

Larisa A. Tarasevič

Zum Problem der Vergleichbarkeit und Repräsentativität von sprachlichem Material

Am Beispiel von deutschen und russischen Präpositionen mit lokaler Bedeutung

1. Einleitung

Bei den sprachvergleichenden Untersuchungen ist aus methodologischen Gründen die Frage zu lösen, was und in welcher Menge zu vergleichen bzw. zu analysieren ist, d.h., dass die Vergleichbarkeit und Repräsentativität des zu analysierenden sprachlichen Materials eine Begründung erfordert. Besonders wichtig ist eine methodologisch fundierte Auswahl des sprachlichen Materials für die Forschungen, die den quantitativen Aspekt der zu vergleichenden sprachlichen Erscheinungen mitberücksichtigen. Dabei ist hier der Vergleich nicht als Ziel, sondern als eine Methode zu verstehen. Im Spiegel einer anderen Sprache abgebildet, kommen semantische und pragmatische Aspekte der sprachlichen Erscheinungen zum Vorschein, die im Rahmen einer Sprache unbemerkt bleiben (vgl. Benveniste 1974: 111 ff.). In der Kombination mit den quantitativen Methoden scheint die vergleichende Methode besonders ergiebig zu sein: Die neusten Untersuchungen auf dem Gebiet der Linguistik zeigen, dass qualitative und quantitative Charakteristika der Spracheinheiten gegenseitig bedingt sind (Komarova/Krajev 2008: 122).

Es ist nachgewiesen, dass quantitative Unterschiede in der Verwendbarkeit korrelierender Spracheinheiten mit den Unterschieden in ihrer semantischen Organisation zusammenhängen können (Tarasevich 2003).

Die Aussagekraft der Ergebnisse, die mit Hilfe von quantitativen und vergleichenden Methoden gewonnen werden, hängt aber von der Repräsentativität des Untersuchungsmaterials ab, d.h., dass die zu untersuchenden Korpora, denen die Belege entnommen werden, besonderen Anforderungen entsprechen sollen. In diesem Beitrag soll die eigene Vorgehensweise zur Erstellung der repräsentativen vergleichbaren Korpora, die als Quelle der Belege zum Vergleich der Semantik und Pragmatik der lokalen Präpositionen im Deutschen und Russischen dienen, dargestellt werden. Diese Vorgehensweise kann eine Anwendung in anderen quantitativ-qualitativen Forschungen finden und den Forschern viel Zeit und Arbeit ersparen. Im Folgenden werden die Hauptprinzipien, die an die Untersuchungskorpora in der Korpuslinguistik gestellt werden, diskutiert, dann sollen

eigene Erfahrungen der Erstellung der Untersuchungskorpora zu den Zielen des *Sprachvergleichs zusammengefasst*, sowie ihre *Repräsentativität nachgewiesen* werden. Zuletzt werden einige Daten zur Verwendbarkeit der ausgewählten deutschen und russischen lokalen Präpositionen, die anhand der erstellten Korpora gewonnen wurden, dargestellt.

2. Anforderungen an das Untersuchungskorpus beim Sprachvergleich

Die Anforderungen, die im Allgemeinen an ein Untersuchungskorpus¹ gestellt werden, sind in Baranov 2001 formuliert. Die wichtigsten sind *Repräsentativität, Vollständigkeit und Effizienz* (Baranov 2001: 118-119)².

Die Repräsentativität des Korpus stellt ein wichtiges Problem dar, das in der letzten Zeit hauptsächlich im Rahmen der angewandten Linguistik aktiv diskutiert wird: Die Repräsentativität, die infolge einer besonderen Prozedur der Textauswahl erreicht werden soll, ist eine konstituierende Eigenschaft des Korpus (Rykov 2002). Sehr allgemein wird unter der Repräsentativität die Entsprechung des Korpus zum Gebiet, das es zu vertreten hat, verstanden. Im Idealfall sollte das Korpus umfangreich sein (Jaskevič 2009). Baranov 2001 setzt bestimmte Einschränkungen des Umfangs des Textkorpus voraus: Nach Meinung des Autors soll ein Korpus alle Eigenschaften des Problemgebiets widerspiegeln, die für die gegebene Art der linguistischen Untersuchung relevant sind, und das in einer bestimmten Proportion, die durch die Frequenz der sprachlichen Erscheinung auf dem Problemgebiet bestimmt wird. Das Untersuchungskorpus kann mit anderen Worten abhängig vom Untersuchungsgegenstand proportional eingeschränkt werden (Baranov 2001: 118).

Nach der Anforderung der Effizienz (bei Baranov *экономичность*) soll das Untersuchungskorpus nicht nur eine genau festgelegte Menge an Texten des Problemgebiets beinhalten, sondern sich nach Möglichkeit im Umfang von ihm bedeutend unterscheiden. Die Erfüllung dieser Anforderung ermöglicht es, einerseits repräsentative Ergebnisse zu bekommen, andererseits den zeitlichen Aufwand des Forschers zu verringern.

Wenn es sich wie in meinem Fall um eine vergleichende Untersuchung handelt, entsteht die Notwendigkeit noch eine Anforderung anzuführen, die die Glaubwürdigkeit des Vergleichs sicherstellen soll: Da für eine vergleichende

¹ Die Anforderungen, die an andere Arten der Korpora, z.B. illustrative oder parallele Korpora sowie Monitoringskorpora usw., gestellt werden, können sich von den genannten unterscheiden.

