	<i>Affinität des Stammes</i>	<i>Lemmaangabe</i>
ja	‚gemeinsam‘	<u>ja</u> <u>ja</u>
paperâ	‚weißrussisch‘	<u>papera</u> 0
bumahâ	‚russisch‘	0 <u>bumaga</u>
étyjê	‚hybrid‘	<u>hëty</u> <u>étot</u>

Folgendes ist bei der Notation des Lemmas zu beachten:

- Bei der Transliteration der Lemmata gelten im Wesentlichen die Konventionen der wissenschaftlichen Transliteration, wobei Folgendes zu beachten ist:

<i>Kyryllisch</i>	<i>Transliteration im OK-WRGR:</i>
x	ch (anders als bei der Transkription der Wortformen)
э	é
ѣ	´ (ASCII 180)
ѝ	´´
’	’
ÿ	ú
wr. 2	h
ru. 2	g
wr. 3	ě (ASCII 235)
ru. 3	e

- Alle Lemmata werden kleingeschrieben.
- Die Teile von Komposita, die im Russischen/Weißrussischen nicht zusammengeschrieben werden, sind durch ( / ) verbunden (*Beispiel:* `_0_sjuda/tuda_`).
- Gleiches gilt für Klitika, die im Russischen/Weißrussischen mit Bindestrich geschrieben werden (*Beispiel:* `_dzes´ci_gde/to_`).
- Personennamen erhalten auch auf der Lemma-Ebene die Angabe `_persname_`.
- Als Lemmata für Kurzadjektive (bzw. Kurzpartizipien) werden in der Regel Langadjektive eingetragen, es sei denn, es existiert keine lange Form (wie z.B. bei ru. *dolžen* oder wr. *pavinen*). In solchen Fällen wird die Kurzform als Lemma eingetragen.
- Als Lemmata von Adverbialpartizipien werden nicht Infinitive, sondern Adverbialpartizipien eingetragen (*Beispiel:* `_pryvykšy_0_`).
- Bei anaphorischen Pronomen wird die Genus- und Numerus-Differenzierung beibehalten: *jejo* bekommt `_0_ona_` als Lemma, *jevo* dagegen `_0_on_` oder `_0_ono_`, *ix* erhält `_0_oni_`.

- Bei ‚russischen‘ Komparativen wird die Komparativform als Lemma eingetragen, und zwar mit dem Suffix *-ee* und nicht *-ej*, z.B. `_0_dobree_`.
- Bei Wortformen, die mit *pol-* beginnen, wird *pol-* für die Zuordnung zu einem Lemma ignoriert. Z.B. erhält die Wortform *polstala* als Lemmaangabe `_0_stol_`.
- Bei kurzen Varianten von Partikeln werden volle Varianten als Lemma angegeben. Z.B. erhält die Partikel *b* als Lemmaangabe `_by_by_`.
- Bei langen Varianten von Präpositionen werden die kurzen Varianten als Lemma angegeben. Z.B. wird bei der Präposition *va* als Lemma `_0_v_` eingetragen.
- Bei weißrussisch dialektalen Einträgen werden standardsprachliche Entsprechungen bevorzugt. Z.B. wird bei der dialektalen Wortform *dak* die standardsprachliche Form `_dyk_0_` eingetragen, bei der dialektalen Wortform *ina* die standardsprachliche Form `_jana_0_`. Wenn keine ausdrucksseitige Ähnlichkeit zwischen der dialektalen und der standardsprachlichen Form besteht, wird die dialektale Form als Lemma eingetragen.
- Bei Wortarten-Homonymien können unterschiedliche Lemmata vorliegen: Wenn *étâ* (ru. *éto*) oder *hétâ* (wr. *héta*) als Partikel vorliegen, wird als Lemma `_0_éto_` bzw. `_héta_0_` angegeben, bei einer Verwendungsweise als Demonstrativpronomen dagegen `_0_étot_` bzw. `_hétý_0_`.
- Bei Okkasionalismen, die keine Entsprechung im Wörterbuch haben, für die aber eine Nennform herleitbar ist, wird diese notiert. Z.B. wird bei der Wortform *âtžalelâ* als Lemma `_0_otžalet´_` angegeben.
- Bei abgebrochenen Wortformen, deren Lemma aus dem Kontext konstruiert werden kann, wird dieses notiert. Z.B. erhält `topč@k` (intendiert: ru. *topčetes´*) als Lemmaangabe `_0_toptat´sja_`.
- Abgebrochene Einträge, deren Lemma aus dem Kontext nicht zu konstruieren ist, oder andere nicht bestimmbare Fälle werden keinem

Lemma zugeordnet, d.h., für das russische und das weißrussische Lemma wird ein Nullwert eingetragen (\_0\_0\_).

- Bei Teilen von idiomatischen Ausdrücken wird das Lemma konstruiert. Z.B. wird bei dem idiomatischen Ausdruck *z buchy barachty* für die Wortform *buchy* als Lemma *\_0\_buchta\_* angegeben, für *barachty* wird als Lemma *\_0\_barachta\_* angegeben.
- Bei russischen dialektalen Wortformen werden diese auch als Lemma angegeben. Z.B. wird bei der russisch-dialektalen Wortform *ichnëj* als Lemma *\_0\_ichnij\_* angegeben.
- Bei Pluralia tantum wird die Pluralform notiert.
- Bei reziproken Pronomen (*drug\_drugâ*) wird die Genitivform angegeben: *\_0\_drug\_druga, \_adna\_adnoj\_0\_*.

### 3.5.4 Deklination

Die grammatischen Kategorien von Nomen werden nach folgendem Schema angegeben:

Affinität#Wortart|\_WeißrussischesLemma\_RussischesLemma\_&KASUS-  
GENUS:NUMERUS

Ambiguitäten, die nur paradigmatischer Natur sind, werden, wenn möglich, syntagmatisch aufgelöst, z.B. bei Kasushomonymien über die syntaktische Einbettung, beim Genus von adjektivischen Attributen im Plural über das Genus des Bezugselements. Nur wenn auch syntagmatisch keine Auflösung möglich ist, wird als Wert für die jeweilige Kategorie „ambig“ angegeben.

#### 3.5.4.1 Kasus

Die Angabe des Kasus erfolgt nach der Angabe des Lemmas getrennt durch den Marker ( & ).

Affinität#Wortart|\_WeißrussischesLemma\_RussischesLemma\_&KASUS-  
GENUS:NUMERUS



<i>Kürzel</i>	<i>Bedeutung</i>
NOM	Nominativ
GEN	Genitiv
DAT	Dativ
ACC	Akkusativ
INS	Instrumental
LOC	Lokativ/Präpositiv
VOC	Vokativ
CAMB	Kasus ambig, nicht zu entscheiden
0	nicht relevant

#### 3.5.4.2 Genus

Die Angabe des Genus erfolgt nach der Angabe des Kasus getrennt durch den Marker ( - ).

Affinität#Wortart|\_WeißrussischesLemma\_RussischesLemma\_&KASUS-  
**GENUS**:NUMERUS

<i>Kürzel</i>	<i>Bedeutung</i>
M	Maskulinum
F	Femininum
N	Neutrum
GC	Genus Commune
GAMB	Genus ambig, nicht zu entscheiden
0	nicht relevant

#### 3.5.4.3 Numerus

Die Angabe des Numerus erfolgt nach der Angabe des Genus getrennt durch den Marker ( : ).

Affinität#Wortart|\_WeißrussischesLemma\_RussischesLemma\_&KASUS-  
**GENUS**:**NUMERUS**

<i>Kürzel</i>	<i>Bedeutung</i>
SG	Singular
PL	Plural
0	nicht relevant

### 3.5.5 Konjugation

Die grammatischen Kategorien des Verbs werden nach folgenden Schemata angegeben:

```
Affinität#Wortart|_WeißrussischesLemma_Russisches-
Lemma_&TEMPUS-PERSON_oder_GENUS:NUMERUS
```

#### 3.5.5.1 Tempus und Imperativ

Die Angabe des Tempus erfolgt nach der Angabe des Lemmas getrennt durch den Marker ( & ).

```
Affinität#Wortart|_WeißrussischesLemma_Russisches-
Lemma_&TEMPUS-PERSON_oder_GENUS:NUMERUS
```