² Neben diesen Anforderungen, die die inhaltliche Seite betreffen, werden auch die Notwendigkeit einer bestimmten Strukturierung des Materials und Computerunterstützung genannt, d.h. die technische Ausführung. Da ich in der gegebenen Untersuchung von den existierenden computergestützten Korpora Gebrauch mache, konzentriere ich mich weiterhin nur auf die Anforderungen an den Inhalt.

Untersuchung tatsächlich zwei Korpora, welche die Belege für die zu vergleichenden Spracheinheiten liefern, erstellt werden, formuliere ich hier die Anforderung der *Vergleichbarkeit* der Korpora. Unter der Vergleichbarkeit verstehe ich sowohl qualitative, als auch quantitative Charakteristika der Texte, die die Korpora in den zu vergleichenden Sprachen bilden, d.h. die Texte in beiden Sprachen sollen zum gleichen Genre und zur gleichen Zeitperiode gehören, sowie auch den ungefähr gleichen Umfang haben. Somit ist festzustellen, dass abhängig von den Zielen die Frage der Inhalte und des Umfanges der vergleichbaren Korpora³ zu lösen ist.

3. Das Problem der Erstellung vergleichbarer Korpora

Das Ziel des Projektes, an dem ich arbeite, ist ein Vergleich der Verwendungsweisen der lokalen Präpositionen im Deutschen und Russischen zur Feststellung ihrer semantischen und pragmatischen Besonderheiten. Vom definierten Ziel hängt die inhaltliche Füllung der zu erstellenden Korpora ab, d.h. die Texte der Korpora sollen thematisch maximal mannigfaltig sein, um ein möglichst breites Spektrum der Verwendungsweisen der lokalen Präpositionen widerzuspiegeln, und darüber hinaus thematisch übereinstimmen, um die Vergleichbarkeit zu gewährleisten. In dieser Hinsicht scheint das Textmaterial der Zeitungen optimal zu passen, das in den publizistischen Korpora vorhanden ist: Diese Korpora umfassen alle Zeitungsrubriken und beziehen sich auf eine breite Themenpalette – von gesellschaftlich-politischen, wirtschaftlichen und wissenschaftlich-populären bis zu Sport-, Kultur- und Alltagsthemen. Außerdem ist die Tatsache zu berücksichtigen, dass die Inhalte einer Zeitung von mehreren Autoren geschaffen werden und demzufolge subjektive Einflüsse des zu analysierenden Materials reduziert werden.

Unter den vorhandenen publizistischen Korpora für das Deutsche und Russische sind nach ihren technischen und inhaltlichen Daten die Texte des taz-Korpus (150 Millionen Textwörter, erstellt am Mannheimer Institut für Deutsche Sprache), sowie auch die Texte des publizistischen Teils des Korpus der russischen Sprache (über 20 Millionen Textwörter, erstellt am Moskauer Institut für russische Sprache bei der Russischen Akademie der Wissenschaften) am besten geeignet. Da beide Korpora in wissenschaftlichen Einrichtungen geschaffen wur-

³ Der Begriff „vergleichbare Korpora“ ist in der Korpuslinguistik kaum anzutreffen. Für die Zwecke des Sprachvergleichs werden in erster Linie Parallelkorpora genutzt, in denen Originaltexte und ihre Übersetzungen in eine oder mehrere Sprachen dargestellt sind. Solche Korpora bieten ein fruchtbares Feld für die Untersuchungen in den Bereichen Übersetzungstheorie, Lexikologie, Organisation des Diskurses usw. (Dobrovolskij 2008). Im Unterschied dazu kann man mit vergleichbaren Korpora die Frage erhellen, wie sich Sprachelemente mit gleicher Funktion in verschiedenen Sprachen verhalten.

den, ziehe ich ihre Vollständigkeit und Repräsentativität nicht in Zweifel.⁴ Diesen Korpora wurden die Teile entnommen, deren Texte zur gleichen Zeitperiode – Anfang der neunziger Jahre des vorigen Jahrhunderts – gehören und in etwa den gleichen Umfang (3 153 772 Textwörter für das Deutsche und 3 272 332 für das Russische)⁵ haben. Diese Teile bilden die vergleichbaren Untersuchungskorpora zur Erforschung präpositionaler Semantik und Pragmatik in meinem Projekt.

Während die thematische und zeitbezogene Vergleichbarkeit der Untersuchungskorpora erzielt ist, soll an dieser Stelle ein anderer wichtiger Aspekt, der die Repräsentativität der Korpora beeinflussen kann, überprüft werden – ihr Umfang. In der Literatur finden sich ganz unterschiedliche Angaben zum Mindestumfang des sprachlichen Materials zur Untersuchung der präpositionalen Verwendungen. So begründen Komarova/Krajev (2008: 64) in ihrer quantitativ-qualitativen Untersuchung der Präpositionen in der Fachsprache Informatik den ausreichenden Umfang von hunderttausend Textwörtern. Das Korpus von Di Meola 2001 zur Untersuchung der Grammatikalisierungsprozesse der deutschen Präpositionen beträgt 5 Millionen Textwörter. Ein solcher Umfangsunterschied kann durch die Besonderheiten des Untersuchungsgegenstandes erklärt werden: im ersten Fall sind das sowohl primäre, als auch sekundäre Präpositionen in allen ihren Bedeutungen, verwendet in einem einzigen thematischen Bereich. Dementsprechend ist hier mit einer größeren Verwendbarkeit des Untersuchungsgegenstandes zu rechnen; im zweiten sind das in erster Linie sekundäre Präpositionen und präpositionale Strukturen, die eine Tendenz zur Präpositionalisierung zeigen, also ist eine kleinere Verwendbarkeit anzunehmen. Bei der Einschränkung des Untersuchungsmaterials wird eher intuitiv vorgegangen, während eine auf den mathematischen Methoden basierende Überprüfung der Repräsentativität des Korpus hinsichtlich des Umfangs in einem Experiment zweckmäßig erscheint. Die Organisation und die Resultate eines solchen Experimentes werden unten dargestellt.

⁴ Die Prinzipien der Erstellung des russischen publizistischen Korpus sind ausführlich in Baranov 2001, des deutschen auf der Internetseite des Mannheimer Instituts für Deutsche Sprache beschrieben.

⁵ Ein kleiner Unterschied von ca. 4% ist in unserem Fall unwichtig, da der Untersuchungsgegenstand solche höchst frequentative Spracheinheiten wie Präpositionen bilden.

4. Das Experiment zur Überprüfung der Repräsentativität des Korpus

4.1 Überprüfung der Repräsentativität der Korpora mit Hilfe der Methode der Speicherung und Korrektur der relativen Häufigkeit

Im ersten Schritt wurde aus meinen oben beschriebenen Untersuchungskorpora mit Hilfe der Methode der lückenlosen Auswahl für 47 deutsche und 44 russische Präpositionen, die laut den akademischen Grammatiken über eine lokale Bedeutung verfügen, eine Menge von Kontexten gewonnen, von denen im zweiten Schritt die nichträumlichen (temporalen, kausalen, komitativen u.a.) Verwendungen der polysemen Präpositionen ausgeschlossen wurden. Die Gesamtzahl der gewonnenen Belege betrug für das Deutsche über 80 000, für das Russische über 60 000, was an sich eine sehr große Menge zur Durchführung einer qualitativen Analyse darstellt, welche die Klassifizierung und den Vergleich der mit den lokalen Präpositionen verwendeten Bezeichnungen von räumlichen Objekten im Deutschen und Russischen zum Ziel hat. Besonders hohe Zahlen der Belege ergeben sich für solche korrelierende Präpositionen wie *in* und *в* (über 46 000 und 35 000 Belege entsprechend), *an*, *auf* und *на* (ca. 8000, 8500 und 10 500 Belege) usw. Es ist anzunehmen, dass diese Mengen nicht nur ausreichend, sondern überflüssig sind, was der Forderung nach Effizienz (s.o.) widerspricht. Zur Bestimmung vernünftiger Grenzen für die Arbeit mit frequentativen Präpositionen und zur Überprüfung der Repräsentativität der Korpora für die Präpositionen mit einer geringeren Verwendbarkeit eignet sich die Methode, die in Baranov 2001a entwickelt wurde.

Die genannte Methode beruht auf der Idee der Speicherung und Korrektur der relativen Häufigkeit bei der lückenlosen Auswahl der Belege. Es wird angenommen, dass der Gegenstand der Untersuchung (beim Autor sind das metaphorische Modelle) im Problemgebiet, d.h. in der Gesamtmenge der realen sprachlichen Verwendungen, eine bestimmte (reale) Häufigkeit F_m aufweist, die unbekannt ist. Im Untersuchungskorpus kann sich die Häufigkeit des Untersuchungsgegenstandes der realen Häufigkeit annähern oder sich von dieser bedeutend unterscheiden. Inwiefern die Häufigkeit im erstellten Korpus mit der Häufigkeit im Problemgebiet im realen Sprachgebrauch übereinstimmt, wird auf folgende Weise überprüft: Die Gesamtmenge der gewonnenen Belege wird in einzelne Segmente (genaue Untermengen) unterteilt. In einem frei gewählten Segment M_1 ⁶ erweist der Untersuchungsgegenstand eine Häufigkeit F_1 , die sich höchstwahrscheinlich von der Häufigkeit im Problemgebiet unterscheidet. Wenn das Untersuchungskorpus repräsentativ ist, wird sich bei der weiteren Speicherung des Untersuchungsmaterials die relative Häufigkeit des Untersuchungsgegenstandes in den Segmenten $M_1+M_2+\dots+M_n$ immer mehr an den Wert F_m

⁶ In Baranov 2001a werden die Segmente mit A (entsprechend A1, A2, An) bezeichnet. Die Veränderung der Bezeichnung ist aus Übersetzungsgründen notwendig.

annähern. Diese Tendenz läßt sich auch für jedes weitere Segment zeigen. Auch der Unterschied zwischen den Werten der relativen Häufigkeit in den Nachbarsegmenten, der als *Abweichungsparameter* bezeichnet wird, soll ab einem bestimmten Segment geringer werden. Dieser Parameter ermöglicht laut Autor die Bewertung der Repräsentativität des Korpus. Das zeigt sich graphisch in der Tendenz der Kurve der relativen Häufigkeitswerte in den Segmenten auf der Horizontalen, was davon zeugt, dass mindestens eine relative Repräsentativität des Korpus erreicht wurde.