<i>Kürzel</i>	<i>Bedeutung</i>
IMPER	Imperativ
PRAES	Präsens
PRAET	Präteritum
INF	Infinitiv
0	Nicht relevant

#### 3.5.5.2 Person

Die Angabe der Person erfolgt nach der Angabe des Tempus getrennt durch den Marker ( - ).

```
Affinität#Wortart|_WeißrussischesLemma_Russisches-
Lemma_&TEMPUS-PERSON:NUMERUS
```

<i>Kürzel</i>	<i>Bedeutung</i>
1	1. Person
2	2. Person
3	3. Person
0	Nicht relevant

### 3.5.5.3 Genus

Die Angabe des Genus erfolgt nach der Angabe des Tempus getrennt durch den Marker ( - ).

Affinität#Wortart|\_WeißrussischesLemma\_Russisches-  
Lemma\_&TEMPUS-**GENUS**:NUMERUS

<i>Kürzel</i>	<i>Bedeutung</i>
M	Maskulinum
F	Femininum
N	Neutrum
GAMB	Genus ambig, nicht zu entscheiden
0	Nicht relevant

### 3.5.5.4 Numerus

Die Angabe des Numerus erfolgt nach der Angabe des Genus / der Person getrennt durch den Marker ( : ).

Affinität#Wortart|\_WeißrussischesLemma\_Russisches-  
Lemma\_&TEMPUS-PERSON\_oder\_GENUS:**NUMERUS**

<i>Kürzel</i>	<i>Bedeutung</i>
SG	Singular
PL	Plural
0	Nicht relevant

## 4 Arbeit mit dem OK-WRGR in CLAN

### 4.1 Einleitung

Das OK-WRGR ist mithilfe der im Rahmen des CHILDES-Projekts entwickelten Werkzeuge des Programmes CLAN analysierbar. Im Folgenden werden die nötigen Vorbereitungen dafür vorgestellt, sowie – anhand der Befehle `kwal` und `freq` – einige Beispiele der Analyse gezeigt. Dieser Abschnitt soll keine Einführung in CLAN sein bzw. diese ersetzen. Der Nutzer sei auf das CLAN-Manual (MCWHINNEY 2013b) verwiesen, das unter <http://childes.psy.cmu.edu/manuals/CLAN.pdf> verfügbar ist.

### 4.2 Vorbereitung des Korpus

*Installation von CLAN:*

- Installieren Sie CLAN. Die Installationsdatei und Anleitungen zur Installation können auf folgender Seite heruntergeladen werden:  
<http://childes.psy.cmu.edu/clan/>

*Vorbereitung des OK-WRGR:*

- Laden Sie von der Seite <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> die gewünschten Protokolle herunter (beispielsweise den Ordner `ok-wrgr_fam_alles.zip` für alle Gespräche des Familienkorpus). Entpacken Sie den / die heruntergeladenen Ordner an einen beliebigen Ort.
- Laden Sie von der Seite <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> die Datei `depfile.cut` herunter. Es handelt sich dabei um eine Steuerungsdatei, auf die das Programm CLAN zugreift. Hier sind die möglichen / nötigen Einträge in den einzelnen Zeilen der Transkripte festgelegt. Die im OK-WRGR herunterladbare Version ist eine Erweiterung der auf den CHILDES-Seiten herunterladbaren Version, in der die im OK-WRGR verwendeten Spezifika hinzugefügt sind. Dies sind folgende:
  - Zu den möglichen Einträgen bei **@Participants**:  
Village Migrant Town na Other

- Zu den möglichen Einträgen in den Äußerungszeilen (in der Datei *depfile.cut* ist dies die Zeile \*: )  
`[-] [_/] [_//] [_///] [_////] ["] ["-"] *@r *@h  
 *@pb *@pc *@ph *@dial *@pros *@w *@ls *-*`
- Zu den möglichen Einträgen bei **%mor**:  
`[-] [_/] [_//] [_///] ["] ["-"] :`
- Zudem sind zwei neue mögliche Zeilen definiert, und zwar  
`@MT of #: *` und `@FL of #: *`

Speichern Sie die Datei *depfile.cut* in demselben Ordner, in dem die entpackten Protokolle gespeichert sind.<sup>20</sup>