Nach der oben beschriebenen Methode habe ich das vorhandene Untersuchungsmaterial mit den durch den Gegenstand und Aufgaben meiner Untersuchung bedingten Korrekturen bewertet. Da es sich um die Klassifizierung und den Vergleich der mit den lokalen Präpositionen verwendeten Bezeichnungen von räumlichen Objekten handelt, wurde die Dynamik relativer Häufigkeitswerte für einzelne Klassen der Objektbezeichnungen in den Segmenten für die jeweilige Präposition verfolgt. Die Bewertung der Repräsentativität der vergleichbaren Korpora wurde am Beispiel der deutschen Präpositionen *auf* und *aus*, sowie der russischen *на* und *но* durchgeführt. Die deutsche Präposition *auf* und die russische *на* gelten als Korrelate – mit ihnen wird Bezug auf die Oberfläche des Objektes genommen (z.B. *auf dem Tisch/на столе*, *auf dem Fußboden/на полу*); mit *но* wird auch in Bezug auf eine Oberfläche lokalisiert mit dem Unterschied, dass diese Präposition eine kurzzeitige (auch mehrmalige) Berührung der Oberfläche des Bezugsobjektes bezeichnet (z.B. *идти по улице/дורך die Straße gehen*, *ударить по столу/ауф den Tisch schlagen*); *но* hat im Deutschen keine präpositionale Entsprechung. Die Präposition *aus* bezeichnet die Bewegung aus dem Innenraum und ist aus der Sicht der Opposition „Inklusion – Exklusion“ den Präpositionen *auf*, *на* und *но* gegenübergestellt.

Aus der Gesamtmenge der Belege für die genannten Präpositionen wurden je 1500 nach einer bestimmten Regel genommen: Die Gesamtmengen der Belege für *auf*, *aus* und *на* (ca. 8500, 6300 und 10 500 entsprechend) wurden in 15 gleiche Teile geteilt, von denen die ersten 100 Belege genommen wurden. Für *но* sind fast alle Belege zum Experiment herangezogen, weil ihre Gesamtzahl (ca. 1900) die festgelegte Menge von 1500 geringfügig übertrifft. Die auf solche Weise gewählten Belege wurden weiterhin in 15 Segmente je 100 Belege unterteilt.

Die Belege in jedem Segment habe ich in Entsprechung zur einheitlichen Klassifikation der Bezeichnungen räumlicher Objekte nach ihren ontologischen Eigenschaften aussortiert. Diese Klassifikation, die eine Grundlage für den Vergleich der räumlichen Verwendungsweisen der Präpositionen in beiden Sprachen bildet, umfasst acht Gruppen der Objektbezeichnungen.⁷ Die neunte Gruppe bil-

⁷ Hier werden die Prinzipien und Grundlagen der Unterteilung räumlicher Objekte in Klassen, sowie das Problem der Abgrenzung der Abhängigkeit der Präposition vom Nomen vs. Verb nicht weiter diskutiert, weil das den Rahmen des gegebenen Aufsatzes sprengen würde. Wichtig für das Experiment sind einheitliche Kriterien der Bearbeitung des sprachlichen Materials für beide Sprachen.

den Belege mit der semantisch „durchsichtigen“ verbalen Rektion, wie z.B. dt. *schießen auf*, russ. *стрелять по/в*, usw.

Die Liste der neun ausdifferenzierten Gruppen sieht wie folgt aus:

- 1) „Mensch“ – Bezeichnungen der Menschen (z.B. Personalpronomen, Nomina, die auf Geschlecht oder Alter hinweisen, Namen, Berufsbezeichnungen usw.), Tiere, sowie auch der Körperteile.
- 2) „Geopolitische Objekte“ – Bezeichnungen der Länder, Städte, Siedlungen (Nomina wie *Dorf, Ortschaft, Lager, Friedhof*), usw.
- 3) „Gebäude“ – Bezeichnungen der Gebäude und Gebäudeteile, sowie institutioneller Einrichtungen, wenn es sich um ein Gebäude handelt (z.B. *Haus, Villa, Schule, Tür, Bühne* usw.).
- 4) „Transportmittel“ – Bezeichnungen von Transportmitteln wie Lkws, Pkws, Fahrräder, Züge, Schiffe, Flugzeuge usw.
- 5) „Einrichtungsgegenstände“ – Bezeichnungen von Möbelstücken, Geräten, Geschirr, Kleidung usw.
- 6) „Naturobjekte“ – Bezeichnungen der Ozeane, Seen, Kontinente, Berge, Flüsse, Wälder, Pflanzen, Stoffe, die in der Natur anzutreffen sind usw.
- 7) „Wege“ – Bezeichnungen von Straßen, Plätzen, Höfen, Brücken, sowie auch Nomina wie *Weg, Pfad* usw.
- 8) „Klassifikatoren“ – Nomina, die keine Objekte, sondern ihre räumlichen Eigenschaften bezeichnen wie z.B. *Seite, Ende, Teil, Linie, Stück, Oberfläche, Höhe, Tiefe, Platz, Stelle, Punkt, Ecke, Rand* usw.
- 9) Belege mit der verbalen Rektion.

Auf diese Weise wurden für jede Präposition 15 Werte zu jeder Gruppe gewonnen, die die absolute Häufigkeit der Verwendungen jeweiliger Präpositionen mit der gegebenen Gruppe der Bezeichnungen von Objekten in den Segmenten darstellen.

Nach der Berechnung der relativen Häufigkeitswerte für jedes Segment, sowie auch des Abweichungsparameters wurden insgesamt 34 Graphiken erstellt, die die Dynamik der Veränderung dieses Parameters von Segment zu Segment widerspiegeln: je neun für *auf* und *na* und je acht für *aus* und *no* (für *aus* finden sich im Korpus keine Belege der verbalen Rektion, und für *no* keine Belege für die Objektgruppe „Transportmittel“). Als Beispiel werden unten die Tabelle der Berechnungen und die entsprechende Graphik für die Verwendungen der korrelierenden Präpositionen *auf* und *na* mit der Gruppe „Klassifikatoren“ angeführt (Abb. 1).

Segmente	M1		M1+M2			M1+M2+M3			M1+...M4		
	AF	F1	AF	F2	Q1 (F2-F1)	AF	F3	Q2 (F3-F2)	AF	F4	Q3 (F4-F3)
<i>auf</i> + Klassifikator	25	0,25	41	0,205	-0,045	63	0,21	0,005	81	0,203	-0,008
<i>na</i> + Klassifikator	11	0,11	44	0,22	0,11	55	0,183	-0,037	59	0,148	-0,036

Segmente	M1+...M5			M1+...M6			M1+...M7			M1+...M8		
	AF	F5	Q4 (F5-F4)	AF	F6	Q5 (F6-F5)	AF	F7	Q6 (F7-F6)	AF	F8	Q7 (F8-F7)
<i>auf</i> + Klassifikator	96	0,192	-0,0105	111	0,185	-0,007	128	0,183	-0,0021	142	0,176	-0,005
<i>na</i> + Klassifikator	72	0,144	-0,004	96	0,16	0,016	125	0,179	0,019	155	0,194	0,015

Segmente	M1+...M9			M1+...M10			M1+...M11			M1+...M12		
	AF	F9	Q8 (F9-F8)	AF	F10	Q9 (F10-F9)	AF	F11	Q10 (F11-F10)	AF	F12	Q11 (F12-F11)
<i>auf</i> + Klassifikator	164	0,182	0,005	187	0,187	0,005	218	0,198	0,011	238	0,198	0,0002
<i>na</i> + Klassifikator	194	0,216	0,022	224	0,224	0,008	247	0,225	0,001	266	0,222	-0,003

Segmente	M1+...M13			M1+...M14			M1+...M15		
	AF	F13	Q12 (F13-F12)	AF	F14	Q13 (F14-F13)	AF	F15	Q14 (F15-F14)
<i>auf</i> + Klassifikator	250	0,192	-0,006	257	0,184	-0,009	270	0,18	-0,004
<i>na</i> + Klassifikator	286	0,22	-0,002	305	0,218	-0,002	313	0,209	-0,009

Tabelle 1: Berechnung des Abweichungsparameters für die Verwendungen von *auf* und *na* mit der Gruppe „Klassifikatoren“⁸

⁸ Legende: AF – absolute Häufigkeit; F – relative Häufigkeit; Q – Abweichungsparameter.

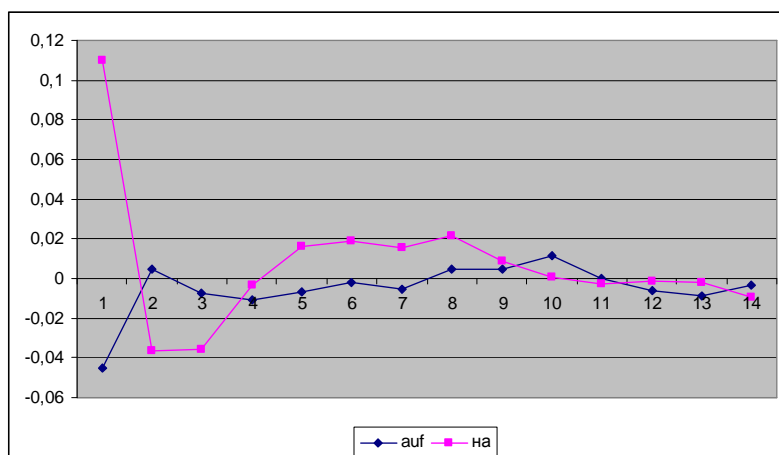


Abb. 1: Dynamik der Veränderung von Abweichungsparametern für *auf* und *на* mit der Objektgruppe „Klassifikatoren“.

In der gegebenen Grafik stabilisiert sich z.B. die Kurve für *auf* ab dem zweiten, für *на* ab dem vierten Segment. Die Analyse der Ergebnisse zeigt, dass die Stabilisierung der Kurve in den meisten Fällen zwischen dem vierten, fünften und zehnten Segment erfolgt, d.h. im Rahmen von 500-1000 Belegen (s. als Beispiel Tabelle 2, in der Segmente für die Objektgruppen angegeben sind, ab denen eine Stabilisierung des Abweichungsparameters festzustellen ist). Somit erweisen sich die vergleichbaren Korpora mindestens für die Präpositionen als repräsentativ, für die sich in den Korpora 500 und mehr räumliche Verwendungen finden.

Objektgruppe	Segmentnummer			
	auf	aus	на	по
Mensch	6	5	4	4
Geopolitische Objekte	9	2	3	3
Gebäude	2	6	4	3
Transportmittel	6	4	5	-
Einrichtungsgegenstände	4	6	3	7
Naturobjekte	3	9	3	3
Wege	3	8	3	5
Klassifikatoren	2	8	6	2
Verbale Rektion	2	-	5	5

Tabelle 2: Die Anfangsgrenzen der Stabilisierung der Kurve für die Präpositionen nach den Objektgruppen.