- Öffnen Sie die in dem Ordner ...\*CLAN\lib\fixes* (an dem Ort, an dem Sie CLAN installiert haben) befindliche Datei *ISO-639.cut* mit einem Texteditor oder mit CLAN. Die Datei *ISO-639.cut* enthält eine Liste mit Sprachkürzeln, auf welche das CLAN-Programm zurückgreift. Fügen Sie manuell folgende Zeilen hinzu (zwischen den Textteilen steht jeweils ein Tabstopp):

```
hb  hybe  HybBel
yr  hyru  HybRus
ru  rus   Rus
hy  hyb   Hyb
na  nan   NoAnswer
brms brms  BelarusianRussianMixedSpeech
```

---

<sup>20</sup> Zum Zeitpunkt der Veröffentlichung des OK-WRGR bestand in CLAN nicht mehr die Möglichkeit, in einer Datei *depadd.cut* zusätzliche Optionen zur Datei *depfile.cut* bereitzustellen. Daher müssen die notwendigen Änderungen in der Datei *depfile.cut* erfolgen.

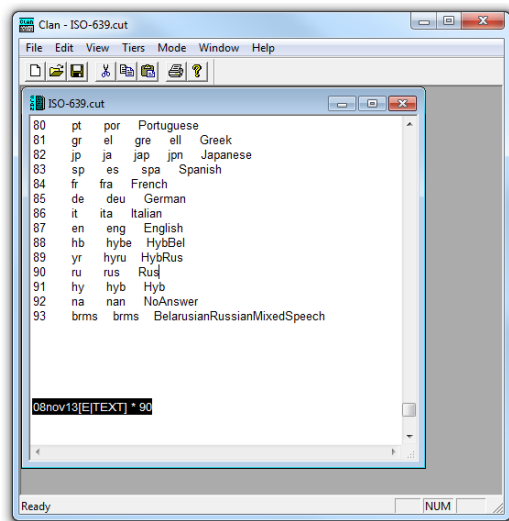


Abb. 4 Die Datei *ISO-639.cut*

Alternativ laden Sie von der Seite <http://diglib.bis.uni-oldenburg.de/bis-verlag/ok-wrgr/> die Datei *ISO-639.cut* herunter. Die dort herunterladbare Datei ist die bei der Installation von CLAN entstehende Datei, erweitert um die für das OK-WRGR verwendeten Sprachbezeichnungen. Ersetzen Sie die bereits bestehende Datei *ISO-639.cut* in dem Ordner `...\CLAN\lib\fixes` durch die von Ihnen heruntergeladene Datei *ISO-639.cut*.

- Nach dem Start von CLAN setzen Sie das Arbeitsverzeichnis (*working directory*) auf den Ordner mit den Gesprächsprotokollen.

### 4.3 Suche nach Wortformen

Da an die Transkription der Wortform Tags für die sprachliche Affinität folgen, muss in dem Suchbefehl mit dem Platzhalter ( \* ) gearbeitet werden. Der Befehl `freq +s"tak"` würde also keine Wortformen finden.

*Beispielsuche:*

```
freq fam_sm_9.cha +s"tak*"
```

*Suchergebnis:*

```
Speaker: *SFA:
```

```
2 tak@c@pc
```

```
1 takojê@c@pc
```

```
-----
```

```
2 Total number of different item types used
```

```
3 Total number of items (tokens)
```

```
0.667 Type/Token ratio
```

Das Zeichen ( @ ) lässt sich im Suchbefehl als Zeichen für das Ende der Wortform benutzen:

*Beispielsuche:*

```
freq fam_sm_9.cha +s"tak@"
```

*Suchergebnis:*

```
Speaker: *SFA:
```

```
2 tak@c@pc
```

```
-----
```

```
1 Total number of different item types used
```

```
2 Total number of items (tokens)
```

```
0.500 Type/Token ratio
```

#### 4.4 Suche nach Annotationen in %mor

Die in 3.5 vorgestellte Struktur der Annotationen in der Zeile %mor erlaubt es, bei der Suche nach grammatischen Werten oder Wortformen bestimmter sprachlicher Affinität den Suchassistenten von CLAN zu benutzen. So ist die Wortart über *part of speech*, das Lemma über *stem*, die sprachliche Affinität über *prefix*, Kasus und Tempus über *suffix*, Genus, Person und Numerus über *fusion* suchbar. Die folgende Abbildung zeigt die Eingabe für eine Suche nach Wortformen ‚russischer‘ Affinität mit ‚weißrussischer‘ Lautung:

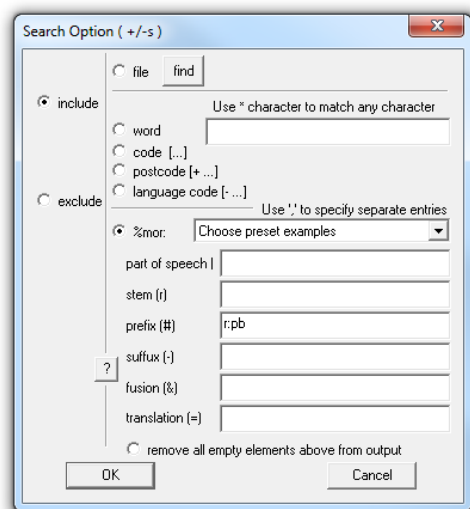


Abb. 5 Suche nach ‚russischen‘ Wortformen mit ‚weißrussischer‘ Lautung mithilfe des Suchassistenten von CLAN

Der entsprechende Suchbefehl im Befehlsfenster von CLAN ist (für den Befehl **freq**):

```
freq *.cha +s"@#-r:pb"
```

Diese Suchmöglichkeiten lassen sich auch kombinieren. Folgender Suchbefehl findet alle Adverbien ‚russischer‘ Affinität:

```
freq *.cha +s"@|-adv,#-r:*
```

#### 4.5 Suche nach Interpunktionszeichen

Nach Interpunktionszeichen kann ebenfalls gesucht werden. Folgender Befehl findet alle Wiederholungen (für den Befehl **kwal**):

```
kwal *.cha +s'[_/]'
```

Auch hier kann mit Platzhaltern gearbeitet werden. Folgender Befehl findet Wiederholungen, Korrekturen, Umformulierungen und Übersetzungen:

```
kwal @ +s'[_/*'
```



#### 4.6 Eingrenzung der Suche auf Teilkorpora / Städte / Rollen

Durch die Benennung der Dateien (vgl. 3.1) ist es einfach, die Suche auf Teile des Korpus zu beschränken. Folgender Befehl sucht nur in Gesprächen aus dem Interviewkorpus.

```
freq int*.cha +s"@|-adv,#-r:pc"
```

Folgender Befehl sucht nur in Gesprächen aus der Stadt *Akcjabrski*.

```
freq *ak*.cha +s"@|-adv,#-r:pc"
```

Auch die Beschränkung der Suche auf Sprecher einer „Rolle“ bzw. eines bestimmten Migrationsstatus ist möglich. Folgender Befehl sucht nur in Äußerungen von Land-Stadt-Migranten:

```
freq *.cha +s"@|-adv,#-r:pc" +t#Migrant
```

Für weitere Optionen sei der Nutzer auf das CLAN-Manual verwiesen (MCWHINNEY 2013b).