Aus dem russischen Korpus habe ich 500 und mehr Belege für die Verwendung in räumlicher Bedeutung solcher Präpositionen wie *в*, *за*, *из*, *к*, *на*, *от*, *по*,

nod, c/co, y (10 von 44) gewonnen. Im deutschen Korpus finden sich nicht weniger als 500 Belege für die räumlichen Verwendungen der Präpositionen *an, auf, aus, bei, bis, durch, hinter, in, nach, neben, über, um, von, vor, zu, zwischen* (16 von 47).

Die Grenze von 500 Belegen kann für bestimmte Gruppen der lokalen Präpositionen unterschritten werden: Die Motiviertheit der inneren Form und eine größere Zahl der eindeutig markierten semantischen Merkmale in der Bedeutung der Präposition beschränken ihre Verwendbarkeit für die Bezeichnungen bestimmter Objektgruppen (Tarasevič 2003). Dementsprechend nehme ich für solche Präpositionen einen geringeren Umfang der repräsentativen Menge der zu analysierenden Belege an. Diese Annahme wird in weiteren Forschungen zur semantischen Organisation der lokalen Präpositionen überprüft.

4.2 Die Überprüfung der Beschaffenheit von Objektgruppen

Es ist anzunehmen, dass abhängig von den im gegebenen Segment überwiegender Themen die gegebene Präposition eine höhere Menge an Belegen für die themenbedingte aktuelle Objektgruppe aufweist (z.B. Einrichtungsgegenstände oder Gebäude, wenn es sich um den Baubereich handelt oder geopolitische Objekte in den sozial-politischen Themen). In einem repräsentativen Korpus soll dieser themenbedingte Einfluss ausgeglichen werden⁹: Bestimmte Schwankungen sind durchaus denkbar, doch wenn die Zahl der Belege für die Verwendungen ein- und derselben Präposition mit den Bezeichnungen der Objekte ein- und derselben Gruppe je nach Segment Schwankungen von Null bis zu den Höchstwerten aufweist, hätte dies zu bedeuten, dass die Texte der vergleichbaren Korpora thematisch nicht ausgewogen sind. Das könnte die Vergleichbarkeit der Korpora in Frage stellen und die Glaubwürdigkeit der Resultate gefährden.

Aus diesen Gründen finde ich die Überprüfung der Repräsentativität aus Sicht der *Beschaffenheit der Objektgruppen* notwendig. Unter der Beschaffenheit der Objektgruppen verstehe ich die Besonderheiten ihrer Füllung mit den Belegen – in verschiedenen Segmenten können verschiedene Gruppen von Objekten vorkommen; die Häufigkeitswerte der Objektgruppen können von Segment zu Segment wesentliche Unterschiede aufweisen. Dementsprechend wurde im Experiment erstens untersucht, ab welchem Segment die maximale Zahl der Objektgruppen erreicht wird, und zweitens, wie hoch die Schwankungen der relativen Häufigkeit der Objektgruppen für eine Präposition sind.

Die Analyse der vorhandenen Daten hat ergeben, dass sich spätestens ab dem dritten Segment Belege für die maximale Zahl der Objektgruppen nachweisen

⁹ Darauf weist indirekt die Untersuchung von Arapov 1988: 24-25 hin, in der nachgewiesen wurde, dass unter der Bedingung der Repräsentativität der Häufigkeitsrang der Wörter (ein hoher oder ein niedriger) in verschiedenen Textmengen erhalten bleibt. Ich nehme diese Tendenz auch für die Verwendbarkeit der lokalen Präpositionen mit den Bezeichnungen verschiedener Objektgruppen an.

lassen. In der Tabelle 3 sind als Beispiel die gewonnenen Mengen der Belege für *aus* in absoluten Zahlen nach den Objektgruppen angegeben.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Gesamt
Geopolitische Objekte	34	49	49	49	34	36	49	46	56	55	65	54	58	48	68	750
Gebäude	40	31	35	40	35	37	24	21	15	14	7	17	14	14	9	353
Klassifikatoren	4	11	8	5	10	6	11	21	12	18	20	14	15	16	8	179
Einrichtungsgg.	9	1	2	2	7	9	6	2	7	4	1	5	3	9	6	73
Naturobjekte	3	3	5	2	5	7	4	1	2	7	6	7	3	4	5	64
Transportmittel	4	2	0	0	5	0	1	3	6	0	1	2	4	2	2	32
Mensch	4	2	0	0	1	4	3	1	2	2	0	1	2	5	2	29
Wege	2	1	1	2	3	1	2	5	0	0	0	0	1	2	0	20

Tabelle 3: Die Beschaffenheit der Objektgruppen für *aus*.

Aus den Graphiken, die die Dynamik der relativen Häufigkeit der Belege für die Objektgruppen nach den Segmenten abbilden, wird ersichtlich, dass sich für jede Präposition Objektgruppen ergeben, deren Häufigkeitswerte im hohen, mittleren oder unteren Bereich liegen. Z.B. wird durch hohe Häufigkeitswerte die Verwendbarkeit von *aus* mit den Bezeichnungen der geopolitischen Objekte charakterisiert: die relative Häufigkeit für diese Gruppe schwankt je nach Segment von 0,34 bis 0,68 ($0,34 < F < 0,68$); weniger frequentativ sind die Verwendungen dieser Präposition mit den Nomina, die die Gruppe „Klassifikatoren“ bilden ($0,04 < F < 0,2$); eine niedrige Häufigkeit erweist sich für die Verwendbarkeit von *aus* mit den Bezeichnungen der Gruppe „Wege“ ($0 < F < 0,05$). Die Verteilung der relativen Häufigkeitswerte nach Segmenten für jede genannte Objektgruppe ist in der Graphik (Abb. 2) abgebildet.

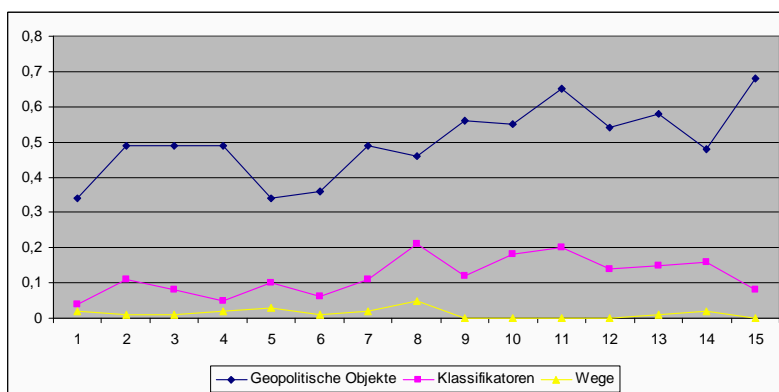


Abb. 2: Verteilung der relativen Häufigkeitswerte der Verwendungen von *aus* mit den Objektgruppen „Geopolitische Objekte“, „Klassifikatoren“, „Wege“ nach Segmenten.

Wie die Graphik zeigt, sind die Schwankungen der Werte der relativen Häufigkeit in den Segmenten relativ stabil und bewegen sich in ziemlich festen Grenzen. Extreme Häufigkeitsschwankungen (vom oberen Bereich bis Null) sind nicht nachgewiesen worden. Die Daten bestätigen noch einmal die Repräsentativität der Korpora und lassen auf bestimmte Gesetzmäßigkeiten schließen, die der Verwendbarkeit der lokalen Präpositionen mit den Bezeichnungen der Objektgruppen sowohl im Deutschen, als auch im Russischen zugrunde liegen.

Ein kurzer Einblick in diese Gesetzmäßigkeiten öffnet sich mit dem Vergleich der Verwendbarkeit von Bezeichnungen der Objektgruppen mit den für das Experiment gewählten Präpositionen (Abb. 3): schon in der ersten Annäherung zeigt sich eine deutliche „Spezialisierung“ der Präpositionen auf bestimmte Objektgruppen. So haben die Verwendungen der Bezeichnungen von geopolitischen Objekten mit *no* und *aus* hohe Häufigkeitswerte, während die Verwendungen mit *auf* und *na* niedrige aufweisen. Stabil hohe Werte kennzeichnen die Verwendungen der Präpositionen mit den Klassifikatoren, was durch die Notwendigkeit der Einschränkung des Lokalisationsraums erklärt werden kann (z.B. *aus der Stadt – aus einem Teil der Stadt, auf dem Tisch – am Rande des Tisches*).¹⁰ Menschen (bzw. Tiere, Körperteile) und Transportmittel werden dagegen seltener als Bezugsobjekte zur Orientierung im Raum gewählt.

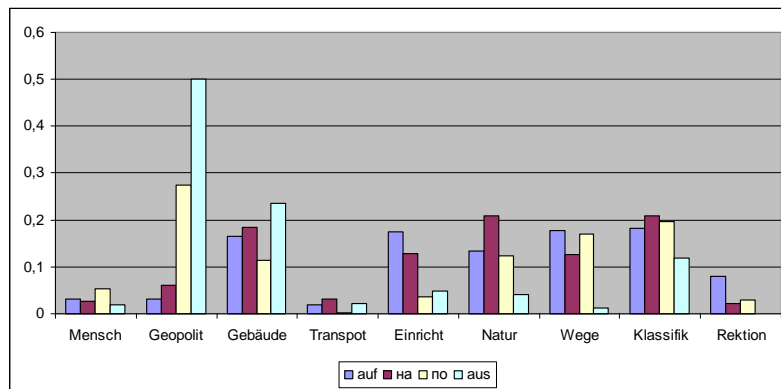


Abb. 3: Relative Häufigkeit der Verwendungen von *no*, *na*, *aus* und *auf* mit den Objektgruppen im Vergleich.

Die gegebene Graphik bestätigt den Zusammenhang der präpositionalen Semantik mit den Prozessen sprachlicher Kategorisierung. Verschiedene Autoren bemerken, dass die Verwendung einer bestimmten Präposition die Erwartungen einer bestimmten Kategorie der Objekte hervorruft, wie z.B. die Verwendung von *in*

¹⁰ Es sei an dieser Stelle auf dieses Problem nicht näher eingegangen. Welche Arten der Klassifikatoren, abhängig von den ihnen beschriebenen Raumeigenschaften, unterschieden werden können, sowie die Frage, wie sie mit verschiedenen Präpositionen verwendet werden, sollen den Gegenstand künftiger Untersuchungen bilden.