## 5 Literatur

- BRS 2003: *Belarуска-ruski sloŭnik (u 3 t.)*. Minsk
- CYCHUN, H. 2013 [1998]: *Trasjanka jak ab'ekt linhvistyčnaha dasledavannja*. In: CYCHUN, H. 2013: *Studien zur „Trasjanka“ – Študyi pra „Trasjanku“*. Herausgegeben von Gerd Hentschel. Oldenburg, 19-30 [= *Studia Slavica Oldenburgensia* 22]
- HENTSCHEL, G. & KITTEL, B. 2011a: Zur weißrussisch-russischen Zweisprachigkeit in Weißrussland – nicht zuletzt aus der Sicht der Weißrussen. In: Bohn, T. (Hrsg.), *Ein weißer Fleck in Europa. Die Imagination der Belarus als Kontaktzone zwischen Ost und West*. Bielefeld, 49-67
- HENTSCHEL, G. & KITTEL, B. 2011b: Weißrussische Dreisprachigkeit? Zur sprachlichen Situation in Weißrussland auf der Basis von Urteilen von Weißrussen über die Verbreitung „ihrer Sprachen“ im Lande. In: *Wiener Slavistischer Almanach* 67, 107-135
- HENTSCHEL, G. & ZELLER, J. P. 2013: Gemischte Rede, gemischter Diskurs, Sprechertypen: Weißrussisch, Russisch und gemischte Rede in der Kommunikation weißrussischer Familien. In: *Wiener Slavistischer Almanach* 70 (2012), 127-155
- HENTSCHEL, G. & ZELLER, J. P. 2014: Belarusians' pronunciation: Belarusian or Russian? Evidence from Belarusian-Russian mixed speech. In: *Russian Linguistics*. DOI: 10.1007/s11185-014-9126-1 [online unter: <http://link.springer.com/article/10.1007/s11185-014-9126-1>; erscheint im Druck in: *Russian Linguistics* 38:2]
- HENTSCHEL, G. (= XENČEL', G.) 2012: Nekotorye različija i schodstva v smešannoju belorussko-ruskoju reči v raznych belorusskich gorodach i u raznych govorjaščich. In: Zaprudski, S. & Cychun, G. (red.), *Novae slova ŭ belarusicysty. Movaznaŭstva*. Minsk, 220-232
- HENTSCHEL, G. 2008: Zur weißrussisch-russischen Hybridität in der weißrussischen „Trasjanka“. In: Kosta, P. & Weiss, D. (Hrsg.), *Slavistische Linguistik 2006/2007*. [= *Slavistische Beiträge* 464], 169-219
- HENTSCHEL, G. 2013 [= CHENTŠEL', G.]: Belorusskij, russkij i belorussko-russkaja smešannaja reč'. In: *Voprosy Jazykoznanija* (2013), 53-76
- HENTSCHEL, G. & ZELLER, J. P. (= CHENŠEL', G. / CELLER, JA. P.) 2011: Jakan'e, ekan'e, ikan'e v belorussko-ruskoju smešannoju reči: nabljudenija na osnove eksperimental'no-akustičeskogo analiza. In:

- Ševčenko, G. (red.), *Aktual'nye problemy filologii: antičnaja kul'tura i slavjanskij mir*. Minsk, 228-234
- KITTEL, B. & LINDNER, D. 2011: Der soziale Hintergrund von Sprachwahlen in Belarus. Eine sprachsoziologische Analyse der „gemischten Rede“. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 63, 623-647
- KITTEL, B., LINDNER, D., TESCH, S. & HENTSCHEL, G. 2010: Mixed Language Usage in Belarus. The Sociostructural Background of Language Choice. In: *International Journal of the Sociology of Language* 206, 47-71
- LISKOVEC, I.V. 2005: *Russkij i belorusskij jazyki v Minske: problemy bilingvizma i otnošenija k jazyku*. Dissertation, Europäische Universität St. Peterburg.
- MACWHINNEY, B. 2013a: *The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Part 1: The CHAT Transcription Format*. Carnegie Mellon University [<http://childes.psy.cmu.edu/manuals/chat.pdf>]
- MACWHINNEY, B. 2013b: *The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Part 2: The CLAN Programs*. Carnegie Mellon University [<http://childes.psy.cmu.edu/manuals/clan.pdf>]
- MEČKOVSKAJA, N.B. 2006: Belorusskaja trasjanka i ukrainskij suržik: surrogaty etničeskogo substandarta v ich otnošenijach k massovoj kul'ture i literaturnym jazykam. In: *Problemy zistavnoj semantyky* 7, 109-115
- MJAČKOŪSKAJA, N.B. 2007: Trasjanka ŭ kantynume Belaruska-Ruskich idyjalektaŭ: Chto I kali razmaŭljaje na Trasjancy? In: *Vesnik BDU. Ser. 4* (2007), 91-97
- RAMZA, T. 2008: Die Evolution der Trasjanka in literarischen Texten. In: *Zeitschrift für Slawistik* 53, 305–325
- RBS 2005: *Russko-belorusskij slovar' (v 3 t.)*. Minsk
- TSBM 1977-1984: *Tlumačal'ny sloŭnik belaruskaj movy (y 5 t.)*. Minsk
- ZELLER, J. P. (i.Vorb.): *Phonische Variation in weißrussisch-russischer gemischter Rede*.