ein Behälter als Bezugsobjekt erwarten lässt (z.B. *im Haus*), und die Verwendung von *neben* auf die Dimensionalität des Objektes aufmerksam macht (*neben dem Haus*) (Wunderlich 1985: 74, Schwarze 1989: 316). Diese Eigenschaft der präpositionalen Semantik, die aktuelle sprachliche Kategorisierung des Objektes anzukündigen, wird als kategorisierende Leistung bezeichnet (Grießhaber 1999: 91). Üblicherweise stimmt die durch die Präposition induzierte Kategorisierung mit einer der „typischen“ Kategorisierungen des Objektes überein. So wird *aus* am häufigsten mit den Bezeichnungen der „typischen“ Behälter wie Länder und Städte (Lebensräume des Menschen), Gebäude und Einrichtungsgegenstände wie Schränke, Tische usw. verwendet, *no* kommt mit den Bezeichnungen der „typischen“ Wege oder Routen vor, zu denen wiederum Länder und Städte (vgl. *no странам и континентам, no городам и весям*), geographische Objekte (Wälder, Felder, Berge usw.) gehören. Die Fälle der „untypischen“ Verwendbarkeit bedürfen einer tieferen Forschung. Ihre Ursachen können sowohl in einer spezifischen sprachlichen Kategorisierung der Bezeichnung liegen (vgl. *auf dem Lande/на селе*), als auch durch das eigene Kategorisierungspotenzial der Präposition bedingt sein.

5. Resümee

Mit dem durchgeführten Experiment ist mindestens die relative Repräsentativität der vergleichbaren Korpora nachgewiesen worden. Es hat sich herausgestellt, dass sich die Mindestmenge der zu analysierenden Belege, die nach einem bestimmten Verfahren (in meinem Fall ist das die lückenlose Auswahl) gewonnen werden, berechnen lässt. Der mit mathematischen Methoden fundierte Umfang des Untersuchungsmaterials steigert sowohl die Effizienz, als auch die Aussagekraft der Untersuchung (vgl. Baranov 2001a). Die Arbeit mit den vergleichbaren Korpora hat ihre Ergiebigkeit für die sprachvergleichenden Forschungen gezeigt. Die aus den Korpora gewonnenen Daten ermöglichen es, ein komplexes Bild von den Besonderheiten des Funktionierens der Systeme lokaler Präpositionen im Deutschen und Russischen zu bekommen, z.B. welchen Rang der Verwendbarkeit in beiden Sprachen die korrelierenden Präpositionen in der lokalen Bedeutung haben, in welchem Verhältnis dieser Rang zum Rang in der Gesamtmenge ihrer Bedeutungen steht, mit Bezeichnungen welcher Objektklassen korrelierende Präpositionen angewandt werden usw. Der Behandlung dieser Probleme widmen sich meine weiteren Forschungen zur präpositionalen Semantik und Pragmatik.

Literatur

Arapov, Michail V. (1988): *Quantitativnaja lingvistika*. Moskva.

- Baranov, Anatolij N. (2001): *Vvedenije v prikladnuju lingvistiku*. Moskva.
- Baranov, Anatolij N. (2001a): Problema representativnosti korpusa dannyh (na primere političeskoj metaforiki). <http://www.dialog-21.ru/materials/archive.asp?id=6812&y=2001&vol=6078> [zuletzt eingesehen am 15.09.2010].
- Benveniste, Émile (1974): *Obščaja lingvistika*. Moskva.
- DiMeola, Claudio (2000): *Die Grammatikalisierung deutscher Präpositionen. (Studien zur deutschen Grammatik: Bd.62)*. Tübingen.
- Dobrovol'skij, Dmitrij O. (2008): Korpus parallel'nych textov v issledovanii kul'turno-spezifičeskoj lexiki. <http://ruscorpora.ru/sbornik2008/20.pdf> [zuletzt eingesehen am 3.01.2011].
- Freitag, York, Vandermeeren, Sonja (2005): Deutsche Präpositionen: eine fehleranalytische Untersuchung. In: *Das Wort, Germanistisches Jahrbuch Russland 2005*; 155-182.
- Griebhaber, Wilhelm (1999): *Die relationierende Prozedur*. Münster u.a.
- Jaskevič, Aleksandr A. (2009): Korpusnaja lingvistika. <http://www.philology.by/page/corpling> [zuletzt eingesehen am 6.10.2010].
- Komarova, Soja I., Krajev, Semën V. (2008): *Jadernyje služebnyje slova v ruskom podjasyke informatiki: kvantitativno-kvalitativnoje issledovanije*. Jekaterinburg.
- Rykov, Vladimir V. Korpus tekstov kak realizazyja objektno-orientirovannoj paradigmy. <http://www.dialog-21.ru/materials/archive.asp?id=7373&y=2002&vol=6077> [zuletzt eingesehen am 6. 10. 2010].
- Schwarze, Christoph (1989): Polysemie als Prozedur, am Beispiel von frz. à travers und chez. In: Habel, Christopher / Herweg, Michael / Rehkaemper, Klaus (Hrsg.) (1989). *Raumkonzepte in Verstehensprozessen. Interdisziplinäre Beiträge zu Sprache und Raum*. Tübingen. 310–338.
- Tarasevič, Larisa (2003): *Dimensionale Präpositionen. Eine kontrastiv-semantische Untersuchung von vor und neped*. Münster u.a.
- Tarasevič, Larisa A. (2005): Wenn vor hinter bedeutet. Die Qual der Wahl der passenden Präposition zur Versprachlichung räumlicher Konstellationen, in: *Das Wort. Germanistisches Jahrbuch Russland 2005*; 277-295.
- Wunderlich, Dieter (1985): Raum, Zeit und das Lexikon. In: Schweizer, Harro / Ehlich, Konrad (Hrsg.) (1985). *Sprache und Raum. Psychologische und linguistische Aspekte der Aneignung und Verarbeitung von Räumlichkeit; ein Arbeitsbuch für das Lehren von Forschung*. Stuttgart. 66–89.